Uncovering the Properties and Binding Mechanisms of Artificial ATP-Binding Proteins Made from Primordial, Reduced Amino Acid Alphabets

Alexander D. Soltau

Submitted under the supervision of Dr. Burckhard Seelig to the University Honors Program at the University of Minnesota-Twin Cities in partial fulfillment of the requirements for the degree of Bachelor of Science, summa cum laude in Biochemistry.

ABSTRACT:

In early life forms, the first proteins were likely composed of a limited set of amino acid building blocks. How these simplistic proteins may have functioned and conferred a selective advantage to early cells remains relatively unknown. Previous work in the Seelig lab has generated random libraries of 85 amino acid-long proteins and isolated variants with adenosine triphosphate (ATP) binding affinity, made from varying subsets of the 20 modern amino acids. By characterizing these proteins' binding mechanisms and properties, we hope to gain insight into the nature of early functional proteins. Here, we took initial steps to analyze how the presence of specific metal ions affected the ATP binding of such proteins and what metal ions might be directly bound. Our preliminary data suggest that magnesium and manganese are required for ATP binding for some of our model primordial proteins composed of just five amino acid types. If confirmed, this would indicate that early proteins exhibited metal preferences akin to those of modern nucleotide-binding proteins. Ongoing work will expand these analyses to additional protein variants, further exploring how metal ion interactions shaped early protein function and drawing comparisons to extant natural proteins.

INTRODUCTION:

How did the first single-celled organism arise from a soup of abiotic organic molecules? A question whose answer resides more than 3.5 billion years ago may seem challenging to answer. However, by using models of *primordial* proteins, we may be able to figure out a small piece of the giant puzzle that is the origin of life. At the heart of the origin of life debate is the question of how the first replicating life began. Some theories focus on an RNA-world hypothesis, where an early RNA-based replication system was encapsulated by a lipid membrane to form a protocell (Joyce & Szostak, 2018)—notably with an initial lack of protein. Other theories may or may not include proteins at the origin of life and focus on other elements of the first protocells, such as a membrane or metabolism first (Świeżyński, 2016). Regardless of how protocells initially formed and gained replicative abilities, proteins—essential components of all modern life—were certainly integrated into the cell before the emergence of the last universal common ancestor (LUCA). Still, it remains unclear how the earliest *functional* proteins were produced and how they initially functioned in conjunction within early cells. My work aimed to study model primordial proteins with ATP binding affinity to help understand how the first proteins may have utilized divalent metal ions to aid their functions.

Where did the first proteins come from? The Miller-Urey experiment (Miller, 1953) provided the earliest evidence of the production of 5 types of amino acids, the basic building blocks of proteins, under simulated possible early Earth conditions. Since then, the specific conditions of the experiment have been revised, but updated versions nonetheless resulted in the abiotic synthesis of many building blocks of life, including amino acids for proteins, sugars, lipid-like molecules, and nucleotides (Shaw, 2016). Vials from Miller's original experiment were also re-analyzed using modern methods to identify nine additional amino acids in the original

experiment. Re-analyzed samples of simulated conditions of a volcanic lightning storm revealed a total of 22 amino acids. In addition to arising from the early Earth itself, amino acids and other essential building blocks from life may have accumulated from extraterrestrial sources. Notably, carbonaceous meteorites have been found to contain several amino acids and nitrogenous bases (Kvenvolden et al., 1970; Oba et al., 2023). Regardless of their precise origin, it is clear that amino acids were abiotically synthesized.

In aqueous solutions, a small portion of amino acids will spontaneously polymerize at equilibrium, though this process alone cannot produce many peptides longer than dimers (Danger et al., 2012). Peptide formation has been demonstrated in high temperatures >200 °C and in dehydrating conditions during dry and wet cycling with simultaneous heating and cooling (Campbell et al., 2019; Imai et al., 1999). While prebiotically plausible, such conditions generate low yields of <1% of oligomers longer than ~9 amino acids. Extreme temperatures and dry-wet cycling could have also been stressful conditions for early life (Campbell et al., 2019; Danger et al., 2012). Peptide synthesis could have been promoted by numerous activating agents, such as cyanamide and trimetaphosphate, as well as many metal ions and prebiotically plausible gases—allowing synthesis in less stressful conditions (Frenkel-Pinter et al., 2020; Ponnamperuma & Peterson, 1965; Rabinowitz et al., 1969).

Once synthesized, model early peptides composed of small numbers of amino acids have been seen to stabilize RNA and membrane-less protocells and can even serve functional catalytic roles (Cornell et al., 2019; Despotovic & Tawfik, 2021; Van Der Gulik & Speijer, 2015; Wieczorek et al., 2017). Peptides of increasing length may have self-assembled into repeating amyloid structures. Alternatively, spontaneous separation of peptides into bubble-like coacervates may have enabled early cell structures and compartmentalization (Despotovic &

Tawfik, 2021). Phenomena such as these could have provided selective pressure for the addition of proteins to the macromolecular content of early life. Today, proteins serve as the main functional players in almost every aspect of modern life. How early functional proteins evolved into the life we know today is one of the key questions in understanding not just primordial life, but all of life as we know it. In a step towards solving this, recent work has evaluated which amino acids primordial proteins may have been composed of.

Evolution of the Genetic Code

Today, the universal genetic code provides a system for the translation of genetic DNA/RNA sequences into proteins composed of the 20 *canonical* amino acids. However, this was not always the case (Trifonov, 2004). The genetic code is not stagnant. Rather, it appears to have continuously changed over time. We can see this change by looking at various clues, one such clue being the composition of amino acids in proteins that are conserved across domains of life (Trifonov, 2004). This change continues in recent history, as newly evolved 21st or 22nd genetically encoded amino acids have been discovered in nature where redundant stop codons are reassigned to encode for selenocysteine or pyrrolysine (Turanov et al., 2011; Zhang & Gladyshev, 2007).

Trifonov (2004) put forth a consensus order for the addition of amino acids to the genetic code through a comprehensive meta-analysis of over 60 criteria ranging from metabolic pathways and chemical complexity to genetic code analysis and tRNA evolutionary reconstruction. This most plausible consensus order provides a useful model for experimental work on early protein evolution by allowing the construction of proteins that may more accurately reflect the amino acid diversity of early life.

The Protein World

Proteins, which are composed of chains of amino acid building blocks, are distinct from one another by their *primary sequence*—the order of these building blocks. In most, but evidently not all, modern life, there are 20 types of amino acids. It is the unique ordering of these amino acids into long chains that produces the vast diversity of functional proteins we observe. For a protein of 300 amino acids, which is roughly average across the domains of life (Nevers et al., 2023), there are 20^{300} (or about $\sim 2*10^{390}$) possible unique primary sequences. This is **far** more than the total estimated number of protons and electrons in the observable universe combined (Eddington, 1944). One may reason that this incomprehensibly large possible sequence space of proteins could not have been fully explored by the limited number of molecules in existence on the planet. To support this statement, let us consider the opposite. Some have argued, rather, that the functional space of protein folds is much smaller and has been searched throughout evolutionary history (Dryden et al., 2008). This notion indeed has merit, as evidence suggests that 90% of potentially functional random sequence proteins fail to properly fold (Keefe & Szostak, 2001). However, based on cumulative gross primary productivity since the origin of life, it is estimated that the cumulative number of cells to ever exist on Earth is between 10^{39}-10^{40} (Crockford et al., 2023). Even if each cell had hundreds of thousands of genes, this would still not even approach the possible number of protein sequences. The same argument stands even if only the estimated 10% properly folded protein sequences are considered. Thus, much of the possible protein sequence space is still unexplored.

Protein Screening and Selection

To probe this unexplored protein space, we can sample protein *libraries*—composed of many different sequences of individual proteins—for variants with a particular function or property. A wide variety of methods for testing such libraries exist, all with varying throughput and unique limitations. Rather than screening each variant one at a time, higher throughput selection methods are commonly used.

One such selection technique is mRNA display. The mRNA display technique creates a link between the protein and its corresponding mRNA sequence during protein translation. During translation, the ribosome is stalled by a DNA linker region attached to a translation inhibitor called puromycin. Puromycin enters the ribosome and is covalently connected to the peptide chain at its 3' end. Since the puromycin remains connected to the mRNA encoding the same protein, it forms a stable protein-mRNA linkage (Keefe, 2001). With each protein linked to its genetic material, many variants can be tested in parallel. To detect binding affinity, a single test tube of ~10^{12}-10^{14} unique mRNA-displayed proteins can be applied to a gravity column filled with the target ligand immobilized within a resin (Newton et al., 2020). Proteins that bind to the ligand will stay in the column even as it is washed with buffer solution. Bound proteins are then separately eluted off the column. Each variant's linked mRNA can then be reverse-transcribed into complementary DNA (cDNA), which is easily sequenced. These sequencing results can be used to identify the successful protein variants.

Work by Keefe & Szostak (2001) utilized mRNA display to explore the frequency of ATP binding within a pool of completely random sequences. ATP, often called the *energy currency of life*, is a common binding target for essential proteins in modern life. They found that roughly 1 in 10^{11} sequences have ATP-binding ability between the nanomolar to micromolar range.

Model Primordial Proteins That Bind ATP

In the research on which this thesis builds, work by Newton et al. (2019) in the Seelig lab aimed to study how early proteins may have initially gained *de novo* function—a function not

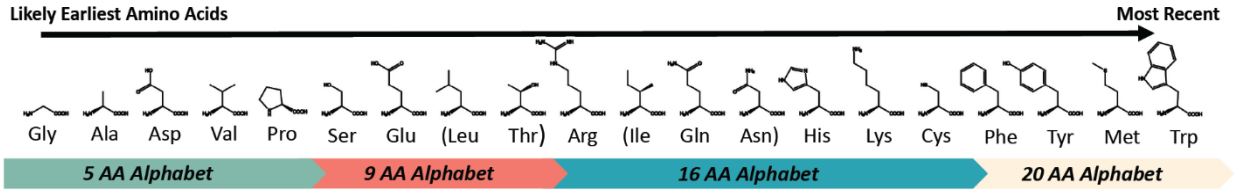

Figure 1. Consensus sequence order for the integration of amino acids into the genetic code

Corresponding alphabets shown below include smaller encompassed alphabets. Successive alphabets encompass all smaller alphabets. Figure adapted with permission from Peter Winslow.

previously existent—and what those intermediary functional primordial proteins may have looked like. To do this, libraries of random sequences like those used by Keefe & Szostak were created. Four separate libraries were synthesized, each composed of successively reduced amino acids alphabets (Fig. 1). Each alphabet corresponded to a stage in the hypothetical chronological emergence of amino acids, as explained by Trifonov (2004). For each library, mRNA display was used to screen for ATP-binding proteins. Interestingly, unpublished data has revealed ATP binding proteins in all four libraries composed of the 5, 9, 16, and 20 earliest consensus order amino acids.

Potentially Novel Mechanisms of ATP Binding for Model Primordial Proteins

With hundreds of putative ATP binding proteins and >5 confirmed binders successfully chosen from each library, it is remarkable that proteins composed of only five types of amino acids can already bind to ATP. In modern proteins, ATP is most commonly bound through a combination of 1) hydrogen bonding, 2) π-π with aromatic amino acids (Fig. 2) and cation–π interactions between ATP's adenine base and aromatic or positively charged residues, and 3)

magnesium ion coordination of the ATP phosphates (Mao et al., 2004). However, in the work by Newton et al. (2019), the 16, 9, and 5-amino-acid libraries do not contain aromatic side chains for π-stacking, and the 9- and 5-amino-acid libraries do not contain positively charged residues. Thus, the modes of binding for these early protein models with limited amino acids are very likely to be quite different from those observed in modern ATP binders.

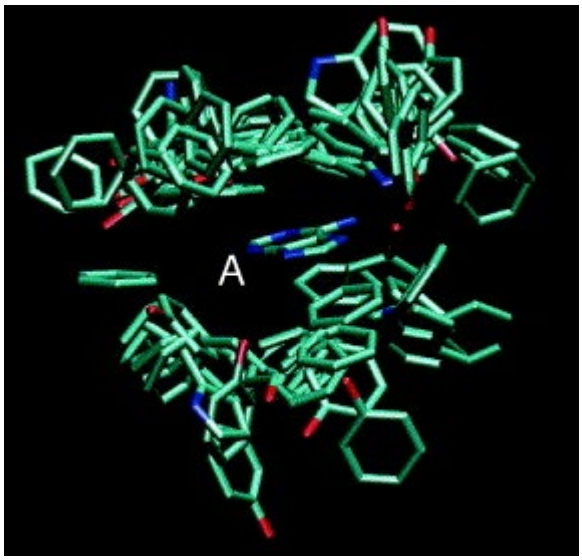

Figure 2. Superposition of all π- π stacking interactions observed across 44 adenylate-protein complexes.

Letter A indicates adenine base. Residues pictured include phenylalanine, tyrosine, and tryptophan. Adapted from Mao et al. (2004).

For protein variants consisting of 5 and 9 types of amino acids, with no positively charged or aromatic residues, binding is likely to occur through hydrogen bonding and coordination of divalent metal ions since cation-π and π-π stacking interactions cannot occur. Modern proteins often utilize magnesium cations to interact with negatively charged phosphate groups of ATP (Mao et al., 2004). Coordination of cations like magnesium by negatively charged side chains in the limited-amino-acid libraries will likely compensate for the lack of positive amino acids in cation- π interactions with the adenine base of ATP and cation-phosphate interactions with the ATP phosphates (Giacobelli et al., 2022; Mahadevi & Sastry, 2013). While natural proteins primarily utilize magnesium for ATP binding (Mao et al., 2004), other divalent metals could potentially be utilized, even for functions unrelated to ATP binding. Keefe & Szostak's (2001) work creating synthetic ATP binders yielded a protein requiring a zinc ion to stabilize its functional, folded structure. Whether by directly coordinating the phosphates of ATP,

participating in salt bridge interactions, or stabilizing the folded structure of our protein variants, it is likely that they could utilize divalent metals in some way—especially in the case of the more limited five and nine amino acid libraries.

Research Objective

This study aimed to characterize the selected ATP binders from the 5, 9, 16, and 20-amino-acid libraries to elucidate which, if any, divalent metal cations are required for ATP binding.

Hypothesis

Because the majority of modern ATP binding proteins use magnesium ions for the coordination of ATP, we hypothesized that our model primordial proteins—particularly the five and nine amino acid proteins—may also utilize magnesium ions to bind to ATP

To allow for biophysical characterization, >5 proteins selected from each library were expressed as fusions with maltose binding protein (MBP) and a 6x-His tag. Fusion proteins were purified from cultures of *Escherichia coli.* Each protein was cleaved from its MBP and 6x-His tags at a recognition sequence using recombinant TEV protease before further characterization. ATP binding assays were used to probe reliance on salt concentration and divalent metal ions. Although not addressed here, ^{1}H and ^{15}N nuclear magnetic resonance spectroscopy (NMR), circular dichroism spectroscopy (CD), and x-ray crystallography were pursued for structural characterization and determination. Elemental analysis of bound divalent metals was performed with inductively coupled plasma optical emission spectrometry (ICP-OES).

We found that our variants from the 5- and 9-amino-acid libraries, which were all highly negatively charged (net charge <-10 at pH 7) and lacked aromatic amino acids, had altered

properties that created significant difficulties in protein purification, quantification, and analysis. The unique properties of our selected ATP-binding proteins may reflect those of actual primordial proteins.

Additionally, we found preliminary evidence suggesting that two five-amino-acid proteins (5A and 5D) require magnesium and manganese to bind to ATP—or for some essential structural function required for binding. Future work will validate these preliminary findings without possibly confounding contamination and further probe the possible mechanisms of binding utilized by the many chosen variants. Ongoing efforts will continue to determine the 3D structure of each selected protein by x-ray crystallography or by ^{15}N NMR and determine precise binding mechanisms in detail.

METHODS:

Fusion Protein Expression

Each protein variant was expressed as a fusion protein with maltose binding protein (MBP) and 6x-His tag, both separated from the variant by a TEV Protease recognition sequence. This schema allows for orthogonal dual purification methods followed by the cleavage of purification tags.

Fusion proteins were expressed from 50% glycerol stocks of BL21 *E. coli.*, which were cultured overnight in 4 mL of luria broth (LB). The next day, each overnight culture was added to 200 mL of terrific broth (TB) in ultra-yield flasks. Cultures were incubated at 37 °C until the OD600 reached ~0.7-1. Fusion protein expression was induced via the lac operon with IPTG (1mM). Induction was allowed to continue for 2 hours at 37 °C before centrifugation at 2831 x g and 4 °C for 20 minutes into a cell pellet. Cell pellets were stored at -30 °C until further

purification.

Fusion Protein Purification

Each cell pellet was allowed to thaw and was resuspended with lysis buffer (20 mM HEPES, 400 mM KCl, Protease Inhibitor Cocktail (Sigma), 0.625 mg/mL lysozyme, 16 U/mL benzonase) with intermittent agitation until a homogeneous consistency was achieved. Resuspend cell pellets were then sonicated with a one fourth inch diameter tip at 35% amplitude, 20/20 on/off cycles, for a sonication time of 6 minutes while submersed in an ice bath. Cell lysates were centrifugated at 30k x g for 15-20 minutes.

Lysate supernatant containing soluble protein was purified via Ni-NTA agarose and amylose affinity gravity columns. HisPur Ni-NTA agarose resin (Thermo Scientific) was used to bind 6x-His tags on the fusion protein. Gravity columns (20 mL, Bio-Rad) of Ni-NTA agarose resin were prepared with 2mLs of bed volume per 200mL TB culture. Lysate supernatant was applied to the column, collected, and reapplied twice. Columns were washed with 10 bed volumes of wash buffer (20mM HEPES, 400mM KCl) and eluted with 5 bed volumes of nickel elution buffer (20mM HEPES, 400mM KCl, 200mM imidazole). The estimated protein concentration in all fractions was determined with A280 absorbance on a Nanodrop Spectrophotometer.

Gravity columns (20 mL, Bio-Rad) of amylose resin (NEB) were prepared with 1 mL of bed volume for every 3.6 mg/mL of protein in nickel elution samples as measured by A280. Nickel elution fractions were applied to the amylose columns, re-collected, and reapplied twice. Amylose columns were washed with 10 bed volumes of wash buffer and eluted with 5 bed volumes of amylose elution buffer (20mM HEPES, 400mM KCl, 10mM maltose). The estimated protein concentration in all fractions was determined with A280 absorbance on a Nanodrop

Spectrophotometer.

TEV Protease Treatment

Amylose elution samples were centrifuged at 2800 x g at 4 °C in spin filters (Amicon Ultra, 30000 MW) to reach an approximate concentration of 5 mg/mL. Concentrated protein samples were treated with 1 mg of TEV protease (either from a prepared TEV storage solution or purchased from Sigma Aldrich) for every 100 mg of protein. TEV-treated protein solutions were incubated at room temperature for 2 hours with infrequent agitation and overnight at 4 °C.

Secondary Nickel Column for Removal of 6x-His Tagged Components

Another HisPur Ni-NTA agarose resin (Thermo Sci.) column (20 mL, Bio-Rad) was prepared with a 5-fold excess resin capacity compared to total protein mass with a 16.4 mg/mL resin capacity. Post-TEV treatment samples were applied to the column and reapplied twice. Approximately half a bed volume of wash buffer (20mM HEPES, 400mM KCl) was applied to the column and collected with the flow-through. Columns were washed with an additional 10 bed volumes of wash buffer and eluted with 5 bed volumes of nickel elution buffer (20mM HEPES, 400mM KCl, 200mM imidazole). Protein concentration in the flow-through was determined via BCA assay for non-UV-absorbing free protein. The concentration of UV-absorbing protein (primarily MBP) was determined with A280 absorbance on a Nanodrop Spectrophotometer.

Protein Concentration Determination and Purity Analysis

Protein concentration in elution fractions was determined with A280 absorbance on a Nanodrop Spectrophotometer. For precise measurements of non-UV absorbing proteins, a Pierce™ bicinchoninic acid (BCA) assay was performed with a heating step for 5 minutes at 60 °C. Bovine serum albumin (BSA) protein standards were used to make a standard curve.

Purity analysis was conducted using SDS-PAGE (4–12 % Bis·Tris, NuPage gels, Life Technologies) in MES buffer (Life Technologies). 4x LDS or SDS (NuPAGE) with 5% beta-mercaptoethanol (BME) was used to load protein samples for denaturing gels. Gels were stained with either Coomassie Brilliant Blue or silver staining with the Pierce™ Silver Stain Kit. Gel purity analysis was conducted with Gel Analyzer (Istvan Lazar Jr., PhD & Istvan Lazar Sr., PhD, CSc, n.d.).

Affinity Tag-free Protein Concentration

Protein samples were prepared for further characterization with concentration and buffer exchange at 2800 x g and 4 °C in spin filters (Amicon Ultra, 3000 MW). The concentration of tag-free protein was measured via BCA assay.

FPLC Size Exclusion

Additional purification of affinity tag-free, cleaved protein from co-purified contaminants was conducted with a Superdex 75x 101300 SEC column with a volume of 23.56 mL. Wash buffer (20 mM HEPES, 400 mM KCl) was used as a mobile phase throughout sample injection and fraction collection. Fraction analysis was conducted with BCA assay and Pierce™ silver stained SDS-PAGE due to poor UV absorption of the tag-free proteins.

^{32}P ATP Binding Assay

Buffer conditions of the ATP binding assay were intended to mimic those of the initial selection. The buffer contained 4 mM $MgCl_2$, 10 mM maltose, 400 mM KCl, 20 mM HEPES, 2 mM glutathione, 1 mM glutathione disulfide, and 0.05 mM $ZnCl_2$, $CaCl_2$, $CuSO_4 \cdot 5H_2O$, $MnSO_4 \cdot H_2O$, $CoCl_2 \cdot 6H_2O$, $NiSO_4 \cdot 7H_2O$, and $NaMoO_4 \cdot 2H_2O$. The concentration of protein was determined via BCA assay. ^{32}P labeled ATP was added to each sample to reach a concentration of 1 nM and incubated with the protein sample for one hour. 150 uL of each sample was added to

a Microcon 30K cutoff 1.5 mL filter unit. The filter units were centrifuged for 10 seconds at 16873 rcf to wet the membranes. Samples were centrifuged for 1 minute at 16873 rcf to separate the retentate and filtrate. 15 uL aliquots taken from the retentate and filtrate were measured with a Beckman LS 6500 scintillation counter. Counts per minute (cpm) were compared in the retentate (R) and filtrate (F) to calculate fractions of protein-bound ligand ([PL]) and total free ligand ([L]).

$$R_{cpm} = [PL]_{cpm} + \left[L_{free}\right]_{cpm} \qquad F_{cpm} = \left[L_{free}\right]_{cpm}$$

$$\frac{R_{cpm} - F_{cpm}}{R_{cpm}} = \frac{[PL]}{[L]}$$

Inductively Coupled Plasma-Optical Emission Spectrometry (ICP-OES)

ICP-OES data was obtained from the Research Analytical Laboratory at the University of Minnesota. Before dialysis, each divalent metal ion was added to 0.291 mM concentration in the sample and incubated for one hour at room temperature. Protein samples were dialyzed overnight against 1L of wash buffer (20mM HEPES, 400mM KCl) with 5 g of Chelex 100 resin (Bio-Rad). After dialysis, the ~0.5 mL sample was diluted to 15 mL with 20 mM HEPES and filled to 2% nitric acid. Protein concentrations of prepared samples were determined via BCA assay prior to dilution. Elemental analysis from ICP-OES was reported in parts per million (ppm) and converted to mole equivalents to compare to known protein concentrations.

RESULTS:

Overview:

25 protein variants from the four libraries have been characterized to varying degrees. At least 12 of these variants shared a common 74kDa contaminant protein that was confirmed via mass spectrometry analysis to be HSP70 family protein DnaK (Fig. 3). Therefore, characterization efforts have focused on variants without detectable DnaK contamination on a silver-stained gel, which included all 5-alphabet variants with the exception of SE8. However, recent silver staining of these variants has found that they may still contain <1% DnaK contamination. Initial evidence in an ^{32}P ATP binding assay has shown that 5A and 5D may require magnesium or manganese to bind ATP, although such findings must be validated in the complete absence of DnaK.

ATP binding assay of tag-free variant 9C shows the effect of DnaK contaminant on binding affinity. I used an ATP binding assay to characterize DnaK's effect on the measurements and to determine the severity of the DnaK contamination problem. Initial efforts were taken to quantify the effect of trace amounts of contaminant DnaK on the ATP-binding measured in a ^{32}P-ATP binding assay. A sample of tag-free variant 9C was purified further via size exclusion chromatography to remove putative DnaK contamination. An SDS-PAGE gel of the tag-free 9C sample pre- and post-size exclusion was silver-stained, and gel bands were quantified (Fig. 4). The pre-size exclusion sample of protein 9C was determined to be 19.7% DnaK whereas the post-size exclusion sample was determined to have <1% DnaK contamination, as it was undetected. A ^{32}P-ATP binding assay measured the comparative ATP binding of these two samples. It provided evidence that DnaK increased measured binding of otherwise identical tag-less 9C protein samples (Fig. 5). For accurate determination of ATP binding of our protein

variants, confounding factors should have a minimal effect on the measured binding, which was not the case here. This was problematic for further ATP-binding characterization, and thus, contaminated samples were avoided through continuing characterization efforts. This finding directed our study to the five amino acid library proteins, which had no detectable DnaK contamination on a silver-stained gel except for SE8.

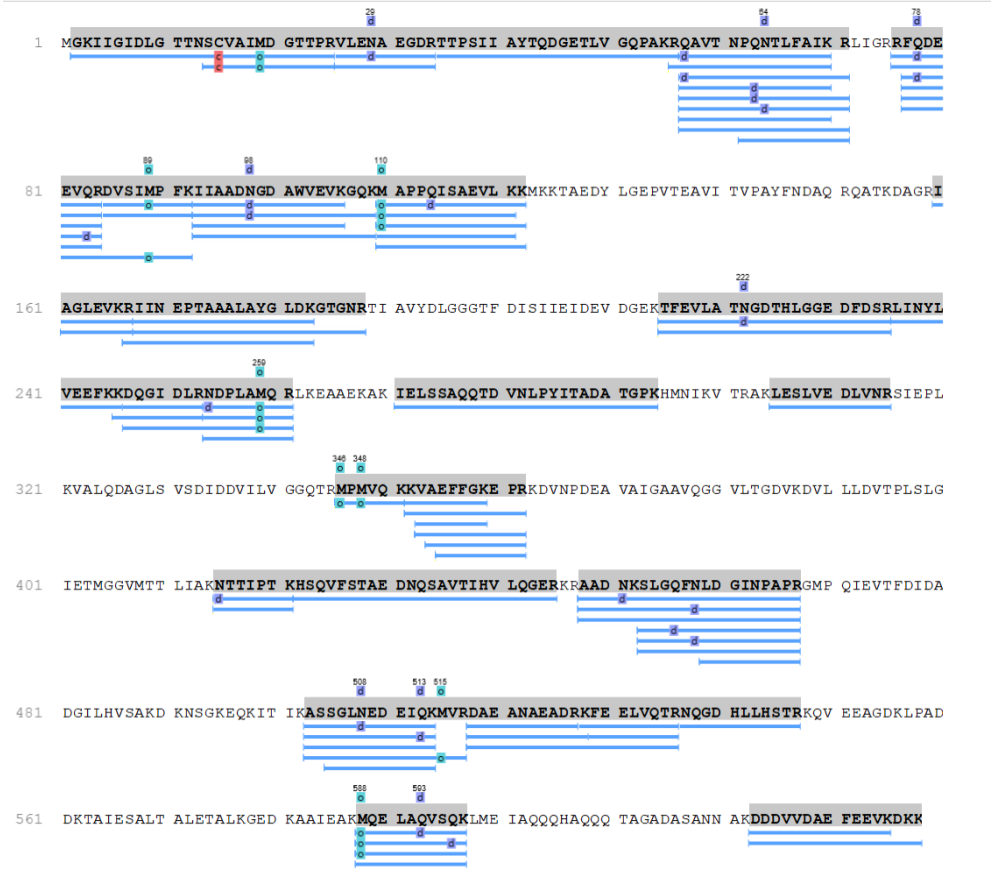

Figure 3. Mass spectrometry analysis reveals contaminant HSP70 DnaK

Mass spectrometry of ~74 kDa band excised from the rightmost lane marked “DnaK” in the SDS-PAGE gel in Figure 4. The coverage of trypsin-digested fragments over the HSP70 DnaK protein sequence is shown. Total protein coverage was 58%. Additional annotations indicate post-translational modifications.

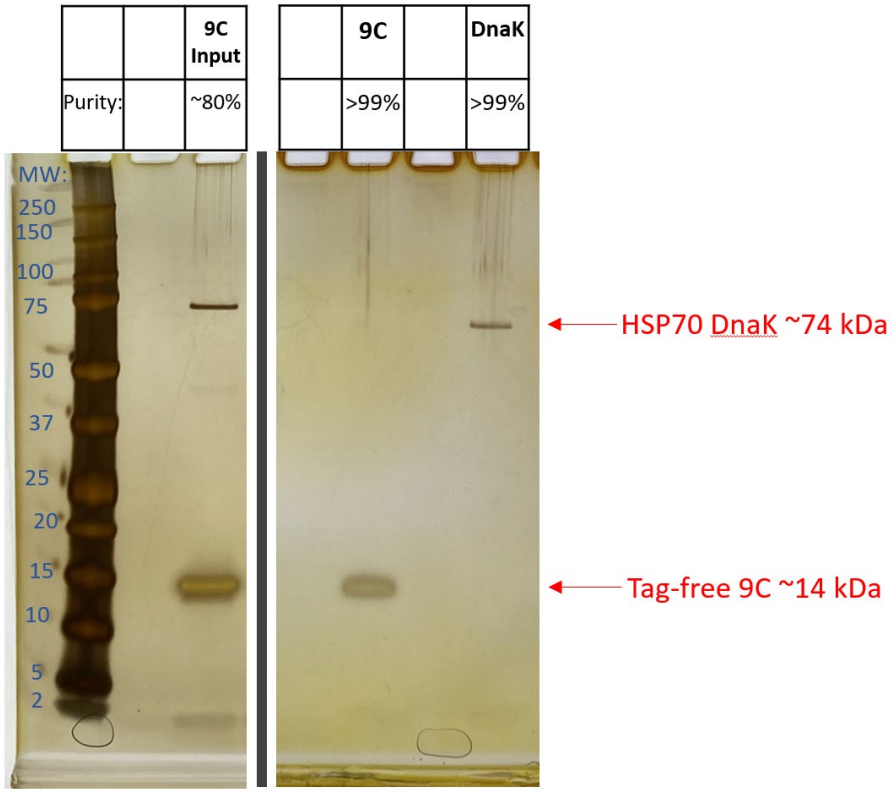

Figure 4. Gel electrophoresis of affinity tag-free protein 9C before and after size exclusion chromatography to separate 9C from HSP70 protein DnaK

Two precast NuPAGETM 4-12% Bis-Tris polyacrylamide gels were run for 35 minutes at 200V. Each lane, except the DnaK lane, contains ~5 ug of protein as determined by BCA assay. MW: molecular weight as determined by Bio-rad Precision Plus Protein™ Dual Xtra Standards ladder. The gel to the right did not have a ladder, but RF values remained comparable. Each gel was silver stained with a Pierce™ Silver Stain Kit, and band percentages were determined with Gel Analyzer.

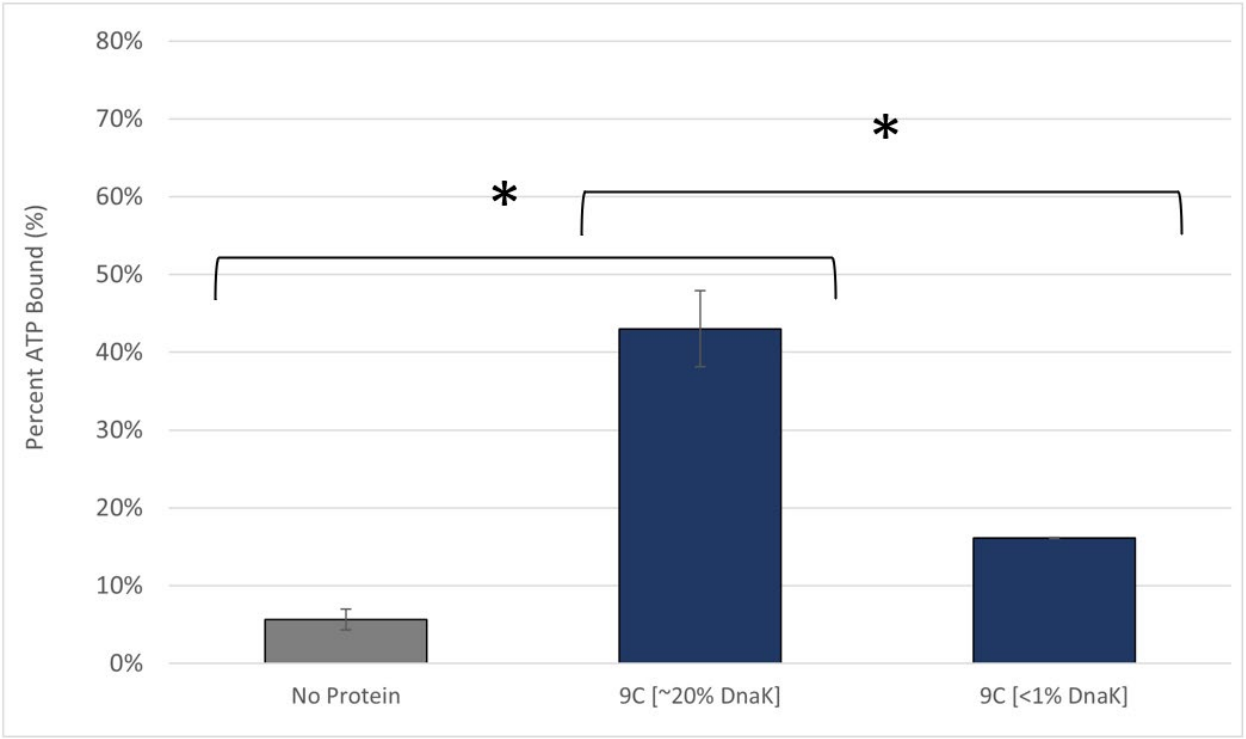

Figure 5. Variant 9C's ATP-Binding putatively increased by contaminant DnaK

Percent ATP bound was calculated as the difference in measured radioactivity of the retentate and filtrate out of total retentate radioactivity in counts per minute. Each sample was run in duplicate with error bars indicating +/- standard error of the mean. Pairwise student's t-tests were performed for 9C+DnaK vs. no protein, p=0.007; vs. 9C, p=0.002; and no protein vs. 9C, p=0.059, * = <0.05.

The first results from ICP-OES were inconclusive. We investigated which divalent metals were bound by each protein in ICP-OES trace metal analysis with a dialyzed sample of tag-free variant 5D (Fig. 6). Prior to dialysis, each divalent metal ion was added to 0.291 mM concentration in the sample and incubated for one hour at room temperature. Protein samples were dialyzed overnight, diluted with 20 mM HEPES, and brought up to 2% nitric acid. If a given divalent metal ion was bound by protein in the dialyzed sample, a larger amount of such metal would be retained within the dialysis membrane as compared to a sample of wash buffer. Unfortunately, the positive control, which should have contained a concentration of 3μM of each

divalent metal cation, was measured to be between 6-13 µM for each divalent metal ion. In general, they contained an elevated amount of each metal—likely due to evaporation in the metal stock solutions that had been stored for ~7 years.

Additionally, the dialyzed wash buffer sample contained higher levels of Cu^{2+} than the 5D treatment sample, invalidating this data. While further ICP-OES characterization was not undertaken, the next step will be to repeat the same ICP-OES protocol with new divalent metal stock solutions and a more stringent two-step dialysis. Characterization efforts continued to pursue metal dependence data for ATP binding in an ATP binding assay.

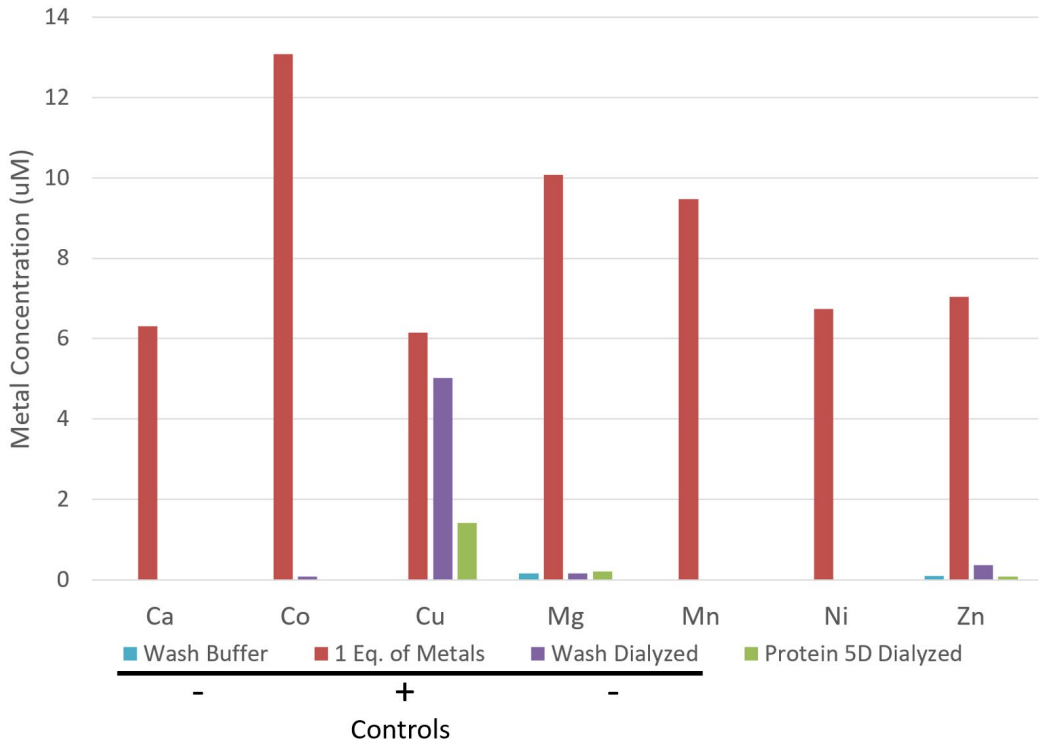

Figure 6. Inconclusive ICP-OES trace metal analysis of tag-free 5D

Wash buffer (20 mM HEPES, 0.4 M KCl), a one molar equivalent (3µM) metals solution, post-dialysis 5D incubated with metals (0.291 mM each), and post-dialysis wash buffer incubated with metals (0.291 mM each). For each divalent metal, the positive control sample was found to be variable and higher than the intended concentration. Copper levels appeared elevated in the dialyzed samples, and the negative control dialyzed wash buffer had elevated metal content.

5A's ATP binding affinity is modulated by magnesium (II) and manganese (II) ions in the absence of other divalent metals. Characterizing the divalent metal ions required for the function of each variant was vital for informing the general mechanisms of binding. The effect of individual divalent metal ions on the ATP-binding affinity of each protein was investigated for a sample of MBP-fusion 5A (Fig. 7). Either all eight divalent metal cations, no divalent metals, or individual metals were added to samples of MBP-5A prior to the ATP incubation step in a ^{32}P-ATP binding assay. The addition of solely magnesium or manganese yielded ATP-binding slightly higher than the addition of all divalent metals. In contrast, all other individual metals measured exhibited binding at levels close to the no divalent metals control (Fig. 8). Interestingly, the no divalent metals control had higher observed binding than the no protein negative control. This could reflect ATP-binding due to trace metals in the purified protein samples. Preliminary results suggested a similar metal preference for protein 5D, where magnesium and manganese recovered ATP-binding (data not shown). These results suggested that 5A and 5D can utilize magnesium or manganese ions in their ATP-binding mechanisms or that those ions provide some essential structural stability necessary for ATP-binding.

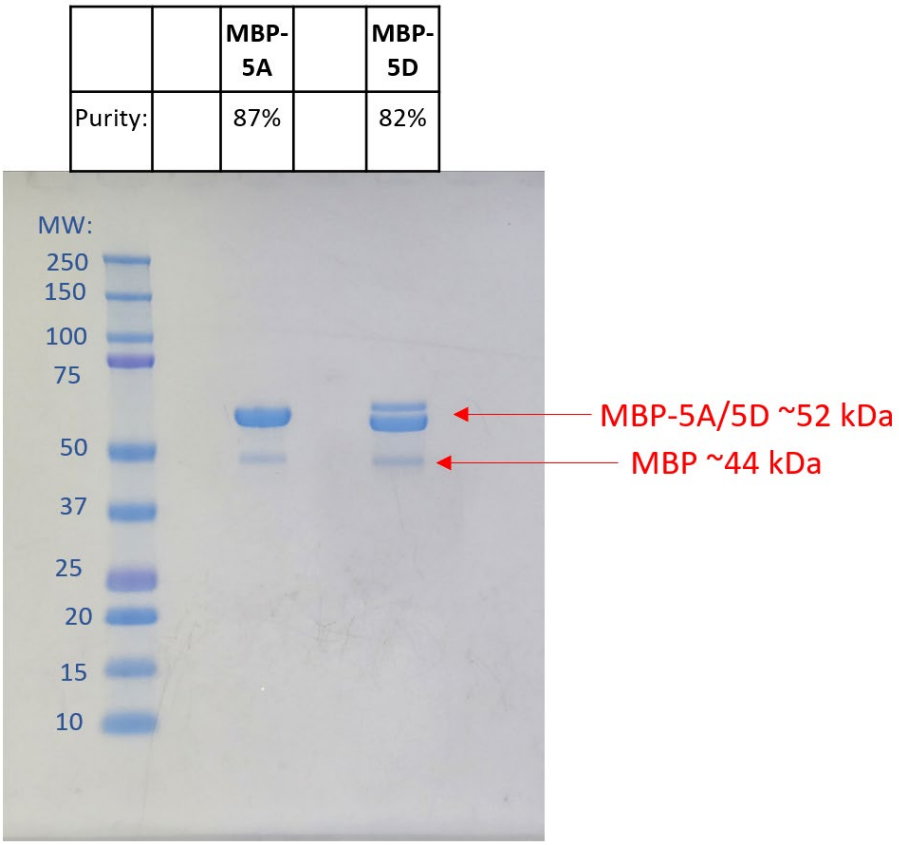

Figure 7. Gel electrophoresis of purified MBP-5A and MBP-5D

A precast NuPAGETM 4-12% Bis-Tris polyacrylamide gel was run for 35 minutes at 200V. Each lane contains ~5 ug of protein as determined by A280 absorbance. MW: molecular weight as determined by Bio-rad Precision Plus Protein™ Dual Color Standards ladder. The gel was Coomassie-stained and band percentages were determined with Gel Analyzer. Both proximal bands in the MBP-5D sample were counted as 5D for purity calculations.

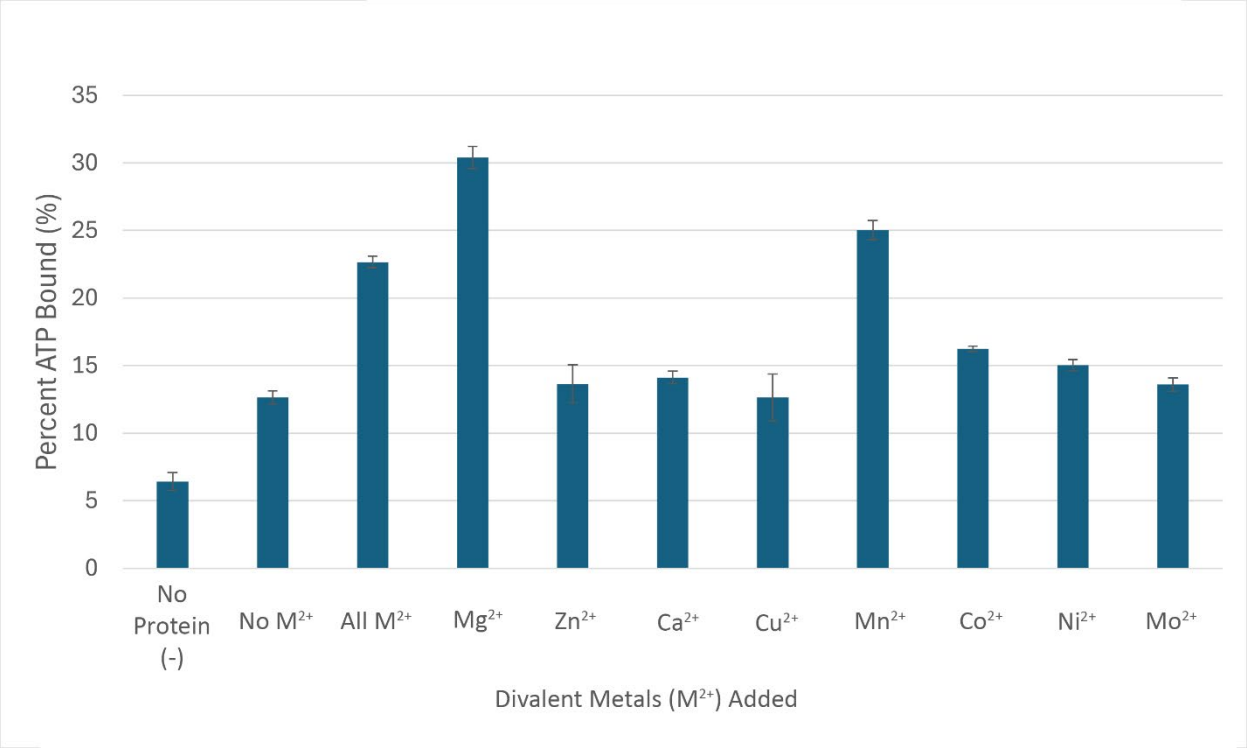

Figure 8. Metal dependence of ATP-binding by variant 5A

The percentage of ATP bound for each condition is shown as a bar plot with error bars indicating positive and negative standard error based on triplicate samples. The "No protein" sample contains only wash buffer (20 mM HEPES, 400 mM KCl) instead of a protein sample. Each metal was provided at 50 µM concentration except Mg2+, which was 4 mM. Each sample contained 7.5 µM protein and 1 nM 32P-ATP.

The random sequence model primordial proteins exhibited unique behaviors in comparison to natural proteins. All selected five and nine amino acid proteins had exceptionally high negative charges, with isoelectric points between two and three. This was due to the unique composition of the proteins, which were made to proportionally mimic the modern distribution of amino acid abundance, but restricted within the limited alphabets (Newton et al., 2019). Notably, the five amino acid proteins contained the least amount of aspartate, and the nine amino acid proteins contained similarly low amounts of aspartate and glutamate. Still, the proteins generally were highly negatively charged due to their lack of or lower composition of

positively charged amino acids in the five and nine amino acid alphabets respectively. Primordial proteins with similar compositions may have served a role in forming coacervates that could have been functionally important in early life (Despotovic & Tawfik, 2021). In practice, however, this manifested in proteins that behaved abnormally in many molecular biology techniques as compared to natural proteins. For the five and nine amino acid proteins, we observed, among other behaviors, a lack of interaction with SDS when running SDS-PAGE gels, an inability to quantify our proteins by A280 absorbance, poor effectiveness of biuret and BCA quantification methods, and a lack of staining or a halo-like inverted staining of proteins bands on SDS-PAGE gels when Coomassie or silver stained. Additionally, ongoing characterization efforts have suggested the possibility that the model primordial proteins are intrinsically disordered yet possess micromolar affinity for ATP.

DISCUSSION:

Overview:

Our ATP-binding assay data confirmed our hypothesis and indicates that early functional proteins may indeed have used magnesium and manganese ions to bind to ATP. However, there is a possibility that our current ATP-binding measurements have been confounded by contamination with ~0.1% *E. coli* protein DnaK. It is possible that the divalent metal-dependent ATP-binding measured for proteins 5A and 5D (Fig. 8) could be wholly or partly due to this contaminant DnaK. Therefore, the next crucial step will be the purification of all proteins of interest using the DnaK deletion strain of BL21(DE3) (Ratelade et al., 2009) to eliminate the potential for DnaK contamination.

DnaK Contamination

While our ATP-binding assay data confirms our hypothesis, concern remains that the metal preference measured was that of the DnaK protein rather than of variants 5A and 5D. It is known that DnaK utilizes Mg^{2+} ions to bind ATP (Skowyra & Wickner, 1995), and it is plausible that Mn^{2+} could substitute for Mg^{2+} as this has been observed in other chaperone proteins (Diamant et al., 1995). Even more concerningly, the ~1 nM binding affinity of DnaK for ATP (Russell et al., 1998) could allow even trace amounts of DnaK to distort binding measurements in otherwise pure (99%) protein samples. This 1 nM dissociation constant (K_d) of DnaK for ATP was measured in conditions with 11 mM Mg^{2+} (Russell et al., 1998). In contrast, similar characterizations have found Kd values ranging between 44 nM and 90 nM for different ATP and ADP analogues without magnesium ions (Theyssen et al., 1996). In our binding assay, 4 mM Mg^{2+} was present in the buffer, which could cause a lower observed affinity of DnaK for ATP compared to that with 11 mM Mg^{2+}. However, both in the case of the measured affinity of DnaK for ATP and in our case, Mg^{2+} would be in significant molar excess of DnaK and total protein, respectively.

To simplify the following thought experiment, we will assume that the 1 nM K_d will remain unchanged at 4 mM Mg^{2+}. Based on the ~1 nM affinity of DnaK for ATP with 11 mM Mg^{2+}, 1 nM DnaK constituting ~0.013% of the 7.5 uM protein sample could theoretically bind ~50% of the 1 nM ATP present in the binding assay conditions. However, this is an estimate based on the linearity of binding assuming an excess of protein and ligand over the bound complex, which is not the case. Rather, using equation one, we may substitute free protein with $[P]=[P_{total}]-[PL]$ and the analogous case for ligand. Then, distributing terms, we solve the resultant quadratic binding equation with the discriminant shown in equation two (Hulme &

Trevethick, 2010). Assuming 1 nM DnaK is present with a K_d of 1 nM, we calculate a protein-ligand complex [PL] of 0.303 nM. Thus, ~30% of the ATP would be bound. With 1% contamination of the total 7.5 uM protein sample with DnaK, the effective concentration of DnaK would be 75 nM. In this case, the estimated binding would be ~98.6%.

$$\mathrm{Kd} = \frac{k_{off}}{k_{on}} = \frac{[\mathrm{P}][\mathrm{L}]}{[\mathrm{PL}]} \tag{1}$$

$$[PL] = \frac{([P]_{tot} + [L]_{tot} + K_d) - \sqrt{([P]_{total} + [L]_{total} + K_d)^2 - 4[P]_{total}[L]_{total}}}{2} \tag{2}$$

However, this was not observed to be the case experimentally. Pictured in Figure 4, a sample of protein 9C with ~20% DnaK contamination was determined by silver-stained SDS-PAGE gel band intensity quantification. Since only ~40% ATP binding was observed in this sample, some assumptions must have been incorrect. Two primary reasons likely underlie the discrepancy between theoretical ATP-binding proportions and our experimental data.

First, as previously mentioned, for many of the studied protein variants, particularly those from the five- and nine-amino-acid libraries, silver stain quantification of band intensity underestimated the total protein mass on SDS-PAGE gels. In Figure 4, a transparent halo-like stain of the 9C band was observed, which suggests that the stain intensity integration may significantly underestimate the actual sample composition of the band compared to the well-defined and darkly stained DnaK band. Second, while fewer assumptions have been made with the quadratic model for DnaK's ATP-binding—which accounts for the lowered effective concentration of ATP as DnaK binds it—the actual situation is more complex. We have neglected to take into account the ATPase activity of DnaK. DnaK has a slow rate of hydrolysis (k_{hyd}) of 0.02 min^{-1}(Russell et al., 1998). Thus, DnaK could feasibly hydrolyze a significant fraction of

the ATP in solution over the course of the one-hour incubation period. For example, if the 1 nM ATP was completely bound by DnaK, over the course of one hour, approximately $1- e^{-0.02.60}=$ 70% of the initial ATP would be hydrolyzed. This could introduce further binding interactions between both DnaK and ADP and between the studied variants and ADP, which we have not fully characterized. As future characterization proceeds with a DnaK knockout strain of BL21(DE3) (Ratelade et al., 2009), we hope to avoid the complications presented by DnaK contamination entirely.

Divalent Metal Dependence of Proteins 5A and 5D

Considering these inconsistencies, we can extrapolate the observed binding trend to the ATP-binding data collected for protein MBP-5A in the metal dependence ATP binding assay (Fig. 8). We estimate that our MBP-5A and MBP-5D samples, on which ATP binding assays were performed, contain approximately 0.1% DnaK contamination. This was estimated by comparing band intensities on a silver-stained gel to a standard curve of known masses. Even with such little contamination, DnaK would be projected to bind ~87% of the 1 nM ATP by the quadratic binding equation (2). As shown in Figure 8, no conditions of MBP-5A bound close to this much ATP, with the highest condition of Mg^{2+} only reaching ~30% ATP bound. Thus, we can be certain that some experimental complexity is not adequately accounted for in our model calculations. It is still unclear why this discrepancy remains. Because these samples were of fusion proteins, it is less likely that the discrepancy in the projected binding of DnaK to ATP arose from a discrepancy in stain effectiveness, as the MBP affinity tag stains well. Thus, the discrepancy likely arose from a combination of the effects of ATP hydrolysis and DnaK's altered affinity for ATP under the conditions of our binding assay. Regardless of the precise cause, we aim to avoid these confounding factors moving forward with a DnaK knockout strain of *E. coli*.

As noted before, the sample of protein 5A with no divalent metals added (Fig. 8) had higher ATP-binding than the "no protein" control. This difference is most likely due to the presence of trace metals in the no M^{2+} sample derived from the purified protein sample. However, suppose the measured metal preference for 5A and 5D was entirely or partially due to the DnaK contaminant. In that case, it is possible that the remaining fraction of binding present above the negative control could feasibly be due to the 5A/5D variants themselves. In future work, dialyzed samples of 5A and 5D—such that unbound trace metal ions should be effectively eliminated from the solution—will be directly tested via ^{32}P ATP-binding assay.

5A and 5D may be confirmed to require Mg^{2+} or Mn^{2+}. Alternatively, future testing may reveal that they do not require divalent metal cations at all. In either case, such data would have major implications for the nature of early functional proteins and how they compare to modern proteins. If such model primordial proteins require Mg^{2+} or Mn^{2+} metal ions to bind ATP, it would suggest a continuity between the mechanisms of ATP binding in extant life and the mechanisms early, now-extinct forms of life could have used. Were these findings to be validated, it could be proposed that the magnesium coordination of ATP may be a universal optimum solution for ATP-binding across the possible sequence space of proteins. Early life likely utilized far larger shares of reduced iron, manganese, cobalt, and nickel, which were all more enriched in the ocean prior to the Great Oxidation Event approximately 2.4 billion years ago (Dupont et al., 2010). Evolutionary evidence suggests that life evolved proteins that preferentially utilized the metals available during the evolutionary period in which they evolved (Dupont et al., 2010; Wang et al., 2007). Thus, it is likely that our laboratory-generated protein variants composed of random amino acid sequences might adopt a variety of the divalent metals provided during their ATP-binding selection.

ICP-OES Results

While the data collected from initial ICP-OES testing was deemed unusable, it still served as a valuable indicator for the sensitivity of such measurements and as a baseline for future work. The primary reason for the invalidation of the ICP-OES data was the failure of the dialyzed negative control sample, which was likely due to incomplete dialysis. The sample was composed of wash buffer containing 400 mM KCl and 20 mM HEPES. Before dialysis, each divalent metal ion was added to 0.291 mM concentration in the sample. Dialysis proceeded overnight, but in an oversight, the dialysis buffer was not stirred, and only one round of dialysis was done. Both factors likely contributed to the incomplete dialysis. Interestingly, an excess of copper, in particular, was observed in the dialyzed wash buffer and protein samples (Fig. 6). This excess of copper was not due to a systematic error of excess copper initially added to the samples since copper was not in excess of other divalent metals in the one molar equivalent metals sample. Thus, the anomalous copper remaining after dialysis could either be due to a separate unknown source of copper contamination—such as from the 20mM HEPES buffer used to dilute the samples or due to a selective lack of copper dialysis—perhaps due to selective inefficiency of copper chelation by the Chelex 100 resin used in the dialysis buffer.

There was also a noticeable inconsistency in the amount of each divalent metal present in the one molar equivalent sample. The sample was intended to contain each divalent metal cation at a ~3mM concentration. However, all divalent metal cations were experimentally measured to be in at least two-fold excess of this intended concentration. This inaccuracy is likely due to a combination of the hydration of divalent metal salt stocks and the gradual evaporative concentration of dissolved divalent metal stocks used to formulate the sample. Several metal salt

stocks are hygroscopic and were later examined to find clumping and other evidence of deliquescence, suggesting this as the correct explanation.

Current Outlook and Future Directions

Throughout the course of working with our model primordial-like proteins, evidence has accumulated to suggest that they may be intrinsically disordered. The protein variants were selected from completely random sequences. Although multiple rounds of mRNA display selection were required to enrich ATP binding sequences, those sequences were not mutagenized and thus remained unevolved. As mentioned previously, the unique composition of the five and nine amino acid proteins caused numerous difficulties in applying standard molecular biology techniques. We also found putative affinity of DnaK for many, if not all, of the protein variants characterized due to its co-purification. Additionally, efforts to crystallize the protein variants have been met with great difficulty. In summary, many lines of evidence point towards some form of disorder across all four libraries of variants studied.

Current efforts, particularly those by Seelig lab members Eli Blascyk, Samuel Erickson, and Peter Winslow, seek to further confirm this potential for intrinsic disorder across the different libraries through circular dichroism (CD) and ^{15}N nuclear magnetic resonance spectroscopy (NMR). If the selected variants, particularly those from the five- and nine-amino-acid libraries, are found to be intrinsically disordered, it might suggest that early functional proteins were also generally intrinsically disordered.

In addition to exploring disorder, once structural data has been collected, we aim to evaluate the accuracy of protein fold prediction models such as AlphaFold3 (Abramson et al., 2024). These models are essentially pattern recognition tools and, as such, may only accurately

predict within the sequence space they have been trained on. Whether their predictive accuracy can generalize beyond this into the sequence space of model primordial proteins remains to be seen.

ACKNOWLEDGMENTS:

I want to thank Peter Winslow for his priceless mentorship and teaching throughout my work with him and Dr. Burckhard Seelig for his instrumental role in improving my abilities to practice and communicate complex science during my time at his lab. I would also like to thank my parents, Pamela and Dave, for their unending support and my girlfriend, Olivia, for her care and commitment to my well-being. I could not have accomplished this work without a team of support behind me. I would also like to thank Eli Blascyk for their collaborative role in the ongoing project.

FUNDING:

This work was supported by NASA under award 80NSSC21K0595 and in part by the University of Minnesota's Office of Undergraduate Research.

CONFLICTS OF INTEREST:

I declare no conflicts of interest in writing this thesis.

References:

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w

Campbell, T. D., Febrian, R., McCarthy, J. T., Kleinschmidt, H. E., Forsythe, J. G., & Bracher, P. J. (2019). Prebiotic condensation through wet–dry cycling regulated by deliquescence. *Nature Communications*, *10*(1), 4508. https://doi.org/10.1038/s41467-019-11834-1

Cornell, C. E., Black, R. A., Xue, M., Litz, H. E., Ramsay, A., Gordon, M., Mileant, A., Cohen, Z. R., Williams, J. A., Lee, K. K., Drobny, G. P., & Keller, S. L. (2019). Prebiotic amino acids bind to and stabilize prebiotic fatty acid membranes. *Proceedings of the National Academy of Sciences*, *116*(35), 17239–17244. https://doi.org/10.1073/pnas.1900275116

Crockford, P. W., Bar On, Y. M., Ward, L. M., Milo, R., & Halevy, I. (2023). The geologic history of primary productivity. *Current Biology*, *33*(21), 4741-4750.e5. https://doi.org/10.1016/j.cub.2023.09.040

Danger, G., Plasson, R., & Pascal, R. (2012). Pathways for the formation and evolution of peptides in prebiotic environments. *Chem. Soc. Rev.*, *41*(16), 5416–5429. https://doi.org/10.1039/C2CS35064E

Despotovic, D., & Tawfik, D. S. (2021). Proto-proteins in Protocells. *ChemSystemsChem*, *3*(4), e2100002. https://doi.org/10.1002/syst.202100002

Diamant, S., Azem, A., Weiss, C., & Goloubinoff, P. (1995). Effect of Free and ATP-bound Magnesium and Manganese Ions on the ATPase Activity of Chaperonin GroEL14. *Biochemistry*, *34*(1), 273–277. https://doi.org/10.1021/bi00001a033

Dryden, D. T. F., Thomson, A. R., & White, J. H. (2008). How much of protein sequence space has been explored by life on Earth? *Journal of The Royal Society Interface*, *5*(25), 953–956. https://doi.org/10.1098/rsif.2008.0085

Dupont, C. L., Butcher, A., Valas, R. E., Bourne, P. E., & Caetano-Anollés, G. (2010). History of biological metal utilization inferred through phylogenomic analysis of protein structures. *Proceedings of the National Academy of Sciences*, *107*(23), 10567–10572. https://doi.org/10.1073/pnas.0912491107

Eddington, A. S. (1944). The evaluation of the cosmical number. *Mathematical Proceedings of the Cambridge Philosophical Society*, *40*(1), 37–56. https://doi.org/10.1017/S030500410001817X

Frenkel-Pinter, M., Samanta, M., Ashkenasy, G., & Leman, L. J. (2020). Prebiotic Peptides: Molecular Hubs in the Origin of Life. *Chemical Reviews*, *120*(11), 4707–4765. https://doi.org/10.1021/acs.chemrev.9b00664

Giacobelli, V. G., Fujishima, K., Lepšík, M., Tretyachenko, V., Kadavá, T., Makarov, M., Bednárová, L., Novák, P., & Hlouchová, K. (2022). In Vitro Evolution Reveals Noncationic Protein–RNA Interaction Mediated by Metal Ions. *Molecular Biology and Evolution*, *39*(3), msac032. https://doi.org/10.1093/molbev/msac032

Hulme, E. C., & Trevethick, M. A. (2010). Ligand binding assays at equilibrium: Validation and interpretation. *British Journal of Pharmacology*, *161*(6), 1219–1237. https://doi.org/10.1111/j.1476-5381.2009.00604.x

Imai, E., Honda, H., Hatori, K., Brack, A., & Matsuno, K. (1999). Elongation of Oligopeptides in a Simulated Submarine Hydrothermal System. *Science*, *283*(5403), 831–833. https://doi.org/10.1126/science.283.5403.831

Istvan Lazar Jr., PhD & Istvan Lazar Sr., PhD, CSc. (n.d.). *GelAnalyzer* (Version 19.1) [Computer software]. www.gelanalyzer.com

Joyce, G. F., & Szostak, J. W. (2018). Protocells and RNA Self-Replication. *Cold Spring Harbor Perspectives in Biology*, *10*(9), a034801. https://doi.org/10.1101/cshperspect.a034801

Keefe, A. D. (2001). Protein Selection Using mRNA Display. *Current Protocols in Molecular Biology*, *53*(1). https://doi.org/10.1002/0471142727.mb2405s53

Keefe, A. D., & Szostak, J. W. (2001). Functional proteins from a random-sequence library. *Nature*, *410*(6829), 715–718. https://doi.org/10.1038/35070613

Kvenvolden, K., Lawless, J., Pering, K., Peterson, E., Flores, J., Ponnamperuma, C., Kaplan, I. R., & Moore, C. (1970). Evidence for Extraterrestrial Amino-acids and Hydrocarbons in the Murchison Meteorite. *Nature*, *228*(5275), 923–926. https://doi.org/10.1038/228923a0

Mahadevi, A. S., & Sastry, G. N. (2013). Cation−π Interaction: Its Role and Relevance in Chemistry, Biology, and Material Science. *Chemical Reviews*, *113*(3), 2100–2138. https://doi.org/10.1021/cr300222d

Mao, L., Wang, Y., Liu, Y., & Hu, X. (2004). Molecular Determinants for ATP-binding in Proteins: A Data Mining and Quantum Chemical Analysis. *Journal of Molecular Biology*, *336*(3), 787–807. https://doi.org/10.1016/j.jmb.2003.12.056

Miller, S. L. (1953). A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science*, *117*(3046), 528–529. https://doi.org/10.1126/science.117.3046.528

Nevers, Y., Glover, N. M., Dessimoz, C., & Lecompte, O. (2023). Protein length distribution is remarkably uniform across the tree of life. *Genome Biology*, *24*(1), 135. https://doi.org/10.1186/s13059-023-02973-2

Newton, M. S., Cabezas-Perusse, Y., Tong, C. L., & Seelig, B. (2020). *In Vitro* Selection of Peptides and Proteins—Advantages of mRNA Display. *ACS Synthetic Biology*, *9*(2), 181–190. https://doi.org/10.1021/acssynbio.9b00419

Newton, M. S., Morrone, D. J., Lee, K., & Seelig, B. (2019). Genetic Code Evolution Investigated through the Synthesis and Characterisation of Proteins from Reduced-Alphabet Libraries. *ChemBioChem*, *20*(6), 846–856. https://doi.org/10.1002/cbic.201800668

Oba, Y., Koga, T., Takano, Y., Ogawa, N. O., Ohkouchi, N., Sasaki, K., Sato, H., Glavin, D. P., Dworkin, J. P., Naraoka, H., Tachibana, S., Yurimoto, H., Nakamura, T., Noguchi, T., Okazaki, R., Yabuta, H., Sakamoto, K., Yada, T., Nishimura, M., … Hayabusa2-initial-analysis SOM team. (2023). Uracil in the carbonaceous asteroid (162173) Ryugu. *Nature Communications*, *14*(1), 1292. https://doi.org/10.1038/s41467-023-36904-3

Ponnamperuma, C., & Peterson, E. (1965). Peptide Synthesis from Amino Acids in Aqueous Solution. *Science*, *147*(3665), 1572–1574. https://doi.org/10.1126/science.147.3665.1572

Rabinowitz, J., Flores, J., Krebsbach, R., & Rogers, G. (1969). Peptide Formation in the Presence of Linear or Cyclic Polyphosphates. *Nature*, *224*(5221), 795–796. https://doi.org/10.1038/224795a0

Ratelade, J., Miot, M.-C., Johnson, E., Betton, J.-M., Mazodier, P., & Benaroudj, N. (2009). Production of Recombinant Proteins in the *lon* -Deficient BL21(DE3) Strain of *Escherichia coli* in the Absence of the DnaK Chaperone. *Applied and Environmental Microbiology*, *75*(11), 3803–3807. https://doi.org/10.1128/AEM.00255-09

Russell, R., Jordan, R., & McMacken, R. (1998). Kinetic Characterization of the ATPase Cycle of the DnaK Molecular Chaperone. *Biochemistry*, *37*(2), 596–607. https://doi.org/10.1021/bi972025p

Shaw, G. H. (2016). The Miller-Urey Experiment and Prebiotic Chemistry. In G. H. Shaw, *Earth's Early Atmosphere and Oceans, and The Origin of Life* (pp. 41–48). Springer International Publishing. https://doi.org/10.1007/978-3-319-21972-1_6

Skowyra, D., & Wickner, S. (1995). GrpE Alters the Affinity of DnaK for ATP and Mg2+. *Journal of Biological Chemistry*, *270*(44), 26282–26285. https://doi.org/10.1074/jbc.270.44.26282

Świeżyński, A. (2016). Where/when/how did life begin? A philosophical key for systematizing theories on the origin of life. *International Journal of Astrobiology*, *15*(4), 291–299. https://doi.org/10.1017/S1473550416000100

Theyssen, H., Schuster, H.-P., Packschies, L., Bukau, B., & Reinstein, J. (1996). The Second Step of ATP Binding to DnaK Induces Peptide Release. *Journal of Molecular Biology*, *263*(5), 657–670. https://doi.org/10.1006/jmbi.1996.0606

Trifonov, E. N. (2004). The Triplet Code From First Principles. *Journal of Biomolecular Structure and Dynamics*, *22*(1), 1–11. https://doi.org/10.1080/07391102.2004.10506975

Turanov, A. A., Xu, X.-M., Carlson, B. A., Yoo, M.-H., Gladyshev, V. N., & Hatfield, D. L. (2011). Biosynthesis of Selenocysteine, the 21st Amino Acid in the Genetic Code, and a Novel Pathway for Cysteine Biosynthesis. *Advances in Nutrition*, *2*(2), 122–128. https://doi.org/10.3945/an.110.000265

Van Der Gulik, P., & Speijer, D. (2015). How Amino Acids and Peptides Shaped the RNA World. *Life*, *5*(1), 230–246. https://doi.org/10.3390/life5010230

Wang, M., Yafremava, L. S., Caetano-Anollés, D., Mittenthal, J. E., & Caetano-Anollés, G. (2007). Reductive evolution of architectural repertoires in proteomes and the birth of the tripartite world. *Genome Research*, *17*(11), 1572–1585. https://doi.org/10.1101/gr.6454307

Wieczorek, R., Adamala, K., Gasperi, T., Polticelli, F., & Stano, P. (2017). Small and Random Peptides: An Unexplored Reservoir of Potentially Functional Primitive Organocatalysts. The Case of Seryl-Histidine. *Life*, *7*(2), 19. https://doi.org/10.3390/life7020019

Zhang, Y., & Gladyshev, V. N. (2007). High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm Olavius algarvensis. *Nucleic Acids Research*, *35*(15), 4952–4963. https://doi.org/10.1093/nar/gkm514