Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2015, 7(10):142-146

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Bioinformatic analysis of RpoH sigma factors from the anoxygenic phototrophic *Rhodobacter* species

Karli Geethanjali, K. Madhuri and Ramchander Merugu*

University College of Science and Informatics, Mahatma Gandhi University, Nalgonda

---

ABSTRACT

The composition of alanine, leucine, glutamic acid and arginine was the highest while lowest concentrations of histidine, threonine and phenylalanine and tryptophan. Cysteine was not found in any of the Rpo H sigma factors. There was no much variation in the content in proline and histidine of all the sigma factors. The number of negative charged residues is more compared to positively charged residues. The instability index of all the sigma factors was also same the highest being that of sigma factor sgf2. It was less than 40 showing that most of them are stable. Aliphatic index was found to span within a range of 82 to 87.

Key words: Sigma factors, *Rhodobacter* species, in silico analysis

---

INTRODUCTION

RNA polymerase is considered as one of the master protein in the living organisms reflecting the gene functions. In bacteria, though a single type of RNA Polymerases is known to be responsible for expression of all protein coding genes, they slightly vary at the initiation of transcription by virtue of a protein called sigma factor (σ factor). The type of sigma factor binding to the core RNA polymerase and then to a promoter vary according the type of the genes and several environmental signals. Sigma factors responding to the environmental stress/effects result in increase in the transcription of corresponding genes, results in adaptability of such bacteria to that particular environment. Several different types of sigma factors are identified and characterized across the bacteria. Among them σ^{70}(RpoD) / σ^{A} is the primary sigma factor which is also called as the "housekeeping" sigma factor responsible for the transcription of majority of the genes. They identify the consensus sequences –10 and –35 before the start point of the genes. The remaining sigma factors are called alternate sigma factors. A σ^{24} (RpoE) which is known as the extracytoplasmic/extreme heat stress sigma factor, σ^{28} (RpoF) the flagellar sigma factor,σ^{32} (RpoH) the heat shock sigma factor, $^{\sigma 38}$ (RpoS) - the starvation/stationary phase sigma factor, σ^{54} (RpoN) - the nitrogen-limitation sigma factor are well studied [1,2]. During a heat shock response sigma factors belonging to the RpoH family in bacterial RNA Polymerases usually control gene expression. It initiates the transcription of heat shock proteins like are chaperones, proteases and DNA-repair enzymes.

Anoxygenic photosynthetic bacteria possess two or more paralogs of RpoH, namely Rpo HI and Rpo HI. Rpo HI is associated with heat shock response, while RpoHII is associated with the response to the reactive oxygen species. Both of these are presented with a region C, a stretch of 23 amino acids [3,4]. Subsequent studies indicated that region C directly interacts with core RNA polymerase [5]. *R. sphaeroides* RpoHI is a member of the family of alternate sigma factors [6]. RpoH family of alternative sigma factors is characterized by a conserved amino acid sequence that is involved in RNA polymerase interactions [7,8]. RpoH family members have conserved amino acid

sequences in regions 2.4 and 4.2 that interact with promoter sequences [9]. Several alpha proteobacteria [10-15] possess two or more RpoH homologs suggesting that they may have roles in other stress responses.

EXPERIMENTAL SECTION

Retrieval of Rpo H sequences was done from UniProtKB/Swiss-Prot [16]. These sequences were used for further analysis. ExPASy's ProtParam tool was used for the computation of various physical and chemical parameters [17]. SOPMA tool (Self-Optimized Prediction Method with Alignment) server was used to characterize the secondary structural features [18]. Clustal Omega was used for multiple sequence alignment and percentage identity matrix [19].

Table 1: Different RNA polymerase sigma factor H from *Rhodobacter along with the sequence*

Sigma factor H	*Rhodobacter* Species /strain	Accession no.
Sgf1	Rhodobacter capsulatus Y262 > Sgf1 MSSYANLPAPSPEQGLNRYLQEIRKFPLLEPEEEYMLAKAWVDHQDPKAAHRLVTSHLRLAAKIAMGYRGYGLPQ AEVISEANVGLMQAVKRFDPERGFRLATYAMWWIRAAIQEYILRSWSLVKLGTTSAQKKLFFNLRKAKSKIGALDE GDLRPDAVAKIAHDLNVSEDDVIEMNRRLAGSDASLNAQVGASDGESATQWQDWLEDEDADQAEAYAEADELQA RRAMLVAAMDVLNERERDILMARRLRDEPVTLEELSSQYDVSRERIRQIEVRAFEKLQGRMKALAKEKGMSLPG	ETE 52832.1.
Sgf2	*Rhodobacter sphaeroides (strain KD131 / KCTC 12085* >Sgf2 MSTYTSLPAPSPEQGLNRYMQEIRKFPLLEPEEEYMLAKRWVDHQDNRAAHKLVTSHLRLAAKIAMGYRGYGLPQ AEVISEANVGLMQAVKRFDPEKGFRLATYAMWWIRASIQEYILRSWSLVKLGTTSAQKKLFFNLRKAKAKLGALEE GDLRPENVAQIAKDLGVSETEVIDMNRRLSGSDASLNATIGSDGEGSTQWQDWLEDEDSDQAADYAERDELEIRRE LLAQSMSVLNDREKDILVQRRLTDDPVTLEELSEGYGVSRERIRQIEVRAFEKLQAKMRELARSKGMTIPA	ACM 00585.1.
Sgf3	Rhodobacter sphaeroides (strain ATCC 17025 / ATH 2.4.3) >Sgf3 MSSYTSLPAPSPEQGLNRYMQEIRKFPLLEPEEEYMLAKRWVDHQDNRAAHRLVTSHLRLAAKIAMGYRGYGLPQ AEVISEANVGLMQAVKRFDPEKGFRLATYAMWWIRASIQEYILRSWSLVKLGTTSAQKKLFFNLRKAKAKLGALEE GDLRPENVAQIAKDLSVTEAEVIDMNRRLSASDASLNATIGSDGDGATQWQDWLEDEDSDQASDYAERDELEIRRE LLAQAMSVLNDREKDILVQRRLTDDPVTLEELSDGYGVSRERIRQIEVRAFEKLQARMRELARGKGMTIPA	ABP 70465.1
Sgf4	Rhodobacter sp. CACIA14H1 >Sgf4 MSTYQNLPAPSPEQGLNRYLQEIRKFPLLEPEQEYMLAKRWVDHQDAGAAHQLVTSHLRLAAKIAMGYRGYGLPQ AEVISEANVGLMQAVKRFDPEKGFRLATYAMWWIRASIQEYILRSWSLVKLGTTSAQKKLFFNLRKAKAKVGALE DGDLRPENVAQIAKDLSVSEDEVVEMNRRLAGSDASLNATVGSDGDSATQWQDWLEDEDSDQAGAYEERDELDA RRALLVQAMAVLNDREKDILMQRRLADTPVTLEELSESYGVSRERIRQIEVRAFEKLQSRMRELAKGKGMVIPA	ESW 60438.1.
Sgf5	Rhodobacter sp. SW2 > Sgf5 MSTYTNLPAPSPEQGLNRYMQEIRKFPLLEPEEEYMLAKRWVDHQDATAAHKMVTSHLRLAAKIAMGYRGYGLPQ AEVISEANVGLMQAVKRFDPEKGFRLATYAMWWIRASIQEYILRSWSLVKLGTTSAQKKLFFNLRKAKAKVGALEE GDLRPENVARIAHDLNVSETEVIDMNRRLSGGDASLNATVGSDGEGSTQWQDWLEDEDSDQANDYAERNELEMR RALLVQAMAVLNDREKDILMQRRLADEPVTLEDLSASYGVSRERIRQIEVRAFEKLQERMRELARGKGMAIPA	EEW 25714.1.
Sgf6	Rhodovulum sulfidophilum (Rhodobacter sulfidophilus) > Sgf6 MSNYANLPAPSPEQGLNRYLQAIRKFPLLEPEEEYMLAKSWVDHHDTEAAHRLVTSHLRLAAKIAMGYRGYGLPQ AEVISEANVGLMQAVKRFDPEKGFRLATYAMWWIRAAIQEYILRSWSLVKLGTTSAQKKLFFNLRKAKARIGALED GDLRPENVERIAHDLNVTETEVIAMNRRLSGGDASLNAMVGSDGDTTTEWQDWLEDEDANQAEAYAEKDELDSR RALLTEAMDVLNDREKDILMQRRLQEQPVTLEDLSTVYNVSRERIRQIEVRAFEKLQKRMKELASQKGLLAQA	BAQ 68451.1.

Table 2: Amino acid composition of different RNA polymerase sigma factor H from *Rhodobacter*

Sigma factor H / Amino acids	Sgf1	Sgf 2	Sgf 3	Sgf 4	Sgf 5	Sgf 6
Ala	40	29	32	34	33	34
Arg	26	26	28	25	27	25
Asn	8	8	8	8	11	12
Asp	20	18	20	19	17	18
Cys	0	0	0	0	0	0
Gln	15	15	15	18	14	15
Glu	28	30	28	27	29	29
Gly	15	17	16	17	17	14
His	4	3	3	3	4	5
Ile	12	14	14	11	12	12
leu	33	33	33	33	31	35
Lys	16	7	15	16	15	16
Met	11	10	10	10	13	10
Phe	6	6	6	6	6	6
Pro	11	10	10	10	10	9
Ser	18	21	20	19	17	15
Thr	6	12	11	9	11	13
Trp	6	6	6	6	6	6
Tyr	9	9	9	9	9	9
Val	15	14	14	18	16	15
Pyl	0	0	0	0	0	0
sec	0	0	0	0	0	0

Table 3: Physico chemical characteristics of different RNA polymerase sigma factor H from *Rhodobacter*

Sigma factor H	No of amino acids	Molecular weight	P I	-ve charged residues	+ve charged residues	Instability index	Aliphatic index	GRAVY
Sgf1	299	33862.4	5.50	48	42	53.13	86.62	-0.495
Sgf2	298	33980.5	5.54	48	43	56.77	84.87	-0.575
Sgf3	298	33976.5	5.53	48	43	50.97	85.87	-0.554
Sgf4	298	33743.3	5.51	46	41	50.95	86.51	-0.505
Sgf5	298	33961.6	5.78	46	42	51.78	82.92	-0.546
Sgf6	298	34062.6	5.59	47	41	50.64	87.52	-0.541

Table 4: Secondary structure analysis of different RNA polymerase sigma factor H from *Rhodobacter*

Sigma factor H	Alpha helix	310 helix	Pi helix	Beta bridge	Extended strand	Beta turn	Bend region	Random coil	Ambiguous state	Other states
Sgf1	64.88	0.00	0.00	0.00	6.69	6.69	0.00	21.74	0.00	0.00
Sgf2	57.72	0.00	0.00	0.00	10.07	7.72	0.00	24.50	0.00	0.00
Sgf3	60.74	0.00	0.00	0.00	10.07	7.38	0.00	21.81	0.00	0.00
Sgf4	59.73	0.00	0.00	0.00	9.06	8.05	0.00	23.15	0.00	0.00
Sgf5	58.05	0.00	0.00	0.00	11.07	6.38	0.00	24.50	0.00	0.00
Sgf6	66.11	0.00	0.00	0.00	6.71	6.38	0.00	20.81	0.00	0.00

Table 5: CLUSTAL O(1.2.1) multiple sequence alignment results for sigma factors of RpoH

Sgf1	MSSYANLPAPSPEQGLNRYLQEIRKFPLLEPEEEYMLAKAWVDHQDPKAAHRLVTSHLRL
Sgf6	MSNYANLPAPSPEQGLNRYLQAIRKFPLLEPEEEYMLAKSWVDHHDTEAAHRLVTSHLRL
Sgf4	MSTYQNLPAPSPEQGLNRYLQEIRKFPLLEPEQEYMLAKRWVDHQDAGAAHQLVTSHLRL
Sgf5	MSTYTNLPAPSPEQGLNRYMQEIRKFPLLEPEEEYMLAKRWVDHQDATAAHKMVTSHLRL
Sgf2	MSTYTSLPAPSPEQGLNRYMQEIRKFPLLEPEEEYMLAKRWVDHQDNRAAHKLVTSHLRL
Sgf3	MSSYTSLPAPSPEQGLNRYMQEIRKFPLLEPEEEYMLAKRWVDHQDNRAAHRLVTSHLRL
	.* .********:* **********:****** ****:* ***::*******
Sgf1	AAKIAMGYRGYGLPQAEVISEANVGLMQAVKRFDPERGFRLATYAMWWIRAAIQEYILRS
Sgf6	AAKIAMGYRGYGLPQAEVISEANVGLMQAVKRFDPEKGFRLATYAMWWIRAAIQEYILRS
Sgf4	AAKIAMGYRGYGLPQAEVISEANVGLMQAVKRFDPEKGFRLATYAMWWIRASIQEYILRS
Sgf5	AAKIAMGYRGYGLPQAEVISEANVGLMQAVKRFDPEKGFRLATYAMWWIRASIQEYILRS
Sgf2	AAKIAMGYRGYGLPQAEVISEANVGLMQAVKRFDPEKGFRLATYAMWWIRASIQEYILRS
Sgf3	AAKIAMGYRGYGLPQAEVISEANVGLMQAVKRFDPEKGFRLATYAMWWIRASIQEYILRS
	************************************:**************:********
Sgf1	WSLVKLGTTSAQKKLFFNLRKAKSKIGALDEGDLRPDAVAKIAHDLNVSEDDVIEMNRRL
Sgf6	WSLVKLGTTSAQKKLFFNLRKAKARIGALEDGDLRPENVERIAHDLNVTETEVIAMNRRL
Sgf4	WSLVKLGTTSAQKKLFFNLRKAKAKVGALEDGDLRPENVAQIAKDLSVSEDEVVEMNRRL
Sgf5	WSLVKLGTTSAQKKLFFNLRKAKAKVGALEEGDLRPENVARIAHDLNVSETEVIDMNRRL
Sgf2	WSLVKLGTTSAQKKLFFNLRKAKAKLGALEEGDLRPENVAQIAKDLGVSETEVIDMNRRL
Sgf3	WSLVKLGTTSAQKKLFFNLRKAKAKLGALEEGDLRPENVAQIAKDLSVTEAEVIDMNRRL
	**********************:::***::*****: * :**:** *:* :*: *****
Sgf1	AGSDASLNAQVGASDGESATQWQDWLEDEDADQAEAYAEADELQARRAMLVAAMDVLNER
Sgf6	SGGDASLNAMVG-SDGDTTTEWQDWLEDEDANQAEAYAEKDELDSRRALLTEAMDVLNDR
Sgf4	AGSDASLNATVG-SDGDSATQWQDWLEDEDSDQAGAYEERDELDARRALLVQAMAVLNDR
Sgf5	SGGDASLNATVG-SDGEGSTQWQDWLEDEDSDQANDYAERNELEMRRALLVQAMAVLNDR
Sgf2	SGSDASLNATIG-SDGEGSTQWQDWLEDEDSDQAADYAERDELEIRRELLAQSMSVLNDR
Sgf3	SASDASLNATIG-SDGDGATQWQDWLEDEDSDQASDYAERDELEIRRELLAQAMSVLNDR
	:..******* .* ***: .:*:.**********::** * * .**: ** :*. :* ***:*
Sgf1	ERDILMARRLRDEPVTLEELSSQYDVSRERIRQIEVRAFEKLQGRMKALAKEKGMSLPG
Sgf6	EKDILMQRRLQEQPVTLEDLSTVYNVSRERIRQIEVRAFEKLQKRMKELASQKGLLAQA
Sgf4	EKDILMQRRLADTPVTLEELSESYGVSRERIRQIEVRAFEKLQSRMRELAKGKGMVIPA
Sgf5	EKDILMQRRLADEPVTLEDLSASYGVSRERIRQIEVRAFEKLQERMRELARGKGMAIPA
Sgf2	EKDILVQRRLTDDPVTLEELSEGYGVSRERIRQIEVRAFEKLQAKMRELARSKGMTIPA
Sgf3	EKDILVQRRLTDDPVTLEELSDGYGVSRERIRQIEVRAFEKLQARMRELARGKGMTIPA
	*:***: *** : ******:** * ********************* :*: ** **: .

Table 6: Percent Identity Matrix as calculated by Clustal 2.1

100.00	82.55	84.90	83.22	81.21	81.88
82.55	100.00	83.22	84.23	81.21	82.21
84.90	83.22	100.00	89.93	87.92	88.59
83.22	84.23	89.93	100.00	90.94	89.60
81.21	81.21	87.92	90.94	100.00	95.64
81.88	82.21	88.59	89.60	95.64	100.00

RESULTS AND DISCUSSION

Sigma factors of RpoH which are known to play a crucial role in stress response. Sigma factors of different *Rhodobacter* species obtained from database are presented in Table 1. Table 2 shows that the amino acid composition of six different RpoH sigma factors of *Rhodobacter* species found in biological databases. The composition of alanine, leucine, glutamic acid and arginine was the highest while lowest concentrations of histidine, threonine and phenylalanine and tryptophan. Cysteine was not found in any of the Rpo H sigma factors. Tryptophan, tyrosine and phenylalanine were found to be constant throughout all the sigma factors which is very unique. There was no much variation in the content in proline and histidine of all the sigma factors. The number of negative charged residues is more compared to positively charged residues (Table 3). Molecular weights of all the sigma factors were almost same. The instability index of all the sigma factors was also same the highest being that of sigma factor sgf2. It was less than 40 showing that most of them are stable. Aliphatic index was found to span within a range of 82 to 87. The GRAVY value, calculated by adding the hydropathy values indicates the level of

hydrophobicity. Negative scores in this study show a lower hydrophobicity of the sigma factors. From Table 4, Secondary structural analysis of the sigma factors shows the dominance of α-helices and random coils equally for all the sigma factors. Beta turns were almost same except for sigma factor sgf4. The percent identity gives information about the percentage of comparison at a base-to-base or residue-to-residue level which includes sequence gaps and mismatches [20-22]. These *In silico* findings can be used for working on properties of sigma factors H in solution.

REFERENCES

[1] Tanja M. Gruber; Carol A. Gross, **2003** ,*Annu Rev Microbiol*, 57: 441-466
[2] Sharma UK; Chatterji D , **2010**,*FEMS Microbiol Rev*. **2010**,34(5):646-57.
[3] J.S. McCarty; S. Rudiger; H.J. Schonfeld; J. Schneider-Mergener; K. Nakahigashi; T. Yura; B. Bukau, **1996**, *J. Mol. Biol*, 256:829–837
[4] H. Nagai; H. Yuzawa; M. Kanemori; T. Yura, *Proc. Natl. Acad. Sci. USA*, **1994**, 91:10280–10284.
[5] D.M. Joo; A. Nolte; R. Calendar; Y. N. Zhou; D. J. Jin, *J. Bacteriol*,**1998**, 180,1095–1102.
[6] R.K. Karls; J. Brooks; P. Rossmeissl; J. Luedke; T. J. Donohue, *J. Bacteriol*, **1998**, 180,10–19.
[7] F. Arsene; T. Tomoyasu ; A. Mogk; C. Schirra; A.Schulze-Specking, *J Bacteriol*, **1999**, 181, 3552–3561.
[8] D.M. Joo; A. Nolte; R. Calendar;Y.N. Zhou ,D.J. Jin, *J Bacteriol*, **1998**, 180, 1095–1102.
[9] M. Wosten, FEMS Microbiology Reviews, **1998**, 22,127–150.
[10] M. Delory; R. Hallez; J.J. Letesson; X. De Bolle, *J Bacteriol*, **2006**, 188, 7707–7710.
[11] Y.Ono ; H. Mitsui; T. Sato ; K.Minamisawa, *Molecular and General Genetics*, **2001**, 264, 902–912.
[12] V.Oke; B.G. Rushing; E.J. Fisher;M. Moghadam Tabrizi; S.R. Long, *Microbiology-Sgm*, **2001**,147, 2399–2408.
[13] F. Narberhaus; P. Krummenacher; H.M.Fischer; H. Hennecke, *Molecular Microbiology*, **1997**, 24, 93–104.
[14] T. Kaneko; Y. Nakamura; S. Sato; K. Minamisawa ; T. Uchiumi, *DNA Res*, **2002**, 9, 189–197.
[15] J.M. Martinez Salazar; M. Sandoval Calderon ; X.W. Guo ; S. Castillo Ramirez; A. Reyes, *Microbiology-Sgm*, **2009**, 155, 386–397.
[16] Apweiler et al, *Fold Des*, **1996**;1, 3.
[17] E.Gasteiger et al, *Nucleic Acids Res*, **2003**, 31(13), 3784.
[18] C. Geourjon; G. Deleage, *Comput Appl Biosci*, **1995**, 11(6), 681.
[19] http://www.clustal.org/clustal2/
[20] SB Needleman; Wunsch CD, *J Mol Biol* **1970**, **48:**443-453.
[21] WR, Pearson; DJ Lipman, *Proc Natl Acad Sci* **1988**, **85:**2444-2448.
[22] EG Shpaer; M Robinson; D Yee; JD Candlin; R Mines; T Hunkapiller, *Genomics* **1996**, **38:**179-191.