Supplementary Material (ESI) for Molecular BioSystems
This journal is (c) The Royal Society of Chemistry, 2007

Reconstruction, Modeling & Analysis of *Halobacterium salinarum* R-1 Metabolism (Supplements)

Biomass Determination The cellular concentration of each amino acid except for cysteine, tryptophan, glutamine and asparagine was experimentally measured at various optical densities. Glutamine and asparagine appear in the data as glutamate and aspartate respectively because they are immediately deaminated upon protein hydrolysis. Consequently, theoretical values were used for both amino acids and glutamate and aspartate were corrected accordingly. Theoretical values were also used for cysteine and tryptophan as they could not be reliably measured.

The deoxyribonucleotides were counted directly from the main chromosome and the 4 megaplasmids assuming one copy of each. Ribonucleotides were approximated to comprise 20% of a cell's organic mass and the individual copy numbers of AMP, UMP, GMP, and CMP were computed accordingly based on the chromosome's 68% GC composition.

Rather than employing a true cell wall, *H. salinarum* uses what is called a surface-layer as its outermost covering (Lechner et al, 1987; Blauroock et al, 1976). This S-layer contains a glycoprotein that accounts for approximately 50% of the protein and all of the nonlipid carbohydrates in this outer layer. We estimated the glycoprotein's copy number using structural data from the closely related organism *H. volcanii* (Kessel et al, 1988) and the average surface area of *H. salinarum* cells (9.99 x 10^{-8}cm^2). Since there is still no assembly pathway for the complex, its components were included separately in the biomass using a charaterization found in the literature (Lechner et al, 1987; Mescher et al, 1976).

The membrane of *H. salinarum* is composed of proteins and a lipid bilayer composed of isoprenyl chains attached to glycerol (Kamekura, 1993, 1998; Kamekura and Kates, 1988). For simplicity, we only used archaeol since it is a known late precursor of most archeal membrane lipids. With the assumption that lipids occupy 50% of the total cellular surface area and proteins the remaining 50%, the lipid's copy number was estimated by dividing the surface area by the molecule's area of $65A^2$.

In addition, the organism is also composed of other compounds which may be essential to its survival but contribute only little with respect to cell mass. These compounds were assumed to constitute 0.01% of the total organic mass. Currently, these are siroheme, FAD, NAD+, NADP+, tetrahydrofolate (THF), thiamin, coenzyme-A, menaquinone, and cobamide coenzyme.

The Growth Function The growth function, defined as a pseudo-reaction that consumes cellular components in the proper ratio and produces a cell, was used as the Flux Balance Analysis (FBA) optimization basis. Given that this preudo-reaction corresponds to column j of the stoichiometric matrix S, then the value of each s_{ij} is negative and reflects how much of the compound i is present in a cell. For example, if $i = a$ is the index for the compound alanine, then $s_{aj} = -186.38$ since there are about 186.38 attomoles of this amino acid in each cell. The only exception is the index $i = c$ which corresponds to a cell. This is set to 1 because the equivalent of one cell is produced by reaction j. Note

that the biomass composition was calculated using population averages. Consequently, variations between cells in the population are not modeled. Moreover, using this reaction as the basis for optimization to determine fluxes carries the implicit assumption that cells optimize their metabolism for biomass production (growth).

Energy Consumption An important aspect of the dynamic model is the quantification of the energy used by cells which are not explicitly accounted for in synthesizing the compounds defined in the biomass composition (as defined above). Contributing endergonic processes include protein turnover, cellular movement, and gene regulation. We lumped these processes together and modeled the aggregate in terms of the equivalent energy released in the reduction of a specific amount of O_2. As a theoretical upperbound for this value, we calculated the differences between the measured uptake and biomass incorporation rates of the supplied nutrients. With the assumption that nutrients that do not turn up in the biomass are converted to energy, the residuals are the equivalent of 45 nmol O_2 Klett^{-1} ml^{-1} hr^{-1}. Note that cultures were given abundant oxygen supply and arginine is the only known fermentative substrate of *H. salinarum*. The best fit of the model was achieved with a value of 41.6 nmol O_2 Klett^{-1} ml^{-1} hr^{-1}. Comparing this with the maximum value of 12.16 nmol O_2 Klett^{-1} ml^{-1} hr^{-1} from oxygen consumption data taken from cells grown in rich medium (Oesterhelt et al, 1973), the higher values indicate that cells grown in synthetic media achieve lower growth rates even with higher energy consumption. In the future, distinguishing between growth related and non-growth related energy consumption should increase the model's accuracy (Varma and Palsson, 1994).

Oxaloacetate conversion to Acetyl-CoA It was stated in the main text (Section 2.7.1) that cells under the specified conditions rely heavily on the conversion of oxaloacetate to acetyl-CoA. As this process represents a substantial flux into the TCA cycle, we analyzed it in more detail and querried the model for alternative routes. We did this by restricting flux into the described pathway and subsequently inspecting how the other fluxes compensated. We found an alternative route which the organism can utilize without any loss of optimality. The alternative begins by transaminating oxaloacetate to form aspartate, which is then decarboxylated to form β-alanine. The product then loses the amonium moiety through another transamination reaction to form 3-oxopropanoate, which is then oxidatively decarboxylated to form acetyl-CoA. Although this alternative pathway is more complex, it yielded the identical net reaction

$$\mathsf{oxaloacetate} + \mathsf{CoA} + \mathsf{carrier\text{-}H} \rightarrow \mathsf{acetyl\text{-}CoA} + 2\ \mathsf{CO_2} + \mathsf{carrier}$$

Flux Variability Analysis: Suboptimal Pathways and Processes As stated in the main text (Section 2.7.2), we performed flux variability analysis with the constraint on the objective function relaxed to $\geq$90% of the optimal value to analyze possible suboptimal processes which may occur given the specified conditions. Again, we note that the computed variabilities are indicative of the confidence we may attribute to a particular flux prediction. In the main text, we discussed processes for which the predicted fluxes are at least qualitatively accurate, and noted that the majority of the reactions in the network are of this

type. Here we discuss processes with variabilities which are less constrained. In this context, an interesting pathway is the TCA cycle. While the calculated variabilities showed that most of the reactions are constrained to flow in the classical (oxidative) direction, the arc from isocitrate to succinyl-CoA exhibited the possibility of reversing in direction. Inspection of nearby reactions revealed that this is due to the action of isocitrate-glyoxylate lyase. Thus, the glyoxylate shunt is a sub-optimal alternative to the complete TCA cycle. Note however that currently, only at most very weak activity has been detected of any of the glyoxylate cycle enzymes (Oren and Gurevich, 1995a; Aitken and Brown, 1969). It is likely that very little is diverted from the TCA cycle if the glyoxylate cycle is active at all. Similarly, variabilities also showed that degradation of threonine through 2-oxobutanoate is a suboptimal alternative to the route indicated in Figure 3. Through interconversions this route is also possible for glycine and serine.

In addition to suboptimal pathways, two other consequences of relaxing the constraints on optimality are: (1) complex biomass constituents supplied in the medium can be synthesized rather than transported. For example, folate can be synthesized *de novo* rather than transported even though transport is more efficient as its biosythetic pathway is quite complex. And (2), energy dissipating futile cycles can operate. For instance, cells can "waste" ATP through the combined action of a kinase and a hydrolase. A particular example of this is glucose which can be phosphorylated by glucokinase to glucose 6-phosphate using ATP. Glucose 6-phosphate can then be converted back to glucose via glucose 6-phosphate phosphohydrolase. Note that such futile cycles, if involved in larger pathways such as gluconeogenesis in the case of the example, do not affect the net flux through the pathway.

Computational Details The core of our simulation code was written and executed in Matlab (Matlab, http://www.mathworks.com/products/matlab/). The default linear programming package however was unable to handle the scale of our network so we integrated the GNU Linear Programming Kit (GLPK, http://www.gnu.org/software/glpk/) to perform Flux Balance Analysis.

During simulations, exchange fluxes were provided for all amino acids, all ribonucleotides, ornithine, acetate, NH_3, thiamin, sulfate, phosphate, biotin, nitrate, cobamide coenzyme, iron, folate, fatty acids, cobalt, lactate, potassium, ribose, glycerol, succinate, methanethiol, glucose, galactose, sodium, and chloride. The positive flow of these fluxes into the metabolic network is controlled by the nutrients layer which is initialized based on the medium used. In addition, free exchange fluxes were provided for H+, CO_2, O_2 and H_2O.

References

Aitken DM, Brown AD (1969) Citrate and glyoxylate cycles in the halophil, *Halobacterium salinarum. Biochim. Biophys. Acta* 177:351-354.

Blaurock AE, Stoeckenius W, Oesterhelt D, Scherphof GL (1976) Structure of the cell envelope of *Halobacterium Halobium. The Journal of Cell Biology* 71:1-22.

GLPK (GNU Linear Programming Kit). *http://www.gnu.org/software/glpk/* accessed March, 2007.

Kamekura M (1993) in *The biology of halophilic bacteria.*, eds Vreeland RH, Hochstein LI (CRC Press, Boca Raton), pp. 135-161.

Kamekura M (1998) Diversity of extremely halophilic bacteria. *Extremophiles* 2:289-295.

Kamekura M, Kates M, (1988) in *Halophilic bacteria Vol. II*, eds Rogriguez-Valera, F (CRC Press, Boca Raton), pp. 25-54.

Kessel M, Wildhaber I, Cohen S, Baumeister W (1988) Three dimensional structure of the regular surface glycoprotein layer of *Halobacterium volcanii* from the Dead Sea. *The EMBO Journal* 7(5):1549-1554.

Lechner J, Sumper M (1987) The Primary Structure of Procaryotic Glycoprotein. *The Journal of Biological Chemistry* 262(20):9724-9729.

Matlab. http://www.mathworks.com/products/matlab/ accessed March, 2007.

Mescher MF, Strominger JL (1976) Purification and Characterization of a Prokaryotic Glycoprotein from the Cell Envelope of Halobacterium salinarum. *The Journal of Biological Chemistry* 251(7):2005-2014.

Oesterhelt D, Krippahl G (1973) Light inhibition of respiration in *Halobacterium halobium*. *FEBS Letters* 36(1):72-76.

Oren A, Gurevich P (1995) Isocitrate lyase activity in halophilic archaea. *FEMS Microbiol. Lett.* 130:91-95.

Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* 60:3724-3731.

Table III: *H. salinarum* average cellular biomass composition.

Cmpd.	Amount (μg/ODml)	Cmpd.	Amount (μg/ODml)
Amino Acids		Nucleotides	
Ala	22.5 ± 5.2	dAMP	0.6
Arg	26.9 ± 5.6	dTMP	0.6
Asp	42.0 ± 9.2	dGMP	1.3
Asn	9.6	dCMP	1.1
Cys	2.9	AMP	15.7
Glu	77.8 ± 23.2	UMP	14.7
Gln	13.4	GMP	34.9
Gly	15.9 ± 2.9	CMP	31.1
His	15.1 ± 3.0		
Ile	13.1 ± 3.0	S-Layer non AA	
Leu	24.1 ± 5.0	GalNAc	0.5
Lys	12.3 ± 2.5	GlcNAc	0.5
Met	5.4	Gal	1.7
Phe	13.4 ± 2.6	Glc	1.5
Pro	18.5 ± 4.5		
Ser	13.9 ± 2.8	Membrane	
Thr	19.3 ± 4.2	Archaeol	20.0
Trp	7.7		
Tyr	6.3 ± 3.0	Others	
Val	21.4 ± 4.9	ATP	2.0

Amino acid compositions were measured using an amino acid analyzer at different optical densities. Each showed a good linear correlation with the optical densitiy indicating that the average amino acid composition during growth remains constant. Methionine measurements were extremely noisy. Other constituents were approximated from available genomic, proteomic, and physiological data. Experimentally determined values are provided with error margins. 1 ODml corresponds to about 732 μgDW.

Table IV: Summary of amino acid biosynthetic and degradative pathways.

AA	**Synthesis**	**Degradation**	**AA**	**Synthesis**	**Degradation**
Ala	yes	yes	Leu	no	probable
Arg	no	yes	Lys	no	probable
Asp	yes	yes	Met	no	yes
Asn	yes		Phe	yes	probable
Cys	yes		Pro	yes	yes
Glu	yes	yes	Ser	yes	yes
Gln	yes	yes	Thr	yes	yes
Gly	yes	yes	Trp	yes	no
His	yes	probable	Tyr	yes	probable
Ile	no	probable	Val	no	probable

Summary of amino acid biosynthetic and degradative pathways in *H. salinarum*. "Probable" indicates that none or only an incomplete set of degradative enzymes was found for an amino acid, but simulation results predict that the pathway should be there. The most common reason is that the amino acid's uptake rate significantly exceeds its biomass incorporation rate.

Table V: Compound Abbreviations

Abbrev.	**Name**	**Abbrev.**	**Name.**
2-akg	α-ketoglutarate (2-Oxoglutarate)	iasp	Iminoaspartate
2-obut	2-Oxobutanoate	i-cit	Isocitrate
5,10mtthf	5,10-Methenyltetrahydrofolate	Ile	L-Isoleucine
ac-CoA	Acetyl-CoA	ipdp	Isopentenyl diphosphate
acac	Acetoacetate	Leu	L-Leucine
ADP	ADP	Lys	L-Lysine
Ala	L-Alanine	mal	Malate
β-Ala	β-Alanine	Met	L-Methionine
Arg	L-Arginine	mev	Mevalonate
Asn	L-Asparagine	NAD+	NAD+
asp4sa	L-Aspartate	NADH	NADH
asp4sa	L-Aspartate 4-semialdehyde	NH_3	Ammonia
ATP	ATP	orn	Ornithine
cbp	Carbamoyl phosphate	oxal	Oxaloacetate
CH4S	Methanethiol	pep	Phosphoenolpyruvate
chor	Chorismate	Phe	L-Phenylalanine
cit	Citrate	pnto	(R)-Pantothenate
ctrl	Citruline	pphn	Prephenate
CO_2	Carbon dioxide	prec2	Precorrin 2
Cys	Cysteine	Pro	L-Proline
dkfrc-p	6-deoxy-5-Ketofructose 1-P	prop-CoA	Propionyl-CoA
fbp	Fructose 1,6-bisphosphate	prpp	5-Phospho-ribose 1-PP
fum	Fumarate	pyr	Pyruvate
g3p	Glyceraldehyde 3-phosphate	ru5p	Ribulose 5-phosphate
g6p	Glucose 6-phosphate	Ser	L-Serine
gal	Galactose	skm	Shikimate
glc	Glucose	succ	Succinate
glcn	Gluconate	succ-CoA	Succinyl-CoA
Gln	L-Glutamine	thf	Tetrahydrofolate
Glu	L-Glutamate	Thr	L-Threonine
Gly	Glycine	Trp	L-Tryptophan
glyc	Glycerol	Tyr	L-Tyrosine
His	L-Histidine	uppg3	Uroporphyrinogen III
hmg-CoA	Hydroxymethylglutaryl-CoA	Val	L-Valine

Compound abbreviations used in Figure 1.

Table VI: Synthetic medium with 15 amino acids

NaCl	4 M	L-Ala	7.4 mM
KNO_3	1 mM	L-Arg	3.5 mM
KCl	27 mM	L-Ile	3.4 mM
$MgSO_4$ * $7H_2O$	81 mM	L-Met	2.3 mM
Sodium Citrate * H_2O	1.7 mM	L-Pro	2.4 mM
K_2HPO_4	0.42 mM	L-Phe	0.8 mM
KH_2PO_4	0.58 mM	L-Ser	12.6 mM
$FeSO_4$ * $7H_2O$	11.6 μM	L-Thr	8.3 mM
$CuSO_4$ * $5H_2O$	0.2 μM	L-Tyr	1.5 mM
$MnCl_2$ * $4H_2O$	1.8 μM	L-Val	1.0 mM
$ZnSO_4$ * $7H_2O$	1.5 μM	L-Lys	2.7 mM
Na_2MoO_4 * $2H_2O$	0.1 μM	Thiamin	16.5 μM
L-Asp	6.2 mM	Folate	11.5 μM
L-Glu	16.9 mM	Biotin	2.1 μM
L-Leu	10 mM	pH	7.2
L-Gly	2.9 mM		

Synthetic growth medium.

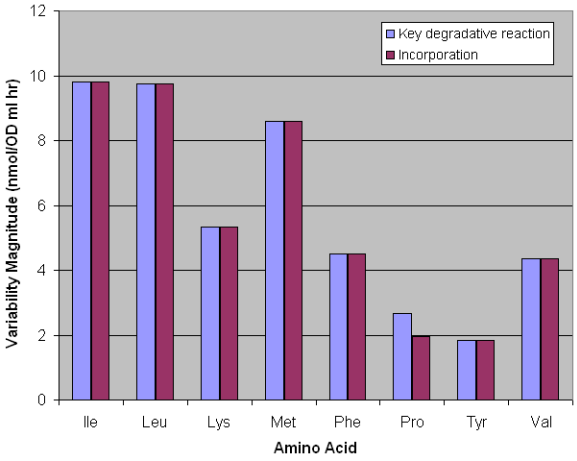

Figure 4: Variability magnitudes of key reactions in degradative pathways of some amino acids and the variabilities of the corresponding incorporation rate (variability of growth function multiplied by the appropriate amino acid coefficient) using a 90% suboptimality threshold. Variabilities indicated that the 8 degradative processes are active under all suboptimal distributions within the threshold. Moreover, the fact that the variability of each of these pathways can be attributed mostly to the variability of the corresponding and competing incorporation rate implies that the the quantitative aspects of the predicted fluxes for these pathways should also be accurate even beyond the threshold used. Similar conclusions could not be reached for the other amino acids as their variabilities indicated the presence of alternative suboptimal pathways.