Construction of Biomass Reaction for *Halobacterium Salinarum* NRC-1 using whole genome sequence

*[1]R. V. Kavitha, [2]Suresh R, [3]Makam V Roshan, [4]Arjun S Pejathaya, [5]Rhea Harish, [6]Sanjana Karthik
[1]Assistant Professor, [2]Professor and Dean, [3]Professor and Ex-Chairperson
[1,3,4,5,6]Biotechnology Department, PES University
100 Feet Ring Road, BSK 3rd Stage, Bangalore – 560085, India
[2]Chemical Engineering Department, RVCE
Mysore Road, Bangalore, India

ABSTRACT
With the development of whole genome sequencing and the use of metabolic engineering to optimize the production of biomass or a metabolite, it is very necessary to know the components and its relation to biomass reaction. This work uses whole genome sequence information to develop a biomass reaction with components such as DNA, RNA, proteins, S-layer non-amino acids, membrane lipids, ATP from growth and non-growth associated and others using the building blocks associated with it respectively. Stoichiometric reactions for DNA, RNA, proteins, S-later non-amino acids, membrane lipids, growth and non-growth associated ATP and others were developed. A biomass reaction was constructed and arrived at for producing 1 g of dry weight of biomass theoretically. Also, a stoichiometric biomass equation was developed. Both the biomass reaction and the equation can be used in in-silico metabolic engineering for optimization of biomass or a metabolite.

Keywords: *Biomass, Reaction, halobacterium, genome, sequence*

INTRODUCTION

With the advent of whole genome sequencing and metabolic engineering there is an ever need to determine the biomass equation since optimizing the metabolite or the biomass is the goal of metabolic engineering [1], [2]. *Halobacterium Salinarum* NRC-1 is a model organism which thrives well in high salt concentrations and is usually found in the Dead Sea as well as Salt taverns [5]. It membrane contains 70-80% of bacteriorhodopsin (bR) which has a wide variety of applications such as optical devices, memory chips, water purification, etc., since it acts both as a light harvesting membrane protein and a proton pump [6]. With the whole genome of *Halobacterium Salinarum* NRC-1 sequenced its use is in metabolic engineering to produce bR or enhancing biomass hence, there is the need for determining the biomass reaction [3], [4]. This research aims at developing theoretically the biomass reaction and stoichiometric biomass equation using the whole genome sequence intended for metabolic engineering.

DATA ANALYSIS

Halobacterium Salinarum NRC-1 (ATCC 70092) complete genome ASM680v1consisting of the chromosome (2,014,239 nucleotides), plasmid pNRC100 (191,346 nucleotides) and plasmid pNRC200 (365,425 nucleotides) and the FASTA sequences of the proteins were downloaded from National Center for Biotechnology Information (NCBI). The deoxy-ribose nucleotides numbers from the DNA sequence, the ribose nucleotides numbers for the RNA sequence, and the amino acids numbers from all the FASTA sequences of the proteins were counted and tabulated in Table 1, Table 2 and Table 3 respectively. The building blocks (BB) of DNA, RNA and proteins were converted to grams of BB per gram Dry Weight (DW) using the following equations:

$$Total_{BB} = \sum BB_{Number} \quad (1)$$

$$Fraction_{BB} = \frac{BB_{Number}}{Total_{BB}} \quad (2)$$

$$g_{BB} = Fraction_{BB}\ x\ MW_{BB} = \left(\frac{g_{BB}}{mol_{macromolecule}}\right) \quad (3)$$

$$Total\ Weight_{BB} = \sum g_{BB} \quad (4)$$

$$Weight\ Fraction_{BB} = \frac{g_{BB}}{Total\ Weight_{BB}} = \left(\frac{g_{BB}}{g_{macromolecule}}\right) \quad (5)$$

$$\left(\frac{mmol_{BB}}{g_{DW}}\right) = \frac{Weight\ Fraction_{BB}\ x\ (g_{macromolecule}/g_{Dw})\ x\ 1000}{MW_{BB}} \quad (6)$$

$$\left(\frac{g_{BB}}{g_{DW}}\right) = \frac{\left(mmol_{BB}/g_{DW}\right) x\ MW_{BB}}{1000} \quad (7)$$

Similarly, the above equations were used for S-Layer non amino acids, Membrane lipid (Archaeol), Adenosine Triphosphate from associated growth and non-growth and others. The S-layer non amino acids and others are listed in Table 4-7.

The elemental composition of each building block consisting of Carbon, Hydrogen, Nitrogen, Oxygen, Sulphur and Phosphorus were calculated and the final stoichiometric composition of biomass was determined.

RESULTS AND DISCUSSION

Based on the above data analysis and using the building blocks composition for DNA, RNA, Protein, S-Layer non AA, Membrane, ATP and others were arrived at using equation 1 through equation 7. They are listed in Table 1 through Table 7. The total number of nucleotide triphosphates for DNA were 2571012 while that of RNA were 2574423. On a similar basis the total number of amino acids were 725226 for protein. *Halobacterium Salinarum* NRC-1 consists of an S-layer comprised of glycoproteins which are listed in Table 4. Archeol was considered as the main membrane lipid present. The others consisted of Seroheme, FAD, NAD+, NADP+, Tetrahydrofolate, Thiamin, Coenzyme A, Menaquinone and Cobamide Coenzyme and are listed in Table 7. Each of compositions were computed both in mmol/g-DW and g/g-DW. The final composition was computed and listed in Table 8.

Using the compositions and the moles of each of the building blocks, reactions for DNA (equation 8), RNA (equation 9), Protein (equation 10), S-Layer non-amino acids (equation 11) were arrived at. Finally, the theoretical biomass reaction was constructed using the mole percentage composition of all the components and is given in equation 12. It was found that approximately 90% of the biomass comprised of proteins. Further, using the chemical formulae of all the building blocks of the components, a stoichiometric biomass equation (chemical formula) was developed. Equation 8 through equation 12 can be used for in-silico metabolic engineering studies.

Table 1 : DNA composition in *Halobacterium Salinarum* NRC-1

DNA						
Nucleotide triphosphates	**No. per genome**	**% per genome**	**g /mol macromolecule**	**g /g macromolecule**	**mmol /g DW**	**g/g DW**
dATP	436710	0.170	83.43	0.171	0.005	0.002
dGTP	847575	0.330	167.20	0.343	0.010	0.005
dCTP	847585	0.330	154.01	0.316	0.010	0.004
dTTP	439142	0.171	82.36	0.169	0.005	0.002
Total	2571012	1	487.00	1	0.029	
g Nut/g DW	0.014					

Table 2 : RNA composition in *Halobacterium Salinarum* NRC-1

RNA						
Nucleotide triphosphates	**No. per genome**	**% per genome**	**g /mol macromolecule**	**g /g macromolecule**	**mmol /g DW**	**g/g DW**
ATP	439957	0.171	86.79	0.173	0.045	0.023
GTP	848682	0.330	172.70	0.345	0.087	0.045
CTP	848462	0.330	159.45	0.318	0.087	0.042
UTP	437322	0.170	82.35	0.164	0.045	0.022
Total	2574423	1	501.29	1	0.264	
g Nut/g DW	0.132					

Table 3 : Amino Acid composition in *Halobacterium Salinarum* NRC-1 protein

Protein	**# of aa per genome**	**% of aa per genome**	**g aa/mol macromolecule**	**g aa/g macromolecule**	**mmol of aa/g DW**	**g/g DW**
Alanine	101185	0.140	12.417	0.090	0.526	0.047
Arginine	51506	0.071	12.358	0.089	0.268	0.047
Asparagine	17836	0.025	3.246	0.023	0.093	0.012

Aspartic acid	72320	0.100	13.263	0.096	0.376	0.050
Cysteine	5760	0.008	0.961	0.007	0.030	0.004
Glutamine	21995	0.030	4.428	0.032	0.114	0.017
Glutamic acid	56626	0.078	11.478	0.083	0.294	0.043
Glycine	66819	0.092	6.910	0.050	0.347	0.026
Histidine	17785	0.025	3.801	0.027	0.092	0.014
Isoleucine	30002	0.041	5.419	0.039	0.156	0.020
Leucine	67585	0.093	12.208	0.088	0.351	0.046
Lysine	14501	0.020	2.919	0.021	0.075	0.011
Methionine	14312	0.020	2.940	0.021	0.074	0.011
Phenylalanine	24731	0.034	5.627	0.041	0.128	0.021
Proline	36935	0.051	5.857	0.042	0.192	0.022
Serine	42214	0.058	6.112	0.044	0.219	0.023
Threonine	54252	0.075	8.902	0.064	0.282	0.034
Typtophan	8467	0.012	2.382	0.017	0.044	0.009
Tyrosine	20395	0.028	5.090	0.037	0.106	0.019
Valine	74457	0.103	12.012	0.087	0.387	0.045
Total	725226	1.000	138.331	1.000	4.155	0.521
g protein/g DW	0.5212					

Table 4 : S-Layer non-Amino acid composition in *Halobacterium Salinarum* NRC-1

S-Layer non AA	No. per genome	% per genome	g /mol macromolecule	g /g macromolecule	mmol /g DW	g/g DW
GalNAc	2.21E-006	0.108	23.78	0.119	0.003	0.001
GlcNAc	2.21E-006	0.108	23.78	0.119	0.003	0.001
Gal	8.55E-006	0.417	80.89	0.405	0.012	0.002
Glc	7.54E-006	0.368	71.35	0.357	0.011	0.002
Total	2.05E-005	1.000	199.81	1.000		0.006

Table 5 : Membrane composition in *Halobacterium Salinarum* NRC-1

Membrane	No. per genome	% per genome	g /mol macromolecule	g /g macromolecule	mmol /g DW	g/g DW
Archaeol	2.99E-05	1.000	652.00	1.000	0.042	0.027

Table 6 : Growth and Non-Growth related ATP and Others composition in *Halobacterium Salinarum* NRC-1

ATP and Others	No. per genome	% per genome	g /mol macromolecule	g /g macromolecule	mmol /g DW	g/g DW
ATP	5.04E-06	1.000	510.00	1.000	0.007	0.004
Others	9.98E-06	1.000	715.51	1.000	0.014	0.01

Table 7 : Other components composition in *Halobacterium Salinarum* NRC-1

Others
Seroheme
FAD
NAD+
NADP+
Tetrahydrofolate (THF)
Thiamin
Coenzyme A
Menaquinone
Cobamide Coenzyme

Table 8 : Average Biomass composition of *Halobacterium Salinarum* NRC-1 from whole genome

Component	g/g DW	wt %	mol	mol %
Protein	0.5212	0.7300	0.003811	0.8985
RNA	0.1320	0.1849	0.000264	0.0623

DNA	0.0141	0.0197	0.000029	0.0068
S-Layer non AA	0.0057	0.0080	0.000029	0.0067
Membrane	0.0273	0.0383	0.000042	0.0099
ATP	0.0036	0.0050	0.000007	0.0017
Others	0.0100	0.0140	0.000060	0.0141
Total	0.7140	1.0000	0.004242	1.0000

The average molecular weights of the BB for DNA, RNA, Protein, S-Layer non AA, Membrane, ATP and others are 486.92 g/mol, 499.41 g/mol, 136.75 g/mol, 200.69 g/mol, 653.17 g/mol, 507.18 g/mol and 166.67 g/mol. Using these molecular weights and the individual BB (g/g-DW Biomass) gave the following equations for each component of the Biomass:

$$0.170(dATP) + 0.330(dGTP) + 0.330(dCTP) + 0.171(dTTP) = DNA \quad (8)$$

$$0.171(ATP) + 0.330(GTP) + 0.330(CTP) + 0.170(UTP) = RNA \quad (9)$$

$$0.127(Ala) + 0.064(Arg) + 0.022(Asn) + 0.090(Asp) + 0.007(Cys) + 0.028(Gln) + 0.071(Glu) + 0.084(Gly) + 0.022(His) + 0.038(Ile) + 0.085(Leu) + 0.018(Lys) + 0.018(Met) + 0.031(Phe) + 0.046(Pro) + 0.053(Ser) + 0.068(Thr) + 0.011(Trp) + 0.026(Tyr) + 0.093(Val) = Protein \quad (10)$$

$$0.108(GalNAc) + 0.108(GlcNAc) + 0.417(Gal) + 0.368(Glc) = S\text{-}Layer\ non\ AA \quad (11)$$

The overall stoichiometric reaction for 1 g Dry Weight of Biomass is

$$0.0068(DNA) + 0.0623(RNA) + 0.8985(Protein) + 0.0067(S\text{-}Layer\ non\text{-}AA) + 0.0099(Membrane) + (0.0017ATP) + 0.0141(Others) = Biomass \quad (12)$$

The final chemical formula for Halobacterium Salinarum NRC-1 is $C_{3.699}H_{6.724}NO_{1.775}S_{0.065}P_{0.035}Fe_{0.001}Co_{0.001}$ giving a biomass of 160.71 yocto grams.

CONCLUSIONS

This work has presented theoretical data on constructing and developing a biomass reaction and stoichiometric biomass equation from whole genome sequence which can be used in metabolic engineering.

REFERENCES

[1] Gonzalez Orlando, Reconstruction, Modeling and Analysis of Haloarchaeal Metabolic networks, PhD Thesis, Ludwig-Maximilans-Universitat, Munchen, **2009,** 83-88.

[2] Thiele Ines and Paisson Bernhard O, A protocol for generating a high-quality genome-scale metabolic reconstruction, National Protocols, **2010**, 5(1): 93-121.

[3] Kamekura Masahiro and Kates Morris, Structural Diversity of Membrane Lipids in Members of halobacteriaceae, Bioscience, Biotechnology and Biochemistry, **1999**, 63(6): 969-972.

[4] Noguchi Yuuki et al, Composition Analysis of Polar Lipids in halobacteria with Mass Spectrometry, Journal of Mass Spectrometry Society Japan, **2004**, 52(6): 307-315.

[5] Baas Becking LGM, Historical notes on salt and salt manufacture, Sci Monthly, **1931**, 32: 434-446.

Tables