Cell Geometry and Membrane Protein Crowding Constrain Growth Rate, Overflow Metabolism, Respiration, and Maintenance Energy.

Ross P. Carlson[1*], Ashley E. Beck[2], Mauricio Garcia Benitez[3], William R. Harcombe[4], Radhakrishnan Mahadevan[3], Tomáš Gedeon[5]

1. Department of Chemical and Biological Engineering, Center for Biofilm Engineering, Montana State University, Bozeman, MT USA
2. Department of Biological and Environmental Sciences, Carroll College, Helena, MT USA
3. Department of Chemical Engineering and Applied Chemistry, University of Toronto, Canada
4. Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN USA
5. Department of Mathematical Sciences, Montana State University, Bozeman, MT USA

[*]corresponding author: R.P.C., rossc@montana.edu

Keywords: systems biology, cell biology, metabolism, cell geometry

Abstract

A metabolic theory is presented for predicting maximum growth rate, overflow metabolism, respiration efficiency, and maintenance energy flux based on the intersection of cell geometry, membrane protein crowding, and metabolism. The importance of cytosolic macromolecular crowding on phenotype has been established in the literature but the importance of surface area has been largely overlooked due to incomplete knowledge of membrane properties. We demonstrate that the capacity of the membrane to host proteins increases with growth rate offsetting decreases in surface area-to-volume ratios (SA:V). This increase in membrane protein is hypothesized to be essential to competitive *Escherichia coli* phenotypes. The presented membrane-centric theory uses biophysical properties and metabolic systems analysis to successfully predict the phenotypes of *E. coli* K-12 strains, MG1655 and NCM3722, which are genetically similar but have SA:V ratios that differ up to 30%, maximum growth rates on glucose media that differ by 40%, and overflow phenotypes that start at growth rates that differ by 80%. These analyses did not consider cytosolic macromolecular crowding, highlighting the distinct properties of the presented theory. Cell geometry and membrane protein crowding are significant biophysical constraints on phenotype and provide a theoretical framework for improved understanding and control of cell biology.

INTRODUCTION

Bacterial cell geometry is highly regulated and constrains the surface area available for acquiring nutrients and the volume available for synthesizing proteins (1-4). Cell geometry also impacts molecular crowding by fixing the dimensions of the spaces occupied by macromolecules (5-9). Three-dimensional, cytosolic molecular crowding has been analyzed both explicitly and implicitly in systems biology studies to predict catabolite repression and phenotype shifts including overflow metabolism (5-7, 10, 11). However, phenotypic constraints based on two-dimensional molecular crowding on a surface are not well studied (12, 13). Systems biology studies have applied three-dimensional enzyme metrics such as total protein mass to approximate crowding on a two-dimensional membrane surface, but these approaches generally miss the significant difference between a two- and three-dimensional crowding constraint (10, 14, 15). A quantitative and predictive understanding of the intersection of cellular geometry, two-dimensional membrane protein crowding, and phenotype is largely unexplored (16-18). These are the foci of the current study.

Membranes provide a barrier against environmental stressors, retain macromolecules, provide a platform for the selective transport of molecules, and play a critical role in energy metabolism including respiration (15, 19)(Fig. 1A). The lipid bilayer is a major building block of the membrane and has a finite capacity to host embedded and adsorbed proteins due to a combination of protein crowding and the loss of membrane integrity at high protein densities (19-23). While theoretical studies have addressed the role of surface area on metabolism (12, 13), no study has developed a quantitative and predictive molecular level theory that accounts for strain-specific differences in SA:V ratios, growth rate dependent changes in SA:V ratios, and growth rate dependent changes in membrane protein crowding (24). Additionally, the current study makes a distinction between two- and three-dimensional protein crowding on the two-dimensional membrane surface (10, 14). Enzyme requirements for membrane surface area are not readily predictable from the enzyme mass or volume. In fact, enzyme volume and occupied membrane surface area have a poor correlation, as we will show later from compiled literature data, highlighting the need to study each separately.

Membranes support major fluxes for macromolecule synthesis and maintenance energy generation. Maintenance energy is cellular energy consumed for functions other than the production of biomass components (25). Meeting maintenance energy requirements accounts for a significant fraction (30-90+%) of substrate fluxes (26-29). Metabolic pathways for primary macromolecules are well documented, but maintenance energy is inferred from calculations requiring several assumptions (26, 30). It can be difficult to compare maintenance energy values across studies because the assumptions necessary for their calculation are often not reported (30, 31). Furthermore, maintenance energy calculations have not considered the impact of biophysical constraints like finite membrane surface area and membrane protein crowding.

Escherichia coli is a facultative anaerobe and a convenient host for studying tradeoffs between cell geometry, membrane protein crowding, and phenotype. Two well studied and genetically similar *E. coli* K-12 strains, MG1655 and NCM3722, have distinct phenotypes including different maximum growth rates on minimal salts medium, different acetate overflow metabolisms, and different cellular geometries. NCM3722 has a specific growth rate on glucose salts media that is ~40% faster than MG1655 (0.97 ± 0.06 h^{-1} vs. 0.69 ± 0.02 h^{-1}, respectively, supplementary data S1). MG1655 displays an acetate overflow metabolism at growth rates $\geq 0.4 \pm 0.1$ h^{-1} while NCM3722 displays acetate overflow at growth rates $\geq 0.75 \pm 0.05$ h^{-1}, which are faster than μ_{max} for MG1655 (supplementary data S1). Additionally, MG1655 has a cellular volume that is approximately two times larger than NCM3722 at a growth rate of 0.65 h^{-1}, whereas NCM3722 has a surface area to volume (SA:V) ratio ~30% larger than MG1655 at this growth rate (supplementary data S2). Cells display exquisite control over their geometry, and we propose that this helps balance the flux of nutrients and energy between surface area- and volume-associated processes (1-3) and present a systems biology theory centered around membrane surface area and protein crowding to test and support this proposal.

RESULTS

Membrane protein capacity increases with growth rate, offsetting decreases in SA:V ratio.

The inner membrane provides finite surface area for nutrient transport and energy generation (Fig. 1A). The two-dimensional occupancy of proteins is not necessarily indicative of their three-dimensional volume (Fig. 1B), and analyses of 22 central metabolism enzymes shows no clear

correlation between membrane surface area requirements and enzyme volume (Fig. 1C)(supplemental data S1). *E. coli* SA:V ratios decrease with increasing growth rate (Fig. 1D) due to increases in cell length and diameter (2, 16)(supplementary data S2). While the SA:V ratio decreases with growth rate, the capacity of the membrane to host proteins per cell volume increases with growth rate (Fig. 1E)(32-35)(supplementary data S3). We hypothesize this increase in membrane protein density is central to optimal phenotypes as it enables essential transport of substrates, cycling of redox cofactors, and generation of cellular energy. Membrane protein loading in Fig. 1E is presented on a cell volume basis to facilitate comparison with the SA:V data in Fig. 1D. Other normalizations are presented in supplementary data S2. The areal density of membrane proteins for *E. coli* K-12 strain MG1655 was determined using three curated proteomics data sets (32-35) and cell geometry data (2). Proteomics data from *E. coli* K-12 strains MG1655 and BW25113 were pooled for the presented analyses as the related strains have similar growth rates, cell geometries, and phenotypes (36). Copy number of membrane-associated proteins per cell was converted into an occupied surface area using enzyme properties and cell geometry (2, 32-35)(supplementary data S3). Membrane protein hosting increased faster than the decrease in surface area per volume, resulting in a net gain of membrane catalytic potential with growth rate (Fig. 1D, 1E). The increase in membrane protein is proposed to be a result of the growth rate dependent increase in cell diameter which helps stabilize the membrane at higher protein loading (2, 3, 22, 37, 38). The presented trends are limited to growth rates supported by glucose salts media. The cited proteomics studies used in Fig. 1E included data for faster growing cultures. However, faster growth requires complex media with multiple substrates. The cited studies did not measure which of the available substrates were consumed, preventing an analysis of utilized surface area. SA:V data in Fig. 1D do not require substrate consumption data and therefore are plotted over a larger range of growth rates.

Growth rate and biomass yield are constrained by membrane surface area, protein crowding, and maintenance energy fluxes.

Membrane surface area and membrane capacity to host proteins are finite (1, 12, 13, 17, 18). We hypothesize these biophysical properties constrain phenotype and developed a systems biology theory to test the hypothesis. The specific membrane surface area (*sMSA*, units: nm^2 $(g\ cdw)^{-1}$)

required for membrane-associated enzyme fluxes was quantified using enzyme properties including flux per g cell dry weight (cdw), surface area requirements, and k_{cat}. Total *sMSA* required for a phenotype was tabulated by summing the contribution of all active membrane-associated enzymes and by accounting for the growth rate dependent SA:V ratio. Mathematical development is presented in Box 1 while parameters are reviewed in supplementary data S4.

The *sMSA* constraint was evaluated using a metabolic model of *E. coli* (39, 40) and flux balance analysis (FBA)(41). The *sMSA* constraint was implemented by incorporating a flux proportional membrane surface area requirement for each membrane-associated reaction, termed MRE_i (membrane real estate)(Box 1)(12). Flux through membrane-associated enzymes was constrained by a finite *sMSA* pool which varied in magnitude based on growth-rate dependent SA:V ratios, strain-specific SA:V ratios, and protein crowding (quantified as the fraction of the Membrane Surface Area occupied by central metabolism protein, f_{MSA})(supplementary data S5, Box 1). The *sMSA* balance was enforced concurrently but orthogonally to standard FBA metabolite mass balances (mmol (g cdw h)$^{-1}$). The *in silico* theory was termed specific Membrane Surface Area-constrained Flux Balance Analysis (*sMSAc-FBA*) (41). This formulation contrasts with previous studies where the weights on the membrane-associated fluxes were the molecular masses of the protein (10, 14, 15, 42) as opposed to the occupied membrane surface area as in the presented work. Case specific constraints and objective functions are described below and in the supplementary material.

Cell growth requires fluxes for biomass synthesis and maintenance energy. The theoretical impact of finite *sMSA* on these fluxes was quantified for *E. coli* strain MG1655 growing on glucose salts medium (Fig. 2A). These analyses were entirely predictive. *sMSAc-FBA* simulations applied a general matrix of biomass synthesis rates and maintenance energy fluxes as constraints while the FBA objective function minimized glucose flux. Metabolic processes competing for a finite pool of *sMSA* resulted in rate-yield tradeoffs (13, 43). Phenotypes with high biomass yields (g biomass (g glucose)$^{-1}$) required larger investments of *sMSA* for high efficiency ETC components, quantified by the P/O number (Fig. 2B). The P/O number is the ratio of ATP produced via respiration per two electrons transferred to O_2. *E. coli* can theoretically operate ETC configurations with P/O numbers from ~0 - 2 as discussed in more detail below (44-46). ETC efficiency is not influenced by substrate-level phosphorylation such as ATP synthesis

associated with acetate overflow. Phenotypes with low growth rates and small maintenance energy fluxes can support high P/O numbers because of the modest *sMSA* requirements for transporters and ATP synthase (upper left Fig. 2B). Phenotypes with high growth rates require larger investments of *sMSA* for glucose transporters (PtsG) which necessitates submaximal P/O numbers (Fig. 2B, Fig. 2C, supplementary data S6). These tradeoffs generate gradients in other phenotypic properties including the potential for maintenance energy fluxes (q_{ATP})(Fig. 2A), use of parallel ETC enzymes Nuo (Fig. 2D) or Ndh II (Fig. 2E), and acetate overflow metabolism ($q_{acetate}$)(Fig. 2F). Secretion of acetate generates ATP via substrate-level phosphorylation and reduces the need for *sMSA*-intensive ETC components and ATP synthase complexes (12, 13). Presented q_{ATP} fluxes account for the capacity to generate the sum of growth- and nongrowth-associated maintenance energy and did not include ATP required to polymerize monomers into macromolecules for biomass synthesis (supplementary data S5).

Extracellular fluxes predict a specific membrane surface area usage.

E. coli K-12 strains MG1655 and NCM3722 are closely related (24) yet have distinct phenotypes and cell geometries (supplementary data S1, S2). Experimental data from six MG1655 batch cultures (47) and two NCM3722 batch cultures (48), all grown on glucose salts media, were mapped to strain-specific, rate-yield phenotype spaces to interpret the experimental data through the lens of cell geometry and membrane protein crowding. *In silico* simulations for MG1655 and NCM3722 used the same metabolic model with the same enzyme parameters to facilitate interpretation and to increase transparency. However, strain-specific SA:V ratios which arise from different cell dimensions were applied (Fig 1D). The similarity between experimental and *in silico* extracellular fluxes was quantified using Euclidean distance. Each experimental flux was evaluated 100 times while randomly perturbing the fluxes within an assumed error range of ±7%. The mean Euclidean distance from the 100 simulations is plotted in Fig 3A. The *sMSA* constraint defines a relationship between growth rate, glucose consumption rate, and overflow metabolism ($q_{acetate}$) due to enzymes competing for limited *sMSA*. Experimental fluxes for both K-12 strains were most similar to simulations with a membrane protein crowding of f_{MSA} = 0.05-0.07 (5-7% of surface area occupied by central metabolism enzymes, Box 1)(Fig. 3A). These values are lower than values estimated by Szenk *et al*. (12). The current study adjusted enzyme k_{cat} values to a culture temperature of 37°C, resulting in faster k_{cat} values and therefore smaller

f_{MSA} requirements (49). Both strains had similar predicted protein crowding despite a 40% difference in maximum growth rate. MG1655 and NCM3722 have different SA:V ratios as a function of growth rate, and the predicted f_{MSA} values also occur at different growth rates potentially providing NCM3722 an advantage for hosting membrane protein. The f_{MSA} parameter range for both *E. coli* K-12 strains is further tested and interpreted in the sections to follow.

μ_{MAX} is a Pareto tradeoff between growth rate and maintenance energy generation.

Finite *sMSA* creates phenotypic tradeoffs providing a new theory to interpret maximum growth rate. Consensus extracellular flux data were assembled for *E. coli* strains MG1655 and NCM3722 growing on glucose salts media. Data span growth rates from 0.1 h^{-1} to μ_{max} (supplementary data S1)(6, 34, 42, 50, 51). These data did not include the data used in Fig. 3A to keep the analyses independent. The experimental consensus data were interpreted using *sMSA* theory by applying the fluxes as constraints to *sMSAc-FBA* analyses and quantifying maximum possible maintenance energy generation (q_{ATP}) as a function of growth rate, cell geometry, and f_{MSA} (0.04-0.08)(Fig 3B, 3C). Maintenance energy was defined here as any ATP produced that was not directly required to synthesize biomass precursors nor to polymerize monomers into macromolecules and would include growth and nongrowth-associated maintenance energy. q_{ATP} was calculated 100 times for each scenario by randomly and independently perturbing each experimental consensus flux within an assumed error interval (±7% of flux) and using these perturbed fluxes as *sMSAc-FBA* constraints. Data points in Fig. 3B and 3C are the mean from 100 simulations; error bars are the standard deviation.

Membrane protein crowding (f_{MSA}) has major biological significance as it constrains the energetic efficiency of respiration by altering the capacity of the cell to extract energy from substrates. For example, there was ~4-fold increase in MG1655 maintenance energy potential (10 ± 3 vs. 39 ± 4 mmol ATP (g cdw h)$^{-1}$) for the same consensus substrate fluxes (μ = 0.5 h^{-1}) if the enforced protein crowding increased from 4% to 8% (f_{MSA} = 0.04, 0.08)(Fig. 3B, supplementary data S6). Similar trends were observed for strain NCM3722 at μ = 0.7 h^{-1}. The maximum possible maintenance energy fluxes increased ~5-fold from f_{MSA} = 0.04 to 0.08 (7 ± 2 vs. 37 ± 4 mmol ATP (g cdw h)$^{-1}$), Fig. 3C, supplementary data S6). The difference in maintenance energy potential is based on the availability of *sMSA* required to conserve more of the substrate energy as ATP via the use of higher

efficiency ETC components and expressing more ATP synthase complexes. We hypothesize that *E. coli* operates at a maximum membrane protein capacity as this significantly increases the possible cellular energy fluxes, thereby improving fitness.

Positive slopes between growth rate and maintenance energy flux (Fig. 3B, 3C) indicate both rates can increase simultaneously without penalty as the membrane can accommodate a balanced increase of substrate transporters, ETC complexes, and ATP synthase complexes. A slope of zero indicates the growth rate can increase with no reduction in q_{ATP}; the membrane hosts more substrate transporters while holding the other processes constant. A negative slope indicates the growth rate can increase only with a concurrent decrease in q_{ATP}; the increased substrate transporter demand requires a decrease in ETC components and ATP synthase complexes. Negative slopes define a Pareto front between simultaneous increases in growth rate and q_{ATP} (Fig. 3B, 3C)(52). Experimental *E. coli* MG1655 cultures have a μ_{max} of 0.69 ± 0.02 h^{-1} on glucose salts media (blue shaded area, Fig. 3B) which coincides with the start of a negative slope between growth rate and q_{ATP} (f_{MSA} = 0.07). Experimental *E. coli* NCM3722 cultures growing on glucose salts media have a μ_{max} of 0.97 ± 0.06 h^{-1} (red shaded area, Fig. 3C). μ_{max} for NCM3722 also corresponds with the initiation of a phenotypic tradeoff between growth rate and q_{ATP} when f_{MSA} is between 0.07 and 0.08. The predicted membrane protein crowding range is similar to the independent data presented in Fig. 3A. Interestingly, strains MG1655 and NCM3722 have similar potential for q_{ATP} at their respective μ_{max} (38 ± 5 and 34 ± 4 mmol ATP (g cdw h)$^{-1}$) despite NCM3722 having a specific growth rate ~40% faster than MG1655 (Fig. 3B, 3C). All reactions in the *in silico* model were balanced for atoms and electrons (supplementary material S4). Therefore, all simulations necessarily balanced consensus fluxes with appropriate CO_2 and O_2 fluxes as no other electron donors or acceptors were available for the simulations.

Maximum growth rate maximizes the areal density of ATP synthase and rate of ATP hydrolysis.

Maximizing q_{ATP} corresponds with a maximal areal density of ATP synthase complexes, which provides a molecular basis for interpreting μ_{max}. The *in silico* areal density of the enzyme reached a maximum just prior to the start of overflow metabolism, coinciding with strain-specific μ_{max} (Fig. 3D). Predicted areal density of ATP synthase (complexes μm^{-2}) for strain MG1655 correlated well with experimental values from curated proteomics data (Fig. 3D)(32-35); there was no analogous

proteomics dataset for wildtype strain NCM3722. The MG1655 *in silico* rate-yield space was analyzed for the areal density of ATP synthase (Fig 3F, f_{MSA} = 0.07). The zone of maximal ATP synthase density corresponded with commonly observed *E. coli* rate-yield phenotypes (43). Strain MG1655 has an experimental μ_{max} of 0.69 h^{-1} while the available *sMSA* could support a specific growth rate of 1.4 h^{-1}, and it has an experimental biomass yield on glucose of 0.4 g g^{-1} while it is theoretically possible to support yields as high as 0.7 g g^{-1} (Fig 2A, supplementary data S1). Neither of these maxima are realized because maintenance energy fluxes divert substrate energy. Instead, *E. coli* K-12 realizes an intermediate growth rate and yield while maximizing the rate of substrate energy dissipation per area of membrane via ATP (53). Interestingly, this corresponds with maximizing biomass yield for a particular maintenance energy flux (white q_{ATP} isocline, Fig 3F). Systems biology studies typically adjust the growth- and nongrowth-associated maintenance energy parameters so that simulations and growth data align. Fig. 3F shows the range of solutions available from a stoichiometric model and how the maintenance energy parameters select a small subset of the possible solution space.

Submaximal P/O numbers optimize energy conserving potential of finite membrane surface area.

E. coli has a branched ETC that can convey electrons using multiple enzymatic routes to acceptors including O_2 (44-46). As mentioned previously, *E. coli* ETC components can theoretically operate at P/O numbers between ~0 and 2 (44-46). P/O number can significantly influence maintenance energy calculations (30). *E. coli* cultures typically operate at submaximal P/O numbers in the range of 1 to 1.5 even at O_2 sufficiency (reviewed in supplementary data S4). Few theories have been proposed to explain this phenotype (54, 55). We hypothesize *sMSA* theory is relevant to the experimentally observed submaximal P/O numbers. Here, we use a combination of predictions and interpretations to further develop the biological significance of *sMSA* theory.

Data from four proteomics studies (32-35, 51) and enzyme k_{cat} numbers (supplementary data S4) were used to calculate experimental P/O numbers as a function of growth rate (Fig. 4A, 4B, supplementary data S7). Experimental values were compared to *in silico* P/O numbers calculated using *sMSAc-FBA* (Fig. 4A, 4B). *In silico* P/O numbers for both strains at μ_{max} were similar to the experimental values (1.0-1.03 ± 0.06 compared to 1.03-1.07 ± 0.04). We propose the submaximal

P/O numbers represent an optimized use of the finite *sMSA* that enables optimal substrate fluxes for growth and cellular energy generation and now present additional evidence for this conclusion.

Membrane surface area and membrane protein crowding predict initiation of overflow metabolism.

The intersection of metabolism, membrane protein crowding, and cellular geometry predicts the onset of acetate overflow based on optimizing the use of finite *sMSA*. The presented analysis is designed to further support the *sMSA* theory by highlighting its potential significance as a mediator of major phenotypic strategies. *sMSAc-FBA* was used to calculate growth rate dependent P/O numbers for strains MG1655 and NCM3722 using the experimental consensus fluxes, SA:V ratios, and growth rate dependent membrane protein crowding (Fig. 4A, 4B, supplementary data S7). Theoretical P/O number vs. growth rate curves were fit empirically (analyses of different empirical fittings are presented in supplementary data S7). The inflection points were determined from the second derivative with respect to growth rate. Inflection points identified the growth rate where available *sMSA* exceeded the minimal requirements to satisfy the consensus fluxes. Inflection points at $\mu = 0.5$ h^{-1} and 0.72 h^{-1} for MG1655 and NCM3722, respectively, correlate well with the experimental values for overflow initiation, 0.4 ± 0.1 h^{-1} and 0.75 ± 0.05 h^{-1} (supplementary data S1). At growth rates less than the inflection point, cells have sufficient *sMSA* to operate substrate transporters, ETC apparatus, and ATP synthase without requiring the overflow tradeoff (Fig. 4A, 4B). The P/O number inflection also coincides with the maximum experimental biomass yields, providing a basis for defining an optimal biomass phenotype (Fig. 4A, 4B, supplementary data S1).

Finite surface area and membrane protein crowding precludes the constitutive expression of all substrate transporters. Cells dynamically regulate the membrane proteome as a function of environment (32, 56). At specific growth rates less than the P/O number inflection, predicted P/O numbers increased, deviating from experimental values (Fig. 4A, 4B). The consensus fluxes quantify energy-limited chemostat cultures (glucose-limited). While the cells have the theoretical capacity to express higher efficiency ETC components at low growth rates, the cultures instead use the surplus membrane capacity to increase expression of alternative, high

affinity transporters. These alternative substrates are not present in the medium, yet their transporters are likely expressed at the cost of *sMSA* as a hedge strategy to acquire new energy sources (32, 56). Areal density of ABC transporters for alternative sugars increased substantially at growth rates less than the P/O number inflection (Fig. 4C)(see also supplementary data S8) (32-35). Differences in the theoretically possible and actual P/O numbers permit calculation of an opportunity cost paid for the promise of an alternative energy source. At a growth rate of 0.1 h^{-1}, MG1655 forgoes the capacity to produce an additional ~4.5 mmol ATP (g cdw h^{-1}) from the same glucose fluxes for the promise of an untapped energy source (Fig. 4D, supplementary data S6).

Combining the *E. coli* consensus fluxes and the *sMSA* constraint creates a nonlinear relationship between growth rate and the potential for maintenance energy fluxes (Fig. 4D). *E. coli* K-12 demonstrates a linear relationship between specific glucose consumption rates and growth rate even while shifting to an overflow metabolism (supplementary data S1). The nonlinear relationship predicted between q_{ATP} and growth rate is not surprising given the changes in phenotype with overflow metabolism and the growth rate dependent changes in SA:V and membrane protein crowding. We propose a maintenance energy definition that includes the cellular SA:V ratio and the membrane protein capacity (Fig. 4D, supplementary data S6). Defining a P/O number without defining the cell geometry or membrane protein capacity is not sufficient to identify a unique q_{ATP} (supplementary data S9).

sMSAc-FBA predictions are robust to perturbations in enzyme parameters.

Enzyme k_{cat} numbers varied substantially between published studies (supplementary data S4) (57). The robustness of the presented results was analyzed by perturbing the MRE_i parameter which accounted for enzyme properties including k_{cat}, degree of saturation, and surface area requirements (Box 1). The MRE_i parameter was randomly and independently perturbed within a range of ±0% to ±95% of the base value. Analysis considered *E. coli* MG1655 fluxes from batch growth on glucose salts medium (47), not the consensus data. Experimental fluxes were also perturbed randomly and independently within a standard deviation of the measured flux. Each scenario was analyzed with 100 simulations (Fig. 4E).

Capacity for q_{ATP} was robust to perturbations. Calculated q_{ATP} varied ≤ 10% for random perturbations of all MRE_i parameters up to ± 95% of the base value (Fig. 4E). *sMSA* theory is robust to the presented assumptions, likely based on the compensating nature of the redundant and parallel structures in metabolic networks (58). Given the robustness of the results, it is probable the parameters can be used in other systems with reasonable accuracy.

Additional robustness analyses examined the effect of perturbing one MRE_i parameter at a time to assess the effect on q_{ATP}. These analyses considered five membrane-associated enzymes which accounted for ~87% of the utilized surface area during batch growth (ATP synthase, glucose transporter (PtsG), ammonium transporter (AmtB), NADH dehydrogenase (Nuo), and cytochrome oxidase (Cyo) (supplementary data S9)). Each enzyme MRE_i was perturbed up or down 35% to increase or decrease the *sMSA* requirements, respectively. ATP synthase accounted for 45% of the surface area utilized by central metabolism enzymes and was most sensitive to MRE_i perturbations. Increasing the ATP synthase MRE_i parameter 35% while holding all other MRE_i parameters constant, decreased the potential for q_{ATP} fluxes by ~26%, whereas decreasing the MRE_i parameter 35% increased the potential for q_{ATP} fluxes by ~8%. Perturbing the MRE_i parameter for the other four enzymes individually, either up or down 35%, did not change the capacity for q_{ATP} fluxes more than ±5% (supplementary data S9).

DISCUSSION

The rules of cell design remain elusive despite decades of research. For example, the design rules that link cell geometry and membrane protein capacity with central metabolism have not been defined previously on a quantitative basis. Bacteria display exquisite control over cellular geometry and membrane properties which are proposed here to be biologically significant as these properties are major mediators of possible phenotypes (1, 3, 15). *E. coli* cell length and diameter increase with growth rate, reducing the SA:V ratio (Fig. 1D)(2, 16). However, as shown, the reduction in SA:V is offset by an increase in the capacity of the membrane to host proteins (Fig 1E, supplementary data S3). This property is hypothesized to be critical to competitive phenotypes; if membrane protein content were not essential to fitness, this capacity would likely remain independent of, or decrease with increasing growth rate (15). This membrane property was integrated into a new, molecular-level theory which combines metabolism, cell

geometry, and membrane protein crowding. *sMSA* theory predicted maximum growth rates, acetate overflow metabolism, electron transport chain efficiency, and maintenance energy fluxes without accounting for cytosolic proteome investment or a cytosolic molecular crowding (6, 14, 42). Much effort has documented the importance of volume-associated proteome investment and molecular crowding on phenotype (6, 14, 16). Volume-based processes are certainly central to phenotype (5, 6, 9), but we hypothesize that both surface area- and volume-associated processes are critical and concurrently influence phenotype (1). Surface area requirements for membrane enzymes do not correlate well with volume proxies like enzyme mass (Fig. 1C, supplementary data S1); membrane surface area requirements are proposed to be distinct biophysical properties which impose unique constraints (14, 42). Tight regulation of cellular geometry and therefore SA:V ratio would arguably be futile if surface area-associated substrate transport and volume-associated protein synthesis were not also tightly regulated. If a single geometric aspect were limiting, evolution could select for changes in either cell length or cell diameter to overcome the surface area or volume limitation, respectively. A potential strategy to overcome the limits of geometry is to evolve membranes with higher protein hosting capacities. Mitochondria are eukaryotic organelles specialized in ATP generation (16). While the geometries of mitochondria are similar to bacteria, their membranes can host proteins at areal densities approximately two-fold higher than an *E. coli* membrane (24 ± 5.6 % vs. 40-50%) (4, 16, 59)(reviewed in supplementary data S4). The role of cell geometries in different bacterial species, *e.g.* bacilli vs. cocci, or the membrane protein capacities of Gram-negative vs. Gram-positive membrane architectures are open questions that can be addressed using *sMSAc-FBA* (16).

The presented analyses only consider growth on minimal salts medium which limits maximum growth rates. Growth on rich media enables faster growth (2). Extending the current theory to rich media would require measuring the uptake of numerous substrates. Surface area constraints likely prohibit the simultaneous use of all available substrates (32, 56). Instead, an animalcule will sequentially catabolize preferred substrates based on catabolite repression schemes (8, 60). Theories governing preferred substrates for co-metabolism are still being developed (61, 62), and it will be interesting to assess the extent to which surface area constraints inform these theories.

We note that ~7 - 11% of the MG1655 inner membrane surface area is occupied by central metabolism enzymes based on four proteomics studies and assume NCM3722 has a similar capacity (Fig. 2D, 2E, supplementary data S3)(32-35). Central metabolism accounts for approximately one third of membrane-associated proteins; thus, these proteomics-based values are consistent with literature review estimates of 24 ± 5.6% of the inner membrane surface area being occupied by total protein (19, 32, 63, 64)(Fig. 1F)(supplementary material S4). The experimentally measured protein crowding is higher than *sMSAc-FBA* predictions (≤7% for both MG1655 and NCM3722, Fig. 3A, 3B, 3C, 3D). *sMSAc-FBA* predictions assume an idealized scenario including 1) all enzymes are saturated for substrate including ETC components and O_2 cytochromes, 2) cells produce optimal levels of each enzyme with no overproduction, 3) all protein complexes are perfectly assembled with exact subunit ratios, 4) all enzymes are flawlessly embedded in the membrane, 5) no allosteric regulation of enzyme activity occurs based on membrane curvature (3, 38), and 6) the enzyme k_{cat} numbers increase two-fold for every 10°C increase in temperature (Q_{10} = 2) (49). The curated proteomics data provide a basis for estimating deviations from this idealized scenario. Using a protein offset (Δ, Box 1) of ~3% quantitatively aligns the *sMSAc-FBA* simulations with the experimental proteomics data without altering any presented interpretations or conclusions. The presented study implicitly includes a perturbation analysis of Δ. For example, perturbing Δ for batch growth can be analyzed by examining results for f_{MSA} values larger or smaller than 0.07 (Fig. 3B, 3C).

A quantitative, systems-level understanding of phenotype is still elusive for model organisms including *E. coli*. Biological optimality principles are of keen interest as they illuminate the (pseudo) rules of life and are readily implemented using systems biology methods. The presented study demonstrates that the commonly observed submaximal P/O numbers are likely not suboptimal strategies. Instead, they represent an optimal tradeoff between surface area requirements for substrate transporters and ETC components. Other commonly observed submaximal, yet likely optimized, phenotypes include overflow metabolisms occurring at both high (6, 65) and low growth rates (50), use of the Entner-Doudoroff (ED) glycolysis rather than Embden-Meyer-Parnas (EMP) glycolysis (58), and use of a partial citric acid cycle under nutrient limitation (66). Numerous *in silico* studies have replicated the overflow and ED vs. EMP glycolysis phenotypes using a variety of criteria (6, 12, 13, 18, 42, 58, 67). It is probable that many of these

criteria are concurrently relevant to cell geometry (14, 42). All the criteria explicitly or implicitly define tradeoffs between resource sufficiency and resource scarcity and can be viewed through a lens of cellular economics. The relative value of a limiting resource is higher than for resources present in excess. Therefore, fitness is increased when use of the limiting resource is optimized (39, 58). The presented *sMSA* theory quantifies a metabolic resource defined by cell geometry and membrane protein crowding and elucidates cell design rules which impact selection of competitive phenotypes.

MATERIALS AND METHODS

Enzyme parameters

Computational parameters including enzyme k_{cat} (*a.k.a*. turnover) numbers, enzyme surface area, cellular water content, and cell density for *E. coli* were estimated using literature reviews. Parameter values and references are provided in supplementary data S4. k_{cat} numbers were temperature corrected to 37°C using a Q_{10} number of 2 (49). Analyses assumed all membrane-associated enzymes were saturated ($\varepsilon_i = 1$) during batch growth, as typical concentrations of medium components during batch growth are ~2 orders of magnitude greater than the average K_M values of *E. coli* enzymes, ~0.1 mM. During chemostat growth, the saturation parameter (ε_i) for glucose transport (PtsG) was adjusted based on the dilution rate.

Metabolic model

The *E. coli* metabolic model was based on published models (39, 40). Every model reaction was balanced for atoms and electrons. The biomass reaction was constructed using theory developed by Neidhardt *et al.,* as described previously (39, 68). The biomass macromolecular composition on a dry mass basis was 78% protein, 10% RNA, 6% DNA, and 6% lipid and polysaccharide (68), the elemental composition was $C_1H_{1.96}O_{0.52}N_{0.28}P_{0.03}$, and the degree of reduction was 4.24 oxidizable electrons per carbon mole of biomass (NH_3 basis). The biomass ATP demand for monomer synthesis and macromolecule polymerization was 36.9 mol ATP per kg dry biomass (39). The biomass reaction did not account for any maintenance energy requirements as those fluxes were calculated separately. The maximum potential for maintenance energy fluxes (q_{ATP}) was calculated from experimental data by constraining

exchange fluxes to consensus data and maximizing flux through the ATP hydrolysis reaction (reaction identifier: *ATPm*). Thus, the calculated maintenance energy fluxes account for both growth- and nongrowth-associated ATP. The same *E. coli* model was used for both *E. coli* strain MG1655 and NCM3722 simulations while applying strain appropriate SA:V ratios and consensus data (supplementary data S1, S2, S5). Unless otherwise noted, enzyme saturation (ε_i), enzyme over capacity factor (ω_i), and allosteric regulation based on membrane properties (σ_i) were set to 1.

The metabolic model is provided in supplementary data S5. All metabolic simulations used COBRA Toolbox (41) with Gurobi as the solver.

Data Availability: All MATLAB code, models, and supplementary data can be found at github.com/rosspcarlson/carlson2024_sMSAcFBA.

Author Contributions: R.P.C. conceived project, performed theory synthesis, developed code, analyzed data, wrote document, A.E.B. conceived project, developed code, analyzed data, read and edited document, M.G.B., W.R.H., and R.M. analyzed data, developed interpretation, read and edited document, T.G. developed theoretical concepts, analyzed data, read and edited document.

Competing Interest Statement: The authors have no competing interests.

ACKNOWLEDGEMENTS

R.P.C. would like to acknowledge Terrance Hwa, Adrienne Arnold, and Campbell Putnam for insightful discussions.

R.P.C. was supported in part by NIH award U01EB019416 and AFOSR award FA9550-23-1-0589. A.E.B. was supported in part by USDA National Institute of Food and Agriculture grant no. 2021-67013-34537. T.G. was supported in part by NSF award DMS 1361240

Summary of Supplementary Materials

Files available at: github.com/rosspcarlson/carlson2024_sMSAcFBA

Supplementary Data S1: Literature survey of extracellular flux data for *E. coli* K-12 strains MG1655 and NCM3722 including consensus data sets. Comparison of enzyme volume and membrane surface area requirements for 22 central metabolism enzymes.

Supplementary Data S2: Literature values for cell geometry as a function of growth rate for *E. coli* K-12 strains MG1655 and NCM3722.

Supplementary Data S3: *E. coli* K 12 membrane proteomics data from three independent studies collated by Belliveau et al. 2021.

Supplementary Data S4: Literature review of enzyme and biophysical parameters used in study.

Supplementary Data S5: *In silico*, stoichiometric model and associated data files.

Supplementary Data S6: Summary of simulation output for *E. coli* K 12 strains MG1655 and NCM3722.

Supplementary Data S7: Summary of experimental and in silico P/O number calculations.

Supplementary Data S8: Experimental proteomics data quantifying areal density of high affinity ABC transporters.

Supplementary Data S9: Analysis of maintenance energy fluxes as a function of growth rate, membrane surface area, protein hosting capacity, and P/O number. Perturbation analysis of enzyme *MRE* parameters.

Supplementary Data S10: Summary of MATLAB code used for each figure in study.

References

1. Harris LK & Theriot JA (2018) Surface Area to Volume Ratio: A Natural Variable for Bacterial Morphogenesis. *Trends Microbiol* 26(10):815-832.
2. Si F*, et al.* (2017) Invariance of Initiation Mass and Predictability of Cell Size in Escherichia coli. *Curr Biol* 27(9):1278-1287.
3. Shi H, Bratton BP, Gitai Z, & Huang KC (2018) How to Build a Bacterial Cell: MreB as the Foreman of E. coli Construction. *Cell* 172(6):1294-1305.
4. Lowe M, Kalacheva M, Boersma AJ, & Kedrov A (2020) The more the merrier: effects of macromolecular crowding on the structure and dynamics of biological membranes. *Febs J* 287(23):5039-5067.
5. Beg QK*, et al.* (2007) Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences of the United States of America* 104(31):12663-12668.
6. Basan M*, et al.* (2015) Overflow metabolism in Escherichia coli results from efficient proteome allocation. *Nature* 528(7580):99-104.
7. Vazquez A & Oltvai ZN (2016) Macromolecular crowding explains overflow metabolism in cells. *Sci Rep* 6:31007.
8. van den Berg J, Boersma AJ, & Poolman B (2017) Microorganisms maintain crowding homeostasis. *Nature reviews. Microbiology* 15(5):309-318.
9. Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci* 26(10):597-604.
10. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, & Palsson BO (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular systems biology* 9:693.
11. Carlson RP (2007) Metabolic systems cost-benefit analysis for interpreting network structure and regulation. *Bioinformatics* 23(10):1258-1264.
12. Szenk M, Dill KA, & de Graff AMR (2017) Why Do Fast-Growing Bacteria Enter Overflow Metabolism? Testing the Membrane Real Estate Hypothesis. *Cell Syst* 5(2):95-104.
13. Zhuang K, Vemuri GN, & Mahadevan R (2011) Economics of membrane occupancy and respiro-fermentation. *Molecular systems biology* 7:500.
14. Goelzer A & Fromion V (2011) Bacterial growth rate reflects a bottleneck in resource allocation. *Biochim Biophys Acta* 1810(10):978-988.
15. Guigas G & Weiss M (2016) Effects of protein crowding on membrane systems. *Biochim Biophys Acta* 1858(10):2441-2450.
16. Lynch M & Marinov GK (2017) Membranes, energetics, and evolution across the prokaryote-eukaryote divide. *Elife* 6.
17. Phillips R & Milo R (2009) A feeling for the numbers in biology. *Proceedings of the National Academy of Sciences of the United States of America* 106(51):21465-21471.
18. Bruggeman FJ, Planque R, Molenaar D, & Teusink B (2020) Searching for principles of microbial physiology. *FEMS Microbiol Rev* 44(6):821-844.
19. Papanastasiou M*, et al.* (2013) The Escherichia coli peripheral inner membrane proteome. *Mol Cell Proteomics* 12(3):599-610.
20. Sueki A, Stein F, Savitski MM, Selkrig J, & Typas A (2020) Systematic Localization of Escherichia coli Membrane Proteins. *mSystems* 5(2).

21. Budin I*, et al.* (2018) Viscous control of cellular respiration by membrane lipid composition. *Science* 362(6419):1186-1189.
22. Sun J, Rutherford ST, Silhavy TJ, & Huang KC (2022) Physical properties of the bacterial outer membrane. *Nature reviews. Microbiology* 20(4):236-248.
23. Dupuy AD & Engelman DM (2008) Protein area occupancy at the center of the red blood cell membrane. *Proceedings of the National Academy of Sciences of the United States of America* 105(8):2848-2852.
24. Lyons E, Freeling M, Kustu S, & Inwood W (2011) Using genomic sequencing for classical genetics in E. coli K12. *Plos One* 6(2):e16717.
25. Hellingwerf K*, et al.* (1982) Energetics of microbial growth: an analysis of the relationship between growth and its mechanistic basis by mosaic non-equilibrium thermodynamics. *FEMS Microbiol. Lett* 15(1):7-17.
26. Pirt SJ (1965) The maintenance energy of bacteria in growing cultures. *Proc R Soc Lond B Biol Sci* 163(991):224-231.
27. van Bodegom P (2007) Microbial maintenance: a critical review on its quantification. *Microbial ecology* 53(4):513-523.
28. Carlson R & Srienc F (2004) Fundamental Escherichia coli biochemical pathways for biomass and energy production: creation of overall flux states. *Biotechnol Bioeng* 86(2):149-162.
29. Hoehler TM & Jorgensen BB (2013) Microbial life under extreme energy limitation. *Nature reviews. Microbiology* 11(2):83-94.
30. Taymaz-Nikerel H, Borujeni AE, Verheijen PJ, Heijnen JJ, & van Gulik WM (2010) Genome-derived minimal metabolic models for Escherichia coli MG1655 with estimated in vivo respiratory ATP stoichiometry. *Biotechnol Bioeng* 107(2):369-381.
31. Zeng AP, Ross A, & Deckwer WD (1990) A method to estimate the efficiency of oxidative phosphorylation and biomass yield from atp of a facultative anaerobe in continuous culture. *Biotechnol Bioeng* 36(9):965-969.
32. Belliveau NM*, et al.* (2021) Fundamental limits on the rate of bacterial growth and their influence on proteomic composition. *Cell Syst* 12(9):924-944 e922.
33. Schmidt A*, et al.* (2016) The quantitative and condition-dependent Escherichia coli proteome. *Nat Biotechnol* 34(1):104-110.
34. Valgepea K, Adamberg K, Seiman A, & Vilu R (2013) Escherichia coli achieves faster growth by increasing catalytic and translation rates of proteins. *Mol Biosyst* 9(9):2344-2358.
35. Peebo K*, et al.* (2015) Proteome reallocation in Escherichia coli with increasing specific growth rate. *Mol Biosyst* 11(4):1184-1193.
36. Volkmer B & Heinemann M (2011) Condition-dependent cell volume and concentration of Escherichia coli to facilitate data conversion for systems biology modeling. *Plos One* 6(7):e23126.
37. Ramachandran A, Anderson TH, Leal LG, & Israelachvili JN (2011) Adhesive interactions between vesicles in the strong adhesion limit. *Langmuir* 27(1):59-73.
38. Andersen OS & Koeppe RE, 2nd (2007) Bilayer thickness and membrane protein function: an energetic perspective. *Annu Rev Biophys Biomol Struct* 36:107-130.
39. Carlson RP (2009) Decomposition of complex microbial behaviors into resource-based stress responses. *Bioinformatics* 25(1):90-97.

40. Schuetz R, Zamboni N, Zampieri M, Heinemann M, & Sauer U (2012) Multidimensional optimality of microbial metabolism. *Science (New York, N.Y.)* 336(6081):601-604.
41. Heirendt L*, et al.* (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc* 14(3):639-702.
42. Mori M, Hwa T, Martin OC, De Martino A, & Marinari E (2016) Constrained Allocation Flux Balance Analysis. *PLoS computational biology* 12(6):e1004913.
43. Baldazzi V, Ropers D, Gouze JL, Gedeon T, & de Jong H (2023) Resource allocation accounts for the large variability of rate-yield phenotypes across bacterial strains. *Elife* 12.
44. Grauel A*, et al.* (2021) Structure of Escherichia coli cytochrome bd-II type oxidase with bound aurachin D. *Nat Commun* 12(1):6498.
45. Borisov VB*, et al.* (2011) Aerobic respiratory chain of Escherichia coli is not allowed to work in fully uncoupled mode. *Proceedings of the National Academy of Sciences of the United States of America* 108(42):17320-17324.
46. Anraku Y (1988) Bacterial electron transport chains. *Annual review of biochemistry* 57(1):101-132.
47. Beck AE*, et al.* (2022) Environment Constrains Fitness Advantages of Division of Labor in Microbial Consortia Engineered for Metabolite Push or Pull Interactions. *mSystems* 7(4):e0005122.
48. Mori M, Cheng C, Taylor BR, Okano H, & Hwa T (2023) Functional decomposition of metabolism allows a system-level quantification of fluxes and protein allocation towards specific metabolic functions. *Nature Communications* 14(1):4161.
49. Elias M, Wieczorek G, Rosenne S, & Tawfik DS (2014) The universality of enzymatic rate-temperature dependency. *Trends Biochem Sci* 39(1):1-7.
50. Folsom JP & Carlson RP (2015) Physiological, biomass elemental composition and proteomic analyses of Escherichia coli ammonium-limited chemostat growth, and comparison with iron- and glucose-limited chemostat growth. *Microbiology* 161(8):1659-1670.
51. Mori M*, et al.* (2021) From coarse to fine: the absolute Escherichia coli proteome under diverse growth conditions. *Molecular systems biology* 17(5):e9536.
52. Beck A, Hunt K, Bernstein HC, & Carlson R (2016) Interpreting and designing microbial communities for bioprocess applications, from components to interactions to emergent properties. *Biotechnology for Biofuel Production and Optimization*, eds Eckert CA & Trinh CT (Elsevier, Amsterdam), pp 407-432.
53. Rabbers I & Bruggeman FJ (2022) Escherichia coli robustly expresses ATP synthase at growth rate-maximizing concentrations. *Febs J* 289(16):4925-4934.
54. Westerhoff HV, Hellingwerf KJ, & Van Dam K (1983) Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. *Proceedings of the National Academy of Sciences of the United States of America* 80(1):305-309.
55. Werner S, Diekert G, & Schuster S (2010) Revisiting the thermodynamic theory of optimal ATP stoichiometries by analysis of various ATP-producing metabolic pathways. *J Mol Evol* 71(5-6):346-355.
56. Lendenmann U & Egli T (1998) Kinetic models for the growth ofEscherichia coli with mixtures of sugars under carbon-limited conditions. *Biotechnol Bioeng* 59(1):99-107.

57. Chen Y & Nielsen J (2021) In vitro turnover numbers do not reflect in vivo activities of yeast enzymes. *Proceedings of the National Academy of Sciences of the United States of America* 118(32).
58. Carlson RP & Taffs RL (2010) Molecular-level tradeoffs and metabolic adaptation to simultaneous stressors. *Curr Opin Biotech* 21(5):670-676.
59. Schlame M (2021) Protein crowding in the inner mitochondrial membrane. *Biochim Biophys Acta Bioenerg* 1862(1):148305.
60. Park H, McGill SL, Arnold AD, & Carlson RP (2019) Pseudomonad reverse carbon catabolite repression, interspecies metabolite exchange, and consortial division of labor. *Cellular and molecular life sciences : CMLS*.
61. McGill SL*, et al.* (2021) Pseudomonas aeruginosa reverse diauxie is a multidimensional, optimized, resource utilization strategy. *Scientific Reports* 11(1).
62. Hermsen R, Okano H, You C, Werner N, & Hwa T (2015) A growth-rate composition formula for the growth of E. coli on co-utilized carbon substrates. *Molecular systems biology* 11(4):801.
63. Stachowiak JC*, et al.* (2012) Membrane bending by protein-protein crowding. *Nat Cell Biol* 14(9):944-949.
64. Duncan AL*, et al.* (2017) Protein crowding and lipid complexity influence the nanoscale dynamic organization of ion channels in cell membranes. *Sci Rep* 7(1):16647.
65. El-Mansi M (2023) Control of central metabolism's architecture in Escherichia coli: An overview. *Microbiol Res* 266:127224.
66. Fischer E & Sauer U (2003) A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *The Journal of Biological Chemistry* 278:46446-46451.
67. Carlson RP*, et al.* (2018) Competitive resource allocation to metabolic pathways contributes to overflow metabolisms and emergent properties in cross-feeding microbial consortia. *Biochemical Society Transactions* 46(2):269-284.
68. Carlson R & Srienc F (2004) Fundamental Escherichia coli biochemical pathways for biomass and energy production: identification of reactions. *Biotechnol Bioeng* 85(1):1-19.

Box and Figures

Box 1: The membrane surface area required for a phenotype can be tabulated by considering enzyme fluxes and enzyme properties. Literature reviews of enzyme parameters used in this study can be found in supplementary data S4. The surface area required for enzyme *i*, per g cdw, is calculated as:

$$SA_{Ei} = \left[\frac{v_i \gamma_i}{\tau_i \varepsilon_i}\right]\left[\frac{\omega_i}{\sigma_i}\right]\left[\frac{mol}{10^3\ mmol}\right][N_{Avo}]\left[\frac{h}{3600\ s}\right] \qquad \textit{eqn. 1}$$

SA_{Ei} = surface area occupied by enzyme *i*, units: nm^2 $(g\ cdw)^{-1}$
v_i = flux of reaction *i*, units: mmol $(g\ cdw\ h)^{-1}$
γ_i = membrane area occupied by single enzyme *i*, units: nm^2 $(enzyme)^{-1}$
τ_i = enzyme *i* turnover number (*a.k.a.* k_{cat}), units: molecules $(s*enzyme)^{-1}$
ε_i = enzyme *i* degree of saturation, unitless (0 - 1)
σ_i = membrane allosteric regulation , unitless (0 - 1)
ω_i = enzyme overcapacity factor , unitless (≥1)
N_{Avo} = Avogadro's number, units: molecules $mole^{-1}$

Total membrane surface area used by central metabolism enzymes is determined by summing the contributions from all active enzymes and applying biophysical parameters (reviewed in supplementary data S4). *Eqn. 2* quantifies membrane protein crowding as the fraction of membrane surface area (f_{MSA}) occupied by active central metabolism enzymes. Conversions include, from left to right, water content of live cells, density of live cells, growth rate dependent SA:V ratio of cells ($c_1*\mu + c_2$), and unit conversions. δ_i accounts for nonidealities in surface area usage including misfolded proteins or misassembled protein complexes:

$$f_{MSA} = \sum_i \left(SA_{Ei}\left[\frac{g\ cdw}{3.85\ g\ wet\ cell}\right]\left[\frac{1.1\ g\ wet\ cell}{cm^3\ wet\ cell}\right]\left[\frac{\mu m^3\ wet\ cell}{(c_1*\mu+c_2)\ \mu m^2}\right]\left[\frac{cm}{10^4\ \mu m}\right]\left[\frac{cm^2}{10^{14}\ nm^2}\right] + \delta_i\right) \qquad \textit{eqn. 2}$$

Eqn. 2 can be arranged to define the surface area resource MRE_i (membrane real estate for enzyme *i*) which is incorporated into the presented metabolic model. θ is a constant comprised of biophysical parameters and conversion factors (note: c_1 has units of h^{-1}, c_2 is unitless):

$$f_{MSA} = \sum_i \left(\left[\frac{v_i \gamma_i}{\tau_i \varepsilon_i}\right]\left[\frac{\omega_i}{\sigma_i}\right]\left[\frac{\theta}{c_1*\mu+c_2}\right] + \delta_i\right) = \sum_i \left(v_i MRE_i\left[\frac{\theta}{c_1*\mu+c_2}\right] + \delta_i\right), \Theta = 0.048\ \text{g cdw nm}^{-2} \qquad \textit{eqn. 3}$$

Summing δ_i over all the enzymes quantifies the f_{MSA} nonideality offset Δ.

$$(f_{MSA} - \Delta)\left[\frac{c_1*\mu+c_2}{\theta}\right] = \sum_i v_i MRE_i \qquad \textit{eqn. 4}$$

Which can be arranged as:

$$(f_{MSA} - \Delta)\left[\frac{c_2}{\theta}\right] = \sum_i v_i MRE_i - (f_{MSA} - \Delta)\left[\frac{c_1*\mu}{\theta}\right] \qquad \textit{eqn. 5}$$

Collecting constants results in the linear constraint accounting for finite surface area:

$$\varphi = \sum_i v_i MRE_i - \beta * \mu \qquad \textit{eqn. 6}$$

φ = surface area available for enzymes, constraint during simulations, nm^2 $(g\ cdw)^{-1}$

MRE_i = membrane area used by active enzyme *i*, $h*nm^2$ $mmol^{-1}$

β = growth rate dependent change in surface area due to change in SA:V, $h*nm^2$ $(g\ cdw)^{-1}$

The constraint was applied by including an appropriate MRE_i cost to membrane-associated reactions and incorporating the β parameter into the biomass synthesis reaction. Note, the SA:V slope, c_1, is negative making the contribution of $\beta*\mu$ positive in *eqn. 6*. See supplementary data S4 and S5 for parameter values and literature reviews.

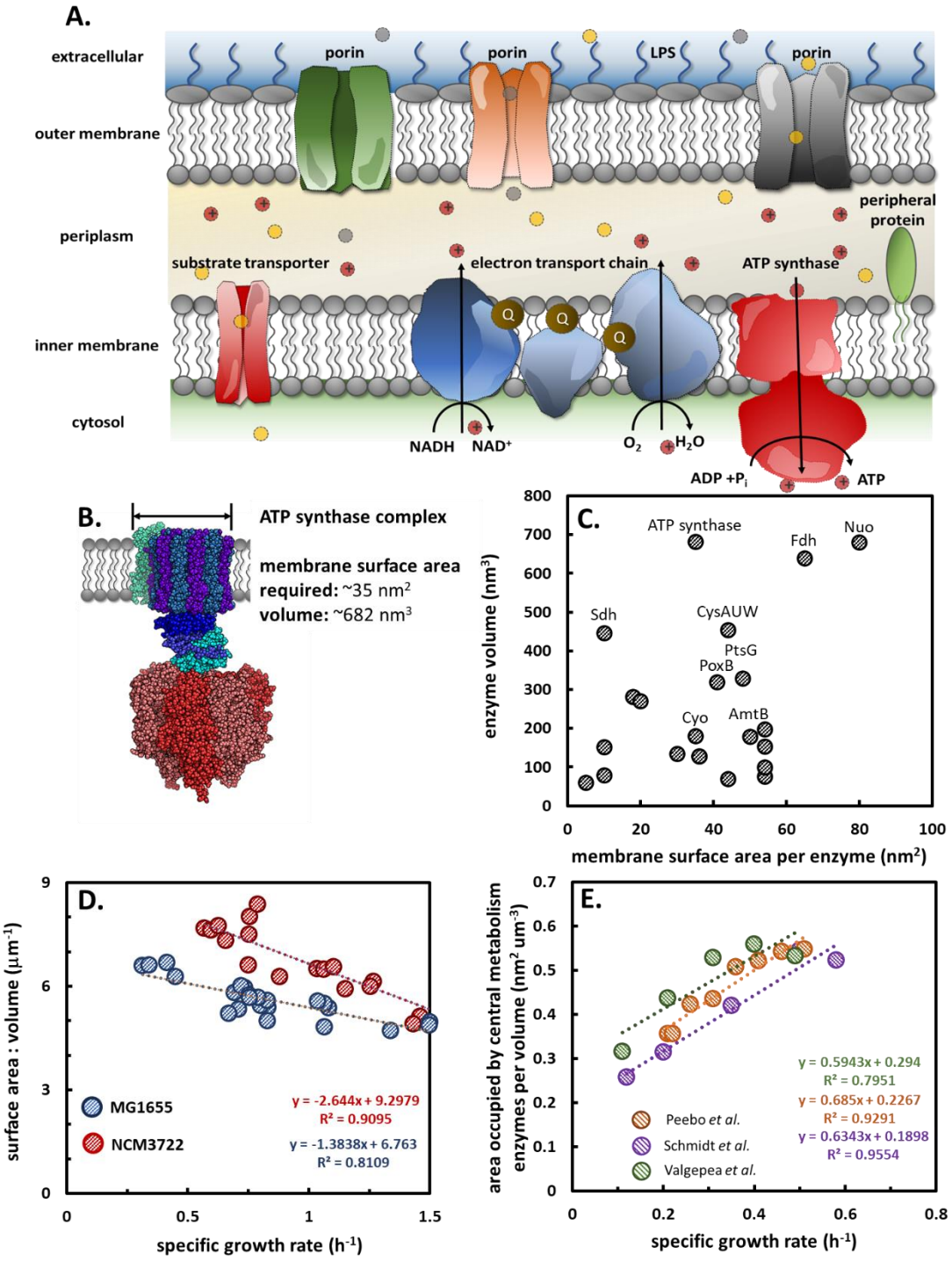

Figure 1. *E. coli* membrane and enzyme properties. A) The inner membrane enables critical cellular functions including substrate transport and oxidative phosphorylation. Membrane surface area is finite. **B)** The physical dimensions of membrane-associated enzymes influence both two- and three-dimensional molecular crowding. ATP synthase image modified from Wiki

commons. **C)** Membrane surface area requirements and enzyme volume for 22 central metabolism enzymes do not correlate well. Data available in supplementary data S1 and S4. **D)** Experimental surface area to volume (SA:V) ratios for *E. coli* K-12 strains MG1655 and NCM3722 as a function of specific growth rate. *E. coli* geometry data from (2). **E)** The capacity of the *E. coli* MG1655 inner membrane to host proteins, expressed here as nm^2 occupied by central metabolism enzymes per cell volume (μm^{-3}) for comparison with Figure 1D, increases with specific growth rate offsetting the decrease in SA:V. Data from *E. coli* cultures grown on glucose salts medium. Proteomics data from (32-35). Enzyme properties and calculations can be found in the supplementary data S3 and S4. Analysis was limited to cultures grown on glucose minimal data. Calculations were not possible for complex medium because the consumption rates of the numerous substrates were unknown.

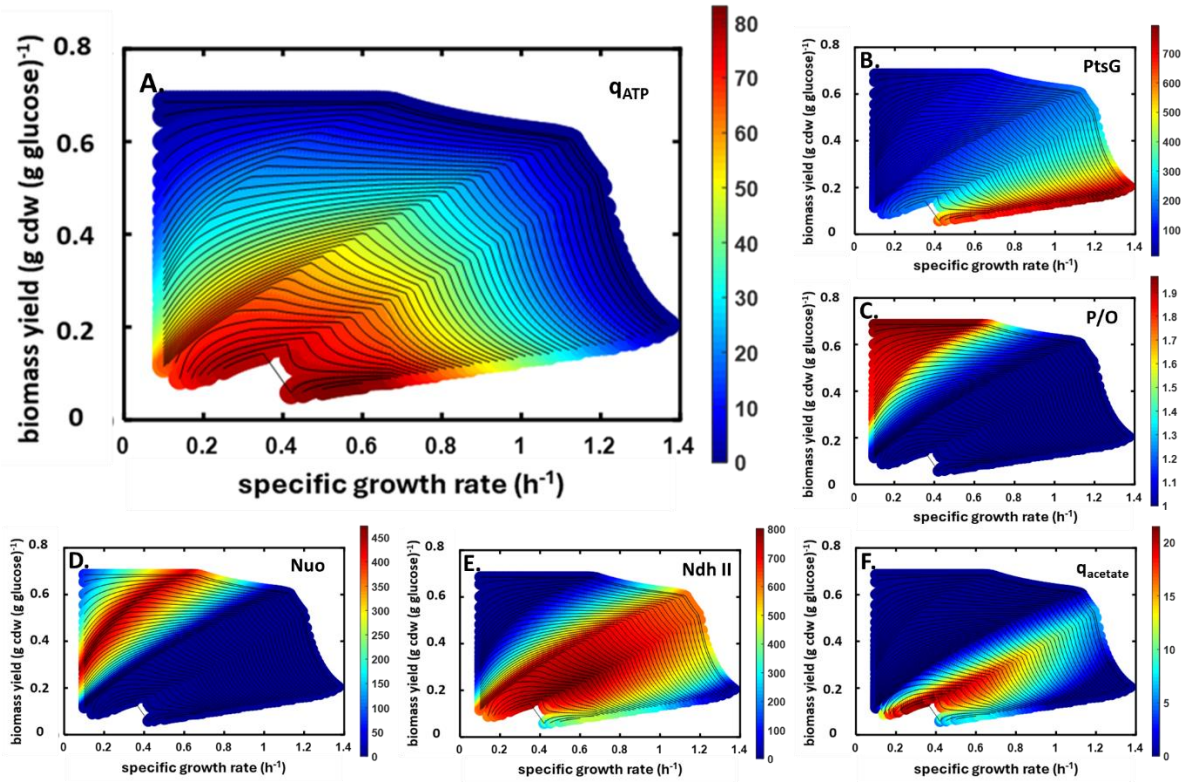

Figure 2. Predictive growth rate – biomass yield phenotype space for *E. coli* strain MG1655 constrained by membrane surface area and membrane protein crowding. **A)** Maintenance energy flux (mmol ATP (g cdw h)$^{-1}$) shown by the color gradient (protein crowding, f_{MSA} = 0.07). Black contour lines are maintenance energy isoclines. Maintenance energy flux accounts for ATP in excess of ATP used for macromolecule polymerization. **B)** Color gradient represents the P/O number which quantifies the efficiency of oxidative phosphorylation. **C)** Color gradient quantifies the areal density of glucose transporter PtsG (complexes μm^{-2}). **D)** Color gradient quantifies areal density of the electron transport chain component Nuo (complexes μm^{-2}). **E)** Color gradient quantifies areal density of the electron transport chain component Ndh II (complexes μm^{-2}). **F)** Color gradient quantifies overflow metabolism $q_{acetate}$ (mmol acetate (g cdw h)$^{-1}$). Color gradient scales, simulation data, and additional phenotypic representations can be found in supplementary data S6.

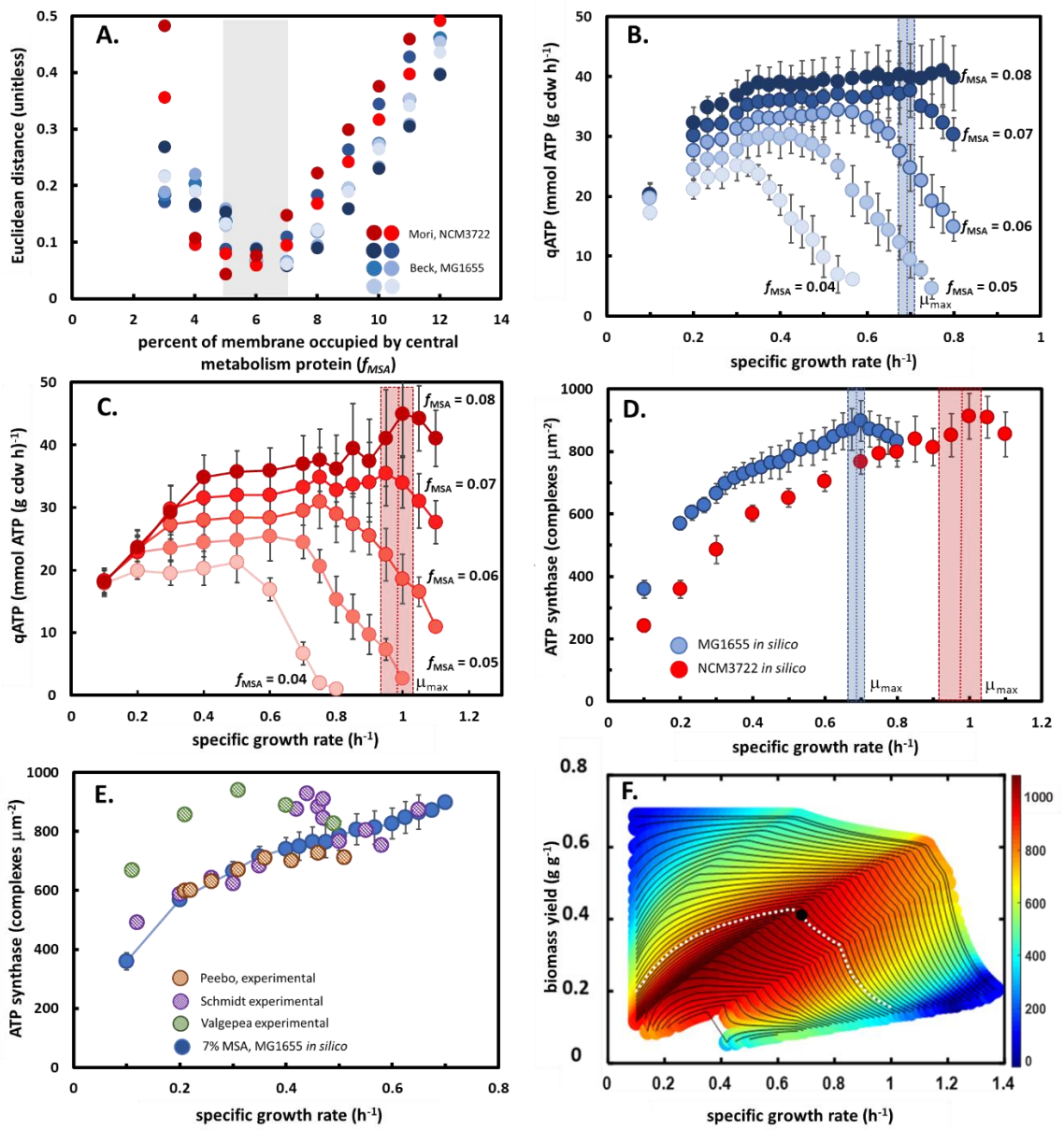

Figure 3. Maximum specific growth rate is constrained by membrane surface area and protein crowding. A) Experimental fluxes from *E. coli* K-12 strains MG1655 (6 data sets, blue points (47)) and NCM3722 (2 data sets, red points (48)) correlate best to a narrow range of occupied membrane area (5-7%) calculated using simulations enforcing a finite membrane surface area and membrane protein crowding. **B)** The maximum specific growth rate for *E. coli* strain MG1655 and **C)** *E. coli* strain NCM3722 on glucose salts medium correlates with a Pareto tradeoff between growth rate and maintenance energy flux potential. Shaded area quantifies experimental maximum growth rates and standard deviations. A series of simulations were

performed enforcing different membrane protein crowding (f_{MSA} = 0.04-0.08, or 4-8% of occupied membrane area). Error bars represent standard deviation of 100 simulations; see text for details. **D)** *E. coli* K-12 strains MG1655 and NCM3722 maximize the areal density of ATP synthase (complexes μm^{-2}) near the maximum specific growth rate. Error bars represent standard deviation of 100 simulations; see text for details. **E)** Predicted areal density of ATP synthase (complexes μm^{-2}) for strain MG1655 compared to experimental data from three proteomics studies (32-35). **F)** *E. coli* MG1655 rate-yield space with predicted ATP synthase areal density (complexes μm^{-2}) quantified by the color gradient. Black lines are maintenance energy isoclines (see Fig. 2A). Black point denotes experimental *E. coli* strain MG1655 maximum specific growth rate and biomass yield during batch growth on glucose salts medium.

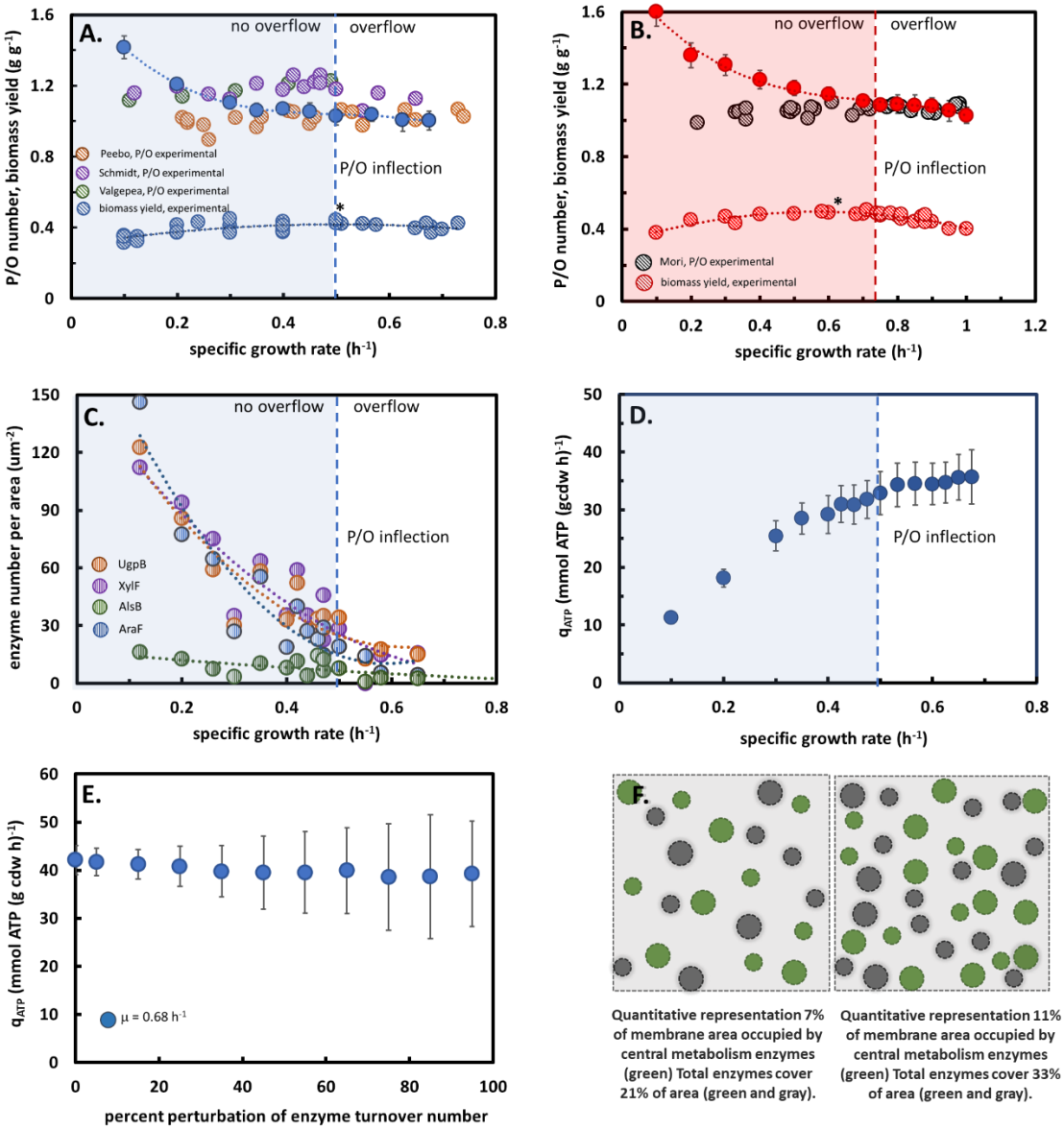

Figure 4. Membrane surface area and protein hosting capacity predicts the onset of acetate overflow metabolism. Predicted *E. coli* MG1655 P/O numbers (solid blue points) (**A**) and *E. coli* NCM3722 P/O numbers (solid red points) (**B**) plotted with experimentally determined P/O numbers as a function of growth rate. Predicted P/O numbers have an inflection point denoting the change in scarcity of the membrane surface area. Overflow metabolism occurs at severe surface area scarcity, right of inflection point. The P/O number inflection corresponds with higher experimental biomass yields on glucose. Biomass yield maximum highlighted with (*). Error bars represent standard deviation of 100 simulations using strain-specific consensus

fluxes. Experimental proteomics data for MG1655 from (32-35). Experimental proteomics data for NCM3722 from (44). **C)** *E. coli* MG1655 experimental areal density of high affinity ABC substrate transporters (complexes μm^{-2}) as a function of growth rate. Data from (33). See also supplementary data S8. **D)** *E. coli* MG1655 potential for maintenance energy generation (q_{ATP}) as a function of growth rate considering glucose-limited consensus fluxes. The simulation data applied P/O = 1.04 for results to the left of the P/O number inflection point to reflect experimental proteome trends (supplementary data S6). **E)** Perturbation analysis of enzyme parameters and their effect on the potential for maintenance energy (q_{ATP}) production. Experimental fluxes from batch growth of *E. coli* MG1655 on glucose salts media with f_{MSA} = 0.07. Enzyme parameters were independently perturbed between ± 0 and 95% (abscissa) while experimental fluxes were simultaneously and independently perturbed between ± one standard deviation. Each point is the mean of 100 simulations, and error bars represent one standard deviation. **F)** Quantitative representation of surface area fractions relevant to *E. coli* K-12 MG1655 at low (0.1 h^{-1}, 21% of area, left) and high (0.7 h^{-1}, 33% of area, right) specific growth rates. Areas are based on analyses in main text. Green circles represent central metabolism enzymes while gray circles represent other membrane-associated proteins.