ABSTRACT: Herein, we examine the complicated network of interactions among RNA, the metabolome, and the metalome in conditions that mimic the *E. coli* cytoplasm. First, we determined Mg^{2+} binding constants for the top 15 *E. coli* metabolites, comprising 80% of the total metabolome, at physiological pH and monovalent ion concentrations. Then, we used this information to inform creation of artificial cytoplasms that mimic *in vivo E. coli* conditions, termed Eco80. We empirically determined that the mixture of *E. coli* metabolites in Eco80 approximates single sit binding behavior towards Mg2+ in the biologically relevant free Mg2+ range of ~0.5 to 3 mM Mg^{2+}, using a Mg^{2+} binding fluorescent dye (8-Hydroxy-5-quinolinesulfonic acid). Furthermore, we examined the effects of Eco80 conditions on the thermodynamic stability, chemical stability, catalysis, and compactness of RNA. We find that these Eco80 conditions lead to opposing effects, wherein thermodynamic stability of RNA helices were weakened but chemical stability, compactness, and catalysis were enhanced. We propose a mechanism where increased RNA compactness and catalysis is facilitated in Eco80.

Introduction

Summary of progress on in vivo-like conditions.

Studies that consider cellular components one at a time.

Studies that consider cellular components together in artificial cytoplasm.

In contrast, we take a bottom up, *aufbau*, approach that builds up complexity, to make an artificial cytoplasm that contains 80% of *E. coli* metabolites with biologically relevant concentrations of monovalent ions and free Mg^{2+} ions. This *aufbau* approach allows us to understand the effects of 80% of metabolite and metal ion species that compose a part of the network of interactions that RNA experiences in *E. coli* cells.

Eco80: An artificial cytoplasm containing 80% of *E. coli* metabolites

E. coli cells contain hundreds of metabolites (about 243 mM total), which is too many metabolites to test systematically. However, 15 abundant metabolites, an experimentally manageable number, comprise 80% (195 mM) of total metabolites (Figure 1A). Thus, we sought to prepare Eco80, an artificial cytoplasm containing the biological concentrations of the 15 most abundant metabolites in *E. coli* (Table 1).

All of the metabolites in Eco80 are zwitter ions, or negatively charged, near physiological pH (~7) and require electrostatic neutralization with metal ions. We prepared Eco80 so that the final monovalent ion concentration was the physiological value of 240 mM Na^+ and 140 mM K^+ (Supplementary information (SI) Table 1). Metabolite salts and free acids were prepared to a final 2xconcentration, and the amount of Na^+ and K^+ added with each metabolite was recorded. Next, the pH of the 2xstock was adjusted to pH 7.0 using NaOH, and the amount of Na^+ was recorded. Lastly, NaCl and KCl were added to a final 240 mM Na^+ and 140 mM K^+. We thus created Eco80, at a 2x final concentration so that it could be diluted into other reagents for experiments.

Next, we considered how metabolites effect the speciation of free and chelated Mg^{2+}. All 15 Eco80 metabolites have functional groups, carboxylates and phosphates, that drive chelating interactions with Mg^{2+} ions (Table 1), and we have previously estimated that the metabolite pool in *E. coli* has potential to chelate 51 mM Mg^{2+}, assuming 2 mM free Mg^{2+}, at an ionic strength of 0.15 M and a pH of 7.5. While extensive literature exists on chelating interactions between Mg^{2+} and small molecules, our previous estimates are putative as Mg^{2+} binding affinity is dependent on environmental factors such as pH, ionic strength, the composition of background ions, and temperature. Thus, we sought to better characterize Mg^{2+} chelation by the metabolites in Eco80, at the physiological background.

We determined apparent disassociation constants (K_D) for Eco80 metabolites in 240 mM NaCl, 140 mM KCl, pH 7.0 buffer at 37 °C (Table 1). Isothermal titration calorimetry (ITC) was used to determine K_Ds for phosphorylated metabolites (SI figure 1, SI Table 2). A fluorescence assay, that measures the free Mg^{2+} concentration in a sample using the metal ion binding dye 8-Hydroxy-5-quinolinesulfonic (HQS) acid, was used to estimate the K_D for Mg^{2+} of metabolites that did not produce enough heat on binding Mg^{2+} to measure

with ITC (SI figure 2, SI Table 3). For the HQS assay, Mg^{2+} is titrated into HQS solutions in the absence and presence of chelators. HQS emission as a function of the total Mg^{2+} in the absence of chelators is then fit to a binding model (SI figure 2A, top blue data and black fit). The free Mg^{2+} concentration is then then calculated from the fluorescence emission for each data point using the binding model, providing the free Mg^{2+} concentration as a function of the total Mg^{2+} concentration (SI figure 2A, bottom). Mg^{2+} binding by metabolites is thus observed by fitting the free Mg^{2+} concentration as a function of the total Mg^{2+} concentration, which is shifted to the right as Mg^{2+} is sequestered by metabolites.

The four nucleotide triphosphates, ATP, UTP, GTP, and dTTP, were classified as strong Mg^{2+} binders, with K_D values less than the approximate free Mg^{2+} concentration in *E. coli,* 2 mM (Table 1). Conversely, 8 other metabolites, L-glutamic acid, fructose 1,6-BP, UDP-N-acetylglucosamine, Glucose 6-phosphate, L-aspartic acid, 6-Phospho-gluconic acid, dihydroxyacetone phosphate, and pyruvic acid were classified as weak Mg^{2+} binders, with a K_D value greater than 2 mM (Table 1). Three metabolites, glutathione, L-valine, and L-glutamine had negligible Mg^{2+} binding properties, as measured with HQS (SI figure 2). We thus broke Eco80 down into two other artificial cytoplasm, NTP chelated Mg^{2+} (NTPCM), and weak metabolite chelated Mg^{2+} (WMCM), composed of the strong Mg^{2+} chelators (NTPs), and weak Mg^{2+} chelators, respectively (Table 1).

We next used two methods to estimate how Eco80 metabolites effect the speciation of free and chelated Mg^{2+} as a mixture. The first method was experimental, using HQS emission to estimate the free Mg^{2+} concentration in the presence of metabolites (Figure 1B-C, SI Table 3). The second method was a statistical model that accounts for experimental uncertainty in metabolite concentrations and uncertainty in K_D determination, based on single-site binding (meaning that one metabolite associates one Mg^{2+}). The statistical model is described in detail in the Supplementary Methods. Briefly, concentration errors were propagated from uncertainties in reagent masses and volumes used during sample preparation, and K_D uncertainties were obtained from the fits (Table 1). Both uncertainties were then randomly seeded into Equation 1 to create 1000 virtual artificial cytoplasms, where $[Mg]_T$ is the total Mg^{2+} concentration, $[Mg]$ is the free Mg^{2+} concentration, i is an integer representing each metabolite in a mixture, N is the total number of metabolites in a mixture, $[L_i]_T$ is the concentration of the "*i'th*" metabolite in a mixture, and K_D is the disassociation constant.

$$[Mg]_T = [Mg] + \sum_{i=1}^{N} \frac{[L_i]_T [Mg]}{K'_{Di} + [Mg]} (10)$$

Then, equation 1 is solved numerically to determine the free Mg^{2+} concentration produced at a given total Mg^{2+} concentration in a virtual artificial cytoplasm.

The two methods indicate that Mg^{2+} speciates in artificial cytoplasms according to a single-site model within or below the biological free Mg^{2+} range of 0.5 to 3 mM Mg^{2+}, but not at higher free Mg^{2+} concentrations (Figure 1 E-F). For example, in Eco80, the statistical model suggests that the free Mg^{2+} should increase slowly as the total Mg^{2+} concentration is increased, until the strong Mg^{2+} chelators (NTPs) become saturated at about 27 mM total Mg^{2+} (Figure 1E, hex bins). At total Mg^{2+} concentrations higher than 27 mM, the free Mg^{2+} should increase faster because the NTPs are saturated by Mg^{2+} and the weak chelators sequester less Mg^{2+}. Free Mg^{2+}, calculated using HQS emission shows a similar trend to the statistical model below 3 mM Mg^{2+} free (Figure 1E, data points). However , the free Mg^{2+} concentration calculated from HQS emission does not increase with the total Mg^{2+} as fast as the single-site model would predict above 3 mM free Mg^{2+}, indicating that multivalent interactions, where one metabolite interacts with several Mg^{2+} molecules, dominate the equilibrium. Non-single-site behavior above 3 mM free Mg^{2+} is also observed in the NTPCM and WMCM artificial cytoplasms (Figure 1 F & G).

Lastly, we sought to empirically determine how much total Mg^{2+} is required to maintain a free Mg^{2+} concentration of 2 mM in Eco80, NTPCM, and WMCM. The relationship between the free Mg^{2+} calculated from HQS emission and the total Mg^{2+} concentration was fit to a polynomial to empirically approximate the data (Figure 1 E-G, colored lines), and the total Mg^{2+} concentration required to produce 2 mM Free Mg^{2+} was calculated from the polynomial fit. This resulted in a predicted 31.6, 25.0, and 6.5 mM total Mg^{2+} concentration to produce 2 mM free Mg^{2+} in Eco80, NTPCM, and WMCM, respectively (Table 2).

Thermodynamic analysis of RNA helices in Eco80

We sought to understand how Eco80 effects the stability of RNA helices composed of Watson-Crick base pairs. RNA helix stability is traditionally measured with UV-absorbance melting curves, usually at 260 or 280 nm, which allows for accurate calculation of helix folding energies. However, absorbance melting curves were not appropriate for measuring helix stability in Eco80 because of the high absorbtivity of ATP, UTP, GTP, dTTP, and UDP-N-acetylglucosamine. Thus, we used a fluorescence binding isotherm assay, which is optically orthogonal to Eco80.

Helix stability was measured using the emission of a 5'-fluorophore labeled RNA strand (FAM-RNA) in equilibrium with a complementary 3'-quencher labeled RNA strand (RNA-BHQ1), (Figure 2A). High emission indicates that the FAM-RNA is single-stranded and low emission indicates that the FAM-RNA is bound in duplex with a RNA-BHQ1 strand (Figure 2A). We used a binding isotherm method, where RNA-BHQ1 is titrated into a constant concentration of FAM-RNA at a single temperature (SI figure 3), resulting in an apparent binding isotherm (Figure 2B). Fluorescence binding isotherms were favored over fluorescence melts because of the unpredictable dependence of FAM emission on temperature. FAM emission was monitored in a Real-Time PCR instrument at different temperatures, resulting in a isotherm every 0.5 °C from 20 to 80 °C (Figure 2B). Fluorescence isotherms were fit at each temperature to SI equation X, thus determining the K_D between the FAM-RNA strand and the RNA-BHQ1 strand.

Raw fluorecence was fit with a new program called MeltR, created by the authors, to determine folding energies. MeltR is a package of functions in the popular R programming language, that allows facile conversion of raw data to folding energies. Importantly, MeltR handles two sources of experimental error, uncertainty in RNA concentration and inaccurate K_Ds collected at low and high temperatures. MeltR then calculates folding energies using two Van't Hoff methods, directly fitting a Van't hoff plot (Figure 3C) and globally fitting raw fluorescence emission to SI equation X.

We found that helix energies from fitting fluorescence binding isotherms are highly dependent on the errors in the determination of RNA concentrations in stock solutions, which is propagated systematically during sample preparation (SI Figure 3). To understand why, we modeled data assuming a folding enthalpy (ΔH), entropy (ΔS), and Gibb's free energy at 37 °C ($\Delta G^{37°C}$) of -56.2 kcal/mol, -136. 4 cal/mol/K, and -13.9 kcal/mol respectively, 5% random fluorescence error, and perfectly accurate determination of RNA concentrations in concentrated stocks (a 200 nM FAM-RNA concentration, and 0, 1, 10, 50, 100, 150, 200, 250, 400, 600, 800, and 1000 nM RNA-BHQ1 concentrations). We then used MeltR to fit the modeled data, resulting in accurate determination of ΔH = -56.0 kcal/mol, ΔS = -135.7 cal/mol/K, and ΔG = -13.9 kcal/mol. We next considered how assuming incorrect RNA stock concentrations, thus incorporating a systematic error, could effect the accuracy of the fits. Systematic error (-50% to +50%) was seeded into virtual stock concentrations and the data were refit (SI figure 4A). We found that fit accuracy what highly dependent on error in stock concentrations, unless the FAM-RNA error and the RNA-BHQ1 error compensated for each other, e.g. %RNA-BHQ1 error = %FAM-RNA error.

We next considered a more realistic scenario, where the experiment assumes perfectly accurate determination of RNA concentrations but there is actually +20% FAM-RNA concentration error. Data was modeled using the same folding energies and RNA-BHQ1 concentrations, but with a 240 nM FAM-RNA

concentration (+20% error). We then used MeltR to fit the modeled data, assuming a 200 nM FAM-RNA concentration, resulting in inaccurate determination of ΔH = -35.8 kcal/mol, ΔS = -75.5 cal/mol/K, and ΔG = -12.3 kcal/mol. Once again, we seeded systematic error (-50% to +50%) into virtual stock concentrations and the data were refit (SI figure 4B). Similar to SI figure 4A, we found that fit accuracy what highly dependent on errors in stock concentrations. However, the fits were most accurate according to Equation 1 instead of where %RNA-BHQ1 error = %FAM-RNA error:

$$\%BHQ1_{error} = \frac{1}{X}\%FAM_{error} + \frac{100 - 100\,X}{X}$$

Where X is the actual FAM-RNA concentration divided by the estimated FAM-RNA concentration (240 nM/200 nmol = 1.2 in this example). Thus, MeltR does not need perfectly accurate concentrations, just an optimization algorithm that finds the FAM-RNA concentration correction factor X. To find X, MeltR selects an isotherm (usually the lowest temperature), where the K_D is more than 10 times less than the FAM-RNA labeled concentration (SI figure 4C). At this K_D range, the shape of the binding curve is independent of K_D, and MeltR uses the isotherm as a Job plot to determine X. MeltR then uses X to correct the FAM-RNA concentration. We next tested the MeltR optimization algorithm. The modeled data, with +20% FAM-RNA concentration error, was fit using MeltR with the concentration optimization algorithm on, resulting in an accurate determination of ΔH = -51.9 kcal/mol, ΔS = -123.5 cal/mol/K, and ΔG = -13.6 kcal/mol. We then seeded additional error into the data set (-50% to +50%), refit the data using the MeltR concentration optimization algorithm, and found that MeltR calculates accurate folding energies (within 0.2 kcal/mol in terms of the ΔG) even with large inaccuracies in reagent concentration estimates (SI Figure 4D).

MeltR then filters out isotherms that produce inaccurate K_Ds, according to user specifications. Fits are first filtered by magnitude because fits K_Ds are most accurate in the range of more than the FAM-RNA concentration/10 and less than the FAM-RNA concentration times 10 (SI figure 5). K_Ds below the FAM-RNA concentration/10 are inaccurate because the shape of the isotherm curve is independent of the K_D, as the FAM-RNA stand is stoichiometrically bound by RNA-BHQ1. Likewise, K_Ds above 10 times the FAM-RNA concentration are also inaccurate because not enough FAM-RNA is binding to RNA-BHQ1 to generate a curve. After filtering by magnitude, MeltR filters K_Ds by the standard error in the fit, so that the most accurate K_Ds are used to determine folding energies. Both the MeltR K_D range, and K_D error threshold can be adjusted by the user to refine fits. For example, with the MeltR fit of the modeled data with +20% FAM-RNA concentration error and the concentration optimization algorithm on, can be improved to more accurate helix energies, ΔH = -56.4 kcal/mol, ΔS = -136.9 cal/mol/K, and ΔG = -13.9 kcal/mol

Eco80 destabilizes RNA helices

Paragraph 1: Summary of helix energies and error propagation for values reported in Table 1

We used fluorescence binding isotherms to determine helix folding energies in background monovalent metal ion control (240 mM NaCl 140 mM KCl), Eco80, NTPCM, and WMCM, for a data set of five representative RNA helices. All solutions contain 2 mM Mg^{2+}, as per Table 2. We used both methods to determine folding energies in MeltR, where method 1 was fitting the Van't Hoff relationship between the disassociation constant and temperature and method 2 was directly globally fitting the raw fluorescence binding isotherms (SI table 4). Standard errors estimated from the fit for method 1 were 1.7%, 2%, and 0.3 % on average for the ΔH, ΔS, and ΔG$^{37°C}$. Standard errors estimated from the fit for method 2 were larger, at 21.7%, 26.5%, and 3.4 % on average for the ΔH, ΔS, and ΔG$^{37°C}$. However, the difference in helix folding energies between the two methods between the two methods was much smaller, at 1.2%, 1.4%, and 0.2% on average for the ΔH, ΔS, and ΔG$^{37°C}$. Given that the two methods provide similar helix folding energies, but different standard errors, the discrepancy in standard error likely reflects differences in the number of parameters that must be estimated

by the fit (two for method 1, ΔH and ΔS, and 2+2*number of raw isotherms for method 1, ΔH, ΔS, and a F_{max}/F_{min} for each raw isotherm), more than it reflects systematic and random errors in helix folding energy estimation. Turner and colleagues estimated an uncertainty of 12%, 13.5%, and 4% for the ΔH, ΔS, and $\Delta G^{37°C}$, respectively, conservatively reflect systematic and random errors in error determination for absorbance melting curves by comparing helix energies collected on the same sequences by different labs. Given that the discrepancy between method 1 and method 2 were smaller than the Turner uncertainty for all fluorescence binding isotherms in this study, and that we are determining differences between conditions on the same sequences collected in the same lab, the 4% value in terms of the is too conservative. We determined that there was on average 0.2 kcal/mol, or 1.5%, error for MeltR fitting modeled fluorescence data using the concentration optimization algorithm (SI Figure 4), Thus, we reported the $\Delta G^{37°C}$ of helix formation from method 1 with an uncertainty of 1.5% in Table 3 and the 1.5% uncertainty was propagated to the association constant in Figure 2D.

All five representative helices were significantly destabilized in Eco80, meaning the the $\Delta\Delta G^{37°C}$ between the background monovalent condition and Eco80 was larger than the propagated uncertainty in $\Delta\Delta G^{37°C}$, by 0.44 to 1.12 kcal/mol (Table 3, Figure 2D). To better understand how components of Eco80 destabilize RNA helices, we then analyzed the effects of strong Mg2+ chelating metabolites and weak Mg2+ chelating metabolites, separately. NTPCM, which is composed of strong Mg2+ chelating metabolites, consistently destabilized RNA helices (Figure 2D). Helices 1, 2, 3, and 5 were significantly destabilized by 0.41 to 0.60 kcal/mol (Table 3). NTPCM did not significantly destabilize Helix 4, however, the the 0.32 kcal/mol destabilization effect is consistent with the other 5 helices. In contrast to NTPCM, WMCM, which is composed of strong Mg2+ chelating metabolites, destabilized, had no effect, or stabilized RNA helices depending on the helix identity (Figure 2D). WMCM significantly destabilized helix 3 and 4 by 0.82 and 0.40 kcal/mol respetively. No significant change in stability was observed for helices 2 and 5 in WMCM. Lastly, helix 1 was significantly stabilized in WMCM by about -0.4 kcal/mol. Thus, the net effect of Eco80 on RNA helices is destabilization, with destabilizing interactions dominating for strong Mg^{2+} chelating metabolites, and a mixture of stabilizing and destabilizing interactions for weak Mg^{2+} chelating metabolites.

Eco80 protects RNA from degradation

E. coli metabolites promote RNA compactness and RNA catalysis

Discussion

Making Eco80 artificial cytoplasm

-Complex but manageable

Effects on helices

-Mg stabilizes (from lit)

-Polor osmolytes destabilize (from lit)

-Here we analyze their effects together and see opposing effects

Propose model that summaries effects