Supplementary Text, Figures and Tables

A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information

Adam M. Feist[1], Christopher S. Henry[2], Jennifer L. Reed[1], Markus Krummenacker[3], Andrew R. Joyce[1], Peter D. Karp[3], Linda Broadbelt[2], Vassily Hatzimanikatis[4], Bernhard Ø. Palsson[1]

1 Dept. of Bioengineering, University of California San Diego, 9500 Gilman Drive, Mail Code 0412, La Jolla, CA, 92093, USA
2 Dept. of Chemical and Biological Engineering, McCormick School of Engineering and Applied Sciences, 2145 Sheridan Rd, Northwestern University, Evanston, IL, 60208, USA
3 SRI International, 333 Ravenswood, Menlo Park, CA 94025, USA
4 Institute of Sciences and Chemical Engineering, Ecole polytechnique fédérale de Lausanne, EPFL SB ISIC LCSB, BCH 3110 (Bât. BCH), CH-1015 Lausanne, Switzerland

Table of Content

Supplementary Text

Specific advancements in *i*AF1260 content

In addition to changes in reaction content between *i*AF1260 and *i*JR904, 22 metabolites were also removed from the previous reconstruction (18 of these were totally removed and the electrical charge of 4 were altered). Common examples of proteins included as metabolites in *i*AF1260 were acyl carrier protein (ACP, an important carrier in fatty acid biosynthesis) and thioredoxin (an electron carrier). tRNA synthase reactions were also included in this reconstruction for later integration with additional protein synthesis pathways (I Thiele and B Palsson, personal communication).

Gap-filling analysis

Through FBA simulations utilizing the BOF_{WT}, we performed a gap-filling procedure on *i*AF1260 to determine if the reaction network could generate all of the BOF_{WT} metabolites from common media substrates or alternatively, if reactions were missed or needed to fulfill the biomass requirements placed on the system. During this procedure, we found that one of the common cofactors in *E. coli*, biotin, could not be synthesized from common media substrates. Therefore, it was removed from the BOF_{WT}. During the gap-filling process, one common type of reaction that was found to be missing was the anaerobic complement of reactions in certain cofactor biosynthesis pathways. These reactions were often added with no gene-association, but are known to exist (Alexander and Young, 1978; Beale, 1996; Meganathan, 1996). Additionally, it was necessary to add five 'sink' reactions to the network so that certain byproducts could leave the system without active transport. This was necessary since there are no known pathways for excretion or degradation of these byproducts. For example, thiamine diphosphate could not be synthesized because there is no degradation pathway or known transport mechanism for the byproduct, 4-Hydroxy-benzyl alcohol (Leonardi and Roach, 2004). Therefore a 'sink' reaction was added for 4-Hydroxy-benzyl alcohol. Similarly, under anaerobic conditions, there is no degradation pathway or transporter for 5'-deoxyadenosine, a byproduct in heme biosynthesis (Choi-Rhee and Cronan, 2005). It was also necessary to add a methionine salvage pathway to the metabolic network based on the pathway present in *Bacillus subtilis* in order for the recycling generation of spermidine (Sekowska and Danchin, 2002). Some of the reactions in this pathway are catalyzed by known ORFs in the current annotation (Choi-Rhee and Cronan, 2005; Riley *et al*, 2006), whereas others are added with no gene-association. After the gap-filling procedure, all of the metabolites in the BOF_{WT} could be made aerobically or anaerobically from the defined *in silico* minimal media (see Supplementary Table 2) supplemented with D-glucose, D-ribose or glycerol. This gap-filling procedure was similar to that of Green and Karp (Green and Karp, 2004).

Common biomass metabolites that could not be synthesized from minimal media

In *i*AF1260, there are some metabolites that cannot be synthesized from common minimal media substrates. *E. coli* is auxotrophic for a cobalamin precursor since it lacks the full biosynthetic pathway for vitamin B_{12} (O'Toole *et al*, 1996). To fulfill this need, cob(I)alamin (vitamin B12s) (cbl1) can be added to the minimal media. Alternatively, cobinamide (cbi) could also theoretically serve as a precursor except there is no biosynthetic pathway for the adenosylcobalamin precursor, 5,6-dimethylbenzimidazole (dmbzid), characterized in *E. coli*. Biotin could not be synthesized because there is no known biosynthetic pathway for its precursor, pimeloyl-coA, characterized in *E. coli*.

Additional issues that can affect computational simulations

Another parameter that has been found to have potential for significantly affecting the solutions produced using metabolic modeling with FBA (Varma and Palsson, 1995) is the stoichiometry of the proton translocating enzymes of the ETS. For the most part, the stoichiometric rules that govern the conversion of metabolites in *E. coli* are hard constraints (i.e. only 2 units of pyruvate can be generated from 1 unit of glucose), but this is not the case for the ETS components which can couple metabolic conversion to proton translocation across the membrane. The *i*AF1260 reconstruction includes the efficiencies of the ETS based on characterization studies and some debate exists over the actual values (Puustinen *et al*, 1991; Weidner *et al*, 1993; Gennis and Stewart, 1996). Additionally, the current reconstruction incorporates the same proton translocation efficiency for all of three different quinones, regardless of the potential energy difference between the quinone/quinole couple (Tran *et al*, 1997).

Additional remarks on the gene essentiality analysis in *i*AF1260

A categorization of the computational essentiality predictions by COG functional class reveals the extent of agreement for each area of metabolism and under which medium conditions essentiality for a particular ORF occurs. Figure 5 shows the ORFs that were predicted to be essential under rich medium conditions, under both glucose and glycerol minimal medium conditions and under glucose minimal medium conditions experimentally by Baba et al and Joyce et al (Baba *et al*, 2006; Joyce *et al*, 2006) on the background of the computationally predicted essential ORFs under aerobic glucose conditions. Whereas most essential ORFs from the amino acid (E) and nucleotide (F) transport and metabolism classes are essential under minimal media conditions, those associated with cell wall (M) and lipid (I) metabolism are essential under rich media conditions (note that ORFs that are essential under rich medium conditions are also essential under minimal medium conditions). This suggests that most amino acids can be transported into the cell under rich medium conditions and even if present in the medium, cell wall and lipid molecules cannot be transported or need additional modification once in the cell. Coenzyme (H) metabolism and transport ORF essentiality is mostly at the level of rich media conditions and this can be explained by realizing that some cofactors cannot be transported into the cell and have to be synthesized *de novo* even under rich media conditions (e.g., riboflavin, coenzyme A, hemes).

*i*AF1260 information contained in the SBML file format

- reactions (format: 'R_<*reaction abbreviation* >')
  - reaction name, reversibility, reaction stoichiometry, gene-protein-reaction (GPR) association, subsystem, E.C. number
- metabolites (format: 'M_<*metabolite abbreviation*>_<*compartment abbreviation*>')
  - metabolite name, compartment, charge, formula (appended to the end of the name, <*metabolite name*>_FORMULA)
- a flux distribution associated with a steady-state modeling simulation
  - lower bound, upper bound, objective coefficient, flux value, reduced cost

SBML File Properties	
file name	*E. coli* _iAF1260_flux1.xml
organism	E. coli K-12 MG1655
model	iAF1260
Biomass Objective Function (BOF)	Ec_biomass_iAF1260_core_59p81M (E. coli biomass objective function (iAF1260) - core - with 59.81 GAM estimate)
flux balance analysis objective	maximize BOF
Growth Associated Maintenance (GAM)	59.81 mmol ATP gDW-1
Non-Growth Associated Maintenance (NGAM)	8.39 mmol ATP gDW-1 hr-1
media conditions	computational minimal media
carbon source	8 mmol glucose gDw-1 hr-1
aerobic or anaerobic	18.5 mmol O2 gDw-1 hr-1
additional constraints	
reactions constrained to zero	CAT, SPODM, SPODMpp, FHL
flux split between reaction pairs	none

SBML File Properties	
file name	*E. coli* _iAF1260_flux2.xml
organism	E. coli K-12 MG1655
model	iAF1260
Biomass Objective Function (BOF)	Ec_biomass_iAF1260_core_59p81M (E. coli biomass objective function (iAF1260) - core - with 59.81 GAM estimate)
flux balance analysis objective	maximize BOF
Growth Associated Maintenance (GAM)	59.81 mmol ATP gDW-1
Non-Growth Associated Maintenance (NGAM)	8.39 mmol ATP gDW-1 hr-1
media conditions	computational minimal media
carbon source	11.0 mmol glucose gDw-1 hr-1
aerobic or anaerobic	18.2 mmol O2 gDw-1 hr-1
additional constraints	
reactions constrained to zero	152 reactions identified to be unavailable to the cell under glucose aerobic conditions
flux split between reaction pairs	NDH-1:NDH-2 is 1:1 (3 pairs of reactions - different quinone usage) NADH10:NADH17pp, NADH5:NADH16pp, NADH9:NADH18pp

Supplementary Table I - Unique ORFs from each COG functional class in *i*AF1260

The table gives the total and unique number of ORFs from each COG functional class for both *E. coli* (based off the latest genome annotation (Riley et al, 2006)) and those included in *i*AF1260.

COG Functional Class	Abbr.	iAF1260 unique	iAF1260 total	iAF1260 % unique	E. coli unique	E. coli total	E. coli % unique
Nucleotide transport and metabolism	F	65	72	90%	70	81	86%
Cell wall/membrane/envelope biogenesis	M	90	102	88%	179	214	84%
Energy production and conversion	C	156	177	88%	233	269	87%
Carbohydrate transport and metabolism	G	160	194	82%	239	339	71%
Lipid transport and metabolism	I	43	58	74%	63	90	70%
Amino acid transport and metabolism	E	183	256	71%	240	373	64%
Coenzyme transport and metabolism	H	75	106	71%	97	132	73%
Inorganic ion transport and metabolism	P	87	134	65%	157	241	65%
J, K, L, O, S, T, U, V	OT	59	99	60%	799	1126	71%
Secondary metabolites biosynthesis, transport and catabolism	Q	10	23	43%	28	65	43%
General function prediction only	R	41	96	43%	273	419	65%

Supplementary Table II - The *in silico* (computational) minimal media

Metabolite Name	Metabolite Abbreviation	Excahnge Reaction	Reaction Equation	Note
calcium	ca2	EX_ca2(e)	[e] : ca2 <==>	ion
chloride	cl	EX_cl(e)	[e] : cl <==>	ion
carbon dioxide	co2	EX_co2(e)	[e] : co2 <==>	other
cobalt (2+)	cobalt2	EX_cobalt2(e)	[e] : cobalt2 <==>	ion
copper (2+)	cu2	EX_cu2(e)	[e] : cu2 <==>	ion
iron (2+)	fe2	EX_fe2(e)	[e] : fe2 <==>	ion
iron (3+)	fe3	EX_fe3(e)	[e] : fe3 <==>	ion
proton	h	EX_h(e)	[e] : h <==>	ion
water	h2o	EX_h2o(e)	[e] : h2o <==>	other
potassium	k	EX_k(e)	[e] : k <==>	ion
magnesium (2+)	mg2	EX_mg2(e)	[e] : mg2 <==>	ion
manganese (2+)	mn2	EX_mn2(e)	[e] : mn2 <==>	ion
molybdate	mobd	EX_mobd(e)	[e] : mobd <==>	ion
sodium	na1	EX_na1(e)	[e] : na1 <==>	ion
ammonium	nh4	EX_nh4(e)	[e] : nh4 <==>	N source
phosphate	pi	EX_pi(e)	[e] : pi <==>	P source
sulfate	so4	EX_so4(e)	[e] : so4 <==>	S source
tungstate	tungs	EX_tungs(e)	[e] : tungs <==>	ion
zinc	zn2	EX_zn2(e)	[e] : zn2 <==>	ion
Necessary for the BOFWT simulations				
cob(i)alamin	cbl1	EX_cbl1(e)	[e] : cbl1 <==>	precursor (vitamin B12)

Supplementary Table III - Essentiality predictions using *i*JR904

Comparison of computational essentiality predictions using *i*JR904 (Reed *et al*, 2003) and high-throughput experimental data (Baba *et al*, 2006; Joyce *et al*, 2006). The table is similar to Table 4 which is based on results from *i*AF1260.

		experimental	
		essential	non-essential
computational	essential	139 (15%)	40 (4%)
	nonessential	69 (8%)	656 (73%)

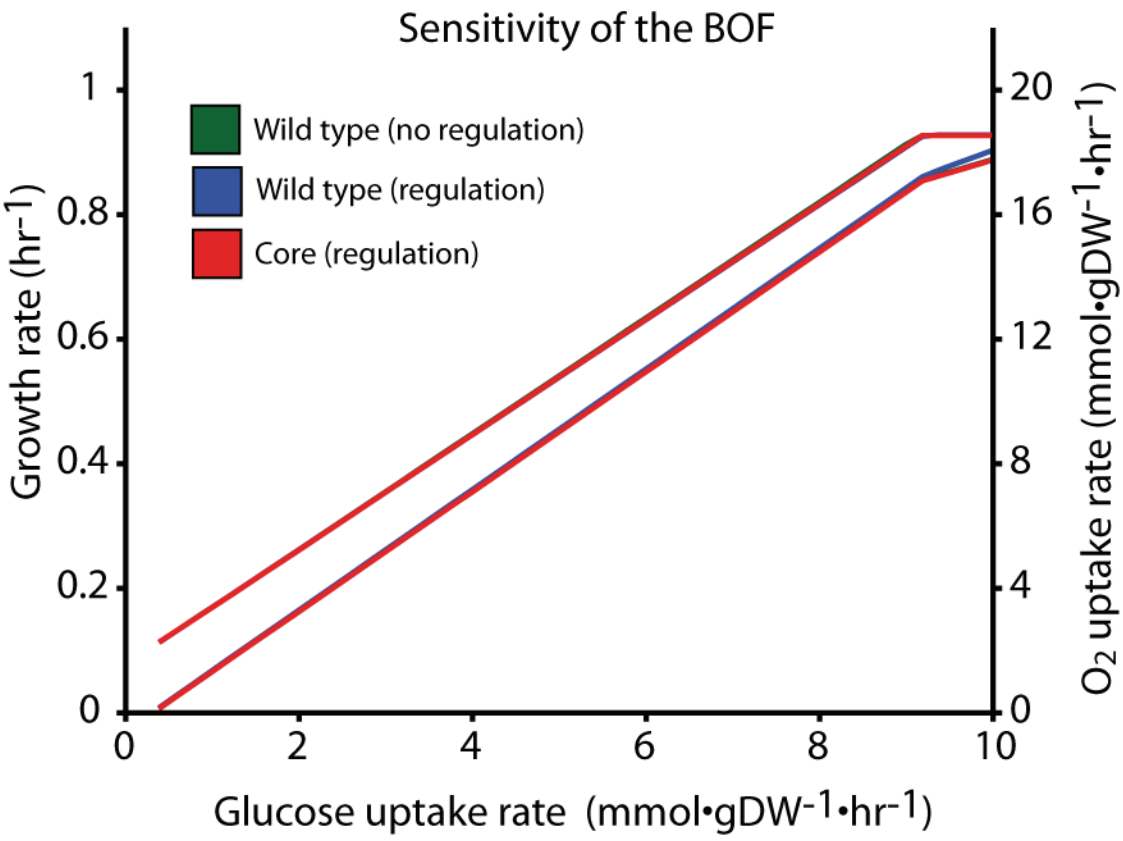
Supplementary Table IV – Results of the mapping between *i*AF1260 and EcoCyc

compounds	**number**	**%**	**manual**	**%**
total mapped	**945**	**91%**	**267**	**26%**
mapped to EcoCyc IDs	883		205	
mapped to MetaCyc IDs	62		62	
total unmapped *	**94**	**9%**		
total	**1039**			

reactions	**number**	
total mapped	**1308**	**63%**
intracellular or transport	1080	
diffusion	228	
total unmapped	**769**	**37%**
contain unmapped compounds	232	
unmapped	537	
total	**2077**	

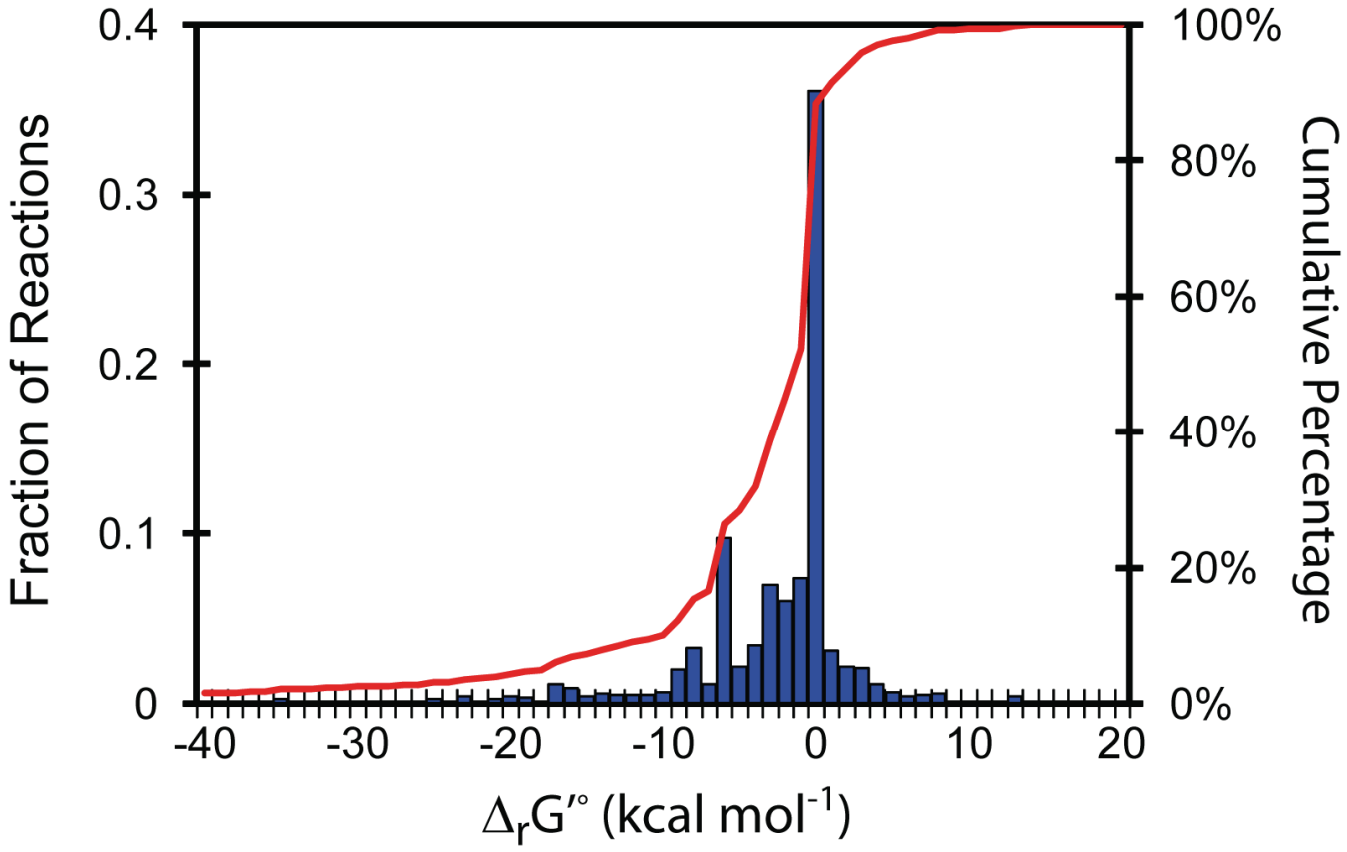
* All unmapped compounds have been evaluated manually and have either been matched to a future release of EcoCyc or have been categorized for future inclusion.

Supplementary Figure 1 - Sensitivity analysis varying the biomass objective function



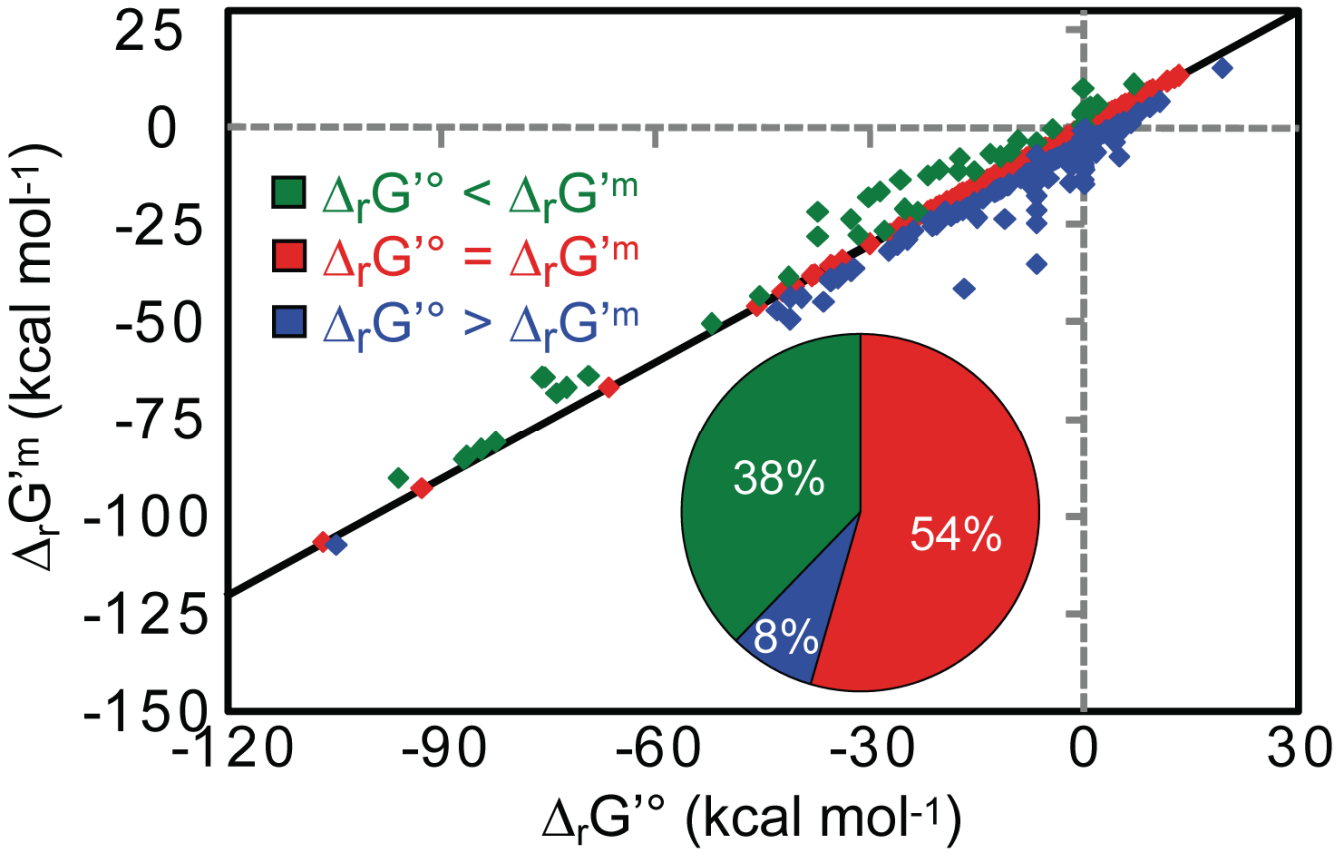
The relationship between the glucose uptake rate (mmol gDW^{-1} hr^{-1}) (bottom axes, the dependant variable) and the resulting 1) growth rate (hr^{-1}) (left axes) and 2) oxygen uptake rate (mmol gDW^{-1} hr^{-1}) (right axes) produced during the sensitivity analysis using *i*AF1260. Using FBA and *i*AF1260, optimal growth was simulated under glucose aerobic conditions while varying which biomass objective function (BOF) was used along with the number of reactions available to the network due to transcriptional regulation. Two different BOFs were used, a core biomass objective function and wild-type biomass objective function, and regulation was imposed by not allowing any flux through reactions unavailable to the network due to transcriptional regulation (Covert *et al*, 2004). The results show that the predicted optimal growth rate and O_2 uptake rate are insensitive to the BOF used or level of transcriptional regulation imposed under these conditions.

Supplementary Figure 2 – Estimated $\Delta_r G'^{o}$ values for the reactions in *i*AF1260



The distribution of estimated $\Delta_r G'^{o}$ values for the reactions in *i*AF1260. This histogram compliments Figure 2A which contains estimated $\Delta_r G'^{m}$ values. The distribution of $\Delta_r G'^{m}$ values for the reactions in *i*AF1260 differs significantly from the distribution of $\Delta_r G'^{o}$ values as there are significantly fewer $\Delta_r G'^{m}$ that are between zero and 1 kcal mol^{-1} and many more $\Delta_r G'^{m}$ values that are less and 0 kcal mol^{-1}. Overall, this indicates that most reactions in the *i*AF1260 are more thermodynamically favorable at the mM concentration conditions that exist within the cell.

Supplementary Figure 3 – Relationship between $\Delta_r G'^{\circ}$ and $\Delta_r G'^{m}$ for *i*AF1260 reactions.



Shown is a plot of the estimated $\Delta_r G'^{\circ}$ values against $\Delta_r G'^{m}$ values for each reaction in *i*AF1260. From the scatter plot and the pie chart, $\Delta_r G'^{m}$ is most often lower than $\Delta_r G'^{\circ}$ for the majority of the reactions in *i*AF1260. This indicates that the majority of the reactions in *i*AF1260 are more favorable in direction of operation required for near optimal growth at the 1 mM concentrations levels found in the cell than at 1 M concentrations. This also indicates that most biological reactions have more products than reactants in the direction of operation required for near optimal growth on the 174 carbon sources examined. This result is to be expected as the purpose of metabolism is to produce all of the metabolites required for cell growth from a small number of nutrient compounds.

References

Alexander K, Young IG (1978) Alternative hydroxylases for the aerobic and anaerobic biosynthesis of ubiquinone in *Escherichia coli*. *Biochemistry* **17:** 4750-4755.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2:** 2006.0008.

Beale SI (1996) Biosynthesis of Hemes. In *Escherichia coli and Salmonella*, Neidhardt FC (ed) pp 731 - 748. ASM Press, Washington, DC.

Choi-Rhee E, Cronan JE (2005) A nucleosidase required for in vivo function of the S-adenosyl-L-methionine radical enzyme, biotin synthase. *Chem Biol* **12:** 589-593.

Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429:** 92-96.

Gennis RB, Stewart V (1996) Respiration. In *Escherichia coli and Salmonella*, Neidhardt FC (ed) pp 217-261. ASM Press, Washington, DC.

Green ML, Karp PD (2004) A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC Bioinformatics* **5:** 76.

Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, Mori H, Lesely SA, Palsson BO, Agarwalla S (2006) Experimental and Computational Assessment of Conditionally Essential Genes in *Escherichia coli*. *J Bacteriol* **188:** 8259-8271.

Leonardi R, Roach PL (2004) Thiamine biosynthesis in Escherichia coli: in vitro reconstitution of the thiazole synthase activity. *J Biol Chem* **279:** 17054-17062.

Meganathan R (1996) Biosynthesis of Isoprenoid Quinones Menaquinone (Vitamin K_2) and Ubiquinone (Coenzyme Q). In *Escherichia coli and Salmonella*, Neidhardt FC (ed) pp 642-656. ASM Press, Washington, DC.

O'Toole GA, Rondon MR, Trzebiatowski JR, Suh S-J, Escalante-Semerena JC (1996) Biosynthesis and Utilization of Adenosyl-Cobalamin (Coenzyme B12). In *Escherichia coli and Salmonella*, Neidhardt FC (ed) pp 710-720. ASM Press, Washington, DC.

Puustinen A, Finel M, Haltia T, Gennis RB, Wikstrom M (1991) Properties of the two terminal oxidases of Escherichia coli. *Biochemistry* **30:** 3936-3942.

Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of *Escherichia coli* K-12 (*i*JR904 GSM/GPR). *Genome Biology* **4:** R54.51-R54.12.

Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, Plunkett G, 3rd, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D, Wanner BL (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot-2005. *Nucleic Acids Res* **34:** 1-9.

Sekowska A, Danchin A (2002) The methionine salvage pathway in Bacillus subtilis. *BMC Microbiol* **2:** 8.

Tran QH, Bongaerts J, Vlad D, Unden G (1997) Requirement for the proton-pumping NADH dehydrogenase I of *Escherichia coli* in respiration of NADH to fumarate and its bioenergetic implications. *Eur J Biochem* **244:** 155-160.

Varma A, Palsson BO (1995) Parametric sensitivity of stoichiometric flux balance models applied to wild-type *Escherichia coli* metabolism. *Biotechnol Bioeng* **45:** 69-79.

Weidner U, Geier S, Ptock A, Friedrich T, Leif H, Weiss H (1993) The gene locus of the proton-translocating NADH: ubiquinone oxidoreductase in Escherichia coli. Organization of the 14 genes and relationship between the derived proteins and subunits of mitochondrial complex I. *J Mol Biol* **233:** 109-122.