Exploring Kinetic Controlled Protein Solubility under Physiologically Relevant Conditions

by

Hung (Justin) Dang

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(CHEMISTRY)

at the

UNIVERSITY OF WISCONSIN-MADISON

2024

Date of Final Oral Examination: 09/07/2023

The dissertation is approved by the following members of the Final Oral Committee:

Silvia Cavagnero, Professor, Chemistry

Tina Wang, Professor, Chemistry

Judith Burstyn, Professor, Chemistry

Ivan Rayment, Professor, Biochemistry

I dedicate this Thesis to those who think they aren't good enough in life…

You are. Love & believe in yourself. We are here to make memories and mistakes.

Only then, we can learn to be better versions of ourselves, for ourselves and those who are important to you.

Table of contents

Acknowledgements

Though I have many achievements in my life, none of which are accomplished without the tremendous support and kindness from many important people in my life. In my life and in my career. I dedicated this section to all those who have an impact.

To my PI and Committee members: I would first like to thank my doctoral advisor, Prof. Silvia Cavagnero, for the academic support throughout my 4 years of graduate studies. I would also like to thank my thesis committee members: Prof. Tina Wang, Prof. Ivan Rayment, and Prof. Judith Burstyn. Your contributions, advice, and support have helped me shape my graduate school advancements throughout the years.

To my Cavagnero family: I cannot have gotten through my graduate school career without my lab members: Anna Allen, Andrew Fuchs, Matthew Dalphin, Clayton Thompson, Shaneen Britton, Heather Allaman, Ummay Shapla Mahfuza, Natalie Feider, Wanting Wei, Yuzhu Liu, Mridula Srivathsan, and Lucas Bartel. Collectively, thank you for helping me over the years. However, I would like to give special thanks to the following individuals:

Miranda Mecha, Rachel Hutchinson, and Meranda Masse, for being my lab moms over the years. You guys took me in as part of the lab and helped me every time. I remembered always coming late to your class Miranda and then showed up at lab sitting next to you… I remembered hanging out with you guys at the lab and my life in graduate school would not be the same without you.

Hanming Yang, for being a friend and a mentor. You taught me all I know about NMR spectroscopy and your approach to science. I look up to you a lot. Though you have graduated, you are still helping me. Thank you so much.

Siyu Li, for being an amazing colleague and a friend. Sorry I keep calling your cat fat (a bit true though). You made the last 2 years of my Graduate school life much better. Despite being younger, you are a better scientist than I am. I will always cherish the many times we went to hot pot together, drank boba tea when work was harsh, and laughed at the lab. Thank you for always going with my wimps.

Kevin England, for being an amazing Undergraduate, a mentor, and a friend. You were my first mentor in my Graduate School experience, despite being an Undergraduate. I love your passion for science and strive to carry myself throughout my career with as much passion you have. We helped each other when projects were not going well. And because of that, I learnt many things that have made me the scientist I am today. For that, thank you!

To those that have significant importance in my life and career: I would like to give special thanks to the important people who have contributed greatly to my life, to my career as a scientist, and to my graduate school experience. You guys have helped my through a lot of tough times and celebrated with me during my time of achievements. We have shared many happy moments that I will always remember and cherish for the rest of my life. I am glad that I joined the lab to meet all of you. Specifically, I would like to dedicate this section to the following individuals:

Jang Wonjo, for being one of two of my first ever mentors. We spent many nights doing microscopy on mouse neurons in the dark at Dr. Xu's lab. You caught me sleeping in the lab and let me keep the lights off while doing your patch-clamping experiments in the dark. I learnt so much of how to take my first steps as a scientist within the short 3 months at Augusta University that summer of 2018. You are a brilliant and strong person, who made me feel belonged.

Tzu-Wei, for being the other first ever mentors. You taught me all I know about how to do experiments and how to approach science. I poured my first ever gels (with water once…) with you. I remembered you watching baseball while teaching me how to do neuron culturing… It was one of my treasured memories. Though I am pretty fast, accurate, and efficient in my experiments,

I still don't hold a candle to what you could do when we worked together. Nowadays, I still look up to you both as a scientist as I reflect on the past times working with you and Wonjo. It felt like a family during the short time in our small lab. I can confidently say that I owed my entire career to both you and Wonjo. For that, I will carry all you both have taught me every day and in the future.

Ms. Carlin Glennon, for being the best teacher and mentor at St. Laurence High School. I still remember sitting in your AP biology class and falling asleep. I remembered helping you take care of the animals and plants in your class, though they did pass away often there… You inspired me to go into science in the first place. Now, I have come this far thanks to you. I hope I can see you again once I am in Burbank.

Prof. Harry Price, for being the best mentor at Stetson University. You inspired me to study proteins, to enroll in UW-Madison for my Ph.D., and to strive to go see the world. You taught me all I know about Biochemistry and how to be humble. You taught me not to roll over and to stand my ground. Here, I can confidently say I owe my career to you. I will visit you in Florida soon. Thank you.

Dr. Valeria Guzman-Luna, for being the main lab mom, a mentor, and a friend. You are the reason why I lasted so long in Graduate School. I learnt so, so, so many things from you as a scientist. You always take care of me in the lab. Even when both of us were miserable in the lab, you were a point of peace for me. I would not have been the scientist I am today without you. We share many fond memories in the same office. I knew you tried to save my in my first year too. But I am glad I joined and was able to meet you. You take care of me even when you are in a better position, away from academia. Let's keep being friends! Thank you.

Luu Nguyen Phuc, for being my best friend. We have been friends for over 10 years now. Looking back, you helped me through the lowest points of my life. I do not think I would have been here if we had not met. Even though we rarely meet because of the country differences, the way we are, has not changed. I have made many of my fondest memories with you: the late-night street food hunt, traveling to weird places, spending time together at your parents' house. Your

parents basically also took me in… To you, I dedicated myself to live upright, to live as myself, and to live happily. To me, you have been family for a long time already. There are no words I can use to thank you for what our friendship have given me…

Tran Dang Khoa, for being my close friend. I always run to you whenever my life goes wrong. And every time, you give me your friendship and unconditional help. For that, I can never repay your kindness enough… Over the 10 years we have been friends, I have learnt to be as well put-together as you are. You have supported me throughout the years, along with Phuc, even though the dumbest decisions I have made. You and Phuc will always be part of my life so I can't imagine how to repay you both. We will be friends for the rest of this life…

Kya Flukers, for being a bro, the missing 4th Destiny child, and my BL best friend. I have definitely killed many of your brain cells over the 6 years of friendship. Yet, you still come back so we can spend time talking or enjoying BL shows and "literature" together. I must apologize first for the bed incident now… Anyway, you have encouraged me to be myself and like whatever I want. We will have many more memories together, don't worry~ Thank you!

Philip Zhou, for being an amazing colleague and my close friend in my graduate school. I learn a lot on how to become a better human being and reflect on my shortcomings. You have given me a lot of advice that contributed to getting me to where I am. I will always cherish the late-night dinners and the complains about everything weird happing around us… I still need to take you to Vietnam, and you still need to take me around Sichuan and Tianjin. Let's continue to make great memories regardless of where we are.

Amanda Spiewak and Louis Morris, for being department Mom and Dad. The first person I met in graduate school and helped my take my first steps was you, Amanda. Over the years in graduate school, you both have helped me tremendously. We have many amazing memories and okay ones too (sorry for being so wasted on many occasions) together, like amazing parent-son dinners, Unikist concerts, League nights, etc. For that, I can say that I am a more mature person, able to handle many things coming my way, thanks to Mom and Dad.

Dr, Nicholas Hill and Dr. Brian Esselman, for being the best bosses in the Organic Teaching lab. Over the last 4.5 years, I have learnt so many things from you both about work ethics, teaching, collaborating, and having fun teaching Organic Chemistry. You always have supported me throughout my experience here. With TA being a big part of my experience, I would do it all over again to have a chance to work with you both again. Thank you!

Collectively, I want to thank my at-home friends as well: Micky Kwon, Tam Pham, Son Dong, Thong Phan, Linh Dan, Duc Phan, Dat Hoang, Khoa Bui and Charlotte Xu. You made me feel belonged in Madison. We have made so many memories in my apartments, good and bad… I am the person I am today with your help. I will not have survived Graduate School without you guys. Duc, your smile, is best. Let's make more memories in the future!

To Rylie Morris, Emma Santa, Jamorious Smith, and Emma Eisenbraun, you guys are amongst the closest department friends I have. I always go to complain and gossip, and you guys always listen. You supported my decisions in life and in my career. I wanted to graduate with everyone if I had the chance. Thank you as well to other colleagues that I have not had the chance to acknowledge.

To my family, I would like to thank all of my family, including my grandparents and my cousins, for everything I have since I was born. Specifically, I want to give appreciation to:

My mom, for single handedly raising me. We did not have a lot of money growing up, but you gave me everything I needed to be the person I am today. I never thought I would make it this far, being so bad at school in Vietnam… I know I am a difficult person growing and am not as filial as I should be. But I would like you to know that you care deeply about me, and I will carry myself with love and pride in my next steps.

My ma, for helping mom raise me. Despite our family situation, you always help me unconditionally. I would not be here without your support and love, from both you and mom. I am

still a stubborn son and will probably stay like that. Yet you still understand me with love. For that, I can never thank you enough.

If I have not specifically mentioned you, I would like to let you know that you have impacted my life and helped shaped me who I am today. I have had the privilege of receiving the love, friendship, teaching, and help from so many people in my life. I will never take this for granted!

THANK YOU !

Abstract

Proteins are responsible for driving a vast majority of biological processes in the cell. Therefore, it is important that proteins remain mostly folded to carry out their important functions, after transcription-translation. Prior proteome-wide studies showed that proteins in the *Escherichia coli* proteome are at risk of misfolding and extensive aggregation after translation, especially under cellular or environmental stresses. In addition, aggregates have been shown to be more thermodynamically stable than the native states, for many proteins. Therefore, if misfolding and aggregation in the cell were thermodynamically driven processes, many proteins in the *E. coli* proteome would spontaneously form aggregates in the cell. Despite a plethora of protein quality control and degradation systems employed by *E. coli*, it is clear that these systems are energetically expensive and can be overwhelmed, especially if many proteins in the proteome are at risk of thermodynamically driven aggregation. One popular proposed concept in the folding field is that the proteome is kinetically protected from aggregation, avoiding significant reliance on ATP-expensive protein quality control and degradation systems, under cellular conditions. However, there have not been experiments directly proving kinetic stability of proteins relative to aggregation on a proteome-wide scale. Moreover, aggregates are initially formed as soluble non-native oligomers, or soluble aggregates, These soluble aggregates may cause impairments of important biological processes in the cell and can eventually form larger insoluble assemblies. Yet, little is known about how soluble aggregates fit in the life cycle of *E. coli* proteins. To address the above fascinating biological questions, in this thesis, I will explore the kinetic stability and aggregation of a representative *E. coli* proteome (A19 cell strain).

This Ph.D. thesis includes three Chapters. Chapter 1 includes background information on protein folding and aggregation in the cellular context. In addition, it discusses the importance of proteome-wide studies and the important findings on aggregation from current

proteomic analyses and introduces the hypothesis of proteome kinetic stability relative to aggregation. This Chapter serves as the conceptual basis for the subsequent Chapters.

Chapter 2 explores the proteome-wide kinetic stability of proteins in *E. coli* under physiologically relevant conditions. Here, I demonstrate that the free-energy landscape of the *E. coli* proteome includes extensive insoluble aggregates under physiologically relevant conditions. Further, the soluble and insoluble aggregates can exchange reversibly among each other and the apparent critical concentration for soluble-aggregate formation on a proteome-wide scale is very low (c.a., 0.04 μg/mL). I demonstrate that over 80% of the proteome is kinetically protected from forming these soluble and insoluble aggregates on timescales longer than this organism's doubling time. Using bottom-up proteomics, I report that over 800 *E. coli* proteins are kinetically stable relative to aggregation regardless of structure, function, and cellular location. Finally, my results show that cytoplasmic/periplasmic molecular chaperones are amongst the most soluble proteins, both at higher temperature and under physiologically relevant conditions. This finding demonstrates that molecular chaperones are extremely robust members of the proteome.

Chapter 3 focuses on the development of a novel isotopolog of tryptophan to enhance the sensitivity of NMR spectroscopy and enable monitoring protein behaviors in complex environments. Here, a novel selectively isotopic-labeled tryptophan was synthesized and successfully detected at very low concentrations in buffer and in a complex S30 cell extract (c.a., 20 nM and 1 μM respectively) by low-concentration photochemically induced dynamic nuclear polarization (LC-photo-CIDNP) NMR. Our new selective labeling approach to LC-photo-CIDNP, in combination with existing biophysical analyses, can be utilized to study the effect of heterologous expression. For instance, this Trp isotopolog will enable the monitoring of folding and aggregation of model recombinant proteins within in complex cell-like environments.

Overall, I propose that proteome-wide kinetic stability is an effective strategy to maintain a healthy and aggregation-free cellular environment in living systems, without reliance on energetically expensive degradation and disaggregation processes. These results also provide

insights into the structural and functional determinants of bacterial kinetic stability and aggregation. In conclusion, it is hoped that the knowledge gained from this work will ultimately benefit the design and discovery of new strategies to prevent protein aggregation in the cell and to improve the shelf life of many protein-based biologics (e.g., monoclonal antibodies, etc).

Chapter 1

Introduction

1.1. Protein folding is a fundamental process for life.

Proteins are biological machines that drive cellular functions and ensure an organism survival in nature. The journey of all proteins in the cell starts with their translation on ribosome as polypeptide chains. Upon release, protein can fold into their native, three-dimensional structures that are able perform important cellular functions. Folding is proposed to be driven mainly by the hydrophobic effect, through burying the hydrophobic amino acids in the core of a protein and shielding them from unfavorable interactions with the aqueous cellular environment.[1] The simplest model of protein folding describes that the native state of a protein is the most thermodynamically stable and is spontaneously adopted by proteins during folding in solution.[2] However, large and complex proteins may require molecular chaperones to help folding intermediates cross kinetic barriers to access the native states.[3-5] This short period from translation on the ribosome and upon reaching the functional native states is defined as the early stages in protein's life (See Figure 1). Afterwards, proteins spend the rest of their time performing their associated biological functions in the complex cellular environment *in vivo*. This period, called the late stages of protein's life, is estimated to range from minutes to hours in bacteria.[6] Finally, protein's life ends with degradation by cellular proteolytic systems.[7] The dynamic balance between translation, folding, transport, and degradation throughout different life stages is key in maintaining the integrity of an organism's proteome, or protein homeostasis (See Figure 1.1).[8-10]

1.2. Protein aggregation in the *Escherichia coli* cells.

Yet, the life cycle of a protein is fallible in *E. coli*. During translation and folding in the early stages, proteins can form un/misfolded states where hydrophobic residues are more exposed as compared to that of native structures. The prolonged non-native exposure of the hydrophobic residues in these states can lead to association as large protein assemblies called aggregates under the crowded conditions of the cell.[1] Similarly, in the late stages, proteins may be at risk of unfolding or misfolding, especially under stress conditions, that can lead to aggregation.[11]

Aggregates are initially formed as soluble non-native oligomers, or soluble aggregates, and can eventually sequester more soluble proteins to form large, insoluble aggregates.[11] In most cases, aggregates are non-functional assemblies and prevent proteins from performing their native functions, leading to functional deficits to many biological processes *in vivo.* Some aggregates can also be biologically active and illicit toxic responses to the surrounding environment.[12-13] Accumulation of aggregates can lead to complete disbalance of protein homeostasis on a proteome-wide level and eventually cell death.

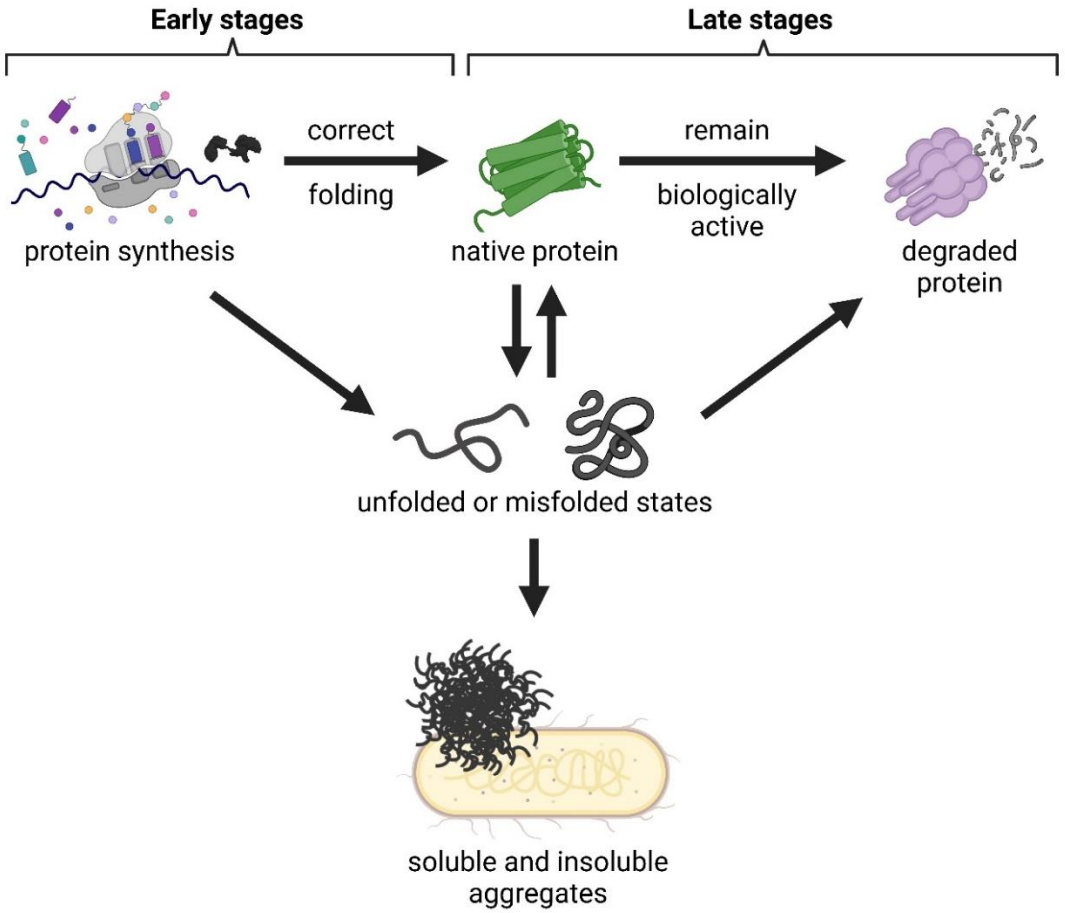


Figure 1.1. The life cycle of protein and the potential risks of un/misfolding and aggregation. Proteins life cycle begins from synthesis on the ribosome till its eventual degradation by the proteolytic system. Created with BioRender.com.

1.3. Cellular response to aggregation.

The cellular machinery of *E. coli* employs many molecular chaperones, disaggregases, and proteases in the cytoplasm to minimize aggregation and maintain protein homeostasis.[14-15] Molecular chaperones are essential helpers for protein folding throughout the protein's life cycle (See Figure 2, panel 1). In the early stages, *de novo* synthesis of proteins on the ribosome is assisted by the co-translational chaperone trigger factor[16] and the Hsp70 chaperone system, including the 70 kDa DnaK chaperones, 40 kDa DnaJ co-chaperone (Hsp40), and nucleotide exchange factor GrpE.[17] Subsequent folding to native states in *E. coli* cytoplasm upon release from the ribosome is further assisted mainly by the Hsp70 system[17], HptG chaperone (Hsp90)[18], and GroEL/S (Hsp60/Hsp10) system.[19] In particular, the Hsp70 and GroEL/S systems have also been demonstrated to rescue proteins from misfolding in the late stages of protein's life.[20-21] These molecular chaperones bind preferentially to hydrophobic motifs exposed on non-native conformations of client proteins to potentially prevent association to aggregates and promote folding to the native states.[5] Except for TF, major chaperone systems in *E. coli* cytoplasm rely on ATP hydrolysis to power their chaperone functions.[19, 22-23]

In concert with the quality control aspect of molecular chaperones, proteases also participate in maintaining protein homeostasis through the selective degradation of un/misfolded or unwanted proteins (See Figure 2, panel 1).[6-7] In *E. coli*, the major proteases in the cytoplasm are ATP-dependent Lon protease and the ClpP-type 2-protease systems (i.e., ClpXP and ClpAP).[15, 24] In the cytoplasm of *E. coli*, Lon protease is responsible for the degradation of over 50% misfolded proteins.[25] For 2 protease systems, they composed of the hexameric ClpP that oligomerize with the ATPase subunits of ClpA and ClpX to participate in the removal of un/misfolded proteins.[15] These proteases feature an AAA+ domain that degrades the client proteins via ATP binding and hydrolysis. In concert, the flux to un/misfolded states are degraded to potentially minimize prolong persistence in the cells, minimizing subsequent aggregation.

In addition to refolding chaperones and proteases, the *E. coli* cell also employs disaggregation machineries to rescue proteins from aggregates (See Figure 2, panel 1). The Hsp100 ClpB forms a hexameric ring that assists disaggregation via the ATP-dependent threading of unfolded proteins through the central core.[26] However, ClpB is not capable of carrying out disaggregation activities on its own and requires the cooperation of the K/J/E chaperone system to associate with protein aggregates and modulate its ATP-dependent activities.[11] Furthermore, in *E. coli*, the disaggregation activities of this bi-chaperone system (ClpB-K/J/E) is assisted by small heat shock proteins (sHsps) (See Figure 2, panel B). These sHsps can bind to the unfolded states of proteins inside aggregates and stabilize native-like states for the downstream disaggregation and refolding of the ClpB-K/J/E system.[27] Proteins recovered from aggregates by ClpB-K/J/E system are refolded and regained their biological functions *in vitro*.[28] The system is also capable of reverting protein aggregation on a proteome-level in *E. coli* cells after heat shock, suggesting that most of protein aggregates *in vivo* are processed by the ClpB-K/J/E system in the cytoplasm.[29]

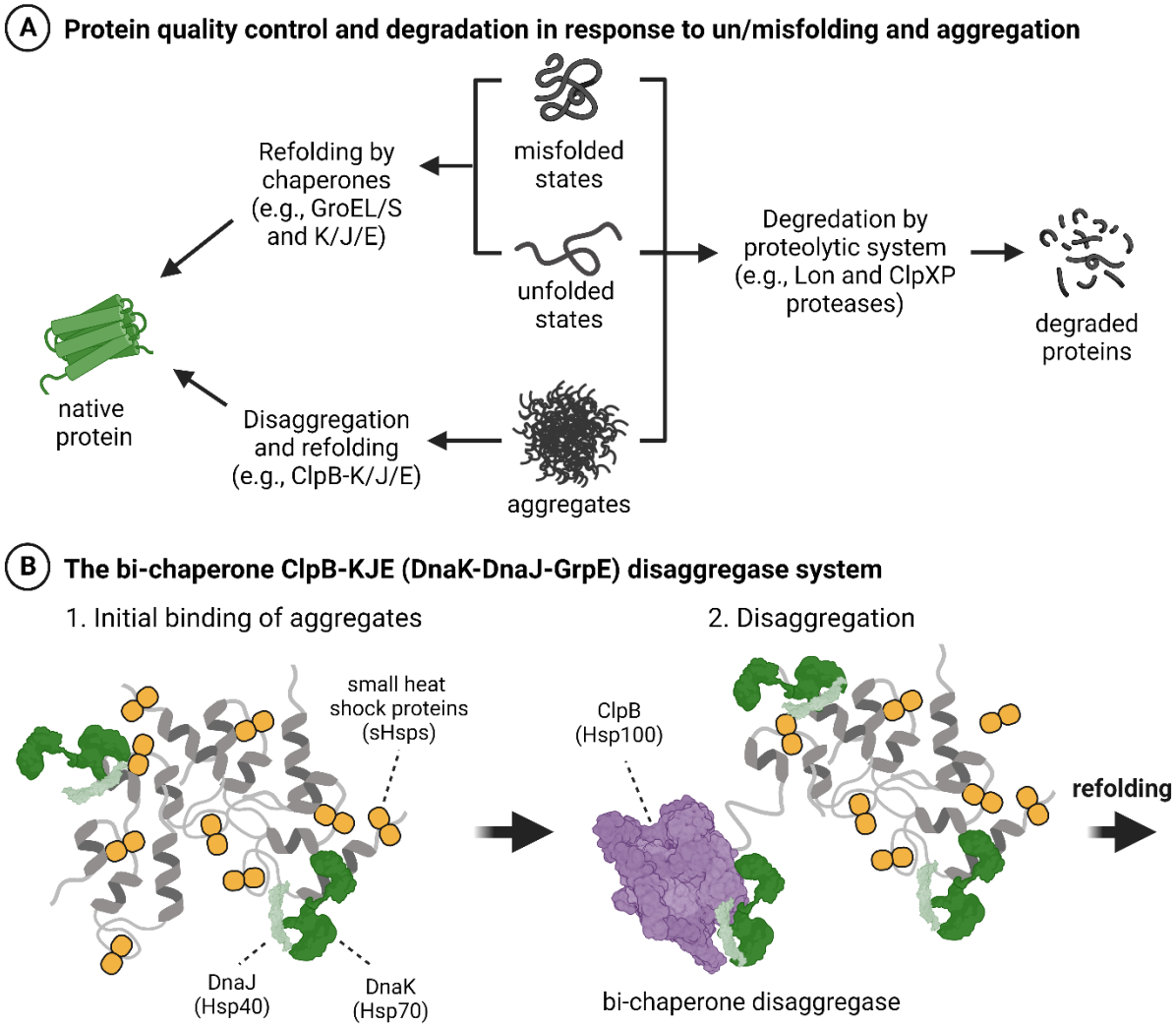


Figure 1.2. *E. coli* employs a plethora of proteins for quality control and degradation to maintain protein homeostasis. (A) Molecular chaperones, proteases, and disaggegases can protect the cell by minimizing misfolding and aggregation. (B) The bi-chaperone system ClpB-KJE works synergistically with small heat shock proteins (sHsps) to rescue functional proteins from aggregates. Adapted from Schram et. al. (2019). PDB: 2KHO (DnaK), 6JZB (DnaJ), and 5OG1(ClpB). Created with BioRender.com.

1.4. Stress-induced protein aggregation on a proteome-wide level.

In vitro studies show that the native states of proteins are sensitive to environmental changes, including protein concentrations, temperature, pH, ionic strengths (i.e., salt concentrations), cellular aging, molecular crowding, and heterologous overexpression.[2] Changes to these conditions in the cytoplasm and periplasm of *E. coli* can impede protein folding and promote proteins un/misfolding and aggregations. In cases where global misfolding and aggregation persist and accumulate, the cellular quality control and degradation machineries can be overwhelmed, leading to proteotoxicity and eventual cell death.

Environmental stresses, such as elevated temperatures or pH, can promote protein un/misfolding and aggregation on a proteome-wide level. Elevated temperatures from the basal growth temperature of an organism can induce the Gibbs free energy of the unfolded state to become more stable than the folded states[1, 30], leading many proteins to unfold on a proteome-scale. Under high temperature stress, the hydrophobic effect is stronger and drive unfolded proteins to self-associate through non-native hydrophobic interactions to reduce hydrophobic exposure, leading to formation amorphous or insoluble aggregates in *E. coli* under these conditions.[20, 31-34] Unlike the reversible nature of temperature-dependent un/misfolding and aggregation, exposures to chemicals (e.g., hypochlorous acid) associated to oxidative stress can lead to irreversible covalent modifications of protein sequences.[35] These modifications on amino acid side chains of a protein can disrupt native folding trajectories, increase un/misfolding, promote aggregation, and lead to resistance to proteolytic degradation.[36]

Besides environmental stresses that directly affect protein folding, impairment of ATP dependent systems (i.e., molecular chaperones) may lead to proteome-wide misfolding and aggregation.[32] Cell growth at stationary phase is associated with a decrease in ATP levels.[32] Many important quality control and degradation systems discussed above are powered through ATP hydrolysis to maintain folding and minimize aggregation *in vivo*. [19, 22-23] ATP depletion during the

transition into stationary phase can lead to functional impairment of these quality control and degradation systems, causing proteome-wide aggregation in *E. coli* cytoplasm.[32]

On the other side of the spectrum, cellular stress can also be associated with the heterologous expression of recombinant proteins in *E. coli*. This induced overexpression of exogenous proteins can promote aggregation due to (i) the tendency to self-aggregate, especially at high concentration, and/or (ii) the burden imposed on the endogenous quality control and degradation systems of bacteria by the demand of producing non-endogenous proteins.[37] Moreover, the process of heterologous expression is often carried out till stationary phase which contributes to additional environmental stress to the cells in the form of depletion of important metabolic and energetic resources (e.g., ATP). Accumulation of exogenous protein aggregates can lead to subsequent formation of large inclusion bodies, consisting of mainly the recombinant protein of interest in amyloid-like or native-like conformations.[38] Interestingly, inclusion bodies from heterologous overexpression can be benign and pose little toxicity to the host cell.[38] These cases are examples demonstrating that protein insolubility does not always associate with proteotoxicity.

1.5. Aggregate localization and inheritance in *E. coli*.

The nucleoid, mainly made up of DNA, in an *E. coli* cell is an example of a large intracellular structure that can impose spatial constraints on the aggregation of proteins. Studies that probe the localization of proteome-level aggregation induced by environmental stresses demonstrate a preferential organization of these aggregates at the poles in *E. coli*.[39-42] For example, in anucleoid cell strains lacking the chromosome partitioning protein MukB, a single large aggregate is observed at the middle of a cell in the absence of a centralized nucleoid.[40] Changes to the conformation and size of the nucleoid resulted in the redistribution of the pole-localized aggregates to the spaces between the nucleoid.[42] In addition, experimental increase in viscosity

of the crowded cytoplasm can promote aggregation within the nucleoid.[43] Overall, these studies suggest that the existence of single aggregates localized at the poles are driven by the aggregation size constraints from the nucleoid preventing formation of large aggregates in this space and free diffusion of smaller aggregates, instead of the active localized formation of aggregates at the pole. The persistence of localized aggregates in *E. coli* influences how it is distributed to the next generations during cell growth.

When the aggregates persist on longer timescales than the cell cycle and division time, they are distributed to the emerging daughter cells.[39] As previously discussed, aggregates are preferentially localized at the poles of a progenitor cell. [39-42] Upon the first event of cell division, each of the two daughter cells inherit one aggregate at the old pole originated from the progenitor cell (See Figure 3).[44-45] A second cell division event leads to four daughter cells, where two cells are made up of new cell poles entirely and escaped inheritance of pole-localized aggregates. [44-45] The other two daughter cells contain the original, oldest poles with the associated pole-localized aggregates. [44-45] This mode of aggregate distribution during cell division is called “asymmetric inheritance” of protein aggregation.[46] Since these long-lived aggregates rarely move away from the cell poles, aggregates are retained within the newer generations as cells divide and inherit the original old cell pole of the progenitors.[39-40]

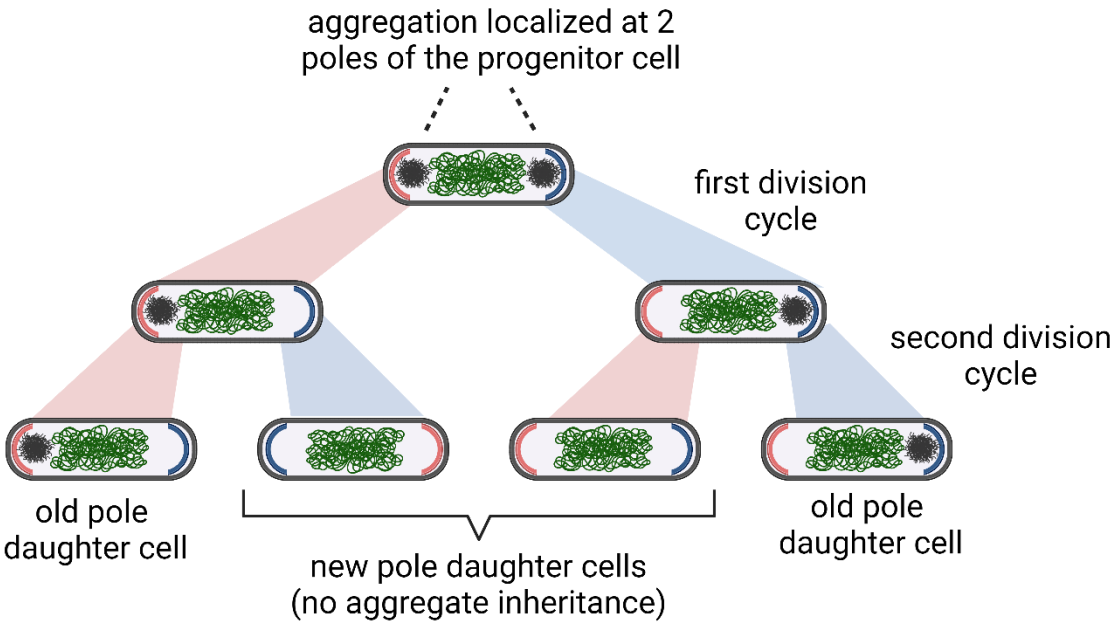


Figure 1.3. Asymmetric inheritance of pole-localized protein aggregates in *E coli*. Asymmetric inheritance result in 2 daughter cells that inherited the original poles and their associated pole-localized aggregates while 2 other daughter cells escaped aggregate inheritance. Adapted from Schram et. al. (2019). Created with BioRender.com.

1.6. Why should we study protein aggregation at a proteome level?

Despite intensive focus on protein folding and aggregation, the ability to accurately predict protein solubility, aggregation propensity, and stability is still limited. In order to resolve potential determinants for accurate predictive power, model proteins in research must represent the diversity of proteins in the proteome.[2] Much of the current knowledge on these topics was derived from the unfolding and refolding *in vitro* of simple "well-behaved" model proteins that only represent a small subset (8.4%) of structural and functional diversity in the *E. coli* proteome.[47] Little is known about the folding and aggregation of large and multimeric proteins. There is also currently a lack of representation of proteins located in the periplasm or the membrane as much emphasis was placed to probing those native to the aqueous cytoplasm.[47]

In addition, many studies have implicated the potential global effects on protein folding, misfolding, and aggregation imposed on the proteome *in vivo* under different growth conditions and environmental stresses.[20, 34, 48] Much can still be learnt on how the cells maintain protein homeostasis in response to these global events. Therefore, it is important to conduct studies on a proteome-wide level, under normal conditions or environmental stresses, to fully understand the fundamentals of protein folding and aggregation for all proteins in the cell. The following sections will highlight important findings to folding and aggregation at different life stages from current proteomics studies.

1.7. *In vitro* proteomics to study the early and late stages of protein's life.

Many current proteomic studies are done *in vitro* to study misfolding and aggregation across protein's life cycle, with focuses on environmental stress conditions (e.g., thermal treatment, chemical denaturation, etc.). In 2009, Niwa et. al. probe the risk of aggregation on a proteome-wide scale at the early stages.[49] Over 800 out of 2100 *E coli* cytosolic proteins had low

solubility (ca., <30%) and formed insoluble aggregates upon synthesis from the ribosome *in vitro* with the PURE cell-free expression system lacking endogenous chaperones.[49] A follow-up study from the same group revealed that the synthesis of a subset of highly insoluble proteins in the previous study in the presence of major cytoplasmic molecular chaperone systems (i.e., GroEL/S, K/J/E, and TF) significantly decreased aggregation.[50] Though no solubility determinant was identified, these series of studies highlight how the proteome of *E. coli* is at risk of aggregation at the early stages and how molecular chaperones play a key role during this period by rescuing newly synthesized proteins from aggregation.

Many other *in vitro* proteomic studies focus on monitoring misfolding and aggregation at the late stages of protein's life. In To et. al. (2019), *E coli* proteome samples were subjected to unfolding-refolding in chemical denaturants with dilution under conditions optimized for minimal formation of insoluble aggregates (i.e., refolding at low total protein concentrations and higher pH of 8.2, and depletion of ribosomal proteins).[20] Upon refolding with dilution, over 33% of the proteome cannot refold to the native state as monitored with limited proteolysis LC-MS/MS approach. These soluble non-native states may be unfolded, misfolded, or soluble aggregated states. A follow up study was done to show that refolding in the presence of GroEL/S or K/J/E chaperone systems rescues 50% of the non-refoldable proteins previously observed.[21]

Moreover, the *E. coli* proteome was also shown to be at risk of forming aggregation under heat-stress conditions. Leuenberger et. al. (2017) subjected the *E. coli* lysates to increasing temperature and proteome-wide denaturation curves (i.e., the change in solubility of soluble proteins to insoluble aggregates as a function of temperature) were constructed with limited proteolysis LC-MS.[51] Most *E.* coli proteins formed heat-induced insoluble aggregates upon heating at 70 °C. Thermostable proteins (i.e., proteins with melting temperature >60 °C) are correlated to having higher abundances and faster protein synthesis rate. On the other hand, thermostability is found to be inversely correlated to increasing protein length and heat-induced

aggregation propensity. Similarly, Varela et. al. (2018), 85% of the total proteins in the *E. coli* proteome sample also form heat-induced insoluble aggregates upon heating at 70 °C.[34]

Overall, these *in vitro* studies reveal how fallible the *E. coli* proteome can be in response to environmental stresses at different life stages and highlight the potential roles of molecular chaperones in alleviating protein misfolding and aggregation. However, even with extensive efforts to mimic the cellular environment in the test tube, *in vitro* experiments cannot fully recapitulate the complex native cellular conditions in the cells. Specifically, many proteins fold efficiently *in vivo* cannot refold *in vitro*, highlighting the different fundamentals of protein folding and aggregation in the cells. Therefore, it is crucial for proteomic works to characterize misfolding and aggregation in the cellular context.

1.8. First steps to revealing folding and aggregation with *in vivo* proteomics.

Despite the on-going progress of developing more advanced technologies to monitor biological processes in the cells, the efforts to elucidate protein aggregation on a proteome-wide scale *in vivo* has begun to gain traction for *E coli* cells.

Mateus et. al. (2018) evaluated the melting profiles of the *E. coli* proteome in vivo at different growth phases.[48] Intact *E. coli* cells were subjected to heat treatment at increasing temperatures and the solubility as a function of temperature (i.e., to construct melting curve) for each protein in the proteome was monitored with quantitative tandem-mass tagging LC-MS approach. Under cellular conditions, *E. coli* proteins has high solubility or do not form insoluble aggregates even upon heating at 50 °C. Overall, the *E. coli* proteome is demonstrated to be thermostable and non-aggregation prone under thermal stress at 15 °C higher than the optimal growth temperature, consistent with the bacterial ability to proliferate at a wide range of

temperatures. However, it is difficult to characterize how the observed thermal stability influences protein aggregation under environmental stress at different stages of protein's life.

One such work to probe protein aggregation at either the early stages or late stages of protein's life is Gyorkei et. al. (2022). In this work, GFP-fusion of cytosolic *E. coli* proteins were individually overexpressed, and the subsequent pole-localized aggregates in the cells were visualized with fluorescent microscopy.[52] The group revealed that 18.6% and 43.6% of cytosolic proteins formed aggregates (i) co-translationally or immediately upon release from the ribosome (i.e., early stages), or (ii) after translation (i.e., late stages) when expressed above native concentrations under physiologically relevant conditions. Interestingly, proteins that aggregated slowly after translation are correlated to having faster folding rates compared to proteins that aggregated in the early stages. This is consistent with previous studies demonstrating that slow aggregation rates and rapid folding may contribute to efficient protein folding.[53] Overall, Gyorkei et. al. is one of the first to demonstrate that *E coli* proteome is at risk of aggregation throughout the protein's life cycle *in vivo*. One caveat to this work is that the observed aggregation profiles for some proteins may be caused or influenced by the expression of endogenous proteins as a fusion with GFP. However, with the advancing technologies to monitor biological processes in the cells, future *in vivo* proteomics may be close to observing protein folding and aggregation of unmodified native proteins.

1.9. The unexplored biological role of proteome kinetic stability.

If aggregation was a thermodynamically favorable process, a global aggregating event, potentially triggered by environmental stress (e.g., temperature, ionic strength pH, etc.), can overwhelm the quality control and degradation systems capacities to maintain protein homeostasis.[54] Therefore, many studies allude to a quality control phenomenon, driven by kinetics, to minimize the flux towards misfolded and aggregated states under physiologically relevant conditions.[20-21, 34, 55-58] However, no study has been conducted to directly probe this idea

of proteome-wide kinetic protection as native states during the late stages of proteins life. Moreover, it has also been suggested that these protein aggregates may start as soluble aggregates before forming insoluble aggregates.[59-60] Recent works suggest that soluble aggregates may exert cytotoxic effects, impairing the biological functions and folding of proteins in many protein-related diseases.[11-13, 61] Yet, there is also a lack of knowledge on how soluble aggregates fit on the free-energy conformational landscape on a proteome-wide scale.

1.10. This thesis work.

The main focus is to further the understanding of protein solubility and aggregation in the cells by probing the kinetic stability of *E. coli* proteins across the proteome during the late stages.

Chapter 2 explores the proteome-wide kinetic stability of proteins in *E. coli* under physiologically relevant conditions. Here, we demonstrate that over 80% of the proteome is kinetically protected from forming soluble and insoluble aggregates on timescales longer than the organism's division time. Furthermore, the soluble and insoluble aggregates in our experiments can exchange reversibly under thermodynamic equilibrium. Finally, we show how kinetic stability is pervasive in the *E. coli* proteome, across a protein collection diverse in structures, functions, and cellular locations.

Chapter 3 focuses on the development of a novel isotopolog of tryptophan to enhance the sensitivity of NMR spectroscopy to monitor protein behaviors in complex environments. Here, we synthesized a novel selective isotopic labeling approach of tryptophan by modifying the previously developed enzymatic synthesis. We then successfully detected this isotopolog of Trp at very low concentrations in buffer and in a complex S30 cell extract (c.a., 20 nm and 1 μM respectively) using photo-chemically induced dynamic nuclear polarization (photo-CIDNP) NMR.

1.11. Innovations.

The following thesis is characterized by several innovative aspects, as summarized below.

a. Little is known about the biological importance of kinetic stability for an organism's proteome relative to protein folding and aggregation. The findings in Chapter 2 will be the first to report on kinetic stability across the *E. coli* proteome and its biological relevant to preventing protein aggregation under physiological conditions.

b. This is also the first to demonstrate that heat-induced insoluble aggregates of over 1,700 *E. coli* proteins can reversibly exchange with soluble aggregates. This phenomenon, which is widely characterized in amyloid-like proteins, is general for proteins across an bacterial proteome and may hold important biological implications.

c. Finally, this work will also be the first to measure the apparent free-energy changes associated with the unfolding of soluble states and insoluble aggregates of a complex proteomic sample. These denaturant unfolding experiments will provide context on the thermodynamics of aggregation to further enhance the importance of kinetic stability.

1.12. References

1. Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R., The Protein Folding Problem. *Annual Review of Biophysics* **2008,** *37* (1), 289-316.
2. Anfinsen, C. B., Principles that govern the folding of protein chains. *Science* **1973,** *181* (4096), 223-30.
3. Oberhauser, A. F.; Marszalek, P. E.; Carrion-Vazquez, M.; Fernandez, J. M., Single protein misfolding events captured by atomic force microscopy. *Nature Structural Biology* **1999,** *6* (11), 1025-1028.
4. Ramshini, H.; Parrini, C.; Relini, A.; Zampagni, M.; Mannini, B.; Pesce, A.; Saboury, A. A.; Nemat-Gorgani, M.; Chiti, F., Large proteins have a great tendency to aggregate but a low propensity to form amyloid fibrils. *PLoS One* **2011,** *6* (1), e16075.
5. Balchin, D.; Hayer-Hartl, M.; Hartl, F. U., In vivo aspects of protein folding and quality control. *Science* **2016,** *353* (6294), aac4354.
6. Maurizi, M. R., Proteases and protein degradation in Escherichia coli. *Experientia* **1992,** *48* (2), 178-201.
7. Sauer, R. T.; Baker, T. A., AAA+ proteases: ATP-fueled machines of protein destruction. *Annual review of biochemistry* **2011,** *80*, 587-612.
8. Powers, E. T.; Morimoto, R. I.; Dillin, A.; Kelly, J. W.; Balch, W. E., Biological and chemical approaches to diseases of proteostasis deficiency. *Annual review of biochemistry* **2009,** *78*, 959-991.
9. Richter, K.; Haslbeck, M.; Buchner, J., The heat shock response: life on the verge of death. *Molecular cell* **2010,** *40* (2), 253-266.
10. Schramm, F. D.; Schroeder, K.; Jonas, K., Protein aggregation in bacteria. *FEMS Microbiology Reviews* **2019,** *44* (1), 54-72.
11. Mogk, A.; Bukau, B.; Kampinga, H. H., Cellular Handling of Protein Aggregates by Disaggregation Machines. *Molecular Cell* **2018,** *69* (2), 214-226.
12. Larson, M. E.; Lesné, S. E., Soluble Aβ oligomer production and toxicity. *Journal of Neurochemistry* **2012,** *120* (s1), 125-139.
13. De, S.; Whiten, D. R.; Ruggeri, F. S.; Hughes, C.; Rodrigues, M.; Sideris, D. I.; Taylor, C. G.; Aprile, F. A.; Muyldermans, S.; Knowles, T. P. J.; Vendruscolo, M.; Bryant, C.; Blennow, K.; Skoog, I.; Kern, S.; Zetterberg, H.; Klenerman, D., Soluble aggregates present in cerebrospinal fluid change in size and mechanism of toxicity during Alzheimer's disease progression. *Acta Neuropathologica Communications* **2019,** *7* (1), 120.
14. Sweeny, E. A.; Shorter, J., Mechanistic and Structural Insights into the Prion-Disaggregase Activity of Hsp104. *Journal of Molecular Biology* **2016,** *428* (9, Part B), 1870-1885.
15. LaBreck, C. J.; May, S.; Viola, M. G.; Conti, J.; Camberg, J. L., The Protein Chaperone ClpX Targets Native and Non-native Aggregated Substrates for Remodeling, Disassembly, and Degradation with ClpP. *Frontiers in Molecular Biosciences* **2017,** *4*.
16. Maier, R.; Scholz, C.; Schmid, F. X., Dynamic association of trigger factor with protein substrates. *Journal of molecular biology* **2001,** *314* (5), 1181-1190.
17. Calloni, G.; Chen, T.; Schermann, Sonya M.; Chang, H.-c.; Genevaux, P.; Agostini, F.; Tartaglia, Gian G.; Hayer-Hartl, M.; Hartl, F. U., DnaK Functions as a Central Hub in the E. coli Chaperone Network. *Cell Reports* **2012,** *1* (3), 251-264.
18. Corteggiani, M.; Bossuet-Greif, N.; Nougayrède, J.-P.; Byrne, D.; Ilbert, M.; Dementin, S.; Giudici-Orticoni, M.-T.; Méjean, V.; Oswald, E.; Genest, O., Uncoupling the Hsp90 and DnaK chaperone activities revealed the in vivo relevance of their collaboration in bacteria. *Proceedings of the National Academy of Sciences* **2022,** *119* (37), e2201779119.
19. Hayer-Hartl, M.; Bracher, A.; Hartl, F. U., The GroEL–GroES Chaperonin Machine: A Nano-Cage for Protein Folding. *Trends in Biochemical Sciences* **2016,** *41* (1), 62-76.

20. To, P.; Whitehead, B.; Tarbox, H. E.; Fried, S. D., Nonrefoldability is Pervasive Across the E. coli Proteome. *Journal of the American Chemical Society* **2021,** *143* (30), 11435-11448.
21. To, P.; Xia, Y.; Lee, S. O.; Devlin, T.; Fleming, K. G.; Fried, S. D., A proteome-wide map of chaperone-assisted protein refolding in a cytosol-like milieu. *Proceedings of the National Academy of Sciences* **2022,** *119* (48), e2210536119.
22. Pearl, L. H.; Prodromou, C., Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu. Rev. Biochem.* **2006,** *75*, 271-294.
23. Mayer, M. P.; Gierasch, L. M., Recent advances in the structural and mechanistic aspects of Hsp70 molecular chaperones. *Journal of Biological Chemistry* **2019,** *294* (6), 2085-2097.
24. Sauerbrei, B.; Arends, J.; Schünemann, D.; Narberhaus, F., Lon Protease Removes Excess Signal Recognition Particle Protein in Escherichia coli. *Journal of Bacteriology* **2020,** *202* (14), 10.1128/jb.00161-20.
25. Chung, C. H.; Goldberg, A. L., The product of the lon (capR) gene in Escherichia coli is the ATP-dependent protease, protease La. *Proceedings of the National Academy of Sciences* **1981,** *78* (8), 4931-4935.
26. Rizo, A. N.; Lin, J.; Gates, S. N.; Tse, E.; Bart, S. M.; Castellano, L. M.; DiMaio, F.; Shorter, J.; Southworth, D. R., Structural basis for substrate gripping and translocation by the ClpB AAA+ disaggregase. *Nature communications* **2019,** *10* (1), 2393.
27. Mogk, A.; Ruger-Herreros, C.; Bukau, B., Cellular functions and mechanisms of action of small heat shock proteins. *Annual review of microbiology* **2019,** *73*, 89-110.
28. Weibezahn, J.; Tessarz, P.; Schlieker, C.; Zahn, R.; Maglica, Z.; Lee, S.; Zentgraf, H.; Weber-Ban, E. U.; Dougan, D. A.; Tsai, F. T., Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cell* **2004,** *119* (5), 653-665.
29. Tomoyasu, T.; Mogk, A.; Langen, H.; Goloubinoff, P.; Bukau, B., Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the Escherichia coli cytosol. *Molecular microbiology* **2001,** *40* (2), 397-413.
30. Dill, K. A.; Ghosh, K.; Schmit, J. D., Physical limits of cells and proteomes. *Proceedings of the National Academy of Sciences* **2011,** *108* (44), 17876-17882.
31. Khodaparast, L.; Khodaparast, L.; Gallardo, R.; Louros, N. N.; Michiels, E.; Ramakrishnan, R.; Ramakers, M.; Claes, F.; Young, L.; Shahrooei, M., Aggregating sequences that occur in many proteins constitute weak spots of bacterial proteostasis. *Nature communications* **2018,** *9* (1), 866.
32. Pu, Y.; Li, Y.; Jin, X.; Tian, T.; Ma, Q.; Zhao, Z.; Lin, S.-y.; Chen, Z.; Li, B.; Yao, G., ATP-dependent dynamic protein aggregation regulates bacterial dormancy depth critical for antibiotic tolerance. *Molecular cell* **2019,** *73* (1), 143-156. e4.
33. Kwon, S.; Jung, Y.; Lim, D., Proteomic analysis of heat-stable proteins in Escherichia coli. *BMB Rep* **2008,** *41* (2), 108-11.
34. Varela, A. E.; Lang, J. F.; Wu, Y.; Dalphin, M. D.; Stangl, A. J.; Okuno, Y.; Cavagnero, S., Kinetic Trapping of Folded Proteins Relative to Aggregates under Physiologically Relevant Conditions. *The Journal of Physical Chemistry B* **2018,** *122* (31), 7682-7698.
35. Santra, M.; Dill, K. A.; de Graff, A. M., How do chaperones protect a cell's proteins from oxidative damage? *Cell systems* **2018,** *6* (6), 743-751. e3.
36. Dahl, J.-U.; Gray, M. J.; Jakob, U., Protein quality control under oxidative stress conditions. *Journal of molecular biology* **2015,** *427* (7), 1549-1563.
37. Yuan, A. H.; Garrity, S. J.; Nako, E.; Hochschild, A., Prion propagation can occur in a prokaryote and requires the ClpB chaperone. *Elife* **2014,** *3*, e02949.
38. de Marco, A.; Ferrer-Miralles, N.; Garcia-Fruitós, E.; Mitraki, A.; Peternel, S.; Rinas, U.; Trujillo-Roldán, M. A.; Valdez-Cruz, N. A.; Vázquez, E.; Villaverde, A., Bacterial inclusion bodies are industrially exploitable amyloids. *FEMS Microbiology Reviews* **2018,** *43* (1), 53-72.

39. Lindner, A. B.; Madden, R.; Demarez, A.; Stewart, E. J.; Taddei, F., Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proceedings of the National Academy of Sciences* **2008,** *105* (8), 3076-3081.
40. Winkler, J.; Seybert, A.; König, L.; Pruggnaller, S.; Haselmann, U.; Sourjik, V.; Weiss, M.; Frangakis, A. S.; Mogk, A.; Bukau, B., Quantitative and spatio-temporal features of protein aggregation in Escherichia coli and consequences on protein quality control and cellular ageing. *The EMBO journal* **2010,** *29* (5), 910-923.
41. Govers, S. K.; Dutré, P.; Aertsen, A., In vivo disassembly and reassembly of protein aggregates in Escherichia coli. *Journal of bacteriology* **2014,** *196* (13), 2325-2332.
42. Neeli-Venkata, R.; Martikainen, A.; Gupta, A.; Gonçalves, N.; Fonseca, J.; Ribeiro, A. S., Robustness of the process of nucleoid exclusion of protein aggregates in Escherichia coli. *Journal of bacteriology* **2016,** *198* (6), 898-906.
43. Oliveira, S. M.; Neeli-Venkata, R.; Goncalves, N. S.; Santinha, J. A.; Martins, L.; Tran, H.; Mäkelä, J.; Gupta, A.; Barandas, M.; Häkkinen, A., Increased cytoplasm viscosity hampers aggregate polar segregation in Escherichia coli. *Molecular microbiology* **2016,** *99* (4), 686-699.
44. Coquel, A.-S.; Jacob, J.-P.; Primet, M.; Demarez, A.; Dimiccoli, M.; Julou, T.; Moisan, L.; Lindner, A. B.; Berry, H., Localization of Protein Aggregation in Escherichia coli Is Governed by Diffusion and Nucleoid Macromolecular Crowding Effect. *PLOS Computational Biology* **2013,** *9* (4), e1003038.
45. Gupta, A.; Lloyd-Price, J.; Neeli-Venkata, R.; Oliveira, Samuel M. D.; Ribeiro, Andre S., In Vivo Kinetics of Segregation and Polar Retention of MS2-GFP-RNA Complexes in Escherichia coli. *Biophysical Journal* **2014,** *106* (9), 1928-1937.
46. Nyström, T.; Liu, B., Protein quality control in time and space–links to cellular aging. *FEMS yeast research* **2014,** *14* (1), 40-48.
47. Braselmann, E.; Chaney, J. L.; Clark, P. L., Folding the proteome. *Trends Biochem Sci* **2013,** *38* (7), 337-44.
48. Mateus, A.; Bobonis, J.; Kurzawa, N.; Stein, F.; Helm, D.; Hevler, J.; Typas, A.; Savitski, M. M., Thermal proteome profiling in bacteria: probing protein state in vivo. *Molecular Systems Biology* **2018,** *14* (7), e8242.
49. Niwa, T.; Ying, B.-W.; Saito, K.; Jin, W.; Takada, S.; Ueda, T.; Taguchi, H., Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proceedings of the National Academy of Sciences* **2009,** *106* (11), 4201-4206.
50. Niwa, T.; Kanamori, T.; Ueda, T.; Taguchi, H., Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proceedings of the National Academy of Sciences* **2012,** *109* (23), 8937-8942.
51. Leuenberger, P.; Ganscha, S.; Kahraman, A.; Cappelletti, V.; Boersema, P. J.; von Mering, C.; Claassen, M.; Picotti, P., Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **2017,** *355* (6327), eaai7825.
52. Györkei, Á.; Daruka, L.; Balogh, D.; Őszi, E.; Magyar, Z.; Szappanos, B.; Fekete, G.; Fuxreiter, M.; Horváth, P.; Pál, C.; Kintses, B.; Papp, B., Proteome-wide landscape of solubility limits in a bacterial cell. *Sci Rep* **2022,** *12* (1), 6547.
53. de Groot, N. S.; Ventura, S., Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. *Journal of Biotechnology* **2006,** *125* (1), 110-113.
54. Hipp, M. S.; Park, S.-H.; Hartl, F. U., Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends in Cell Biology* **2014,** *24* (9), 506-514.
55. Nissley, D. A.; Jiang, Y.; Trovato, F.; Sitarik, I.; Narayan, K. B.; To, P.; Xia, Y.; Fried, S. D.; O'Brien, E. P., Universal protein misfolding intermediates can bypass the proteostasis network and remain soluble and less functional. *Nature Communications* **2022,** *13* (1), 3081.

56. Vecchi, G.; Sormanni, P.; Mannini, B.; Vandelli, A.; Tartaglia, G. G.; Dobson, C. M.; Hartl, F. U.; Vendruscolo, M., Proteome-wide observation of the phenomenon of life on the edge of solubility. *Proceedings of the National Academy of Sciences* **2020,** *117* (2), 1015-1020.
57. Gazit, E., The “Correctly Folded” State of Proteins: Is It a Metastable State? *Angewandte Chemie International Edition* **2002,** *41* (2), 257-259.
58. Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-Cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M., Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *Journal of the American Chemical Society* **2011,** *133* (36), 14160-14163.
59. Morell, M.; Bravo, R.; Espargaró, A.; Sisquella, X.; Avilés, F. X.; Fernàndez-Busquets, X.; Ventura, S., Inclusion bodies: Specificity in their aggregation process and amyloid-like structure. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2008,** *1783* (10), 1815-1825.
60. Carrió, M. M.; Cubarsi, R.; Villaverde, A., Fine architecture of bacterial inclusion bodies. *FEBS Lett* **2000,** *471* (1), 7-11.
61. Lee, S.; Choi, M. C.; Al Adem, K.; Lukman, S.; Kim, T.-Y., Aggregation and Cellular Toxicity of Pathogenic or Non-pathogenic Proteins. *Scientific Reports* **2020,** *10* (1), 5120.

Chapter 2

Exploring kinetically controlled protein solubility on a proteome-wide scale in *E. coli* under physiologically relevant conditions

This Chapter reproduces a manuscript *in prepartions* by Hung H. Dang, Lucas E. Bartel, Grzegorz Sabat, Yuzhu Liu, Mridula Srivathsan, Kevin A. England, and Silvia Cavagnero. H.D. designed and performed all experiments, analyzed the data, including bioinformatics and statistical analysis, and wrote the manuscript. L.B. designed models for aggregation, fitted solubility data to said models, produced solubility correlation raw data, and contributed 8 bioinformatic analyses in the supporting information. G.S. prepared all samples for LC-MS, carried out experiments on the LC-MS instrument, and analyzed the raw peptide data. Y.L. contributed to supporting unfolding in chemical denaturant experiments, performing bioinformatic analyses, and participated in constructing figures and tables in the supporting information. M.S. provided the identifications and descriptions to the raw solubility data for downstream bioinformatic analysis. K.E. contributed to the method and experimental designs in the early stages of the project. S.C. participated in data analysis, supported the writing and contributed edits of the manuscript.

2.1. Abstract

The interplay between protein folding and aggregation is still poorly understood, despite its importance in basic science, disease, as well as the efficient production and shelf-life of pharmaceuticals. Here, we show that most of the *E. coli* proteome is kinetically protected from forming thermodynamically stable aggregates under physiologically relevant conditions, over timescales longer than the bacterial doubling time. Kinetic trapping from aggregation applies across diverse biophysical properties, functions, and cellular locations. However, proteins with the highest thermodynamic tendency to aggregate are most strongly trapped. Cytoplasmic and periplasmic chaperones are particularly soluble at high temperature and optimally kinetically trapped from aggregation under physiological conditions, proving to be extremely robust members of the proteome. Insoluble aggregates serve as dangerous reservoirs, as they are in thermodynamic equilibrium with small concentrations of soluble aggregates, but not with native states. Overall, this study highlights the coexistence of kinetic trapping and reversible equilibria across protein free-energy landscapes of proteins on a proteomic scale.

2.2. Introduction

The early stages of protein life in living cells involve conformational sampling and interactions with a variety of cellular components including molecular chaperones, the ribosome and crowding agents[62-64]. Transcription and translation modulate protein abundance and conformational quality[65-67]. In addition, co- and post-translational interactions with molecular chaperones[5, 17, 68-70] further promote the adjustment of structural characteristics and regulate the extent of co- and post-translational aggregation. Shortly after completion of protein biosynthesis and after having attained the correct folded state, proteins spend the remainder of their life exploiting their biological activity. The overall duration of this stage typically dominates protein life. Interestingly, the effective *in vivo* lifetime of proteins varies widely, ranging from a few minutes to multiple days and years, depending on protein type and parent organism.[6, 71-73] In healthy cells, proteins keep reversibly folding and unfolding[2, 74] during their life span, and populate aggregation-free, covalently intact biologically relevant states.

While undesirable insoluble aggregates are known to exist in most bacterial proteins across their complex free-energy landscapes,[33-34, 48-49] these aggregates are hardly ever populated, under physiologically relevant conditions. This phenomenon is due in part to low extent of insoluble-aggregate formation in healthy cells, and in part to the presence of protein machineries devoted to protein degradation[15] and disaggregation.[14, 75] These molecular machines, including the degradation-devoted ClpXP[15] and the Lon protease, and the ClpB, ClpG and Hsp104 disaggregases[14, 75] in *E. coli*, often work in concert with chaperones and are extremely free-energy expensive. However, proteases and disaggregases are often unable to adequately cope with environmental stress (e.g., anomalous temperature, ionic strength pH, etc.) and aging.[54] Fortunately, under non-stress conditions, the flux towards stable aggregated states, which could thermodynamically become highly populated, is limited by kinetic factors. Consequently, aggregation is often extremely slow relative to protein lifetime.[20-21, 34, 55-58] It I

important to note at this juncture that aggregates are usually more thermodynamically stable than the respective native states.[58, 76]

However, no study has been conducted to directly probe the idea of proteome-wide kinetic protection as native states during the late stages of proteins life. Moreover, it has also been suggested that protein aggregates may start as soluble aggregates before forming insoluble aggregates.[59-60] Recent works suggest that soluble aggregates may exert cytotoxic effects, impairing the biological functions and folding of proteins in many protein-related diseases.[11-13, 61] Yet, there is a lack of knowledge on how soluble aggregates fit on the free-energy conformational landscape on a proteome-wide scale.

In this work, we employ a combination of solubility assays, dynamic light scattering and mass spectrometry to identify the specific extent of kinetic trapping (relative to soluble and insoluble aggregates) experienced by different members of the *E. coli* proteome. Our three-pronged approach enables understanding how Nature employs kinetic characteristics of the proteome to maintain cell viability. First, we identify soluble and insoluble aggregates populated under physiologically relevant conditions via a novel anaerobic heating and cooling procedure. Second, we perform progressive dilution experiments to test whether soluble and insoluble aggregates are in equilibrium relative to each other. Third, we perform incubation experiments at ideal physiological growth temperature for *E. coli* (i.e., 37 °C) to show that only part of the proteome is kinetically trapped in its native state, relative to aggregates. Tandem mass-tagging liquid-chromatography mass spectrometry (LC-MS) was carried out upon heating-and-cooling and incubation to identify proteins falling into individual categories.

We find that most *E. coli* proteins are kinetically protected from aggregation and that ATP-driven processes do not contribute to the existence of this trapping. Kinetic protection from aggregation applies to proteins across diverse biophysical characteristics, functions, and subcellular location. In addition, surprisingly, we also find that soluble and insoluble aggregates are in equilibrium relative to each other. Remarkably, we want to highlight that cytoplasmic and

periplasmic molecular chaperones are among the most intrinsically soluble and best kinetically trapped (relative to aggregates) members of the proteome. This result is consistent with chaperone's important cellular functions. Overall, this study highlights the essential role played by protein kinetic trapping from formation of soluble and insoluble aggregates under physiologically relevant conditions. Our results apply over timescales longer than *E. coli* doubling time and longer than typical bacterial-protein lifetimes. Kinetic trapping relative to aggregates maintains most of the proteome in a biologically competent state, thus minimizing the need for energetically expensive disaggregation.

2.3. RESULTS AND DISCUSSION

2.3.1 Development of an anaerobic heating and cooling procedure for conformational landscape exploration. We first aim to develop a thermal unfolding-refolding treatment to probe the extent of aggregation on a proteomewide scale without introducing heat-related modifications to protein sequence. An aliquot of S100 protein sample at 3 mg/mL was heated at 70°C for 1 hour and subsequently cooled to 22°C for 15 minutes for detection (Figure 2.1, panel A). At 70°C, most proteins in the *E. coli* proteome sample have reached their melting temperatures and unfolded.[48, 51] The procedure was carried out in a glove bag under anaerobic conditions, ca. ~300 ppm of O_2 (See Methods section). The before and after anaerobic heating-cooling samples were then analyzed with label-free LC-MS/MS for the presence of methionine oxidation and glutamine/asparagine deamination (See Methods section). We expected that the depletion of O_2 will minimize potential oxidative modifications at high temperatures. Three biological repeats were carried out for this experiment.

We were able to detect about 1,000 proteins in both the before and after heating-cooling samples. For each protein in a sample, we measured the count of intact and modified peptides (via oxidation and/or deamination) to calculate a percent modified peptide value (See Methods).

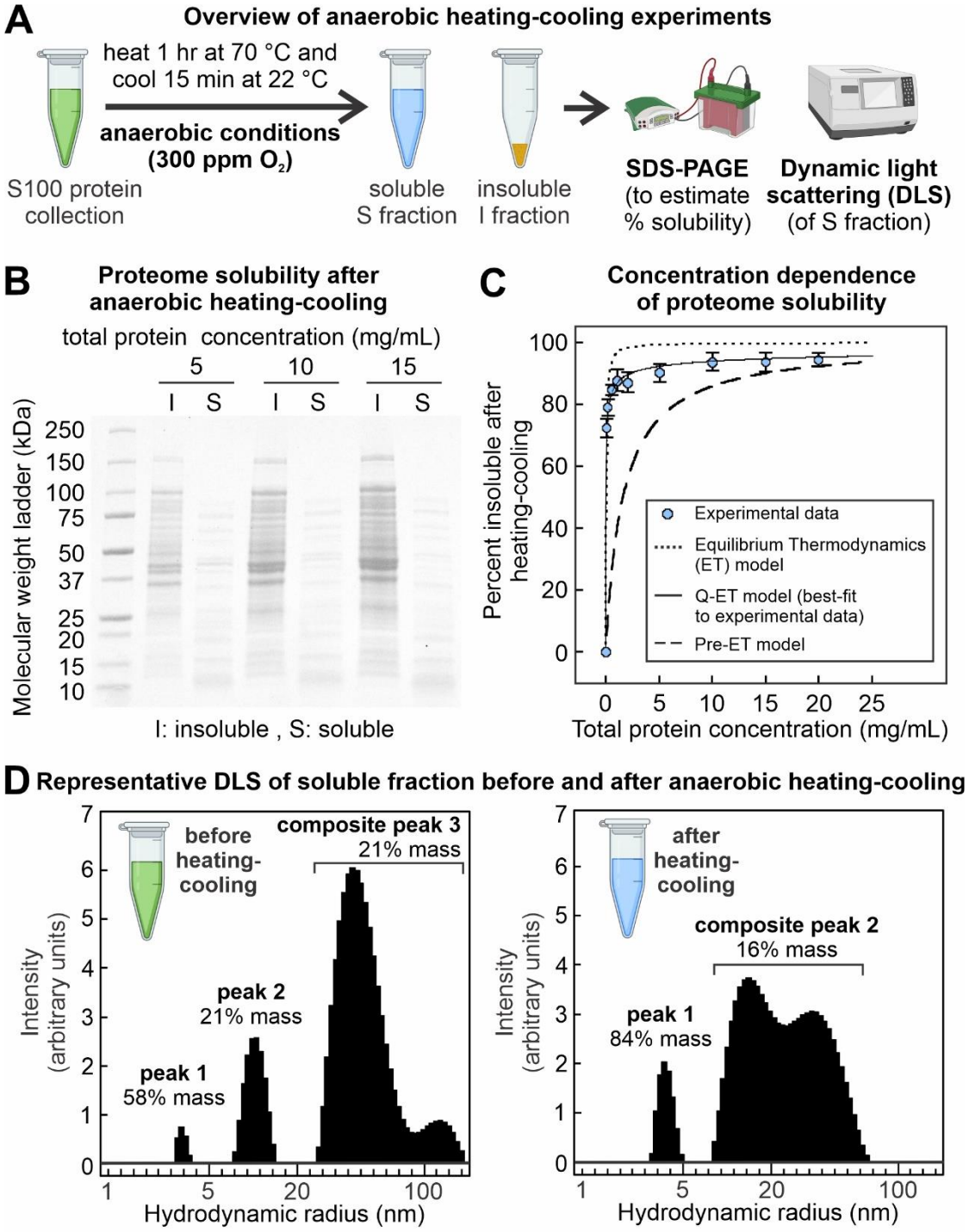


Figure 1. A majority of E. coli proteome forms insoluble aggregates under quasi-equilibrium thermodynamic upon heating-and-cooling. (A) Anaerobic heating-cooling of S100 samples allows for the exploration conformational landscapes for aggregation. (B) Majority of the E. coli S100 sample forms insoluble aggregate after heating-and-cooling at different total protein concentrations monitored by SDS-PAGE solubility assays (n = 5).

(C) Computational model fitting of densitometry analysis shows aggregation upon heating-cooling treatment is driven by quasi-thermodynamic. (D) DLS analysis reveals the soluble fraction after heating-cooling consists of small non-aggregated species (n = 3, total protein concentration before heating-cooling: 5 mg/mL; estimated soluble fraction concentration: 0.5 mg/mL). Panel (A) is created with BioRender.com.

2.3.2 .Most members of the *E. coli* proteome have free-energy landscapes bearing insoluble aggregates. The extent of aggregation from the S100 protein samples using the previously described anaerobic heating-cooling treatment were assessed at increasing total protein concentrations. Aliquots of S100 protein samples at different total concentrations (i.e., 0.1 to 25 mg/mL) were prepared and measured by BCA assay before anaerobic heating-cooling treatment. We performed this experiment on 5 biological repeats and characterized the samples with a combination of SDS-PAGE solubility assay and Dynamic Light Scattering (DLS)

The solubility, in terms of soluble proteins to insoluble aggregates, was estimated via SDS-PAGE analysis. The results shown in Figure 1 (panel B and C) indicate that increasing total protein concentrations led to an increase in insoluble aggregates after heating-cooling. The percent of insoluble aggregate plateaus at 95% at total protein concentration larger than 5 mg/mL. This result is consistent with the low proteome solubility of *E. coli* lysates observed in proteome-wide thermal profiling studies.[33, 48, 51]

To probe the apparent particle size in the soluble samples before and after heating-cooling treatment, DLS analysis was carried out and the resulting intensity decay data was analyzed with regularization method. We detected the presence of small (R_h < 5 nm) and larger particles (R_h = 8.4 ± 1.5 nm and 47 ± 2 nm) in the soluble S100 sample before treatment. By assuming uniform spherical shapes of all particles in solution, we roughly estimated the mass contributions of smaller particles to be 61 ± 1% across 5 biological repeats (Figure 1, panel A and Supplementary Table S1, Peak 1). We interpreted that the small particles of R_h < 5 nm are small soluble non-aggregated states, consistent with previous experimentally-measured hydrodynamic radii of many folded proteins.[30, 77] The hydrodynamic radius of one of the largest complexes in *E. coli*, the 70S ribosome (2.5 MDa), has been estimated to be 17 nm.[77-79] Upon preparation of the S100, it is possible that the large particles with R_h = 47 ± 4 nm are residual polysomes (i.e., complex of multiple ribosomes along a single RNA molecules, see SI Figure S2)[80] not removed from or

soluble aggregates induced by our S100 preparation procedures. Detailed discussion on DLS usage for complex samples can be found in SI Methods, Section 1.9.

Upon analysis of the remaining proteins in the soluble fraction (5-10% of the total protein concentration) after heating-cooling treatment, the sample shows smaller particles contributing majority of mass%, ca., 79 ± 3% (Figure 1, panel D and Supplementary Table S1, Peak 1), with negligible mass contributions from larger particles (ca., 2.6 ± 0.6%, Peak 3). Additional cumulant analysis of the DLS profile also suggested that the average R_h is also small (Supplementary Table S5). We propose that these large particles of R_h > 40 nm are due to large soluble aggregates, not from any native conformations due to the extensive thermal treatment. Overall, the DLS analysis of the post-thermal treatment shows that the proteins in this soluble fraction are in soluble non-aggregated states. This suggests that this small subset of proteins may be resistant to thermally induced aggregation (i.e., thermal-stable).

We further incubated the S100 samples after anaerobic heating-cooling for 20 hrs at 37°C (See SI Figure S6). No detectable change in solubility or conformation was observed via SDS-PAGE and DLS analysis. The results showed that there is no conformational change after additional incubation following anaerobic heating-cooling treatment under physiological conditions.

Overall, the results suggest that the conformational landscapes of the *E. coli* proteome contain many insoluble aggregates. These aggregates are extensively populated under conditions under thermodynamics at concentrations 10-20 times lower than that *in vivo* (ca. 200-300 mg/mL).[81] We hypothesized that these proteins may be more aggregation-prone when observed at *in vivo* concentrations. However, in reality, it is unlikely that these aggregates are observed in healthy cells. Previous works suggest that this lack of aggregation in the cells may be due to kinetic trapping on a proteome-wide scale.[34] Yet, no direct experiment was conducted to test this hypothesis. Another possible hypothesis is that systems involved in maintaining protein homeostasis (i.e., chaperones, disaggregases, and proteases) rescues proteins as these

aggregates are formed *in vivo*. This proposed scenario would rely significantly on these ATP-expensive systems. We will directly explore the nature of aggregation under our thermal treatment and how proteins in the *E. coli* proteome may avoid aggregation under physiologically relevant conditions in the following sections of this work.

2.3.3 .Concentration-dependent formation of insoluble aggregates is revealed upon equilibrium sampling of free-energy landscapes at high temperature. To further understand the mode of aggregation due to our anerobic heating-cooling treatment, we developed a computational model to fit the experimental solubility data (Figure 1, panel C). First, we made a first assumption that any deviation from the constant temperature stated above will lead to the immediate reintroduction of kinetic barriers and prevent any flux between different proteins concentration. During the heating process at 70°C, thermodynamic equilibrium is reached for each protein where most *E. coli* proteins are unfolded.[51] During the cooling process from 70°C to 22°C, there are kinetic deviations due to the presence of kinetic flux between soluble states and aggregates. Detailed derivations of this model are provided in the SI Methods, Sections 1.8.

This model, named the Quasi-Equilibrium Thermodynamics (Quasi-ET) model, where thermodynamics dominates, fits our experimental data well with 95% confidence. From the computational modeling results, we proposed that upon heating, proteins can sample conformational landscapes freely under thermodynamics equilibrium at high temperatures followed by a small amount of flux back to soluble states during cooling for downstream analysis. Therefore, insoluble aggregates must be more thermodynamically stable than soluble states.

To validate this hypothesis, we directly measured the thermodynamic stabilities of the insoluble aggregates relative to soluble states further support that the aggregation upon anaerobic heating-cooling was thermodynamically favored. Soluble proteins from the S100 sample before heating-cooling and insoluble aggregates generated after heating-cooling were unfolded in guanidine hydrochloride (GdnHCl) and the fraction of unfolding was monitored with either

tryptophan fluorescent emission or SDS-PAGE analysis (See SI Figure S7 and S8). The unfolding curves (Figure 2, panel A) were fitted with either the two-state model for soluble states[82-83] (i.e., native, unfolded, misfolded, or soluble aggregated states) or the linear polymerization model[58, 76, 84] for insoluble aggregates.

The fitting results show that the apparent unfolding Gibbs free energy of the insoluble aggregates from our proteomic sample after thermal treatment (app. $\Delta G°_{unfolding}$ = 54.4 ± 2.6 kJ/mol, Figure 2, panel B) is statistically higher (p = 0.012) than that of the soluble states in the S100 before thermal treatment (app. $\Delta G°_{unfolding}$ = 26.6 ± 4.4 kJ/mol). This result demonstrated that insoluble aggregates are more thermodynamically stable than the soluble states upon global unfolding of the E. coli proteome. Similarly, the unfolding m-value of the insoluble aggregates is also statistically higher (p = 0.079) than that of the soluble S100. We interpret that the insoluble aggregates display a higher change in solvent-accessible area compared to that of the soluble states.[82, 85] The measured app. $\Delta G°_{unfolding}$ and m-values for our complex protein samples (c.a., >2,000 proteins) as soluble states and insoluble aggregates are on the same order of magnitude with the previously reported thermodynamic values from one-protein denaturation experiments on native proteins and their associated aggregates.[58, 76] We also show that the unfolded states at 6 M GdnHCl of the soluble states and insoluble aggregates are nearly identical in behaviors by tryptophan fluorescent emission and thioflavin T assay (see SI Figure S9).

Overall, we propose that proteins in the *E. coli* proteome are thermodynamically driven to form insoluble aggregates upon heat-induced unfolding (i.e., from anaerobic heating-cooling). In addition, our observations also align with current views that protein aggregates are generally more thermodynamically stable comparing to the non-aggregated states.[86] Overall, these results directly support the extensive aggregation upon unfolding during anaerobic heating-cooling treatment and our predicted thermodynamic-driven model of aggregation (i.e., quasi-ET model, Figure 1, panel B and C). Furthermore, we believe that future interrogations are needed to

elucidate the individual contribution of each protein to the apparent thermodynamic values associated with the global unfolding of a complex proteome sample.

2.3.4 . Soluble and insoluble aggregates are in thermodynamic equilibrium relative to each other. We were interested in probing the relationships between soluble and insoluble aggregates from our previous experiments using a progressive dilution approach. These diluting experiments have been routinely used previously to study the equilibria between native states and fibrils of amyloidal proteins.[87-90] Similarly, we theorized that the soluble and insoluble aggregated states sampled for many *E. coli* proteins can exchange under thermodynamic equilibrium. To probe this hypothesis, we modified the derivations of the ET-model (See SI Methods, Section 1.15), originally to describe behaviors of soluble monomers and insoluble aggregated states under thermodynamic equilibrium, to provide theoretical basis to develop our subsequent experiments. Theoretically, if there is equilibrium thermodynamics between the soluble aggregates and insoluble aggregates, thermodynamics predicts that the chemical potentials of the two types of aggregates should be equal.

$$\mu_{\mathrm{A-soluble}} = \mu_{\mathrm{A-insoluble}} \tag{x}$$

where $\mu_{\text{A-soluble}}$ is the per mole chemical potential of the soluble aggregated state and $\mu_{\text{A-insoluble}}$ is the per mole chemical potential of the insoluble aggregated state. The activity of the aggregated phase is constant, while the activity of the soluble phase can be expressed in terms of the concentration of soluble proteins.

$$\mu_{\mathrm{M}}^{0} + \mathrm{RTln}\left(\gamma_{\mathrm{A-soluble}}\frac{\mathrm{c_{A-soluble}}}{\mathrm{c_{tot}}}\right) = \mu_{\mathrm{A}}^{0} + \mathrm{RTln}(\mathrm{a_A}) \ , \tag{x}$$

where $\gamma_{A-soluble}$ is the activity coefficient of the soluble aggregated proteins, $c_{A-soluble}$ is the concentration of soluble aggregated proteins, c_{tot} is the total concentration of the soluble solution which is approximately equal to the concentration of water, and $a_{\text{A-insoluble}}$ is the activity value of insoluble aggregated states. At constant temperature and pressure, the concentration of soluble

aggregated proteins will be constant. We denoted this concentration as the critical concentration of soluble aggregates, $[c_{A\text{-soluble}}]_{crit}$.

$$c_{A-soluble} = \frac{c_{tot}}{\gamma_M} \exp\left(\frac{\mu_A^0 + RT\ln(a_A) - \mu_M^0}{RT}\right) = \text{constant} = [c_{A-soluble}]_{crit} \ . \quad (4)$$

Additional formation of soluble aggregates will precipitate into insoluble aggregates to keep the concentration of soluble aggregates constant, according to equilibrium thermodynamics. From this theoretical background, we reason that dilution of the protein concentration in the soluble phase below this $[c_{A\text{-soluble}}]_{crit}$ will shift the system away from equilibrium, where $\mu_{A-soluble} < \mu_{A-insoluble}$. In Case I (Figure 3, panel A), if the system is in rapid thermodynamic equilibrium under a relevant timescale for equilibration, the insoluble aggregates will re-solubilize to maintain the constant $[c_{A\text{-soluble}}]_{crit}$. In Case II (Figure 3, panel B), if the system is under pre-equilibrium conditions, there will be no observable change in the protein concentrations in the soluble and insoluble fractions.

We carried out diluting experiments on insoluble aggregates formed under physiologically relevant conditions. Specifically, aliquots of S100 protein samples at 5 mg/mL were subjected to anaerobic incubation procedure, combined, re-aliquoted in new Eppendorf tubes, and centrifuged to isolate insoluble aggregates. Each sample was diluted with PBS buffer. Increasing diluting volumes (ca, 1:1 to 1:100), relative to the initial soluble phase volume were applied. The changes in insoluble aggregates relative to a non-diluted control and structural characteristics of the soluble phase upon dilution were monitored with SDS-PAGE and DLS analyses respectively.

The level of insoluble aggregates upon each diluting condition was compared to an undiluted insoluble aggregate control from the SDS-PAGE analysis (See SI Figure S11). The assay showed that the insoluble aggregates decreased as a higher dilution factor was applied (Figure 3, panel B and C). Upon reaching over 1:20 dilution factor, the percent of insoluble aggregate decreased significantly. In the regularization analysis of the DLS profiles from soluble fractions upon dilution, only one species with a large R_h value of 79 ± 5 nm was observed,

suggesting that insoluble proteins were re-solubilized as large soluble aggregates (See Figure 3, panel D and Supplementary Table S1, Peak 3). Additional cumulant analysis of the DLS profile also supports this observation (Rh = 68 ± 1 nm, see Supplementary Table S5).

The results align with the proposed Case I in Figure 3, panel A. where there exists a thermodynamic equilibrium between soluble and insoluble aggregates under physiologically relevant conditions. Furthermore, close inspection of the SDS-PAGE analysis (Figure 3, panel B) shows that all proteins in the insoluble aggregates resolubilized. We interpreted this to be that observed exchange under thermodynamic equilibriums happens to most proteins in our S100 protein collection. Previous work demonstrated that insoluble aggregates formed from thermal treatment of *E. coli* S100 samples were not amyloidal in structure using Transmission Electron Microscopy.[34] Here, we showed that most of our *E. coli* proteins in the non-amyloidal insoluble aggregates can exchange to large soluble aggregates under thermodynamics equilibrium at physiologically relevant conditions.

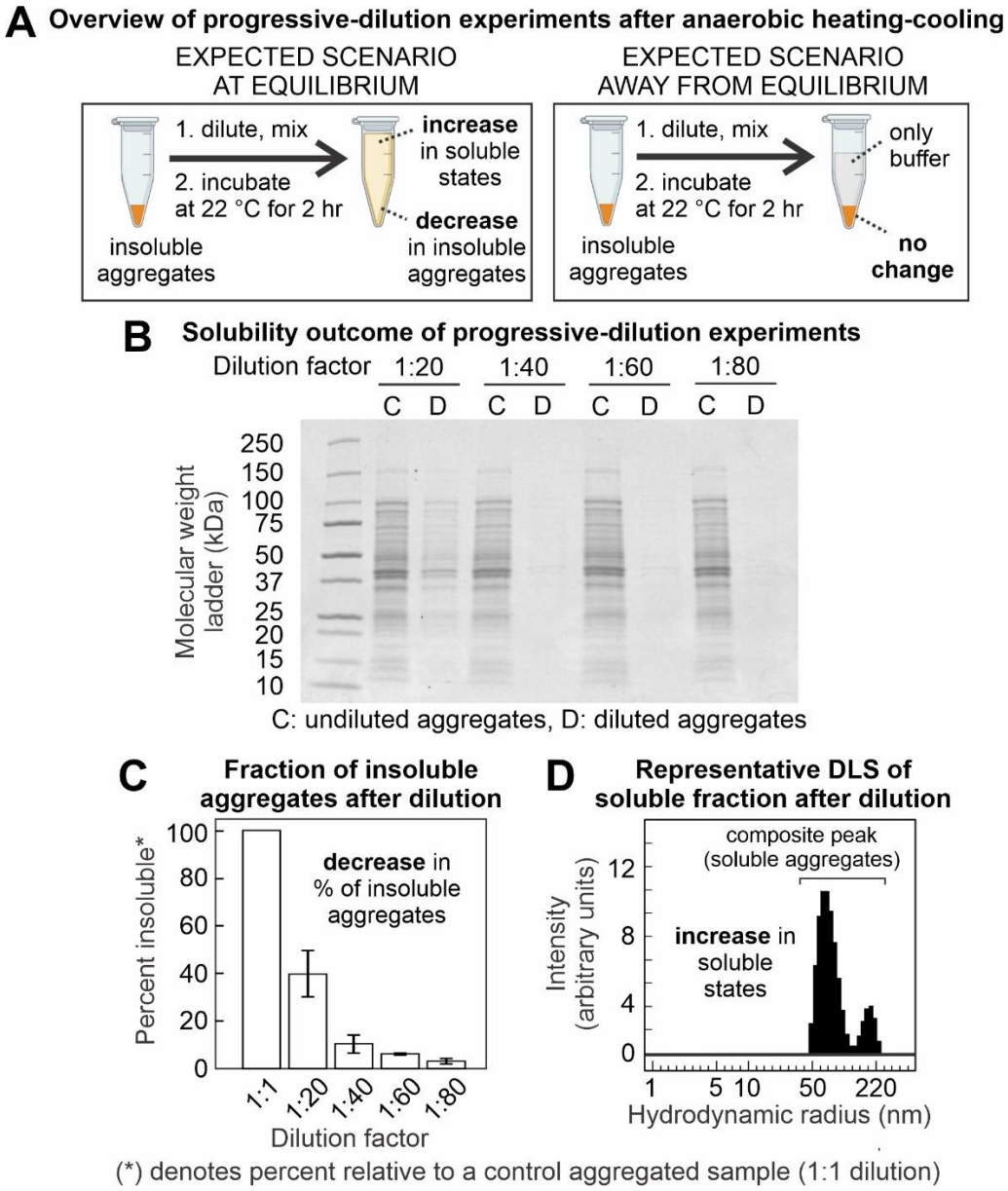


Figure 2. Thermodynamic equilibrium exists between insoluble and soluble aggregates formed from anaerobic heating-cooling experiments. (A) Expected scenarios of the progressive dilution experiments to reveal potential equilibrium exchange between aggregates. (B) SDS-PAGE analysis of insoluble aggregates shows a decrease in insoluble aggregates upon progressive dilution (n = 3). PBS buffer was used to dilute based on increasing dilution factors relative to the initial post-heating-cooling soluble phase volume removed (ca., 90 µL).

(C) Densitometry analysis confirms that most insoluble aggregates were resolubilized after 1:60 or more dilution. (D) DLS analysis shows that resolubilized insoluble aggregates exchanged to large soluble aggregated in solution (n = 3, total protein concentration upon heating-cooling at 37°C: 5 mg/mL). Panel (A) is Created with BioRender.com.

2.3.5 . The *E. coli* proteome is kinetically protected from aggregate formation under physiologically relevant conditions. We assessed the extent of the proposed kinetic control on a proteome-wide scale by carrying out incubation experiments with our S100 samples at increasing total protein concentrations. Aliquots of S100 samples ranging from 0.1-25 mg/mL were analyzed with BCA assay for concentration. All samples were then incubated for 20 hours at 37°C (i.e., physiologically relevant conditions) in the same set up developed for the anaerobic heating-cooling treatment (Figure 4, panel A). We performed the incubation experiment on 5 biological repeats and characterized the samples with a combination of SDS-PAGE solubility assay and DLS.

Solubility assays with SDS-PAGE revealed that over 90% of the total proteins were soluble upon incubation across different total protein concentrations (Figure 4, panel B) while less than 10% of the remaining total proteins formed insoluble aggregates. Regularization analysis of the DLS profile in the soluble fraction after incubation showed that over 85 ± 3% of the mass contribution comes from small particles of R_h = 4 ± 0.3 nm (Figure 4, panel C and Supplementary Table S2, Peak 1) with negligible amounts of soluble aggregates. Additional cumulant analysis of the DLS profile also supports these observations (Supplementary Table S6).

We also carried out these experiments in the presence of cellular concentrations of ATP to test whether the observed high solubility as non-aggregated states from the anaerobic incubation treatment may be due to active endogenous de-aggregation and chaperone systems in our S100 protein collection. Detailed discussion regarding these control experiments is in SI Discussion, Section 2.4. In short, the results from these experiments suggest that ATP-dependent systems, including the Hsp70 chaperone system, do not contribute to the observed kinetic protection of *E. coli* proteome as soluble non-aggregated states relative to aggregation at physiologically relevant conditions (See SI Figure S__).

Finally, we carried out the same diluting experiment on anaerobically heated-cooled insoluble aggregates from Figure 3 on those formed from anaerobic incubation from Figure 5.

Similarly, upon increasing dilution, more insoluble aggregates resuspended into large soluble aggregates as shown in SDS-PAGE solubility assay and DLS analysis (Figure 5, panel A and B and Supplementary Table S6). Therefore, we extend our kinetic protection of the *E. coli* proteome as non-aggregated states relative to soluble aggregates as well. Furthermore, the lack of soluble aggregates detected in our previous incubation experiments also suggests that these soluble aggregates are less populated compared to insoluble aggregates at physiologically relevant conditions.

In summary, upon anaerobic incubation at 37° C, we observed minimal flux from small non-aggregated states to both soluble and insoluble aggregates. We concluded that most of the total proteins in our proteome sample are kinetically protected as small soluble non-aggregated states under physiologically relevant conditions relative to soluble and insoluble aggregates previously observed in Figure 1. This kinetic protection is on a timescale much longer than that of *E. coli* doubling time [91] and is not due to any active endogenous ATP-dependent folding and degradation systems. Since our results in previous experiments suggest that aggregates are more thermodynamically stable, this novel demonstration of kinetic stability on a proteomewide scale may hold significant biological implications for *E. coli*.

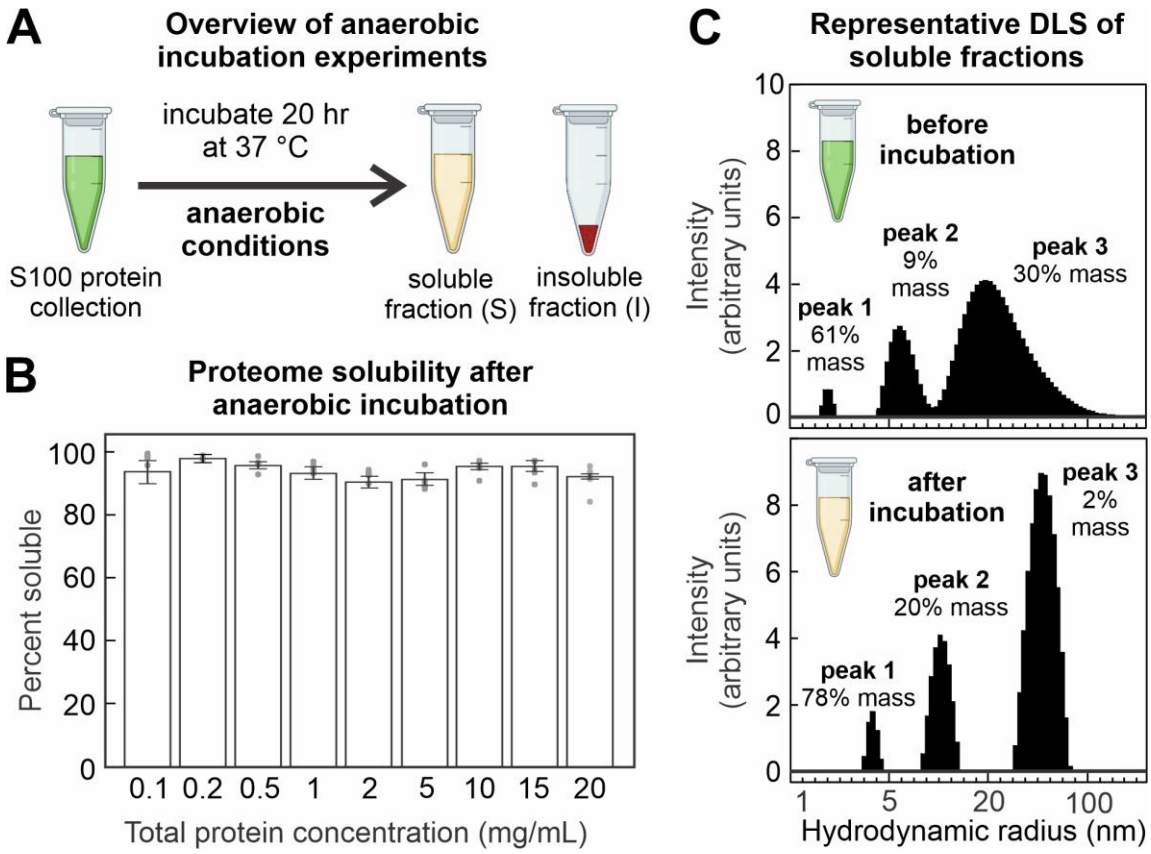


Figure 3. E. coli proteome are kinetically protected from aggregation under physiologically relevant conditions. (A) Anaerobic incubation of S100 samples 37°C for 20 hours to probe proteome-wide kinetic protection from aggregation. (B) Majority of the E. coli S100 sample are kinetically protected from insoluble aggregate upon incubation at different total protein concentrations via SDS-PAGE solubility assay (n = 5). (C) DLS analysis reveals the soluble fraction after incubation consists of small non-aggregated species (n = 3, total protein concentration before incubation: 5 mg/mL; soluble fraction concentration after incubation: 4.5 mg/mL). Panel (A) is created with BioRender.com.

2.3.6 . Kinetic protection from aggregation is pervasive across the *E. coli* proteome. To probe the extent of this proposed kinetic protection as non-aggregated states in the *E. coli* proteome, we correlated the solubilities of all proteins between quasi-ET control from anaerobic heating-cooling treatment and kinetic control from anaerobic incubating treatment. We chose a bottom-up proteomic approach to identify proteins and their abundances in a sample before and after treatment. 5 biological repeats were generated for each treatment and samples were analyzed with LC-MS/MS experiments with tandem-mass tagging approach (See SI Method, Sections 19-21 and SI Figure S17). The LC-MS/MS analysis was able to identify 1,700 *E. coli* proteins from each treatment data set for downstream analysis. We calculated a percent solubility value for each protein from the presence of peptides in the soluble and insoluble fractions from each experiment and an average of solubility values from 5 biological repeats.

Here, we chose to focus on a cohort of 1197 proteins, under standards reported in SI Methods, Section 1.21, that were present in both data sets generated from different treatments. We plotted the values for percent solubility of each protein to correlate and assess the differences in solubility and aggregation under different controls (Figure 6). Total average percent solubilities were calculated as 17% and 70% for proteins under quasi-ET and kinetic control respectively to serve as limits to describe four quadrants on our scatter plot. Overall, 562 out of 832 proteins have low solubility of <17% (i.e., high propensity to form insoluble aggregates; Figure 6, panel A) under quasi-ET control. This result aligns with our previous results in Figure 1, that a majority of proteins formed insoluble aggregates after anaerobic thermal treatment. Similarly, 602 out of 832 proteins have high solubility of <70% after anaerobic incubation under physiologically relevant conditions, (i.e., high kinetic protection as non-aggregated soluble states; Figure 6, panel A), consistent with the results in Figure 5. Detailed discussion on the structural and functional composition of our subset of the *E. coli* proteome is in SI Discussion, Section 2.6. For each protein in a quadrant, we evaluate the observed solubilities relative to subcellular location.

562 out of 1,197 proteins (ca., 47%), are in quadrant I. This quadrant consists of mainly cytoplasmic proteins and displays an important inverse solubility relationship. Specifically, these proteins have low solubility under quasi-ET control, but high solubility under kinetic control. This result directly aligns with our hypothesis that many *E. coli* proteins at risk of aggregation under thermodynamics, in turn, are kinetically protected as soluble non-aggregated states under physiologically relevant conditions (Figure 6, panel C).

264 proteins (ca., 22%) in quadrant II have high solubility under both quasi-ET and kinetic controls. These proteins are resistant to thermodynamically driven aggregation, so they may not require kinetic protection for solubility. Interestingly, we found that most periplasmic proteins are found in this quadrant besides cytoplasmic proteins (see SI Figure S19, Panel B). We propose that proteins within this quadrant have intrinsic resistance to aggregation, lessening the reliance on kinetic protection to remain soluble.

318 proteins (ca., 27%) in quadrant III have low solubility under both solubility controls, suggesting that proteins in this subgroup are not kinetically protected under physiologically relevant conditions from aggregation (Figure 6, panel C). These insoluble proteins in this subset consist of membrane, ribosomal, and cytoplasmic proteins (see SI Figure S19). However, it is important to note that the observed aggregation is under 10% of the total protein concentration after incubation (Figure 3, panel B). Specifically, amongst the small number of membrane proteins detected in our sample (ca., 81, see SI Figure S19, panel C), over 60% fall were detected in Quadrant III. We also identified 51 ribosomal proteins within this low solubility quadrant (see SI Figure S19, panel D). The low quasi-ET solubility is consistent with current findings that ribosomal proteins tend to form insoluble aggregates upon thermal denaturation.[20, 48] Potentially, due to the low concentrations divalent magnesium ion concentration (ca., 0.8 mM), it is unlikely that ribosomal proteins exist as stable S70 complex under our conditions.[92-94] As a result, due to the lack of native conditions provided within ribosome structures or cell membranes, these membrane

and ribosomal proteins may be more prone to aggregation under our conditions compared to cytoplasmic or periplasmic proteins (Figure 6, panel A and SI figure xx).

52 proteins (ca., 4%) 264 in quadrant IV have higher quasi-ET solubility value than the associated kinetic solubility. The lack of protein representation in this quadrant suggests that it is unlikely to observe both high solubility at non-physiological conditions and low solubility at physiological conditions. Therefore, no conclusion was made for proteins in this quadrant.

Overall, we found that our proposed kinetic protection as soluble non-aggregated states apply to over 50% of the 1,197 identified proteins in our *E. coli* proteome sample.

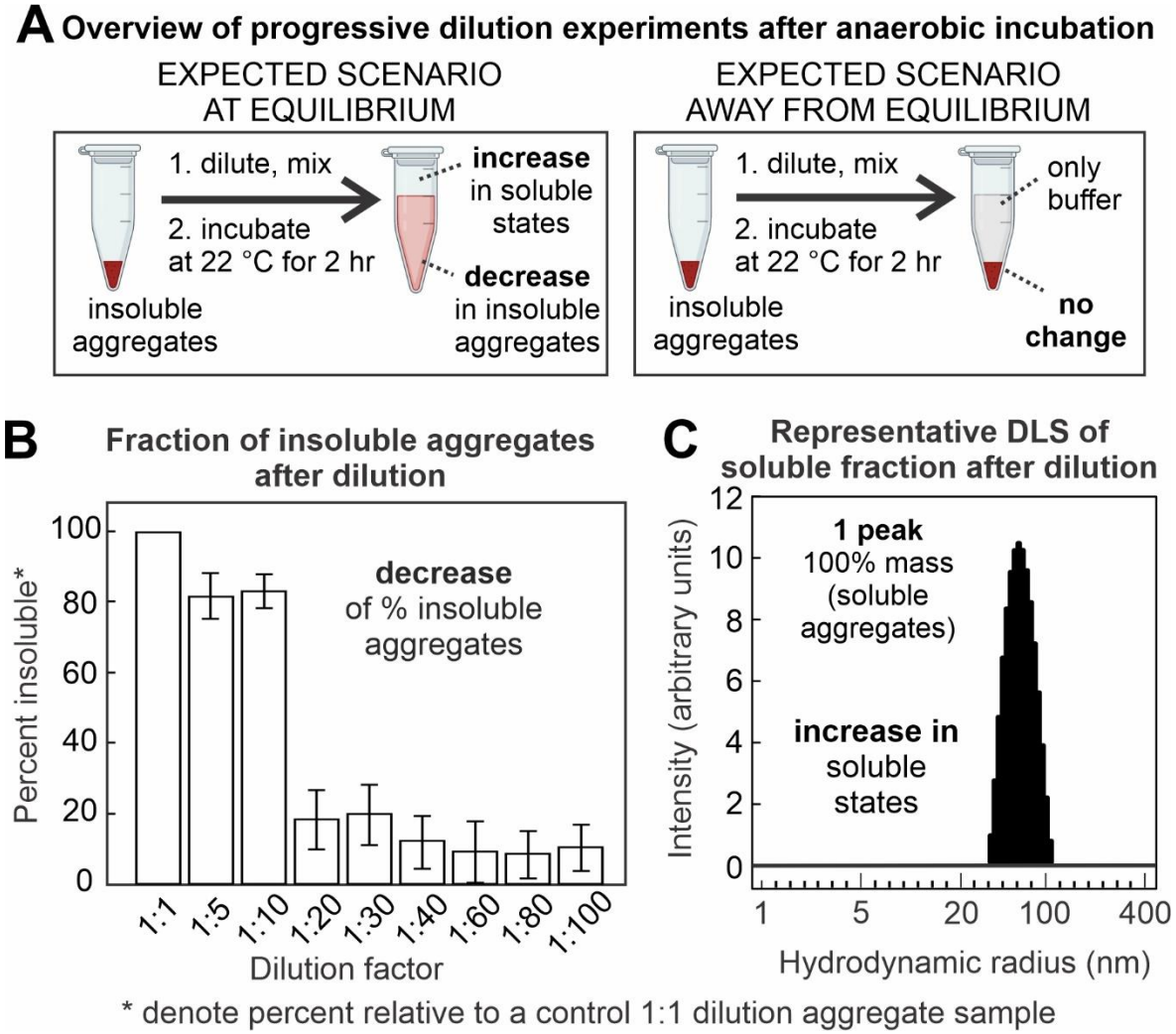


Figure 4. Thermodynamic equilibrium exists between insoluble and soluble aggregates from anaerobic incubation experiments. (A) SDS-PAGE analysis of insoluble aggregates shows a decrease in insoluble aggregates upon progressive dilution. (B) DLS analysis shows that the insoluble aggregates exchanged to large soluble aggregates in solution (total protein concentration for incubation at 37°C: 5 mg/mL; estimated soluble fraction volume removed from incubation at 37°C: 180 µL). Panel A is created with BioRender.com.

2.3.7 .The generality of kinetic protection from aggregation relative to protein properties.

We probe the relationship between protein solubilities and biophysical properties or cellular functions. In general, we found that proteins have the inverse relationship between Quasi-ET and kinetic solubilities as seen in Quadrant I (Figure 6, panel A) with no observable correlation to primary structures (i.e., molecular weights, amino acid contents, copy numbers, isoelectric points and cofactors; see SI Figure S20 - S23 and S27), tertiary structures (i.e., number of domains, see SI Figure S24), quaternary structures (i.e., subunits, see SI Figure S28 and S29). Overall, we conclude that the proteome-wide scale kinetic protection from aggregation applies to many proteins in *E. coli*, across diverse structural properties.

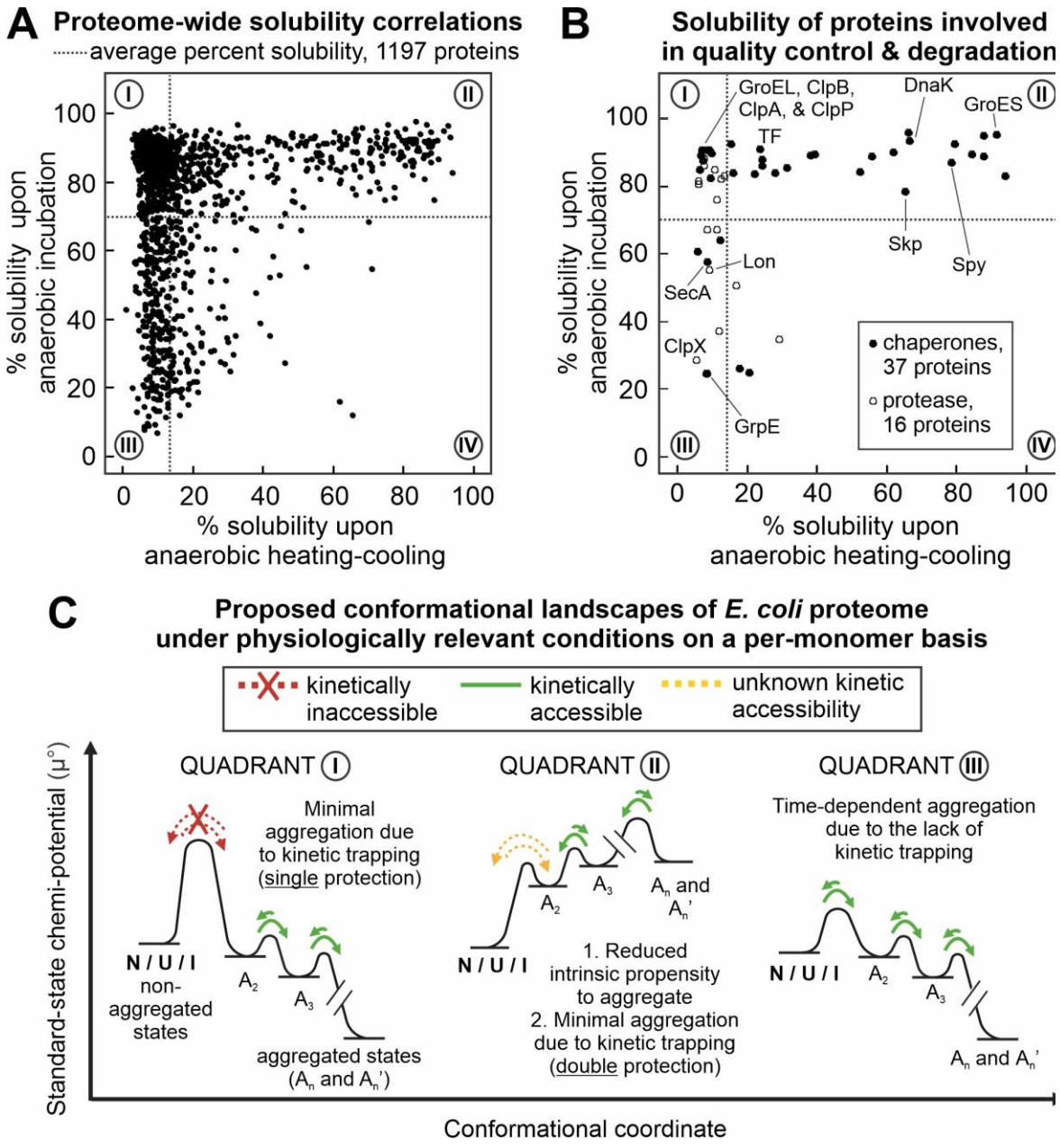


Figure 5. More than 50% of *E. coli* proteins are aggregation-prone, yet kinetically protected from aggregation. (A) Diagram illustrating protein solubility deduced from LC-MS analysis of anaerobically heated-cooled and physiological temperature (37 °C) incubated S100 proteome. Data pertain to 833 *E. coli* proteins (n = 3). (B) Diagram illustrating the % solubility of *E. coli* proteins involved in quality control and degradation. (C) Qualitative representation of free-energy landscape of proteins (on a per-monomer basis) in quadrants I, II, and III of panel A. These plots highlight distinct principles underlying protection from aggregation, for proteins belonging to the different quadrants.

2.3.8 . Protein solubility and kinetic stability correlates with cellular functions. Next, we assigned proteins to specific cellular functions according to the Gene Ontology database (see SI Figure 32).[95-96] We also assign essentiality to our protein dataset based on previous characterizations from the KEIO collection. Satisfyingly, on a global scale, many essential proteins populate Quadrant I across different cellular function groups, suggesting a potential link between kinetic protection and protein essentiality. Our results show that most proteins (c.a., 54 proteins) involved in central carbohydrate metabolic pathways, such as glycolysis, tricarboxylic acid cycle, and oxidative pentose phosphate pathway, are kinetically protected from aggregation (see SI Figure S32, part 1, panel A, B and E).[97-98] In addition, we also identified 21 out of 23 essential tRNA synthetases and many proteins involved in DNA transcription, important in protein biosynthesis alongside with molecular chaperones, to be kinetically stable (see SI Figure 32, part 2, panel I-K).[99-100] We would also like to note that these proteins are enriched in the previously determined high solubility domain fold-types, such as class II aaRS and biotin synthetase, and adenine dinucleotide α-hydrolase-like (see SI Figure 27, panel C and D). On the other hand, proteins associated with cellular anaerobic and aerobic responses, all assigned as non-essential, are aggregation-prone under our conditions (see SI Figure 32, part 2, panel N). Overall, we conclude that kinetic stability may be an important mechanism to protect proteins involved in important cellular functions, such as metabolism/catabolism and protein translation and transcription, from aggregation throughout the late stages of protein's

2.3.9 . Molecular chaperone is amongst the most soluble protein groups across the *E. coli* proteome. Next, we reported the associated solubilities and the extent of kinetic protection for proteins important to the maintenance of protein homeostasis in *E. coli*. In response to proteotoxic stress, upregulation of molecular chaperones and proteases as part of the heat

shock response promotes protein folding and minimizes aggregation.[9] We hypothesized that they must be highly soluble and kinetically protected to perform their important biological functions.

Major housekeeping chaperone systems responsible for assisting de novo folding of a majority of proteins in the cytoplasm for *E. coli* include two groups: 1) the ATP-dependent chaperone systems of this group include GroEL/S (Hsp60/10),[19] DnaK / DnaJ / GrpE,[17] and HptG (Hsp90).[68, 101]; 2) the ATP-independent co-translational chaperone Trigger Factor (TF). We found all these chaperones, except for DnaJ, to be either intrinsically soluble (i.e., not aggregation-prone under physiologically relevant conditions in Quadrant II) or kinetically protected (Quadrant I) as non-aggregated states (Figure 6, panel B). In the case of DnaJ, we hypothesize that the depletion of available zinc ions, which has been suggested to have destabilizing effects to the protein's structure[102-103], may contribute to its low solubility even under physiologically relevant conditions.

Similarly, we identified in Quadrant II all major ATP-independent periplasmic chaperones (HdeA, HdeB, FkpA, DsbA, DsbC, and Skp)[104], previously found to have high thermostabilities [33, 48], to have high intrinsic resistant to aggregation. Finally, SurA, known for assisting the folding of outer-membrane porins[105], is kinetically protected (Quadrant I).

Amongst the major cytoplasmic ATP-dependent proteases, ClpB, known to be the primary machinery responsible for the rescue of functional proteins form aggregation in *E. coli* cytoplasm when acting cooperatively with Hsp70[11], relies on kinetic protection for solubility (Quadrant I) under physiologically relevant conditions (Figure 6, panel B). However, other major *E.* coli proteases such as Lon and ClpX are found in Quadrant III with lower solubility under kinetic protection.[15]

Overall, consistent with our hypothesis, we found that many proteins essential for stress-response rely on kinetic protection to remain soluble and non-aggregated under physiologically relevant conditions. This result implies that these "helper" proteins must be non-aggregation prone under physiological conditions to carry out their important biological functions.

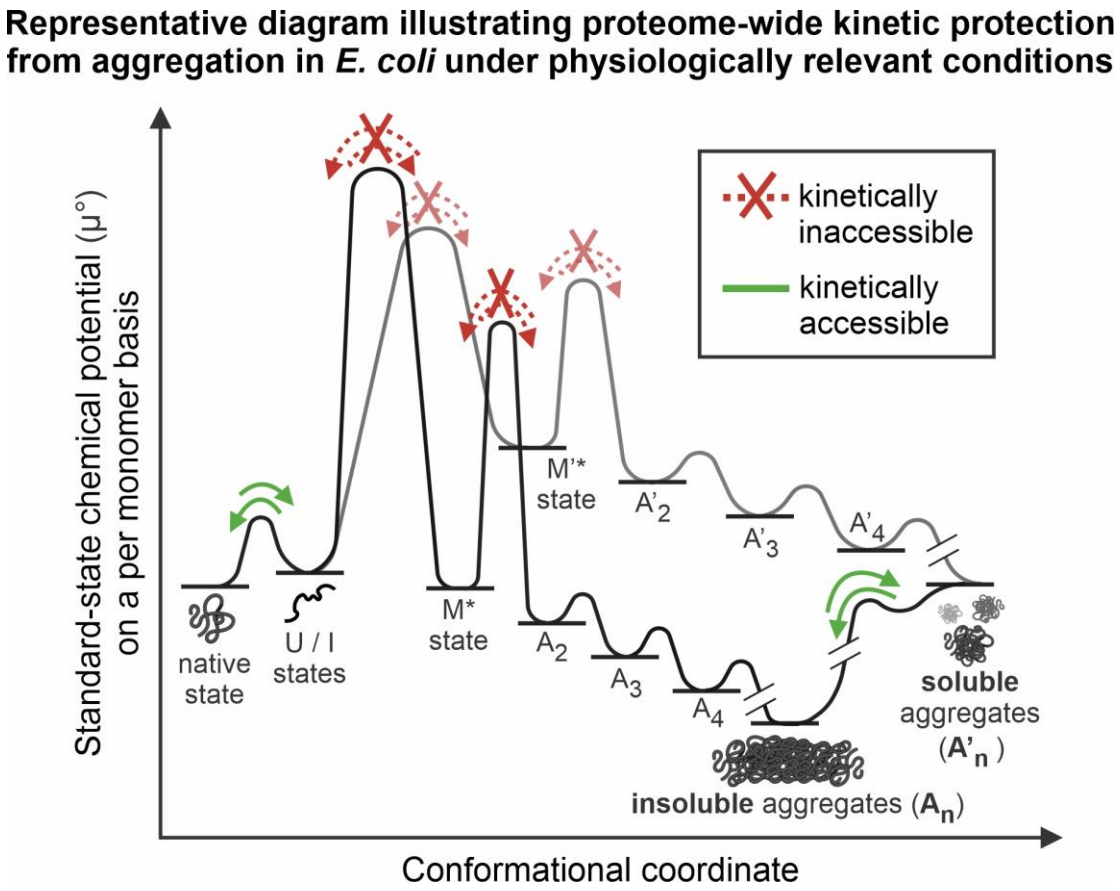


Figure 6. Non-aggregated states are kinetically trapped from soluble and insoluble aggregates under physiologically relevant conditions. Proposed conformational landscape on a per-monomer-basis of many *E. coli* proteins under kinetic protection as non-aggregated states from soluble and insoluble aggregates under physiologically relevant conditions. U / I: unfolded / intermediate; M* or M'*: misfolded states; A_n and A'_n: aggregates with n-th monomers.

2.4. Experimental approach

For our proteome-wide experiments, *E. coli* cells (A19, strain K-12) are grown in Luria Broth to the start of growth phase (OD_{600} 0.8). Cells are then lysed with high pressure and the clarified cell lysate is subjected to high-speed centrifugation (100,000 g) to deplete intact 70S ribosomes.[20, 106-107] The ribosomal-depleted sample is referred to as an S100 protein collection for downstream experiments. Similar to previous reports, we estimated 25-30% of were removed via this ultracentrifugation step with BCA assay.[20] Our S100 protein collection is predicted to contain mainly cytoplasmic proteins (ca., 995) and has low representation of membrane proteins (see SI Discussion, Section 2.6).

With this representative *E. coli* proteome sample, we first aim to explore the free-energy conformational landscape for the extent of aggregation on a proteome-wide scale. We chose a thermal unfolding-refolding approach to provide thermal energy for proteins to overcome energy barriers and unfold while keeping the total protein concentration constant throughout our experiments. We theorized that during this procedure, proteins can explore conformational landscapes freely, which was limited by kinetic barriers under physiologically relevant conditions. We developed an anaerobic heating-cooling treatment to prevent introducing heat-related modifications to the protein samples. Following the developed thermal treatment, each sample is characterized for solubility and aggregation with SDS-PAGE assay and Dynamic Light Scattering (DLS). With computational modeling, we aim to see whether the aggregation process in this experiment can be described with thermodynamics.

Then, we proceed to verify the concept of kinetic protection on a proteome-wide scale as non-aggregated states relative to the aggregation. Here, the representative proteome sample was incubated at physiologically relevant conditions on timescales longer than the organism's doubling time. The sample was characterized with the same methods for solubility and aggregation.

Finally, we explore how pervasive the proposed kinetic protection is across the *E. coli* proteome. We also we aim to see the extent in which proteins and their associated characteristics are 1) kinetically protected under physiologically relevant conditions on a proteome-wide scale, and 2) at risk of thermodynamic-driven aggregation under thermal treatment. Protein identities from either anaerobic thermal treatment or incubation treatment were obtained through analyses with liquid chromatography / tandem mass spectrometry (LC/MS-MS) with a 16-plex tandem-mass tagging (TMT) approach.[108] Then, we use bioinformatics to probe the relationships between different types of protein solubility (i.e., under quasi-ET or kinetic control) and their associated characteristics (i.e., biophysical, co-factors, chaperone dependence, subcellular locations, and cellular processes, etc.).

2.5. Conclusions and Outlook

There are two caveats to the current experimental data and interpretations of this study. First, it is important to acknowledge that the biophysical techniques used in this work cannot provide resolutions to the structural characteristics of soluble non-aggregated states after incubation experiments. These soluble states may include native, unfolded, misfolded, or small oligomeric states that are not resolvable via DLS characterizations. As shown in many proteome-wide scale works, proteins may misfold under physiological conditions.[20] Though we proposed a kinetic protection from aggregated states, our data does not identify where this kinetic trapping is amongst the possible soluble states. Recent coarse-grain computational studies of previously published experimental data suggest that some proteins may form long-lived soluble misfolded states, kinetically trapped from both the native states and aggregation.[55, 109] These misfolded states are also suggested to possess high structural similarity to that of the native states, resulting in their escape from chaperone systems. This is a possible scenario where soluble proteins in our

work are kinetically trapped as these described misfolded states from aggregated landscapes (Figure 7). Further structural studies are needed to resolve the conformations of soluble proteins that are kinetically protected from aggregation in our experiments.

In conclusion, we report a proteome-wide kinetic protection that significantly increases the solubility as small states for many *E. coli* proteins, especially for complex “non-model” proteins, protecting them from thermodynamically stable aggregation, as illustrated in the proposed chemical-potential conformational landscape of Figure 7. This proteome-wide phenomenon may be a strategy employed by the cells to ensure protein homeostasis after synthesis and folding, avoiding reliance on ATP-expensive refolding, disaggregation, and degradation pathways. However, this does not guarantee an aggregation-free environment as a small percent of insoluble aggregates were observed in our experiments despite this proposed kinetic protection. Instead of viewing this result as a pathological aspect, this result highlights that it requires the synergistic actions of cellular features, such as transcription/translation on the ribosome, molecular chaperones, protein degradation, and the native crowded environment in different cell compartments, to ensure an aggregation-free proteome *in vivo*.[10, 110]

We also report that insoluble aggregates across the proteome can exchange with soluble aggregates under physiologically relevant conditions under thermodynamics equilibrium. This result supports the notion that insoluble aggregates may be a source of soluble aggregates *in vivo*. Though further investigation is needed to explore this phenomenon, We suggest that the observed phenomenon may have important implications in the studies of the possible interplay and toxicity of soluble and insoluble aggregates in many proteinopathies.[111]

2.6. Associated Content

Supporting Information is available free of charge at [URL TBD].

2.7. Author Information

2.7.1 Corresponding Author

Silvia Cavagnero - Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA.

2.7.2 Other Authors

Hung H. Dang – Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA.

Lucas E. Bartel – Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA.

Greg Sabat – Biotechnology Center Mass Spectrometry Core Facility, University of Wisconsin - Madison, 425 Henry Mall, Madison, Wisconsin, 53706, USA.

Yuzhu Liu – Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA.

Mridula Srivathsan – Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA.

Kevin E. England – Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA.

2.8. Acknowledgements

We thank the National Science Foundation (NSF) for funding (grants MCB-2124672 and CBET 1912259 to S.C.). We are grateful for Siyu Li and Meranda Masse for insightful discussions and Heather Allaman for providing the S30 cell-free extract for DLS control experiments. Finally, we would like to thank BioRender.com for providing the tools to create publication quality figures of this Chapter and manuscript.

2.9. References

1. Zhou, H.-X.; Rivas, G.; Minton, A. P., Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. *Annu. Rev. Biophys.* **2008,** *37* (1), 375-397.
2. Elcock, A. H., Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr. Opin. Struct. Biol.* **2010,** *20* (2), 196-206.
3. Christiansen, A.; Wang, Q.; Samiotakis, A.; Cheung, M. S.; Wittung-Stafshede, P., Factors Defining Effects of Macromolecular Crowding on Protein Stability: An in Vitro/in Silico Case Study Using Cytochrome c. *Biochemistry* **2010,** *49* (31), 6519-6530.
4. Ciryam, P.; Morimoto, R. I.; Vendruscolo, M.; Dobson, C. M.; O'Brien, E. P., *In vivo* translation rates can substantially delay the cotranslational folding of the *Escherichia coli* cytosolic proteome. *Proceedings of the National Academy of Sciences of the United States of america* **2013,** *110* (2), E132-E140.
5. Netzer, W. J.; Hartl, F. U., Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **1997,** *388* (6640), 343-349.
6. Addabbo, R. M.; Dalphin, M. D.; Mecha, M. F.; Liu, Y.; Staikos, A.; Guzman-Luna, V.; Cavagnero, S., Complementary Role of Co- and Post-Translational Events in De Novo Protein Biogenesis. *J. Phys. Chem. B* **2020,** *124* (30), 6488-6507.
7. Calloni, G.; Chen, T.; Schermann, Sonya M.; Chang, H.-c.; Genevaux, P.; Agostini, F.; Tartaglia, Gian G.; Hayer-Hartl, M.; Hartl, F. U., DnaK Functions as a Central Hub in the E. coli Chaperone Network. *Cell Rep.* **2012,** *1* (3), 251-264.
8. Kerner, M. J.; Naylor, D. J.; Ishihama, Y.; Maier, T.; Chang, H.-C.; Stines, A. P.; Georgopoulos, C.; Frishman, D.; Hayer-Hartl, M.; Mann, M.; Hartl, F. U., Proteome-wide Analysis of Chaperonin-Dependent Protein Folding in Escherichia coli. *Cell* **2005,** *122* (2), 209-220.
9. Balchin, D.; Hayer-Hartl, M.; Hartl, F. U., In vivo aspects of protein folding and quality control. *Science* **2016,** *353* (6294), aac4354.
10. Reichmann, D.; Suss, O., Protein plasticity underlines activation and function of ATP-independent chaperones. *Frontiers in Molecular Biosciences* **2015,** *2*.
11. Koubek, J.; Schmitt, J.; Galmozzi, C. V.; Kramer, G., Mechanisms of Cotranslational Protein Maturation in Bacteria. *Frontiers in Molecular Biosciences* **2021,** *8*.
12. Maurizi, M. R., Proteases and protein degradation in Escherichia coli. *Experientia* **1992,** *48* (2), 178-201.
13. Rusilowicz-Jones, E. V.; Urbe, S.; Clague, M. J., Protein degradation on the global scale. *Molecular Cell* **2022,** *82* (8), 1414-1423.
14. Chen, W. X.; Smeekens, J. M.; Wu, R. H., Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chemical Science* **2016,** *7* (2), 1393-1400.
15. Truscott, R. J. W.; Schey, K. L.; Friedrich, M. G., Old Proteins in Man: A Field its Infancy. *Trends in Biochemical Sciences* **2016,** *41* (8), 654-664.
16. Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H., The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences* **1961,** *47* (9), 1309-1314.
17. Anfinsen, C. B., Principles that govern the folding of protein chains. *Science* **1973,** *181* (4096), 223-30.
18. Varela, A. E.; Lang, J. F.; Wu, Y.; Dalphin, M. D.; Stangl, A. J.; Okuno, Y.; Cavagnero, S., Kinetic Trapping of Folded Proteins Relative to Aggregates under Physiologically Relevant Conditions. *The Journal of Physical Chemistry B* **2018,** *122* (31), 7682-7698.

19. Niwa, T.; Ying, B.-W.; Saito, K.; Jin, W.; Takada, S.; Ueda, T.; Taguchi, H., Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proceedings of the National Academy of Sciences* **2009,** *106* (11), 4201-4206.
20. Mateus, A.; Bobonis, J.; Kurzawa, N.; Stein, F.; Helm, D.; Hevler, J.; Typas, A.; Savitski, M. M., Thermal proteome profiling in bacteria: probing protein state in vivo. *Mol. Syst. Biol.* **2018,** *14* (7), e8242.
21. Kwon, S.; Jung, Y.; Lim, D., Proteomic analysis of heat-stable proteins in Escherichia coli. *BMB Rep* **2008,** *41* (2), 108-11.
22. LaBreck, C. J.; May, S.; Viola, M. G.; Conti, J.; Camberg, J. L., The Protein Chaperone ClpX Targets Native and Non-native Aggregated Substrates for Remodeling, Disassembly, and Degradation with ClpP. *Frontiers in Molecular Biosciences* **2017,** *4*.
23. Katikaridis, P.; Bohl, V.; Mogk, A., Resisting the Heat: Bacterial Disaggregases Rescue Cells From Devastating Protein Aggregation. *Frontiers in Molecular Biosciences* **2021,** *8*.
24. Sweeny, E. A.; Shorter, J., Mechanistic and Structural Insights into the Prion-Disaggregase Activity of Hsp104. *Journal of Molecular Biology* **2016,** *428* (9, Part B), 1870-1885.
25. Hipp, M. S.; Park, S.-H.; Hartl, F. U., Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.* **2014,** *24* (9), 506-514.
26. Nissley, D. A.; Jiang, Y.; Trovato, F.; Sitarik, I.; Narayan, K. B.; To, P.; Xia, Y.; Fried, S. D.; O'Brien, E. P., Universal protein misfolding intermediates can bypass the proteostasis network and remain soluble and less functional. *Nat. Comm.* **2022,** *13* (1), 3081.
27. To, P.; Whitehead, B.; Tarbox, H. E.; Fried, S. D., Nonrefoldability is Pervasive Across the E. coli Proteome. *Journal of the American Chemical Society* **2021,** *143* (30), 11435-11448.
28. To, P.; Xia, Y.; Lee, S. O.; Devlin, T.; Fleming, K. G.; Fried, S. D., A proteome-wide map of chaperone-assisted protein refolding in a cytosol-like milieu. *Proceedings of the National Academy of Sciences US* **2022,** *119* (48), e2210536119.
29. Vecchi, G.; Sormanni, P.; Mannini, B.; Vandelli, A.; Tartaglia, G. G.; Dobson, C. M.; Hartl, F. U.; Vendruscolo, M., Proteome-wide observation of the phenomenon of life on the edge of solubility. *Proceedings of the National Academy of Sciences US* **2020,** *117* (2), 1015-1020.
30. Gazit, E., The "Correctly Folded" State of Proteins: Is It a Metastable State? *Angew. Chem. Int. Ed.* **2002,** *41* (2), 257-259.
31. Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-Cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M., Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *Journal of the American Chemical Society* **2011,** *133* (36), 14160-14163.
32. Narimoto, T.; Sakurai, K.; Okamoto, A.; Chatani, E.; Hoshino, M.; Hasegawa, K.; Naiki, H.; Goto, Y., Conformational stability of amyloid fibrils of beta2-microglobulin probed by guanidine-hydrochloride-induced unfolding. *FEBS Lett.* **2004,** *576* (3), 313-9.
33. Morell, M.; Bravo, R.; Espargaró, A.; Sisquella, X.; Avilés, F. X.; Fernàndez-Busquets, X.; Ventura, S., Inclusion bodies: Specificity in their aggregation process and amyloid-like structure. *Biochim. Biophys. Acta* **2008,** *1783* (10), 1815-1825.
34. Carrió, M. M.; Cubarsi, R.; Villaverde, A., Fine architecture of bacterial inclusion bodies. *FEBS Lett.* **2000,** *471* (1), 7-11.
35. Mogk, A.; Bukau, B.; Kampinga, H. H., Cellular Handling of Protein Aggregates by Disaggregation Machines. *Molecular Cell* **2018,** *69* (2), 214-226.
36. Larson, M. E.; Lesné, S. E., Soluble Aβ oligomer production and toxicity. *Journal of Neurochemistry* **2012,** *120* (s1), 125-139.
37. De, S.; Whiten, D. R.; Ruggeri, F. S.; Hughes, C.; Rodrigues, M.; Sideris, D. I.; Taylor, C. G.; Aprile, F. A.; Muyldermans, S.; Knowles, T. P. J.; Vendruscolo, M.; Bryant, C.; Blennow,

K.; Skoog, I.; Kern, S.; Zetterberg, H.; Klenerman, D., Soluble aggregates present in cerebrospinal fluid change in size and mechanism of toxicity during Alzheimer's disease progression. *Acta Neuropathologica Communications* **2019,** *7* (1), 120.
38. Lee, S.; Choi, M. C.; Al Adem, K.; Lukman, S.; Kim, T.-Y., Aggregation and Cellular Toxicity of Pathogenic or Non-pathogenic Proteins. *Sci. Rep.* **2020,** *10* (1), 5120.
39. Mehta, P.; Woo, P.; Venkataraman, K.; Karzai, A. W., Ribosome purification approaches for studying interactions of regulatory proteins and RNAs with the ribosome. *Methods Mol Biol* **2012,** *905*, 273-89.
40. Felske, A.; Engelen, B.; Nübel, U.; Backhaus, H., Direct ribosome isolation from soil to extract bacterial rRNA for community analysis. *Appl Environ Microbiol* **1996,** *62* (11), 4162-7.
41. Zecha, J.; Satpathy, S.; Kanashova, T.; Avanessian, S. C.; Kane, M. H.; Clauser, K. R.; Mertins, P.; Carr, S. A.; Kuster, B., TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Mol Cell Proteomics* **2019,** *18* (7), 1468-1478.
42. Leuenberger, P.; Ganscha, S.; Kahraman, A.; Cappelletti, V.; Boersema, P. J.; von Mering, C.; Claassen, M.; Picotti, P., Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **2017,** *355* (6327), eaai7825.
43. Kalwarczyk, T.; Tabaka, M.; Holyst, R., Biologistics—Diffusion coefficients for complete proteome of Escherichia coli. *Bioinformatics* **2012,** *28* (22), 2971-2978.
44. Dill, K. A.; Ghosh, K.; Schmit, J. D., Physical limits of cells and proteomes. *Proceedings of the National Academy of Sciences* **2011,** *108* (44), 17876-17882.
45. Stark, H.; Mueller, F.; Orlova, E. V.; Schatz, M.; Dube, P.; Erdemir, T.; Zemlin, F.; Brimacombe, R.; Vanheel, M., The 70S Escherichia coli ribosome at 23 Angstrom resolution -fitting the ribosomal RNA. *Structure* **1995,** *3* (8), 815-821.
46. Bakshi, S.; Siryaporn, A.; Goulian, M.; Weisshaar, J. C., Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Molecular Microbiology* **2012,** *85* (1), 21-38.
47. Qin, D. M.; Fredrick, K., Analysis of Polysomes from Bacteria. In *Laboratory Methods in Enzymology: Rna*, Lorsch, J., Ed. 2013; Vol. 530, pp 159-172.
48. Wiśniewski, J. R.; Hein, M. Y.; Cox, J.; Mann, M., A "Proteomic Ruler" for Protein Copy Number and Concentration Estimation without Spike-in Standards*. *Mol. Cell. Proteomics* **2014,** *13* (12), 3497-3506.
49. Scholtz, J. M.; Grimsley, G. R.; Pace, C. N., Chapter 23 Solvent Denaturation of Proteins and Interpretations of the m Value. In *Methods in Enzymology*, Academic Press: 2009; Vol. 466, pp 549-565.
50. Santoro, M. M.; Bolen, D. W., Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl .alpha.-chymotrypsin using different denaturants. *Biochemistry* **1988,** *27* (21), 8063-8068.
51. Oosawa, F.; Kasai, M., A theory of linear and helical aggregations of macromolecules. *J Mol Biol* **1962,** *4*, 10-21.
52. Myers, J. K.; Nick Pace, C.; Martin Scholtz, J., Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **1995,** *4* (10), 2138-2148.
53. Buell, A. K., Stability matters, too - the thermodynamics of amyloid fibril formation. *Chem. Sci.* **2022,** *13* (35), 10177-10192.
54. Vettore, N.; Buell, A. K., Thermodynamics of amyloid fibril formation from chemical depolymerization. *Physical Chemistry Chemical Physics* **2019,** *21* (47), 26184-26194.
55. Shahnawaz, M.; Soto, C., Microcin amyloid fibrils A are reservoir of toxic oligomeric species. *J Biol Chem* **2012,** *287* (15), 11665-76.
56. Nespovitaya, N.; Gath, J.; Barylyuk, K.; Seuring, C.; Meier, B.; Riek, R., Dynamic Assembly and Disassembly of Functional β-Endorphin Amyloid Fibrils. *Journal of the American Chemical Society* **2015,** *138*.

57. O'Nuallain, B.; Shivaprasad, S.; Kheterpal, I.; Wetzel, R., Thermodynamics of Aβ(1−40) Amyloid Fibril Elongation. *Biochemistry* **2005,** *44* (38), 12709-12718.
58. Smith, C. A., Physiology of the Bacterial Cell. A Molecular Approach: by F C Neidhardt, J L Ingraham and M Schaechter. pp 507. Sinauer Associates, Sunderland, MA. 1990. $43.95. *Biochemical Education* **1992,** *20* (2), 124-125.
59. Petrov, A. S.; Bernier, C. R.; Hsiao, C.; Okafor, C. D.; Tannenbaum, E.; Stern, J.; Gaucher, E.; Schneider, D.; Hud, N. V.; Harvey, S. C.; Dean Williams, L., RNA–Magnesium–Protein Interactions in Large Ribosomal Subunit. *The Journal of Physical Chemistry B* **2012,** *116* (28), 8113-8120.
60. Drygin, D.; Zimmermann, R. A., Magnesium ions mediate contacts between phosphoryl oxygens at positions 2122 and 2176 of the 23S rRNA and ribosomal protein L1. *RNA* **2000,** *6* (12), 1714-1726.
61. Cox, R. A.; Pratt, H.; Huvos, P.; Higginson, B.; Hirst, W., A study of the thermal stability of ribosomes and biologically active subribosomal particles. *Biochem J* **1973,** *134* (3), 775-93.
62. Lo Conte, L.; Ailey, B.; Hubbard, T. J.; Brenner, S. E.; Murzin, A. G.; Chothia, C., SCOP: a structural classification of proteins database. *Nucleic Acids Res* **2000,** *28* (1), 257-9.
63. Andreeva, A.; Kulesha, E.; Gough, J.; Murzin, A. G., The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Res* **2020,** *48* (D1), D376-d382.
64. Harris, M. A.; Clark, J.; Ireland, A.; Lomax, J.; Ashburner, M.; Foulger, R.; Eilbeck, K.; Lewis, S.; Marshall, B.; Mungall, C.; Richter, J.; Rubin, G. M.; Blake, J. A.; Bult, C.; Dolan, M.; Drabkin, H.; Eppig, J. T.; Hill, D. P.; Ni, L.; Ringwald, M.; Balakrishnan, R.; Cherry, J. M.; Christie, K. R.; Costanzo, M. C.; Dwight, S. S.; Engel, S.; Fisk, D. G.; Hirschman, J. E.; Hong, E. L.; Nash, R. S.; Sethuraman, A.; Theesfeld, C. L.; Botstein, D.; Dolinski, K.; Feierbach, B.; Berardini, T.; Mundodi, S.; Rhee, S. Y.; Apweiler, R.; Barrell, D.; Camon, E.; Dimmer, E.; Lee, V.; Chisholm, R.; Gaudet, P.; Kibbe, W.; Kishore, R.; Schwarz, E. M.; Sternberg, P.; Gwinn, M.; Hannick, L.; Wortman, J.; Berriman, M.; Wood, V.; de la Cruz, N.; Tonellato, P.; Jaiswal, P.; Seigfried, T.; White, R., The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **2004,** *32* (Database issue), D258-61.
65. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **2021,** *49* (D1), D325-d334.
66. Durica-Mitic, S.; Göpel, Y.; Görke, B., Carbohydrate Utilization in Bacteria: Making the Most Out of Sugars with the Help of Small Regulatory RNAs. *Microbiology Spectrum* **2018,** *6* (2), 10.1128/microbiolspec.rwr-0013-2017.
67. Hollinshead, W. D.; Rodriguez, S.; Martin, H. G.; Wang, G.; Baidoo, E. E. K.; Sale, K. L.; Keasling, J. D.; Mukhopadhyay, A.; Tang, Y. J., Examining Escherichia coli glycolytic pathways, catabolite repression, and metabolite channeling using Δpfk mutants. *Biotechnology for Biofuels* **2016,** *9* (1), 212.
68. Rubio Gomez, M. A.; Ibba, M., Aminoacyl-tRNA synthetases. *RNA* **2020,** *26* (8), 910-936.
69. Wagner, L.; Jules, M.; Borkowski, O., What remains from living cells in bacterial lysate-based cell-free systems. *Computational and Structural Biotechnology Journal* **2023,** *21*, 3173-3182.
70. Richter, K.; Haslbeck, M.; Buchner, J., The heat shock response: life on the verge of death. *Molecular cell* **2010,** *40* (2), 253-266.
71. Hayer-Hartl, M.; Bracher, A.; Hartl, F. U., The GroEL–GroES Chaperonin Machine: A Nano-Cage for Protein Folding. *Trends in Biochemical Sciences* **2016,** *41* (1), 62-76.
72. Grindle, M. P.; Carter, B.; Alao, J. P.; Connors, K.; Tehver, R.; Kravats, A. N. Structural Communication between the E. coli Chaperones DnaK and Hsp90 *Int. J. Mol. Sci.* [Online], 2021.

73. Banecki, B.; Liberek, K.; Wall, D.; Wawrzynów, A.; Georgopoulos, C.; Bertoli, E.; Tanfani, F.; Zylicz, M., Structure-Function Analysis of the Zinc Finger Region of the DnaJ Molecular Chaperone*. *Journal of Biological Chemistry* **1996,** *271* (25), 14840-14848.
74. Perales-Calvo, J.; Lezamiz, A.; Garcia-Manyes, S., The Mechanochemistry of a Structural Zinc Finger. *The Journal of Physical Chemistry Letters* **2015,** *6* (17), 3335-3340.
75. Goemans, C.; Denoncin, K.; Collet, J.-F., Folding mechanisms of periplasmic proteins. *Biochim. Biophys. Acta* **2014,** *1843* (8), 1517-1528.
76. Behrens, S.; Maier, R.; de Cock, H.; Schmid, F. X.; Gross, C. A., The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *The EMBO Journal* **2001,** *20* (1-2), 285-294.
77. Halder, R.; Nissley, D. A.; Sitarik, I.; Jiang, Y.; Rao, Y.; Vu, Q. V.; Li, M. S.; Pritchard, J.; O'Brien, E. P., How soluble misfolded proteins bypass chaperones at the molecular level. *Nat. Comm.* **2023,** *14* (1), 3689.
78. Schramm, F. D.; Schroeder, K.; Jonas, K., Protein aggregation in bacteria. *FEMS Microbiol. Rev.* **2019,** *44* (1), 54-72.
79. Mecha, M. F.; Hutchinson, R. B.; Lee, J. H.; Cavagnero, S., Protein folding in vitro and in the cell: From a solitary journey to a team effort. *Biophys. Chem.* **2022,** *287*, 106821.
80. Hnath, B.; Chen, J.; Reynolds, J.; Choi, E.; Wang, J.; Zhang, D.; Sha, C. M.; Dokholyan, N. V., Big versus small: The impact of aggregate size in disease. *Protein Sci.* **2023,** *32* (7), e4686.

Chapter 2

Appendix: Supplementary Information

Exploring kinetically controlled protein solubility on a proteome-wide scale in *E. coli* under physiologically relevant conditions

2.10. Materials and Methods

2.10.1 Preparation of *E. coli* cytoplasmic S100 protein collection Preparation of *E. coli* S100 protein collection. *E. coli* cells (A19 strain[112]) were inoculated in a 5 mL culture of Luria Broth (LB) medium and grown overnight to saturation at 37 °C, under shaking at 250 rotations per min (rpm). The culture was then centrifuged at 4,000 g for 10 min, resuspended in fresh LB medium, and added to a 1 L flask containing LB medium. Cells were then grown at 37 °C until the start of growth phase of this cell strain at OD_{600} = 0.8 (with 250 rpm shaking), and harvested by centrifugation at 6,690 g for 10 min. The supernatant was discarded, and the combined cell pellet resulting from multiple growths following the above procedure was stored at − 20 °C until further use. Cell lysis was carried out after homogenizing the pellet in lysis buffer (50 mM Tris Base, 2 mM EDTA, 5 mM benzamidine HCl, 10 mM $MgCl_2$, 7 mM β-mercaptoethanol, pH 8.3), employing 1.5 mL of lysis buffer per gram of cells. The homogenized sample was lysed with a French-press cell disrupter (Thermo Scientific, Rockford, IL) at 800 psi. This cell-lysis procedure is essentially equivalent to the one employed to prepare S30 extracts for transcription-translation cell-free protein expression in *E. coli*.[112] The resulting lysate was subject to two consecutive rounds of centrifugation at 30,000 g for 25 min at 4 °C, with transfer to a new 50 mL centrifuge tube after each centrifugation, to remove insoluble materials. In order to deplete ribosomal proteins as intact 70S ribosomes, the clarified lysate was transferred to a fresh 50 mL ultracentrifuge tube and spun at 100,000 g for 2 hrs at 4 °C (Ti-70 rotor, Beckman Ultra floor centrifuge).[20-21, 107, 113] The supernatant was then dialyzed twice in phosphate saline buffer (PBS, 66.8 mM K_2HPO_4, 10 mM NaCl, 0.8 mM $MgCl_2$, 100 nM $CaCl_2$, pH 7.3), each time for 2 hrs at 4 °C (2,000 MWCO 3-12 mL Slide-A-Lyzer Dialysis Cassette, Thermo Scientific, Rockford, IL). A final buffer exchange dialysis step was carried out for an additional 15 hrs at 4 °C. The expected dilution factor was 5 x 10^8 times relative to the initial buffer concentrations. The resulting S100 protein sample was aliquoted, flash frozen, and stored at − 80 °C. All buffers were filtered through a 0.22 µM filter (Milipore-

Sigma), prior to use. The concentration of the S100 protein sample was determined via the bicinchoninic acid assay (BCA assay, See details in next section, n=3).

Overall, 4 L of A19 cell cultures harvested at OD_{600} of 0.8 typically yields 6 mL of soluble S100 sample with of 25 mg/mL total protein concentration. Ten independent preparations of S100 protein samples were carried out, to serve as biological repeats.

2.10.2 BCA assays. The total protein concentration of S100 protein collections was measured with the BCA assay (Abcam, BCA protein assay kit reducing agent compatible - Microplate format, ab207003) according to manufacturer protocols.[114] Absorbance was measured at 562 nm on a microplate reader (Multimode Plate Reader EnVision 2105, Perkin Elmer). A standard curve was constructed from a 2 mg/mL stock Bovine Serum Albumin (BSA, Sigma Aldrich) stock dissolved in the solutions listed below. BSA stock concentration was determined by electronic absorption at 280 nm (8452A Diode Array Spectrophotometer, Hewlett Packard). Triplicate measurements were carried out for each standard curve (standard errors were ± 1%). Three independent BSA standard curves were generated in a) PBS buffer, b) PBS buffer containing 1 mM DTT, and c) 1% (v/v) sodium dodecyl sulfate (SDS) in water.

2.10.3 Anaerobic heating-cooling. The S100 protein-collection stock solutions were thawed, centrifuged at 15,490 g for 5 min, and transferred to fresh 1.5 mL Eppendorf tubes. Samples were diluted with PBS in 1 mM dithiothreitol (DTT) to a final total-protein concentration of 20 mg/mL, measured via the BCA assay. This was the starting concentration for each experiment. Solutions were transferred to the experimental set-up for anaerobic experiments, encompassing a thermocycler housed in an Aldrich-Atmos bag (ThermoFisher) equipped with an oxygen-probe (FireSting) to monitor oxygen levels. N_2 gas was flushed into the Atmos bag until oxygen level dropped from 20 to 0.02%, expressed as percent by volume of dry air.

Under the above anaerobic conditions, 7 additional samples at lower total protein concentration (see Fig. 1) were prepared via dilution of the 20 mg/mL S100 stock solution and distributed with 200 µL per PCR tube. Samples were subjected to heating at 70 °C for 1 hr followed by slow cooling to 22 °C for 15 min, using a T100™ Thermocyler (BioRad). The resulting 2-phase (insoluble and soluble) samples were resuspended and combined into fresh 1.5 mL Eppendorf tubes, for each total-protein concentration in Figure 1. Insoluble protein was separated from the soluble fraction by centrifugation at 15,490 g for 20 min at 4 °C.

The supernatant was carefully removed. For lower initial total protein concentrations of 0.1 to 1 mg/mL, the soluble fractions were concentrated to 100-200 µL with 3,500 MWCO Amicon centrifugation units (Milipore Sigma), according to manufacturer's protocols, to reach detectable concentrations for downstream analysis. The volume of each soluble fraction was estimated and recorded. Insoluble fractions were washed by pipetting 80 µL of fresh PBS buffer on top without resuspension followed by brief centrifugation at 15,490 g for 5 min at 4°C, to remove any residual soluble fraction. The above procedure was carried out for data in Figure 1, Table 1 and 4, and Supplementary Figures S4-S5 (n=5 independent biological repeats).

2.10.4 Anaerobic incubation. Similarly to the procedures followed for anaerobic heating-cooling, stock S100 protein samples were diluted to a unified concentration of 20 mg/mL, measured with BCA assay. Samples were further diluted in the anaerobic set up to generate 7 additional samples at lower total protein concentrations as shown in Figure 4. and distributed with 200 uL per PCR tube. The samples were subjected to heating at 37 °C (incubation) for 20 hrs followed by slow cooling to 22 °C for 15 min using the Thermocycler. The resulting 2-phase samples were prepared for subsequent analysis with identical procedures as in the anaerobic heating-cooling experiments. These soluble and insoluble fractions, at different total protein concentrations for this section, were denoted as the anaerobic incubated sample under

physiologically relevant conditions for downstream analysis (See Figure 4, Table 2 and 5 and SI Figures 12). 5 independent biological repeats were done.

2.10.5 Solubility assays and densitometry. Soluble and insoluble fractions from either anaerobic heating-cooling or anaerobic incubation experiments were prepared for SDS-PAGE analysis using the following procedures. Each insoluble fraction was resuspended in a gel loading buffer (0.14 M SDS, 2.7 M glycerol, 0.72 M BME, and 0.1 M Tris base at pH 6.8) and diluted with a factor 2 to 6 of the original volume of the supernatant. Higher dilution with gel loading buffer was needed to fully dissolve the insoluble aggregates at higher initial total protein concentrations and to minimize insoluble fraction volume contribution. The volume of the soluble fraction was diluted appropriately to match the dilution of the insoluble fraction gel sample for each protein concentration. In other words, the soluble and insoluble fraction gel samples were tailored to match their relative concentrations before phase separation.

The samples were heated at 90-95 °C for 2 min to fully unfold proteins and centrifuged at 15,490 g for 2 min to pellet any insoluble materials. The heated samples were then analyzed with 4-20 % Mini-PROTEAN® TGX™ Precast Protein SDS-PAGE gradient gels (10 or 15 wells, 30 µL format, Bio-Rad, California) using Tris/Glycine electrophoresis buffer (25 mM Tris, and 192 mM glycine at pH 8.3). The molecular weight ladder used in all gel analysis was Precision Plus Protein All Blue Standards (Bio-Rad, California).

SDS-PAGE gels were analyzed with densitometry analysis using the ImageJ Software version 1.53k (NIH, USA). For a given concentration, we reported a fraction of insoluble or soluble proteins, $F_{insoluble}$ or $F_{soluble}$, from intensity measurements with densitometry analysis via relation

$$F_{insoluble} = \left(\frac{I_{insoluble}}{I_{soluble} + I_{insoluble}}\right) \; or \; F_{soluble} = \left(\frac{I_{soluble}}{I_{soluble} + I_{insoluble}}\right) \quad (1)$$

where $I_{insoluble}$ is the total lane intensity from the soluble fraction subjected to a 1:x soluble phase dilution, and $I_{soluble}$ the total lane intensity of the insoluble fraction. Figure 1 and 2 showed the change in $F_{insoluble}$ or $F_{soluble}$ proteins upon anaerobic heating-cooling and incubation

respectively at increasing total protein concentrations. Standard errors and student-t test were determined across 5 independent biological repeats.

2.10.6 Data fitting according to equilibrium thermodynamics (ET) model. To assess the mode of aggregation during the heating-cooling experiments, we fit the solubility data at different protein concentrations to multiple models of aggregation (See Figure 1, panel C). The simplest model, adopted from previous works,[115] would be that at high temperature (70 °C) there is an equilibrium reached between soluble and aggregated states. Then, during cooling to ambient temperature (22 °C, there is a sudden reintroduction of kinetic barriers, stopping any flux between the two states. Under this model the measured solubilities are representative of the equilibrium values reached at high temperature for subsequent model derivations. Thermodynamics predicts that at equilibrium the chemical potential of the soluble state and the aggregated state will be equal[116].

$$\mu_M = \mu_A \, , \tag{2}$$

where μ_M is the per mole chemical potential of the soluble state and μ_A is the per mole chemical potential of the aggregated state. The chemical potentials can be rewritten in terms of activities[116].

$$\mu_M^0 + RTln(a_M) = \mu_A^0 + RTln(a_A) \, , \tag{3}$$

where μ_M^0 and μ_A^0 are the reference chemical potentials under standard state conditions, T is the temperature of the solution, R is the gas constant, and a_M and a_A are the activities of proteins in the soluble and aggregated phases respectively. The activity of the aggregated phase is constant, while the activity of the soluble phase can be expressed in terms of the concentration of soluble proteins.

$$\mu_M^0 + RTln\left(\gamma_M \frac{c_M}{c_{tot}}\right) = \mu_A^0 + RTln(a_A) \, , \tag{4}$$

where γ_M is the activity coefficient of the soluble proteins, c_M is the concentration of soluble proteins, and c_{tot} is the total concentration of the soluble solution which is approximately equal to

the concentration of water. At constant temperature and pressure, the concentration of soluble proteins, referred to as the critical concentration ($[M]_{crit}$), will be constant.

$$C_M = \frac{c_{tot}}{\gamma_M} \exp\left(\frac{\mu_A^0 + RTln(a_A) \ - \mu_M^0}{RT}\right) = const. = [M]_{crit} \ . \quad (5)$$

Any soluble proteins in excess of this critical concentration will precipitate into the aggregate state.[115] This same theory can be represented via reaction equilibrium. The overall conversion from the soluble phase to the aggregate phase can be modeled as a reaction of n soluble species to form an aggregate,

$$nM \rightleftharpoons A_n \ , \quad (6)$$

or identically:

$$M \rightleftharpoons \frac{1}{n} A_n \ . \quad (7)$$

The concentration of these two species can be related to an equilibrium constant,

$$K = \frac{[A_n]^{1/n}}{[M]} \ . \quad (8)$$

When the aggregate state is large $n \to \infty$ and the numerator approaches 1,

$$K = \frac{1}{[M]} \ . \quad (9)$$

From this it is again clear that the concentration of the soluble state is constant (see SI Figure 3, panel A),

$$[M] = \frac{1}{K} = const. = [M]_{crit} \ . \quad (10)$$

Using this the percent of insoluble proteins can be expressed as (see SI Figure 3, panel B),

$$F_{insol} = 100 * \frac{[P]_{tot} - [M]_{crit}}{[P]_{tot}} \ , \quad (11)$$

where F_{insol} is the percent of proteins in the aggregate state, $[P]_{tot}$ is the total amount of protein per volume of solution (including any in the aggregate state), and $[M]_{crit}$ is the critical concentration. Equation (10) was fitted to experimental measurements of percent insoluble as a

function of total protein concentration, where $[M]_{crit}$ is varied as a parameter. The fit was done using the fitnlm function in MATLAB R2022a. The initial guess was 0.1 mg/mL, and standard optimization parameters were used. The fit value of $[M]_{crit}$ is 0.0332 mg/mL.

The equilibrium model does a poor job of describing the observed solubility behavior (See Figure 1, panel C). While this ET model predicts that the fraction of insoluble proteins will quickly approach 100%, the experimental solubility data plateaus at 95%.

2.10.7 Pre-equilibrium thermodynamics (Pre-ET) model. The second simplest model is one in which we do not assume equilibrium is reached at high temperatures. Instead, there is flux between the soluble and aggregate states, but not enough to reach equilibrium within the 1 hr. This can be described using a pre-ET model of aggregation. It is assumed that the rate-determining step for aggregation would be the nucleation of two monomers,

$$M \rightarrow \frac{1}{2} M_2 \ , \tag{12}$$

where M is the soluble state and M_2 is a dimer that will lead to the aggregate state. Assuming all subsequent steps are fast enough to be ignored, the rate of the forward reaction can be defined as that of the elementary step,

$$r_1 = k_1 * [M]^2 \ , \tag{13}$$

where r_1 is the rate of the forward reaction, k_1 is the corresponding rate constant, and [M] is the moles of soluble protein per volume of solution. We can assume disaggregation will occur on the surface of protein aggregates. Thus, this rate can be represented in terms of the surface area of aggregated pellets[117],

$$r_{-1} = k_{-1} * N * SA \ , \tag{14}$$

where r_{-1} is the rate of disaggregation, k_{-1} is the corresponding rate constant, N is the number of aggregate particles, and SA is the surface area of each aggregate particle. If we model the particles as spherical then their surface area is,

$$SA = 4\pi R^2 \ , \tag{15}$$

where R is the particle radius. Likewise, we can relate the volume of a spherical particle to the total number of proteins in the aggregate state using,

$$\frac{V}{\rho_A} * \frac{[A]}{N} = \frac{4}{3}\pi R^3 \ , \tag{16}$$

where V is the volume of the solution, ρ_A is the molar density of the aggregate, and [A] is the moles protein in the aggregate state per volume of solution. Combining this with equation (14) we get,

$$SA = (4\pi)^{\frac{1}{3}}(\frac{3V * [A]}{\rho_A * N})^{\frac{2}{3}} \ . \tag{17}$$

This gives us the equation for the rate of disaggregation,

$$r_{-1} = N^{\frac{1}{3}} * k_{-1} * (4\pi)^{\frac{1}{3}}(\frac{3V}{\rho_A})^{\frac{2}{3}} * [A]^{\frac{2}{3}} \ . \tag{18}$$

Most of these constants can be combined into an apparent rate constant,

$$k_{-1}^{app} = k_{-1} * (4\pi)^{\frac{1}{3}}(\frac{3V}{\rho_A})^{\frac{2}{3}} \ . \tag{19}$$

This gives a simplified form of the disaggregation reaction,

$$r_{-1} = N^{\frac{1}{3}} * k_{-1}^{app} * [A]^{\frac{2}{3}} \ . \tag{20}$$

Here we must decide how the number of aggregate particles varies with the total aggregate concentration. There are two limiting cases of this. First, the number of particles does not change with total aggregate concentration. The $N^{1/3}$ would then be constant and could be absorbed into the k^{app}, yielding at ⅔ order reaction,

$$r_{-1} = k_{-1}^{app} * [A]^{\frac{2}{3}} \ . \tag{21}$$

Second, we could have the opposite limiting case; the number of particles could be a linear function of the total aggregate concentration,

$$r_{-1} = (c * [A])^{\frac{1}{3}} * k_{-1}^{app} * [A]^{\frac{2}{3}} \ , \tag{22}$$

where c is a constant of proportionality that relates the total aggregate concentration to the number of aggregate particles. In this limiting case the linear constant can be absorbed into the apparent rate constant and the [A]s can be combined giving a 1st order reaction,

$$r_{-1} = k_{-1}^{app} * [A] \ . \tag{23}$$

The real-world behavior of the system will lie somewhere between these two extreme cases. However, we are trying to model a phenomenon where there is more flux to the soluble state than we would expect if the system were at equilibrium. With this goal in mind, the first limiting case is the more restrictive of the two. In other words, if a ⅔ order model can describe our observations, then a 1st order model will also be able to. Thus, the more limiting ⅔ order model was used to describe the rate of disaggregation, giving the overall reaction rate law,

$$r = k_1 * [M]^2 - k_{-1}^{app} * [A]^{\frac{2}{3}} \ , \tag{24}$$

where r is the overall rate of reaction in terms of moles protein per time. This derivation was done for spherical aggregate particles, but the same exponents come about for particles of arbitrary shape. The total reaction law can be converted to the differential equations,

$$\frac{d[A]}{dt} = k_1 * [M]^2 - k_{-1}^{app} * [A]^{\frac{2}{3}} \ , \tag{25}$$

$$\frac{d[M]}{dt} = k_{-1}^{app} * [A]^{\frac{2}{3}} - k_1 * [M]^2 \ . \tag{26}$$

The kinetic model was composed of these two differential equations. At each protein concentration the differential equations were numerically solved using the initial condition that all proteins start in the soluble state,

$$[M]_i = [P]_{tot} \ , \tag{27}$$

where $[M]_i$ is the initial concentration in the soluble state. This model was fitted to our measurements of fraction of soluble proteins at multiple starting concentrations. The differential equations were solved using a numerical ODE solver with k_1 and k_{-1} as fitting parameters. The fitting was done using the fitnlm function in MATLAB R2022a, using 1 or 0.3 for all initial guesses

and standard optimization parameters. This fitting assumes that k_1 and k_{-1} are constant with time. In reality these will vary with temperature, however the sample spends the vast majority of its time as 70^{O} C, so the assumption of constant temperature and rate constants is likely valid. This model yielded two possible results (as shown in Figure 1) When starting guesses for k1 and k-1 are below 0.322 the solver converges on the first kinetic model. This first model describes a system that does not reach equilibrium within 1 hr. The rate constants associated with it are 0.608 $\frac{mL}{mg*hr}$ and 0.00103 $\frac{(\frac{mg}{mL})^{1/3}}{hr}$ for k1 and k-1 respectively. When the starting guesses for k1 and k-1 are above 0.322 the model converges on an equilibrium model, in which the rate constants are large enough to reach equilibrium within the 1 hr. This fits the data as poorly as the previous equilibrium model (See Figure 1, panel C). The rate constants associated with this are 17945 $\frac{mL}{mg*hr}$ and 117.89 $\frac{(\frac{mg}{mL})^{1/3}}{hr}$ for k1 and k-1 respectively.

2.10.8 Quasi-equilibrium thermodynamics (Quasi-ET) model. Previous models describing the behavior of the sample at only one temperature fail to fit the observed data and a more nuanced model was derived. For this Quasi-ET model, it is reasonable to assume that the sample reaches equilibrium at high temperatures, as above 60°C virtually all globular proteins are unfolded[118]. However, during the cooling process there may not be immediate kinetic trapping. Instead, there will be a range of temperatures in which the flux between the soluble state and aggregates is present, yet there is not enough to reach equilibrium.

According to this model, proteins reach equilibrium at high temperatures (70 °C), but kinetic deviations that occur during cooling to 22 °C make the measured solubilities different from their high temperature equilibrium values. The equilibrium at high temperature was modeled with equation (10), and the kinetic deviations were modeled with equations (23) and (24). We can assume that the region where kinetic deviations occur has a small temperature range, allowing

the use of unvarying rate constants. The result is a model that works similarly to the kinetic model, but uses the initial condition,

$$[M]_i = [M]_{crit} \ . \tag{28}$$

This model was fit by varying k_1, k_{-1}, and $[M]_{crit}$ as parameters. The fitting was done using the fitnlm function in MATLAB R2022a, using 1 for all initial guesses and standard optimization parameters. This model does a good job of describing the data across all measured protein concentrations (See Figure 1, panel C). The fit parameters are 0.1932 $\frac{mL}{mg*time}$, 0.13794 $\frac{(\frac{mg}{mL})^{1/3}}{time}$, and 0.0003 $\frac{mg}{mL}$ for k_1, k_{-1} and $[M]_{crit}$ respectively.

2.10.9 Dynamic light scattering (DLS). We will use to estimate the general particle size distributions within different soluble proteomic samples, containing >2,000 proteins. Though DLS has traditionally been used to analyze simpler protein or other macromolecule samples, at least one study has reported its analysis to complex lysates for *E. coli* cell-free systems. [119] In general, protein translational diffusion gives rise to fluctuations in scattered-light intensity. The fluctuation frequency is a function of molecular size. Scattered light was collected as an autocorrelation function (ACF), which is a measurement of the fluctuating intensity of the scattered light due to the continuously changing distance between particles in a solution. ACFs were then analyzed according to the cumulant and regularization method. Specifically, the regularization method fits the experimental autocorrelation function to a sum of exponentials according to the DYNALS algorithm and resolves the potential presence of multiple species of different hydrodynamic radii in the solution, assuming all particle motions are exclusively due to translational diffusion.[120] Translational diffusion coefficient D_t for the different exponentially decaying components of the ACF were obtained from the DYNALS algorithm and converted to apparent hydrodynamic radii (R_h) via the Stokes-Einstein equation:

$$D_t = \frac{k_B T}{6\pi\eta R_h} \quad (29)$$

where k_B is the Boltzmann constant (in kg m^2 s^{-2} K^{-1}), T is the temperature (in K), η is the measured viscosity of the solution in the absence of protein (in kg s^{-1} m^{-1}), and R_h is the hydrodynamic radius (in m). R_h represents the average radius of a hypothetical hard sphere with a translational diffusion coefficient identical to the measured value of a protein. Approximate estimates of the percent mass corresponding to each peak were generated assuming spherical shape for each of the detected particles.

Data were collected at 163.5° (angle between incident light and QELS detector) at an inner cell temperature of 25 °C on a Mobius instrument (Wyatt Technology – Waters Inc.) equipped with an avalanche photodiode detector and a GaAs laser operating at 532 nm. Data for each sample were signal averaged (20 acquisitions, 5 s each) and analyzed with the DYNAMICS 7.4 software (Wyatt Technology). Prior to the measurements, samples were centrifuged at 15,490 g for 10 min at 4 °C to remove dust or other large particles. 20 µL of the top portion of the supernatant was used to wash a quartz 45 µL microcuvette (HE0171, Wyatt Technology, 1 mm, size range resolution of 0.2-200 nm). Then, 20 µL of the sample was used for the measurements. For each sample, we allowed for 10 min before any data collection to allow for both the sample and cuvette cell to reach a stable 25 °C. Three consecutive measurements were done 3 min apart for each sample to ensure no laser power or temperature fluctuation has altered across DLS measurements. Each sample was also diluted and re-measured to prevent multiple scattering from skewing the R_h analysis without compromising the quality and sensitivity of the ACFs. Refractive index and viscosity values of all solutions without proteins were measured with an Abbe refractometer (Thermo Spectronic, Fisher Scientific) and microviscometer (RheoSense, Inc.) respectively for subsequent calculations of hydrodynamic radii. Both cumulant (SI Tables 5-6) and regularization (SI Tables 1-4) analyses were reported for each sample.

2.10.10 Steady-state fluorescence emission measurements. All steady-state fluorescent emission measurements in this study, including tryptophan fluorescent emission and thioflavin T assay (see SI Figure S9, panel B) were done on a PC1 photo-counting steady-state fluorimeter (ISS Inc.). Excitation and emission monochromators had slit widths of 4 nm. No filters were used. 15 µL of each sample was used to wash an ultra-micro-Quartz cuvette with pathlength of 10 mm (105.210-QS, Hellma Analytics) and discarded. Then, 15 µL of the same sample was used to perform steady-state fluorescent emission measurements at 22.5 °C. To eliminate contributions of background fluorescent emission and Rayleigh/Raman scatterings from buffers/solutions, the emission spectrum of each buffer/solution were independently collected and subtracted from the associated sample emission spectrum.

2.10.11 Thermodynamic stability of soluble proteins before anaerobic heating-cooling. To measure the apparent thermodynamic stability of the soluble protein (relative to unfolded states), we unfolded the soluble S100 samples in guanidine hydrochloride (GdnHCl). We chose GdnHCl because a 9.5 M urea solution was unable to completely unfold our proteomic samples. Soluble S100 samples were thawed and prepared as 5 mg/mL aliquots, determined by BCA assay for subsequent denaturation experiments. A stock of GdnHCl was prepared by dissolving solid GdnHCl (Sigma-Aldrich) in the previously described PBS buffer (SI Methods 1.1.) and the concentration assessed by refractometry as previously described.[82] The stock GdnHCl was subsequently passed through a sterile syringe 0.22 micron filter. Then, 45 µL aliquots of S100 samples were incubated in LoBind® Eppendorf tubes in the presence of 0-6 M GdnHCl for 24 hrs at 22.5 °C with a constant solution volume of 200 µL. The tryptophan fluorescence emission of each denatured sample (from 295 to 500 nm, see SI Figure S9, panel B) was monitored at 22.5 °C with steady-state fluorescent emission with excitation at 285 nm. Excitation and emission monochromators had slit widths of 4 nm. No filters were used. The emission spectrum at each denaturant concentration was blanked with an independently collected spectrum of the associated

denaturant-PBS solution. After baseline correction, a spectral center of mass, λ_{SCOM}, of each emission spectrum was assessed via the relation[67, 121]:

$$Spectral\ center\ of\ mass\ (nm) = \lambda_{SCOM} = \frac{\sum \lambda_i \times I_i}{\sum I_i} \quad (30)$$

where I_i stands for the fluorescence intensity emitted at wavelength λ_i and the summation is performed across the wavelength range of 300-400 nm of each emission spectrum, excluding lower wavelengths due to Rayleigh scattering at the 285 nm excitation. The apparent fraction of unfolded proteins at each GdnHCl concentration (F_{unfold}) was determined according to the following relation[67, 121]:

$$F_{unfold} = \left[1 + Q \times \left(\frac{\lambda_{SCOM} - \lambda_{6M}}{\lambda_{0M} - \lambda_{SCOM}}\right)\right]^{-1} \quad (31)$$

where Q denotes the ratio between the average quantum yields of the unfolded and native states (i.e., average highest intensities at 0-1 M and 4-6 M respectively), λ_{SCOM} is the spectral center of mass calculated from Equation 28 at a given GdnHCl concentration, and λ_{0M} and λ_{6M} denotes the spectral centers of mass of the native and unfolded states respectively. The resulting unfolding curve was fitted with a two-state model, using 6 variables fitting to account for non-flat baselines of folding and unfolding, according to Santoro and Bolen,[83, 122] with Kaleidagraph (version 4.5.2, Synergy Software):

$$F_{unfold} = \frac{\{y_f + m_f[GdnHCl]\} + \{y_u + m_u[GdnHCl]\} \times \{e^{-\left(\frac{\Delta G^\circ_{H2O,unfold} - m_{unfold}}{RT}\right)}\}}{1 + e^{-\left(\frac{\Delta G^\circ_{H2O,unfold} - m_{unfold}}{RT}\right)}} \quad (32)$$

where y_f and m_f, and y_u and m_u are intercept and slope of the pre- and post- transition baselines respectively, $\Delta G^\circ_{H2O,unfold}$ in J·mol^{-1} is the apparent standard state Gibbs free energy change of unfolding, m_{unfold} in J·mol^{-1}·M^{-1} is the cooperativity constant in respect to denaturant concentration, R is the Boltzmann constant in J·mol^{-1}·K^{-1}, and T is temperature in K. Unfolding experiments were performed on 3 biological repeats of the S100 soluble samples to account for potential variations

in protein compositions (see SI Figure S6). Apparent $\Delta G_{H2O,unfold}$ and m_{unfold} were deduced from the fits and reported as an average ± SE for $n_{biological} = 3$ (see Figure 2, panel B and SI Figure S9, panel A). It is important to note that the deduced apparent thermodynamic stability values are represented of a complex proteome sample, containing over 2,000 *E. coli* proteins. Further investigation is needed to probe the individual contribution of the unfolding of each individual protein. However, it has been demonstrated that a protein's thermodynamic stability does not increase significantly when the denaturing experiments were done in the presence of concentrated complex lysate (c.a., < 100 mg/mL).[123-124]

2.10.12 Thermodynamic stability assessment of insoluble aggregates after anaerobic heating-cooling. Insoluble aggregates were generated from S100 aliquots (5 mg/mL, determined by BCA assay) upon anaerobic heating-cooling treatment. The resulting 2-phase aliquots were first combined to normalize out potential variations in aggregation from different aliquots and redistributed as 45 µL aliquots to new LoBind® Eppendorf tubes. Samples are then centrifuged at 15,490 g for 10 min and the soluble phase removed. The insoluble aggregates are incubated in the presence of 0-6 M GdnHCl, as previously prepared in SI Methods S1.11, for 24 hrs at 22 °C with a constant solution volume of 200 µL (i.e., total protein concentration upon unfolding: 1.1 mg/mL). Post-incubation samples were centrifuged at 15,490 g for 20 min and the soluble GdnHCl phase removed. The insoluble aggregate sample for each concentration of GdnHCl was washed with 90 µL of PBS buffer to remove residual soluble proteins and GdnHCl, centrifuged at 15,490 g for 5 min, and the wash buffer removed. The washed aggregates were analyzed with SDS-PAGE solubility assay as described in SI Method 1.5 (see SI Figure S6). For a given unfolding condition, we calculated a fraction of unfolded protein, F_{unfold}, from intensity measurement with densitometry analysis via the relation:

$$F_{unfold} = 1 - \left(\frac{I_{xM}}{I_{0M}}\right) \quad (33)$$

where I_{xM} is the total lane intensity from insoluble aggregates incubated in a specific concentration of GdnHCl, and I_{0M} the total lane intensity of the insoluble aggregates incubated at 0 M GdnHCl (i.e., no unfolding of aggregates). The resulting curve was fitted with the isodesmic form of the linear polymerization model.[58, 76, 87] In this model, the overall aggregation is described by a series of reversible addition of a monomer, M, to an aggregate, A_i, of size i, under a single association equilibrium constant K_a:

$$[A_{i-1}] + [M] \rightleftharpoons [A_i] \; with \; K_a = \frac{[A_i]}{[M][A_{i-1}]} \quad (34)$$

This isodesmic model assumes a single association equilibrium constant for all addition of a monomer to an existing aggregate, or to another monomer. Then, the total concentration of monomer, $[M]_0$, is defined as:

$$[M]_0 = \sum_{i=1}^{\infty} i[A_i] = \sum_{i=1}^{\infty} iK_a^{i-1}[M]^i = \frac{[M]}{(1 - K_a[M])^2} \quad (35)$$

In the presence of chemical denaturants, such as GdnHCl, the aggregates are destabilized and unfold, like the state transition in the two-state model previously described in SI Methods 1.10. Therefore, by solving Equation 30. with the quadratic formula, the concentration of monomers in solution, $[M]$, is:

$$[M] = \frac{1 + 2[M]_0K_a - \sqrt{(-1 - 2[M]_0K_a)^2 + 4[M]_0^2K_a^2}}{2[M]_0K_a^2} \text{ with } K_a = e^{\frac{-\Delta G^0_{water,unfold} - m[GdnHCl]}{RT}} \quad (36)$$

Therefore, the fraction of unfolding under this linear polymerization model can be defined as the following relation and used to fit the unfolding profile of aggregates, with Kaleidagraph (Figure 2, panel A):

$$F_{unfold} = \frac{[M]}{[M]_0} = \frac{\frac{1}{2} + [M]_0 e^{\frac{-(\Delta G^0_{water,unfold} + m[GdnHCl])}{RT}} - \sqrt{[M]_0 e^{\frac{-(\Delta G^0_{water,unfold} + m[GdnHCl])}{RT}} + \frac{1}{4}}}{[M]_0^2 e^{\left(\frac{-(\Delta G^0_{water} + m[denaturant])}{RT}\right)^2}} \quad (37)$$

where $\Delta G°_{H2O,unfold}$ in J·mol^{-1} is the apparent standard state Gibbs free energy change of unfolding, m_{unfold} in J·mol^{-1}·M^{-1} is the cooperativity constant in respect to denaturant concentration, R is the Boltzmann constant in J·mol^{-1}·K^{-1}, and T is temperature in K. Unfolding experiments were performed on the insoluble aggregates generated from 3 biological S100 sample replicates to account for potential variations in protein compositions. Apparent $\Delta G_{H2O,unfold}$ and m_{unfold} were deduced from the fits and reported as an average ± SE for $n_{biological} = 3$ (See Figure 2 and SI Figure S9). It is important to note that the deduced apparent thermodynamic stability values are represented of a complex proteome sample, containing over 2,000 *E. coli* proteins. Similar to SI Methods 1.10., further investigation is needed to probe the individual contribution of the unfolding of each individual protein.

2.10.13 Thioflavin T assay to assess proteome unfolding in the absence and presence of 6 M GdnHCl. ThT solid (Sigma-Aldrich) was dissolved in PBS buffer and the concentration of the stock 40 µM ThT was determined on an HP 8452A Diode Array UV-vis spectrophotometer (Hewlett-Packard) at 412 nm using the extinction coefficient of 36 mM^{-1} cm^{-1}.[125] For measurement of ThT fluorescence, 40 µL of each sample was mixed with 40 µL of the stock ThT solution to a working final ThT concentration of 20 µM.[126] Fluorescence emission spectra was measured at 22.5 °C with a steady-state fluorimeter as described in SI Methods 1.10. upon excitation at 430 nm, to minimize the contribution of Rayleigh scattering at the excitation wavelength to the emission profile at 485 nm. Excitation and emission monochromators had slit widths of 4 nm. No filters were used. To eliminate the potential self-fluorescent property of ThT above 20 µM and background scattering from the sample buffer,[126] ThT emission spectra in PBS

buffer and GdnHCl solutions were collected and subtracted to the associated samples for subsequent analysis. Samples from the unfolding experiments (i.e., the soluble S100 sample and insoluble aggregates before and after anaerobic heating-cooling treatment respectively) were analyzed with ThT assay in the presence and absence of 6 M GdnHCl (see SI Figure S9, panel B). Total protein concentrations in the working ThT assay were estimated to be 2.5 mg/mL.

2.10.14 Circular dichroism (CD) to assess proteome unfolding in the absence and presence of 6 M GdnHCl. All CD measurements in this study were done on a MOS-450 spectrophotometer (Bio-Logic Science Instruments) in CD mode over 200-300 nm range with 0.5 nm steps and a constant data collection time of 10s per step, at a temperature of 22.5 °C. 100 µL of each sample was used to wash a micro-absorption Quartz glass cuvette with pathlength of 1 mm (Hellma Analytics) and discarded. Then, 150 µL of the sample was used to perform CD absorption measurement. To eliminate contributions of background absorption and Rayleigh/Raman scatterings from buffers/solutions, the CD spectrum of each buffer/solution were independently collected and subtracted from the associated sample CD spectrum. The resulting sample spectra were reported as θ, or experimentally measured ellipticity in milidegrees, because it is not possible calculate molar ellipticity residues for a complex proteome S100 samples (i.e., containing over 2,000 *E coli* proteins). Samples from the unfolding experiments (i.e., the soluble S100 sample and insoluble aggregates before and after anaerobic heating-cooling treatment respectively) were analyzed with CD absorption in the presence and absence of 6 M GdnHCl (see SI Figure 10). The estimated total protein concentration in the soluble samples is 1.1 mg/mL, as in SI Method 1.12.

2.10.15 Progressive dilution experiments to probe the proteome-wide reversibility of aggregation. To probe reversibility of aggregation as insoluble aggregates, we subjected the aggregates to progressive dilution experiments. Theoretically, dilution of the soluble phase brings the system away from the initial equilibrium and result in the change in the chemical potentials of the two phases, as described in a modified version of Equation 2 under dilution:

$$\mu_{soluble} < \mu_{insoluble} \quad (38)$$

where $\mu_{soluble}$ is the per mole chemical potential of the soluble state and $\mu_{insoluble}$ is the per mole chemical potential of the insoluble aggregated state. If the two phases are under exchangeable equilibrium, proteins in insoluble phase can exchange with the soluble states to establish a new equilibrium on a biologically relevant timescale (e.g., 2 or 20 hrs). The protein concentration in the soluble phase under this equilibrium exchange tends to a constant critical aggregated concentration as described in Equation.[87, 90, 127] This type of dilution experiment has been routinely used to probe equilibrium reversibility of fibril formations.[87-88, 90, 128]

Insoluble aggregates were generated from S100 aliquots (5 mg/mL, determined by BCA assay) upon anaerobic heating-cooling treatment. The resulting 2-phase aliquots were first combined to normalize out potential variations in aggregation from different aliquots and redistributed as 90 µL aliquots to new Eppendorf tubes. Samples are then centrifuged at 15,490 g for 10 min and the 90 µL soluble phase removed without disturbing the insoluble aggregates. The insoluble aggregate samples were diluted with increasing volume PBS buffer with dilution factor from 1:1 (i.e., no dilution) to 1:100 dilution relative to the initial 90 µL soluble phase volume removed (Figure 3, panel B and C). Diluted samples were incubated at 22.5 °C for 2 hrs and centrifuged at 15,490 g for 20 min to collect the insoluble aggregates. The aggregates were analyzed with SDS-PAGE solubility assay as described in SI Method 1.5 (see SI Figure S11). For a given dilution condition, we calculated a fraction of insoluble protein, $F_{insoluble}$, from intensity measurement with densitometry analysis via the relation:

$$F_{insoluble} = \left(\frac{I_{1:x\ diution}}{I_{1:1\ dilution}}\right) \quad (39)$$

where $I_{1:x}$ is the total lane intensity from insoluble aggregates subjected to a 1:x soluble phase dilution, and $I_{1:1}$ the total lane intensity of the control insoluble aggregate sample under no dilution (i.e., 1:1). Figure 2 showed the change in fraction of insoluble proteins upon progressive dilution. The soluble phases generated from the dilution of insoluble aggregates were probed with DLS as

described in SI Methods 1.9 (Figure 2, panel D and SI Table S1). Previous experiment on the dilution of fibrils suggest that the dissociated proteins consists of small, non-aggregated states upon longer incubation following dilution.[89] Therefore, we proceeded to carry out the progressive dilution experiment followed by a longer period of incubation of 20 hrs at 22.5 °C and the resulting soluble phase probed with DLS (i.e., soluble phase from 1:20 dilution of insoluble aggregates, see SI Figure S16). Lastly, due to the low sensitivity of available concentration measurement methods for the soluble phase upon dilution of insoluble aggregates from over 2,00 proteins, we resulted to estimating the apparent critical aggregate concentration of the soluble phase, $app.[CAC]$, with the initial insoluble aggregate concentration, $[A_{insoluble}]$ (c.a., 4.5 mg/mL as determined by BCA assay), and the fraction of insoluble protein, $F_{insoluble}$, at each dilution condition:

$$app.[CAC] = [A_{insoluble}] \times d_{1:x}(1 - F_{insoluble}) \quad (40)$$

where $d_{1:x}$ is the dilution factor of the soluble phase. The estimated apparent critical concentration of the soluble phase, for the complex proteome sample containing over 2,000 proteins, is reported in SI Table S8.

We carried out identical experiments and characterizations on the small fraction of insoluble aggregates generated from anaerobic incubation (See Figure 4, panel B and SI Table S2).

2.10.16 Label-free Enzymatic "In Liquid" Digestion. 3 biological repeats were subjected to liquid chromatography - mass spectrometry analysis to assess oxygen-related modifications from anaerobic heating-cooling treatment (See SI Figure 1). Each biological repeat consists of 2 samples: 1) control untreated, and 2) supernatant-pellet samples before and after heating-cooling treatment. A total of 6 samples from 3 repeats were subjected to downstream analysis. *E. coli* samples (300 µL treated or untreated) underwent TCA/Acetone incubation to precipitate proteins

(10% TCA v/v for 30 min on ice followed by addition of 400 µL of acetone and further 30 min incubation on ice). Samples were spun for 10 min at 22°C with max speed (16,000 g) and the generated pellets were washed twice with cold acetone and once with cold methanol. The generated pellets were re-solubilized and denatured in 40 µL of 8 M Urea in 50 mM NH_4HCO_3 (pH 8.5). Subsequently, 10 µL was used for protein concentration determination (Pierce™ 660 nm micro–Protein Assay Kit) and 30 µL to set up 200 µL trypsin digestion with addition of the following: 12 µL of 25 mM DTT, 116 µL of 25 mM NH_4HCO_3 (pH 8.5). 2 µL of 100 x digestion enhancer (Promega Corporation) and 40 µL of Promega's Trypsin Platinum solution (100 ng/µL in 25 mM NH_4HCO_3) to 200 µL final volume. Digestion was conducted at 50°C for 3 hrs. The reaction was terminated by acidification with 2.5% Trifluoroacetic Acid (TFA) to 0.3% final concentration (v/v).

2.10.17 Label-free Nano-LC-MS/MS to assess the degree of covalent modifications upon anaerobic heating-cooling treatment. Digested label-free samples were desalted using Agilent Bond Elut OMIX C18 SPE pipette tips per manufacturer protocol and eluted in 20 µL of 70/30/0.1% (v/v) acetonitrile/water/TFA (ACN/H_2O/TFA). The soluble protein sample was dried to completion in the speed-vac and reconstituted in 50 µL of 0.1% (v/v) formic acid. Peptides were analyzed by nano-LC-MS/MS using the Agilent 1100 nanoflow system (Agilent) connected to hybrid linear ion trap-orbitrap mass spectrometer (LTQ-Orbitrap Elite™, ThermoFisher Scientific) equipped with an EASY-Spray™ electrospray source (held at constant 35°C). Chromatography of peptides prior to mass spectral analysis was accomplished using capillary emitter column (PepMap® C18, 3 µM, 100 Å, 150 x 0.075mm, ThermoFisher Scientific) onto which 2 µL of extracted peptides was automatically loaded. NanoHPLC system delivered solvents A: 0.1% (v/v) formic acid , and B: 99.9% (v/v) acetonitrile, 0.1% (v/v) formic acid at 0.50 µL/min to load the peptides (over a 30 min period) and 0.3 µL/min to elute peptides directly into the nano-electrospray with gradual gradient from 0% (v/v) B to 30% (v/v) B over 150 min followed by 10

min fast gradient from 30% (v/v) B to 50% (v/v) B and concluded with a 7 min flash-out from 50-95% (v/v) B.

As peptides eluted from the HPLC-column/electrospray source survey MS scans were acquired in the Orbitrap with a resolution of 120,000 followed by CID-type MS/MS fragmentation of 30 most intense peptides detected in the MS1 scan from 350 to 1800 m/z; redundancy was limited by dynamic exclusion. We assessed the degree of modification of methionine oxidation and asparagine/glutamine deamination on the detected peptides in samples before and after anaerobic heating-cooling procedure. A percent of modified peptide was calculated for each protein and the averages % of modified peptide for all detected proteins from before and after heating-cooling were compared (See SI Figure 1).

2.10.18 Enzymatic "In Liquid" Digestion and Tandem-Mass Tagging labeling. 3 biological repeats from 2 treatments (anaerobic heating-cooling and anaerobic incubation) were subjected to liquid chromatography - mass spectrometry analysis. Each biological repeat consists of 3 samples: 1) control untreated sample, 2) supernatant, and 3) pellet upon treatment, with a total of 9 samples for each treatment. Untreated, and treated soluble samples, 100 µL each in PBS, were denatured with addition of 2 M Urea in 25 mM NH_4HCO_3 (pH 8.5). 3kDa MWCO spin filters (Millipore Amicon® Ultra) were used to spin samples down to 60 µL (14,000 g for 20 min) and 350 µL of 2M Urea in 25mM NH_4HCO_3 (pH 8.5) was added to buffer exchange and spun again at 14,000 g for 20 min. Recovered, unfiltered sample (60 µL) was transferred to a new LoBind® 1.5 mL Eppendorf tube and volume adjusted to final 120 µl with 2M Urea in 25mM NH_4HCO_3 (pH 8.5). Insoluble aggregated/pellet sample was solubilized in 8 M Urea in 25 mM NH_4HCO_3 (pH 8.5) and diluted to 120 µL final volume with 25 mM NH_4HCO_3 (pH 8.5) to a 2 M Urea final concentration in the same prepared volume as soluble samples. All 15 the individual samples were spiked with 5 µL of 0.05 µg/µL (250 ng total) of BSA (internal digestion normalization standard) and formulated for tryptic digestion which was initiated first by the

reduction step with 5 µL of 25 mM DTT and incubation at 56°C for 15 min. After cooling on ice to ambience temperature (22°C), 6 µL of 55mM chloroacetamide (CAA) was added for alkylation where samples were incubated in darkness at ambience temperature for 15 min. This reaction was quenched by 16 µL addition of 25 mM DTT. Subsequently 15 µL of Trypsin/LysC solution (100 ng/µL 1:1 Trypsin, Promega : LysC, FujiFilm mix in 25 mM NH_4HCO_3) with 33 µL of 25 mM NH_4HCO_3 (pH 8.5) was added to treated pellet samples. 30 µL of Trypsin/LysC solution with 18µl of 25mM NH_4HCO_3 (pH 8.5) was added to untreated and treated soluble samples for a final 200 µL volume digestion. Digestions were conducted for 6 hrs at 37°C. Reactions were terminated by acidification with 2.5% TFA to 0.3% (v/v) final.

Digests were cleaned up using Strata™-X 33µm Polymeric Reversed Phase 10 mg SPE cartridges (Phenomenex), where cartridges were first conditioned with 400 µL of 100% methanol, equilibrated in 400 µL of MilliQ water. Then, samples were treated with 0.5% (v/v) heptafluorobutyric acid (HFBA), washed with 400 µL of 0.05% (v/v) TFA. The peptide samples were eluted with 300 µL of 70%:30%:2% (v/v) ACN/H_2O/Formic acid, dried to completion with the speed-vac, and reconstituted in 108 µL of 100mM TEAB. 2 µL of beta-galactosidase digest (1 pmole/µL) was added as an internal standard for labeling and LC-MS instrument analysis. 10 µL of each sample was pooled together for an internal pooled standard as a 16^{th} sample. 100 µL of each of the 16 samples was labeled with 40 µL of labeling reagent from a TMT-16plex™ Labeling Reagent Set (10 µg/µL per labeling reagent, Lot#WB313469, ThermoFisher Scientific) at ambience temperature (22°C) for 1 hr with intermittent gentle vortexing. The labeling reaction was terminated by 8 µL addition of 5% hydroxylamine and 15-min incubation. All 16 labelled samples were pooled together for a 40-µg total master pool, diluted to 4.5% (v/v) final ACN concentration, and acidified with 0.35% TFA (v/v) and 2% HFBA (v/v). The samples were then subjected to solid phase extraction using OMIX C18 SPE cartridges (Agilent) per manufacturer protocol and eluted in 20 µL of 70/30/0.1% (v/v) ACN/H_2O/TFA and dried to completion in the

speed-vac. This 16-multiplexed sample was subsequently resolubilized in 30 µL of 0.1% (v/v) Formic acid and 3% (v/v) ACN.

2.10.19 Bottom-up proteomics via TMT- Nano-LC-MS/MS to assess protein solubility under different conditions. Peptides were analyzed by Orbitrap Fusion™ Lumos™ Tribrid™ platform, where 2 µL was injected using Dionex UltiMate™3000 RSLCnano delivery system (ThermoFisher Scientific) equipped with an EASY-Spray™ electrospray source (held at constant 50°C). Chromatography of peptides prior to mass spectral analysis was accomplished using capillary emitter columns (PepMap® C18, 2 µM, 100 Å, 500 x 0.075 mm, ThermoFisher Scientific). NanoHPLC system delivered solvents A: 0.1% (v/v) formic acid , and B: 80% (v/v) acetonitrile, 0.1% (v/v) formic acid at 0.30 µL/min to load the peptides at 2% (v/v) B, followed by quick 2 min gradient to 5% (v/v) B and gradual analytical gradient from 5% (v/v) B to 62.5% (v/v) B over 203 min when it concluded with rapid 10 min ramp to 95% (v/v) B for a 9 min flash-out. As peptides eluted from the HPLC-column/electrospray source survey MS scans were acquired in the Orbitrap with a resolution of 60,000 followed by HCD-type MS2 fragmentation into Orbitrap (36% collision energy and 30,000 resolution) with 0.7 m/z isolation window in the quadrupole under ddMSnScan 1 s cycle time mode with peptides detected in the MS1 scan from 400 to 1400 m/z; redundancy was limited by dynamic exclusion and MIPS filter mode ON.

2.10.20 Mass spectrometry data analysis. Raw data was directly imported into Proteome Discoverer 2.5.0.400 where protein identifications and quantitative reporting was generated. Seaquest HT search engine platform was used to interrogate Uniprot *Escherichia coli* proteome database (UP000000625, 10/06/2020 download, 4,361 total entries) along with a cRAP common lab contaminant database (116 total entries). Cysteine carbamidomethylation and TMT-16plex specific labeling were selected as static modifications whereas protein N-terminal Met-loss, methionine oxidation and asparagine/glutamine deamidation were selected as dynamic modifications. Peptide mass tolerances were set at 10 ppm for MS1 and 0.6 Da for MS2. Peptide

and protein identifications were accepted under strict 1% false discovery rate (FDR) cut offs with high confidence XCorr thresholds of 1.9 for z=2 and 2.3 for z=3. Reporter Ion Quantifier settings were used on unique and razor peptides, protein grouping was considered for uniqueness. Reporter abundance was based on normalized total peptide amount intensity values with co-isolation threshold filter set at ≤20. Peptide abundances across biological repeats for TMT nano-LC-MS/MS analysis of samples upon a) anaerobic heating-cooling treatment, and upon b) anaerobic incubation treatment was assessed and compared as box plots (See SI Figure 17, panel A).

2.10.21 Bioinformatic analysis of TMT LC-MS data. From each LC-MS data analysis from each treatment (anaerobic heating-cooling and anaerobic incubation), we generated a list of identified proteins with their Uniprot IDs and their peptide abundances in each of the 9 labeled sample. An average percent solubility was calculated for each protein across biological repeats based on its associated peptide abundances in the insoluble and soluble fractions from a specific treatment (i.e., heating-cooling or incubation). The standard errors associated with the calculated solubilities for all proteins were assessed under different treatments as box plots and reproducibility plot (see SI Figure 16, panel B and D). The proteins detected from each treatment with the highest sum PEP score, at least 2 or more peptides detected, and sequence coverages above 20% were selected for analysis. Amongst the detected proteins, we considered 1197 proteins that were identified in both treatments. For each set of proteins proteomic information was collected and analyzed with Microsoft Excel Suite 2019. All data is provided in the Data Availability section.

Information on associated length, amino acid sequence, molecular weight, cofactors, protein families, number of domains, number of subunits, and all Gene Ontology terms (cellular function, biological processes) were collected from the Uniprot database using their Rest API.[129]

This data was collected from Uniprot in August of 2023. All reported numbers for isoelectric points were calculated based on sequences obtained from the Uniprot database.

Information on domain folds were gathered from the SCOP database[130] for a given UniProtKB accession in October of 2022. Assigning of essential and non-essential proteins was done using scoring reported in the Keio collection.[131] Scoring in the KEIO collection is originally reported using gene IDs. To translate these for our analysis we used the NCBI Entrez API[132] to convert gene IDs to Uniprot IDs.

Information on kinetic stability from unfolded states in detergent, thermal stability detected via 2D-SDS PAGE – LC/MS, thermal stability detected via thermal profiling in terms of melting temperature - T_m, protein copy number, solubility upon cell-free synthesis, non-refoldabiity / refoldability relative to non-native states, and usage of solubility tags is assigned using the following publications respectively: Xia et. al. (2007), Kwon et. al. (2008), Mateus et. al. (2018), Wisniewski et. al. (2014), Niwa et. al. (2009), To et. al. (2021), and Costa et. al. (2014). [20, 33, 48-49, 81, 133-134]

2.10.22 Statistical Analysis. Statistical significance was determined by two tailed Student's t-tests with a p value of 0.05, assuming unequal variance, and done for small non-LC-MS/MS data sets (c.a., n < 10). Volcano plots to compare changes in protein abundances across biological repeats were done with p-value (Student's t-test) cutoff of 0.05 and $\log_2$ (change in abundances) cutoff of 1.20. The Wilconox-Mann-Whitney's test with a p-value of 0.05, assuming unpaired data sets, was carried out for all LC-MS/MS data sets.[135-136]

We also represent large sample sets, with at least n = 10, as Tukey style box-and-whisker plots to enhance the visualization of data point distribution (e.g., skewness) for each data set and the downstream comparisons across different data sets.[137-138] Overall, the box-and-whisker plot is represented by a box, defined by sample values located in the 25th lower quartile (Q1) and the 75th upper quartile (Q3). This interquartile range, or IQR = Q3 – Q1, contains 50% of the data. A line inside the IQR box denotes the median of the data. When necessary, a notch is added to

denote the 95% confidence interval relative to the median. The lower and upper whiskers in our study are constructed by the most extreme data points that is within 1.5 × IQR limit (Tukey style). Outliers are defined as data points exceeding of 1.5 × IQR limit from Q1 and Q3. The number of data points, n, is also indicated for each sample set in this study. In addition, when necessary, the average and standard error (SE) of each data set was independently determined and overlapped on the appropriate box plot. All box-and-whiskers plots in this study were constructed with the BoxPlotR web-tool developed by Spitzer et. al. (Nature Methods, 2014).[138]

2.9.1

2.11. Supplementary Results

2.11.1 Homo- and Hetero-aggregation. Our anaerobic heating-cooling experiments induce global unfolding of many proteins in our samples,[48] it is possible that the observed aggregates may be formed only in the presence of other protein aggregates (i.e., heteroaggregation). Many studies suggest that protein aggregation is highly specific (i.e., homoaggregation) both *in vitro* on purified proteins[59, 139-140] and *in vivo* of live cells.[141] We predict that most of the observed aggregates are formed via homoaggregation due to the low sequence identity among *E. coli* proteins[142] and the high specificity of aggregations. Further studies are still required to address the specific structural characteristics of the homo and heteroaggregation. Nevertheless, this does not change our interpretations of the proteome-wide kinetic protection as non-aggregated states relative to homoaggregates or any potential amount of heteroaggregates.

2.11.2 Measurement of protein solubility. The overall solubility of a proteomic sample from each treatment (i.e., anaerobic heating-cooling or anaerobic incubation) was determined by different detection methods, including SDS-PAGE via Coomassie Blue G250 staining, BCA assay, and LC=MS/MS (see SI Table S7). Coomassie Blue G250 dye interacts with protein surfaces, preferably to positively charged residues (i.e., Lys, His, and Arg) and the protein-dye complex formations are quantified via absorption at 595 nm.[143] Protein-to-protein variations of Coomassie Blue G250 detection stem from different amino acid compositions.[114] On the other hand, in BCA assay, in the presence of protein peptide bonds, under alkaline conditions, Cu^{2+} is reduced to Cu^{+}, via oxidation of the protein bonds. The resulting Cu^{+} can be detected by bicinchoninic acid (BCA) via absorption at 562 nm of the Cu^{+}-(BCA)2 complex formation.[114] Protein-to-protein variations of the BCA assay stem from the different lengths of proteins (i.e., number of peptide bonds). This method of quantifying protein concentrations offers higher accuracy than the traditional Bradford assay via Coomassie Blue G250.[144] Despite this, we chose to use the SDS-PAGE solubility assay,

detected with Coomassie Blue dye, due to the ability to visualize the whole proteomic samples in both soluble and insoluble phases. In addition to providing a measurement of protein solubility, this method also allows the tracking of protein changes from different phases of the whole proteome samples, under different treatments (i.e., unfolding or dilution of insoluble aggregates; see Figure 2 and 3). Upon comparison of overall protein solubility of the proteomic samples determined by these two methods, we conclude that the estimated solubility values are similar for the purpose of this work (see SI Table S7).

For individual protein solubility was determined via LC-MS/MS data as detailed in SI Methods 1.21. Average solubility value across the proteomic samples is different from the previous 2 detection methods. However, we argue that the difference in overall solubility resulting from this method does not change our current conclusions. In addition, it allows us to determine specific protein solubility for further downstream analyses and is routinely used in proteome studies.[20-21, 48] For the above reasons, we chose to report the overall solubility from SDS-PAGE assay, detected by Coomassie Blue G250 and estimated by densitometry analysis, and individual protein solubility from LC-MS/MS analysis.

2.11.3 Relationship between solubility and tertiary/quaternary structures. To further validate our unfolding experiments in Figure 2, we compared the extent of unfolding at high GdnHCl concentration of the soluble S100 protein collection and its insoluble aggregates from anaerobic heating-cooling treatments. Tryptophan (Trp) fluorescent emission shows identical spectral center of mass shift for both samples upon unfolding in 6 M GdnHCl (see SI Figure 9, panel A). Similarly, thioflavin T assay shows similar loss of fluorescence of both samples in 6 M GdnHCl (see SI Figure 9, panel B). We concluded from these results that the unfolded states of both soluble S100 protein sample and insoluble aggregate sample, from before and after anaerobic heating-cooling respectively, are quite similar.

We also performed CD measurements on these 2 types of proteomic samples (see SI Figure 10). Soluble S100 protein samples are unfolded in 6 M GdnHCl. No CD data was obtained for insoluble aggregates before unfolding due to high levels of Rayleigh scattering that can significantly contribute to CD absorption profiles. However, we expect our samples to contain residual tRNA, which is known to be CD-active, in addition to >2,000 E. coli proteins. Therefore, due to the complexity of the sample, we simply concluded that, based on the CD profiles, proteins are being unfolded upon exposure to 6M GdnHC

2.11.4 Proteome-wide kinetic trapping is not due to ATP-dependent molecular machines.

We first want to address whether the observed high solubility as non-aggregated states from the anaerobic incubation treatment may be due to active endogenous de-aggregation and chaperone systems in our S100 protein collection. Since major chaperone and disaggregation systems in *E. coli* are ATP-dependent, we hypothesize that these systems should not be active due to the depletion of ATP levels during our S100 sample preparation. Therefore, we expect these systems to have minimal contribution to the observed high solubility from our incubation treatment. To test the following hypotheses, we carried out the anaerobic incubation treatment on S100 samples in the presence of cellular levels of ATP and/or Hsp70 systems. Specifically, the Hsp70 chaperone system was chosen in this experiment because it is known to interact with a large fraction of cytoplasmic proteins in *E. coli*.[17]

S100 samples at total protein concentration of 5 mg/mL were incubated in two additional conditions: 1) in the presence of 5 mM ATP,[145] and 2) in the presence of 5 mM ATP and additional Hsp70 chaperone system, consisting of DnaK : DnaJ : GrpE at concentrations of 54 μM : 135 μM : 270 μM respectively.[146] From SDS-PAGE solubility assay after incubation in the absence and presence of cellular levels of ATP, the solubilities were not different, ca. 82 ± 1% and 83 ± 1% respectively (SI Figure, Panel A). Similarly, incubation of S100 samples in the presence of ATP and additional Hsp70 chaperone yielded a solubility value of 82 ± 1%. Regularization analysis of

the DLS data in all conditions showed that the soluble fractions consisted of small species (R_h < 5 nm) after incubation (SI Figure, panel B).

In summary, we concluded that ATP-dependent systems, including the Hsp70 chaperone system, do not contribute to the observed passive kinetic protection of *E. coli* proteome as soluble non-aggregated states relative to aggregation at physiological conditions. The activation of endogenous ATP dependent systems through incubation in cellular levels of ATP did not change the overall solubility of the S100 sample upon incubation. One possible explanation is that the ATP-dependent chaperone and proteolytic systems were not functional due to the nature of our S100 sample preparation. Nevertheless, this would lead to the same conclusions based on the similar protein solubilities between samples in the presence and absence of ATP. Interestingly, incubation of cellular levels of Hsp70 system and ATP also did not change the overall solubility, or decrease of aggregation, upon incubation at physiologically relevant conditions. This result suggests that the Hsp70 system alone cannot prevent flux to aggregated states. This is consistent with current views that the Hsp70 system has been shown to disaggregate insoluble protein aggregates only in the presence of the protein-disaggregating chaperone ClpB.[147-148]

2.11.5 Reversibility of protein aggregation on long-timescales upon dilution. Results from Figures 3 and 5 show that soluble and insoluble aggregates are under reversible exchange upon progressive dilution. The timescale for this reversible exchange (i.e., 2 hrs equilibration) exceeds that of the E. coli doubling time. As previously mentioned, similar dilution experiments have been carried out on amyloid fibrils to demonstrate reversibility of aggregation. A few studies suggest that fibrils can revert to non-aggregated states upon sufficient dilution and time for equilibration. Therefore, we further allow the diluted insoluble aggregates of the E. coli proteomic samples to equilibrate for 20 hrs (See SI Methods 1.15). The autocorrelation curve from DLS measurement shows presence of a smaller species in the soluble phase upon 20 hrs post-dilution equilibration (see SI Figure 16, panel A). Regularization analysis shows that the estimated apparent

hydrodynamic radius distribution is 68 nm (see SI Figure 16, panel B). Due to the complexity of the proteomic samples, we simply conclude that, upon 20 hrs post-dilution equilibration, the soluble fraction contains mainly large soluble aggregates, exceeding the relevant timescales of this work.

2.11.6 Our *E. coli* proteome is representative subset of cytoplasmic and periplasmic proteins. Our sample was prepared from A19 *E. coli* cells collected at the start of exponential phase (OD_{600} = 0.8) of this strain.[149-150] LC-MS analysis successfully identified a subset of 1197 *E coli* proteins in our S100 proteomic samples, with enrichment of cytoplasmic, periplasmic, and ribosomal proteins upon comparison to a representative proteome of 2,237 proteins in Wisniewski et. al. (2014).[81] Upon inspection of our data based on cellular locations characterized by the GO database, 995 out of 1197 cytoplasmic proteins (c.a., 83%) were identified across biological repeats (See SI Figure S17, panel A). Similarly, we identified 106 periplasmic proteins (c.a., 9%), 43 membrane proteins (c.a., 4%), and 48 proteins with no associated cellular location (c.a., 4%). When compared with the representative proteome, our sample is representative of 90% cytoplasmic *E. coli* proteins (See SI Figure S17, panel B) and representative of 72% periplasmic *E. coli* protein upon comparison to the representative proteome. Since ribosomal proteins are expected to be enriched for A19 cell cultures at growth phase,[79] we chose to reduce the amounts of these proteins through ultracentrifugation. Despite this fact, we still detected 51 out of 54 ribosomal proteins within our samples. In addition, our sample is underrepresenting membrane proteins with only 43 detected, making up of only 6% of *E. coli* membrane proteins. This is consistent with previous reports where most membrane proteins are not present within S100 samples.[20-21]

Next, we analyze the protein compositions of our proteomic samples based on involvement in different biological processes from the SCOPs database. We expect that the S100 proteome sample contains many important cytoplasmic and periplasmic pathways (e.g.,

transcription-translation, catabolism/metabolism, and protein quality control/degradation), indicative of proteomes prepared from exponentially growing A19 cells as previously reported.[149-150] Satisfyingly, the LC-MS analysis detected important proteins required for translation, including 51 out of 54 ribosomal proteins and all 24 essential aminoacyl-tRNA synthetases. We also detected proteins involved in purine / pyrimidine biosynthesis and energy generating metabolic/catabolic pathways involving fatty acids and other carbohydrates (e.g., glycolysis/gluconeogenesis, oxidative pentose phosphate pathway - OPPP, tricarboxylic acid cycle – TCA cycle, etc.). In addition, our proteomic samples contain many major proteins involved in quality control and degradation in the cytoplasm of *E. coli*, including molecular chaperones, disaggregases, and proteases. To our surprise, membrane-associated proteins involved in the respiratory chain, including ATP-synthase complex, and NADH-quinone oxidoreductase were detected.

2.11.7 Protein solubility: comparisons with previous studies. To further validate our results, we compared our solubility data under both treatments (i.e., anaerobic heating-cooling and anaerobic incubation) with previously reported solubility data in the current literature. We identified all 16 proteins that were characterized as heat stable proteins in Kwon et. al. (2009, see Figure 33, panel A). Solubility data upon anerobic heating-cooling shows that 12 out of 16 proteins are more soluble (i.e., heat stable) than the other proteins from the E. coli proteomic sample. Our results further characterized all 16 of these proteins to be kinetically stable upon anaerobic incubation at physiologically relevant conditions. Similarly, 106 out of 126 proteins with low thermal stability, identified by Mateus et. al. in terms of melting temperature (i.e., Tm < 52 °), are also aggregation prone upon anaerobic heating-cooling at 70 °C (see SI Figure 33, panel B and D). However, we found that 99 proteins with high thermal stability (i.e., 65 °C < Tm < 70 °C or Tm > 70 °C, see SI Figure 33, panel A and C) show mixed solubility profiles under both treatments. Finally, we identified 41 out of 50 proteins, which were found to be kinetically stable relative to the

unfolded states in Xia et. al. (2007), to also be kinetically stable relative to aggregates under physiologically relevant conditions (see SI Figure 32, panel B)[133]. Overall, we concluded that our solubility data for both treatments is supported by protein solubility and stability findings from previous studies. However, we would also like to emphasize that the data in this work is the first extensive collection of E. coli proteins (c.a., 1197 proteins) to report on the correlation between each protein's the propensity to aggregate upon thermal treatment and the associated kinetic stability relative to aggregation under physiological conditions.

2.12. Supplementary References

1. 1. Cole, S. D.; Miklos, A. E.; Chiao, A. C.; Sun, Z. Z.; Lux, M. W., Methodologies for preparation of prokaryotic extracts for cell-free expression systems. *Synth Syst Biotechnol* **2020,** *5* (4), 252-267.
2. Cui, Y.; Chen, X.; Wang, Z.; Lu, Y., Ribosome purification from Escherichia coli by ultracentrifugation. *Biotechnology Notes* **2022,** *3*, 118-123.
3. Felske, A.; Engelen, B.; Nübel, U.; Backhaus, H., Direct ribosome isolation from soil to extract bacterial rRNA for community analysis. *Appl Environ Microbiol* **1996,** *62* (11), 4162-7.
4. To, P.; Whitehead, B.; Tarbox, H. E.; Fried, S. D., Nonrefoldability is Pervasive Across the E. coli Proteome. *Journal of the American Chemical Society* **2021,** *143* (30), 11435-11448.
5. To, P.; Xia, Y.; Lee, S. O.; Devlin, T.; Fleming, K. G.; Fried, S. D., A proteome-wide map of chaperone-assisted protein refolding in a cytosol-like milieu. *Proceedings of the National Academy of Sciences US* **2022,** *119* (48), e2210536119.
6. Cortés-Ríos, J.; Zárate, A. M.; Figueroa, J. D.; Medina, J.; Fuentes-Lemus, E.; Rodríguez-Fernández, M.; Aliaga, M.; López-Alarcón, C., Protein quantification by bicinchoninic acid (BCA) assay follows complex kinetics and can be performed at short incubation times. *Analytical Biochemistry* **2020,** *608*, 113904.
7. Linse, S., Toward the equilibrium and kinetics of amyloid peptide self-assembly. *Curr. Opin. Struct. Biol.* **2021,** *70*, 87-98.
8. Hixson, A. W.; Crowell, J. H., Dependence of Reaction Velocity upon surface and Agitation. *Ind. Eng. Chem.* **1931,** *23* (8), 923-931.
9. Galea, C. A.; Pagala, V. R.; Obenauer, J. C.; Park, C.-G.; Slaughter, C. A.; Kriwacki, R. W., Proteomic Studies of the Intrinsically Unstructured Mammalian Proteome. *J. Proteome Res.* **2006,** *5* (10), 2839-2848.
10. Hershewe, J. M.; Warfel, K. F.; Iyer, S. M.; Peruzzi, J. A.; Sullivan, C. J.; Roth, E. W.; DeLisa, M. P.; Kamat, N. P.; Jewett, M. C., Improving cell-free glycoprotein synthesis by characterizing and enriching native membrane vesicles. *Nat. Comm.* **2021,** *12* (1), 2363.
11. Stetefeld, J.; McKenna, S. A.; Patel, T. R., Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys. Rev.* **2016,** *8* (4), 409-427.
12. Scholtz, J. M.; Grimsley, G. R.; Pace, C. N., Chapter 23 Solvent Denaturation of Proteins and Interpretations of the m Value. In *Methods in Enzymology*, Academic Press: 2009; Vol. 466, pp 549-565.
13. de Silva, A. P.; de Silva, S. A., Fluorescent signalling crown ethers; ‘switching on’ of fluorescence by alkali metal ion recognition and binding in situ. *J. Chem. Soc., Chem. Commun.* **1986,** (23), 1709-1710.
14. Addabbo, R. M.; Dalphin, M. D.; Mecha, M. F.; Liu, Y.; Staikos, A.; Guzman-Luna, V.; Cavagnero, S., Complementary Role of Co- and Post-Translational Events in De Novo Protein Biogenesis. *J. Phys. Chem. B* **2020,** *124* (30), 6488-6507.
15. Santoro, M. M.; Bolen, D. W., Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl .alpha.-chymotrypsin using different denaturants. *Biochemistry* **1988,** *27* (21), 8063-8068.
16. Pace, C. N., Measuring and increasing protein stability. *Trends Biotechnol.* **1990,** *8* (4), 93-8.
17. Na, Y. R.; Park, C., Investigating protein unfolding kinetics by pulse proteolysis. *Protein Sci.* **2009,** *18* (2), 268-76.
18. Welte, H.; Kovermann, M., Insights into Protein Stability in Cell Lysate by 19F NMR Spectroscopy. *ChemBioChem* **2020,** *21* (24), 3575-3579.

19. Narimoto, T.; Sakurai, K.; Okamoto, A.; Chatani, E.; Hoshino, M.; Hasegawa, K.; Naiki, H.; Goto, Y., Conformational stability of amyloid fibrils of beta2-microglobulin probed by guanidine-hydrochloride-induced unfolding. *FEBS Lett.* **2004,** *576* (3), 313-9.
20. Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-Cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M., Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *Journal of the American Chemical Society* **2011,** *133* (36), 14160-14163.
21. Vettore, N.; Buell, A. K., Thermodynamics of amyloid fibril formation from chemical depolymerization. *Physical Chemistry Chemical Physics* **2019,** *21* (47), 26184-26194.
22. Groenning, M.; Norrman, M.; Flink, J. M.; van de Weert, M.; Bukrinsky, J. T.; Schluckebier, G.; Frokjaer, S., Binding mode of Thioflavin T in insulin amyloid fibrils. *J Struct Biol* **2007,** *159* (3), 483-97.
23. Xue, C.; Lin, T. Y.; Chang, D.; Guo, Z., Thioflavin T as an amyloid dye: fibril quantification, optimal concentration and effect on aggregation. *R Soc Open Sci* **2017,** *4* (1), 160696.
24. Novo, M.; Freire, S.; Al-Soufi, W., Critical aggregation concentration for the formation of early Amyloid-β (1–42) oligomers. *Sci. Rep.* **2018,** *8* (1), 1783.
25. O'Nuallain, B.; Shivaprasad, S.; Kheterpal, I.; Wetzel, R., Thermodynamics of Aβ(1−40) Amyloid Fibril Elongation. *Biochemistry* **2005,** *44* (38), 12709-12718.
26. Shahnawaz, M.; Soto, C., Microcin amyloid fibrils A are reservoir of toxic oligomeric species. *J Biol Chem* **2012,** *287* (15), 11665-76.
27. Nespovitaya, N.; Gath, J.; Barylyuk, K.; Seuring, C.; Meier, B. H.; Riek, R., Dynamic Assembly and Disassembly of Functional β-Endorphin Amyloid Fibrils. *Journal of the American Chemical Society* **2016,** *138* (3), 846-856.
28. Nespovitaya, N.; Gath, J.; Barylyuk, K.; Seuring, C.; Meier, B.; Riek, R., Dynamic Assembly and Disassembly of Functional β-Endorphin Amyloid Fibrils. *Journal of the American Chemical Society* **2015,** *138*.
29. Consortium, T. U., UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **2022,** *51* (D1), D523-D531.
30. Lo Conte, L.; Ailey, B.; Hubbard, T. J.; Brenner, S. E.; Murzin, A. G.; Chothia, C., SCOP: a structural classification of proteins database. *Nucleic Acids Res* **2000,** *28* (1), 257-9.
31. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H., Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2006,** *2*, 2006.0008.
32. Sayers, E. W.; Bolton, E. E.; Brister, J. R.; Canese, K.; Chan, J.; Comeau, D. C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; Madej, T.; Marchler-Bauer, A.; Lanczycki, C.; Lathrop, S.; Lu, Z.; Thibaud-Nissen, F.; Murphy, T.; Phan, L.; Skripchenko, Y.; Tse, T.; Wang, J.; Williams, R.; Trawick, B. W.; Pruitt, K. D.; Sherry, S. T., Database resources of the national center for biotechnology information. *Nucleic Acids Res* **2022,** *50* (D1), D20-d26.
33. Xia, K.; Manning, M.; Hesham, H.; Lin, Q.; Bystroff, C.; Colón, W., Identifying the subproteome of kinetically stable proteins via diagonal 2D SDS/PAGE. *Proceedings of the National Academy of Sciences* **2007,** *104* (44), 17329-17334.
34. Kwon, S.; Jung, Y.; Lim, D., Proteomic analysis of heat-stable proteins in Escherichia coli. *BMB Rep* **2008,** *41* (2), 108-11.
35. Mateus, A.; Bobonis, J.; Kurzawa, N.; Stein, F.; Helm, D.; Hevler, J.; Typas, A.; Savitski, M. M., Thermal proteome profiling in bacteria: probing protein state in vivo. *Mol. Syst. Biol.* **2018,** *14* (7), e8242.
36. Niwa, T.; Ying, B.-W.; Saito, K.; Jin, W.; Takada, S.; Ueda, T.; Taguchi, H., Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of

Escherichia coli proteins. *Proceedings of the National Academy of Sciences* **2009,** *106* (11), 4201-4206.
37. Costa, S.; Almeida, A.; Castro, A.; Domingues, L., Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Front. Microbiol.* **2014,** *5*.
38. Wiśniewski, J. R.; Hein, M. Y.; Cox, J.; Mann, M., A “Proteomic Ruler” for Protein Copy Number and Concentration Estimation without Spike-in Standards*. *Mol. Cell. Proteomics* **2014,** *13* (12), 3497-3506.
39. Vinaixa, M.; Samino, S.; Saez, I.; Duran, J.; Guinovart, J. J.; Yanes, O., A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites* **2012,** *2* (4), 775-95.
40. Russell, M. R.; Lilley, K. S., Pipeline to assess the greatest source of technical variance in quantitative proteomics using metabolic labelling. *J. Proteomics* **2012,** *77*, 441-454.
41. Krzywinski, M.; Altman, N., Visualizing samples with box plots. *Nat. Methods* **2014,** *11* (2), 119-120.
42. Spitzer, M.; Wildenhain, J.; Rappsilber, J.; Tyers, M., BoxPlotR: a web tool for generation of box plots. *Nat. Methods* **2014,** *11* (2), 121-2.
43. Speed, M. A.; Wang, D. I. C.; King, J., Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition. *Nat. Biotechnol.* **1996,** *14* (10), 1283-1287.
44. Jones, E. M.; Surewicz, W. K., Fibril Conformation as the Basis of Species- and Strain-Dependent Seeding Specificity of Mammalian Prion Amyloids. *Cell* **2005,** *121* (1), 63-72.
45. Morell, M.; Bravo, R.; Espargaró, A.; Sisquella, X.; Avilés, F. X.; Fernàndez-Busquets, X.; Ventura, S., Inclusion bodies: Specificity in their aggregation process and amyloid-like structure. *Biochim. Biophys. Acta* **2008,** *1783* (10), 1815-1825.
46. Rajan, R. S.; Illing, M. E.; Bence, N. F.; Kopito, R. R., Specificity in intracellular protein aggregation and inclusion body formation. *Proceedings of the National Academy of Sciences* **2001,** *98* (23), 13060-13065.
47. Blattner, F. R.; Plunkett, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y., The Complete Genome Sequence of Escherichia coli K-12. *Science* **1997,** *277* (5331), 1453-1462.
48. Candiano, G.; Bruschi, M.; Musante, L.; Santucci, L.; Ghiggeri, G. M.; Carnemolla, B.; Orecchia, P.; Zardi, L.; Righetti, P. G., Blue silver: A very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **2004,** *25* (9), 1327-1333.
49. Glyk, A.; Heinisch, S. L.; Scheper, T.; Beutel, S., Comparison of colorimetric methods for the quantification of model proteins in aqueous two-phase systems. *Analytical Biochemistry* **2015,** *477*, 35-37.
50. Calloni, G.; Chen, T.; Schermann, Sonya M.; Chang, H.-c.; Genevaux, P.; Agostini, F.; Tartaglia, Gian G.; Hayer-Hartl, M.; Hartl, F. U., DnaK Functions as a Central Hub in the E. coli Chaperone Network. *Cell Rep.* **2012,** *1* (3), 251-264.
51. Yaginuma, H.; Kawai, S.; Tabata, K. V.; Tomiyama, K.; Kakizuka, A.; Komatsuzaki, T.; Noji, H.; Imamura, H., Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Sci. Rep.* **2014,** *4* (1), 6522.
52. Mogk, A.; Tomoyasu, T.; Goloubinoff, P.; Rüdiger, S.; Röder, D.; Langen, H.; Bukau, B., Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB. *EMBO J* **1999,** *18* (24), 6934-49.
53. Doyle, S. M.; Shastry, S.; Kravats, A. N.; Shih, Y.-H.; Miot, M.; Hoskins, J. R.; Stan, G.; Wickner, S., Interplay between E. coli DnaK, ClpB and GrpE during Protein Disaggregation. *Journal of Molecular Biology* **2015,** *427* (2), 312-327.

54. Doyle, S. M.; Hoskins, J. R.; Wickner, S., DnaK Chaperone-dependent Disaggregation by Caseinolytic Peptidase B (ClpB) Mutants Reveals Functional Overlap in the N-terminal Domain and Nucleotide-binding Domain-1 Pore Tyrosine*. *Journal of Biological Chemistry* **2012,** *287* (34), 28470-28479.
55. Failmezger, J.; Rauter, M.; Nitschel, R.; Kraml, M.; Siemann-Herzberg, M., Cell-free protein synthesis from non-growing, stressed Escherichia coli. *Sci. Rep.* **2017,** *7* (1), 16524.
56. Foshag, D.; Henrich, E.; Hiller, E.; Schäfer, M.; Kerger, C.; Burger-Kentischer, A.; Diaz-Moreno, I.; García-Mauriño, S. M.; Dötsch, V.; Rupp, S.; Bernhard, F., The E. coli S30 lysate proteome: A prototype for cell-free protein production. *N. Biotechnol.* **2018,** *40*, 245-260.
57. Bakshi, S.; Siryaporn, A.; Goulian, M.; Weisshaar, J. C., Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Molecular Microbiology* **2012,** *85* (1), 21-38.

2.13. Supplementary Tables

SUPPLEMENTARY TABLES

Supplementary Table S1. DLS regularization analysis of soluble samples upon anaerobic heating-cooling. For each entry, 4-6 biological repeats were measured. See SI Methods 1.9 for more details on DLS measurement.

	Peak 1 *		**Peak 2 ****		**Peak 3 *****	
Sample identity	Rh (nm)	% mass	Rh (nm)	% mass	Rh (nm)	% mass
Untreated control	3.2 ± 0.1	61 ± 1	8.4 ± 1.5	14 ± 2	47 ± 2	22 ± 3
Anaerobic heating-cooling	2.8 ± 0.5	79 ± 3	11 ± 2	18 ± 3	38 ± 7	2.6 ± 0.6
Dilution of insoluble aggregates	–	–	–	–	79 ± 5	100

* denotes particle size for non-aggregated states

** denotes particle size for oligomeric and small soluble aggregated states

*** denotes particle size for large soluble aggregates

Supplementary Table S2. DLS regularization analysis of soluble samples upon anaerobic incubation at 37°C (n = 3 to 6). For each entry, 4-6 biological repeats were measured. See SI Methods 1.9 for more details on DLS measurement.

	Peak 1 *		Peak 2 **		Peak 3 ***	
Sample identity	Rh (nm)	% mass	Rh (nm)	% mass	Rh (nm)	% mass
Untreated control	3.2 ± 0.1	61 ± 1	8.4 ± 1.5	14 ± 2	47 ± 2	22 ± 3
Anaerobic incubation at 37 °C	4.0 ± 0.3	85 ± 3	13 ± 3	12 ± 3	47 ± 9	2.5 ± 0.4
Dilution of insoluble aggregates	–	–	–	–	67 ± 6	100

* denotes particle size for non-aggregated states

** denotes particle size for oligomeric and small soluble aggregated states

*** denotes particle size for large soluble aggregates

Supplementary Table S3. Regularization analysis of DLS profile for soluble fractions upon anaerobic heating-cooling with or without additional incubation at 22 °C for 20 hrs (See Figure 1, panel D and Figure 2, panel D). For each entry, 3 biological repeats were measured.

	Peak 1[a]		**Peak 2[b]**		**Peak 3[c]**	
Sample identity	Rh (nm)	% mass	Rh (nm)	% mass	Rh (nm)	% mass
Untreated control	3.2 ± 0.1	61 ± 1	8.4 ± 1.5	14 ± 2	47 ± 2	22 ± 3
Anaerobic heating-cooling	2.8 ± 0.5	79 ± 3	11 ± 2	18 ± 3	38 ± 7	2.6 ± 0.6
Post- heating-cooling and incubate for 20hr	2.6 ± 0.3	78 ± 6	7 ± 1	22 ± 6	23 ± 2	7 ± 4

[a] denotes particle size for non-aggregated states

[b] denotes particle size for oligomeric and small soluble aggregated states

[c] denotes particle size for large soluble aggregates

Supplementary Table S4. Average number of peaks resolved from regularization analysis of Dynamic Light Scattering (DLS) profile of soluble fractions under different treatments (see Figure 1, panel D and Figure 3, panel C). For each entry, 3 biological repeats were measured.

Sample identity	Average number of peaks
Untreated control	3
Anaerobic heating-cooling (Figure 2)	3.3 ± 0.3
Post-heating-cooling and incubate for 20 hr (SI Figure S5)	3
Dilution of heat-cool insoluble aggregates (Figure 2)	1
Anaerobic incubation at 37 °C (Figure 3)	3
Dilution of incubated insoluble aggregates (Figure 5)	1.3 ± 0.3

Supplementary Table S5. DLS cumulant analysis of soluble fractions upon anaerobic heating-cooling (n = 3 to 6).

Incubation at 37 °C experiments		
Sample Identity	Rh (nm)	SOS
Control	29 ± 0.7	10.7
Anaerobic incubation at 37 °C	12 ± 0.3	9.1
Dilution of insoluble aggregates	59 ± 5	1.4

SOS denotes Sum of least squares

Supplementary Table S6. DLS cumulant analysis of soluble fractions upon anaerobic incubation at 37 °C (n = 3 to 6).

Anaerobic heating-cooling experiments		
Sample Identity	Rh (nm)	SOS
Untreated control	29 ± 0.7	10.7
Anaerobic heating-cooling	13 ± 0.3	4.2
Dilution of insoluble aggregates	68 ± 1	3.9

SOS denotes Sum of Least Squares

Supplementary Table S7. Measurements of percent of protein solubility shows similar values when detected with different methods. Protein % solubility measurement for each method was done on 3 biological repeats.

Detection method	% solubility upon anaerobic heating-cooling	% solubility upon anaerobic incubation
SDS-PAGE, Coomassie Blue G250 (SI Methods 1.5)	13 ± 1	93 ± 2
BCA Assay (SI Methods 1.2)	14.0 ± 0.4	88.0 ± 0.4
LC-MS/MS (SI Methods 1.19)	17 ± 2	80 ± 2

Supplementary Table S8. The apparent critical aggregate concentrations, app. [CAC], for soluble aggregate is small relative to the total protein concentration (i.e., 5 mg/mL). Due to the low concentration of the solubilized protein aggregates upon dilution of insoluble aggregates and the complex nature of the proteome samples, it is not possible to determine the concentrations directly using traditional methods. Therefore, we chose to estimate the app. [CAC] from data obtained in dilution of insoluble aggregates in both treatments.

Sample ID	Apparent [CAC] (μg/mL)[1]
Aggregates from anaerobic heating-cooling (c.a., 95% of total protein concentration)	50 ± 9
Aggregates from anaerobic incubation (c.a., 10% of total protein concentration)	11 ± 2

[1] Three biological repeats were measured.

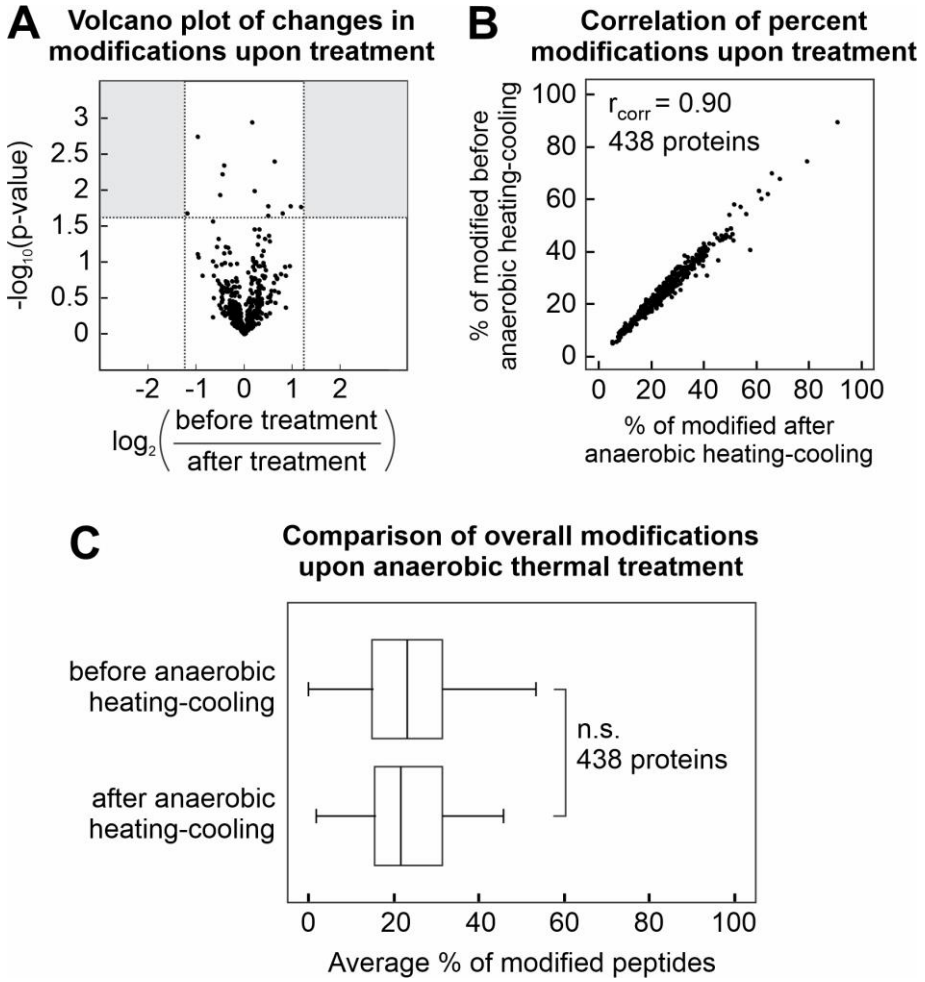
Supplementary Table S9, part 1. Solubilities and functions of important cytoplasmic proteins involved in protein quality control and degradation in the cytoplasm.

Protein name	UniProt ID	Cellular functions	% sol (heating-cooling)	% sol (incubation)
Trigger factor	P0A850	co-translational chaperone, interact with newly synthesized proteins	17.8	94.6
DnaK chaperone (Hsp70)	P0A6Y8	chaperone of the Hsp70 system, facilitates protein folding and prevent aggregation	57.7	96.4
DnaJ chaperone (Hsp40)	P08622	co-chaperone of the Hsp70 system; have disaggregation capabilities autonomously	6.2	38.4
nucleotide exchange factor GrpE (Hsp24)	P09372	nucleotide exchange factor of the Hsp70 chaperone system	12.0	90.1
GroEL (Hsp60)	P0A6F5	part of the GroEL/S chaperone system to assist protein folding	6.6	93.7
GroES (Hsp10)	P0A6F9	part of the GroEL/S chaperone system to assist protein folding	87.9	96.2
HtpG chaperone (Hsp 90)	P0A6Z3	acts as a chaperone with ATPase activity	5.2	93.2
Chaperonin 33 kDa (Hsp33)	P0A6Y5	protects thermally unfolding and oxidatively damaged proteins from irreversible aggregation	43.4	90.1
ClpB chaperone (Hsp100)	P63284	disaggregates in cooperation with Hsp70 system	7.4	93.6
HscA chaperone	P0A6Z1	involve in the maturation of iron-sulfur cluster-containing proteins	5.6	91.6

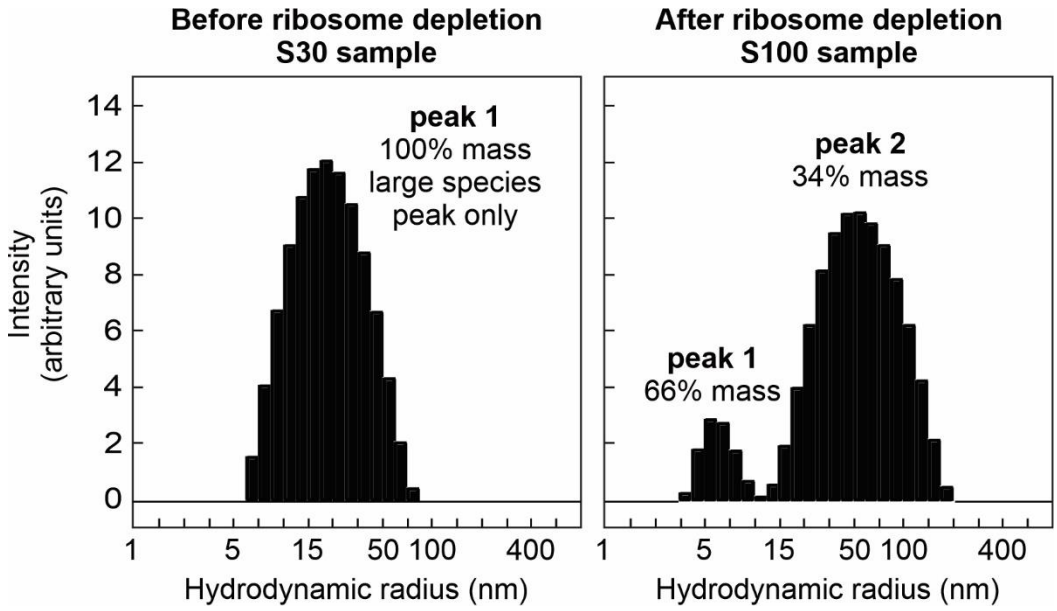
Supplementary Table S10. Solubilities and functions of periplasmic proteins involved in quality control and degradation.

Protein name	UniProt ID	Cellular functions	%Sol (heating-cooling)	%Sol (incubation)
Skp Chaperone	P0AEU7	interacts with outer membrane proteins to maintain solubility of early folding intermediates during passage through periplasm	56.5	87.9
SurA Chaperone	P0ABZ6	involves in the correct folding and assembly of outer membrane proteins	5.4	94.3
HdeA, acid stress chaperone	P0AES9	exhibits chaperone-like activity at pH below 3, suppresses the aggregation of denatured periplasmic proteins	91.1	91.1
HdeB, acid stress chaperone	P0AET2	exhibits chaperone-like activity at pH below 3, suppresses the aggregation of denatured periplasmic proteins	78.8	94.9
FkpA isomerase	P45523	FKBP-type peptidyl-prolyl cis-trans isomerase, accelerates the folding of proteins via cis-trans isomerization	82.9	96.8
DsbA, disulfide isomerase	P0AEG4	involves in disulfide bond formation	57.5	97.7
DsbC, disulfide isomerase	P0AEG6	corrects non-native disulfide bonds in misfolded proteins	52.9	94.5
DsbG, disulfide isomerase	P77202	involves in disulfide bond formation	21.4	90.1
DegP, periplasmic serine endoprotease (Hsp)	P0C0V0	acts as a chaperone at low temperatures; switches to a peptidase (heat shock protein) at higher temperatures	9.5	88.1
PpiD, periplasmic transmembrane chaperone	P0ADY1	facilitates the early periplasmic folding of newly translocated proteins across cell inner membrane	13.4	46.8

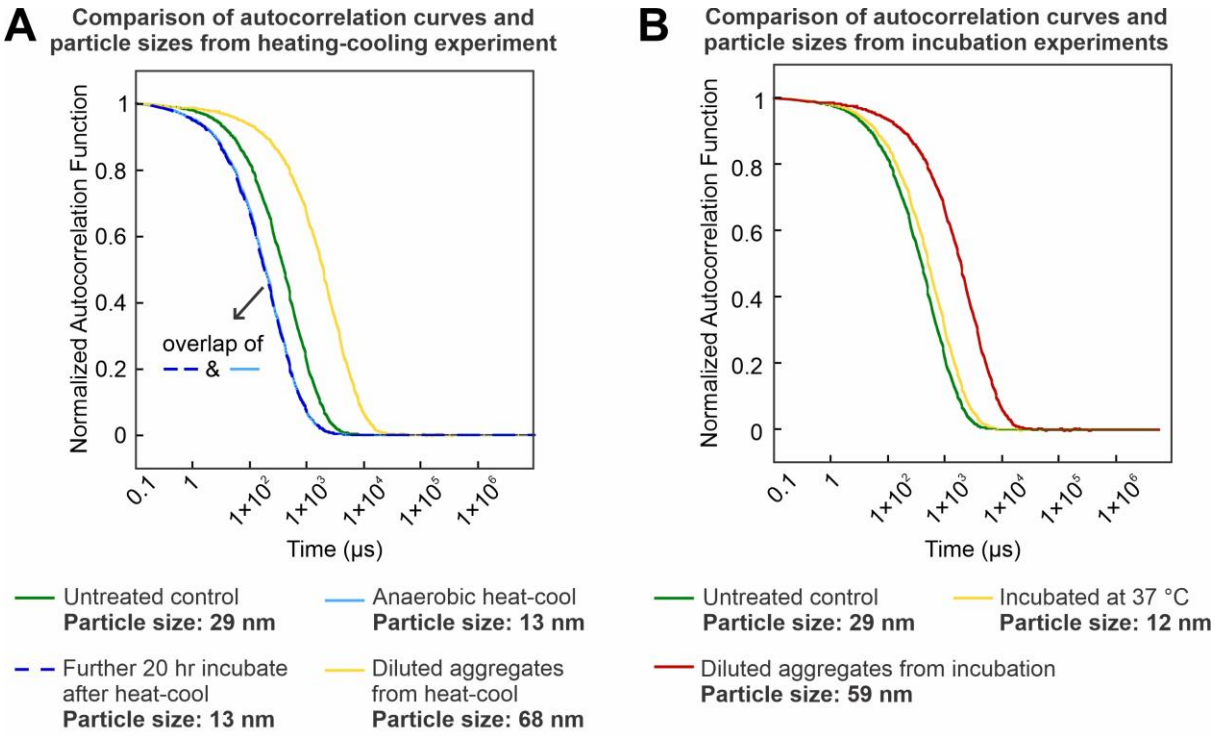
2.14. Supplementary Figures



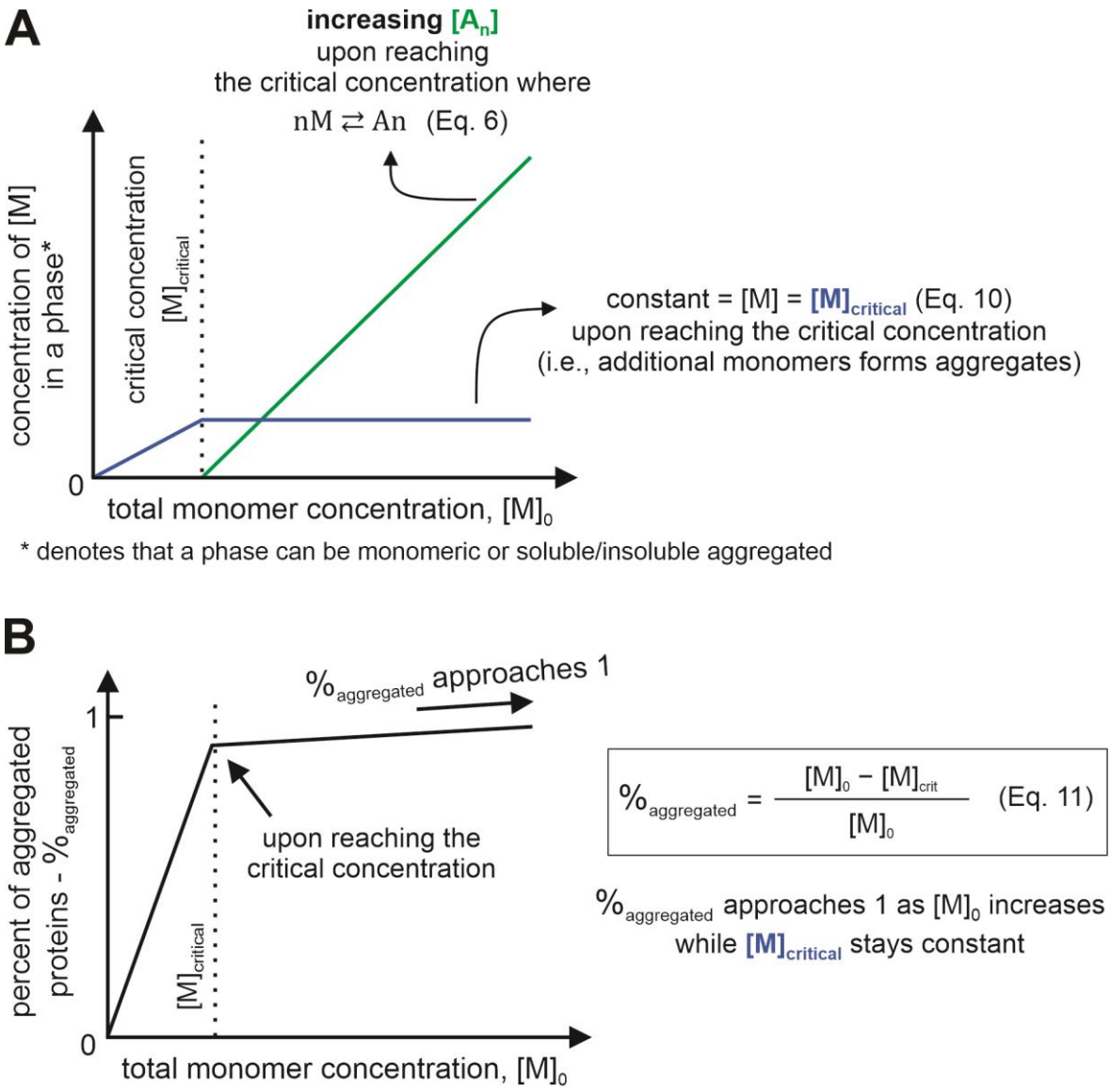
Supplementary Figure S1. Heat-associated covalent modifications before and after anaerobic heating-cooling of *E. coli* S100 protein collection. We examined methionine oxidation and glutamine/asparagine deamination with Mascot. 3 biological repeats were done. (A) Volcano plot shows that all proteins show similar level of heat-associated covalent modifications. P-value cut-off is at 0.05 significant level; cutoff of $\log_2$ (changes upon treatment) is at ± 1.2 (B) Correlation between % modified peptides before and after heating-cooling for each protein (total: 438 proteins) shows a linear profile; r_{corr} = 0.9 (Pearson's correlation test). Proteins with noticeable changes in the level of modification upon treatment will deviate from the linear diagonal of the correlation plot. (C) Comparison of overall % of modifications before and after anaerobic heating-cooling shows similar values; P-values (Wilcoxon-Mann-Whitney's test) = 0.81, n = 100, n.s.: not significant. Detailed methods for label-free LC-MS/MS and peptide data analysis in SI Methods 1.16-17 and 1.20 respectively.



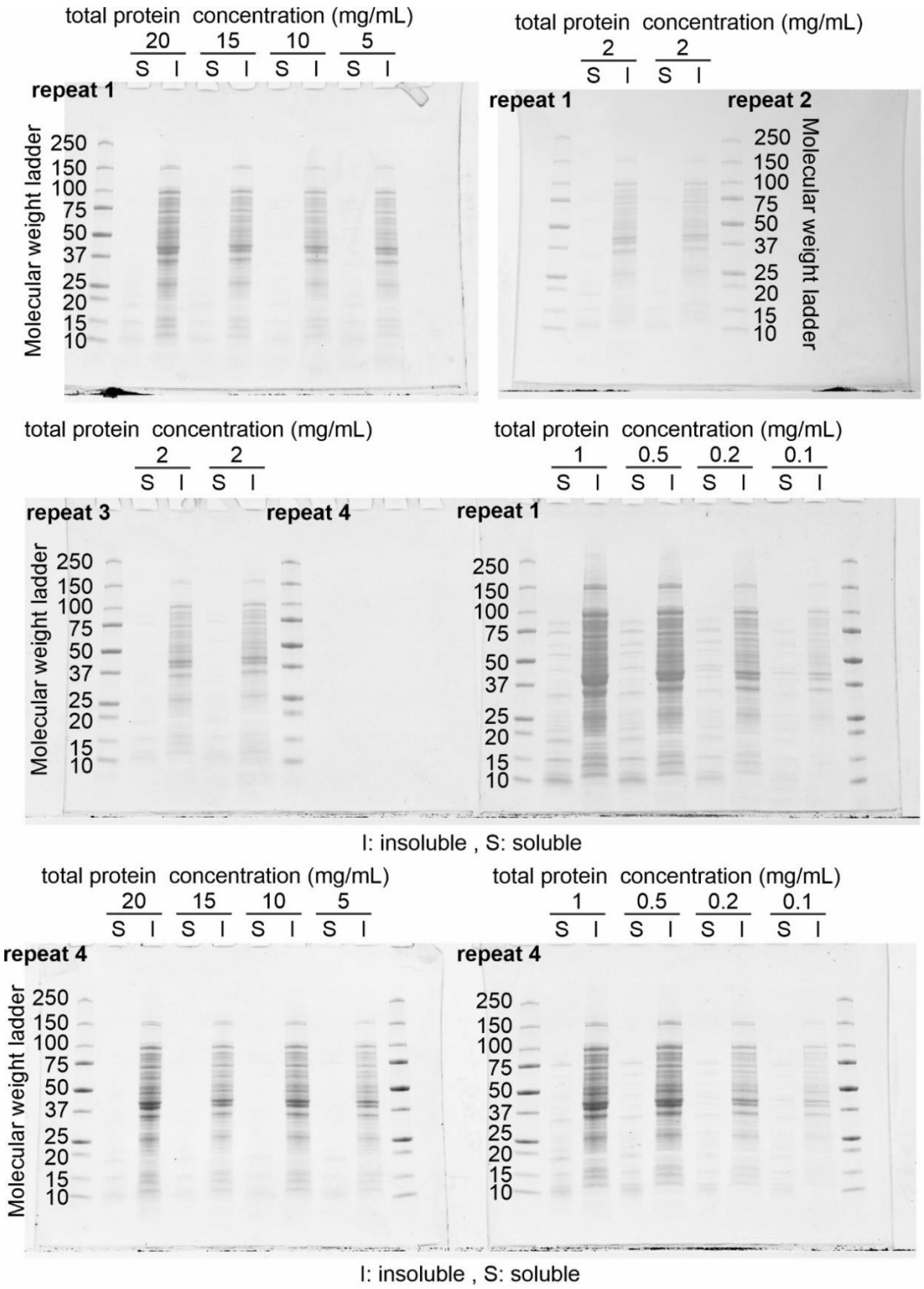
Supplementary Figure S2. *E. coli* protein collection before and after depletion of ribosomal proteins as intact 70S ribosomes via ultracentrifugation. Large biological complexes were depleted and yielded fewer scattering contributions in our S100 samples. We estimated about 25-30% of total ribosomal proteins were depleted upon ultracentrifugation via BCA assays.



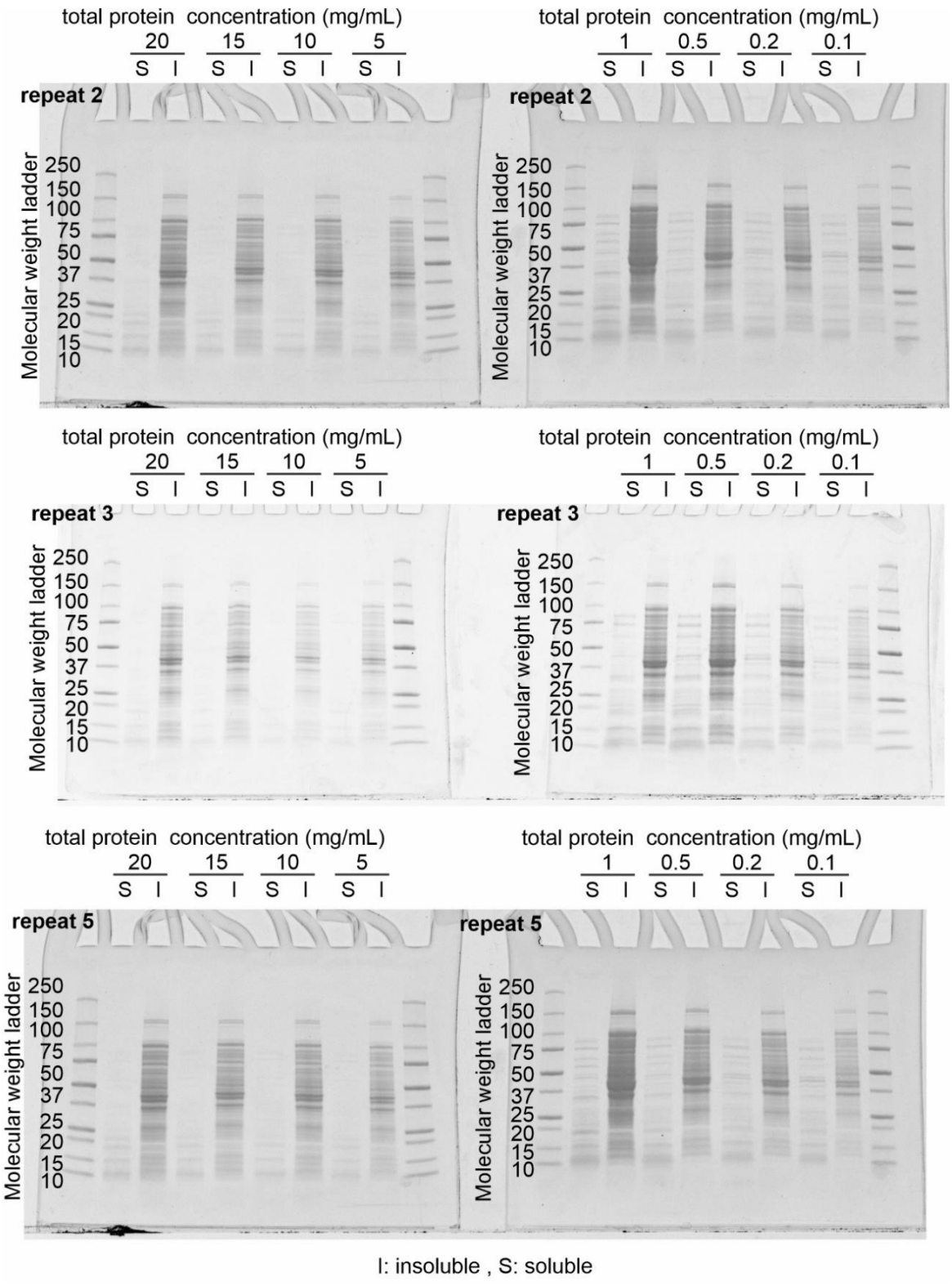
Supplementary Figure S3. Assessment of Autocorrelation Functions (ACFs) of soluble proteomic samples before and after different experimental treatments. Dynamic light scattering measurements were done on soluble samples before and after (A) anaerobic heating-cooling or (B) incubation at 37° C to generate the resulting ACFs. The ACF profiles were then normalized for comparison between samples and determined to be of good quality (i.e., smooth decay of adequate sensitivity, etc.) for downstream regularization or cumulant analyses. Overall, samples show that upon heating-cooling or incubation the soluble samples consisted of smaller species. Interestingly, the soluble fraction generated upon dilution of insoluble aggregates contained large species in both treatments. Total protein concentrations were measured to be 5 mg/mL before treatment.



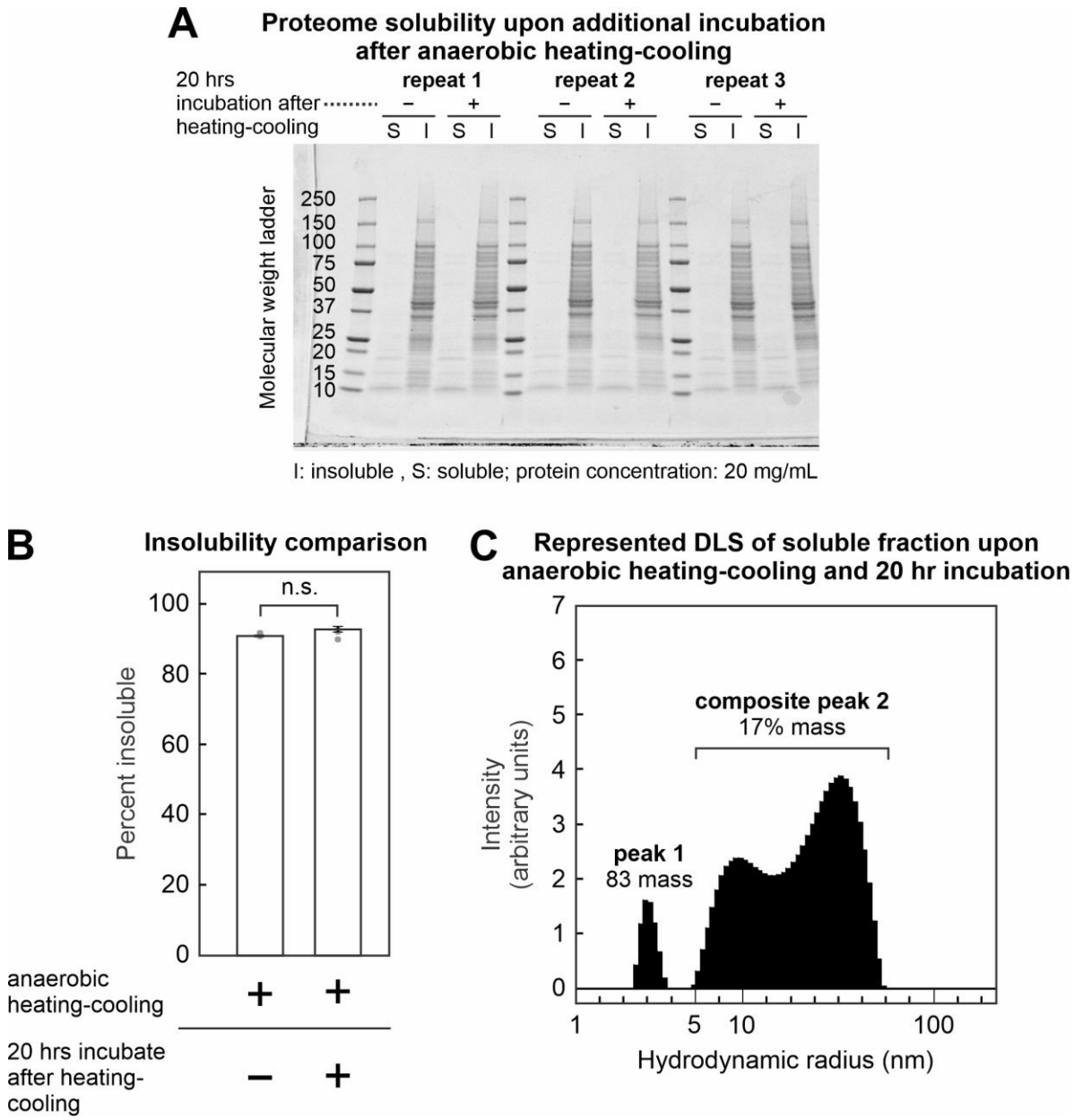
Supplementary Figure S4. Equilibrium Thermodynamics (ET) model shows that additional monomers, past the critical concentration, are incorporated into aggregated phase. (A) The changes in concentration of monomers, [M], in a phase (i.e., monomeric or aggregated) as a function of increasing total protein concentration, where $[M]_0$., $[M]_{crit}$ is the critical concentration; n is the number of monomers; and $[A_n]$ can denote either the concentration of soluble aggregates or insoluble aggregates of size n. (B) Percent of aggregated proteins, $\%_{aggregated}$, as a function of $[M]_0$. See SI Methods 1.6. for more detailed derivations.



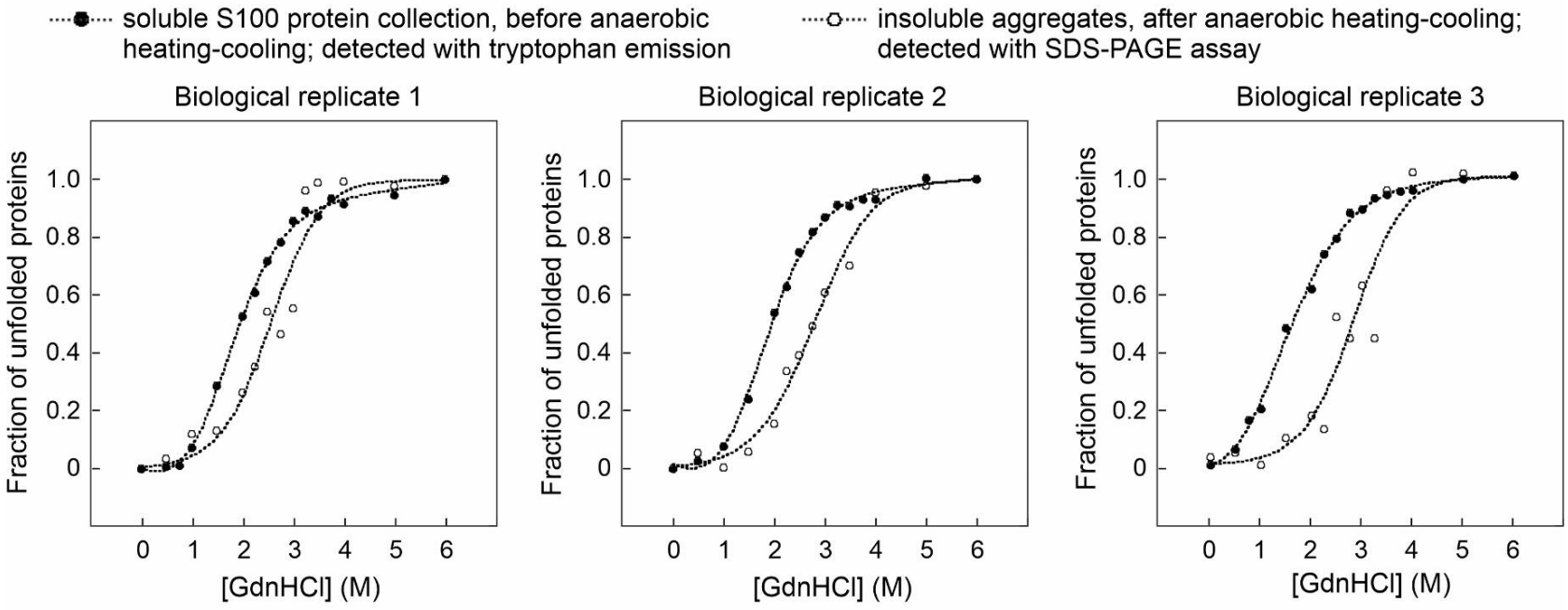
Supplementary Figure S5, part 1. SDS-PAGE analysis of S100 protein samples at different total protein concentration upon anaerobic heating-cooling treatment (See Figure 1 A and B). 5 biological repeats were measured.



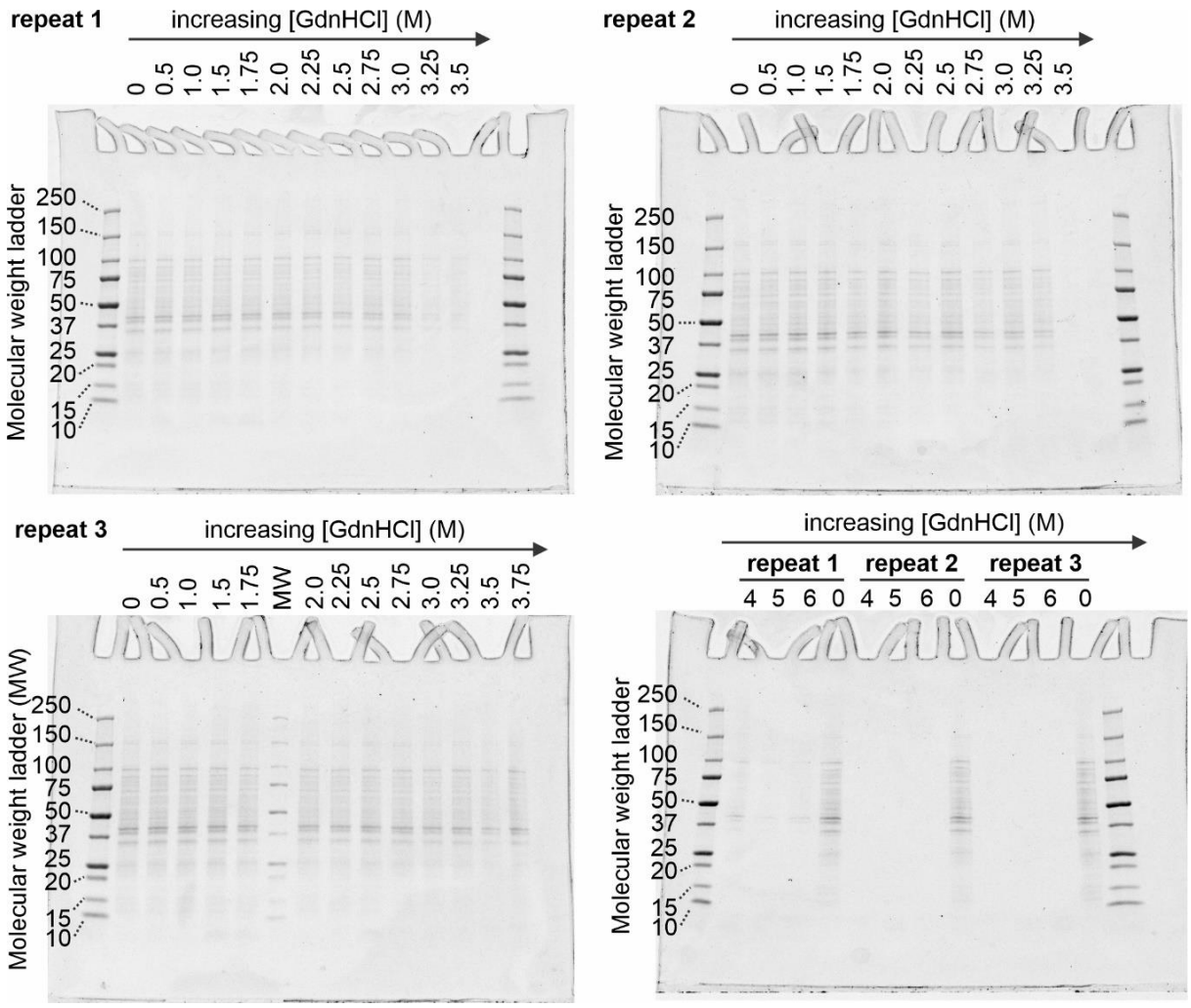
Supplementary Figure S5, part 2. SDS-PAGE analysis of S100 protein samples at different total protein concentration upon anaerobic heating-cooling treatment (See Figure 1 A and B). 5 biological repeats were measured.



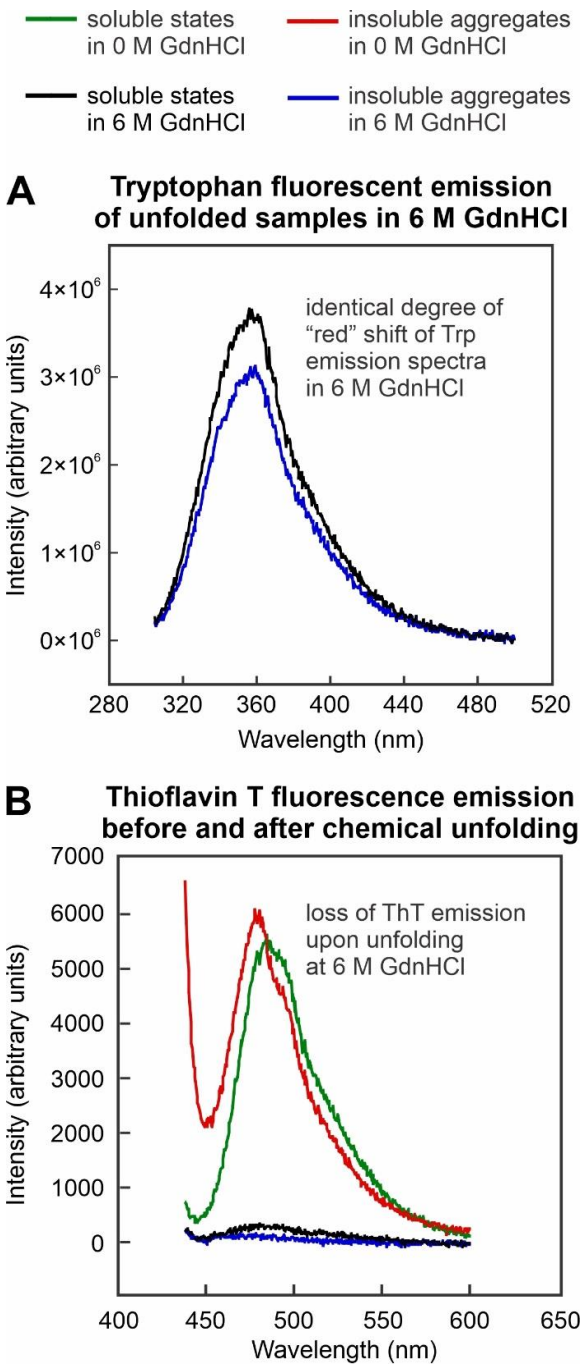
Supplementary Figure S6. Soluble proteins and insoluble aggregates are kinetically trapped relative to each other after anaerobic heating-cooling treatment. (A) SDS-PAGE solubility analysis of S100 proteome sample upon aerobic heating-cooling and additional 20 hr incubation. (B) Densitometry analysis shows the same solubility in the presence and absence of additional post-heating-cooling 20 hr incubation; P-values (Student's t-test) > 0.4, n = 3, n.s.: not significant. (C) DLS analysis shows identical profiles as that of soluble fraction immediately after thermal treatment (See Figure 1, panel D). 3 biological repeats were measured. Total protein concentration was 20 mg/mL. Samples show no change in solubility before and after additional incubation.



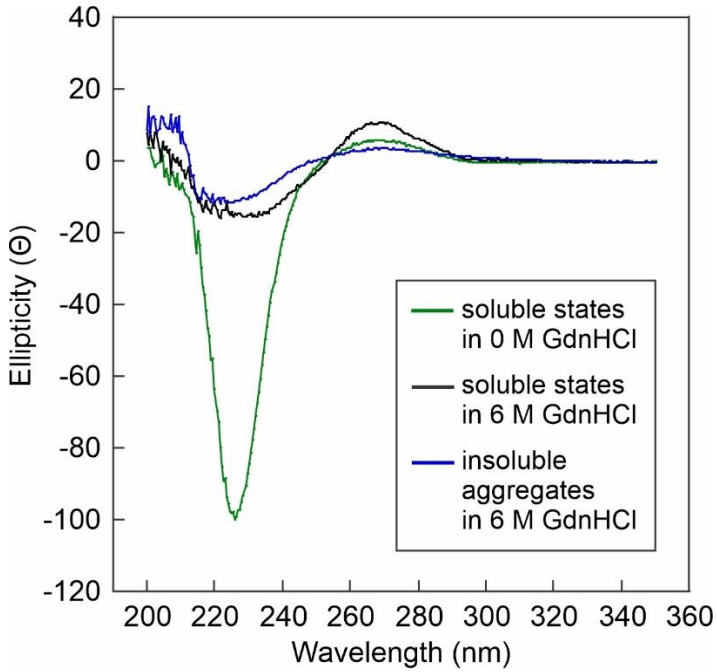
Supplementary Figure S7. Unfolding profiles of *E coli* proteome soluble S100 protein collection and insoluble aggregates before and after anaerobic heating-cooling treatment respectively. 3 biological repeats were done to account for the potential protein composition variation from S100 proteome preparation. The unfolding profiles of the soluble S100 protein collection (total protein concentration: 5 mg/mL) was fitted with the two-state model according to Santoro-Bolen. The unfolding profiles of insoluble aggregates (total protein concentration: 5 mg/mL) were fitted with the isodesmic linear polymerization model. Detailed data collection and model fitting can be found in SI Methods 1.11 and 1.12.



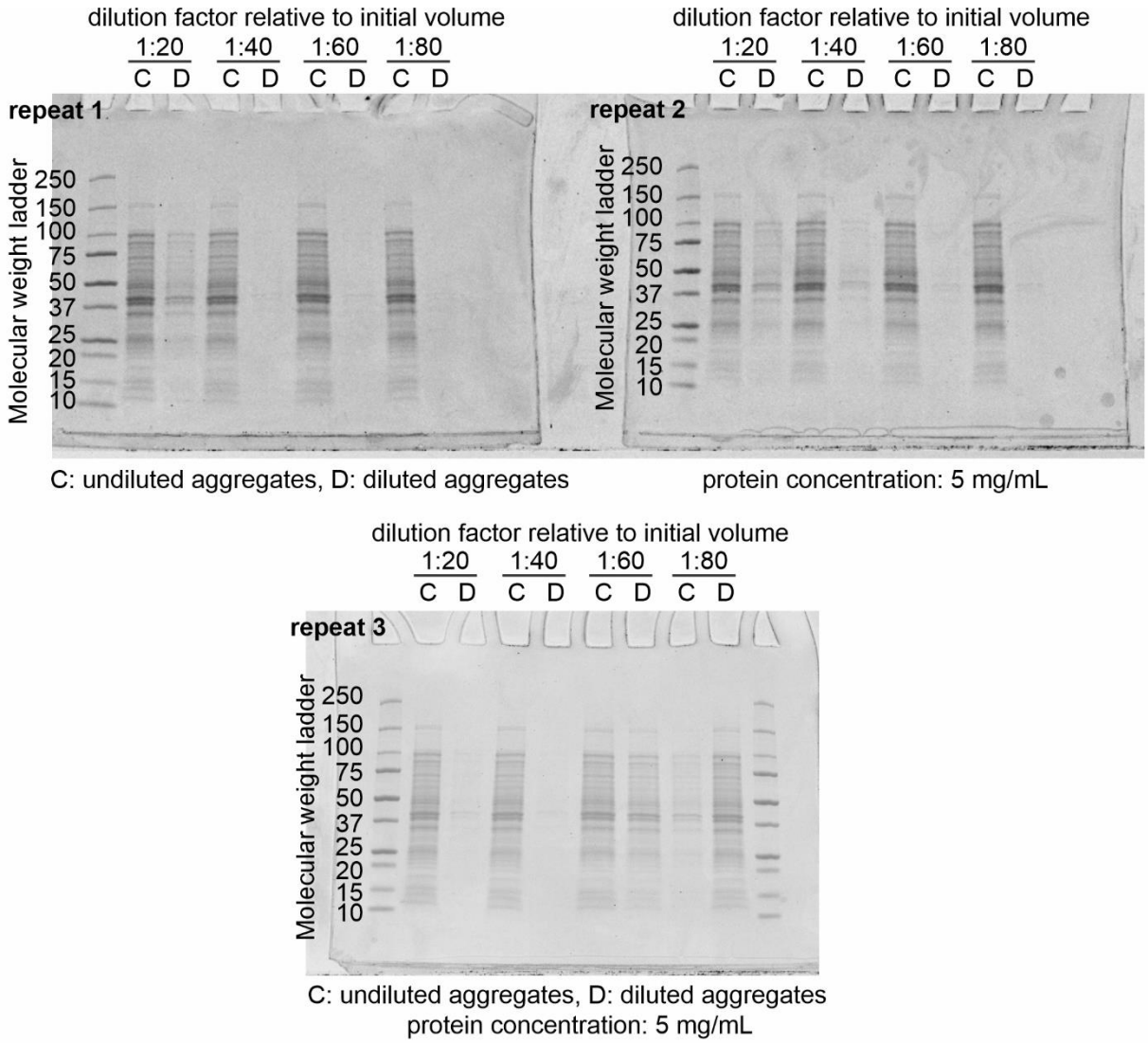
Supplementary Figure S8. SDS-PAGE analysis of insoluble aggregates, generated after anaerobic heating-cooling, in the presence of 0-6M GdnHCl. 3 biological repeats were measured. Total protein concentration was determined to be 5 mg/mL by BCA assay. Samples show no change in solubility before and after additional incubation. Detailed methods can be found in SI Methods 1.12.



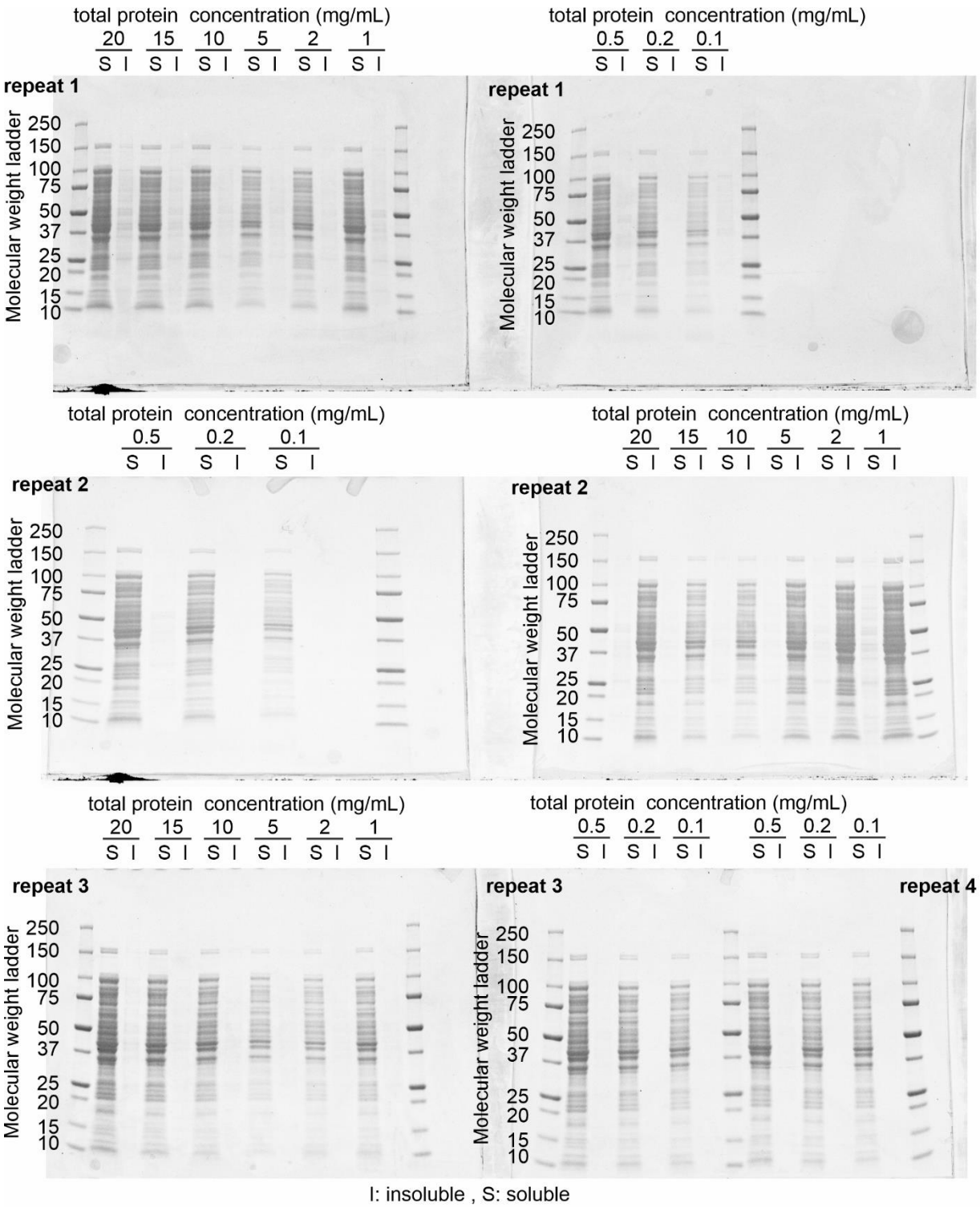
Supplementary Figure S9. Unfolding profiles of soluble S100 protein collection and insoluble aggregates from before and after anaerobic heating-cooling respectively are similar at 6 M GdnHCl. (A) Tryptophan fluorescent emission shows that the unfolded states from both samples are identical in spectral center of mass shift upon denaturation in 6 M GdnHCl (See SI Methods 1.11) . (B) Loss of Thioflavin T fluorescent emission profiles represents the unfolding of β-sheet rich structures from both samples upon denaturation in 6 M (See SI Methods 1.13). Total protein concentration was 5 mg/mL by BCA assay.



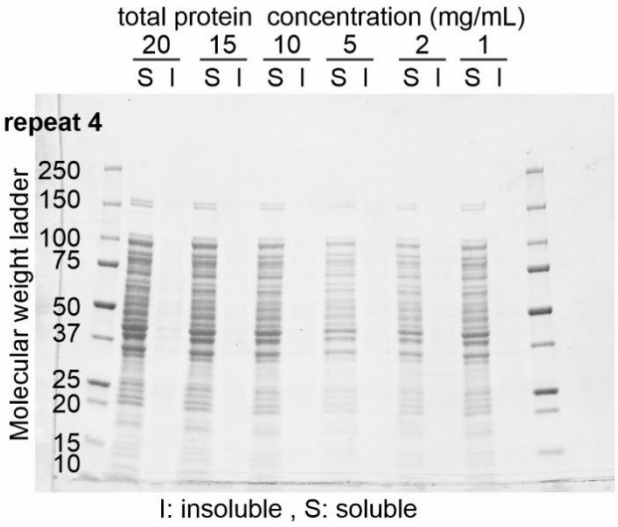
Supplementary Figure S10. Circular dichroism shows similar degree of unfolding of soluble S100 protein collection and insoluble aggregates from before and after anaerobic heating-cooling respectively. However, residual of secondary structure profiles at 215-240 nm was observed, which is reminiscent of β-sheet structures. In addition, CD profile at 250-300 nm suggests the presence of DNA and tRNA within our S100 proteome samples. No further interpretation was done on this CD data due to the complexity of our samples. Detailed methods are in SI Methods 1.14.



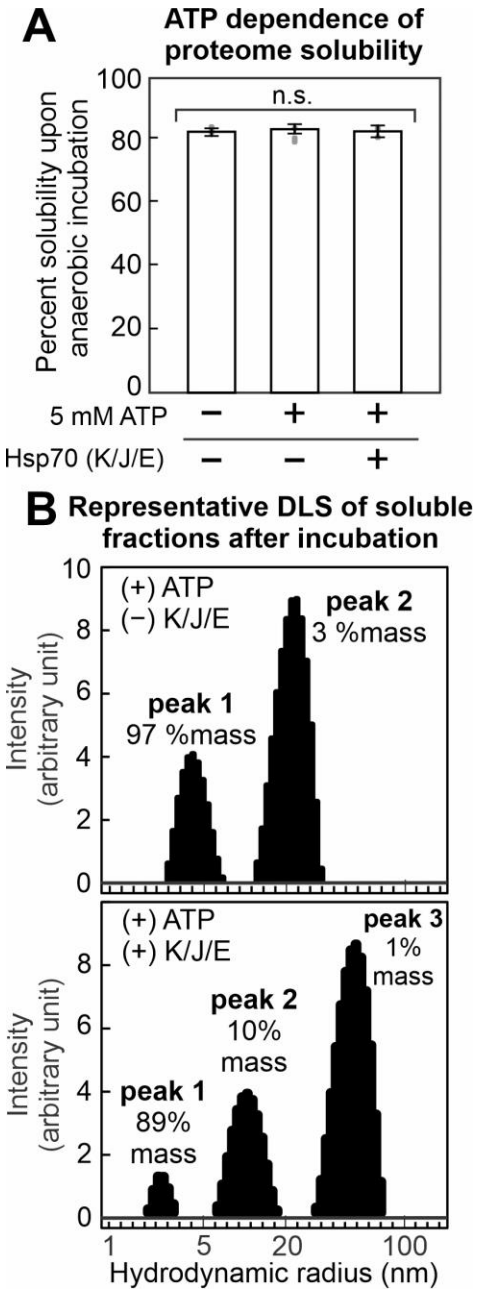
Supplementary Figure S11. Insoluble aggregates from anaerobic heating-cooling treatment are under reversible exchange with soluble states under physiologically relevant conditions. SDS-PAGE analysis of insoluble aggregates, generated from anaerobic heating-cooling experiments, upon progressive dilution experiments (See Figure 2, panel B and C, and SI Methods 1.15). 3 biological repeats were measured. The total protein concentration for anaerobic heating-cooling experiment was 5 mg/mL.



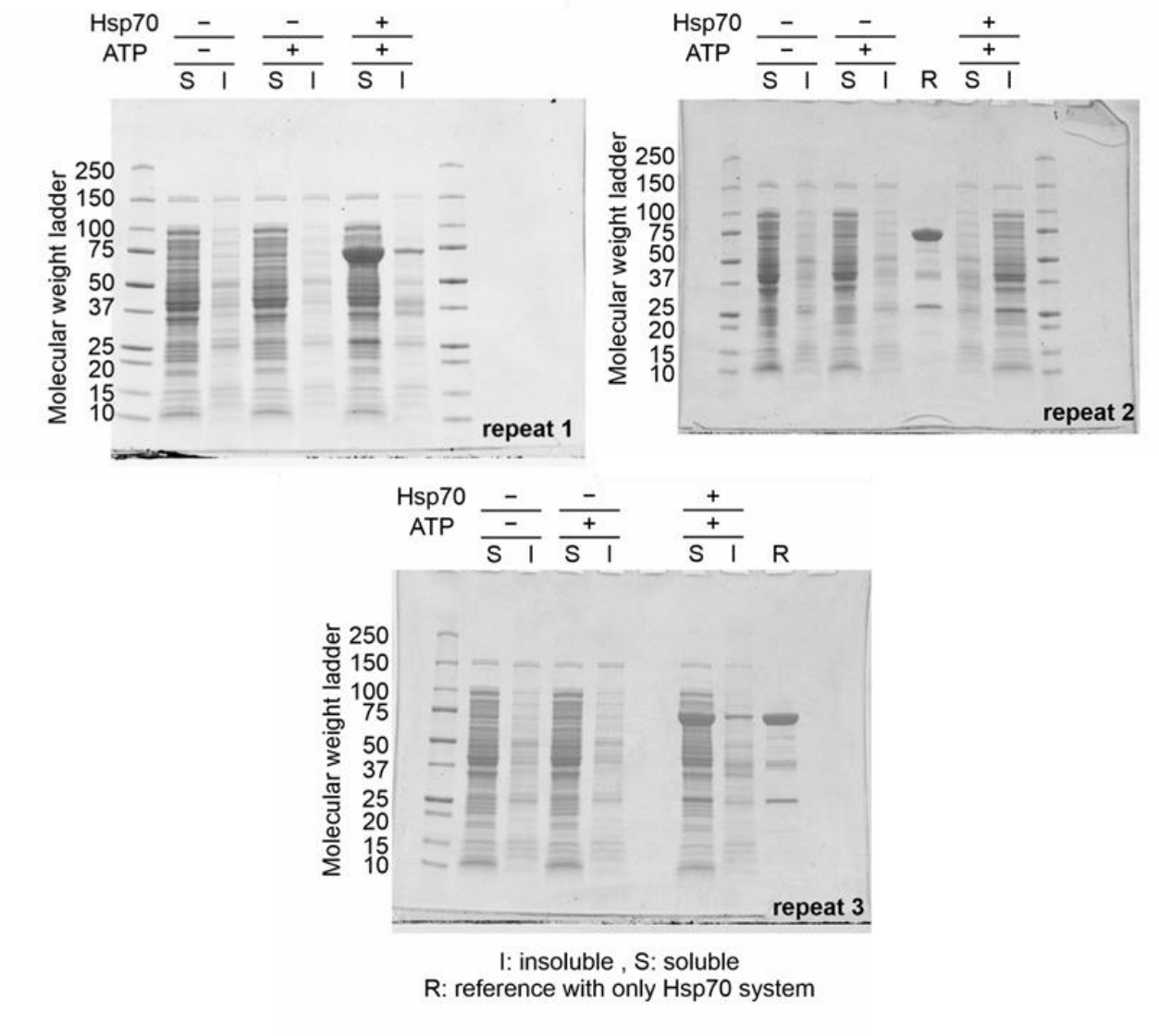
Supplementary Figure S12, part 1. Soluble states are kinetically protected from insoluble aggregate upon anaerobic incubation. SDS-PAGE analysis of S100 protein samples at different total protein concentration upon anaerobic incubation at 37 °C (See Figure 3 A and B). 4 biological repeats were measured.



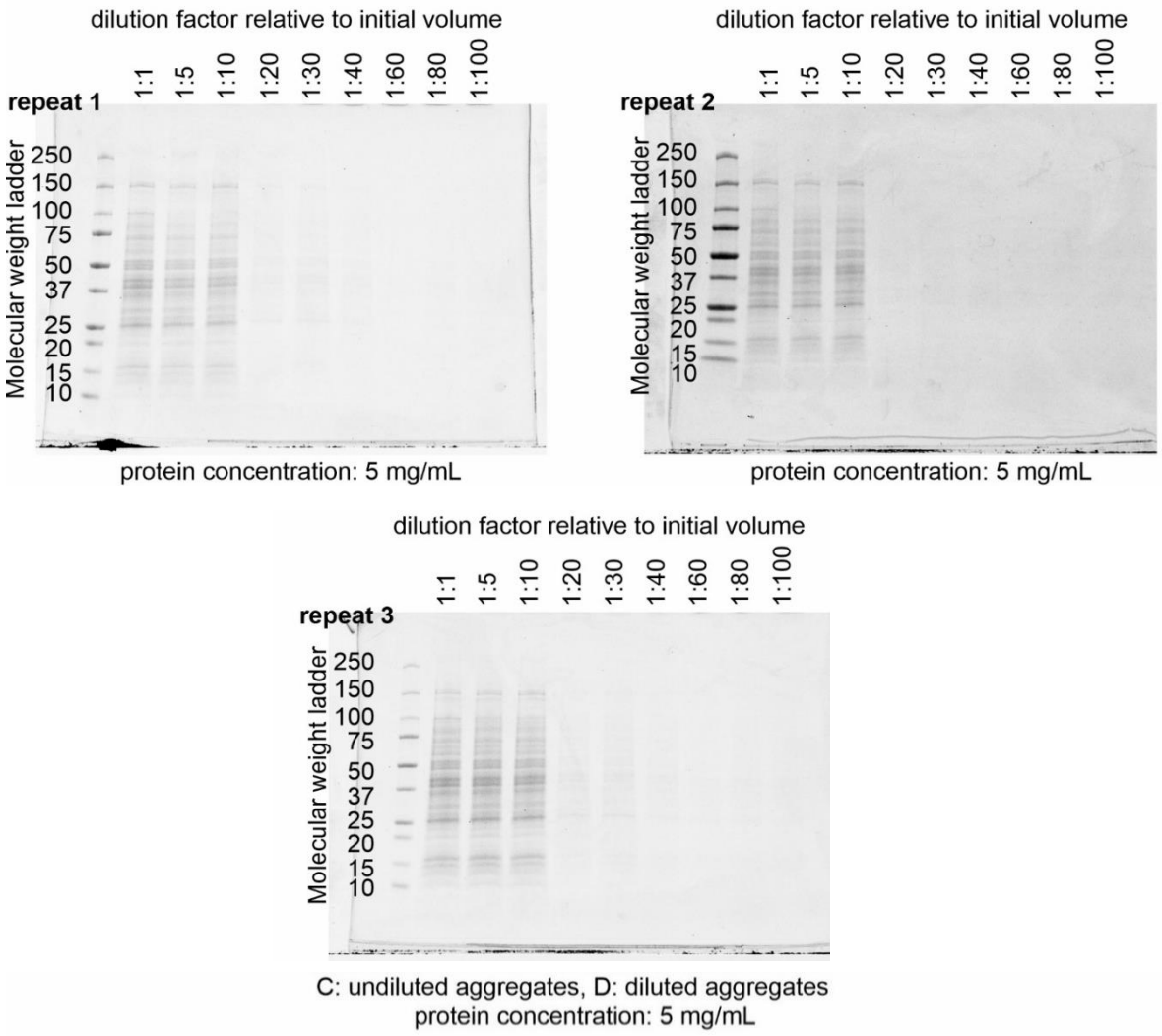
Supplementary Figure S12, part 2. Soluble states are kinetically protected from insoluble aggregate upon anaerobic incubation. SDS-PAGE analysis of S100 protein samples at different total protein concentration upon anaerobic incubation at 37 °C (See Figure 3 A and B). 4 biological repeats were measured.



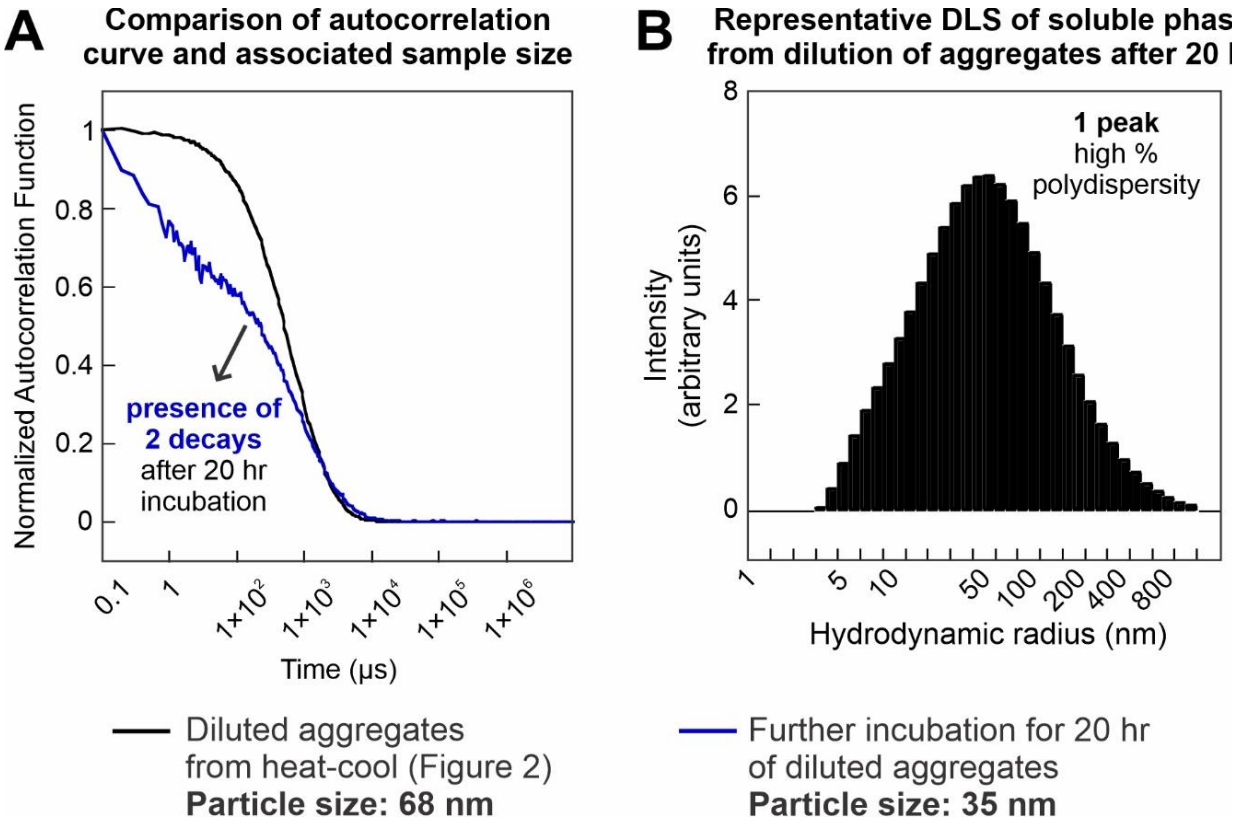
Supplementary Figure S13. **Hsp70 chaperone system does not play a role in the kinetic control solubility of E. coli proteins** (A) SDS-PAGE solubility assay of S100 sample upon incubation at 37°C in the absence and presence of 5 mM ATP and 54 μM DnaK in the Hsp70 system shows statistically similar percent solubilities; P-values (Student's t-test) > 0.7, n = 3, n.s.: not significant. The concentrations of co-chaperones of DnaK (K) are in ratio of 1:2.5:5 for K/J/E. (B) DLS analysis reveals the soluble fractions after incubation in the presence of ATP and Hsp70 system consists of non-aggregated species (total protein concentration before incubation: 5 mg/mL).



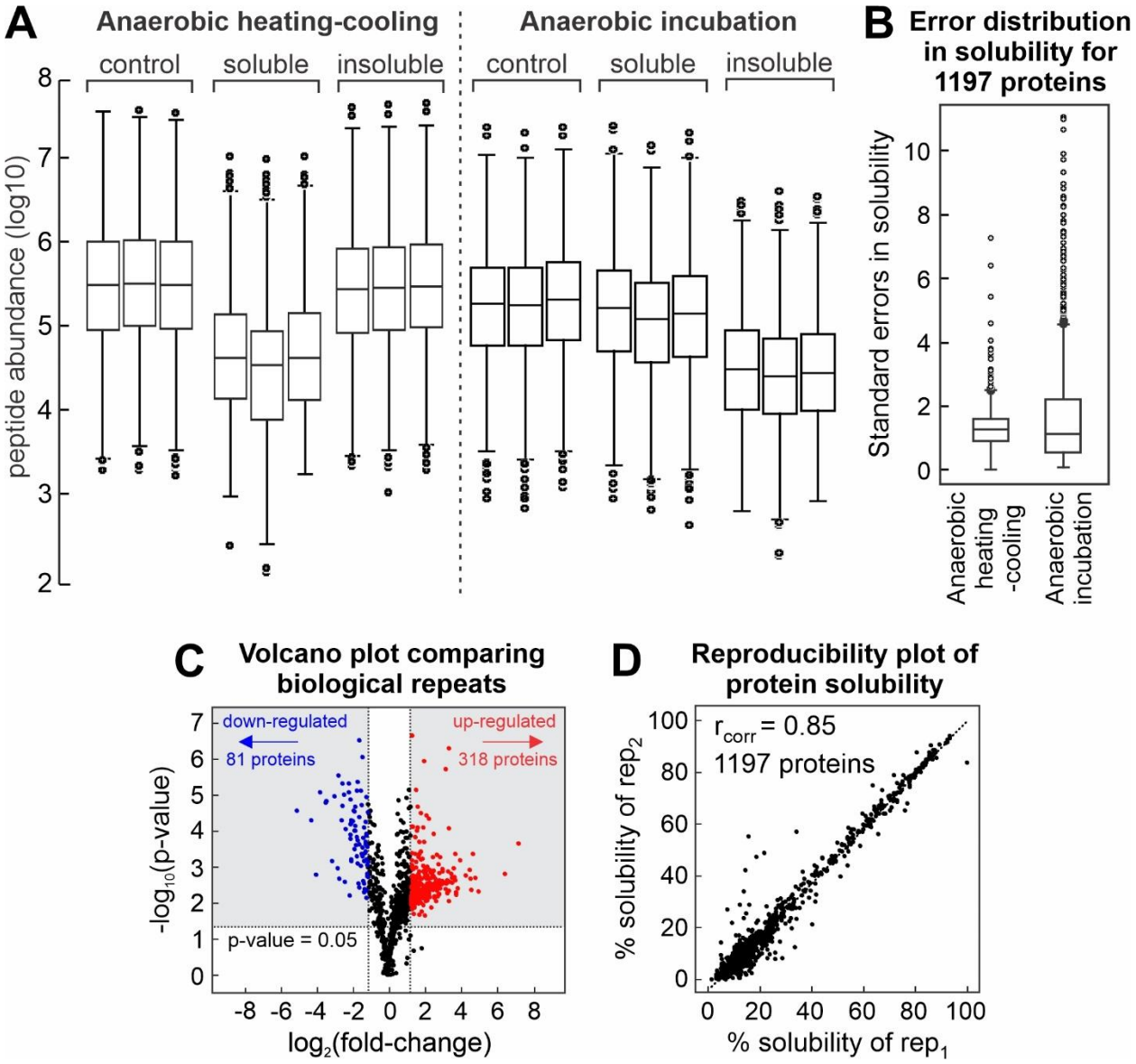
Supplementary Figure S14. Kinetic protection is not dependent on ATP-dependent systems, including the chaperone system Hsp70. SDS-PAGE analysis of S100 protein samples upon anaerobic incubation at 37 °C in the absence and presence of 5 mM ATP and/or Hsp70 in concentration ratio 1 : 2.5: 5 with concentration of DnaK as 54 μM (See Figure 4, panel A). 3 biological repeats were measured. The total protein concentration was 5 mg/mL.



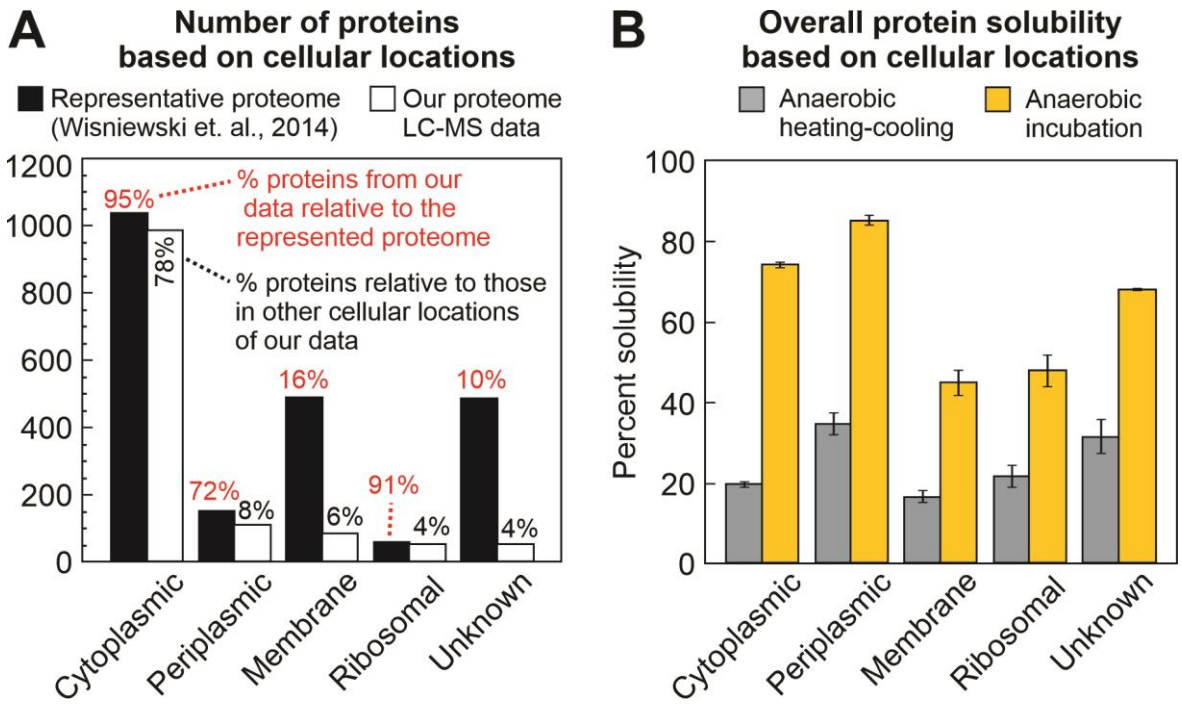
Supplementary Figure S15. Insoluble aggregates from anaerobic heating-cooling treatment are under reversible exchange with soluble states under physiologically relevant conditions. SDS-PAGE analysis of insoluble aggregates, generated from anaerobic incubation experiments, upon progressive dilution experiments (See Figure 5, panel A). 3 biological repeats were measured. The total protein concentration for anaerobic heating-cooling experiment was 5 mg/mL.



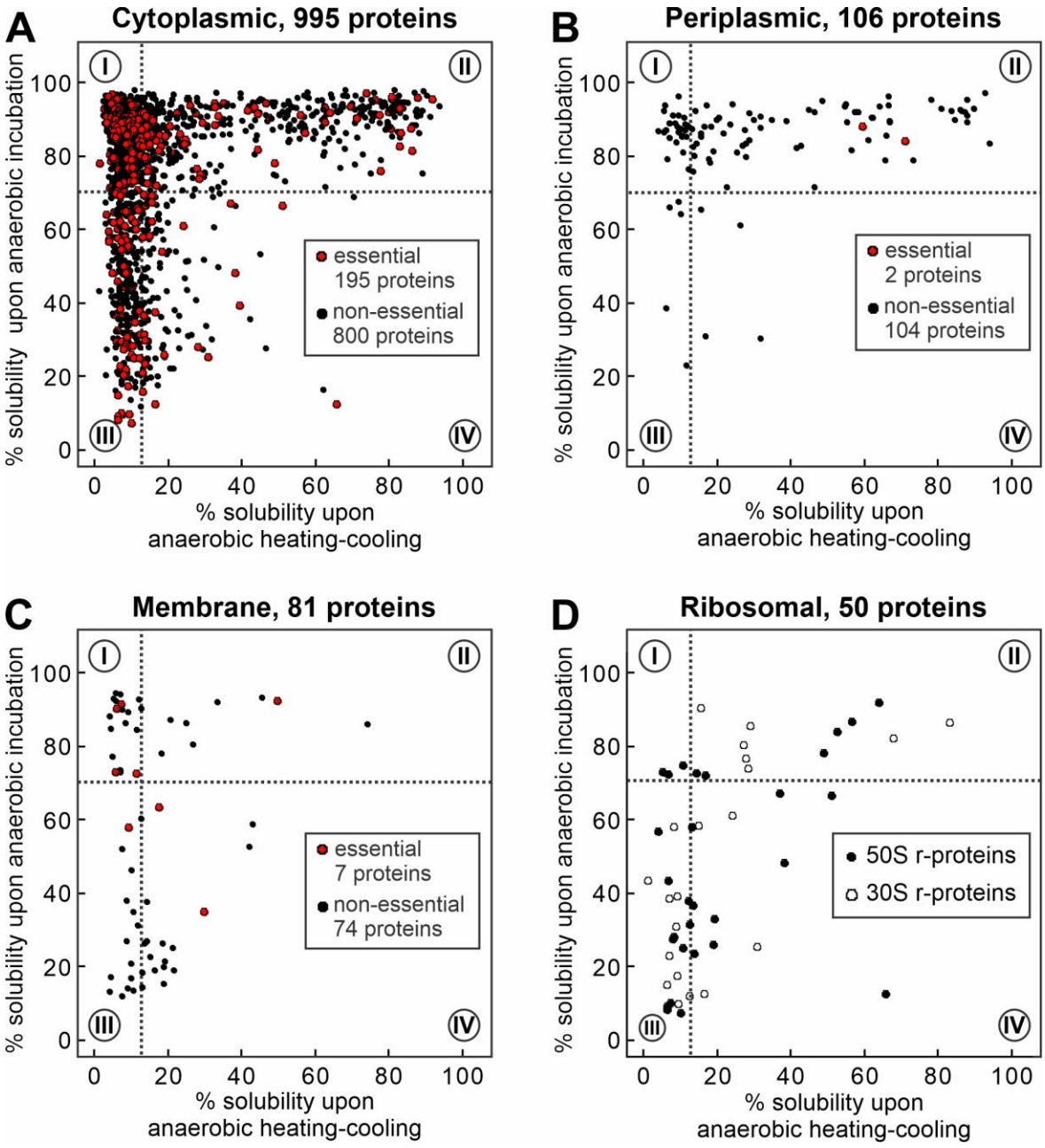
Supplementary Figure S16. Upon dilution and additional incubation of 20 hrs, soluble states, resulted from the reversible exchange as insoluble aggregates, may also contains smaller states besides large soluble aggregates. Insoluble aggregates, generated from anaerobic heating-cooling experiments (See Figure 5, panel A), was subjected to progressive dilution and additional incubation for 20 hrs at 22.5°C. Due to the high presence of large species from the ACF profile and the regularization results, we concluded that, even upon further incubation, the soluble aggregates formed from solubilization of insoluble aggregates. Soluble states generated from 1:20 dilution of insoluble aggregates was analyzed (See SI Methods


A
Anaerobic heating-cooling
Anaerobic incubation
control
soluble
insoluble
control
soluble
insoluble
peptide abundance (log10)
B
Error distribution in solubility for 1197 proteins
Standard errors in solubility
Anaerobic heating -cooling
Anaerobic incubation
C
Volcano plot comparing biological repeats
down-regulated
81 proteins
up-regulated
318 proteins
p-value = 0.05
-log10(p-value)
log2(fold-change)
D
Reproducibility plot of protein solubility
rcorr = 0.85
1197 proteins
% solubility of rep2
% solubility of rep1

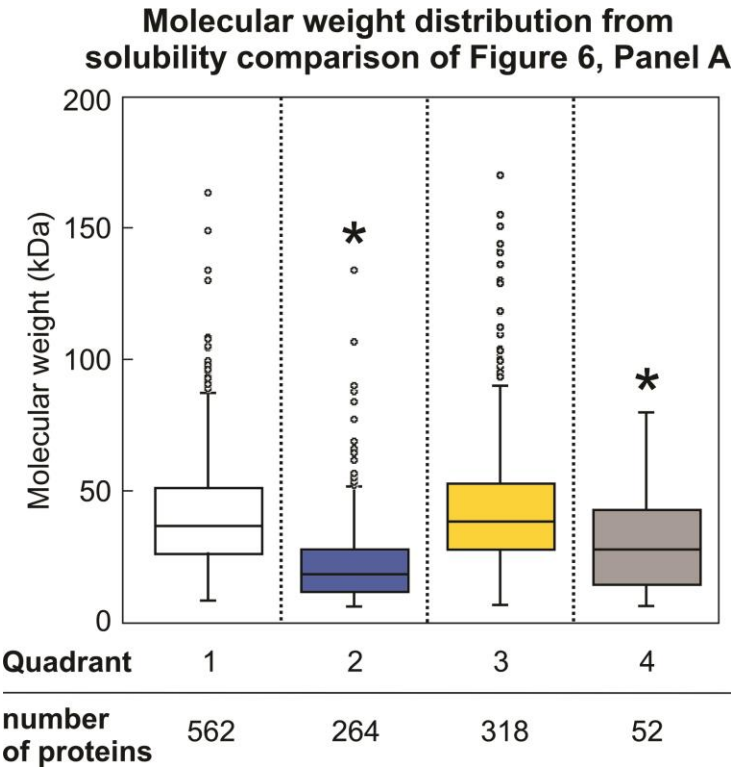
Supplementary Figure S17. Reproducibility of protein abundance and solubility between different samples (control, soluble, and insoluble) across 3 biological repeats from (A) anaerobic heating-cooling experiments and anaerobic incubation experiments. Samples were analyzed with nano-LC MS/MS using tandem mass tagging approach. The analysis was performed using Proteome Discoverer Software. The total protein concentration was 3 mg/mL measured with BCA assay (i.e., a control or b. both soluble and insoluble fractions). 2 internal standards, bovine serum albumin and β-galactosidase, for 1) digestion and TMT-labeling, and 2) LC-MS/MS analysis respectively, were used to normalized peptide abundance prior to downstream bioinformatics. (B) Standard error distribution of protein solubility for each protein across biological repeats from different experimental treatments in panel A. Most proteins have standard error values below 5%. (C) Representative volcano plot shows that 798 of 1197 proteins (c.a. 66%) have similar abundances across 2 biological repeats. Specifically, 318 (red, c.a. 27%) and 81 (blue, c.a. 7%) proteins are expressed at higher or lower abundances respectively across 2 biological replicates. Each biological repeat contains 3 technical repeats. P-value cutoff is at 0.05 confidence interval (Student's t-test); Log_2 of the fold-change cutoff is at ± 1.20. (D) Representative reproducibility plot shows consistent solubility measured across 2 biological repeats for each protein; r_{corr} = 0.85 (Pearson's correlation test).



Supplementary Figure S18. The S100 protein collection consists mainly of cytoplasmic & periplasmic proteins and depletes of membrane proteins. (A) Comparison between 1197 proteins in our data set to a representative *E. coli* proteomic collection of 2,237 proteins identified by LC-MS in Wisniewski et. al. (2014), categorized in terms of cellular locations. Our S100 protein collection is representing a majority of cytoplasmic and periplasmic proteins, under-representing membrane and uncharacterized proteins, and over-representing ribosomal proteins upon comparison to Wisniewski et. al (see percentage in red). Within the proteins in our sample across different cellular location categories, most of our proteome sample consist of cytoplasmic proteins (see percentage in black) with low representation of membrane, periplasmic, and uncharacterized proteins. (B) Solubility comparison on a proteome-wide scales based on cellular locations suggests that membrane and ribosomal proteins have low kinetic stability comparing to that of proteins from other cellular location entries in our *in vitro* conditions; p-value (Student's t-test) < 0.001. Other cellular location entries have p-value (Student's t-test) > 0.05, not significant. (See Figure 6, panel A).



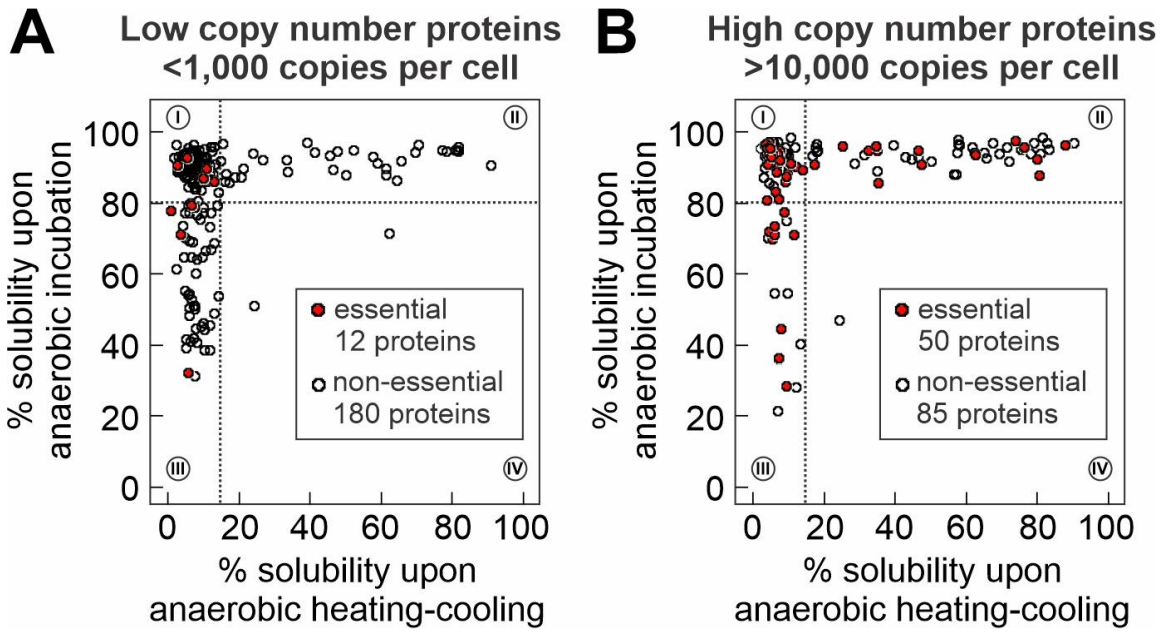
Supplementary Figure S19. Many proteins in the S100 proteome sample, especially cytoplasmic and periplasmic proteins, are kinetically protected from aggregation . Solubility comparison on a proteome-wide scales for (A) cytoplasmic proteins, (B) periplasmic, (C) integral and peripheral membrane proteins, and (D) ribosomal proteins (See Figure 6, panel A).



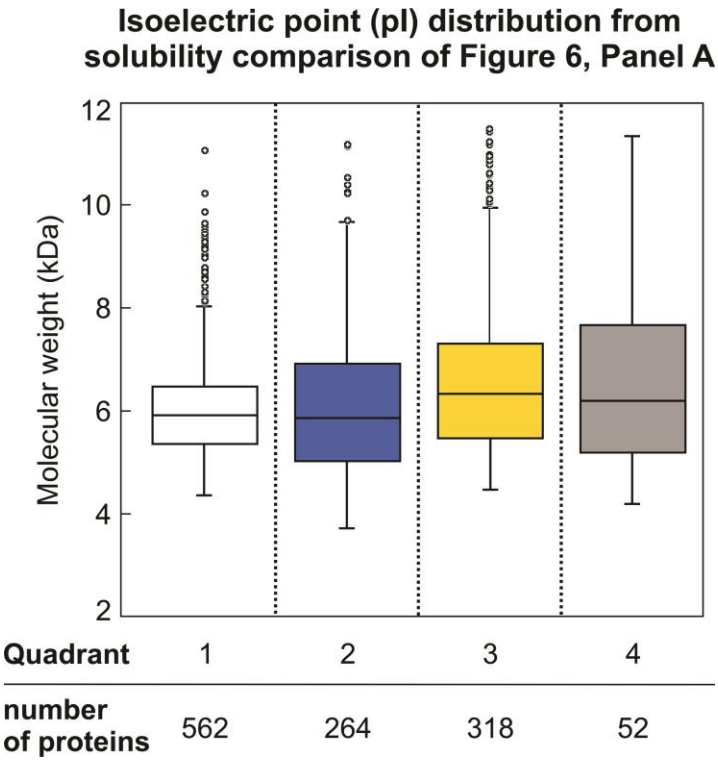
Supplementary Figure S20. Quadrant 2 contains many highly soluble and stable (i.e., relative to aggregates) proteins that tends to be lower in molecular weight. Molecular weight distribution according to quadrants from solubility correlation (See Figure 6, panel A) shows that proteins in quadrants II and IV are smaller in molecular weight than those in other quadrants; P-values (Wilcoxon-Mann-Whitney's test) ≤ 0.05. Correlation analysis shows a weak negative correlation between kinetic stability and propensity to aggregate and molecular weight, $r_{corr} = 0.27$ (Pearson's correlation test).

			NON-POLAR										POLAR					(+) CHARGED			(-) CHARGED	
Quadrant	Type of Solubility	Average % Solubility	Gly G	Ala A	Val V	Cys C	Pro P	Leu L	Ile I	Met M	Trp W	Phe F	Ser S	Thr T	Tyr Y	Asn N	Gln Q	Lys K	Arg R	His H	Asp D	Glu E
I	heating-cooling	10.0	-0.041	0.015	-0.050	0.057	-0.014	0.059	-0.023	-0.032	0.002	-0.120	0.107	-0.015	-0.097	0.007	0.161	-0.061	0.040	-0.002	0.024	-0.046
562 proteins	incubation at 37°c	85.3	0.110	0.054	0.118	-0.126	-0.040	-0.128	0.009	-0.031	0.029	0.019	-0.074	0.163	0.007	0.043	-0.186	0.089	-0.155	-0.106	0.126	-0.005
II	heating-cooling	53.2	-0.030	-0.013	-0.010	-0.005	-0.196	-0.220	-0.005	-0.015	-0.123	0.015	0.024	0.029	-0.087	0.124	0.036	0.290	-0.188	-0.191	0.153	0.152
264 proteins	incubation at 37°c	87.3	0.064	0.024	0.041	0.025	0.002	-0.163	0.015	0.018	-0.047	-0.016	-0.109	0.017	-0.025	0.109	-0.046	0.120	-0.209	-0.117	0.106	0.107
III	heating-cooling	10.7	-0.031	0.122	-0.145	0.076	0.110	-0.002	-0.088	0.041	0.043	-0.030	0.019	-0.002	0.009	0.014	0.143	-0.042	-0.068	0.100	-0.049	-0.049
318 proteins	incubation at 37°c	42.9	-0.097	-0.030	-0.106	0.048	0.181	0.161	-0.090	-0.041	-0.005	-0.060	0.028	-0.021	-0.014	-0.046	0.061	-0.190	-0.055	0.124	0.105	0.190
IV	heating-cooling	33.2	0.037	0.062	0.028	0.140	**-0.352**	0.131	-0.218	-0.196	-0.098	-0.243	0.163	0.243	**-0.314**	0.110	0.153	0.080	-0.117	-0.266	0.070	0.103
52 proteins	incubation at 37°c	44.1	**-0.328**	0.055	-0.285	0.113	-0.198	0.250	-0.010	-0.020	0.016	-0.094	0.193	-0.050	-0.155	0.011	0.096	-0.011	0.077	0.165	-0.121	0.167

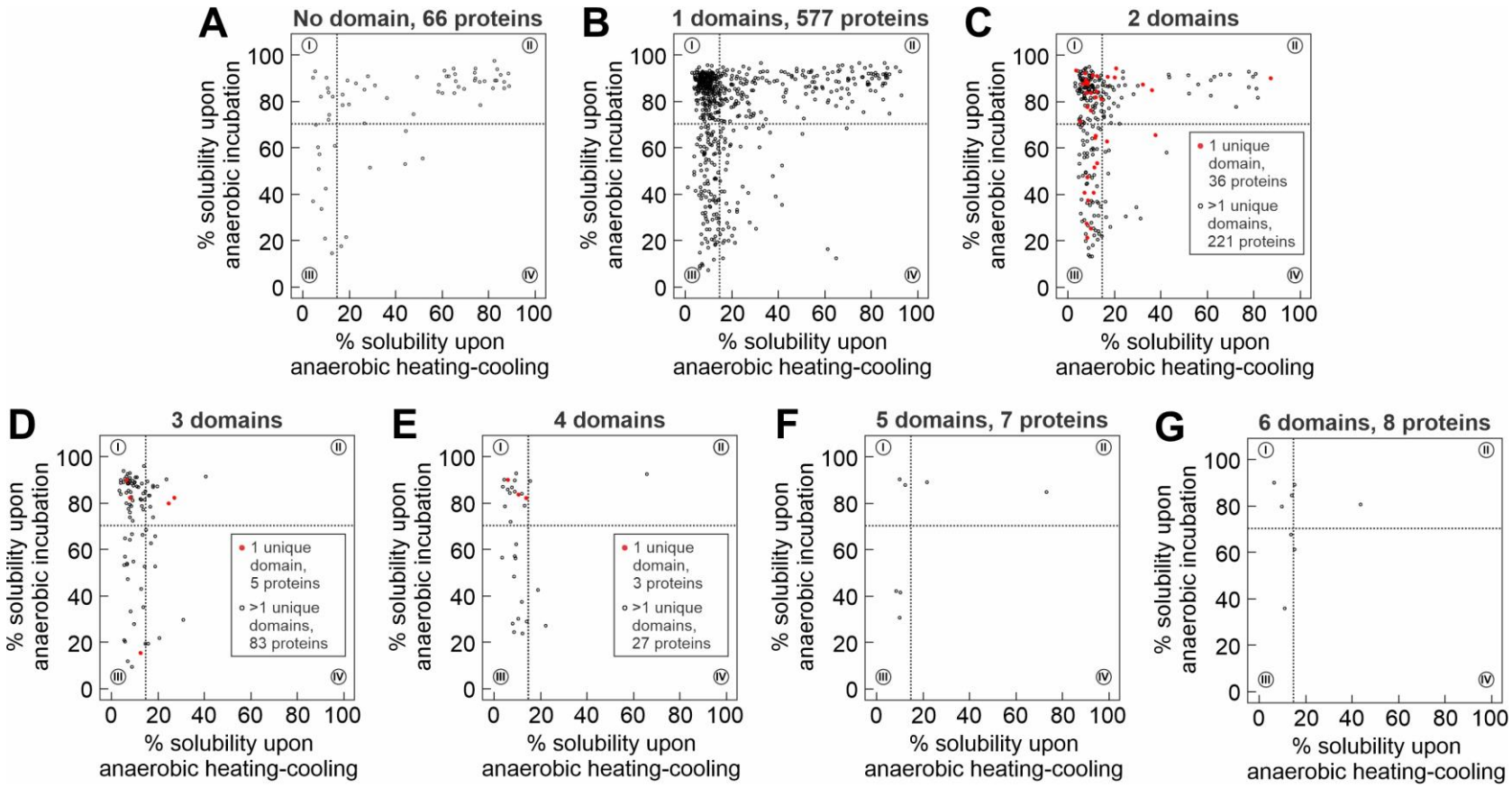
Supplementary Figure S21. Correlation analysis between protein solubility and amino acid contents from different experimental treatments. Pearson's correlation test was used to assess the correlation between the percentage of each canonical amino acid in each protein sequence and experimentally measured solubility of each protein from either a) heating-cooling, or b) incubation at 37°C experiments. r_{corr} values > 0.3 (dark red) were interpreted as weak correlation; r_{corr} values < 0.3 were interpreted as little to no correlation. Correlation analysis was carried out for each quadrant from the solubility correlation (See Figure 6, part A). Average % solubility value across proteins from each experiment was also provided for each quadrant as reference.



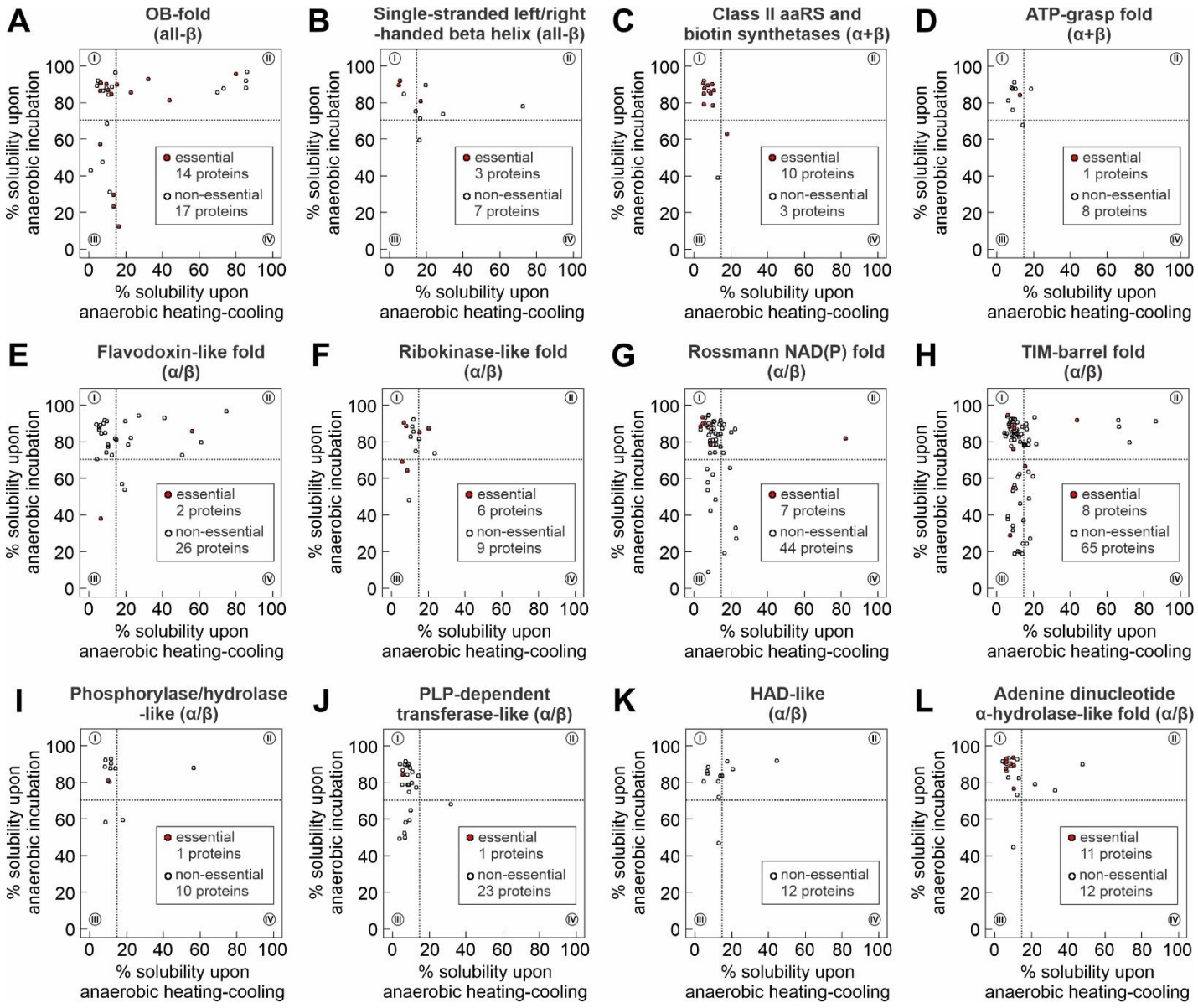
Supplementary Figure S22. Correlation between protein solubility and copy number. Protein solubility was plotted for proteins with (A) low copy number (< 1,000 copies per cell) and (B) high copy number (> 10,000 copies per cell). Proteins that are essential and non-essential, under the characterizations from the KEIO collection, are indicated in red and no-fill dots respectively. Correlation analysis between protein solubility upon anaerobic heating-cooling and copy number shows weak positive correlation for proteins with high copy number (Panel B); r_{corr} = 0.31 (Pearson's correlation test).



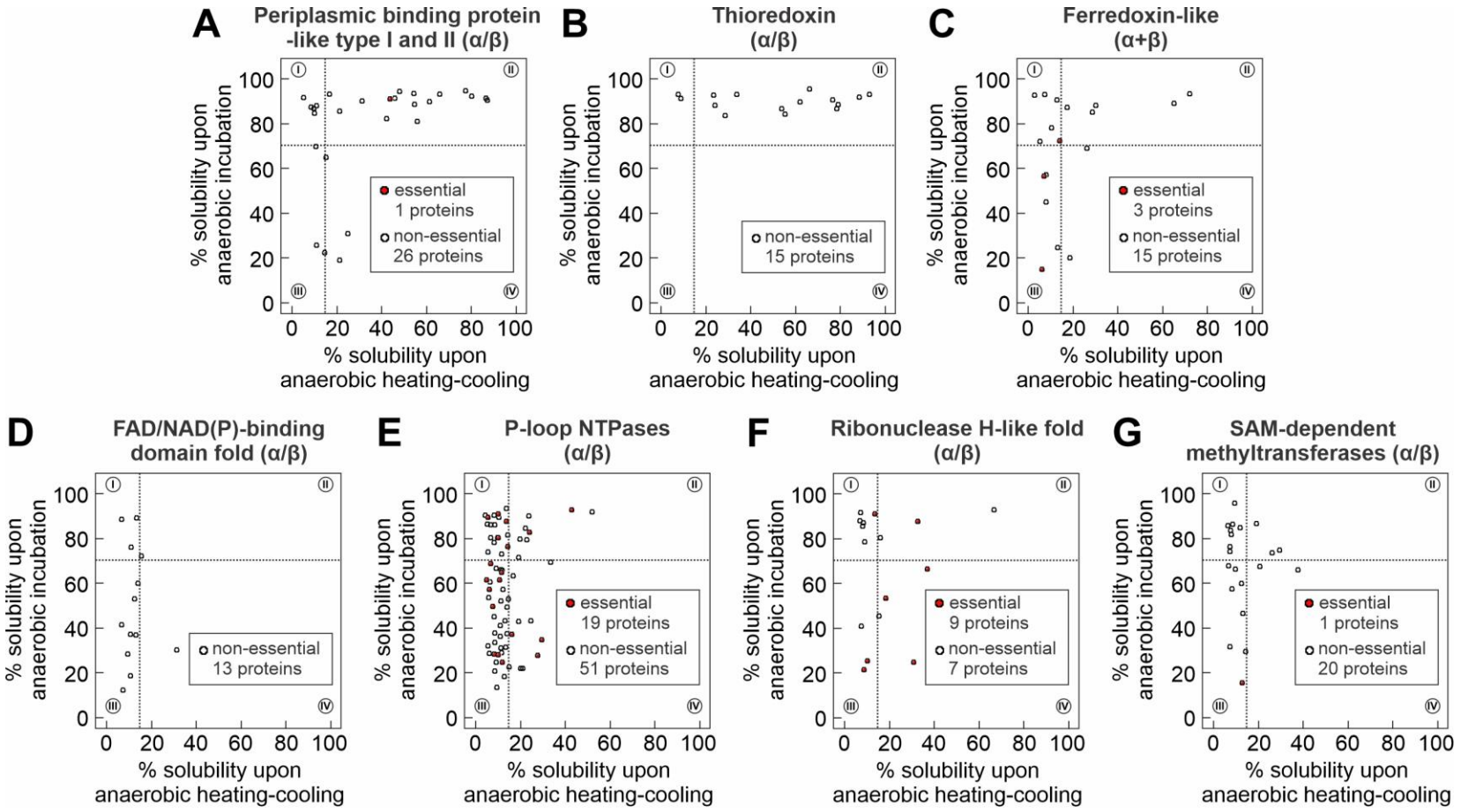
Supplementary Figure S23. Correlation analysis between protein solubility and isoelectric points. There is no correlation between pI values and protein solubilities; 0.1 > r_{corr} > -0.1 (Pearson's correlation test); P-values (Wilconox-Mann-Whitney's test) > 0.1, not significant. pI values were calculated using pI Analysis Tool on the Expasy server.



Supplementary Figure S24. Correlation analysis between protein solubility and number of domain (Gene Ontology database) for 1-7 domains as shown in (A-E)**.** Proteins that are essential and non-essential, under the characterizations from the KEIO collection, are indicated in red and no-fill dots respectively. No direct correlation is observed between number of domains and protein solubility. Interestingly, proteins with high solubility in Quadrant 2 consists of 2 or less domains.

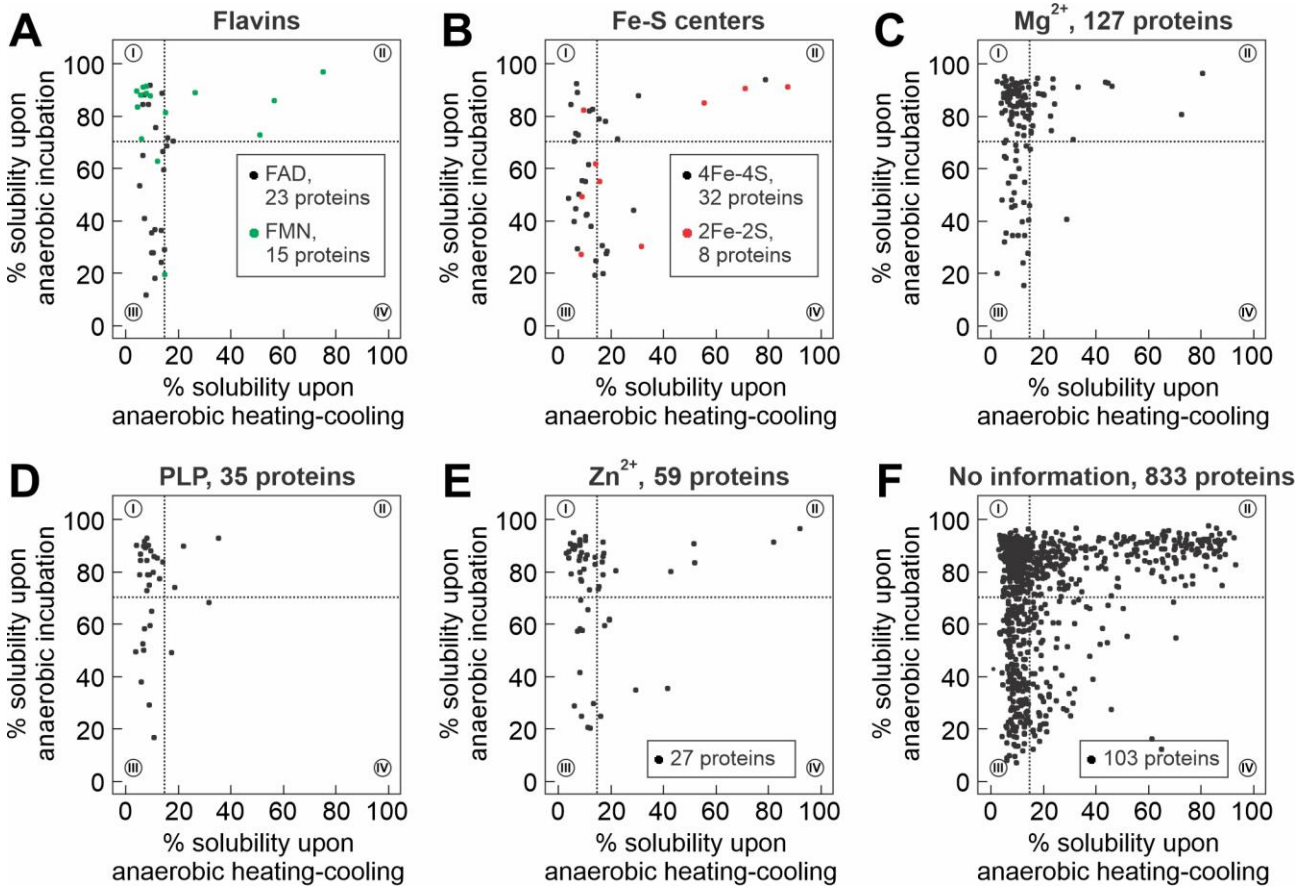

A
OB-fold (all-β)
essential 14 proteins
non-essential 17 proteins
B
Single-stranded left/right -handed beta helix (all-β)
essential 3 proteins
non-essential 7 proteins
C
Class II aaRS and biotin synthetases (α+β)
essential 10 proteins
non-essential 3 proteins
D
ATP-grasp fold (α+β)
essential 1 proteins
non-essential 8 proteins
E
Flavodoxin-like fold (α/β)
essential 2 proteins
non-essential 26 proteins
F
Ribokinase-like fold (α/β)
essential 6 proteins
non-essential 9 proteins
G
Rossmann NAD(P) fold (α/β)
essential 7 proteins
non-essential 44 proteins
H
TIM-barrel fold (α/β)
essential 8 proteins
non-essential 65 proteins
I
Phosphorylase/hydrolase -like (α/β)
essential 1 proteins
non-essential 10 proteins
J
PLP-dependent transferase-like (α/β)
essential 1 proteins
non-essential 23 proteins
K
HAD-like (α/β)
non-essential 12 proteins
L
Adenine dinucleotide α-hydrolase-like fold (α/β)
essential 11 proteins
non-essential 12 proteins
% solubility upon anaerobic incubation
% solubility upon anaerobic heating-cooling
0 20 40 60 80 100
I II III IV

Supplementary Figure S25. Correlation analysis between protein solubility and the domain fold-types (Structure Classification of Protein – SCOP database). Kinetic stability relative to aggregation applies to a diverse collection of protein domain fold-types under all β, α+β, and α/β classes. Only domain fold-types with at least 10 proteins are quantified and reported. All-β fold, (A) OB-fold and (B) Single-stranded left/right-handed beta helx, are kinetically stable. Two α+β folds, (B) ATP-grasp and (C) Class II aaRS and biotin synthetase, are also kinetically stable. Finally, the most represented domain fold class, α/β, contains many fold-types that are kinetically stable: (D) ATP-grasp fold, , (E) Flavodoxin-like, (F) Ribokinase-like, (G) Rossman NAD(), (H) TIM-barrel, (I) Phosphorylase/hydrolase-like, (J) PLP-dependent transferase-like, (K) HAD-like, and (L) Adenine dinucleotide α-hydrolase-like. Generally, essential proteins are kinetically trapped relative to aggregates across different domain folds. We also considered whether a protein has one single domain or multi-domains into our correlation analysis between solubility and domain folds. Overall, we arrived at the same conclusions in either case. Therefore, each correlation plot of a specific domain fold-type contains both single domain proteins and multidomain proteins with at least one domain belonging to the relevant domain fold-type. In addition, proteins that are essential and non-essential, under the characterizations from the KEIO collection, are indicated in red and no-fill dots respectively.

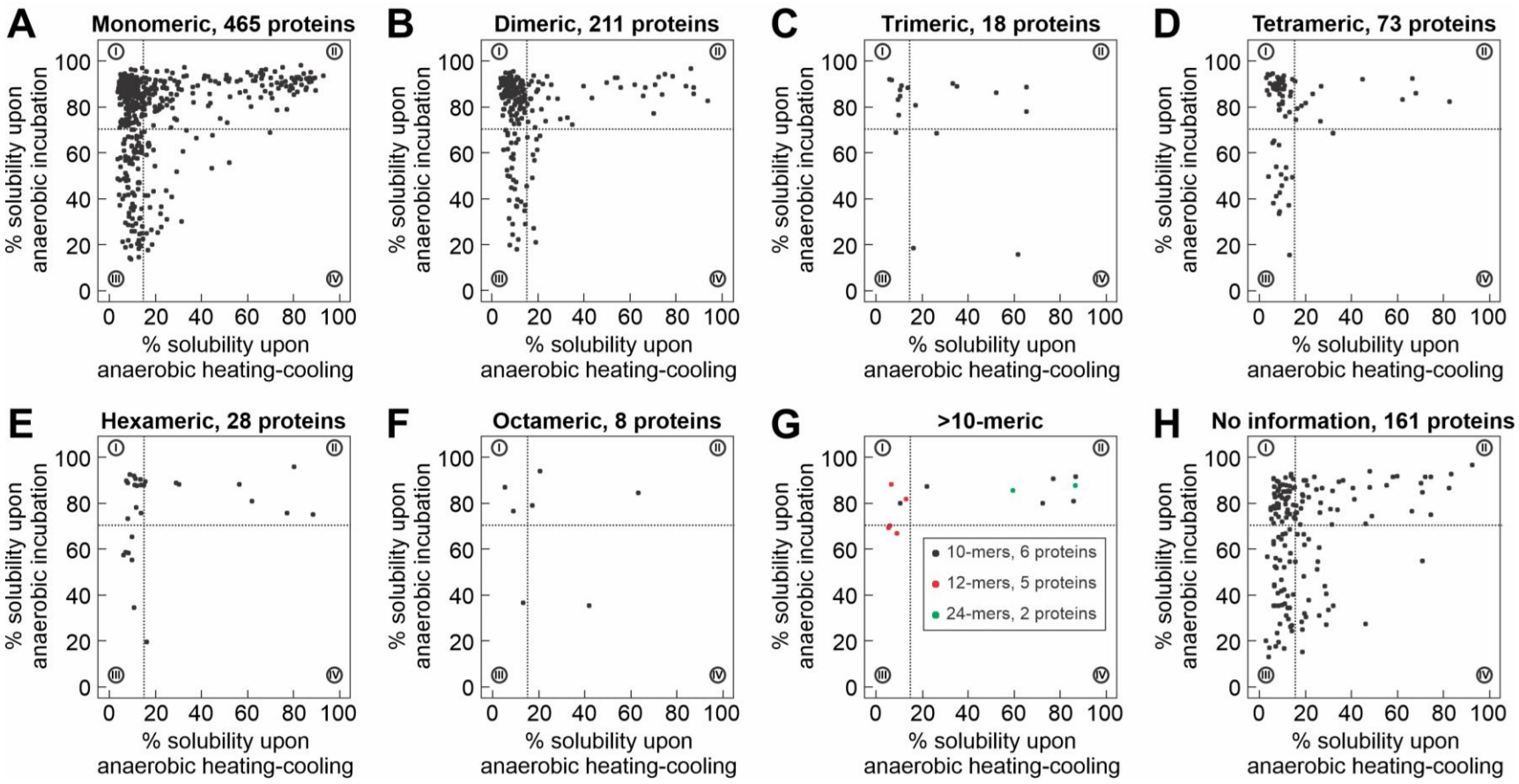


Supplementary Figure S26. Correlation analysis between protein solubility and the domain fold-types of highly soluble proteins in Quadrant II or aggregation-prone proteins in Quadrant III (Structure Classification of Protein – SCOP data base). Only domain fold-types with at least 10 proteins are quantified and reported. Like Figure S24, each correlation plot of a specific domain fold-type contains both single domain proteins and multidomain proteins. Many highly soluble proteins in quadrant II contain domain of fold-types from α/β classes: (A) Periplasmic binding protein-like type I and II, and (B) Thioredoxin.

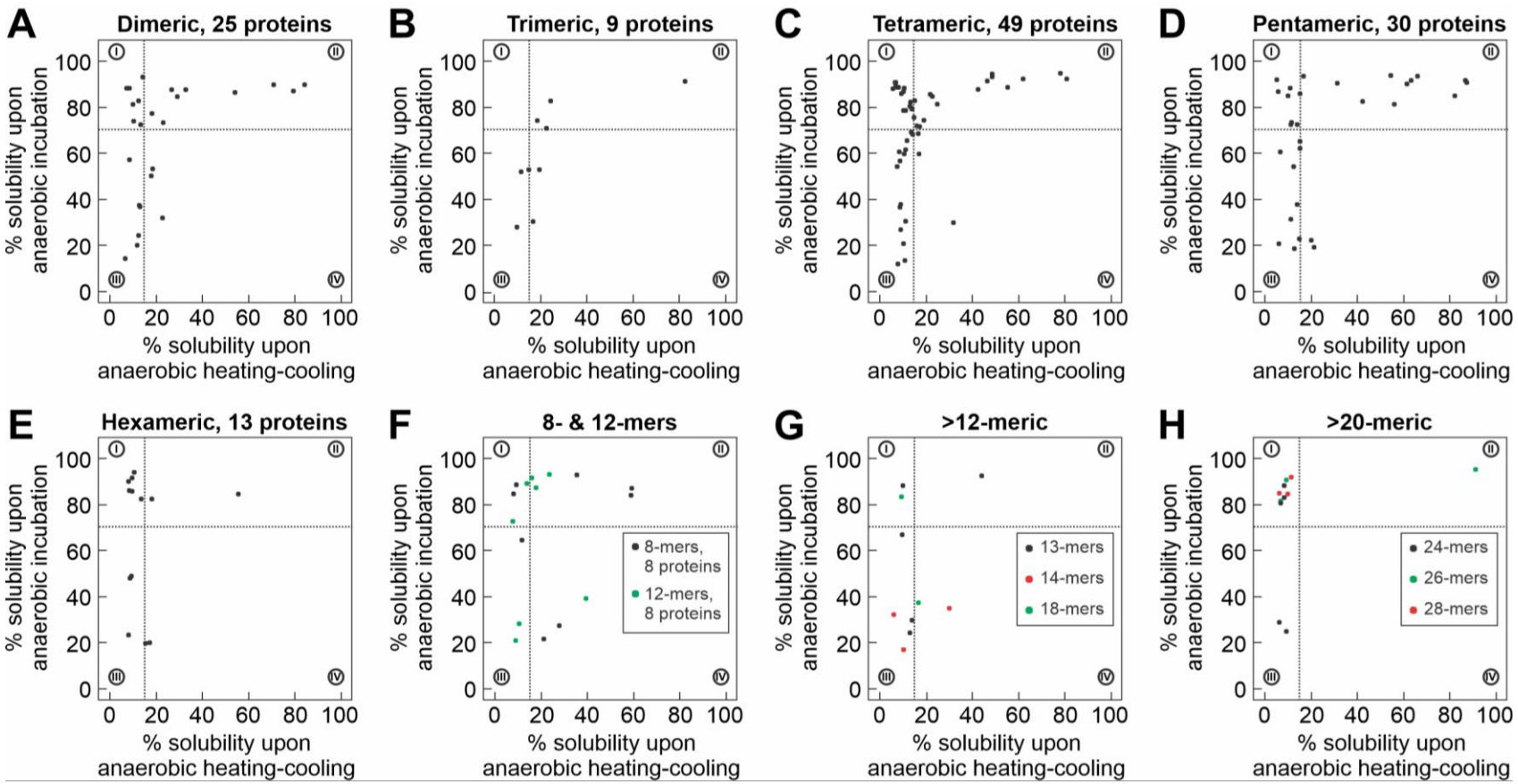
However, many proteins with the following fold-types from the α/β classes are also aggregation-prone under physiologically relevant conditions: (C) Ferredoxin-like, (D) FAD/NAD(P) binding, (E) P-loop NTPases, (F) Ribonuclease H-like, and (G) SAM-dependent methyltransferases. These highly soluble or aggregation-prone fold-types are not well-represented amongst essential proteins, characterized by the KEIO collection.



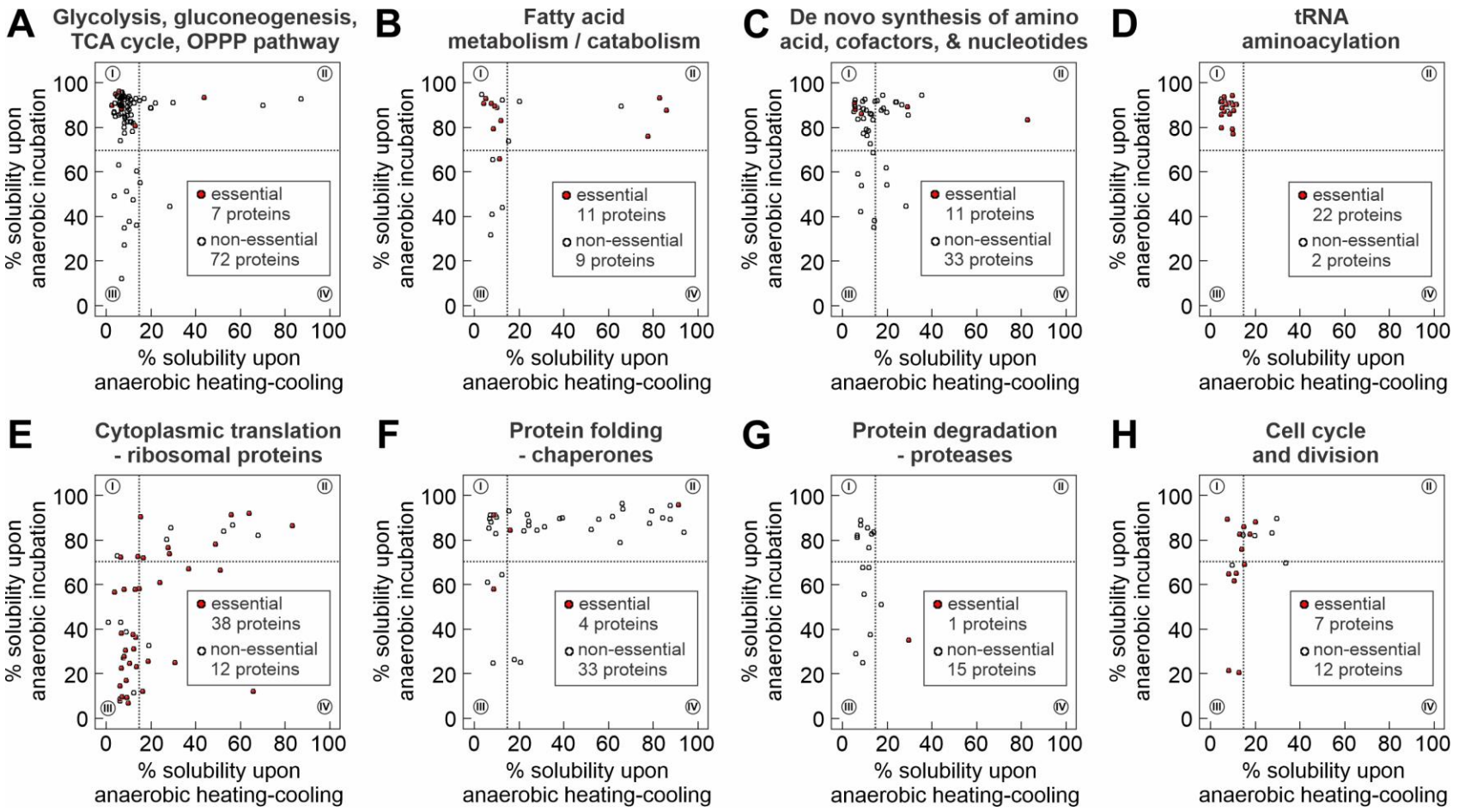
Supplementary Figure S27. Correlation analysis between protein solubility and cofactors: (A) Flavins, (B) 4Fe-4S cluster, (C) Mg^{2+}, (D) Mn^{2+}, (E) Pyridoxal 5-phosphate – PLP, (F) Zn^{2+}. Proteins with (G) no cofactor, and (H) uncharacterized presence or absence of a cofactor are also included. Only cofactor categories with at least 10 proteins are quantified and reported. Proteins with non-covalently bound cofactors, such as PLP and non-Fe-S cluster metal centers, are generally more kinetically stable. Proteins with iron-sulfur clusters are more aggregation prone under our *in vitro* conditions comparing to other cofactor categories.



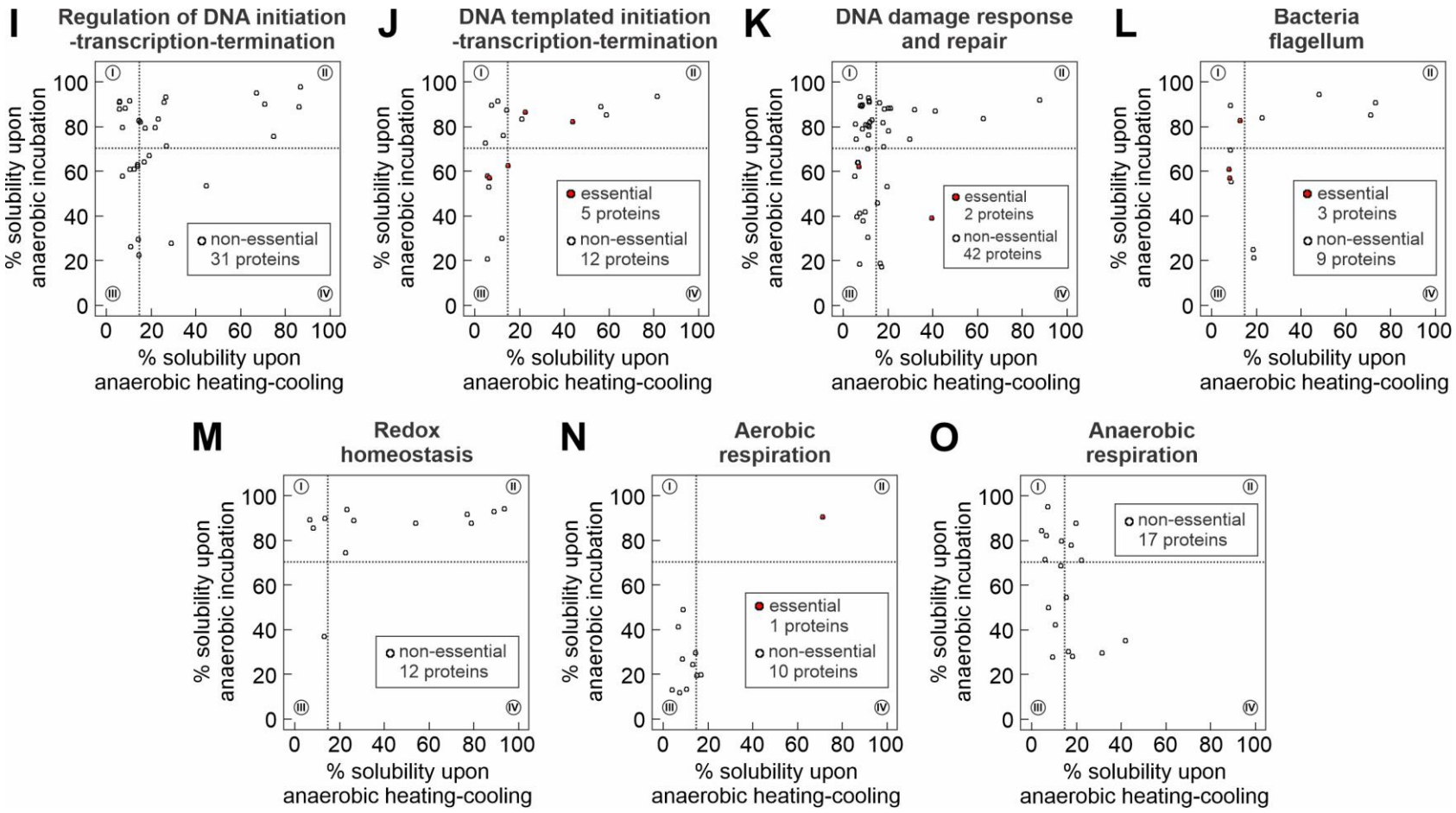
Supplementary Figure S28. Correlation analysis between protein solubility and number of subunits of homomeric proteins. Kinetic stability relative to aggregation applies to many proteins that are (A) monomeric, (B) homodimeric, (C) homotrimeric, (D) homotetrameric, (E) homohexameric, (F) homooctameric. Proteins found to (G) engage in higher order complexes with other proteins or (H) uncharacterized are also shown. Interestingly, homomeric proteins known to form complexes of 6 or more subunits are generally kinetically stable as shown in (E) and (F).



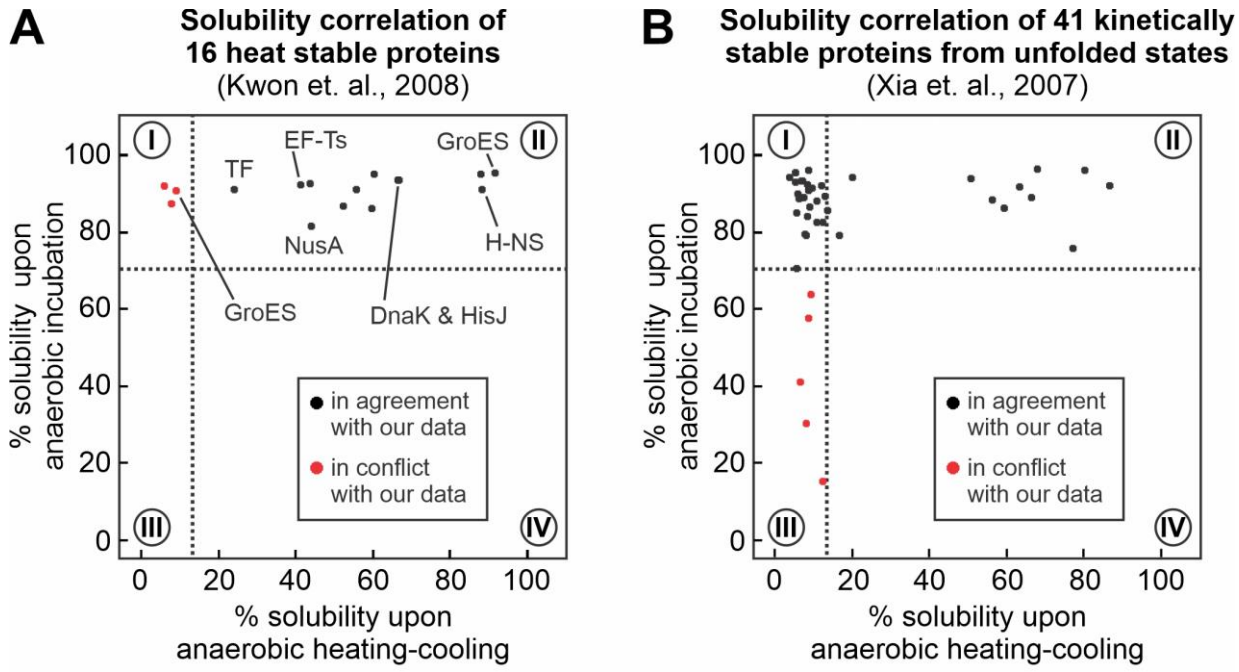
Supplementary Figure S29. Correlation analysis between protein solubility and number of subunits of heteromeric proteins. Kinetic stability relative to aggregation applies to many proteins that are (A) heterodimeric, (B) heterotrimeric, (C) heterotetrameric, (D) hetero pentameric, (E) heterohexameric, (F) hetero-8- & -12-mers, (G) > hetero-12-mers, and (H) > hetero-20-mers.



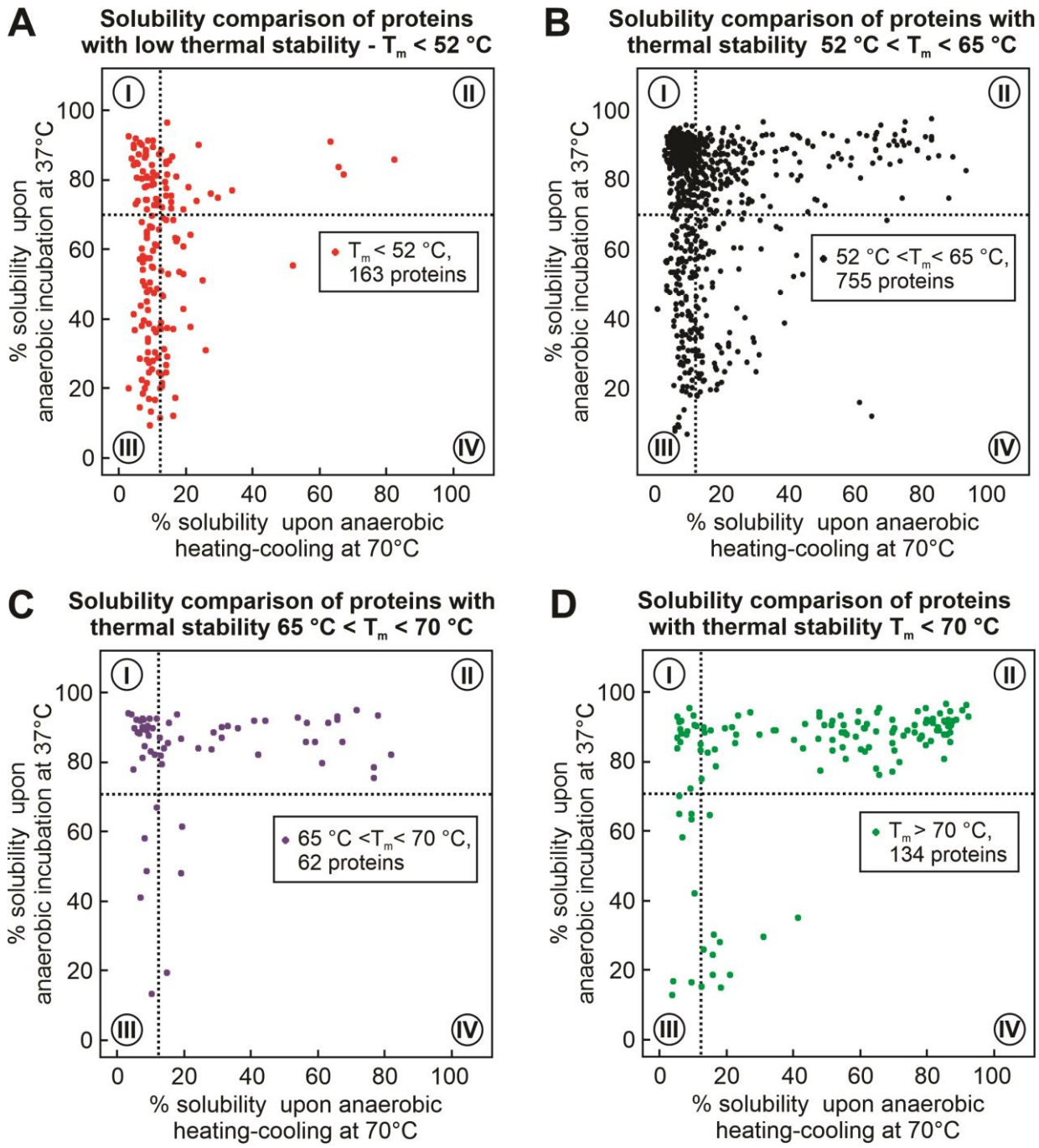
Supplementary Figure S30, part 1. Correlation analysis between protein solubility and cellular functions (Gene Ontology database): (A) Carbohydrate metabolism/catabolism: Glycolysis, Gluconeogenesis, Tricarboxylic acid – TCA cycle, and Oxidative Pentose Phosphate Pathway (OPPP), (B) Fatty acid metabolism/catabolism, (C) de novo synthesis pathways of amino acids, cofactors, and nucleotides, (D) tRNA aminoacylation (23 essential tRNA synthetases), (E) Translation (ribosomal proteins), (F) Protein folding (molecular chaperones), (G) Protein degradation (proteases), and (H) Cell cycle and divisions. Proteins involved in pathways important during growth phase are kinetically protected from aggregation, including (A), (B), (C), (D), (F), and (H).



Supplementary Figure S30, part 2. Correlation analysis between protein solubility and cellular processes (Gene Ontology database): (I) Regulation of DNA initiation-transcription-termination, (J) DNA templated initiation-transcription-termination, (K) DNA repair and damage response, (L) Bacteria flagellum, (M) Redox homeostasis, (N) Aerobic respiration, and (O) Anaerobic respiration.

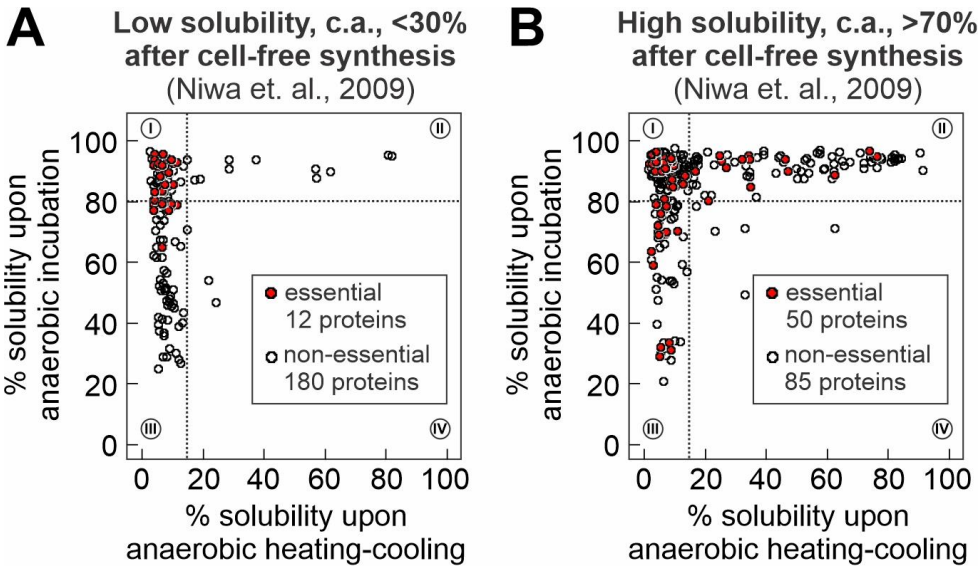


Supplementary Figure S31. Correlation analysis between solubility and kinetic or heat stability of known proteins from literature. (A) Solubility correlation of 16 heat stable proteins that were determined to be resistant to aggregation under harsh thermal treatments by 2D-SDS-PAGE (Kwon et. al., 2008). Here, we show that all previously identified heat stable proteins are kinetically stable as well from aggregation. (B) Solubility correlation of 41 kinetically stable proteins that were determined to be resistant to SDS unfolding by 2D-SDS-PAGE (Xia et. al., 2007). 36 of the previously identified kinetically stable (relative to unfolded states) proteins are also confirmed to be kinetically stable (relative to aggregation) proteins from our work.

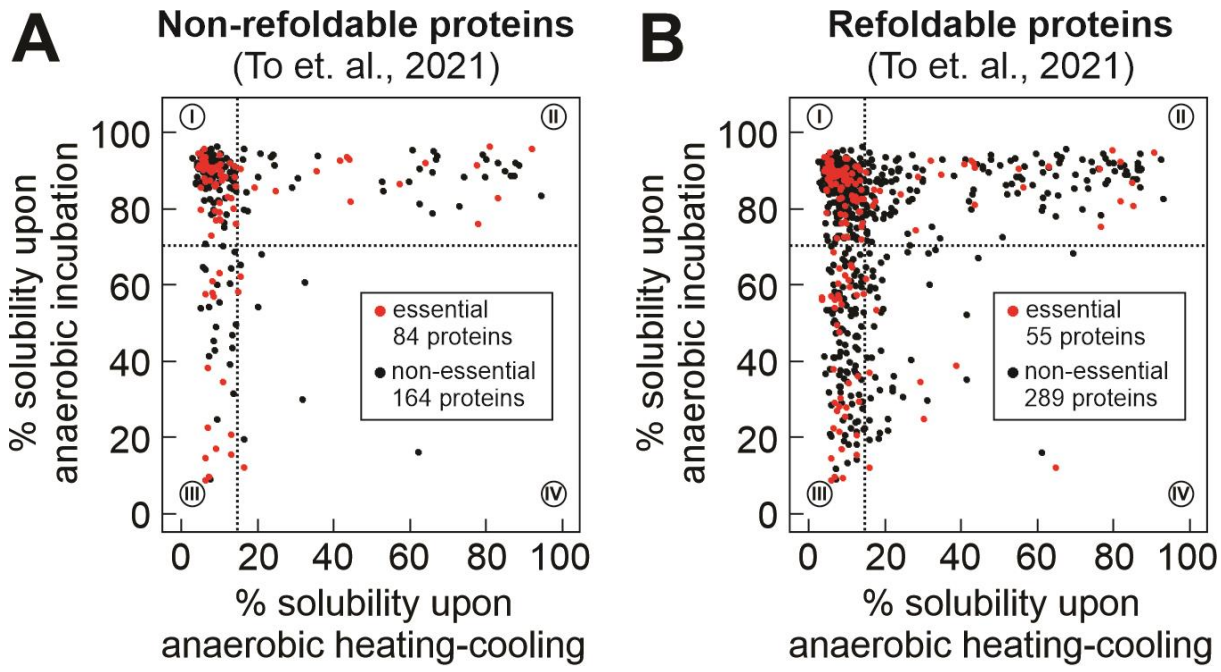


Supplementary Figure S32. Correlation analysis between protein solubility and thermal stability. Thermal stabilities, described by the melting temperature T_m, for proteins from our experiments were obtained from Mateus et. al. (A) Solubility correlation of proteins with high T_m; Despite previous reported resistant to aggregation, proteins with T_m > 70 °C or 65 °C < T_m < 70 °C form extensive insoluble aggregates upon heating at 70 °C during our anaerobic heating-

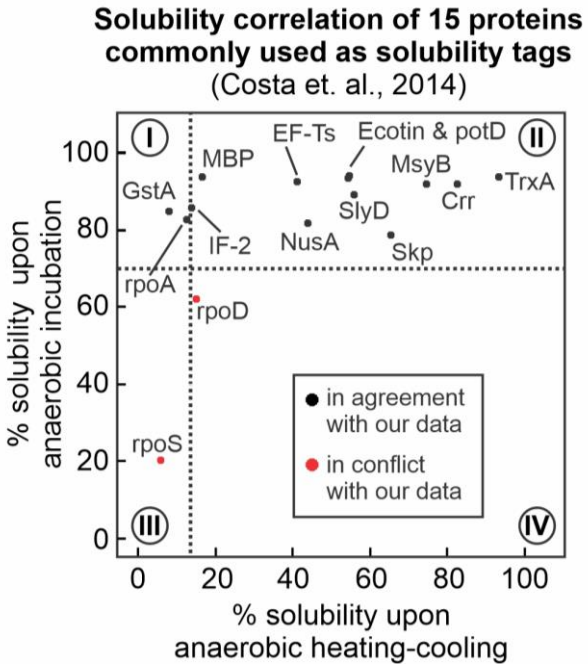
cooling treatment. (B) Solubility correlation of proteins with low T_m. Similar to previous reported levels of aggregation for these proteins, our data show that most proteins with T_m < 52 °C are insoluble upon heating at 70 °C during our treatment. Overall, no correlation is observed between kinetic stability or propensity to form insoluble aggregates and thermal stability; 0.2 > r_{corr} > - 0.2 (Pearson's correlation test). (C) and (D) show the solubility correlation of notable proteins with either high or low thermal stability listed directly in Mateus et. al.



Supplementary Figure S33. Correlation analysis between protein solubility at the late stages and at the early stages of protein's life (i.e., upon translation and release from the ribosome, Niwa et. al). Overall, many proteins are kinetically protected from aggregation regardless of having (A) low solubility (<30%) or (B) high solubility (>70%) upon synthesis at the early stages protein (Niwa et. al., 2009). Generally, essential proteins, characterized under the KEIO collection, are kinetically protected from aggregation relative to non-essential proteins regardless of solubility. Specifically, proteins with low solubility (i.e., form insoluble aggregates) upon cell-free synthesis are also aggregation-prone upon anaerobic thermal treatment. In addition, Quadrant II contains many proteins with high solubility upon cell-free synthesis, suggesting that they are also highly soluble throughout the late-stages.



Supplementary Figure S34. Correlation analysis between protein solubility and refoldability from a chemically denatured state. Many proteins, that are (A) non-refoldable (B) refoldable upon chemical denaturation after 2 hrs (To et. al., 2021), are kinetically stable from aggregation under physiologically relevant conditions.



Supplementary Figure S35. Most solubility tags from *E. coli* are kinetically stable relative to aggregation. Solubility correlation of proteins that are routinely used as solubility tags to enhance protein expression and purifications, according to Costa et. al. (2014).

Chapter 3

Development of highly sensitive NMR techniques with isotopic labelling – study protein folding, misfolding, and aggregation

This Chapter reproduces an article published: Yang, H.; Li, S.; Mickles, C. A.; Guzman-Luna, V.; Sugisaki, K.; Thompson, C. M.; Dang, H. H.; Cavagnero, S., Selective Isotope Labeling and LC-Photo-CIDNP Enable NMR Spectroscopy at Low-Nanomolar Concentration. *Journal of the American Chemical Society* **2022,** *144* (26), 11608-11619. H.Y. carried out NMR experiments, designed the theoretical basis, performed model fitting, and wrote the manuscript. S.L. performed additional model fitting and wrote the manuscript. C.M. performed synthesis and purification of the materials used in this Chapter and participated in writing the manuscript. V.L. prepared cell-free S30 extracts and participated in producing figures in the Chapter . K.S. provided the computational model to predict g-factor values of fluorescein and photo-CIDNP polarization. H.D. contributed to the synthesis, purification, and characterizations of materials used in this Chapter. S.C. designed the project, participated in data analysis, wrote the manuscript and contributed to manuscript editing.

3.1. Abstract.

NMR spectroscopy is a powerful tool to investigate molecular structure and dynamics. The poor sensitivity of this technique, however, limits its ability to tackle questions requiring dilute samples. Low-concentration photochemically induced dynamic nuclear polarization (LC-photo-CIDNP) is an optically-enhanced NMR technology capable of addressing the above challenge by increasing the detection limit of aromatic amino acids in solution up to 1,000-fold, either in isolation or within proteins. Here, we show that the absence of NMR-active nuclei close to a magnetically active site of interest (e.g., the structurally diagnostic H^a-C^a pair of amino acids) is expected to significantly increase LC-photo-CIDNP hyperpolarization. Then, we exploit the spin-diluted tryptophan isotopolog Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 and take advantage of the above prediction to experimentally achieve a ca 4-fold enhancement in NMR sensitivity over regular LC-photo-CIDNP. This advance enables the rapid (within seconds) detection of 20 nM concentrations or the molecule of interest, corresponding to a remarkable 3 ng detection limit. Finally, the above Trp isotopolog is amenable to incorporation within proteins and is readily detectable at 1 mM concentration in complex cell-like media, including *E. coli* cell-free extracts.

3.2. Introduction.

Nuclear magnetic resonance (NMR) spectroscopy is a non-invasive technique capable of yielding invaluable information on molecular structure and dynamics at atomic resolution. On the other hand, NMR is intrinsically insensitive due to small differences in population between nuclear spin eigenstates, even at the highest commercially available magnetic field of 28.2 T [151-152]. Although low sensitivity can be compensated by increased sample concentration, this is often not a feasible option due to limited solubility, small amounts of material and high aggregation propensities. The above challenges are particularly severe in the case of biological samples in aqueous media. Numerous hyperpolarization methods were developed to increase the sensitivity of liquid-state NMR by altering nuclear spin populations at thermal equilibrium. Major developments comprise Overhauser dynamic nuclear polarization (ODNP) [153-155], dissolution dynamic nuclear polarization (DDNP) [156-158], and parahydrogen-induced polarization (PHIP) [159-165], including signal amplification by reversible exchange (SABRE) [166-183].

On the other hand, a few challenges have hampered the wide applicability of the above hyperpolarization technologies, to date. These include expensive instrumentation, harsh sample conditions (e.g., high/low temperatures, freeze-thaw cycles), long hyperpolarization times, applicability to a limited number of molecules of interest, safety considerations, inability

to perform *in situ* hyperpolarization, requirement of high sample concentration (50 mM or more) and unfeasible data collection in physiologically relevant environments.

Earlier identification of chemically induced dynamic nuclear polarization (CIDNP) [184-185] in organic reactions led to the development of a photochemically enhanced NMR approach known as photo-CIDNP. While originally employed to assess solvent exposure [186], this technique has more recently been employed to significantly increase NMR sensitivity in solution at high applied field [187-193]. Photo-CIDNP is devoid of most of the drawbacks of other approaches listed in the previous paragraph. This technique requires aromatic moieties and benefits from an O_2-free environment.

Photo-CIDNP exploits spin-selective processes between a molecule of interest and a photosensitizer dye. The phenomenon proceeds via a radical pair mechanism, as shown in Figure 1A. Photo-CIDNP can be employed for NMR sensitivity enhancement purposes [194] and, when combined with heteronuclear correlation spectroscopy, it enables highly sensitive detection of hyperpolarized ^{13}C and ^{15}N [188-189, 195]. More recently, trace amounts of commercially available glucose oxidase (GO) and catalase (CAT) were employed to remove dissolved molecular oxygen, thus preventing excited-state dye quenching and further increasing NMR sensitivity [187]. The use of fluorescein (Fig. 1B) as photosensitizer, in place of flavin mononucleotide (FMN), led to extension of this technique to low sample concentration (low- to sub-mM) [190]. The use of photosensitizers characterized by a long triplet-state lifetime

(e.g., 20 ms for fluorescein [190, 196]) gave rise to a branch of photo-CIDNP known as low-concentration photo-CIDNP (LC-photo-CIDNP). Replacement of costly and hard-to-maintain high-power laser with LEDs (Fig. 1C) with no sensitivity losses led to enhanced ease of access, portability and safety [192].

LC-photo-CIDNP performs optimally at low-to-sub mM sample concentration [190-191], and it provides fast (0.2 s) *in situ* nuclear hyperpolarization. The latter features enable long-term signal-averaging and multi-dimensional NMR of extremely dilute samples. This technology leads to enhanced signal-to-noise ratio (S/N) of tryptophan (Trp) and tyrosine (Tyr), either as free amino acids or within polypeptides and proteins [188, 191-193]. Remarkably, high LC-photo-CIDNP enhancements are observed for C^a nuclei of aromatic amino acids, likely due to their high hyperfine coupling constants [188, 197], enabling the efficient detection of H^a-C^a pairs. The C^a chemical shifts are particularly useful because they are diagnostic of backbone secondary structure in proteins [198]. To take advantage of proton's higher sensitivity, ^{13}C-to-^{1}H polarization transfer via reverse INEPT is typically employed [188]. In addition, 2D data collection is carried out to preserve information on $^{13}C^a$ chemical shifts [188, 191]. The recently developed ^{13}C RASPRINT pulse sequence is optimized for ultra-fast data collection in aqueous solvent [192].This sequence exploits minimal recycle delays, given that most nuclear-spin polarization is generated during the optical irradiation time, thus abrogating the need for lengthy inter-scan intervals [192]. The ^{13}C RASPRINT sequence is particularly useful for the ultra-rapid detection of

H^a-C^a pairs at low sample concentration, down to ca. 500 nM [192]. More recent studies showed that LC-photo-CIDNP hyperpolarization can be further enhanced in the presence of low-mM concentrations of reductive radical quenchers including ascorbic acid, a.k.a. vitamin C, leading to detection of 200 nM aromatic amino acids [193]. Despite the above advances, it is highly desirable to render liquid-state NMR spectroscopy even more sensitive so that it approaches detection limits of other leading techniques, e.g., fluorescence and mass spectrometry.

In this work, we provide theoretical and experimental evidence illustrating that nearby NMR-active nuclei (e.g., ^{1}H's, ^{13}C and ^{15}N) significantly decrease the extent of observable ^{13}C LC-photo-CIDNP polarization. Upon replacement of these nuclei with either NMR-inactive (^{12}C) or low-gyromagnetic-ratio (^{2}H) nuclei by selective isotope labeling, we were able to achieve an unprecedented improvement in NMR detection level in solution, down to 20 nM Trp, in only a few seconds. In addition, we showed that the above isotopolog-assisted optically enhanced technology is also effective in complex cell-like media, enabling the analysis of aromatic amino acids in physiologically relevant milieu at atomic resolution and extremely high sensitivity.

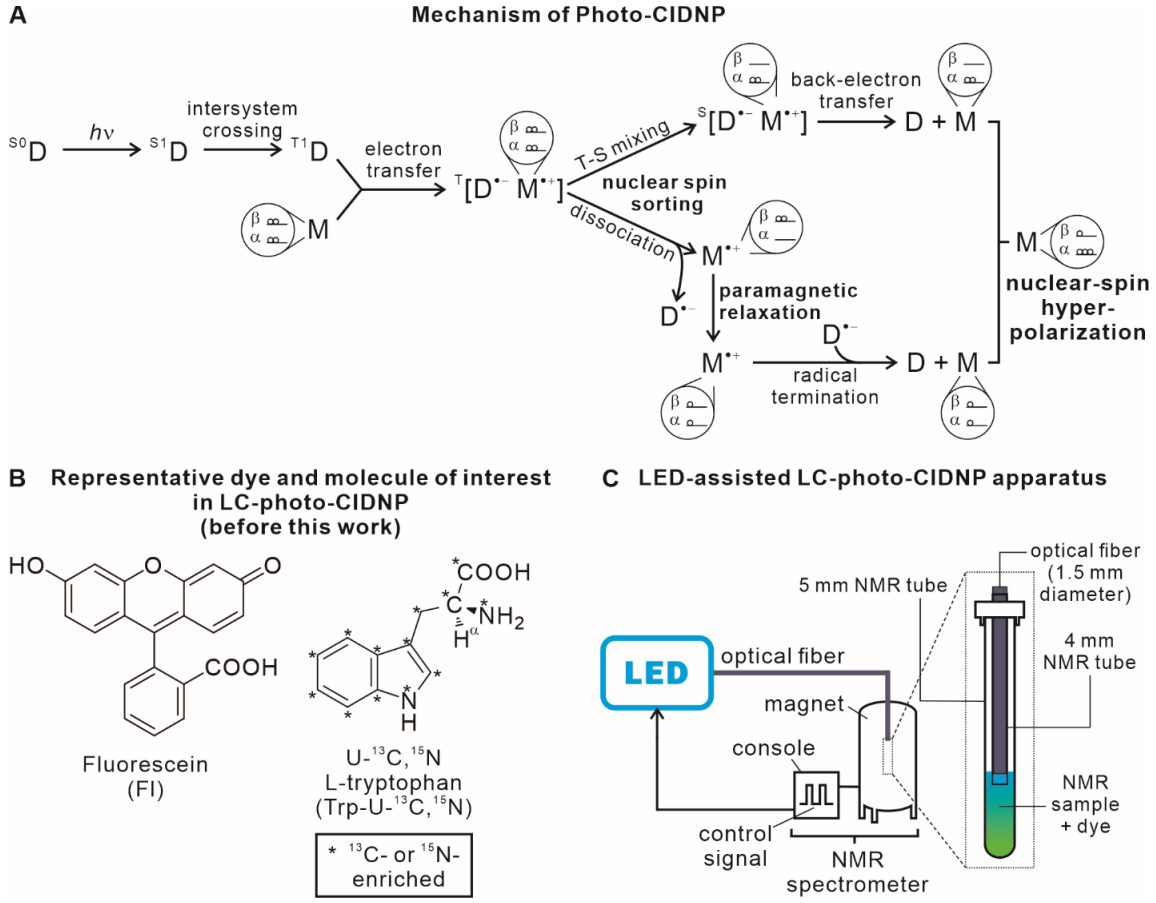


Figure 1. Basic aspects of low-concentration photo-chemically induced dynamic nuclear polarization (LC-photo-CIDNP). A. Schematic illustration of the radical-pair mechanism of photo-CIDNP in liquids. D and M denote the dye and molecule of interest, respectively. S_0, S_1 and T_1 refer to the ground singlet, 1st excited singlet and 1st excited electronic triplet states, respectively. The T-S acronym denotes triplet-singlet mixing. The frequency of T-S mixing of the $^{T}[D^{\bullet-}M^{\bullet+}]$ radical pair depends on the nuclear spin states of the molecule of interest $M^{\bullet+}$, due to hyperfine coupling between unpaired electron and nuclear spins. Therefore, depending on the value of nuclear spin states, molecules can either undergo fast or slow T-S mixing. The population with nuclear spin states undergoing slow T-S mixing experiences radical-pair dissociation followed by paramagnetic relaxation, resulting in thermal nuclear-spin polarization. Conversely, the population with nuclear spin states experiencing fast T-S mixing undergoes T-S mixing followed by fast back electron

transfer. Importantly, this population experiences nuclear-spin hyperpolarization. In practice, one detects the sum of the two populations, in a photo-CIDNP experiment. B. Structure of a typical small molecule of interest and photosensitizer dye employed in LC-photo-CIDNP. C. Scheme illustrating typical LC-photo-CDINP instrumentation.

3.3. Results and Discussions

3.3.1 Theoretical predictions on the dependence of LC-photo-CIDNP-hyperpolarization on isotope substitution patterns and g-factor values. LC-photo-CIDNP polarization depends on the g-factor of the radical form of the reduced photosensitizer dye. Fluorescein is typically employed as a dye. The g-factor of its reduced radical form in solution, however, is not known. Note that different protonation states of fluorescein in solution are expected to affect g-factor values. We carried out density functional theory (DFT) calculations to estimate the pK_a and g-factor of the fluorescein dye radical at 25 ºC and pH 7.0. Water solvation effects were taken into account via the integral equation formalism polarizable continuum model (IEFPCM). Predicted g-factor values were determined for the fully protonated state of the $Fl^{\bullet}$ dye radical and for seven variably deprotonated states, as listed in Figure 2A. The pK_a values of the three titratable protons were employed to determine the population of each of the eight variably protonated species listed in Figure 2A via the Henderson-Hasselbalch equation. The resulting individual populations and corresponding g-factors were employed to predict the overall weighted-average g-factor of the fluorescein dye radical at 25 ºC and pH 7.0. This g-factor value was mostly contributed by $Fl^{\bullet}$ and was found to be 2.003077.

The known g-factor of $Trp^{\bullet+}$ [197] and the weighted-average g-factor of the fluorescein dye radical, determined as described in the previous paragraph, were used as input for

computational predictions of LC-photo-CIDNP polarization as a function of g-factor of the dye radical. We followed known theoretical and practical guidelines [199-201], with some modifications, as detailed below.

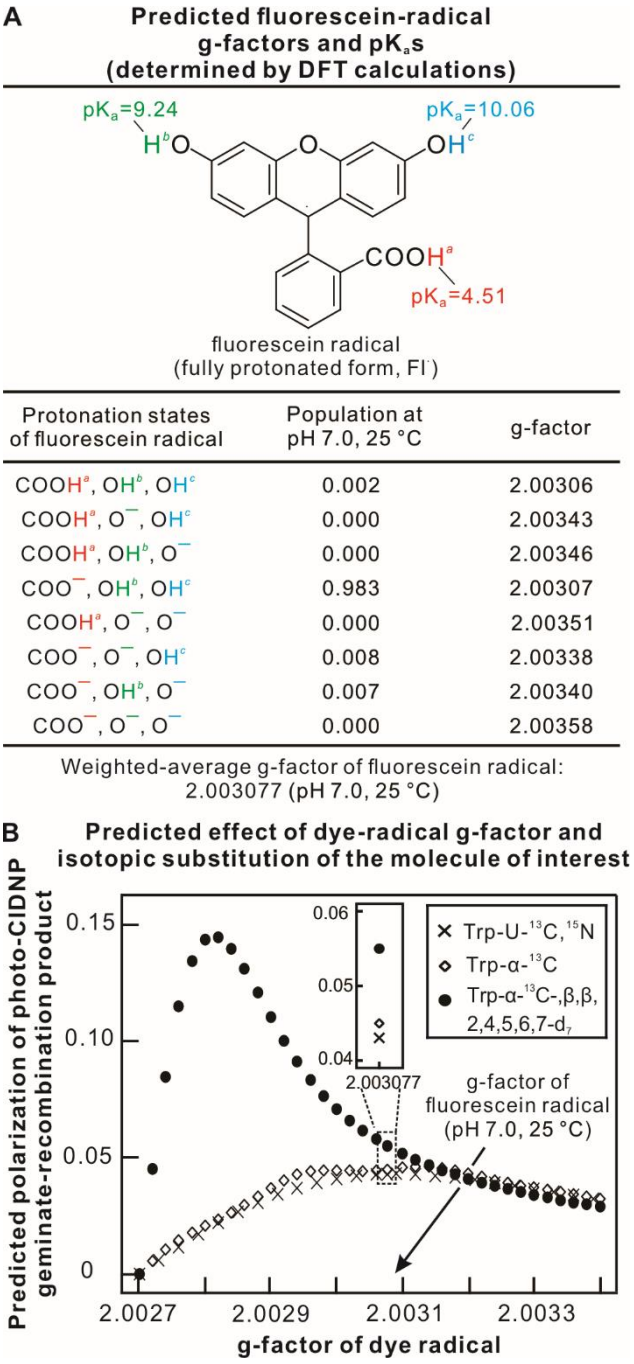
For freely-diffusing radical pairs, the population difference between any two nuclear-spin configurations 1 and 2 of the geminate-recombination product is

$$p_1 - p_2 = p\left(\sqrt{|\omega_{TS1}|\tau_d} - \sqrt{|\omega_{TS2}|\tau_d}\right) \quad , \qquad [1]$$

where p_1, p_2 are the populations with nuclear-spin configuration 1 and 2, respectively; p is a normalization factor (see Supplementary Information), ω_{TS1} and ω_{TS2} are the triplet-singlet mixing frequencies of the nuclear-spin configurations, and τ_d is the average time the radical-pair components remain closely associated (beyond orbital-overlap distances) before long-term dissociation. The triplet-singlet mixing frequency of each of any possible nuclear-spin configurations is

$$\omega_{TS} = \frac{1}{2}\left[\Delta g \mu_B B_0 + \sum_i m_i A_i - \sum_j m_j A_j\right] \quad , \qquad [2]$$

where the *i* and *j* subscripts denote individual nuclei of the dye and molecule of interest, respectively. In addition, D*g* is the difference between the g-factors of the radical-pair components (i.e., the dye and molecule of interest), m_B denotes the Bohr's magneton B_0 is the applied magnetic field, and *m* and *A* are the magnetic nuclear-spin quantum number and hyperfine coupling constant, respectively.



Protonation states of fluorescein radical	Population at pH 7.0, 25 °C	g-factor
$COOH^a$, OH^b, OH^c	0.002	2.00306
$COOH^a$, O^-, OH^c	0.000	2.00343
$COOH^a$, OH^b, O^-	0.000	2.00346
COO^-, OH^b, OH^c	0.983	2.00307
$COOH^a$, O^-, O^-	0.000	2.00351
COO^-, O^-, OH^c	0.008	2.00338
COO^-, OH^b, O^-	0.007	2.00340
COO^-, O^-, O^-	0.000	2.00358

Weighted-average g-factor of fluorescein radical: 2.003077 (pH 7.0, 25 °C)

Figure 2. Computational predictions. A. Predicted pK_a and isotropic g-factor values for fluorescein radicals of different protonation states. B. Predicted photo-CIDNP polarization of three Trp isotopologs as a function of dye g-factor. Simulations were performed using g-factor of $Trp^\bullet$ = 2.0027, and hyperfine coupling constant of Trp and fluorescein radicals from literature sources (see details in Suppl. Inf.).

For different isotope labeling schemes of the molecule of interest, corresponding to different isotopologs, the third term of equation [2] is expected to vary depending on specific A_j values. Therefore, we reasoned that different isotopic substitutions ought to give rise to variable populations of geminate-pair products, resulting into different levels of photo-CIDNP hyperpolarization [202]. Hence we proceeded to explore how the g-factor of dye radicals is expected to affect the theoretically achievable geminate-pair recombination products, in the case of different Trp isotopologs. We focused on three representative labeling schemes of the molecule of interest: namely, *(i)* uniformly ^{13}C-labeled Trp (Trp-U-^{13}C), *(ii)* a hypothetical Trp isotopolog bearing a ^{13}C-^{1}H pair at the a carbon (C^{a}) and extensive deuteration at non-exchangeable sites (Trp-a-^{13}C-b,b,2,4,5,6,7-d$_7$, Fig. 3A), and, finally, *(iii)* a reference Trp isotopolog enriched in ^{13}C at the C^{a} and bearing natural-abundance nuclides at all other sites (Trp-a-^{13}C, Fig. 3B). Importantly, all of the above isotopologs have a ^{13}C at the a position, honoring the fact that the LC-photo-CIDNP hyperpolarization approach is particularly effective at enhancing the ^{13}C^{a} of the Trp and Tyr amino acids, either alone or within proteins [191]. Additional details about the computations, including assessment of overall probabilities of geminate recombination, proportional to nuclear-spin polarization at the Trp a carbon, are provided in the Supplementary Information.

The simulation results for the above three Trp isotopologs are shown in Figure 2B. Even at optimal values of g-factor of the dye radical, the two compounds bearing no

deuteration are predicted to afford only limited sensitivity enhancement. On the other hand, the Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 deuterated isotopolog (solid circles in the plot of Fig. 2B) is expected to undergo drastic polarization enhancements that significantly vary in magnitude depending on the g-factor of the dye radical. These enhancements extend up to nearly 14.5% for dyes that bear an optimal g-factor value.

Under typical LC-photo-CIDNP conditions, fluorescein is employed as photosensitizer [191]. As detailed elsewhere [190], this dye bears a conveniently long triplet-state lifetime that persists across the infrequent collisions between dye and molecule of interest, at moderate sample concentrations (low mM and lower). Therefore, despite its sub-optimal g-factor value in the plot of Figure 2B, fluorescein is the current photosensitizer of choice in LC-photo-CIDNP experiments. When fluorescein is employed as a photosensitizer at room temperature and 25 ºC, the predicted probability of geminate recombination (proportional to nuclear-spin hyperpolarization [202]) of uniformly ^{13}C-labeled Trp (Trp-U-^{13}C) is the lowest (ca. 5%). On the other hand, Trp-a-^{13}C is predicted to experience a 4.4% more geminate-recombination than Trp-U-^{13}C. More strikingly, the selectively labeled isotopolog Trp-a-^{13}C-b,b,2,4,5,6,7-d_7, which includes extensive deuteration at C^b and aromatic carbon sites, is predicted to experience 28.2% more geminate recombination than Trp-U-^{13}C.

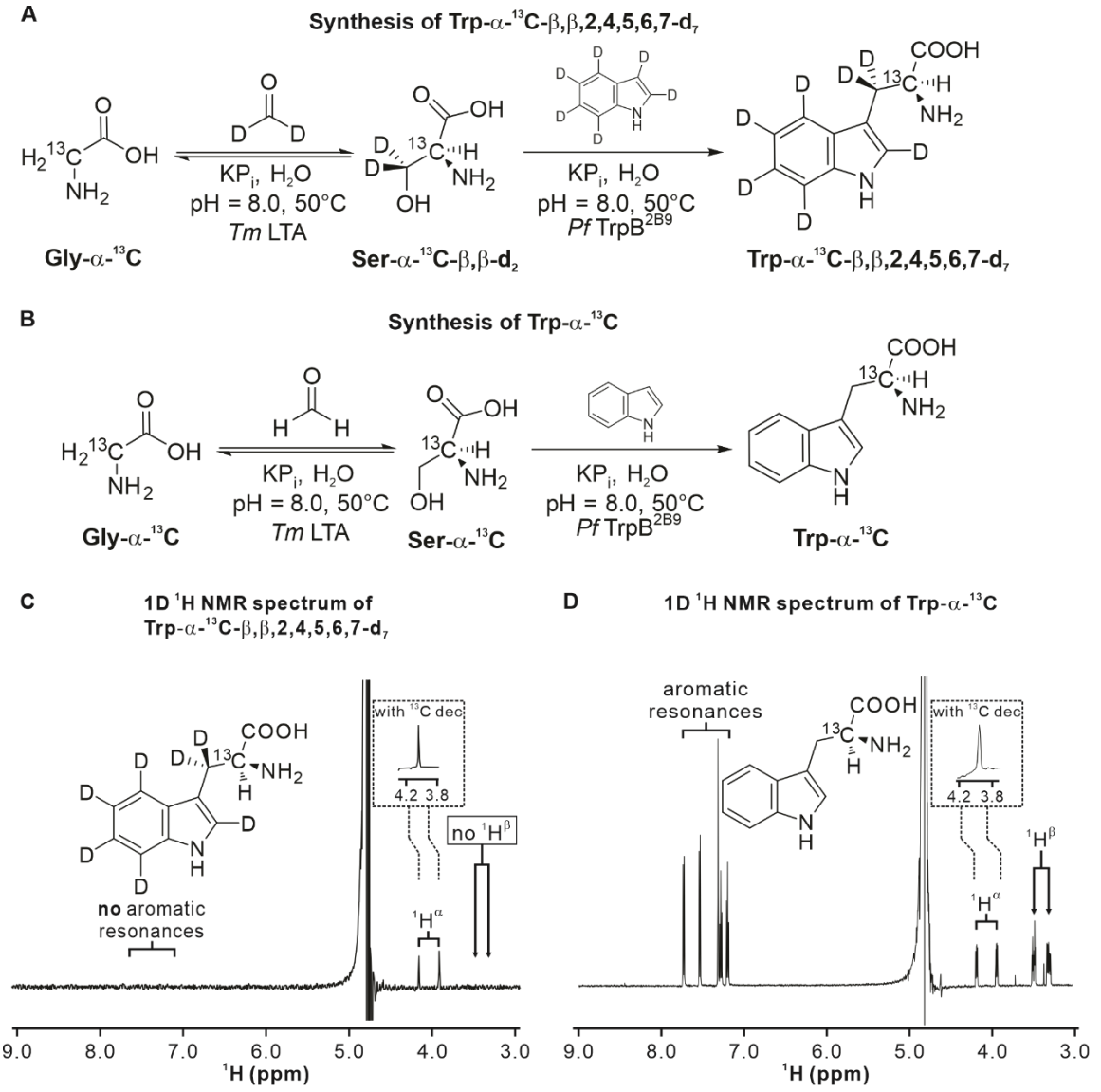


Figure 3. Synthesis and characterization of Trp isotopologs. A. Scheme illustrating the enzyme-cascade synthesis of Trp-a-13C-b,b,2,4,5,6,7-d7. B. Scheme illustrating the enzyme-cascade synthesis of Trp-a-13C. C. 1D 1H NMR characterization of Trp-a-13C-b,b,2,4,5,6,7-d7. A pulse-acquire data-collection scheme with a 121.5 Hz solvent presaturation scheme during the entire 5 s recycle delay was employed. 8 scans were acquired. The sample (ca. 100 mM) was prepared in 90% H2O, 10% D2O with traces of sodium trimethylsilylpropanesulfonate (DSS) used as an internal reference. The inset shows the 1Ha region of a spectrum including 13C decoupling during acquisition using

bi_p5m4sp_4sp.2 composite pulse decoupling (see Exp. Section in Suppl. Inf.). D. 1D 1H NMR characterization of HPLC-purified and lyophilized Trp-a-13C (ca. 200 mM). The inset shows the 1Ha region of the 13C-decoupled spectrum. The same acquisition and processing parameters as in panel C were used.

3.3.2 Trp analogs bearing variable isotopic substitution patterns were prepared and characterized. Encouraged by the computational predictions of Figure 2B, we proceeded to prepare some relevant isotopologs and experimentally test their performance in terms of achievable LC-photo-CIDNP hyperpolarization.

Previous selective labeling schemes were based on supplementing bacterial growth media with [4-^{13}C]-d-aminolevulinic acid [203] or with different singly labeled ^{13}C-glucose isotopologs [204]. These methods, however, suffer from low isotope-incorporation yields and are not easily customizable to the desired labeling patterns. Here, we employed an alternative recently developed enzyme-cascade approach [205] to readily achieve high levels of isotope enrichment starting from affordable and readily available precursors. This approach is particularly convenient due to its modular nature [205]. For instance, it enables the facile synthesis of a wide variety of desired isotopologs by simply changing the labeling pattern of the starting materials (i.e., glycine and indole) and solvent (water).

The synthetic schemes of Figure 3A and 3B detail the route that led to the generation of Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 and Trp-a-^{13}C with 44.7% and 31.8% yield, respectively. Given that the amounts of purified products were more than adequate for our purposes, no further optimization was performed. The identity of purified Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 and Trp-a-^{13}C was confirmed by ESI MS and ^{1}H NMR. The pulse-acquire ^{1}H NMR spectra of the two isotopologs are shown in Figure 3C and 3D and confirm the presence of the expected ^{13}C-^{1}H

scalar coupling. The latter disappears upon ^{13}C decoupling. In addition, no ^{1}H aromatic resonances are observed for Trp-a-^{13}C-b,b,2,4,5,6,7-d_7, consistent with expectations.

3.3.3 Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 isotopolog leads to large LC-photo-CIDNP sensitivity enhancements. The two Trp isotopologs described in the previous section and the commercially available uniformly labeled Trp-U-^{13}C,^{15}N were initially analyzed under dark (i.e., LED-off) conditions. The ^{13}C RASPRINT pulse sequence [192] was employed, except that the typically short recycle delays (50 ms) [192] were replaced by long ones (5 s) to allow for nearly complete longitudinal relaxation. This pulse sequence enables the detection of ^{13}C-bound protons and is routinely employed in LC-photo-CIDNP [192].

Figure 4A, which focuses on H^a protons under dark conditions, shows that the selectively labeled Trp-a-^{13}C and Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 isotopologs display a larger S/N than uniformly labeled Trp (Trp-U-^{13}C,^{15}N). While Figure 4A provides a useful qualitative assessment, the different sample concentrations prevent quantitative comparisons. A more quantitative analysis of NMR behavior under dark conditions is provided in Figure 4B. The plot in this figure examines both intensities at the resonance maximum and areas, and normalizes the data so that they are independent of sample concentration. Here we note that, relative to Trp-U-^{13}C,^{15}N, the Trp-a-^{13}C isotopolog displays 20% increases in resonance area and intensity at the maximum. Conversely, Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 displays a 42% increase in

area and an even more dramatic 192% increase in intensity at the maximum. The latter enhancement results from narrower ^{1}H linewidths, as confirmed by T_2 measurements (Supplementary Information). In summary, perhaps not surprisingly, detection of Trp H^a resonances under dark conditions is rendered more facile by the extensive deuteration of the Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 isotopolog. It is worth noting that H^a resonances are important in biology because they are diagnostic of structure, especially when analyzed in the context of protein backbone.

Next, we performed 1D LC-photo-CIDNP experiments on all three Trp isotopologs under light (i.e., LED-on) conditions. Data were collected at very low sample concentration (1mM). The NMR spectra under light conditions, shown in Figure 4C, display considerable increases in S/N upon going from Trp-U-^{13}C,^{15}N to Trp-a-^{13}C to Trp-a-^{13}C-b,b,2,4,5,6,7-d_7. Remarkably, the latter isotopolog displays a ca. 4-fold increase in S/N relative to the already intense signal of uniformly labeled Trp (Trp-U-^{13}C,^{15}N). More quantitative studies, displayed in Figure 4D, show that Trp-a-^{13}C displays a 70% increase in resonance area and intensity relative to Trp-U-^{13}C,^{15}N, while Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 features a 120% increase in area and a 300% increase in resonance intensity.

At this juncture, it is important to consider a few simple items. First, NMR sensitivity [151, 206] is defined as S/N divided by the square root of experiment time t, i.e., $\frac{S}{N\sqrt{t}}$. Second, the NMR S/N ratio depends on the amplitude of both the resonance of interest and the noise.

Third, all data were collected with identical acquisition and processing parameters. Taking all of the above considerations into account, the results of Figure 4C and 4D clearly show that the Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 isotopolog displays a remarkable 4-fold increase in sensitivity relative to uniformly labeled Trp (Trp-U-^{13}C,^{15}N). Note that all experiments under light conditions must bear equivalent noise values. This result is even more significant if one notices that (a) all the Figure 4C spectra acquired under dark conditions display no detectable signal, and (b) all the Figure 4 light and dark spectra included only 32 scans. In conclusion, the overall sensitivity of the ^{13}C RASPRINT experiment for the Trp-a-13C-b,b,2,4,5,6,7-d_7 isotopolog must be very high. More details on specific enhancements factors and polarization values are provided in a later section

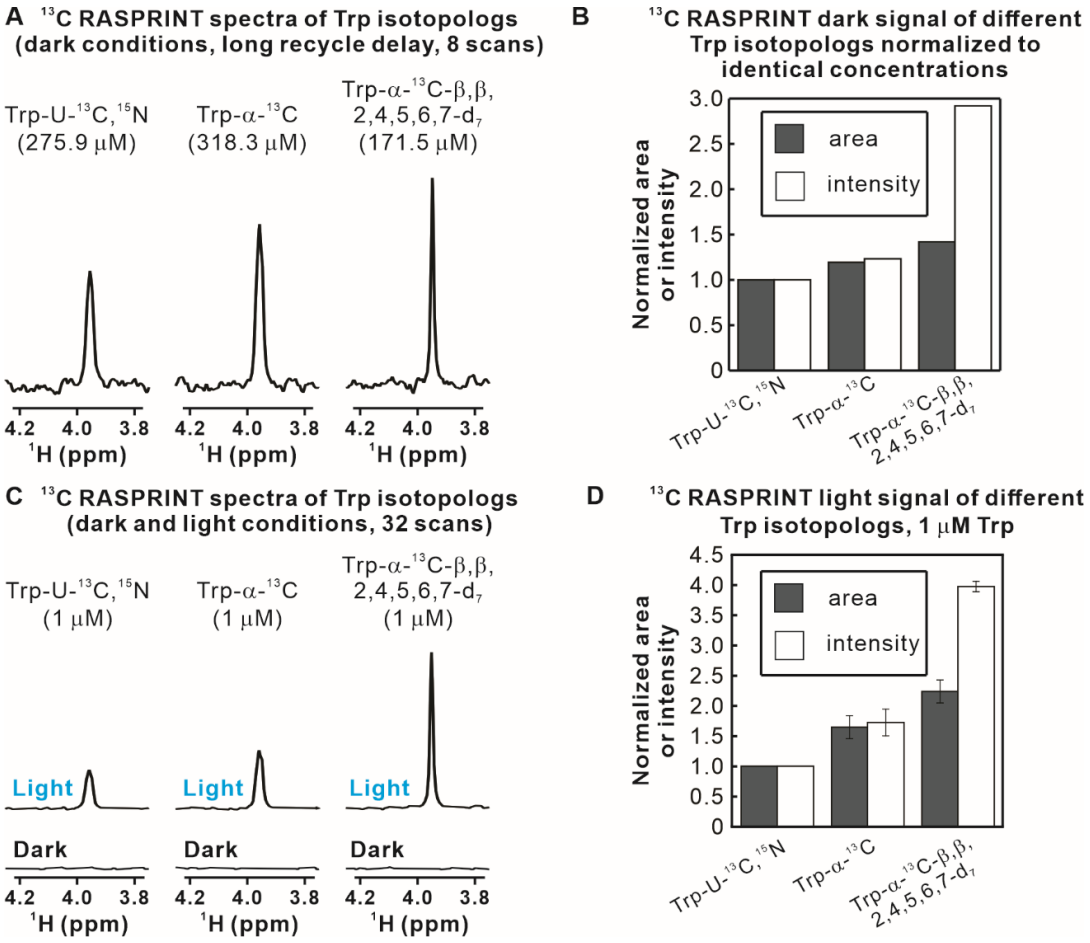


Figure 4. LED-off and LED-on data show enhancements due to both dark and light contributions. A. NMR spectra showing the H$^{\alpha}$ resonance of different Trp isotopologs. Data were collected with the 1D ^{13}C RASPRINT pulse sequence under dark (i.e., LED-off) conditions at the concentrations listed above the spectra. A 5 s recycle delay was employed. B. Quantitative analysis of H$^{\alpha}$-resonance areas and intensities-at-the-maximum relative to Trp-U-^{13}C,^{15}N, after normalization to eliminate differences due to variable concentrations. C. LC-photo-CIDNP spectra acquired with the 1D ^{13}C RASPRINT pulse sequence under light (LED-on) conditions for three Trp isotopologs of equal concentration. D. Quantitative analysis of H$^{\alpha}$ resonance area and intensity-at-the-maximum under light (i.e., LED-on) conditions relative to Trp-U-^{13}C,^{15}N. Data in panels B and D are shown as avg ± S.E. (n=2). All spectra were acquired at 24 °C with 64 scans, 4 dummy scans, and an optical irradiation time of 0.2 s.

3.3.4 The Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 isotopolog enables optically-enhanced NMR at low nanomolar concentration. Encouraged by the results described in the previous sections, we further explored the LC-photo-CIDNP detection limit of the Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 isotopolog. Towards this end, we took a few deliberate simple steps. First, to optimize detection of the target isotopolog, we modified the ^{13}C RASPRINT pulse sequence by removing its constant-time evolution. Note that the absence of ^{13}Cs at the b and carbonyl positions ($^{13}C^{b}$ and ^{13}C') of the Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 isotopolog renders constant-time evolution unnecessary, due to lack of ^{13}C-^{13}C J couplings [207]. Second, we converted all the on-resonance selective pulses into hard pulses and removed the ^{13}C' off-resonance pulse. These modifications are justified by the fact that the Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 isotopolog only bears one ^{13}C and lacks ^{13}C-^{13}C J couplings. Therefore, there is no need for $^{13}C^{a}$ selective excitation or for selective C' pulses to refocus $^{13}C^{a}$-^{13}C' J couplings.

Importantly, all the above modifications are beneficial because they lead to shorter experiment times and reduce signal losses arising from pulse imperfections and transverse relaxation during the pulse scheme. The resulting pulse sequence, displayed in Figure 5A, was denoted as ^{13}C isotopolog-optimized RASPRINT or ^{13}C iso-RASPRINT.

Then, we tested the LC-photo-CIDNP behavior of extremely dilute (50 and 20 nM) Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 upon data collection with the ^{13}C iso-RASPRINT pulse sequence. The results (two outmost-left spectra in Fig. 5B) show that the 50 nM sample yields a detectable resonance of moderate intensity after 64 scans (Fig. 5B). On the other hand, the 20 nM sample displayed no observable signal (data not shown). Given that reductive radical quenchers including ascorbic acid, also known as vitamin C (VC) is known to protect LC-photo-CIDNP molecules of interest from photodamage [193], potentially extending detection limit, we prepared an additional set of samples containing a small concentration of VC (2 mM). The results, shown in the four righthandside panels of Figure 5B, prove that in the presence of the VC

reductive radical quencher even 20 nM Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 can be readily detectable, within only 64 scans. This is a remarkably small concentration, corresponding to an extremely favorable detection limit of 3 ng.

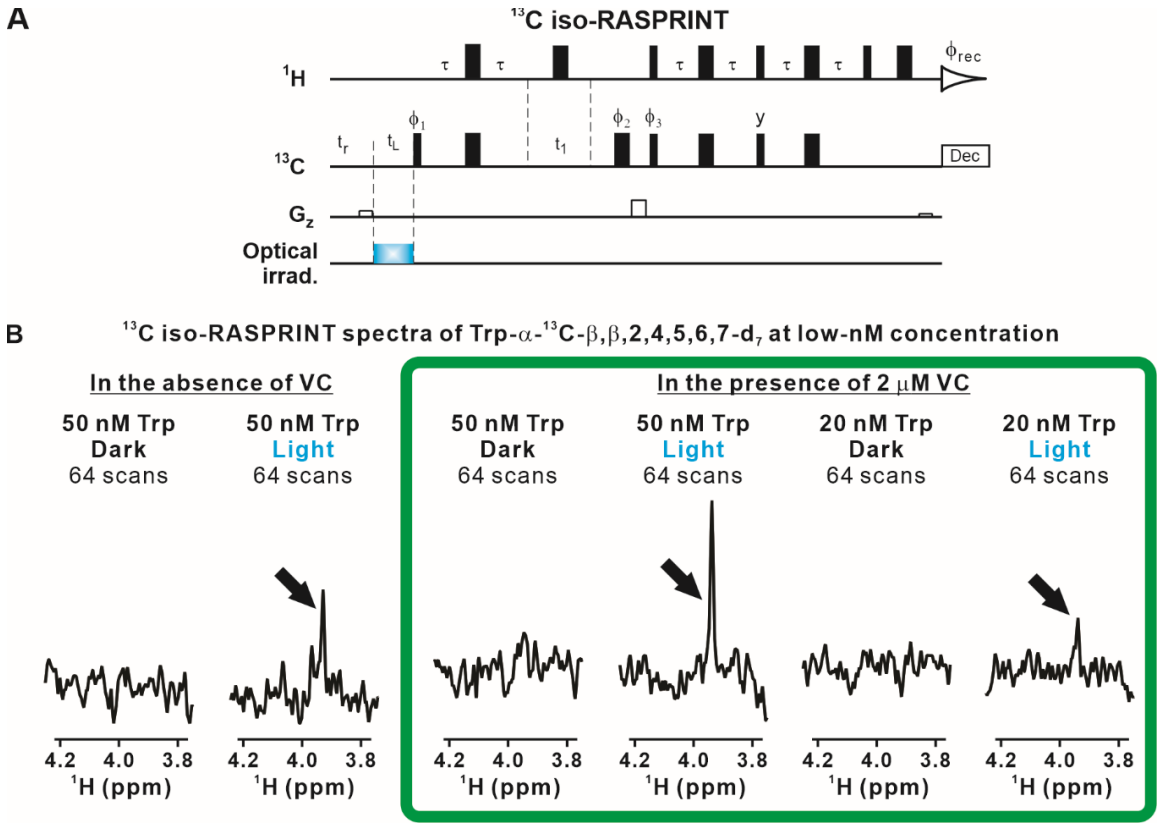


Figure 5. LC-photo-CIDNP in combination with selective isotope labeling enables the detection of Trp at low-nanomolar concentration. A. Scheme illustrating the 13C iso-RASPRINT pulse sequence, optimized for the detection of the Trp-a-13C-b,b,2,4,5,6,7-d7 isotopolog. The 13C iso-RASPRINT sequence differs from the 13C RASPRINT sequence in that on-resonance selective pulses are replaced by hard pulses and off-resonance selective pulses are removed. In addition, the constant-time evolution was removed, yet a short spin-echo was preserved to compensate for the effect of pulse field gradients. B. LC-photo-CIDNP (1D 13C iso-RASPRINT) spectra of 50 and 20 nM Trp-a-13C-b,b,2,4,5,6,7-d7. Data were acquired under dark and light conditions, in the absence and presence of 2 mM vitamin C (VC).. In each spectrum, signal-averaging was carried out across 64-scans preceded by 4 dummy scans. The optical irradiation time was 0.2 s. Data were collected at 24 °C. The spectra enclosed within the green frame show the highest degree of nuclear-spin hyperpolarization.

3.3.5 Quantitative assessment of LC-photo-CIDNP enhancements and polarization values. To more quantitatively evaluate the extent of polarization enhancement generated via LC-photo-CIDNP for each of the isotopologs analyzed in this work, we compared resonance areas under dark and light conditions according to the relation below

$$\varepsilon = \frac{Area_{light}}{Area_{dark}} \frac{[Trp]_{dark}}{[Trp]_{light}} \quad , \qquad [3]$$

where e is the polarization enhancement factor, defined as the ratio between nuclear spin polarization under light and dark conditions. Note that we acquired the dark (LED-off) spectra at much higher concentrations (ca. 100 mM) than the LED-on experiments, and then adjusted for concentration differences according to equation [3]. In addition, we employed long (5 s) recycle delays to ensure nearly complete relaxation between scans. We collected the same number of scans under dark and light conditions. Clearly, if the shorter delays employed in the light experiments had been employed for data collection under dark conditions, the apparent polarization enhancement factors would have been much larger. We refrained from doing this so that we could provide the highest possible reference S/N values under dark conditions.

From percent polarization enhancements e and from thermal-equilibrium percent polarization $P_{\%,te}$, assessed from the Boltzmann distribution in the presence of a 600 MHz applied field [151], we also determined the percent polarization $P_{\%}$ according to

$$P_{\%} = \varepsilon P_{\%,te} \quad . \qquad [4]$$

The resulting e and $P_{\%}$ values of all three isotopologs examined in this work are shown in Figure 6.

The enhancement factor and percent polarization of the two best isotopologs are discussed next. As shown in Figure 6, the e and $P_{\%}$ values of Trp-a-^{13}C are 410 ± 2 and 0.496 ± 0.003%, respectively. The Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 displays the highest values; namely e and $P_{\%}$ of 470 ± 13 and 0.568 ± 0.016%. The Trp-a-^{13}C and Trp-a-^{13}C-b,b,2,4,5,6,7-d_7 isotopologs performed considerably better than uniformly labeled Trp (Trp-U-^{13}C,^{15}N), with increases of 38% and 57%, respectively

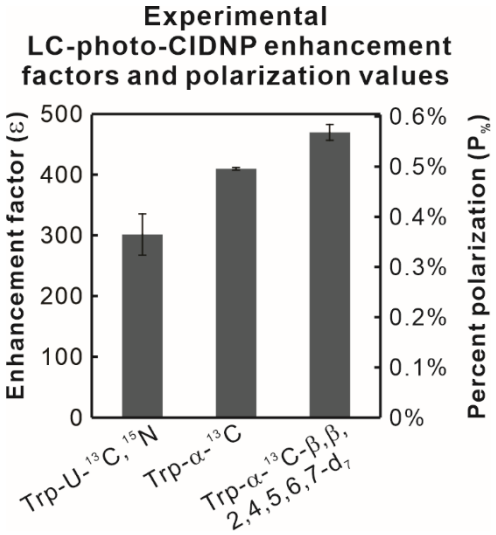


Figure 6. **The Trp-a-13C-b,b,2,4,5,6,7-d7 isotopolog displays unprecedented LC-photo-CIDNP enhancement factors e and percent polarization in solution.** Enhancement factors (e) were determined from experimental data according to equation [3]. Percent polarization values were assessed from enhancement factors and thermal-equilibrium percent polarizations according to equation [4]. Thermal-equilibrium percent polarizations were estimated at room temperature at 600 MHz applied field. Sample concentrations were 1 mM and 100 mM in LED-on and LED-off experiments, respectively. All values are shown as avg ± S.E., with n=2, where n denotes the number of independent experiments.

3.3.6 The observed sensitivity enhancements arise from both LED-on and LED-off effects. The data of Figure 4B, collected under dark (LED-off) conditions, show that selective ^{13}C labeling as well as introduction of deuteration at C□ and across the aromatic ring induces some sensitivity enhancement even in the absence of optical irradiation. Therefore, some non-photo-CIDNP-related factors must contribute to the overall observed effects.

We were able to identify two specific dark-state contributions. First, there was an increase in 1Ha T2 arising from either the removal of some nearby NMR-active nuclei and(or) from their replacement with low-gyromagnetic-ratio (low-g) nuclei due to selective labeling. The results of 1Ha T2 measurements are presented in the Supplementary Information. Clearly, this effect accounts for the sharper resonances, and longer 1Ha T2 displayed by the Trp-a-13C-b,b,2,4,5,6,7-d7 isotopolog even under dark conditions. Second, the elimination of many 13C-13C, multi-bond 1H-13C and 1H-1H J couplings arising from selective labeling prevents loss of coherence via J coupling during the pulse sequence. This effect was quantified upon comparing the area/concentration ratios for the 1Ha resonance of each isotopolog in LC-photo-CIDNP measurements carried out under dark conditions.

In addition, there were also contributions due to LC-photo-CIDNP hyperpolarization effects under light (LED-on) conditions expected from the theoretical arguments discussed in the previous sections. The contributions arising from all three factors (linewidth reduction, J-coupling effects and LC-photo-CIDNP hyperpolarization) were quantified as described in the

Supplementary Text and Supplementary Figure S4. The results are summarized in Table 1 for each of the isotopologs examined in this work. For easier-evaluation purposes, the data were normalized relative to Trp-U-$^{13}C,^{15}N$. The product of all three separately-assessed contributions agrees well with the overall enhancement factor in sensitivity. Overall, it is intriguing to note that the sensitivity enhancements described in this work were achieved by a combination of “dark and light effects”.

Trp isotopolog	Relative sensitivity partitioned into individual contributions			Product of individual contributions [d]	Overall relative sensitivity (assessed via ^{13}C RASPRINT, light cond.) [e]
	LC-photo-CIDNP hyperpolarization [a] (light effect)	Elimination of J-coupling [b] (dark effect)	Linewidth reduction [c] (dark effect)		
Trp-U-^{13}C,^{15}N	100%	100%	100%	100%	100%
Trp-α-^{13}C	138±4%	119%	126 ± 2%	207 ± 7%	172 ± 20%
Trp-α-^{13}C-β,β, 2,4,5,6,7-d_7	157±4%	142%	190 ± 5%	424 ± 16%	397 ± 9%

Table 1. Relative sensitivity of LC-photo-CIDNP NMR experiments (1D ^{13}C RASPRINT pulse sequence) carried out on different Trp isotopologs. The role of individual light (LED-on) and dark (LED-off) contributions is listed. Uniformly labeled Trp (Trp- U-^{13}C,^{15}N) is regarded as reference compound.

a. Determined from data in Fig. 6, including assessment of errors.

b. Determined from data in Fig. 4B. Only one experiment was carried out.

c. Assessed from T_2 experiments in Supplementary Fig. S2. Values are shown as avg ± S.E., with n=3, where n denotes the number of independent experiments.

d. Errors were assessed by error propagation of the individual contributions.

e. Three ^{13}C RASPRINT experiments were carried out for each isotopolog under LED-on conditions. Relative sensitivities were assessed from resonance intensities of experiments run with identical parameters, including number of scans and duration. Relative sensitivities are shown as avg ± S.E.

3.3.7 LC-photo-CIDNP is effective in complex biologically relevant media. In addition to the simple environments discussed so far, including only few purified components, LC-photo-CIDNP can also be applied to samples within extremely complex biological milieux. To demonstrate this capability in the case of the LC-photo-CIDNP technology presented here, we prepared an E. coli S30 bacterial cell extract according to known procedures [208]. This extract includes most E. coli soluble proteins. The composition of the corresponding NMR sample is schematically illustrated in Figure 7A and the procedure followed to generate the cell extract is outlined in Figure 7B. Importantly, this complex biological sample included 1 mM Trp-a-13C-b,b,2,4,5,6,7-d7. Figure 7C illustrates the fact that no signal was observable for this extremely dilute Trp isotopolog under dark conditions. In contrast, 1 mM Trp-a-13C-b,b,2,4,5,6,7-d7 became readily detected after only 8 scans under light conditions. Remarkably, due to the nearly absent recycle delay of 13C iso-RASPRINT under light conditions, the entire data collection, including the hyperpolarization time, only took a total of ca. 4 s (excluding the time for the 4 dummy scans, i.e., 2 additional seconds).

3.3.8 Comparisons with other hyperpolarization approaches in liquids. As a comparison, two representative examples from other leading nuclear-spin hyperpolarization approaches will be listed next. The well-known dissolution DNP (dDNP) technology enabled detection of 20 mM protein samples [209]. This technology has also been used in the context of

protein-ligand interactions [156, 210-215]. In addition, Kouřil et al. [216] recently showed that 32% 13C polarization was rendered possible by bullet-DNP, an interesting variation of dDNP employing small sample-dilution factors during the dissolution process. In both the above studies, considerably higher sample concentrations were required than in the present work. Moreover, unlike LC-photo-CIDNP, the dDNP approach requires a freeze-thaw process that is not always desirable in the case of biomolecules.

Very recent studies showed that parahydrogen-induced polarization (PHIP), especially via the SABRE approach, can be employed for the detection of amino acids in solution [168, 217-218]. While this methodology is promising, identification of suitable amino-acid precursors for SABRE remains difficult, and extension of this technique to proteins has not been feasible.

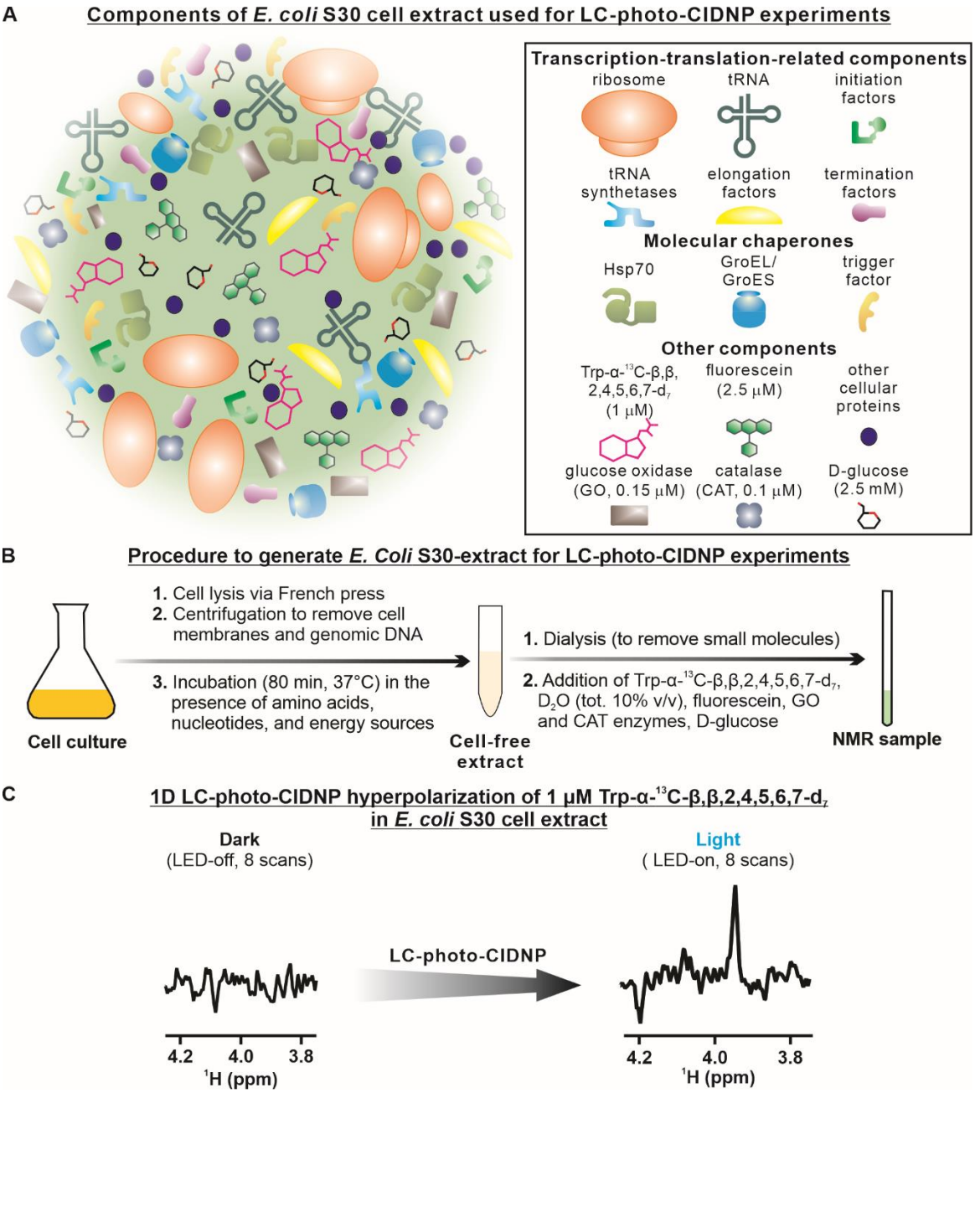

A
Components of E. coli S30 cell extract used for LC-photo-CIDNP experiments
Transcription-translation-related components
ribosome
tRNA
initiation factors
tRNA synthetases
elongation factors
termination factors
Molecular chaperones
Hsp70
GroEL/ GroES
trigger factor
Other components
Trp-α-13C-β,β, 2,4,5,6,7-d7 (1 μM)
fluorescein (2.5 μM)
other cellular proteins
glucose oxidase (GO, 0.15 μM)
catalase (CAT, 0.1 μM)
D-glucose (2.5 mM)
B
Procedure to generate E. Coli S30-extract for LC-photo-CIDNP experiments
1. Cell lysis via French press
2. Centrifugation to remove cell membranes and genomic DNA
3. Incubation (80 min, 37°C) in the presence of amino acids, nucleotides, and energy sources
Cell culture
Cell-free extract
1. Dialysis (to remove small molecules)
2. Addition of Trp-α-13C-β,β,2,4,5,6,7-d7, D2O (tot. 10% v/v), fluorescein, GO and CAT enzymes, D-glucose
NMR sample
C
1D LC-photo-CIDNP hyperpolarization of 1 μM Trp-α-13C-β,β,2,4,5,6,7-d7 in E. coli S30 cell extract
Dark
(LED-off, 8 scans)
Light
(LED-on, 8 scans)
LC-photo-CIDNP
4.2
4.0
3.8
1H (ppm)
4.2
4.0
3.8
1H (ppm)

Figure 7. LC-photo-CIDNP in combination with selective isotope labeling enables hypersensitive data collection in a complex cell-like medium. A. Schematic illustration of the components of the bacterial cell-like medium, i.e., an E. Coli S30 extract, employed in this work. B. Summary of the adopted procedures to generate the E. Coli S30 extract. C. LC-photo-CIDNP spectrum of 1 mM Trp-a-13C-b,b,2,4,5,6,7-d7 in an E. coli S30 cell extract under dark (i.e., LED-off) and light (i.e., LED-on) conditions. Each spectrum was acquired at 24 °C with 8 scans, 4 dummy scans, and 0.2 s of optical irradiation.

3.4. Conclusions and Outlook

In this study we introduce a combination of LC-photo-CIDNP and selective isotope labeling to enable the NMR detection of low-nM biomolecular concentrations by liquid-state NMR. The use of selective isotope labeling enables an additional ca. 4-fold increase in sensitivity relative to a uniformly isotopically enriched reference molecule (Trp-U-13C,15N). This gain in sensitivity was crucial to access the low-nanomolar (20 nM) concentration range, providing a remarkable detection limit of only 3 ng in a few seconds. To the best of our knowledge, this is the lowest concentration achieved by high-resolution biomolecular NMR in liquids to date. Our methodology works well even in extremely complex physiologically relevant environments, including an E. coli S30 cell extract. The capability to assess Trp levels [219-221], metabolism and transport in biological media is important because of the emerging role of this amino acid and its metabolites in health and disease [222-228].

Importantly, the LC-photo-CIDNP technology, including the advances presented here, is also readily amenable to extensions to Trp-containing macromolecules. For instance, LC-photo-CIDNP of polypeptides and proteins has already been established [188, 191, 193], and it yields comparable enhancements to free Trp. Further, given that Trp analogs can be routinely incorporated into proteins via auxotroph strains [229], one can envisage that the same strains can also be employed to incorporate Trp isotopologs like Trp-a-13C-b,b,2,4,5,6,7-d7 into proteins. Ultrasensitive detection of Ha-Ca pairs in proteins is highly desirable because it

provides access to site-specific information on backbone secondary structure [198]. Further, the 13C iso-RASPRINT pulse sequence can readily be run in 2D mode to resolve overlapping resonances within mixtures of aromatic amino acids or in the context of large biomolecules.

In all, the advances described in this work pave the way to a variety of biomolecular NMR applications requiring high sensitivity and fast data collection. It will be exciting to follow the forthcoming progress of this technology in the years to come.

3.5. References

62. Lee, J. H.; Okuno, Y.; Cavagnero, S., Sensitivity enhancement in solution NMR: emerging ideas and new frontiers. *J. Magn. Reson.* 2014, *241*, 18-31.

63. Ardenkjaer-Larsen, J. H.; Boebinger, G. S.; Comment, A.; Duckett, S.; Edison, A. S.; Engelke, F.; Griesinger, C.; Griffin, R. G.; Hilty, C.; Maeda, H.; Parigi, G.; Prisner, T.; Ravera, E.; van Bentum, J.; Vega, S.; Webb, A.; Luchinat, C.; Schwalbe, H.; Frydman, L., Facing and overcoming sensitivity challenges in biomolecular NMR spectroscopy. *Angew. Chem. Int. Ed.* **2015,** *54* (32), 9162-9185.

64. Schops, P.; Spindler, P. E.; Prisner, T. F., Multi-frequency pulsed Overhauser DNP at 1.2 Tesla. *Zeit. Phys. Chem. Intl. J. Res. Phys. Chem. Chem. Phys.* **2017,** *231* (3), 561-573.

65. Denysenkov, V. P.; Prisner, T. F., Liquid-state Overhauser DNP at high magnetic fields. *Emagres* **2019,** *8* (1), 41-54.

66. Bennati, M.; Orlando, T., Overhauser DNP in liquids on ^{13}C nuclei. *Emagres* **2019,** *8* (1), 11-18.

67. Kim, Y.; Hilty, C., Applications of dissolution-DNP for NMR screening. In *Biological NMR, Pt B*, Wand, A. J., Ed. 2019; Vol. 615, pp 501-526.

68. Kagawa, A.; Miyanishi, K.; Ichijo, N.; Negoro, M.; Nakamura, Y.; Enozawa, H.; Murata, T.; Morita, Y.; Kitagawa, M., High-field NMR with dissolution triplet-DNP. *J. Magn. Reson.* **2019,** *309*, 106623.

69. Jahnig, F.; Himmler, A.; Kwiatkowski, G.; Dapp, A.; Hunkele, A.; Kozerke, S.; Ernst, M., A spin-thermodynamic approach to characterize spin dynamics in TEMPO-based samples for dissolution DNP at 7 T field. *J. Magn. Reson.* **2019,** *303*, 91-104.

70. Bowers, C. R.; Weitekamp, D. P., Transformation of symmetrization order to nuclear-spin magnetization by chemical reaction and nuclear magnetic resonance. *Phys. Rev. Lett.* **1986,** *57* (21), 2645-2648.

71. Bowers, C. R.; Weitekamp, D. P., Parahydrogen and synthesis allow dramatically enhanced nuclear alignment. *J. Am. Chem. Soc.* **1987,** *109* (18), 5541-5542.

72. Carson, P. J.; Bowers, C. R.; Weitekamp, D. P., The PASADENA effect at a solid surface: High-sensitivity nuclear magnetic resonance of hydrogen chemisorption. *J. Am. Chem. Soc.* **2001,** *123* (47), 11821-11822.

73. Pravica, M. G.; Weitekamp, D. P., Net NMR alignment by adiabatic transport of parahydrogen addition products to high magnetic field. *Chem. Phys. Lett.* **1988,** *145* (4), 255-258.

74. Wang, W. Y.; Xu, J.; Deng, F. W., Q.; Chu, Y. Y.; Qi, G. D.; Li, S. H., Pairwise stereoselective hydrogenation of propyne on supported Pd-Ag catalysts investigated by parahydrogen-induced polarization. *J. Phys. Chem. C* **2021,** *125* (31), 17144-17154.

75. Salnikov, O. G.; Chukanov, N. V.; Kovtunova, L. M.; Bukhtiyarov, V. I.; Kovtunov, K. V.; Shchepin, R. V.; Koptyug, I. V.; Chekmenev, E. Y., Heterogeneous ^{1}H and ^{13}C parahydrogen-Induced polarization of acetate and pyruvate esters. *ChemPhysChem* **2021,** *22* (13), 1389-1396.
76. Pokochueva, E. V.; Burueva, D. B.; Salnikov, O. G.; Koptyug, I. V., Heterogeneous catalysis and parahydrogen-induced polarization. *ChemPhysChem* **2021,** *22* (14), 1421-1440.
77. Adams, R. W.; Aguilar, J. A.; Atkinson, K. D.; Cowley, M. J.; Elliott, P. I. P.; Duckett, S. B.; Green, G. G. R.; Khazal, I. G.; Lopez-Serrano, J.; Williamson, D. C., Reversible interactions with para-hydrogen enhance NMR sensitivity by polarization transfer. *Science* **2009,** *323* (5922), 1708-1711.
78. Duckett, S. B.; Mewis, R. E., Application of Parahydrogen Induced Polarization Techniques in NMR Spectroscopy and Imaging. *Acc. Chem. Res.* **2012,** *45* (8), 1247-1257.
79. Pravdivtsev, A. N.; Buntkowsky, G.; Duckett, S. B.; Koptyug, I. V.; Hovener, J. B., Parahydrogen-induced polarization of amino acids. *Angew. Chem.* **2021,** *60* (44), 23496-23507.
80. Taylor, D. A.; Natrajan, L. S.; Nilsson, M.; Adams, R. W., SABRE-enhanced real-time pure shift NMR spectroscopy. *Magn. Reson. Chem.* **2021,** *59*, 1244-1252.
81. Spiridonov, K. A.; Kozinenko, V. P.; Nikovsky, I. A.; Pavlov, A. A.; Vol'khina, T. N.; Nelyubina, Y. V.; Kiryutin, A. S.; Yurkovskaya, A. V.; Polezhaev, A. A.; Novikov, V. V.; Ivanov, K. L., Phosphite-containing iridium polarization transfer catalysts for NMR signal amplification by reversible exchange. *Mendeleev Commun.* **2021,** *31* (4), 475-477.
82. Kiryutin, A. S.; Yurkovskaya, A. V.; Petrov, P. A.; Ivanov, K. L., Simultaneous ^{15}N polarization of several biocompatible substrates in ethanol-water mixtures by signal amplification by reversible exchange (SABRE) method. *Magn. Reson. Chem.* **2021,** *59*, 1216-1224.
83. Barskiy, D. A.; Kovtunov, K. V.; Koptyug, I. V.; He, P.; Groome, K. A.; Best, Q. A.; Shi, F.; Goodson, B. M.; Shchepin, R. V.; Coffey, A. M.; Waddell, K. W.; Chekmenev, E. Y., The Feasibility of Formation and Kinetics of NMR Signal Amplification by Reversible Exchange (SABRE) at High Magnetic Field (9.4 T). *J. Am. Chem. Soc.* **2014,** *136* (9), 3322-3325.
84. Barskiy, D. A.; Kovtunov, K. V.; Koptyug, I. V.; He, P.; Groome, K. A.; Best, Q. A.; Shi, F.; Goodson, B. M.; Shchepin, R. V.; Truong, M. L.; Coffey, A. M.; Waddell, K. W.; Chekmenev, E. Y., In Situ and Ex Situ Low-Field NMR Spectroscopy and MRI Endowed by SABRE Hyperpolarization. *ChemPhysChem* **2014,** *15* (18), 4100-4107.
85. Feng, B. B.; Coffey, A. M.; Colon, R. D.; Chekmenev, E. Y.; Waddell, K. W., A pulsed injection parahydrogen generator and techniques for quantifying enrichment. *J. Magn. Reson.* **2012,** *214*, 258-262.

86. Shchepin, R. V.; Barskiy, D. A.; Coffey, A. M.; Theis, T.; Shi, F.; Warren, W. S.; Goodson, B. M.; Chekmenev, E. Y., N-15 Hyperpolarization of Imidazole-N-15(2) for Magnetic Resonance pH Sensing via SABRE-SHEATH. *Acs Sensors* **2016,** *1* (6), 640-644.
87. Shchepin, R. V.; Truong, M. L.; Theis, T.; Coffey, A. M.; Shi, F.; Waddell, K. W.; Warren, W. S.; Goodson, B. M.; Chekmenev, E. Y., Hyperpolarization of "Neat" Liquids by NMR Signal Amplification by Reversible Exchange. *Journal of Physical Chemistry Letters* **2015,** *6* (10), 1961-1967.
88. Shi, F.; Coffey, A. M.; Waddell, K. W.; Chekmenev, E. Y.; Goodson, B. M., Heterogeneous Solution NMR Signal Amplification by Reversible Exchange. *Angewandte Chemie-International Edition* **2014,** *53* (29), 7495-7498.
89. Shi, F.; Coffey, A. M.; Waddell, K. W.; Chekmenev, E. Y.; Goodson, B. M., Nanoscale Catalysts for NMR Signal Enhancement by Reversible Exchange. *J. Phys. Chem. C* **2015,** *119* (13), 7525-7533.
90. Shi, F.; He, P.; Best, Q. A.; Groome, K.; Truong, M. L.; Coffey, A. M.; Zimay, G.; Shchepin, R. V.; Waddell, K. W.; Chekmenev, E. Y.; Goodson, B. M., Aqueous NMR Signal Enhancement by Reversible Exchange in a Single Step Using Water-Soluble Catalysts. *J. Phys. Chem. C* **2016,** *120* (22), 12149-12156.
91. Theis, T.; Truong, M.; Coffey, A. M.; Chekmenev, E. Y.; Warren, W. S., LIGHT-SABRE enables efficient in-magnet catalytic hyperpolarization. *J. Magn. Reson.* **2014,** *248*, 23-26.
92. Theis, T.; Truong, M. L.; Coffey, A. M.; Shchepin, R. V.; Waddell, K. W.; Shi, F.; Goodson, B. M.; Warren, W. S.; Chekmenev, E. Y., Microtesla SABRE Enables 10% Nitrogen-15 Nuclear Spin Polarization. *J. Am. Chem. Soc.* **2015,** *137* (4), 1404-1407.
93. Truong, M. L.; Shi, F.; He, P.; Yuan, B. X.; Plunkett, K. N.; Coffey, A. M.; Shchepin, R. V.; Barskiy, D. A.; Kovtunov, K. V.; Koptyug, I. V.; Waddell, K. W.; Goodson, B. M.; Chekmenev, E. Y., Irreversible Catalyst Activation Enables Hyperpolarization and Water Solubility for NMR Signal Amplification by Reversible Exchange. *J. Phys. Chem. B* **2014,** *118* (48), 13882-13889.
94. Truong, M. L.; Theis, T.; Coffey, A. M.; Shchepin, R. V.; Waddell, K. W.; Shi, F.; Goodson, B. M.; Warren, W. S.; Chekmenev, E. Y., N-15 Hyperpolarization by Reversible Exchange Using SABRE-SHEATH. *J. Phys. Chem. C* **2015,** *119* (16), 8786-8797.
95. Bargon, J.; Fischer, H.; Johnsen, U., Kernresonanz-emissionslinien während rascher radikalreaktionen. *Zeitschrift für Naturforschung A* **1967,** *22* (10), 1551-1555.
96. Ward, H. R.; Lawler, R. G., Nuclear magnetic resonance emission and enhanced absorption in rapid organometallic reactions. *J. Am. Chem. Soc.* **1967,** *89* (21), 5518-5519.
97. Kaptein, R.; Dijkstra, K.; Nicolay, K., Laser photo-CIDNP as a surface probe for proteins in solution. *Nature* **1978,** *274* (5668), 293-4.

98. Lee, J. H.; Cavagnero, S., A novel tri-enzyme system in combination with laser-driven NMR enables efficient nuclear polarization of biomolecules in solution. *J. Phys. Chem. B* **2013,** *117* (20), 6069-6081.
99. Lee, J. H.; Sekhar, A.; Cavagnero, S., ^{1}H-Detected ^{13}C photo-CIDNP as a sensitivity enhancement tool in solution NMR. *J. Am. Chem. Soc.* **2011,** *133* (21), 8062-5.
100. Lyon, C. E.; Jones, J. A.; Redfield, C.; Dobson, C. M.; Hore, P. J., Two-dimensional ^{15}N-^{1}H photo-CIDNP as a surface probe of native and partially structured proteins. *J. Am. Chem. Soc.* **1999,** *121* (27), 6505-6506.
101. Okuno, Y.; Cavagnero, S., Fluorescein: a photo-CIDNP sensitizer enabling hypersensitive NMR data collection in liquids at low micromolar concentration. *J. Phys. Chem. B* **2016,** *120* (4), 715-23.
102. Okuno, Y.; Mecha, M. F.; Yang, H.; Zhu, L.; Fry, C. G.; Cavagnero, S., Laser- and cryogenic probe-assisted NMR enables hypersensitive analysis of biomolecules at submicromolar concentration. *Proc. Natl. Acad. Sci. USA* **2019,** *116* (24), 11602-11611.
103. Yang, H.; Hofstetter, H.; Cavagnero, S., Fast-pulsing LED-enhanced NMR: A convenient and inexpensive approach to increase NMR sensitivity. *J. Chem. Phys.* **2019,** *151* (24), 245102.
104. Yang, H.; Mecha, M. F.; Goebel, C. P.; Cavagnero, S., Enhanced nuclear-spin hyperpolarization of amino acids and proteins via reductive radical quenchers. *J. Magn. Reson.* **2021,** *324*, 106912.
105. Kaptein, R.; Dijkstra, K.; Müller, F.; Van Schagen, C. G.; Visser, A. J. W. G., 360-MHz laser-induced phot-CIDNP in photoreaction of flavins. *J. Magn. Reson.* **1978,** *31* (1), 171-176.
106. Hore, P. J.; Broadhurst, R. W., Photo-CIDNP of biopolymers. *Prog. Nucl. Magn. Reson. Spectrosc.* **1993,** *25*, 345-402.
107. Lindqvist, L., A flash photolysis study of fluorescein. *Arkiv for Kemi* **1961,** *16* (1-2), 79-138.
108. Kiryutin, A. S.; Morozova, O. B.; Kuhn, L. T.; Yurkovskaya, A. V.; Hore, P. J., ^{1}H and ^{13}C Hyperfine coupling constants of the tryptophanyl cation radical in aqueous solution from microsecond time-resolved CIDNP. *J. Phys. Chem. B* **2007,** *111* (38), 11221-11227.
109. Spera, S.; Bax, A., Empirical correlation between protein backbone conformation and C^{a}. and C^{b}. ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **1991,** *113* (14), 5490-5492.
110. Adrian, F. J., Singlet-triplet splitting in diffusing radical pairs and magnitude of chemically induced electron-spin polarization. *J. Chem. Phys.* **1972,** *57* (12), 5107.
111. Ivanov, K. L.; Pravdivtsev, A. N.; Yurkovskaya, A. V.; Vieth, H. M.; Kaptein, R., The role of level anti-crossings in nuclear spin hyperpolarization. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014,** *81*, 1-36.

112. Lyon, C. E.; Lopez, J. J.; Cho, B. M.; Hore, P. J., Low field CIDNP of amino acids and proteins: characterization of transient radicals and NMR sensitivity enhancement. *Mol. Phys.* **2002,** *100* (8), 1261-1269.
113. Okuno, Y.; Cavagnero, S., Photochemically induced dynamic nuclear polarization: basic principles and applications. *eMagRes* **2017,** *6* (2), 283-313.
114. Janssen, G. J.; Daviso, E.; van Son, M.; de Groot, H. J. M.; Alia, A.; Matysik, J., Observation of the solid-state photo-CIDNP effect in entire cells of cyanobacteria synechocystis. *Photosynth. Res.* **2010,** *104* (2), 275-282.
115. Eisenreich, W.; Joshi, M.; Illarionov, B.; Kacprzak, S.; Lukaschek, M.; Kothe, G.; Budisa, N.; Fischer, M.; Bacher, A.; Weber, S., Strategy for enhancement of ^{13}C-photo-CIDNP NMR spectra by exploiting fractional ^{13}C-labeling of tryptophan. *J. Phys. Chem. B* **2015,** *119* (43), 13934-13943.
116. Thompson, C. M.; McDonald, A. D.; Yang, H.; Cavagnero, S.; Buller, A. R., Modular control of L-tryptophan isotopic substitution via an efficient biosynthetic cascade. *Org. Biomol. Chem.* **2020,** *18* (22), 4189-4192.
117. Ernst, R. R.; Bodenhausen, G.; Wokaun, A., *Principles of nuclear magnetic resonance in one and two dimensions*. Oxford University Press: New York, 1989.
118. Cavanagh, J.; Fairbrother, W.; Palmer, A. I.; M, R.; NJ, S., *Protein NMR Spectroscopy Principles and Practice*. 2nd ed.; Elsevier Academic Press: 2007.
119. Bakke, C. K.; Jungbauer, L. M.; Cavagnero, S., *In vitro* expression and characterization of native apomyoglobin under low molecular crowding conditions. *Prot. Eng. Expr. Purif.* **2006,** *45* (2), 381-92.
120. Ragavan, M.; Iconaru, L. I.; Park, C.-G.; Kriwacki, R. W.; Hilty, C., Real-time analysis of folding upon binding of a disordered protein by using dissolution DNP NMR spectroscopy. *Angew. Chem.* **2017,** *56* (25), 7070-7073.
121. Kim, J.; Mandal, R.; Hilty, C., Characterization of Membrane Protein-Lipid Interactions in Unfolded OmpX with Enhanced Time Resolution by Hyperpolarized NMR. *Chembiochem* **2020,** *21* (19), 2861-2867.
122. Kim, Y.; Liu, M. X.; Hilty, C., Parallelized Ligand Screening Using Dissolution Dynamic Nuclear Polarization. *Anal. Chem.* **2016,** *88* (22), 11178-11183.
123. Kim, Y.; Liu, M. X.; Hilty, C., Determination of binding affinities using hyperpolarized NMR with simultaneous 4-channel detection. *J. Magn. Reson.* **2018,** *295*, 80-86.
124. Kim, Y.; Wang, Y. Y.; Chen, H. Y.; Hilty, C., In-Vitro Dissolution Dynamic Nuclear Polarization for Sensitivity Enhancement of NMR with Biological Molecules. In *Protein NMR: Methods and Protocols*, Ghose, R., Ed. 2018; Vol. 1688, pp 155-168.
125. Liu, M. X.; Kim, Y.; Hilty, C., Characterization of Chemical Exchange Using Relaxation Dispersion of Hyperpolarized Nuclear Spins. *Anal. Chem.* **2017,** *89* (17), 9154-9158.

126. Wang, Y. Y.; Kim, J.; Hilty, C., Determination of protein-ligand binding modes using fast multi-dimensional NMR with hyperpolarization. *Chemical Science* **2020,** *11* (23), 5935-5943.
127. Kouřil, K.; Kouřilová, H.; Bartram, S.; Levitt, M. H.; Meier, B., Scalable dissolution-dynamic nuclear polarization with rapid transfer of a polarized solid. *Nature Comm.* **2019,** *10* (1), 1733.
128. Kaltschnee, L.; Jagtap, A. P.; McCormick, J.; Wagner, S.; Bouchard, L. S.; Utz, M.; Griesinger, C.; Glggler, S., Hyperpolarization of amino acids in water utilizing parahydrogen on a rhodium nanocatalyst. *Chem. Eur. J.* **2019,** *25* (47), 11031-11035.
129. Sellies, L.; Aspers, R.; Feiters, M. C.; Rutjes, F.; Tessari, M., Parahydrogen hyperpolarization allows direct NMR detection of alpha-amino acids in complex (bio)mixtures. *Angew. Chem.* **2021,** *60* (52), 26954-26959.
130. Kositz, C.; Schroecksnadel, K.; Grander, G.; Schennach, H.; Kofler, H.; Fuchs, D., High serum tryptophan concentration in pollinosis patients is associated with unresponsiveness to pollen extract therapy. *Int. Arch. All. Immun.* **2008,** *147* (1), 35-40.
131. Menta, R.; Mancheno-Corvo, P.; Del Rio, B.; Ramirez, C.; DelaRosa, O.; Dalemans, W.; Lombardo, E., Tryptophan concentration is the main mediator of the capacity of adipose mesenchymal stromal cells to inhibit T-lymphocyte proliferation in vitro. *Cytother.* **2014,** *16* (12), 1679-1691.
132. Weinlich, G.; Murr, C.; Richardsen, L.; Winkler, C.; Fuchs, D., Decreased serum tryptophan concentration predicts poor prognosis in malignant melanoma patients. *Dermatol.* **2007,** *214* (1), 8-14.
133. Agus, A.; Planchais, J.; Sokol, H., Gut Microbiota Regulation of Tryptophan Metabolism in Health and Disease. *Cell Host & Microbe* **2018,** *23* (6), 716-724.
134. Bosi, A.; Banfi, D.; Bistoletti, M.; Giaroni, C.; Baj, A., Tryptophan Metabolites Along the Microbiota-Gut-Brain Axis: An Interkingdom Communication System Influencing the Gut in Health and Disease. *Int. J. Tryp. Res.* **2020,** *13*, 1-25.
135. Gao, K.; Mu, C. L.; Farzi, A.; Zhu, W. Y., Tryptophan Metabolism: A Link Between the Gut Microbiota and Brain. *Adv. Nutr.* **2020,** *11* (3), 709-723.
136. Gostner, J. M.; Geisler, S.; Stonig, M.; Mair, L.; Sperner-Unterweger, B.; Fuchs, D., Tryptophan Metabolism and Related Pathways in Psychoneuroimmunology: The Impact of Nutrition and Lifestyle. *Neuropsychobiol.* **2020,** *79* (1), 89-99.
137. Kokturk, O.; Kanbay, A., Tryptophan Metabolism and Sleep. In *Tryptophan Metabolism: Implications for Biological Processes, Health and Disease*, Engin, A.; Engin, A. B., Eds. 2015; pp 239-252.
138. Konopelski, P.; Ufnal, M., Indoles - Gut Bacteria Metabolites of Tryptophan with Pharmacotherapeutic Potential. *Curr. Drug Met.* **2018,** *19* (10), 883-890.

139. Roth, W.; Zadeh, K.; Vekariya, R.; Ge, Y.; Mohamadzadeh, M., Tryptophan Metabolism and Gut-Brain Homeostasis. *International Journal of Molecular Sciences* **2021,** *22* (6), 1877-2013.
140. Broos, J., Biosynthetic incorporation of tryptophan analogs in proteins. In *Fluorescence spectroscopy and microscopy - methods and protocols*, Engelborghs, Y.; Visser, A. J. W. G., Eds. Humana Press: New York, 2014; pp 359-370.
141. Pordes, R.; Petravick, D.; Kramer, B.; Olson, D.; Livny, M.; Roy, A.; Avery, P.; Blackburn, K.; Wenaus, T.; Wuerthwein, F.; Foster, I.; Gardner, R.; Wilde, M.; Blatecky, A.; McGee, J.; Quick, R. In *The open science grid*, 3rd Annual Scientific Discovery through Advanced Computing Conference (SciDAC 2007), Boston, MA, Jun 24-28; Boston, MA, 2007.
142. Sfiligoi, I.; Bradley, D. C.; Holzman, B.; Mhashilkar, P.; Padhi, S.; Wurthwein, F., The Pilot Way to Grid Resources Using glideinWMS. *WRI World Congress on Computer Science and Information Engineering* **2009,** *2*, 428-432.

3.6. Associated Content

Supporting Information is available free of charge at [URL TBD]. Description of theoretical predictions on photo-CIDNP polarization, additional experimental details on NMR transvers relaxation measurements, description of assessment of LED-on and LED-off effects used to generate Table 1, experimental section.

3.7. Author Information

3.7.1. *Corresponding Author*

Silvia Cavagnero - Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA.

3.7.2. *Other Authors*

Hanming Yang – Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA; Present Address: Affinity Research Chemicals, 406 Meco Dr., Wilmington, Deleware, 19804, USA.

Clayton A. Mickles – Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA.

Valeria Guzman-Luna – Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA.

Kenji Sugisaki – Department of Chemistry and Molecular Materials Science, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka, 558-8585, Japan.

Siyu Li – Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA.

Clayton M. Thompson – Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA; Present Address: PPD Inc, 8551 Research Way, Middleton, Wisconsin, 53652-4664, USA.

Hung H. Dang – Department of Chemistry, University of Wisconsin - Madison, 1101 University Ave., Madison, Wisconsin, 53706, USA.

3.7.3 Funding Sources

This work was funded by the National Institutes of Health (grants R01GM125995 and S10OD012245), from the National Science Foundation (grant MCB 2124672) and from the University of Wisconsin-Madison (Bridge Funds from the College of Letters & Science).

3.8. Acknowledgements

We thank Alexandra Yurkovskaya and Andrew Buller for insightful discussions, and Greg Sabat for assistance with mass spectrometry. We are grateful to Charles G. Fry, Heike Hofstetter and Cathy Clewett for technical assistance with NMR. Some of this research was performed using the computer resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences [230-231]. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the OSG Consortium, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

Chapter 3

Appendix: Supplementary Information

Development of highly sensitive NMR techniques with isotopic labelling – study protein folding, misfolding, and aggregation

3.9. Supplementary Results

3.9.1 Theoretical prediction of photo-CIDNP polarization arising from geminate recombination. This section focuses on the prediction of photo-CIDNP polarization arising from geminate recombination as a function of the g-factor and hyperfine coupling constants of the radicals of the photosensitizer dye and the molecule of interest. Importantly, the theoretical considerations below will also clarify how the predicted polarization of the molecule of interest depends on the isotope-labeling scheme. As stated in the article text, for freely-diffusing radical pairs, the population difference between two nuclear spin states in the reaction product is

$$p_1 - p_2 = p\left(\sqrt{|\omega_{TS1}|\,\tau_d} - \sqrt{|\omega_{TS2}|\,\tau_d}\right) \quad , \tag{S5}$$

where p_1, p_2 are the population of nuclear spin 1 and 2, p is a normalization factor, ω_{TS} is the triplet-singlet mixing frequency of a given nuclear spin state, and τ_d is the average time the radical-pair component remain within close distance before dissociating. The latter parameter can be estimated via relation

$$\tau_d = \frac{\left(R_D + R_M\right)^2}{D_D + D_M} \quad , \tag{S6}$$

where R_D and R_M are the van der Waals radii of the radicals of the dye and molecule of interest, respectively, and D_D and D_M are the corresponding translational diffusion coefficients.

The triplet-singlet (T_0 – S) mixing frequency ω_{TS} is defined according to the expressions below, where the subscript *0* denotes the nucleus of interest and χ is the spin configuration of the other nuclei

$$\omega_{TS,\alpha,\chi} = \frac{1}{2}\left[\Delta g \mu_B B_0 + \sum_i m_i A_i - \frac{1}{2} A_0 - \sum_j m_j A_j \right] \text{ and} \quad \text{(S7)}$$

$$\omega_{TS,\beta,\chi} = \frac{1}{2}\left[\Delta g \mu_B B_0 + \sum_i m_i A_i + \frac{1}{2} A_0 - \sum_j m_j A_j \right] , \quad \text{(S8)}$$

where $\omega_{TS,\alpha,\chi}$ and $\omega_{TS,\beta,\chi}$ are the triplet-singlet mixing frequencies for the configurations bearing nucleus 0 in either the α and β spin state, respectively. Δg is the difference of g-factor between dye and substrate radicals, μ_B is the Bohr's magneton, B_0 is the applied magnetic field, A_0 is the hyperfine coupling constant of nucleus *0*. In addition, m_i and A_i are the magnetic quantum number and hyperfine coupling constant of the i^{th} nucleus of the dye radical, respectively. Similarly, m_j and A_j are the magnetic quantum number and hyperfine coupling constant of the j^{th} nucleus of the molecule of interest excluding nucleus 0.

According to equation (S1), the population difference between the $\alpha\chi$ and $\beta\chi$ nuclear-spin configurations is

$$p_{\alpha,\chi} - p_{\beta,\chi} = p\left(\sqrt{\left|\omega_{TS,\alpha,\chi}\right| \tau_d} - \sqrt{\left|\omega_{TS,\beta}\right| \tau_d} \right) . \quad \text{(S9)}$$

Now, in order to predict the polarization of nucleus *0* arising from geminate recombination, one needs to sum up all the possible nuclear spin configurations χ according to

$$P_0 = \sum_\chi p_{\alpha,\chi} - \sum_\chi p_{\beta,\chi} = p\sqrt{\tau_d}\left(\sum_\chi \sqrt{\left|\omega_{TS,\alpha,\chi}\right|} - \sum_\chi \sqrt{\left|\omega_{TS,\beta,\chi}\right|} \right) . \quad \text{(S10)}$$

The normalization factor *p* can be determined from the normalization condition according to

$$1 = \sum_{\chi} p_{\alpha,\chi} + \sum_{\chi} p_{\beta,\chi} = p\sqrt{\tau_d}\left(\sum_{\chi}\sqrt{\left|\omega_{\mathrm{TS},\alpha,\chi}\right|} + \sum_{\chi}\sqrt{\left|\omega_{\mathrm{TS},\beta,\chi}\right|}\right) \quad . \tag{S11}$$

The P_0 simulations presented in this work were carried out with a custom-built script in the Python programming language (Python v. 3.8, Python Software Foundation) and yielded P_0 according to equations (S6) and (S7). All calculations were run either on a MacOS laptop computer or at the UW-Madison Center for High-Throughput Computing (CHTC) [230-231]. The following parameters were employed: g-factor of substrate radical g($Trp^{\bullet+}$) = 2.0027 [197], diffusion coefficient of fluorescein dye D_D = 4.2 × 10^{-6} cm^2s^{-1} [232], diffusion coefficient of tryptophan substrate D_M = 6.592 × 10^{-6} cm^2s^{-1} [233], radius of $Fl^{\bullet-}$ R_D = 4.4 Å [234], radius of $Trp^{\bullet+}$ R_M = 4.2 Å [234]. The hyperfine coupling constants of $Fl^{\bullet-}$ and $Trp^{\bullet+}$ were derived from the literature [197, 235]. Additional details are provided in Supplementary Table S1. The hyperfine coupling constant of the Trp amino-group nitrogen was not included in the calculations, given that its value is not available in the literature to the best of our knowledge. We expect this omission not to introduce any appreciable changes to the calculation results, given that we ran a number of control calculations using either ^{15}N, ^{14}N or no N for the indole nitrogen of Trp-α-^{13}C. These control computations yielded variations in predicted polarization values by less than 7.8% (data not shown).

The above methodology enabled us to compute the expected P_0 value of Trp-U-^{13}C. The procedure adopted for the selectively labeled isotopologs Trp-α-^{13}C and Trp-α-^{13}C-

β,β,2,4,5,6,7-d_7 is analogous. In this case, one needs to set the hyperfine coupling constant to 0 when a ^{13}C is replaced by a ^{12}C (ignoring the presence of natural-abundance ^{13}C). In addition, it is necessary to multiply the hyperfine coupling constant by 0.1535 every time a ^{1}H is replaced by a ^{2}H. This multiplicative factor is equal to the ratio of gyromagnetic ratios of deuterium (D) and proton (^{1}H) according to $\gamma_D / \gamma_H = 0.1535$.

3.9.2 Additional considerations on overall LC-photo-CIDNP polarization. In a previous publication [191], we concluded that the overall polarization arising from LC-photo-CIDNP is

$$P_k^{M,SS} = \frac{T_1^M k_{et}^{eff} [{}^{T1}\mathrm{D}^{SS}][\mathrm{M}^{SS}](1+\gamma)\xi^G \Phi_G}{[\mathrm{M}]_0} \quad , \tag{S12}$$

where

$$\gamma = \frac{\xi^F (1-\Phi_G)}{\xi^G \Phi_G} \quad . \tag{S13}$$

In the above expressions, $P_k^{M,SS}$ denotes the steady-state polarization of nucleus *k* of the molecule of interest, T_1^M denotes the longitudinal relaxation time of nucleus *k*, k_{et}^{eff} is the composite rate constant for the bimolecular collisions (followed by electron transfer) of the molecule of interest M and the triplet excited-state dye ^{T1}D. In addition, $[^{T1}D^{SS}]$ and $[M^{SS}]$ denote the steady-state concentration of triplet excited-state dye and non-radical form of the molecule of interest, respectively. Further, γ denotes the ratio between polarization originating from the F-pair and geminate recombination events, and it is defined in equation (S9). Now,

ξ^F and ξ^G denote the normalized probability difference to generate a recombination product in the α and β spin states per recombination event for the F-pair and geminate pair, respectively. Finally, Φ_G denotes the total probability of geminate recombination per geminate radical pair.

The quantity P_0 defined in equation (S6) is defined as the normalized population difference of the α- and β- nuclear-spin states within a geminate radical pair. It is therefore evident that P_0 in equation (S6) is identical to ξ^G in equation (S8).

3.9.3 $^1H^\alpha$ T_2 measurements on Trp isotopologs. In order to quantify the impact of isotope substitution on $^1H^\alpha$ resonance linewidths, we carried out experiments to determine the T_2 relaxation time of different Trp isotopologs. We determined T_2 values for the three isotopologs analyzed in the main article and, in addition, unlabeled Trp, which carries all natural-abundance atoms. The pulse sequence employed for this purpose is shown in Figure S1. Briefly, this sequence features solvent pre-saturation during the recycle delay, a perfect-echo pulse scheme [236] to eliminate contributions from $J_{^1H\text{-}^1H}$ couplings, and multiple CPMG-like cycles [237-238] to prolong the time allowed for spin-spin relaxation.

The results of the $^1H^\alpha$ T_2 measurements are shown in Figure S2. Clearly, Trp-α-^{13}C-β,β,2,4,5,6,7-d_7 has the longest T_2, nearly twice as large as that of Trp-U-^{13}C,^{15}N. This result is consistent with the qualitative observation of sharper linewidths for Trp-α-^{13}C-β,β,2,4,5,6,7-

d_7. This isotopolog is expected to have longer transverse relaxation times than the other isotopologs due to the presence of the multiple deuterium substitutions.

3.9.4 Photo-CIDNP of Trp isotopologs at higher concentration. In addition to the LC-photo-CIDNP spectral data shown in Figure 4C of the main article, featuring 1 μM sample concentrations, similar experiments were also carried out at higher concentration (5 μM). The NMR spectra and normalized area and intensity-at-the-maximum values for the 5 μM experiments are shown in Figure S3. The results of these experiments are qualitatively consistent with those of the experiments at 1 μM concentration (Fig. 4C, main article).

3.9.5 Quantitative assessment of light and dark effects. The effects responsible for the observed LC-photo-CIDNP sensitivity enhancements under light (LED-on) and dark (LED-off) conditions are schematically illustrated in panel A of Supplementary Figure S4. In addition, a block diagram describing the specific experiments and computations that were performed to assess the respective quantitative contributions of LED-on and LED-off effects is provided in panel B of Supplementary Figure S4. These experiments and calculations were carried out to generate all the results shown in Table 1.

3.10. Experimental Section

3.10.1 Materials. The Trp-α-^{13}C-β,β,2,4,5,6,7-d_7 isotopolog was synthesized starting from formaldehyde-$^{2}H_2$ (Cambridge Isotope Laboratories, Inc.), glycine(2-^{13}C) (Cambridge Isotope Laboratories, Inc.), indole-$^{2}H_7$ (CDN Isotopes) and pyridoxal 5'-monophosphate (MilliporeSigma). The Trp-α-^{13}C isotopolog was prepared starting from formaldehyde (MilliporeSigma), glycine(2-^{13}C) (Cambridge Isotope Laboratories, Inc.), indole (MilliporeSigma) and pyridoxal 5'-monophosphate (MilliporeSigma). Both isotopologs were synthesized in the presence of catalytic amounts of the *Pf* TrpB^{2B9} and *Tm* LTA enzymes (see sections below). Natural-abundance Trp was purchased from Advanced ChemTech, and uniformly labeled Trp (Trp-U-^{13}C,^{15}N) was purchased from Cambridge Isotopes Laboratories, Inc. The fluorescein photosensitizer was purchased as sodium salt (MilliporeSigma). The oxygen-scavenging enzymes *Aspergillus niger* glucose oxidase (GO, Enzyme commission classification code EC 1.1.3.4) and bovine liver catalase (CAT, EC 1.11.1.6) were purchased from MilliporeSigma as freeze-dried powders.

3.10.2 Preparation of *Pf* TrpB^{2B9} and *Tm* LTA Enzymes. *Pyrococcus furiosus* tryptophan synthase β-subunit 2B9 variant (*Pf* TrpB^{2B9}) [239] and *Thermotoga maritima* L-threonine aldolase (*Tm* LTA, EC 4.1.2.5) were prepared as described [205].

3.10.3 Preparation of GO and CAT enzyme stock solutions. GO and CAT enzyme stock solutions were prepared upon dissolving the respective enzyme powder in 10 mM potassium phosphate buffer at pH 7.2. Stock-solution concentrations were determined by electronic absorption spectroscopy using extinction coefficients 267,200 $M^{-1}cm^{-1}$ at 280 nm for GO and 912,500 $M^{-1}cm^{-1}$ at 276 nm for CAT [240]. Enzyme stock solutions were aliquoted, flash-frozen with liquid nitrogen, and stored at -80 °C. Prior to LC-photo-CIDNP experiments, individual aliquots were thawed at room temperature upon incubation in a water bath.

3.10.4 Synthesis of Trp Isotopologs. Synthesis of Trp-α-^{13}C-β,β,2,4,5,6,7-d_7 (IUPAC name: (*S*)-2-Amino-3-[(2,4,5,6,7-$^{2}H_5$)-3-indolyl](2-^{13}C,3,3-$^{2}H_2$)propionic acid) was carried out in the presence of the *Pf* TrpB[2B9] and *Tm* LTA enzymes according to a known enzyme-cascade procedure [205], except that we started from glycine(2-^{13}C), formaldehyde-$^{2}H_2$ and indole-$^{2}H_7$ in H_2O All stock solutions were prepared in 50 mM potassium phosphate in H_2O at pH = 8.0. The reaction was performed in 50 mM potassium phosphate in H_2O at pH = 8.0. Following heat-quenching at 90^{o}C for 15 min and centrifugation at 20,000 g for 5 min, the solution was passed through a 0.2 µm filter. Trp-α-^{13}C-β,β,2,4,5,6,7-d_7 was then purified by HPLC with a Shimadzu system comprising two LC-6AD binary pumps as well as Prominence SPD-20A ultraviolet detector and CBM-20A controller. A BioBasic 18 reverse-phase HPLC column (4.6 × 250 mm, ThermoFisher Scientific) was employed and samples were purified with a water/methanol

solvent system via a gradient including a linear increase in methanol, starting from 5 to 15% (v/v) over 20 min at a 1 ml/min flow rate. HPLC purification was followed by lyophilization. Product identity and reaction yield were determined by ^{1}H-NMR. Identity was also confirmed by ESI-MS.

The second isotopolog, Trp-α-^{13}C (IUPAC name: (*S*)-2-Amino-3-(3-indolyl)(2-^{13}C)propionic acid), was synthesized, purified and characterized as above except that glycine(2-^{13}C) and natural-abundance formaldehyde and indole were employed as starting materials. Reaction yield and identity of the final product were determined by ^{1}H-NMR, with ^{13}C decoupling during acquisition using the bi_p5m4sp_4sp.2 composite pulse decoupling sequence [241]. Identity was also confirmed by ESI-MS.

3.10.5 Preparation of Trp stock solutions. Stock solutions of natural-abundance Trp and Trp-U-^{13}C,^{15}N were prepared by dissolving the respective solid powders in distilled deionized water. Stock-solution concentrations were determined by electronic absorption spectroscopy (ext. coeff. 5,579 M^{-1} cm^{-1} at 278 nm [240]). Stock solutions of Trp-α-^{13}C and Trp-α-^{13}C-β,β,2,4,5,6,7-d_7 were also prepared by dissolving the lyophilized powders in distilled deionized water. The stock-solution concentration of these two isotopologs was assessed by ^{1}H NMR, as described below.

3.10.6 DFT calculations. The pK_a values and the g-factors of fluorescein radicals were calculated by density functional theory (DFT). Specifically, pK_a values were computed upon treatment of pK_a and proton energies via the AKB benchmark scheme by Matsui and coworkers [242-243]. According to the AKB procedure, the pK_a corresponding to the deprotonation reaction HA □ A^- + H^+ is determined via

$$pK_a = \frac{s\Delta G^\circ(solv)}{RT\ln 10} = \frac{s\{G^\circ(A^-) - G^\circ(HA)\}}{RT\ln 10} + \frac{sG^\circ(H^+)}{RT\ln 10} \quad . \tag{S10}$$

Here, $G^\circ(A^-)$, $G^\circ(HA)$, and $G^\circ(H^+)$ are the Gibbs free energies of A^-, HA, and H^+, respectively. R denotes the universal gas constant and T is the absolute temperature; s is a scaling factor that depends on the adopted computational method. Equation (S10) can be recast as (S11) and (S12), where ΔG° is the Gibbs free energy difference $G^0(A^-) - G^0(HA)$

$$pK_a = k\Delta G^\circ + C_0 \quad , \tag{S11}$$

$$k = \frac{s}{RT\ln 10}, C_0 = \frac{sG^\circ(H^+)}{RT\ln 10} \quad . \tag{S12}$$

According to the AKB scheme, the scaling factor s and the Gibbs free energy of H^+ are optimized upon fitting experimental pK_a values of known reference molecules. In this study, we used k = 0.05219 mol kJ^{-1} and C_0 = −53.3877 kJ mol^{-1} for phenol groups, and k = 0.06448 mol kJ^{-1} and C_0 = −71.0849 kJ mol^{-1} for COOH groups [243]. Geometry optimizations and Gibbs free energy calculations of fluorescein radicals were performed at the UB3LYP/6-31+G(d) level of theory. The g-factor calculations were carried out at the UB3LYP/EPR-II level in conjunction with the gauge-independent atomic orbital (GIAO) method [244-245]. In all

calculations, water solvation effects were taken into account via the integral equation formalism polarizable continuum model (IEFPCM) [246]. Atomic radii from the universal force field (UFF) [247] were employed to build the cavity for the IEFPCM computations. At the optimized geometries, no imaginary frequency was obtained in the vibrational mode analysis. Cartesian coordinates, energies and free energies of the fluorescein radicals are provided in Supplementary Table S2. All DFT calculations were carried out with the Gaussian 09 software [248]. Overall pK_a values and g-factors at room temperature were assessed upon taking the Boltzmann distribution at 298.15 K into account.

3.10.7 NMR experiments under dark (LED-off) conditions. NMR samples were prepared upon diluting appropriate stock solutions in aqueous buffer (10 mM potassium phosphate adjusted to pH 7.2). D_2O was then added to a total 10% v/v. NMR data were collected on an Avance III HD NMR spectrometer (Bruker Biospin Corp.) equipped with a 5 mm $^{1}H\{^{19}F/^{13}C/^{15}N\}$ 600 MHz triple-resonance cryogenic probe fitted with a z-gradient. Pertinent pulse sequences are as listed in the Figure legends. Unless otherwise stated, a long recycle delay (5 s) was used to ensure nearly complete spin-lattice relaxation between scans. The acquisition time was set to 2 s and 23,922 complex points. Spectra were processed with the MNova software (version 14.1.1), with zero-filling to 131,072 complex points. An exponential-decay window function was employed (0.5 Hz line-broadening). $^{1}H^{\alpha}$ resonance areas were evaluated via

integration. The concentration of Trp-α-^{13}C and Trp-α-^{13}C-β,β,2,4,5,6,7-d_7 samples was determined upon comparing the respective $^{1}H^{\alpha}$ resonance areas to those of a natural-abundance Trp sample of known concentration in experiments employing identical acquisition and processing parameters.

The high-concentration dark ^{13}C RASPRINT experiments of Figures 4A and 4B were performed on all the above Trp samples except for natural-abundance Trp. Again, a long recycle delay (5 s) was employed to ensure nearly complete spin-lattice relaxation between scans. The acquisition time was set to 0.2048 s and 2,048 complex points were used. Spectra were processed with MNova (version 14.1.1) upon zero-filling to 65,536 complex points and 5 Hz exponential decay as window function. Area and intensity-at-the-maximum were determined with MNova. When relevant, data acquired with identical acquisition and processing parameters were normalized upon dividing areas or intensities by sample concentration, to assess concentration-independent relative enhancements.\

3.10.8 LC-photo-CIDNP NMR experiments under light (LED-on) conditions. LC-photo-CIDNP experiments were carried out as described [192] with the same NMR spectrometer, console and probe employed for the NMR experiments under dark conditions. A 1.5 mm-diameter and 3.5 m long polymer optical fiber (POF, Prizmatix, Holon, Israel) was used, with emission centered at 266 nm and 2.3W power at the source (UHP-mic-LED-450). The power

at the fiber tip, to be inserted in a glass coaxial insert in contact with the NMR sample inside the sample tube, was 0.55 W. The ^{13}C RASPRINT pulse sequence was employed [192]. Data on all isotopologs were collected with a 0.2048 s acquisition time and a 0.2 s irradiation time per scan. Experiments involving the use of the reductive radical quencher ascorbic acid (vitamin C, VC) were carried out with freshly prepared (on the same day) VC stock solutions. The final VC concentration of the NMR samples was 2 μM. This value was optimized as described [193]. VC stock-solution concentrations were determined by electronic-absorption spectroscopy using the in-house measured extinction coefficient of 6,956 $M^{-1}cm^{-1}$ at 250 nm (isosbestic point, pH-independent).

3.10.9 LC-photo-CIDNP experiments on bacterial cell extracts. The S30 *E. coli* cell extract was prepared according to known procedures [208], as schematically illustrated in Figure 7B. LC-photo-CIDNP experiments were carried out upon mixing 140 μl of S30 extract with an aqueous solution of Trp-α-^{13}C-β,β,2,4,5,6,7-d_7, 70 μL of D_2O and other LC-photo-CIDNP-required components (see below) up to a 700 μL final volume. The final sample contained 10 mM pH 7.2 potassium phosphate, 10% v/v D_2O, 0.15 μM glucose oxidase (GO), 0.1 μM catalase (CAT), 2.5 μM fluorescein, 1 μM Trp-α-^{13}C-β,β,2,4,5,6,7-d_7, 2.5 mM D-glucose and 5-fold diluted *E. coli* S30 cell extract.

3.11. Supplementary References

Uncategorized References

1. Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R., The Protein Folding Problem. *Annual Review of Biophysics* **2008,** *37* (1), 289-316.
2. Anfinsen, C. B., Principles that govern the folding of protein chains. *Science* **1973,** *181* (4096), 223-30.
3. Oberhauser, A. F.; Marszalek, P. E.; Carrion-Vazquez, M.; Fernandez, J. M., Single protein misfolding events captured by atomic force microscopy. *Nat. Struct. Biol.* **1999,** *6* (11), 1025-1028.
4. Ramshini, H.; Parrini, C.; Relini, A.; Zampagni, M.; Mannini, B.; Pesce, A.; Saboury, A. A.; Nemat-Gorgani, M.; Chiti, F., Large proteins have a great tendency to aggregate but a low propensity to form amyloid fibrils. *PLoS One* **2011,** *6* (1), e16075.
5. Balchin, D.; Hayer-Hartl, M.; Hartl, F. U., In vivo aspects of protein folding and quality control. *Science* **2016,** *353* (6294), aac4354.
6. Maurizi, M. R., Proteases and protein degradation in Escherichia coli. *Experientia* **1992,** *48* (2), 178-201.
7. Sauer, R. T.; Baker, T. A., AAA+ proteases: ATP-fueled machines of protein destruction. *Annual review of biochemistry* **2011,** *80*, 587-612.
8. Powers, E. T.; Morimoto, R. I.; Dillin, A.; Kelly, J. W.; Balch, W. E., Biological and chemical approaches to diseases of proteostasis deficiency. *Annual review of biochemistry* **2009,** *78*, 959-991.
9. Richter, K.; Haslbeck, M.; Buchner, J., The heat shock response: life on the verge of death. *Mol. Cell* **2010,** *40* (2), 253-266.
10. Schramm, F. D.; Schroeder, K.; Jonas, K., Protein aggregation in bacteria. *FEMS Microbiol. Rev.* **2019,** *44* (1), 54-72.
11. Mogk, A.; Bukau, B.; Kampinga, H. H., Cellular Handling of Protein Aggregates by Disaggregation Machines. *Molecular Cell* **2018,** *69* (2), 214-226.
12. Larson, M. E.; Lesné, S. E., Soluble Aβ oligomer production and toxicity. *Journal of Neurochemistry* **2012,** *120* (s1), 125-139.
13. De, S.; Whiten, D. R.; Ruggeri, F. S.; Hughes, C.; Rodrigues, M.; Sideris, D. I.; Taylor, C. G.; Aprile, F. A.; Muyldermans, S.; Knowles, T. P. J.; Vendruscolo, M.; Bryant, C.; Blennow, K.; Skoog, I.; Kern, S.; Zetterberg, H.; Klenerman, D., Soluble aggregates present in cerebrospinal fluid change in size and mechanism of toxicity during Alzheimer's disease progression. *Acta Neuropathologica Communications* **2019,** *7* (1), 120.

14. Sweeny, E. A.; Shorter, J., Mechanistic and Structural Insights into the Prion-Disaggregase Activity of Hsp104. *Journal of Molecular Biology* **2016,** *428* (9, Part B), 1870-1885.
15. LaBreck, C. J.; May, S.; Viola, M. G.; Conti, J.; Camberg, J. L., The Protein Chaperone ClpX Targets Native and Non-native Aggregated Substrates for Remodeling, Disassembly, and Degradation with ClpP. *Frontiers in Molecular Biosciences* **2017,** *4*.
16. Maier, R.; Scholz, C.; Schmid, F. X., Dynamic association of trigger factor with protein substrates. *J. Mol. Biol.* **2001,** *314* (5), 1181-1190.
17. Calloni, G.; Chen, T.; Schermann, Sonya M.; Chang, H.-c.; Genevaux, P.; Agostini, F.; Tartaglia, Gian G.; Hayer-Hartl, M.; Hartl, F. U., DnaK Functions as a Central Hub in the E. coli Chaperone Network. *Cell Rep.* **2012,** *1* (3), 251-264.
18. Corteggiani, M.; Bossuet-Greif, N.; Nougayrède, J.-P.; Byrne, D.; Ilbert, M.; Dementin, S.; Giudici-Orticoni, M.-T.; Méjean, V.; Oswald, E.; Genest, O., Uncoupling the Hsp90 and DnaK chaperone activities revealed the in vivo relevance of their collaboration in bacteria. *Proceedings of the National Academy of Sciences* **2022,** *119* (37), e2201779119.
19. Hayer-Hartl, M.; Bracher, A.; Hartl, F. U., The GroEL–GroES chaperonin machine: a nano-cage for protein folding. *Trends in biochemical sciences* **2016,** *41* (1), 62-76.
20. To, P.; Whitehead, B.; Tarbox, H. E.; Fried, S. D., Nonrefoldability is Pervasive Across the E. coli Proteome. *Journal of the American Chemical Society* **2021,** *143* (30), 11435-11448.
21. To, P.; Xia, Y.; Lee, S. O.; Devlin, T.; Fleming, K. G.; Fried, S. D., A proteome-wide map of chaperone-assisted protein refolding in a cytosol-like milieu. *Proceedings of the National Academy of Sciences US* **2022,** *119* (48), e2210536119.
22. Pearl, L. H.; Prodromou, C., Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu. Rev. Biochem.* **2006,** *75*, 271-294.
23. Mayer, M. P.; Gierasch, L. M., Recent advances in the structural and mechanistic aspects of Hsp70 molecular chaperones. *J. Biol. Chem.* **2019,** *294* (6), 2085-2097.
24. Sauerbrei, B.; Arends, J.; Schünemann, D.; Narberhaus, F., Lon Protease Removes Excess Signal Recognition Particle Protein in Escherichia coli. *J. Bacteriol.* **2020,** *202* (14), 10.1128/jb.00161-20.
25. Chung, C. H.; Goldberg, A. L., The product of the lon (capR) gene in Escherichia coli is the ATP-dependent protease, protease La. *Proceedings of the National Academy of Sciences* **1981,** *78* (8), 4931-4935.
26. Rizo, A. N.; Lin, J.; Gates, S. N.; Tse, E.; Bart, S. M.; Castellano, L. M.; DiMaio, F.; Shorter, J.; Southworth, D. R., Structural basis for substrate gripping and translocation by the ClpB AAA+ disaggregase. *Nature communications* **2019,** *10* (1), 2393.
27. Mogk, A.; Ruger-Herreros, C.; Bukau, B., Cellular functions and mechanisms of action of small heat shock proteins. *Annu. Rev. Microbiol.* **2019,** *73*, 89-110.

28. Weibezahn, J.; Tessarz, P.; Schlieker, C.; Zahn, R.; Maglica, Z.; Lee, S.; Zentgraf, H.; Weber-Ban, E. U.; Dougan, D. A.; Tsai, F. T., Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cell* **2004,** *119* (5), 653-665.

29. Tomoyasu, T.; Mogk, A.; Langen, H.; Goloubinoff, P.; Bukau, B., Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the Escherichia coli cytosol. *Mol. Microbiol.* **2001,** *40* (2), 397-413.

30. Dill, K. A.; Ghosh, K.; Schmit, J. D., Physical limits of cells and proteomes. *Proceedings of the National Academy of Sciences* **2011,** *108* (44), 17876-17882.

31. Khodaparast, L.; Khodaparast, L.; Gallardo, R.; Louros, N. N.; Michiels, E.; Ramakrishnan, R.; Ramakers, M.; Claes, F.; Young, L.; Shahrooei, M., Aggregating sequences that occur in many proteins constitute weak spots of bacterial proteostasis. *Nature communications* **2018,** *9* (1), 866.

32. Pu, Y.; Li, Y.; Jin, X.; Tian, T.; Ma, Q.; Zhao, Z.; Lin, S.-y.; Chen, Z.; Li, B.; Yao, G., ATP-dependent dynamic protein aggregation regulates bacterial dormancy depth critical for antibiotic tolerance. *Mol. Cell* **2019,** *73* (1), 143-156. e4.

33. Kwon, S.; Jung, Y.; Lim, D., Proteomic analysis of heat-stable proteins in Escherichia coli. *BMB Rep* **2008,** *41* (2), 108-11.

34. Varela, A. E.; Lang, J. F.; Wu, Y.; Dalphin, M. D.; Stangl, A. J.; Okuno, Y.; Cavagnero, S., Kinetic Trapping of Folded Proteins Relative to Aggregates under Physiologically Relevant Conditions. *The Journal of Physical Chemistry B* **2018,** *122* (31), 7682-7698.

35. Santra, M.; Dill, K. A.; de Graff, A. M., How do chaperones protect a cell's proteins from oxidative damage? *Cell systems* **2018,** *6* (6), 743-751. e3.

36. Dahl, J.-U.; Gray, M. J.; Jakob, U., Protein quality control under oxidative stress conditions. *J. Mol. Biol.* **2015,** *427* (7), 1549-1563.

37. Yuan, A. H.; Garrity, S. J.; Nako, E.; Hochschild, A., Prion propagation can occur in a prokaryote and requires the ClpB chaperone. *Elife* **2014,** *3*, e02949.

38. de Marco, A.; Ferrer-Miralles, N.; Garcia-Fruitós, E.; Mitraki, A.; Peternel, S.; Rinas, U.; Trujillo-Roldán, M. A.; Valdez-Cruz, N. A.; Vázquez, E.; Villaverde, A., Bacterial inclusion bodies are industrially exploitable amyloids. *FEMS Microbiol. Rev.* **2018,** *43* (1), 53-72.

39. Lindner, A. B.; Madden, R.; Demarez, A.; Stewart, E. J.; Taddei, F., Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proceedings of the National Academy of Sciences* **2008,** *105* (8), 3076-3081.

40. Winkler, J.; Seybert, A.; König, L.; Pruggnaller, S.; Haselmann, U.; Sourjik, V.; Weiss, M.; Frangakis, A. S.; Mogk, A.; Bukau, B., Quantitative and spatio-temporal features of protein aggregation in Escherichia coli and consequences on protein quality control and cellular ageing. *The EMBO journal* **2010,** *29* (5), 910-923.

41. Govers, S. K.; Dutré, P.; Aertsen, A., In vivo disassembly and reassembly of protein aggregates in Escherichia coli. *J. Bacteriol.* **2014,** *196* (13), 2325-2332.
42. Neeli-Venkata, R.; Martikainen, A.; Gupta, A.; Gonçalves, N.; Fonseca, J.; Ribeiro, A. S., Robustness of the Process of Nucleoid Exclusion of Protein Aggregates in Escherichia coli. *Journal of Bacteriology* **2016,** *198* (6), 898-906.
43. Oliveira, S. M.; Neeli-Venkata, R.; Goncalves, N. S.; Santinha, J. A.; Martins, L.; Tran, H.; Mäkelä, J.; Gupta, A.; Barandas, M.; Häkkinen, A., Increased cytoplasm viscosity hampers aggregate polar segregation in Escherichia coli. *Mol. Microbiol.* **2016,** *99* (4), 686-699.
44. Coquel, A.-S.; Jacob, J.-P.; Primet, M.; Demarez, A.; Dimiccoli, M.; Julou, T.; Moisan, L.; Lindner, A. B.; Berry, H., Localization of Protein Aggregation in Escherichia coli Is Governed by Diffusion and Nucleoid Macromolecular Crowding Effect. *PLOS Computational Biology* **2013,** *9* (4), e1003038.
45. Gupta, A.; Lloyd-Price, J.; Neeli-Venkata, R.; Oliveira, Samuel M. D.; Ribeiro, Andre S., In Vivo Kinetics of Segregation and Polar Retention of MS2-GFP-RNA Complexes in Escherichia coli. *Biophysical Journal* **2014,** *106* (9), 1928-1937.
46. Nyström, T.; Liu, B., Protein quality control in time and space–links to cellular aging. *FEMS Yeast Res.* **2014,** *14* (1), 40-48.
47. Braselmann, E.; Chaney, J. L.; Clark, P. L., Folding the proteome. *Trends Biochem Sci* **2013,** *38* (7), 337-44.
48. Mateus, A.; Bobonis, J.; Kurzawa, N.; Stein, F.; Helm, D.; Hevler, J.; Typas, A.; Savitski, M. M., Thermal proteome profiling in bacteria: probing protein state in vivo. *Mol. Syst. Biol.* **2018,** *14* (7), e8242.
49. Niwa, T.; Ying, B.-W.; Saito, K.; Jin, W.; Takada, S.; Ueda, T.; Taguchi, H., Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proceedings of the National Academy of Sciences* **2009,** *106* (11), 4201-4206.
50. Niwa, T.; Kanamori, T.; Ueda, T.; Taguchi, H., Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proceedings of the National Academy of Sciences* **2012,** *109* (23), 8937-8942.
51. Leuenberger, P.; Ganscha, S.; Kahraman, A.; Cappelletti, V.; Boersema, P. J.; von Mering, C.; Claassen, M.; Picotti, P., Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **2017,** *355* (6327), eaai7825.
52. Györkei, Á.; Daruka, L.; Balogh, D.; Őszi, E.; Magyar, Z.; Szappanos, B.; Fekete, G.; Fuxreiter, M.; Horváth, P.; Pál, C.; Kintses, B.; Papp, B., Proteome-wide landscape of solubility limits in a bacterial cell. *Sci. Rep.* **2022,** *12* (1), 6547.
53. de Groot, N. S.; Ventura, S., Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. *J. Biotechnol.* **2006,** *125* (1), 110-113.

54. Hipp, M. S.; Park, S.-H.; Hartl, F. U., Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.* **2014,** *24* (9), 506-514.
55. Nissley, D. A.; Jiang, Y.; Trovato, F.; Sitarik, I.; Narayan, K. B.; To, P.; Xia, Y.; Fried, S. D.; O'Brien, E. P., Universal protein misfolding intermediates can bypass the proteostasis network and remain soluble and less functional. *Nat. Comm.* **2022,** *13* (1), 3081.
56. Vecchi, G.; Sormanni, P.; Mannini, B.; Vandelli, A.; Tartaglia, G. G.; Dobson, C. M.; Hartl, F. U.; Vendruscolo, M., Proteome-wide observation of the phenomenon of life on the edge of solubility. *Proceedings of the National Academy of Sciences US* **2020,** *117* (2), 1015-1020.
57. Gazit, E., The "Correctly Folded" State of Proteins: Is It a Metastable State? *Angew. Chem. Int. Ed.* **2002,** *41* (2), 257-259.
58. Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-Cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M., Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *Journal of the American Chemical Society* **2011,** *133* (36), 14160-14163.
59. Morell, M.; Bravo, R.; Espargaró, A.; Sisquella, X.; Avilés, F. X.; Fernàndez-Busquets, X.; Ventura, S., Inclusion bodies: Specificity in their aggregation process and amyloid-like structure. *Biochim. Biophys. Acta* **2008,** *1783* (10), 1815-1825.
60. Carrió, M. M.; Cubarsi, R.; Villaverde, A., Fine architecture of bacterial inclusion bodies. *FEBS Lett.* **2000,** *471* (1), 7-11.
61. Lee, S.; Choi, M. C.; Al Adem, K.; Lukman, S.; Kim, T.-Y., Aggregation and Cellular Toxicity of Pathogenic or Non-pathogenic Proteins. *Sci. Rep.* **2020,** *10* (1), 5120.
62. Zhou, H.-X.; Rivas, G.; Minton, A. P., Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. *Annu. Rev. Biophys.* **2008,** *37* (1), 375-397.
63. Elcock, A. H., Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr. Opin. Struct. Biol.* **2010,** *20* (2), 196-206.
64. Christiansen, A.; Wang, Q.; Samiotakis, A.; Cheung, M. S.; Wittung-Stafshede, P., Factors Defining Effects of Macromolecular Crowding on Protein Stability: An in Vitro/in Silico Case Study Using Cytochrome c. *Biochemistry* **2010,** *49* (31), 6519-6530.
65. Ciryam, P.; Morimoto, R. I.; Vendruscolo, M.; Dobson, C. M.; O'Brien, E. P., *In vivo* translation rates can substantially delay the cotranslational folding of the *Escherichia coli* cytosolic proteome. *Proceedings of the National Academy of Sciences of the United States of america* **2013,** *110* (2), E132-E140.
66. Netzer, W. J.; Hartl, F. U., Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **1997,** *388* (6640), 343-349.

67. Addabbo, R. M.; Dalphin, M. D.; Mecha, M. F.; Liu, Y.; Staikos, A.; Guzman-Luna, V.; Cavagnero, S., Complementary Role of Co- and Post-Translational Events in De Novo Protein Biogenesis. *J. Phys. Chem. B* **2020,** *124* (30), 6488-6507.
68. Kerner, M. J.; Naylor, D. J.; Ishihama, Y.; Maier, T.; Chang, H.-C.; Stines, A. P.; Georgopoulos, C.; Frishman, D.; Hayer-Hartl, M.; Mann, M.; Hartl, F. U., Proteome-wide Analysis of Chaperonin-Dependent Protein Folding in Escherichia coli. *Cell* **2005,** *122* (2), 209-220.
69. Reichmann, D.; Suss, O., Protein plasticity underlines activation and function of ATP-independent chaperones. *Frontiers in Molecular Biosciences* **2015,** *2*.
70. Koubek, J.; Schmitt, J.; Galmozzi, C. V.; Kramer, G., Mechanisms of Cotranslational Protein Maturation in Bacteria. *Frontiers in Molecular Biosciences* **2021,** *8*.
71. Rusilowicz-Jones, E. V.; Urbe, S.; Clague, M. J., Protein degradation on the global scale. *Molecular Cell* **2022,** *82* (8), 1414-1423.
72. Chen, W. X.; Smeekens, J. M.; Wu, R. H., Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chemical Science* **2016,** *7* (2), 1393-1400.
73. Truscott, R. J. W.; Schey, K. L.; Friedrich, M. G., Old Proteins in Man: A Field its Infancy. *Trends in Biochemical Sciences* **2016,** *41* (8), 654-664.
74. Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H., The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences* **1961,** *47* (9), 1309-1314.
75. Katikaridis, P.; Bohl, V.; Mogk, A., Resisting the Heat: Bacterial Disaggregases Rescue Cells From Devastating Protein Aggregation. *Frontiers in Molecular Biosciences* **2021,** *8*.
76. Narimoto, T.; Sakurai, K.; Okamoto, A.; Chatani, E.; Hoshino, M.; Hasegawa, K.; Naiki, H.; Goto, Y., Conformational stability of amyloid fibrils of beta2-microglobulin probed by guanidine-hydrochloride-induced unfolding. *FEBS Lett.* **2004,** *576* (3), 313-9.
77. Kalwarczyk, T.; Tabaka, M.; Holyst, R., Biologistics—Diffusion coefficients for complete proteome of Escherichia coli. *Bioinformatics* **2012,** *28* (22), 2971-2978.
78. Stark, H.; Mueller, F.; Orlova, E. V.; Schatz, M.; Dube, P.; Erdemir, T.; Zemlin, F.; Brimacombe, R.; Vanheel, M., The 70S Escherichia coli ribosome at 23 Angstrom resolution -fitting the ribosomal RNA. *Structure* **1995,** *3* (8), 815-821.
79. Bakshi, S.; Siryaporn, A.; Goulian, M.; Weisshaar, J. C., Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Molecular Microbiology* **2012,** *85* (1), 21-38.
80. Qin, D. M.; Fredrick, K., Analysis of Polysomes from Bacteria. In *Laboratory Methods in Enzymology: Rna*, Lorsch, J., Ed. 2013; Vol. 530, pp 159-172.

81. Wiśniewski, J. R.; Hein, M. Y.; Cox, J.; Mann, M., A “Proteomic Ruler” for Protein Copy Number and Concentration Estimation without Spike-in Standards*. *Mol. Cell. Proteomics* **2014,** *13* (12), 3497-3506.
82. Scholtz, J. M.; Grimsley, G. R.; Pace, C. N., Chapter 23 Solvent Denaturation of Proteins and Interpretations of the m Value. In *Methods in Enzymology*, Academic Press: 2009; Vol. 466, pp 549-565.
83. Santoro, M. M.; Bolen, D. W., Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl .alpha.-chymotrypsin using different denaturants. *Biochemistry* **1988,** *27* (21), 8063-8068.
84. Oosawa, F.; Kasai, M., A theory of linear and helical aggregations of macromolecules. *J Mol Biol* **1962,** *4*, 10-21.
85. Myers, J. K.; Nick Pace, C.; Martin Scholtz, J., Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **1995,** *4* (10), 2138-2148.
86. Buell, A. K., Stability matters, too - the thermodynamics of amyloid fibril formation. *Chem. Sci.* **2022,** *13* (35), 10177-10192.
87. Vettore, N.; Buell, A. K., Thermodynamics of amyloid fibril formation from chemical depolymerization. *Physical Chemistry Chemical Physics* **2019,** *21* (47), 26184-26194.
88. Shahnawaz, M.; Soto, C., Microcin amyloid fibrils A are reservoir of toxic oligomeric species. *J Biol Chem* **2012,** *287* (15), 11665-76.
89. Nespovitaya, N.; Gath, J.; Barylyuk, K.; Seuring, C.; Meier, B.; Riek, R., Dynamic Assembly and Disassembly of Functional β-Endorphin Amyloid Fibrils. *Journal of the American Chemical Society* **2015,** *138*.
90. O'Nuallain, B.; Shivaprasad, S.; Kheterpal, I.; Wetzel, R., Thermodynamics of Aβ(1−40) Amyloid Fibril Elongation. *Biochemistry* **2005,** *44* (38), 12709-12718.
91. Smith, C. A., Physiology of the Bacterial Cell. A Molecular Approach: by F C Neidhardt, J L Ingraham and M Schaechter. pp 507. Sinauer Associates, Sunderland, MA. 1990. $43.95. *Biochemical Education* **1992,** *20* (2), 124-125.
92. Petrov, A. S.; Bernier, C. R.; Hsiao, C.; Okafor, C. D.; Tannenbaum, E.; Stern, J.; Gaucher, E.; Schneider, D.; Hud, N. V.; Harvey, S. C.; Dean Williams, L., RNA–Magnesium–Protein Interactions in Large Ribosomal Subunit. *The Journal of Physical Chemistry B* **2012,** *116* (28), 8113-8120.
93. Drygin, D.; Zimmermann, R. A., Magnesium ions mediate contacts between phosphoryl oxygens at positions 2122 and 2176 of the 23S rRNA and ribosomal protein L1. *RNA* **2000,** *6* (12), 1714-1726.
94. Cox, R. A.; Pratt, H.; Huvos, P.; Higginson, B.; Hirst, W., A study of the thermal stability of ribosomes and biologically active subribosomal particles. *Biochem J* **1973,** *134* (3), 775-93.

95. Harris, M. A.; Clark, J.; Ireland, A.; Lomax, J.; Ashburner, M.; Foulger, R.; Eilbeck, K.; Lewis, S.; Marshall, B.; Mungall, C.; Richter, J.; Rubin, G. M.; Blake, J. A.; Bult, C.; Dolan, M.; Drabkin, H.; Eppig, J. T.; Hill, D. P.; Ni, L.; Ringwald, M.; Balakrishnan, R.; Cherry, J. M.; Christie, K. R.; Costanzo, M. C.; Dwight, S. S.; Engel, S.; Fisk, D. G.; Hirschman, J. E.; Hong, E. L.; Nash, R. S.; Sethuraman, A.; Theesfeld, C. L.; Botstein, D.; Dolinski, K.; Feierbach, B.; Berardini, T.; Mundodi, S.; Rhee, S. Y.; Apweiler, R.; Barrell, D.; Camon, E.; Dimmer, E.; Lee, V.; Chisholm, R.; Gaudet, P.; Kibbe, W.; Kishore, R.; Schwarz, E. M.; Sternberg, P.; Gwinn, M.; Hannick, L.; Wortman, J.; Berriman, M.; Wood, V.; de la Cruz, N.; Tonellato, P.; Jaiswal, P.; Seigfried, T.; White, R., The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **2004,** *32* (Database issue), D258-61.
96. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **2021,** *49* (D1), D325-d334.
97. Durica-Mitic, S.; Göpel, Y.; Görke, B., Carbohydrate Utilization in Bacteria: Making the Most Out of Sugars with the Help of Small Regulatory RNAs. *Microbiology Spectrum* **2018,** *6* (2), 10.1128/microbiolspec.rwr-0013-2017.
98. Hollinshead, W. D.; Rodriguez, S.; Martin, H. G.; Wang, G.; Baidoo, E. E. K.; Sale, K. L.; Keasling, J. D.; Mukhopadhyay, A.; Tang, Y. J., Examining Escherichia coli glycolytic pathways, catabolite repression, and metabolite channeling using Δpfk mutants. *Biotechnology for Biofuels* **2016,** *9* (1), 212.
99. Rubio Gomez, M. A.; Ibba, M., Aminoacyl-tRNA synthetases. *RNA* **2020,** *26* (8), 910-936.
100. Wagner, L.; Jules, M.; Borkowski, O., What remains from living cells in bacterial lysate-based cell-free systems. *Computational and Structural Biotechnology Journal* **2023,** *21*, 3173-3182.
101. Grindle, M. P.; Carter, B.; Alao, J. P.; Connors, K.; Tehver, R.; Kravats, A. N. Structural Communication between the E. coli Chaperones DnaK and Hsp90 *Int. J. Mol. Sci.* [Online], 2021.
102. Banecki, B.; Liberek, K.; Wall, D.; Wawrzynów, A.; Georgopoulos, C.; Bertoli, E.; Tanfani, F.; Zylicz, M., Structure-Function Analysis of the Zinc Finger Region of the DnaJ Molecular Chaperone*. *Journal of Biological Chemistry* **1996,** *271* (25), 14840-14848.
103. Perales-Calvo, J.; Lezamiz, A.; Garcia-Manyes, S., The Mechanochemistry of a Structural Zinc Finger. *The Journal of Physical Chemistry Letters* **2015,** *6* (17), 3335-3340.
104. Goemans, C.; Denoncin, K.; Collet, J.-F., Folding mechanisms of periplasmic proteins. *Biochim. Biophys. Acta* **2014,** *1843* (8), 1517-1528.
105. Behrens, S.; Maier, R.; de Cock, H.; Schmid, F. X.; Gross, C. A., The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *The EMBO Journal* **2001,** *20* (1-2), 285-294.

106. Mehta, P.; Woo, P.; Venkataraman, K.; Karzai, A. W., Ribosome purification approaches for studying interactions of regulatory proteins and RNAs with the ribosome. *Methods Mol Biol* **2012,** *905*, 273-89.
107. Felske, A.; Engelen, B.; Nübel, U.; Backhaus, H., Direct ribosome isolation from soil to extract bacterial rRNA for community analysis. *Appl Environ Microbiol* **1996,** *62* (11), 4162-7.
108. Zecha, J.; Satpathy, S.; Kanashova, T.; Avanessian, S. C.; Kane, M. H.; Clauser, K. R.; Mertins, P.; Carr, S. A.; Kuster, B., TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Mol Cell Proteomics* **2019,** *18* (7), 1468-1478.
109. Halder, R.; Nissley, D. A.; Sitarik, I.; Jiang, Y.; Rao, Y.; Vu, Q. V.; Li, M. S.; Pritchard, J.; O'Brien, E. P., How soluble misfolded proteins bypass chaperones at the molecular level. *Nat. Comm.* **2023,** *14* (1), 3689.
110. Mecha, M. F.; Hutchinson, R. B.; Lee, J. H.; Cavagnero, S., Protein folding in vitro and in the cell: From a solitary journey to a team effort. *Biophys. Chem.* **2022,** *287*, 106821.
111. Hnath, B.; Chen, J.; Reynolds, J.; Choi, E.; Wang, J.; Zhang, D.; Sha, C. M.; Dokholyan, N. V., Big versus small: The impact of aggregate size in disease. *Protein Sci.* **2023,** *32* (7), e4686.
112. Cole, S. D.; Miklos, A. E.; Chiao, A. C.; Sun, Z. Z.; Lux, M. W., Methodologies for preparation of prokaryotic extracts for cell-free expression systems. *Synth Syst Biotechnol* **2020,** *5* (4), 252-267.
113. Cui, Y.; Chen, X.; Wang, Z.; Lu, Y., Ribosome purification from Escherichia coli by ultracentrifugation. *Biotechnology Notes* **2022,** *3*, 118-123.
114. Cortés-Ríos, J.; Zárate, A. M.; Figueroa, J. D.; Medina, J.; Fuentes-Lemus, E.; Rodríguez-Fernández, M.; Aliaga, M.; López-Alarcón, C., Protein quantification by bicinchoninic acid (BCA) assay follows complex kinetics and can be performed at short incubation times. *Analytical Biochemistry* **2020,** *608*, 113904.
115. Linse, S., Toward the equilibrium and kinetics of amyloid peptide self-assembly. *Curr. Opin. Struct. Biol.* **2021,** *70*, 87-98.
116. Hiemenz, P.; Rajagopalan, R., *Principles of Colloid and Surface Chemistry*. 3 ed.; Marcel Dekker: New York, USA, 1997.
117. Hixson, A. W.; Crowell, J. H., Dependence of Reaction Velocity upon surface and Agitation. *Ind. Eng. Chem.* **1931,** *23* (8), 923-931.
118. Galea, C. A.; Pagala, V. R.; Obenauer, J. C.; Park, C.-G.; Slaughter, C. A.; Kriwacki, R. W., Proteomic Studies of the Intrinsically Unstructured Mammalian Proteome. *J. Proteome Res.* **2006,** *5* (10), 2839-2848.
119. Hershewe, J. M.; Warfel, K. F.; Iyer, S. M.; Peruzzi, J. A.; Sullivan, C. J.; Roth, E. W.; DeLisa, M. P.; Kamat, N. P.; Jewett, M. C., Improving cell-free glycoprotein synthesis by characterizing and enriching native membrane vesicles. *Nat. Comm.* **2021,** *12* (1), 2363.

120. Stetefeld, J.; McKenna, S. A.; Patel, T. R., Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys. Rev.* **2016,** *8* (4), 409-427.
121. de Silva, A. P.; de Silva, S. A., Fluorescent signalling crown ethers; 'switching on' of fluorescence by alkali metal ion recognition and binding in situ. *J. Chem. Soc., Chem. Commun.* **1986,** (23), 1709-1710.
122. Pace, C. N., Measuring and increasing protein stability. *Trends Biotechnol.* **1990,** *8* (4), 93-8.
123. Na, Y. R.; Park, C., Investigating protein unfolding kinetics by pulse proteolysis. *Protein Sci.* **2009,** *18* (2), 268-76.
124. Welte, H.; Kovermann, M., Insights into Protein Stability in Cell Lysate by 19F NMR Spectroscopy. *ChemBioChem* **2020,** *21* (24), 3575-3579.
125. Groenning, M.; Norrman, M.; Flink, J. M.; van de Weert, M.; Bukrinsky, J. T.; Schluckebier, G.; Frokjaer, S., Binding mode of Thioflavin T in insulin amyloid fibrils. *J Struct Biol* **2007,** *159* (3), 483-97.
126. Xue, C.; Lin, T. Y.; Chang, D.; Guo, Z., Thioflavin T as an amyloid dye: fibril quantification, optimal concentration and effect on aggregation. *R Soc Open Sci* **2017,** *4* (1), 160696.
127. Novo, M.; Freire, S.; Al-Soufi, W., Critical aggregation concentration for the formation of early Amyloid-β (1–42) oligomers. *Sci. Rep.* **2018,** *8* (1), 1783.
128. Nespovitaya, N.; Gath, J.; Barylyuk, K.; Seuring, C.; Meier, B. H.; Riek, R., Dynamic Assembly and Disassembly of Functional β-Endorphin Amyloid Fibrils. *Journal of the American Chemical Society* **2016,** *138* (3), 846-856.
129. Consortium, T. U., UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **2022,** *51* (D1), D523-D531.
130. Lo Conte, L.; Ailey, B.; Hubbard, T. J.; Brenner, S. E.; Murzin, A. G.; Chothia, C., SCOP: a structural classification of proteins database. *Nucleic Acids Res* **2000,** *28* (1), 257-9.
131. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H., Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2006,** *2*, 2006.0008.
132. Sayers, E. W.; Bolton, E. E.; Brister, J. R.; Canese, K.; Chan, J.; Comeau, D. C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; Madej, T.; Marchler-Bauer, A.; Lanczycki, C.; Lathrop, S.; Lu, Z.; Thibaud-Nissen, F.; Murphy, T.; Phan, L.; Skripchenko, Y.; Tse, T.; Wang, J.; Williams, R.; Trawick, B. W.; Pruitt, K. D.; Sherry, S. T., Database resources of the national center for biotechnology information. *Nucleic Acids Res* **2022,** *50* (D1), D20-d26.
133. Xia, K.; Manning, M.; Hesham, H.; Lin, Q.; Bystroff, C.; Colón, W., Identifying the subproteome of kinetically stable proteins via diagonal 2D SDS/PAGE. *Proceedings of the National Academy of Sciences* **2007,** *104* (44), 17329-17334.

134. Costa, S.; Almeida, A.; Castro, A.; Domingues, L., Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Front. Microbiol.* **2014,** *5*.
135. Vinaixa, M.; Samino, S.; Saez, I.; Duran, J.; Guinovart, J. J.; Yanes, O., A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites* **2012,** *2* (4), 775-95.
136. Russell, M. R.; Lilley, K. S., Pipeline to assess the greatest source of technical variance in quantitative proteomics using metabolic labelling. *J. Proteomics* **2012,** *77*, 441-454.
137. Krzywinski, M.; Altman, N., Visualizing samples with box plots. *Nat. Methods* **2014,** *11* (2), 119-120.
138. Spitzer, M.; Wildenhain, J.; Rappsilber, J.; Tyers, M., BoxPlotR: a web tool for generation of box plots. *Nat. Methods* **2014,** *11* (2), 121-2.
139. Speed, M. A.; Wang, D. I. C.; King, J., Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition. *Nat. Biotechnol.* **1996,** *14* (10), 1283-1287.
140. Jones, E. M.; Surewicz, W. K., Fibril Conformation as the Basis of Species- and Strain-Dependent Seeding Specificity of Mammalian Prion Amyloids. *Cell* **2005,** *121* (1), 63-72.
141. Rajan, R. S.; Illing, M. E.; Bence, N. F.; Kopito, R. R., Specificity in intracellular protein aggregation and inclusion body formation. *Proceedings of the National Academy of Sciences* **2001,** *98* (23), 13060-13065.
142. Blattner, F. R.; Plunkett, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y., The Complete Genome Sequence of Escherichia coli K-12. *Science* **1997,** *277* (5331), 1453-1462.
143. Candiano, G.; Bruschi, M.; Musante, L.; Santucci, L.; Ghiggeri, G. M.; Carnemolla, B.; Orecchia, P.; Zardi, L.; Righetti, P. G., Blue silver: A very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **2004,** *25* (9), 1327-1333.
144. Glyk, A.; Heinisch, S. L.; Scheper, T.; Beutel, S., Comparison of colorimetric methods for the quantification of model proteins in aqueous two-phase systems. *Analytical Biochemistry* **2015,** *477*, 35-37.
145. Yaginuma, H.; Kawai, S.; Tabata, K. V.; Tomiyama, K.; Kakizuka, A.; Komatsuzaki, T.; Noji, H.; Imamura, H., Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Sci. Rep.* **2014,** *4* (1), 6522.
146. Mogk, A.; Tomoyasu, T.; Goloubinoff, P.; Rüdiger, S.; Röder, D.; Langen, H.; Bukau, B., Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB. *EMBO J* **1999,** *18* (24), 6934-49.

147. Doyle, S. M.; Shastry, S.; Kravats, A. N.; Shih, Y.-H.; Miot, M.; Hoskins, J. R.; Stan, G.; Wickner, S., Interplay between E. coli DnaK, ClpB and GrpE during Protein Disaggregation. *Journal of Molecular Biology* **2015,** *427* (2), 312-327.
148. Doyle, S. M.; Hoskins, J. R.; Wickner, S., DnaK Chaperone-dependent Disaggregation by Caseinolytic Peptidase B (ClpB) Mutants Reveals Functional Overlap in the N-terminal Domain and Nucleotide-binding Domain-1 Pore Tyrosine*. *Journal of Biological Chemistry* **2012,** *287* (34), 28470-28479.
149. Failmezger, J.; Rauter, M.; Nitschel, R.; Kraml, M.; Siemann-Herzberg, M., Cell-free protein synthesis from non-growing, stressed Escherichia coli. *Sci. Rep.* **2017,** *7* (1), 16524.
150. Foshag, D.; Henrich, E.; Hiller, E.; Schäfer, M.; Kerger, C.; Burger-Kentischer, A.; Diaz-Moreno, I.; García-Mauriño, S. M.; Dötsch, V.; Rupp, S.; Bernhard, F., The E. coli S30 lysate proteome: A prototype for cell-free protein production. *N. Biotechnol.* **2018,** *40*, 245-260.
151. Lee, J. H.; Okuno, Y.; Cavagnero, S., Sensitivity enhancement in solution NMR: emerging ideas and new frontiers. *J. Magn. Reson.* **2014,** *241*, 18-31.
152. Ardenkjaer-Larsen, J. H.; Boebinger, G. S.; Comment, A.; Duckett, S.; Edison, A. S.; Engelke, F.; Griesinger, C.; Griffin, R. G.; Hilty, C.; Maeda, H.; Parigi, G.; Prisner, T.; Ravera, E.; van Bentum, J.; Vega, S.; Webb, A.; Luchinat, C.; Schwalbe, H.; Frydman, L., Facing and overcoming sensitivity challenges in biomolecular NMR spectroscopy. *Angew. Chem. Int. Ed.* **2015,** *54* (32), 9162-9185.
153. Schops, P.; Spindler, P. E.; Prisner, T. F., Multi-frequency pulsed Overhauser DNP at 1.2 Tesla. *Zeit. Phys. Chem. Intl. J. Res. Phys. Chem. Chem. Phys.* **2017,** *231* (3), 561-573.
154. Denysenkov, V. P.; Prisner, T. F., Liquid-state Overhauser DNP at high magnetic fields. *Emagres* **2019,** *8* (1), 41-54.
155. Bennati, M.; Orlando, T., Overhauser DNP in liquids on ^{13}C nuclei. *Emagres* **2019,** *8* (1), 11-18.
156. Kim, Y.; Hilty, C., Applications of Dissolution-DNP for NMR Screening. In *Biological Nmr, Pt B*, Wand, A. J., Ed. 2019; Vol. 615, pp 501-526.
157. Kagawa, A.; Miyanishi, K.; Ichijo, N.; Negoro, M.; Nakamura, Y.; Enozawa, H.; Murata, T.; Morita, Y.; Kitagawa, M., High-field NMR with dissolution triplet-DNP. *J. Magn. Reson.* **2019,** *309*, 106623.
158. Jahnig, F.; Himmler, A.; Kwiatkowski, G.; Dapp, A.; Hunkele, A.; Kozerke, S.; Ernst, M., A spin-thermodynamic approach to characterize spin dynamics in TEMPO-based samples for dissolution DNP at 7 T field. *J. Magn. Reson.* **2019,** *303*, 91-104.
159. Bowers, C. R.; Weitekamp, D. P., Transformation of symmetrization order to nuclear-spin magnetization by chemical reaction and nuclear magnetic resonance. *Phys. Rev. Lett.* **1986,** *57* (21), 2645-2648.

160. Bowers, C. R.; Weitekamp, D. P., Parahydrogen and synthesis allow dramatically enhanced nuclear alignment. *J. Am. Chem. Soc.* **1987,** *109* (18), 5541-5542.
161. Carson, P. J.; Bowers, C. R.; Weitekamp, D. P., The PASADENA effect at a solid surface: High-sensitivity nuclear magnetic resonance of hydrogen chemisorption. *J. Am. Chem. Soc.* **2001,** *123* (47), 11821-11822.
162. Pravica, M. G.; Weitekamp, D. P., Net NMR alignment by adiabatic transport of parahydrogen addition products to high magnetic field. *Chem. Phys. Lett.* **1988,** *145* (4), 255-258.
163. Wang, W. Y.; Xu, J.; Deng, F. W., Q.; Chu, Y. Y.; Qi, G. D.; Li, S. H., Pairwise stereoselective hydrogenation of propyne on supported Pd-Ag catalysts investigated by parahydrogen-induced polarization. *J. Phys. Chem. C* **2021,** *125* (31), 17144-17154.
164. Salnikov, O. G.; Chukanov, N. V.; Kovtunova, L. M.; Bukhtiyarov, V. I.; Kovtunov, K. V.; Shchepin, R. V.; Koptyug, I. V.; Chekmenev, E. Y., Heterogeneous ^{1}H and ^{13}C parahydrogen-Induced polarization of acetate and pyruvate esters. *ChemPhysChem* **2021,** *22* (13), 1389-1396.
165. Pokochueva, E. V.; Burueva, D. B.; Salnikov, O. G.; Koptyug, I. V., Heterogeneous catalysis and parahydrogen-induced polarization. *ChemPhysChem* **2021,** *22* (14), 1421-1440.
166. Adams, R. W.; Aguilar, J. A.; Atkinson, K. D.; Cowley, M. J.; Elliott, P. I. P.; Duckett, S. B.; Green, G. G. R.; Khazal, I. G.; Lopez-Serrano, J.; Williamson, D. C., Reversible interactions with para-hydrogen enhance NMR sensitivity by polarization transfer. *Science* **2009,** *323* (5922), 1708-1711.
167. Duckett, S. B.; Mewis, R. E., Application of Parahydrogen Induced Polarization Techniques in NMR Spectroscopy and Imaging. *Acc. Chem. Res.* **2012,** *45* (8), 1247-1257.
168. Pravdivtsev, A. N.; Buntkowsky, G.; Duckett, S. B.; Koptyug, I. V.; Hovener, J. B., Parahydrogen-induced polarization of amino acids. *Angew. Chem.* **2021,** *60* (44), 23496-23507.
169. Taylor, D. A.; Natrajan, L. S.; Nilsson, M.; Adams, R. W., SABRE-enhanced real-time pure shift NMR spectroscopy. *Magn. Reson. Chem.* **2021,** *59*, 1244-1252.
170. Spiridonov, K. A.; Kozinenko, V. P.; Nikovsky, I. A.; Pavlov, A. A.; Vol'khina, T. N.; Nelyubina, Y. V.; Kiryutin, A. S.; Yurkovskaya, A. V.; Polezhaev, A. A.; Novikov, V. V.; Ivanov, K. L., Phosphite-containing iridium polarization transfer catalysts for NMR signal amplification by reversible exchange. *Mendeleev Commun.* **2021,** *31* (4), 475-477.
171. Kiryutin, A. S.; Yurkovskaya, A. V.; Petrov, P. A.; Ivanov, K. L., Simultaneous ^{15}N polarization of several biocompatible substrates in ethanol-water mixtures by signal amplification by reversible exchange (SABRE) method. *Magn. Reson. Chem.* **2021,** *59*, 1216-1224.

172. Barskiy, D. A.; Kovtunov, K. V.; Koptyug, I. V.; He, P.; Groome, K. A.; Best, Q. A.; Shi, F.; Goodson, B. M.; Shchepin, R. V.; Coffey, A. M.; Waddell, K. W.; Chekmenev, E. Y., The Feasibility of Formation and Kinetics of NMR Signal Amplification by Reversible Exchange (SABRE) at High Magnetic Field (9.4 T). *J. Am. Chem. Soc.* **2014,** *136* (9), 3322-3325.
173. Barskiy, D. A.; Kovtunov, K. V.; Koptyug, I. V.; He, P.; Groome, K. A.; Best, Q. A.; Shi, F.; Goodson, B. M.; Shchepin, R. V.; Truong, M. L.; Coffey, A. M.; Waddell, K. W.; Chekmenev, E. Y., In Situ and Ex Situ Low-Field NMR Spectroscopy and MRI Endowed by SABRE Hyperpolarization. *ChemPhysChem* **2014,** *15* (18), 4100-4107.
174. Feng, B. B.; Coffey, A. M.; Colon, R. D.; Chekmenev, E. Y.; Waddell, K. W., A pulsed injection parahydrogen generator and techniques for quantifying enrichment. *J. Magn. Reson.* **2012,** *214*, 258-262.
175. Shchepin, R. V.; Barskiy, D. A.; Coffey, A. M.; Theis, T.; Shi, F.; Warren, W. S.; Goodson, B. M.; Chekmenev, E. Y., N-15 Hyperpolarization of Imidazole-N-15(2) for Magnetic Resonance pH Sensing via SABRE-SHEATH. *Acs Sensors* **2016,** *1* (6), 640-644.
176. Shchepin, R. V.; Truong, M. L.; Theis, T.; Coffey, A. M.; Shi, F.; Waddell, K. W.; Warren, W. S.; Goodson, B. M.; Chekmenev, E. Y., Hyperpolarization of "Neat" Liquids by NMR Signal Amplification by Reversible Exchange. *Journal of Physical Chemistry Letters* **2015,** *6* (10), 1961-1967.
177. Shi, F.; Coffey, A. M.; Waddell, K. W.; Chekmenev, E. Y.; Goodson, B. M., Heterogeneous Solution NMR Signal Amplification by Reversible Exchange. *Angewandte Chemie-International Edition* **2014,** *53* (29), 7495-7498.
178. Shi, F.; Coffey, A. M.; Waddell, K. W.; Chekmenev, E. Y.; Goodson, B. M., Nanoscale Catalysts for NMR Signal Enhancement by Reversible Exchange. *J. Phys. Chem. C* **2015,** *119* (13), 7525-7533.
179. Shi, F.; He, P.; Best, Q. A.; Groome, K.; Truong, M. L.; Coffey, A. M.; Zimay, G.; Shchepin, R. V.; Waddell, K. W.; Chekmenev, E. Y.; Goodson, B. M., Aqueous NMR Signal Enhancement by Reversible Exchange in a Single Step Using Water-Soluble Catalysts. *J. Phys. Chem. C* **2016,** *120* (22), 12149-12156.
180. Theis, T.; Truong, M.; Coffey, A. M.; Chekmenev, E. Y.; Warren, W. S., LIGHT-SABRE enables efficient in-magnet catalytic hyperpolarization. *J. Magn. Reson.* **2014,** *248*, 23-26.
181. Theis, T.; Truong, M. L.; Coffey, A. M.; Shchepin, R. V.; Waddell, K. W.; Shi, F.; Goodson, B. M.; Warren, W. S.; Chekmenev, E. Y., Microtesla SABRE Enables 10% Nitrogen-15 Nuclear Spin Polarization. *J. Am. Chem. Soc.* **2015,** *137* (4), 1404-1407.
182. Truong, M. L.; Shi, F.; He, P.; Yuan, B. X.; Plunkett, K. N.; Coffey, A. M.; Shchepin, R. V.; Barskiy, D. A.; Kovtunov, K. V.; Koptyug, I. V.; Waddell, K. W.; Goodson, B. M.; Chekmenev, E. Y., Irreversible Catalyst Activation Enables Hyperpolarization and Water Solubility for NMR Signal Amplification by Reversible Exchange. *J. Phys. Chem. B* **2014,** *118* (48), 13882-13889.

183. Truong, M. L.; Theis, T.; Coffey, A. M.; Shchepin, R. V.; Waddell, K. W.; Shi, F.; Goodson, B. M.; Warren, W. S.; Chekmenev, E. Y., N-15 Hyperpolarization by Reversible Exchange Using SABRE-SHEATH. *J. Phys. Chem. C* **2015,** *119* (16), 8786-8797.
184. Bargon, J.; Fischer, H.; Johnsen, U., Kernresonanz-emissionslinien während rascher radikalreaktionen. *Zeitschrift für Naturforschung A* **1967,** *22* (10), 1551-1555.
185. Ward, H. R.; Lawler, R. G., Nuclear magnetic resonance emission and enhanced absorption in rapid organometallic reactions. *J. Am. Chem. Soc.* **1967,** *89* (21), 5518-5519.
186. Kaptein, R.; Dijkstra, K.; Nicolay, K., Laser photo-CIDNP as a surface probe for proteins in solution. *Nature* **1978,** *274* (5668), 293-4.
187. Lee, J. H.; Cavagnero, S., A novel tri-enzyme system in combination with laser-driven NMR enables efficient nuclear polarization of biomolecules in solution. *J. Phys. Chem. B* **2013,** *117* (20), 6069-6081.
188. Lee, J. H.; Sekhar, A.; Cavagnero, S., ^{1}H-Detected ^{13}C photo-CIDNP as a sensitivity enhancement tool in solution NMR. *J. Am. Chem. Soc.* **2011,** *133* (21), 8062-5.
189. Lyon, C. E.; Jones, J. A.; Redfield, C.; Dobson, C. M.; Hore, P. J., Two-dimensional ^{15}N-^{1}H photo-CIDNP as a surface probe of native and partially structured proteins. *J. Am. Chem. Soc.* **1999,** *121* (27), 6505-6506.
190. Okuno, Y.; Cavagnero, S., Fluorescein: a photo-CIDNP sensitizer enabling hypersensitive NMR data collection in liquids at low micromolar concentration. *J. Phys. Chem. B* **2016,** *120* (4), 715-23.
191. Okuno, Y.; Mecha, M. F.; Yang, H.; Zhu, L.; Fry, C. G.; Cavagnero, S., Laser- and cryogenic probe-assisted NMR enables hypersensitive analysis of biomolecules at submicromolar concentration. *Proc. Natl. Acad. Sci. USA* **2019,** *116* (24), 11602-11611.
192. Yang, H.; Hofstetter, H.; Cavagnero, S., Fast-pulsing LED-enhanced NMR: A convenient and inexpensive approach to increase NMR sensitivity. *J. Chem. Phys.* **2019,** *151* (24), 245102.
193. Yang, H.; Mecha, M. F.; Goebel, C. P.; Cavagnero, S., Enhanced nuclear-spin hyperpolarization of amino acids and proteins via reductive radical quenchers. *J. Magn. Reson.* **2021,** *324*, 106912.
194. Kaptein, R.; Dijkstra, K.; Müller, F.; Van Schagen, C. G.; Visser, A. J. W. G., 360-MHz laser-induced phot-CIDNP in photoreaction of flavins. *J. Magn. Reson.* **1978,** *31* (1), 171-176.
195. Hore, P. J.; Broadhurst, R. W., Photo-CIDNP of biopolymers. *Prog. Nucl. Magn. Reson. Spectrosc.* **1993,** *25*, 345-402.
196. Lindqvist, L., A flash photolysis study of fluorescein. *Arkiv for Kemi* **1961,** *16* (1-2), 79-138.

197. Kiryutin, A. S.; Morozova, O. B.; Kuhn, L. T.; Yurkovskaya, A. V.; Hore, P. J., ^{1}H and ^{13}C Hyperfine coupling constants of the tryptophanyl cation radical in aqueous solution from microsecond time-resolved CIDNP. *J. Phys. Chem. B* **2007,** *111* (38), 11221-11227.
198. Spera, S.; Bax, A., Empirical correlation between protein backbone conformation and C^{α}. and C^{β}. ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **1991,** *113* (14), 5490-5492.
199. Adrian, F. J., Singlet-triplet splitting in diffusing radical pairs and magnitude of chemically induced electron-spin polarization. *J. Chem. Phys.* **1972,** *57* (12), 5107.
200. Ivanov, K. L.; Pravdivtsev, A. N.; Yurkovskaya, A. V.; Vieth, H. M.; Kaptein, R., The role of level anti-crossings in nuclear spin hyperpolarization. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014,** *81*, 1-36.
201. Lyon, C. E.; Lopez, J. J.; Cho, B. M.; Hore, P. J., Low field CIDNP of amino acids and proteins: characterization of transient radicals and NMR sensitivity enhancement. *Mol. Phys.* **2002,** *100* (8), 1261-1269.
202. Okuno, Y.; Cavagnero, S., Photochemically induced dynamic nuclear polarization: basic principles and applications. *eMagRes* **2017,** *6* (2), 283-313.
203. Janssen, G. J.; Daviso, E.; van Son, M.; de Groot, H. J. M.; Alia, A.; Matysik, J., Observation of the solid-state photo-CIDNP effect in entire cells of cyanobacteria synechocystis. *Photosynth. Res.* **2010,** *104* (2), 275-282.
204. Eisenreich, W.; Joshi, M.; Illarionov, B.; Kacprzak, S.; Lukaschek, M.; Kothe, G.; Budisa, N.; Fischer, M.; Bacher, A.; Weber, S., Strategy for enhancement of ^{13}C-photo-CIDNP NMR spectra by exploiting fractional ^{13}C-labeling of tryptophan. *J. Phys. Chem. B* **2015,** *119* (43), 13934-13943.
205. Thompson, C. M.; McDonald, A. D.; Yang, H.; Cavagnero, S.; Buller, A. R., Modular control of L-tryptophan isotopic substitution via an efficient biosynthetic cascade. *Org. Biomol. Chem.* **2020,** *18* (22), 4189-4192.
206. Ernst, R. R.; Bodenhausen, G.; Wokaun, A., *Principles of nuclear magnetic resonance in one and two dimensions*. Oxford University Press: New York, 1989.
207. Cavanagh, J.; Fairbrother, W.; Palmer, A. I.; M, R.; NJ, S., *Protein NMR Spectroscopy Principles and Practice*. 2nd ed.; Elsevier Academic Press: 2007.
208. Bakke, C. K.; Jungbauer, L. M.; Cavagnero, S., *In vitro* expression and characterization of native apomyoglobin under low molecular crowding conditions. *Prot. Eng. Expr. Purif.* **2006,** *45* (2), 381-92.
209. Ragavan, M.; Iconaru, L. I.; Park, C.-G.; Kriwacki, R. W.; Hilty, C., Real-time analysis of folding upon binding of a disordered protein by using dissolution DNP NMR spectroscopy. *Angew. Chem.* **2017,** *56* (25), 7070-7073.

210. Kim, J.; Mandal, R.; Hilty, C., Characterization of Membrane Protein-Lipid Interactions in Unfolded OmpX with Enhanced Time Resolution by Hyperpolarized NMR. *Chembiochem* **2020,** *21* (19), 2861-2867.
211. Kim, Y.; Liu, M. X.; Hilty, C., Parallelized Ligand Screening Using Dissolution Dynamic Nuclear Polarization. *Anal. Chem.* **2016,** *88* (22), 11178-11183.
212. Kim, Y.; Liu, M. X.; Hilty, C., Determination of binding affinities using hyperpolarized NMR with simultaneous 4-channel detection. *J. Magn. Reson.* **2018,** *295*, 80-86.
213. Kim, Y.; Wang, Y. Y.; Chen, H. Y.; Hilty, C., In-Vitro Dissolution Dynamic Nuclear Polarization for Sensitivity Enhancement of NMR with Biological Molecules. In *Protein NMR: Methods and Protocols*, Ghose, R., Ed. 2018; Vol. 1688, pp 155-168.
214. Liu, M. X.; Kim, Y.; Hilty, C., Characterization of Chemical Exchange Using Relaxation Dispersion of Hyperpolarized Nuclear Spins. *Anal. Chem.* **2017,** *89* (17), 9154-9158.
215. Wang, Y. Y.; Kim, J.; Hilty, C., Determination of protein-ligand binding modes using fast multi-dimensional NMR with hyperpolarization. *Chemical Science* **2020,** *11* (23), 5935-5943.
216. Kouřil, K.; Kouřilová, H.; Bartram, S.; Levitt, M. H.; Meier, B., Scalable dissolution-dynamic nuclear polarization with rapid transfer of a polarized solid. *Nature Comm.* **2019,** *10* (1), 1733.
217. Kaltschnee, L.; Jagtap, A. P.; McCormick, J.; Wagner, S.; Bouchard, L. S.; Utz, M.; Griesinger, C.; Glggler, S., Hyperpolarization of amino acids in water utilizing parahydrogen on a rhodium nanocatalyst. *Chem. Eur. J.* **2019,** *25* (47), 11031-11035.
218. Sellies, L.; Aspers, R.; Feiters, M. C.; Rutjes, F.; Tessari, M., Parahydrogen hyperpolarization allows direct NMR detection of alpha-amino acids in complex (bio)mixtures. *Angew. Chem.* **2021,** *60* (52), 26954-26959.
219. Kositz, C.; Schroecksnadel, K.; Grander, G.; Schennach, H.; Kofler, H.; Fuchs, D., High serum tryptophan concentration in pollinosis patients is associated with unresponsiveness to pollen extract therapy. *Int. Arch. All. Immun.* **2008,** *147* (1), 35-40.
220. Menta, R.; Mancheno-Corvo, P.; Del Rio, B.; Ramirez, C.; DelaRosa, O.; Dalemans, W.; Lombardo, E., Tryptophan concentration is the main mediator of the capacity of adipose mesenchymal stromal cells to inhibit T-lymphocyte proliferation in vitro. *Cytother.* **2014,** *16* (12), 1679-1691.
221. Weinlich, G.; Murr, C.; Richardsen, L.; Winkler, C.; Fuchs, D., Decreased serum tryptophan concentration predicts poor prognosis in malignant melanoma patients. *Dermatol.* **2007,** *214* (1), 8-14.
222. Agus, A.; Planchais, J.; Sokol, H., Gut Microbiota Regulation of Tryptophan Metabolism in Health and Disease. *Cell Host & Microbe* **2018,** *23* (6), 716-724.

223. Bosi, A.; Banfi, D.; Bistoletti, M.; Giaroni, C.; Baj, A., Tryptophan Metabolites Along the Microbiota-Gut-Brain Axis: An Interkingdom Communication System Influencing the Gut in Health and Disease. *Int. J. Tryp. Res.* **2020,** *13*, 1-25.
224. Gao, K.; Mu, C. L.; Farzi, A.; Zhu, W. Y., Tryptophan Metabolism: A Link Between the Gut Microbiota and Brain. *Adv. Nutr.* **2020,** *11* (3), 709-723.
225. Gostner, J. M.; Geisler, S.; Stonig, M.; Mair, L.; Sperner-Unterweger, B.; Fuchs, D., Tryptophan Metabolism and Related Pathways in Psychoneuroimmunology: The Impact of Nutrition and Lifestyle. *Neuropsychobiol.* **2020,** *79* (1), 89-99.
226. Kokturk, O.; Kanbay, A., Tryptophan Metabolism and Sleep. In *Tryptophan Metabolism: Implications for Biological Processes, Health and Disease*, Engin, A.; Engin, A. B., Eds. 2015; pp 239-252.
227. Konopelski, P.; Ufnal, M., Indoles - Gut Bacteria Metabolites of Tryptophan with Pharmacotherapeutic Potential. *Curr. Drug Met.* **2018,** *19* (10), 883-890.
228. Roth, W.; Zadeh, K.; Vekariya, R.; Ge, Y.; Mohamadzadeh, M., Tryptophan Metabolism and Gut-Brain Homeostasis. *International Journal of Molecular Sciences* **2021,** *22* (6), 1877-2013.
229. Broos, J., Biosynthetic incorporation of tryptophan analogs in proteins. In *Fluorescence spectroscopy and microscopy - methods and protocols*, Engelborghs, Y.; Visser, A. J. W. G., Eds. Humana Press: New York, 2014; pp 359-370.
230. Pordes, R.; Petravick, D.; Kramer, B.; Olson, D.; Livny, M.; Roy, A.; Avery, P.; Blackburn, K.; Wenaus, T.; Wuerthwein, F.; Foster, I.; Gardner, R.; Wilde, M.; Blatecky, A.; McGee, J.; Quick, R. In *The open science grid*, 3rd Annual Scientific Discovery through Advanced Computing Conference (SciDAC 2007), Boston, MA, Jun 24-28; Boston, MA, 2007.
231. Sfiligoi, I.; Bradley, D. C.; Holzman, B.; Mhashilkar, P.; Padhi, S.; Wurthwein, F., The Pilot Way to Grid Resources Using glideinWMS. *WRI World Congress on Computer Science and Information Engineering* **2009,** *2*, 428-432.
232. Casalini, T.; Salvalaglio, M.; Perale, G.; Masi, M.; Cavallotti, C., Diffusion and Aggregation of Sodium Fluorescein in Aqueous Solutions. *J. Phys. Chem. B* **2011,** *115* (44), 12896-12904.
233. Longsworth, L. G., Diffusion Measurements, at 25°, of Aqueous Solutions of Amino Acids, Peptides and Sugars. *J. Am. Chem. Soc.* **1953,** *75* (22), 5705-5709.
234. Togashi, D. M.; Szczupak, B.; Ryder, A. G.; Calvet, A.; O'Loughlin, M., Investigating Tryptophan Quenching of Fluorescein Fluorescence under Protolytic Equilibrium. *J. Phys. Chem. A* **2009,** *113* (12), 2757-2767.
235. Shigeya, N.; Yoshiharu, S.; Shiro, K.; Hiroshi, K., ESR study on fluorescein semiquinone radical. *Bull. Chem. Soc. Jpn.* **1974,** *47* (9), 2121-2125.

236. Aguilar, J. A.; Nilsson, M.; Bodenhausen, G.; Morris, G. A., Spin echo NMR spectra without J modulation. *Chem. Commun.* **2012,** *48* (6), 811-813.
237. Carr, H. Y.; Purcell, E. M., Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.* **1954,** *94* (3), 630-638.
238. Meiboom, S.; Gill, D., Modified spin-echo method for measuring nuclear relaxatn times. *Rev. Sci. Instrum.* **1958,** *29* (8), 688-691.
239. Buller, A. R.; van Roye, P.; Cahn, J. K. B.; Scheele, R. A.; Herger, M.; Arnold, F. H., Directed evolution mimics allosteric activation by stepwise tuning of the conformational ensemble. *J. Am. Chem. Soc.* **2018,** *140* (23), 7256-7266.
240. Fasman, G. D., Handbook of biochemistry and molecular biology. CRC Press: Cleveland (OH), 1976; Vol. 1.
241. Boyer, R. D.; Johnson, R.; Krishnamurthy, K., Compensation of refocusing inefficiency with synchronized inversion sweep (CRISIS) in multiplicity-edited HSQC. *J. Magn. Reson.* **2003,** *165* (2), 253-259.
242. Matsui, T.; Oshiyama, A.; Shigeta, Y., A Simple scheme for estimating the pK_a values of 5-substituted uracils. *Chem. Phys. Lett.* **2011,** *502* (4), 248-252.
243. Matsui, T.; Shigeta, Y.; Morihashi, K., Assessment of methodology and chemical group dependences in the calculation of the pK_a for several chemical groups. *J. Theory Comp.* **2017,** *13* (10), 4791-4803.
244. Ditchfield, R., Self-consistent perturbation theory of diamagnetism. *Mol. Phys.* **1974,** *27* (4), 789-807.
245. Schreckenbach, G.; Ziegler, T., Calculation of the G-tensor of electron paramagnetic resonance spectroscopy using gauge-including atomic orbitals and density functional theory. *J. Phys. Chem. A* **1997,** *101* (18), 3388-3399.
246. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum mechanical continuum solvation models. *Chem. Rev.* **2005,** *105* (8), 2999-3094.
247. Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M., UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992,** *114* (25), 10024-10035.
248. Frisch, M.; Trucks, G.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G., Gaussian 09, Revision d. 01, Gaussian. *Inc., Wallingford CT* **2009,** *201*.
249. Connor, H. D.; Sturgeon, B. E.; Mottley, C.; Sipe, H. J. J.; Mason, R. P., L--Tryptophan Radical Cation Electron Spin Resonance Studies: Connecting Solution-Derived Hyperfine Coupling Constants with Protein Spectral Interpretations. *J. Am. Chem. Soc.* **2008,** *130*, 6381-6387.

3.12. Supplementary Tables

Supplementary Table S3.1. Hyperfine coupling constants used for computing triplet-singlet mixing frequencies.

Specific hyperfine coupling constant	Value (mT)	Ref.	Specific hyperfine coupling constant	Value (mT)	Ref.
A ($Trp^{\bullet+}$ $^{13}C^{\alpha}$)	0.5643	197 [a]	A ($Trp^{\bullet+}$ $^{13}C^{\eta 2}$)	0.443	197
A ($Trp^{\bullet+}$ $^{13}C^{\beta}$)	-0.512	197	A ($Trp^{\bullet+}$ $^{1}H^{\alpha}$)	0	197
A ($Trp^{\bullet+}$ $^{13}C'$)	-0.045	197	A ($Trp^{\bullet+}$ $^{1}H^{\beta 1}$)	2.544	197
A ($Trp^{\bullet+}$ $^{13}C^{\delta 1}$)	-0.227	197	A ($Trp^{\bullet+}$ $^{1}H^{\beta 2}$)	1.189	197
A ($Trp^{\bullet+}$ $^{13}C^{\gamma}$)	1.254	197	A ($Trp^{\bullet+}$ $^{1}H^{\delta 1}$)	-0.421	197
A ($Trp^{\bullet+}$ $^{13}C^{\delta 2}$)	-0.891	197	A ($Trp^{\bullet+}$ $^{1}H^{\varepsilon 1}$)	-0.413	197
A ($Trp^{\bullet+}$ $^{13}C^{\varepsilon 2}$)	0.092	197	A ($Trp^{\bullet+}$ $^{1}H^{\varepsilon 3}$)	-0.504	197
A ($Trp^{\bullet+}$ $^{13}C^{\varepsilon 3}$)	0.619	197	A ($Trp^{\bullet+}$ $^{1}H^{\zeta 2}$)	-0.050	197
A ($Trp^{\bullet+}$ $^{13}C^{\zeta 2}$)	-0.182	197	A ($Trp^{\bullet+}$ $^{1}H^{\zeta 3}$)	0.124	197
A ($Trp^{\bullet+}$ $^{13}C^{\zeta 3}$)	-0.499	197	A ($Trp^{\bullet+}$ $^{1}H^{\eta 2}$)	-0.412	197
A ($Fl^{\bullet-}$ ^{1}H 1,8)	0.3287	235	A ($Fl^{\bullet-}$ ^{1}H 14)	0.0190	235
A ($Fl^{\bullet-}$ ^{1}H 2,7)	0.1510	235	A ($Fl^{\bullet-}$ ^{1}H 15)	0.0170	235
A ($Fl^{\bullet-}$ ^{1}H 4,5)	0.0889	235	A ($Fl^{\bullet-}$ ^{1}H 16)	0.0090	235
A ($Fl^{\bullet-}$ ^{1}H 13)	0.022	235	A ($Trp^{\bullet+}$ $^{14}N^{\varepsilon 1}$)	0.20	249
			A ($Trp^{\bullet+}$ $^{15}N^{\varepsilon 1}$)	0.28	249

[a] Scaled according to the relative hyperfine coupling constant of $^{13}C^{\alpha}$ and $^{13}C^{\gamma}$ reported in the reference and DFT-simulated hyperfine coupling constant of $^{13}C^{\gamma}$ in aqueous medium.

Supplementary Table S3.2. Cartesian coordinates, total energies (E), Gibbs free energies (G) and g-factors (g_{iso}) for fluorescein radicals employed in the DFT calculations. All calculations were carried out assuming 1 atm and 298.15 K, therefore G values are equivalent to G^0.

$Fl^{\cdot}$ (neutral radical)

O	-1.100495	-0.561994	2.328877
O	-3.249971	-1.215315	2.387301
O	-1.116715	5.319985	-0.308815
O	5.456917	-1.344209	0.359217
O	1.958609	1.779772	0.027386
C	4.112832	-1.137234	0.192440
C	3.194061	-2.186023	0.042194
C	1.842805	-1.901819	-0.126025
C	1.349562	-0.573205	-0.158120
C	2.317609	0.452046	0.012891
C	3.670267	0.190511	0.180965
C	-0.030974	-0.218594	-0.317952
C	-0.376923	1.174501	-0.320551
C	0.643594	2.145571	-0.143630
C	-1.686378	1.681018	-0.512608
C	-1.961141	3.044726	-0.511081
C	-0.919413	3.964106	-0.320982
C	0.391880	3.510687	-0.138799
C	-1.062126	-1.264547	-0.558974
C	-1.083150	-1.892397	-1.816081
C	-2.046098	-2.848537	-2.143823
C	-3.028861	-3.196173	-1.213153
C	-3.031878	-2.586743	0.039563
C	-2.054137	-1.635287	0.384046
C	-2.052322	-1.077499	1.765897
H	3.534651	-3.218519	0.060776
H	1.143376	-2.724202	-0.234859
H	4.372192	1.008927	0.304589
H	-2.502876	0.983451	-0.669466
H	-2.979991	3.394624	-0.659157
H	1.206320	4.214343	0.000352
H	-0.334701	-1.609579	-2.551282
H	-2.032173	-3.312644	-3.126366
H	-3.785516	-3.935584	-1.459036
H	-3.785846	-2.857802	0.769972
H	-2.056640	5.522451	-0.448615
H	5.652336	-2.295947	0.350028
H	-3.153136	-0.866080	3.294035

E(UB3LYP/6-31+G(d)) = -1146.12842651 a.u.

G(UB3LYP/6-31+G(d)) = -1145.895876 a.u.

g_{iso}(UB3LYP/EPR-II) = 2.00306

$Fl^{\cdot -}$(-H^a) (anion radical, H^a is deprotonated)

O	-0.890153	-1.065936	2.500523
O	-3.131900	-1.056092	2.226870
O	-1.994237	5.034456	-0.348863
O	5.597520	-0.447912	0.323807
O	1.625648	2.053342	0.010574
C	4.234654	-0.466250	0.169586
C	3.500534	-1.652780	0.034027
C	2.118844	-1.597046	-0.120873
C	1.410777	-0.369413	-0.154673
C	2.198857	0.802202	-0.000683
C	3.578076	0.769822	0.155274
C	-0.012486	-0.254161	-0.300626
C	-0.579596	1.064818	-0.314967
C	0.266989	2.194667	-0.157324
C	-1.955260	1.347804	-0.508388
C	-2.450600	2.648090	-0.523441
C	-1.575027	3.728587	-0.347891
C	-0.205982	3.499703	-0.167603
C	-0.863986	-1.459100	-0.504433
C	-0.769909	-2.141580	-1.732786
C	-1.553675	-3.263185	-2.009943
C	-2.470708	-3.714407	-1.055590
C	-2.590822	-3.034672	0.157753
C	-1.791321	-1.919861	0.459885
C	-1.946930	-1.281486	1.842508
H	4.007272	-2.614615	0.053668
H	1.562430	-2.523391	-0.216689
H	4.135604	1.694202	0.267877
H	-2.643247	0.521154	-0.650356
H	-3.513138	2.824019	-0.673392
H	0.481845	4.329720	-0.041178
H	-0.071838	-1.776355	-2.482745
H	-1.453988	-3.773064	-2.964975
H	-3.090718	-4.584804	-1.256495
H	-3.312650	-3.370041	0.897642
H	5.945025	-1.355339	0.323588

H	-2.956197	5.075136	-0.479116

E(UB3LYP/6-31+G(d)) = -1145.67122707 a.u.
G(UB3LYP/6-31+G(d)) = -1145.450992 a.u.
g_{iso}(UB3LYP/EPR-II) = 2.00307

Fl$^{\bullet-}$(-H^{b}) (anion radical, H^{b} is deprotonated)

O	-1.260899	-0.485937	2.364001
O	-3.466975	-0.843494	2.125359
O	-1.378461	5.238389	-0.449632
O	5.540935	-1.095948	0.442016
O	1.904253	1.855609	-0.011791
C	4.184628	-0.956201	0.259138
C	3.315169	-2.047446	0.144880
C	1.951397	-1.828254	-0.042183
C	1.395608	-0.526877	-0.134261
C	2.317109	0.546157	0.011539
C	3.680121	0.347785	0.197745
C	-0.002535	-0.234600	-0.309238
C	-0.407050	1.140876	-0.345566
C	0.564486	2.169242	-0.191856
C	-1.734467	1.604434	-0.568185
C	-2.063471	2.948305	-0.602882
C	-1.079747	3.986066	-0.422908
C	0.263755	3.517960	-0.220444
C	-0.980797	-1.331290	-0.507611
C	-0.880330	-2.122536	-1.669014
C	-1.808414	-3.119547	-1.967954
C	-2.888038	-3.351581	-1.108336
C	-3.012584	-2.588290	0.049670
C	-2.063763	-1.599209	0.371620
C	-2.182625	-0.916601	1.689847
H	3.698811	-3.063415	0.207804
H	1.292258	-2.687108	-0.114836
H	4.342654	1.201776	0.299876
H	-2.521400	0.870533	-0.722678
H	-3.094927	3.249945	-0.776233
H	1.064185	4.243373	-0.094799
H	-0.065845	-1.924798	-2.360624
H	-1.698499	-3.702402	-2.878901
H	-3.622317	-4.118858	-1.336613
H	-3.837778	-2.769547	0.730021
H	5.774689	-2.038360	0.468586
H	-3.453369	-0.440455	3.014757

E(UB3LYP/6-31+G(d)) = -1145.66006120 a.u.
G(UB3LYP/6-31+G(d)) = -1145.440251 a.u.
g_{iso}(UB3LYP/EPR-II) = 2.00343

Fl$^{\bullet-}$(-H^{c}) (anion radical, H^{c} is deprotonated)

O	-1.287650	-0.459824	2.340654
O	-3.476594	-0.915087	2.120742
O	-1.199079	5.247515	-0.439035
O	5.547114	-1.267873	0.458031
O	1.970408	1.797907	-0.003611
C	4.288546	-1.057213	0.288701
C	3.329855	-2.127340	0.173164
C	1.978946	-1.892078	-0.012261
C	1.431619	-0.580888	-0.113212
C	2.378784	0.472745	0.027757
C	3.732170	0.265934	0.213377
C	0.041361	-0.275787	-0.283571
C	-0.349270	1.110602	-0.319726
C	0.645236	2.115316	-0.175723
C	-1.667275	1.583740	-0.538328
C	-1.979216	2.941257	-0.580712
C	-0.965540	3.892112	-0.412103
C	0.354627	3.474704	-0.213157
C	-0.950501	-1.354812	-0.500902
C	-0.834126	-2.146931	-1.660966
C	-1.768715	-3.129270	-1.984121
C	-2.871019	-3.348976	-1.149314
C	-3.010027	-2.588755	0.008128
C	-2.055981	-1.611711	0.355434
C	-2.193395	-0.934467	1.673358
H	3.698117	-3.149163	0.244992
H	1.303020	-2.740365	-0.078234
H	4.392805	1.124115	0.311815
H	-2.467959	0.865967	-0.685753
H	-3.006413	3.257349	-0.748035
H	1.152476	4.201589	-0.095805
H	-0.000354	-1.959365	-2.331746
H	-1.646368	-3.710856	-2.894308
H	-3.610546	-4.105264	-1.396562
H	-3.851464	-2.762513	0.669962
H	-2.143015	5.416657	-0.593836
H	-3.470369	-0.508330	3.008338

E(UB3LYP/6-31+G(d)) = -1145.66019548 a.u.
G(UB3LYP/6-31+G(d)) = -1145.440367 a.u.
g_{iso}(UB3LYP/EPR-II) = 2.00346

$Fl^{\cdot 2-}$(-H^a, -H^b) (dianion radical, H^a and H^b are deprotonated)

O	-1.047034	-1.075935	2.559268
O	-3.235251	-0.690743	2.158858
O	-1.644786	5.198063	-0.379719
O	5.542942	-0.866581	0.330930
O	1.779121	1.947399	0.016070
C	4.176815	-0.777725	0.175184
C	3.351331	-1.899313	0.040239
C	1.975467	-1.731456	-0.116368
C	1.364378	-0.451593	-0.152780
C	2.246194	0.654579	0.002543
C	3.618824	0.507037	0.159208
C	-0.044772	-0.216111	-0.297933
C	-0.503297	1.141964	-0.312695
C	0.425891	2.208821	-0.157652
C	-1.851993	1.549605	-0.511280
C	-2.237457	2.880346	-0.532642
C	-1.296048	3.956163	-0.361019
C	0.068372	3.543808	-0.173337
C	-0.986792	-1.352253	-0.503978
C	-0.937170	-2.047075	-1.729335
C	-1.798846	-3.108352	-2.012624
C	-2.755956	-3.488043	-1.066137
C	-2.833850	-2.797273	0.144545
C	-1.955007	-1.744261	0.451665
C	-2.082398	-1.102270	1.834451
H	3.778535	-2.899692	0.060930
H	1.347824	-2.611621	-0.211440
H	4.249099	1.383969	0.272137
H	-2.605381	0.778329	-0.647173
H	-3.283671	3.139639	-0.686733
H	0.838085	4.302126	-0.048125
H	-0.207142	-1.736989	-2.473996
H	-1.729163	-3.626025	-2.966437
H	-3.438206	-4.309850	-1.270678
H	-3.585298	-3.077611	0.878451
H	5.813275	-1.799483	0.329482

E(UB3LYP/6-31+G(d)) = -1145.20048714 a.u.
G(UB3LYP/6-31+G(d)) = -1144.993028 a.u.
g_{iso}(UB3LYP/EPR-II) = 2.00338

$Fl^{\cdot 2-}$(-H^a, -H^c) (dianion radical, H^a and H^c are deprotonated)

O	-1.033793	-1.114181	2.563555
O	-3.221202	-0.814938	2.092020
O	-1.773075	5.103450	-0.373200
O	5.617785	-0.710754	0.344939
O	1.738377	1.989269	0.017246
C	4.338158	-0.625133	0.205794
C	3.490349	-1.781989	0.080378
C	2.117421	-1.680086	-0.076732
C	1.439171	-0.430046	-0.131497
C	2.278684	0.709738	0.014954
C	3.650273	0.636660	0.171375
C	0.022697	-0.264300	-0.281903
C	-0.500563	1.073432	-0.308524
C	0.387174	2.175122	-0.155132
C	-1.862598	1.409834	-0.517600
C	-2.312011	2.729752	-0.542745
C	-1.401179	3.776918	-0.365145
C	-0.042096	3.496468	-0.175762
C	-0.868991	-1.438752	-0.491996
C	-0.755793	-2.155496	-1.700712
C	-1.568704	-3.253172	-1.988351
C	-2.540213	-3.650523	-1.063811
C	-2.681109	-2.939957	0.129490
C	-1.852001	-1.848840	0.441602
C	-2.044921	-1.189487	1.808751
H	3.961364	-2.763003	0.115210
H	1.527019	-2.588757	-0.160543
H	4.223049	1.555539	0.275427
H	-2.580969	0.609094	-0.658093
H	-3.367024	2.942024	-0.701900
H	0.675367	4.301806	-0.050552
H	-0.014556	-1.832824	-2.428439
H	-1.450559	-3.785859	-2.929033
H	-3.184903	-4.501209	-1.272088
H	-3.443561	-3.234651	0.846328
H	-2.732455	5.174227	-0.507389

E(UB3LYP/6-31+G(d)) = -1145.20065790 a.u.
G(UB3LYP/6-31+G(d)) = -1144.992835 a.u.
g_{iso}(UB3LYP/EPR-II) = 2.00340

$Fl^{\cdot 2-}$(-H^b, -H^c) (dianion radical, H^b and H^c are deprotonated)

O	-1.561907	-0.140467	2.247366
O	-3.707933	-0.771107	2.027936
O	-0.963318	5.302758	-0.525826
O	5.499348	-1.509164	0.503575

O	2.080546	1.714581	-0.043631
C	4.249751	-1.238173	0.336310
C	3.234263	-2.260133	0.282851
C	1.895961	-1.961115	0.099638
C	1.413746	-0.630853	-0.067875
C	2.413722	0.376573	0.027260
C	3.760217	0.103953	0.210107
C	0.030119	-0.262348	-0.235755
C	-0.281423	1.145223	-0.284740
C	0.762718	2.104845	-0.190706
C	-1.572778	1.699813	-0.508098
C	-1.807011	3.060168	-0.593153
C	-0.749769	4.031783	-0.462247
C	0.555316	3.474605	-0.266052
C	-0.992492	-1.295087	-0.446170
C	-0.807213	-2.201052	-1.520183
C	-1.777856	-3.117163	-1.903076
C	-3.000716	-3.179637	-1.205809
C	-3.203101	-2.340508	-0.121628
C	-2.212638	-1.417967	0.299673
C	-2.414944	-0.713863	1.575692
H	3.546823	-3.295708	0.407766
H	1.177445	-2.775892	0.093932
H	4.462945	0.932096	0.271018
H	-2.414319	1.022286	-0.623656
H	-2.817926	3.427494	-0.762921
H	1.410400	4.142335	-0.186641
H	0.109037	-2.126829	-2.099029
H	-1.601588	-3.763440	-2.759529
H	-3.774234	-3.880917	-1.506817
H	-4.129105	-2.403497	0.439192
H	-3.718598	-0.325116	2.895107

E(UB3LYP/6-31+G(d)) = -1145.18697715 a.u.
G(UB3LYP/6-31+G(d)) = -1144.979506 a.u.
g_{iso}(UB3LYP/EPR-II) = 2.00351

C	1.381204	-0.555185	-0.130684
C	2.334942	0.490182	0.016912
C	3.693411	0.268131	0.182820
C	-0.014906	-0.237072	-0.277195
C	-0.390118	1.152043	-0.302346
C	0.606204	2.157654	-0.159997
C	-1.704950	1.649386	-0.516622
C	-2.003829	3.005845	-0.557012
C	-0.998137	4.017915	-0.392783
C	0.332064	3.516234	-0.195894
C	-1.020661	-1.312630	-0.485323
C	-0.954257	-2.070096	-1.674876
C	-1.873474	-3.078256	-1.969016
C	-2.914312	-3.343834	-1.072378
C	-3.008714	-2.595637	0.102475
C	-2.070569	-1.599043	0.424504
C	-2.225503	-0.921563	1.787434
H	3.647410	-3.140641	0.137681
H	1.248156	-2.711016	-0.151224
H	4.358822	1.122883	0.288377
H	-2.509896	0.931280	-0.650149
H	-3.031549	3.328216	-0.720884
H	1.151705	4.222934	-0.080296
H	-0.161485	-1.846752	-2.385529
H	-1.785356	-3.641317	-2.895546
H	-3.644824	-4.120532	-1.286879
H	-3.820109	-2.788052	0.800171

E(UB3LYP/6-31+G(d)) = -1144.72364361 a.u.
G(UB3LYP/6-31+G(d)) = -1144.529698 a.u.
g_{iso}(UB3LYP/EPR-II) = 2.00358

$FI^{\cdot 3-}$ (trianion radical)

O	-1.227879	-0.941179	2.564402
O	-3.368602	-0.444474	2.054834
O	-1.263532	5.286997	-0.426243
O	5.516696	-1.271533	0.381723
O	1.939582	1.819201	0.016537
C	4.247314	-1.055119	0.227548
C	3.285718	-2.113832	0.097058
C	1.928074	-1.866639	-0.068916

3.13. Supplementary Figure Legends

Supplementary Figure S3.1. NMR pulse sequence used for $^1H^\alpha$ T_2 measurements. $^1H^\alpha$ T_2 values of different Trp isotopologs were measured with a pulse sequence employing a perfect-echo CPMG-like pulse scheme [236] including solvent pre-saturation. The duration of each CPMG-like block was 50 ms, with the delay τ set to 12.5 ms. To avoid artifacts, it was important to keep the τ value fairly short. The different time points for the T_2 measurements (see Fig. S2) were obtained upon varying the number of CPMG-like iterations n.

Supplementary Figure S3.2. Results of $^1H^\alpha$ T_2 measurements. Data were collected with the pulse sequence in Figure S1. As part of the data processing, areas of individual $^1H^\alpha$ resonances were determined and plotted as a function of the total CPMG-like duration. The data were fit to a single exponential decay to determine the transverse relaxation time constant T_2. The resulting T_2 values of each isotopolog are annotated on the plot. Data are displayed as avg ± S.E. (n=3). A 121.5 Hz field solvent-presaturation scheme was applied during the entire 5 s recycle delay.

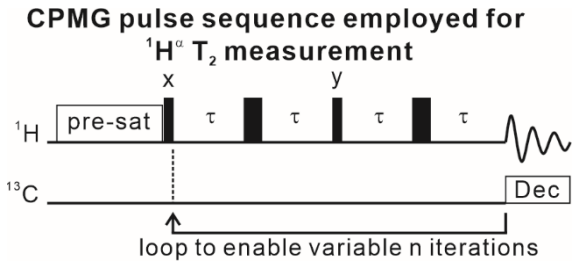
Supplementary Figure S3.3. LC-photo-CIDNP spectra of 5 μM Trp isotopologs. (a) LC-photo-CIDNP spectra of Trp isotopologs acquired with the 1D ^{13}C RASPRINT pulse sequence under both light (LED-on) and dark (LED-off) conditions. The concentration of each sample was 5 μM. (b) Analysis of normalized H^α areas and intensities-at-the-maximum under light (LED-on) conditions. Values were normalized relative to the uniformly labeled Trp-U-^{13}C,^{15}N isotopolog.

Supplementary Figure S4. Quantitative evaluations of light-on and light-off effects. (a) Schematic representation of the effects responsible for the observed LC-photo-CIDNP sensitivity enhancements under light (LED-on) and dark (LED-off) conditions. (b) Block diagram representing the specific experiments and computations carried out to quantitatively

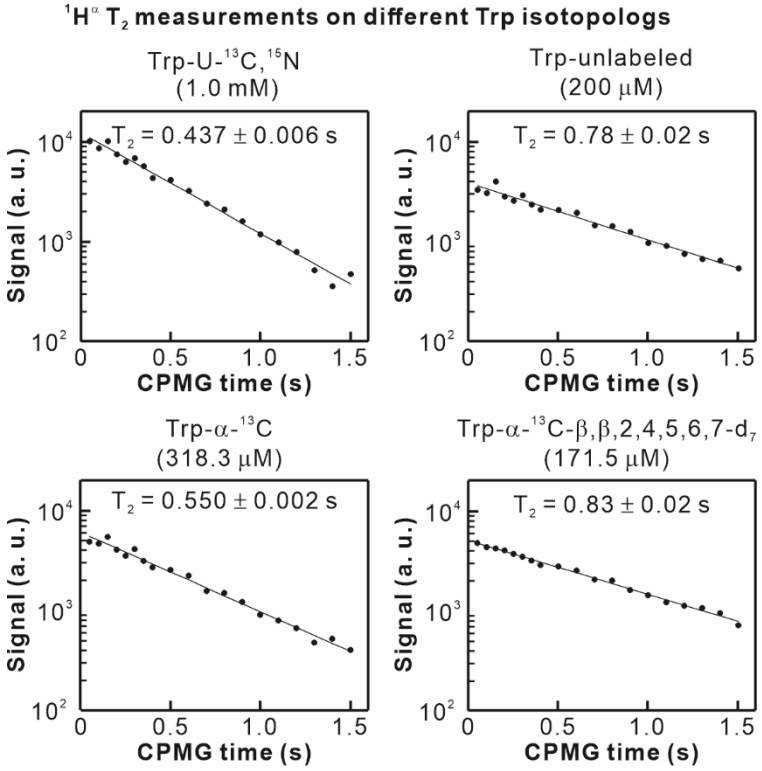
assess the contribution of LED-on and LED-off effects to the observed nuclear-spin hyperpolarization.

3.14. Supplementary Figures

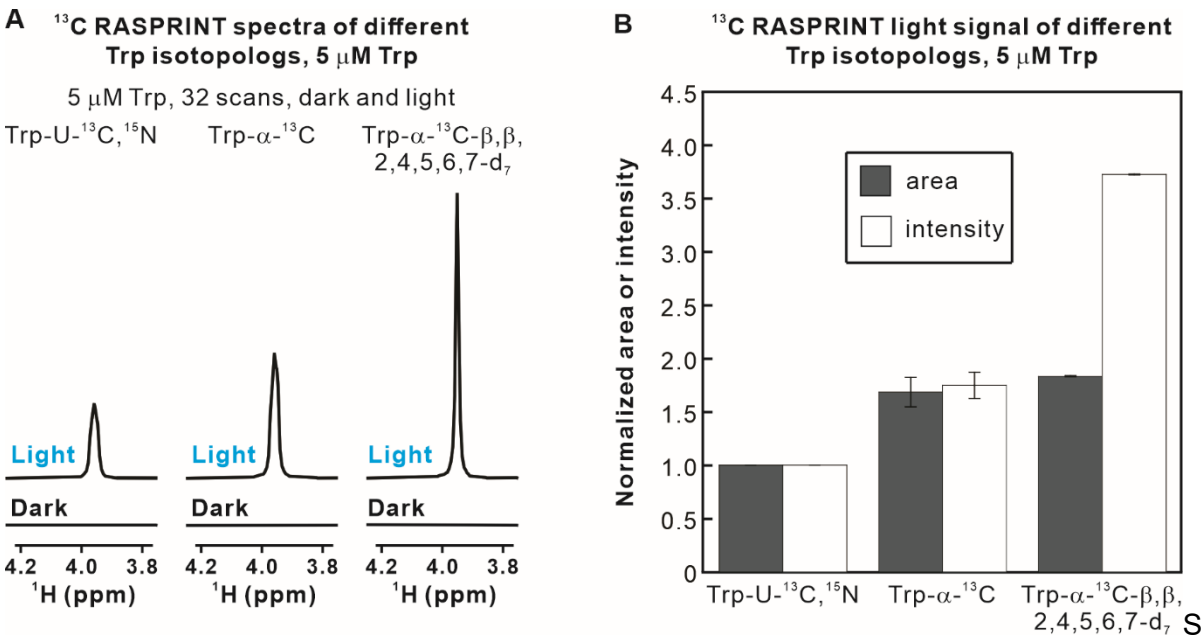
Supplementary Figure S3.1



Supplementary Figure S3.2



Supplementary Figure S3.3



Supplementary Figure S3.4