Comparative Study of DNA Sequences for Genes Encoding Urease in *Klebsiella pneumoniae* and *Proteus mirabilis*

Mohammed Z. Saeed
Biology Department / College of Education
University of Mosul

Received 08 / 02 / 2012
Accepted 02 / 05 / 2012

الملخص

يتكون نظام اوبرون اليوريز من من سبعة مورثات هي, *ureD, ureC,ureB,ureA* *ureG, ureF, ureE* وان دليل P_2 في بكتريا *Proteus mirabilis* لهذه المورثات هو 0.429 ، 0.481 ، 0.409 ، 0.378 ، 0.472 ، 0.392 ، 0.567 على التوالي وهذه القيم اكبر مقارنة ببكتريا *Klebsiella pneumoniae* اذ بلغت لنفس المورثات 0.323 ، 0.369 ، 0.337 ، 0.226 ، 0.365 ، 0.276 ، 0.329. ان محاذاة تسلسلات القواعد للجنسين أظهرت تماثلا بالنسب 58.4%, 36.7%, 26.2%, 67%, 27.8%, 29.7%, 31.3% في حين كان تماثل الاحماض الامينية بالنسب اعلى 81% ، 62 % ، 72.6 %، 42%، 39.36% ، 42.3% ، 77% لنفس المورثات على التوالي. كانت الكودنات المرادفة الثلاثية والرباعية والسداسية النادرة الحدوث بنسب اقل في جرثومة *Pr. mirabilis* ولذا فان الاحماض الامينية الالنين والكلايسين واللوسين والثريونين والفالين بدون كودونات نادرة مقارنة بجرثومة *K. pneumoniae*.

الكلمات المفتاحية: اليوريز، الكودونات المستخدمة، التعبير الجيني

Abstract

Urease genes cluster consist of 7 genes *ureA, ureB, ureC, ureD, ureE, ureF* and *ureG*. P_2 index of *Pr. mirabilis* genes are 0.429, 0.481, 0.409, 0.378, 0.472, 0.392 and 0.567 for *ureA, ureB, ureC, ureD, ureE, ureF* and *ureG* respectively and these values are higher compared with 0.323, 0.369, 0.337, 0.226, 0.365, 0.276 and 0.329 for the same genes of *K. pneumoniae*.Alignment of nucleotide sequences of the two genera showed 58.4%, 36.7%, 26.2%, 67%, 27.8%, 29.7% and 31.3% similarity between *ureA, ureB, ureC, ureD, ureE, ureF* and *ureG* respectively,

while alignments of amino acids sequences results in higher values of 81%, 62%, 72.6%, 42%, 39.36%, 42.3% and 77% for the same genes respectively. The threefold, fourfold and sixfold synonymous codons have less incidence of rare codons in *Pr. mirabilis* so that Ala, Gly, Leu, Thr and Val have no rare codons in their synonymous codons as compared with those in *K. pneumoniae*.

Keywords: urease, codon usage, gene expression

Introduction

Urease (urea amidohydrolase; EC 3.5.1.5) catalyzes the hydrolysis of urea to yield ammonia and carbamate,which spontaneously hydrolyzes to form carbonic acid and a second molecule of ammonia (1).It is a nickel containing enzyme found in many microorganisms and DNA sequencing of the *Klebsiella aerogenes* urease genes (*ureD, ureE, ureF*, and *ureG*) showed that they were located immediately downstream of the three structural genes (*ureA, ureB and ureC*) (2).Studies of *K. aerogenes* showed that nickel is inserted into the apoprotein ureABC in a GTP-depended process that requires the action of the UreD, UreF and UreG and is facilitated by UreE (3).

The *Proteus mirabilis* gene cluster comprises eight contiguous genes; the structural genes *ureABC*, that encode the subunits of the enzyme and are flanked upstream by *ureD* and downstream by *ureEFG* (4).Quantitative variation in urease production among enteric organisms led to the development of biochemical media and tests which could distinguish the rapidly ureolytic *Proteus* spp. from those which hydrolyzes urea more slowly (certain *Klebsiella, Enterobacter, Citrobacter, Yersinia, and Serratia* spp.). The ability to hydrolyse urea with subsequent alkalinization of the urine is a virulence factor that *Pr. mirabilis* shares with other uropathogenic representatives of the family Enterobacteriaceae (5).

Plasmid-encoded urease gene clusters in uropathogenic isolates of *E. coli*, *P. stuartii* and *Salmonella cubana* demonstrated DNA homology, similar positions of restriction endonuclase cleavage sites and manners of urease expression and these genes are closely related to *Pr. mirabilis* urease genes The genetic organization of bacterial urease operons is similar to that of *Pr. mirabilis* and transcribed as polycistronic mRNAs and proceeding these structural genes are the regulatory genes (6). The *K. pneumoniae* urease operon is complex also encoding at least six polypeptides (6,1). *ureA* encode a peptide of 11.1 kilodaltons (KDa), *ureB*(11.7 KDa), *ureC* (60.3 KDa), *ureD*(29.8 KDa), *ureE*(17.6 KDa), *ureG*(21.9 KDa) (7,2).

Materials and Methods

Urease genes of *Klebsiella pneumoniae* MGH78578 and of *Pr. mirabilis* HI4320 were obtained from xbase, a genome database available at: http://xbase.bham.ac.uk (8).

1. Codon usage and relative synonymous codon usage of individual genes was calculated using codonW, available at: http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms, a software developed by Peden(9). Relative synonymous codon usage (RSCU) is defined as the ratio of the observed frequency of a codon to the expected frequency of the same codon within the synonymous group if all synonyms are used equally. RSCU values greater than 1.0 indicate that the corresponding codon is used more frequently than the expected frequency whereas the reverse is true for RSCU values less than 1.0 (10).
2. P_2 index was determined for each gene according to Grosjean and Fiers (11), P_2 = (WWC+SSU)/(WWY+SSY), where W=A or U,S=G or C,Y=C or U, and (for example) WWC is the observed number of codons of that description.
3. Alignments of DNA and protein sequences were determined by ClaustalW, a program developed by Thompson (12).
4. Rare codons of the overall gene cluster were computed as those occurred in a rate less than 10% according to Rosano & Ceccarelli (13).

Results and Discussion

1- Overall base content:

Urease genes cluster of *K. pneumoniae* consist of 7 genes *ureA*(101 codons), *ureB* (107 codons), *ureC* (568 codons), *ureD* (275 codons), *ureE* (159 codons), *ureF* (225 codons) and *ureG*(206 codons) whereas codon content is 101,109,568,275,162,223,206 for the same genes of *Pr. mirabilis* respectively.Table (1) shows the overall base content of both clusters.

Table (1): The overall base content

Base	First base		Second base		Third base	
	K. pnemoniae	*Pr. mirabilis*	*K. pnemoniae*	*Pr. mirabilis*	*K. pnemoniae*	*Pr. mirabilis*
A	352	425	435	490	133	458
G	632	611	334	296	544	321
C	469	322	401	368	738	223
T	188	286	471	490	226	642
A+T	540	711	906	980	359	1100
G+C	1101	933	735	664	1282	544

2- Codon usage:

Tables (2) and (3) show the codon usage and RSCU values of the genes for both bacteria respectively. Comparing the values of the threefold, fourfold and sixfold synonymous codons of each amino acid, the following codons had the highest values in *K. pneumoniae*; CUG, AUC, GUG, AGC, CCG, ACC, GCC, CGC,GGC whereas UUA, AUU, GUU, UCA, CCU, ACC, GCU, CGU, GGU in *Pr. mirabilis* for the amino acids Leu, Ile, Val, Ser, Pro, Thr, Ala, Arg, and Gly respectively. Moreover, the codons UUA,UUG, AUU, AGU, ACA, GCU of *Pr. mirabilis* were higher than those of *K. pneumoniae* in all the genes of the cluster. Most amino acids can be encoded by more than one codon and such codons are called synonymous and usually differ in the third or second nucleotide position. These synonymous codons are not used with equal frequencies; their usage is different in different genomes and even within the same genome (14). Gene sequences from *E. coli, Salmonella typhimurium* and other members of Enterobacteriaceae show a negative correlation between the degree of synonymous codon usage bias and rate of nucleotide substitution at synonymous site (15) and translation selection was thought to be the main factor accounting for codon usage variation among the genes (16). Codon selection was first elucidated in *E. coli* where genes expressed at high levels exhibit a marked bias toward a particular subset of codons and the preferred ones are those that are best recognized by the most abundant tRNA species (17).

Table (2): Codon usage of *K. pneumoniae* MGH78578

Amino acid	codons	*ureA*		*ureB*		*ureC*		*ureD*		*ureE*		*ureF*		*ureG*	
		N	RSCU	N	RSCU	N	RSCU	N	RSCU	N	RSCU	N	RSCU	N	RSCU
Phe	UUU	2	1.33	1	0.33	5	0.71	6	1.20	3	1.20	3	0.86	0	0.00
	UUC	1	0.67	5	1.67	9	1.29	4	0.80	2	0.80	4	1.14	6	2.00
Leu	UUA	0	0.00	0	0.00	1	0.15	1	0.12	1	0.30	2	0.46	0	0.00
	UUG	1	0.50	0	0.00	3	0.44	3	0.38	0	0.00	0	0.00	0	0.00
	CUU	0	0.00	0	0.00	1	0.15	6	0.75	1	0.30	1	0.23	3	0.86
	CUC	2	1.00	0	0.00	9	1.32	13	1.62	3	0.90	4	0.92	4	1.14
	CUA	0	0.00	0	0.00	0	0.00	1	0.12	0	0.00	0	0.00	0	0.00
	CUG	9	4.50	5	6.00	27	3.95	24	3.00	15	4.50	19	4.38	14	4.00
Ile	AUU	2	1.00	1	0.75	9	0.61	3	1.29	2	1.50	2	1.00	1	0.25
	AUC	4	2.00	2	1.50	35	2.39	3	1.29	2	1.50	4	2.00	11	0.75
	AUA	0	0.00	1	0.75	0	0.00	1	0.34	0	0.00	0	0.00	0	0.00
Met	AUG	5	1.00	2	1.00	16	1.00	2	1.00	4	1.00	7	1.00	6	1.00
Val	GUU	1	0.44	4	1.33	5	0.43	2	0.47	3	0.86	1	0.44	0	0.00
	GUC	4	1.78	4	1.33	16	1.36	6	1.41	2	0.57	3	1.33	6	1.71
	GUA	0	0.00	2	0.67	0	0.00	3	0.71	1	0.29	1	0.44	0	0.00
	GUG	4	1.78	2	0.67	26	2.21	6	1.41	8	2.29	4	1.78	8	2.29
Ser	UCU	0	0.00	0	0.00	2	0.52	1	0.43	0	0.00	2	0.80	1	0.55

	UCC	1	1.20	0	0.00	6	1.57	2	0.86	0	0.00	3	1.20	2	1.09
	UCA	0	0.00	0	0.00	3	0.78	0	0.00	0	0.00	0	0.00	1	0.55
	UCG	2	2.40	1	6.00	5	1.35	0	0.00	1	0.75	2	0.80	1	0.55
Pro	CCU	0	0.00	1	0.57	1	0.14	2	0.40	1	0.50	0	0.00	1	0.44
	CCC	1	0.67	3	1.71	4	0.57	6	1.20	2	1.00	3	1.00	1	0.44
	CCA	0	0.00	0	0.00	3	0.34	3	0.60	0	0.00	0	0.00	2	0.89
	CCG	5	3.33	3	1.71	20	2.86	9	1.80	5	2.50	9	3.00	5	2.22
Thr	ACU	0	0.00	0	0.00	0	0.00	3	0.60	1	0.57	1	0.50	1	0.27
	ACC	5	4.00	2	2.67	38	3.71	13	2.60	4	2.29	4	2.00	13	3.47
	ACA	0	0.00	0	0.00	0	0.00	1	0.20	0	0.00	1	0.50	0	0.00
	ACG	0	0.00	1	1.33	3	0.29	3	0.60	2	1.14	2	1.00	1	0.27
Ala	GCU	1	0.44	0	0.00	2	0.12	1	0.17	4	1.23	2	0.24	1	0.21
	GCC	6	2.67	7	2.55	39	2.29	15	2.61	7	2.15	24	2.91	7	1.47
	GCA	0	0.00	1	0.36	3	0.18	1	0.17	0	0.00	0	0.00	1	0.21
	GCG	2	0.89	3	1.09	24	1.41	6	1.04	2	0.62	7	0.85	10	2.11
Tyr	UAU	1	2.00	2	1.33	4	0.89	2	0.80	1	0.50	2	0.67	3	1.50
	UAC	0	0.00	1	0.67	5	1.11	3	1.20	3	1.50	4	1.33	1	0.50
TER	UAA	0	0.00	1	3.00	1	3.00	1	3.00	0	0.00	0	0.00	0	0.00
	UAG	0	0.00	0	0.00	0	0.00	0	0.00	1	3.00	1	3.00	1	3.00
His	CAU	0	0.00	1	0.40	9	0.86	3	0.60	6	0.80	1	1.00	1	1.00
	CAC	2	2.00	4	1.60	12	1.14	7	1.40	9	1.20	1	1.00	1	1.00
Gln	CAA	0	0.00	0	0.00	2	0.21	2	0.31	1	0.40	4	0.50	0	0.00
	CAG	2	2.00	5	2.00	17	1.79	11	1.69	4	1.60	12	1.50	4	2.00
Asn	AAU	0	0.00	2	0.88	6	0.67	2	1.33	0	0.00	0	0.00	2	0.57
	AAC	2	2.00	3	1.20	12	1.33	1	0.67	3	2.00	2	2.00	5	1.43
Lys	AAA	2	0.80	1	0.67	9	1.06	3	1.20	1	0.67	0	0.00	9	1.50
	AAG	3	1.20	2	1.33	8	0.94	2	0.80	2	1.33	1	2.00	3	0.50
Asp	GAU	1	0.50	2	1.33	15	0.79	7	1.08	7	1.40	2	0.50	9	1.29
	GAC	3	1.50	1	0.67	23	1.21	6	0.92	3	0.60	6	1.50	5	0.71
Glu	GAA	6	1.09	2	0.44	14	0.97	4	0.67	5	1.00	8	1.07	11	1.57
	GAG	5	0.91	7	1.56	15	1.03	8	1.33	5	1.00	7	0.93	3	0.43
Cys	UGU	0	0.00	1	2.00	3	0.75	1	0.33	0	0.00	2	0.80	1	1.00
	UGC	0	0.00	0	0.00	5	1.25	5	1.67	2	2.00	3	1.20	1	1.00
TER	UGA	1	3.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Trp	UGG	0	0.00	0	0.00	7	1.00	7	1.00	0	0.00	10	1.00	2	1.00
Arg	CGU	2	1.71	1	0.67	1	0.25	2	0.71	2	1.20	5	1.36	1	0.67
	CGC	3	2.57	6	4.00	18	4.50	10	3.53	6	3.60	11	3.00	8	5.33
	CGA	1	0.86	0	0.00	1	0.25	1	0.35	0	0.00	2	0.55	0	0.00
	CGG	1	0.86	2	1.33	4	1.00	3	1.06	2	1.20	4	1.09	0	0.00
Ser	AGU	0	0.00	0	0.00	1	0.26	1	0.43	0	0.00	0	0.00	1	0.55
	AGC	2	2.40	0	0.00	6	1.57	10	4.29	7	5.25	8	3.20	5	2.73
Arg	AGA	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
	AGG	0	0.00	0	0.00	0	0.00	1	0.35	0	0.00	0	0.00	0	0.00
Gly	GGU	0	0.00	4	1.33	1	0.07	3	0.55	1	0.31	1	0.29	3	0.55
	GGC	6	4.00	8	2.67	42	3.05	16	2.91	11	3.38	9	2.57	16	2.91
	GGA	0	0.00	0	0.00	7	0.51	1	0.18	0	0.00	1	0.29	1	0.18
	GGG	0	0.00	0	0.00	5	0.36	2	0.36	1	0.31	3	0.86	2	0.36

*N is the number of codon

** RSCU is the relative synonymous codon usage

Table (3): Codon usage of *Pr. mirabilis* HI4320

Amino acid	codons	*ureA*		*ureB*		*ureC*		*ureD*		*ureE*		*ureF*		*ureG*	
		N	RSCU	N	RSCU	N	RSCU	N	RSCU	N	RSCU	N	RSCU	N	RSCU
Phe	UUU	1	1.00	5	1.67	11	1.47	13	2.00	3	2.00	7	2.00	4	2.00
	UUC	1	1.00	1	0.33	4	0.53	0	0.00	0	0.00	0	0.00	0	0.00
Leu	UUA	6	3.00	4	3.43	16	2.91	20	4.00	8	2.40	16	3.20	7	2.21
	UUG	2	1.00	1	0.86	3	0.55	3	0.60	3	0.90	4	0.80	4	1.26
	CUU	3	1.50	0	0.00	6	1.09	3	0.60	5	1.50	4	0.80	4	1.26
	CUC	0	0.00	1	0.86	2	0.36	1	0.20	0	0.00	2	0.40	0	0.00
	CUA	0	0.00	0	0.00	2	0.36	3	0.60	2	0.60	2	0.40	1	0.32
	CUG	1	0.50	1	0.86	4	0.73	0	0.00	2	0.60	2	0.40	3	0.95
Ile	AUU	4	2.40	4	2.00	33	1.83	8	1.85	4	2.40	4	1.30	10	1.67
	AUC	0	0.00	1	0.50	15	0.83	5	1.15	0	0.00	4	1.30	7	1.17
	AUA	1	0.60	1	0.50	6	0.33	0	0.00	1	0.60	1	0.33	1	0.17
Met	AUG	5	1.00	3	1.00	21	1.00	8	1.00	3	1.00	7	1.00	7	1.00
Val	GUU	5	2.00	2	0.89	13	1.13	9	2.00	4	1.33	4	1.23	10	2.35
	GUC	1	0.40	1	0.44	12	1.04	2	0.44	1	0.33	1	0.31	1	0.24
	GUA	2	0.80	1	0.44	7	0.61	4	0.89	4	1.33	3	0.92	2	0.47
	GUG	2	0.80	5	2.22	14	1.22	3	0.67	3	1.00	5	1.54	4	0.94
Ser	UCU	0	0.00	1	2.00	4	1.20	1	0.75	1	0.75	3	1.06	1	0.75
	UCC	0	0.00	0	0.00	0	0.00	1	0.75	1	0.75	0	0.00	0	0.00
	UCA	2	3.00	0	0.00	6	1.80	3	2.25	3	2.25	3	1.06	3	2.25
	UCG	0	0.00	0	0.00	5	1.50	2	1.50	0	0.00	1	0.35	0	0.00
Pro	CCU	2	1.60	2	2.00	8	1.03	9	2.00	2	1.33	2	1.00	7	3.11
	CCC	1	0.80	2	2.00	6	0.77	3	0.67	0	0.00	2	1.00	0	0.00
	CCA	2	1.60	0	0.00	12	1.55	3	0.67	3	2.00	3	1.50	1	0.44
	CCG	0	0.00	0	0.00	5	0.65	3	0.67	1	0.67	1	0.50	1	0.44
Thr	ACU	3	1.71	1	1.00	7	0.78	3	0.60	1	0.40	3	0.86	3	1.00
	ACC	2	1.14	1	1.00	15	1.67	4	0.80	6	2.40	3	0.86	2	0.67
	ACA	2	1.14	2	2.00	9	1.00	7	1.40	1	0.40	3	0.86	5	1.67
	ACG	0	0.00	0	0.00	5	0.56	6	1.20	2	0.80	5	1.43	2	0.67
Ala	GCU	3	1.33	5	2.00	15	1.13	4	0.80	6	2.40	8	1.45	7	1.47
	GCC	2	0.89	0	0.00	9	0.68	3	0.60	1	0.40	6	1.09	2	0.42
	GCA	4	1.78	5	2.00	17	1.28	8	1.60	0	0.00	6	1.09	3	0.42
	GCG	0	0.00	0	0.00	12	0.91	5	1.00	3	1.20	2	0.36	7	1.47
Tyr	UAU	0	0.00	2	1.33	12	1.60	9	1.80	4	1.60	5	2.00	5	1.67
	UAC	1	2.00	1	0.67	3	0.43	1	0.20	1	0.40	0	0.00	1	0.33
TER	UAA	0	0.00	0	0.00	1	3.00	1	3.00	1	3.00	0	0.00	1	3.00
	UAG	1	3.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
His	CAU	0	0.00	3	1.50	15	1.58	7	1.56	5	0.83	4	2.00	3	2.00
	CAC	1	2.00	1	0.50	4	0.42	2	0.44	7	1.17	0	0.00	0	0.00
Gln	CAA	3	2.00	2	1.33	13	1.53	12	1.50	6	1.71	14	1.56	4	1.60
	CAG	0	0.00	1	0.67	4	0.47	4	0.50	1	0.29	4	0.44	1	0.40
Asn	AAU	1	2.00	5	2.00	16	1.68	6	2.00	0	0.00	1	2.00	3	0.86
	AAC	0	0.00	0	0.00	3	0.32	0	0.00	1	2.00	0	0.00	4	1.14
Lys	AAA	6	1.71	7	2.00	25	1.85	6	1.09	8	1.60	6	1.50	9	1.50
	AAG	1	0.29	0	0.00	2	0.15	5	0.91	2	0.40	2	0.50	3	0.50
Asp	GAU	3	2.00	2	2.00	29	1.76	6	1.50	9	1.64	7	1.56	15	2.00
	GAC	0	0.00	0	0.00	4	0.24	2	0.50	2	0.36	2	0.44	0	0.00
Glu	GAA	7	1.27	4	0.73	22	1.33	6	0.63	7	1.17	8	1.07	11	1.47
	GAG	4	0.73	7	1.27	11	0.67	13	1.37	5	0.83	7	0.93	4	0.53

Cys	UGU	0	0.00	0	0.00	8	1.78	4	1.60	3	2.00	3	1.50	2	2.00
	UGC	2	2.00	0	0.00	1	0.22	1	0.40	0	0.00	1	0.50	0	0.00
TER	UGA	0	0.00	1	3.00	0	0.00	0	0.00	0	0.00	1	3.00	0	0.00
Trp	UGG	0	0.00	0	0.00	5	1.00	7	1.00	1	1.00	8	1.00	0	0.00
Arg	CGU	2	2.40	2	1.33	11	3.30	4	1.85	2	1.71	3	1.64	5	3.00
	CGC	0	0.00	3	2.00	5	1.50	2	0.92	3	2.57	3	1.64	3	1.80
	CGA	1	1.20	1	0.67	4	1.20	3	1.38	2	1.71	1	0.55	0	0.00
	CGG	0	0.00	0	0.00	0	0.00	2	0.92	0	0.00	3	1.00	1	0.60
Ser	AGU	2	3.00	2	4.00	2	0.60	1	0.75	3	2.25	6	2.12	3	2.25
	AGC	0	0.00	0	0.00	3	0.90	0	0.00	0	0.00	4	1.41	1	0.75
Arg	AGA	2	2.40	2	1.33	0	0.00	2	0.92	0	0.00	0	0.00	1	0.60
	AGG	0	0.00	1	0.67	0	0.00	0	0.00	0	0.00	1	0.55	0	0.00
Gly	GGU	0	0.00	7	2.33	29	1.90	8	1.45	5	1.33	3	1.00	7	1.65
	GGC	3	1.71	4	1.33	18	1.18	7	1.27	4	1.07	3	1.00	2	0.47
	GGA	2	1.14	1	0.33	10	0.66	2	0.36	0	0.00	1	0.55	5	1.18
	GGG	2	1.14	0	0.00	4	0.26	5	0.91	6	1.60	5	1.67	3	0.71

*N is the number of codon
** RSCU is the relative synonymous codon usage

P_2 index of *Pr. mirabilis* genes was 0.429,0.481,0.409, 0.378, 0.472,0.392 and 0.567 for *ureA, ureB, ureC, ureD, ureE, ureF* and *ureG* respectively and these values were higher compared with 0.323, 0.369, 0.337, 0.226, 0.365, 0.276 and 0.329 for the same genes of *K. pneumoniae.* It was suggested that among codons which are either A/T or G/C rich at positions one and two, and which end in a pyrimidine, there is a selection for that base (T or C) which yields the less extreme overall codon-anticodon interaction energy and the P_2 index is devised of this preference, calculated as the frequency of those codons to be preferred (18). Genes thought to be highly expressed and lowly expressed differ in their extent of codon bias, with the bias being more extreme in highly expressed genes (19). Sharp and Li(15) studied 165 *E. coli* genes which were categorized as 27 very highly expressed genes,15 highly expressed genes, 57 moderately low expressed and 58 very lowly expressed genes and the remaining 8 regulatory/repressor genes which were reported to have an extraordinary high frequency of rare codons. They indicated that that values of P_2 index for the highly expressed genes were of the range (0.52-0.82) while lowly expressed genes were of the range (0.3-0.76).

Synonymous codon usage has basically proven to be non-random and species-specific. Several factors including directional mutational bias, translational selection, secondary protein structure, replicational and transcriptional selection and environmental factors have been shown to influence codon usage in a variety of microorganisms (20). Codon bias is now recognized as crucial factor in shaping gene expression and cellular function through its effects on diverse processes, ranging from RNA processing to protein folding. In organisms, including *E. coli, Saccharomyces cerevisiae, Caenorhabditis elegans, Arabidopsis*

thaliana, Drosophila melanogaster and possibly mammals there is a significant positive correlation between a gene's expression and the degree of its codon bias (21). It has been shown that within genomes, highly expressed genes have shifted their codon usage toward a more restricted set of preferred synonymous codons than other less highly expressed genes. In many cases it has been shown that codon usage mirrors the distribution of tRNA abundances indicating that the preferred codons are those that tend to match more abundant anticodons(22).

3- Alignment of sequences:

Alignment of nucleotide sequences yield 58.4%, 36.7%, 26.2%, 67%, 27.8%, 29.7% and 31.3% similarity between *ureA, ureB, ureC, ureD, ureE, ureF* and *ureG* respectively while alignments of amino acids sequences have higher values of 81%, 62%, 72.6%, 42%, 39.36%, 42.3% and 77% for the same genes respectively.

Jones and Mobley (23) indicated that the genes sequences encoding *K. pneumoniae* urease subunits are 72%, 71% and 58% identical to recently reported sequences encoding urease subunits from *Pr. mirabilis, Pr. vulgaris* and *Helicobacter pylori* respectively. 42% of the *K. aerogenes* amino acid residues are present in all three bacterial ureases.In addition, *ureE* and *ureF* genes of *Pr. mirabilis* were 53% and 38% identical to the *K. areogenes* genes. D'Orazio and Collins (6) indicated that the urease structural genes are highly conserved among the members of the family Enterobacteriaceae, so the predicted amino acid sequences of the urease structural subunits of the plasmid –encoded urease gene clusters found in uropathogenic *E. coli* share significant degree of identity. The nucleotide sequences of *ureA, ureB, ureC, ureD, ureE, ureF* were 81%, 76%, 82%, 83%, 83% and 87% identical to those to *Pr. mirabilis* respectively and 69%, 63%, 62%, 50%, 60%, 57% identical to *K. aerogenes* respectively. By contrast amino acid sequences were 94%, 87%, 100%, 73%, 95% and 55% identical to the *Pr. mirabilis* genes respectively and 83%, 63%, 76%, 40%, 64% and 55% identical to *K. aerogenes* genes.

4- Rare codons:

Table (4) shows the codons used rarely that represent less than 10% of the codons chosen from a synonymous codon groups.

Table (4): Rare codons occurring in the two genes clusters of in the two bacterial species

K. pneumoniae (1641)*		*Pr. mirabilis*(1644)*	
Amino acids	codons	Amino acids	codons
Arg	AGA,AGG,CGA,CGG,CGT	Arg	AGA,AGG,CGA,CGG
Ser	AGT,TCA,TCC,TCG,TCT	Ser	AGC,TTC,TCG,TCT
Asn	AAT	Asn	AAC
Ile	ATA	Ile	ATA
Cys	TGT,TGC	Cys	TGC
Pro	CCA,CCT	Pro	CCC,CCG
Gln	CAA	Gln	CAG
Thr	ACA,ACG,ACT,	Lys	AAG
Gly	GGA,GGG,GGT	Asp	GAC
Val	GTA,GTT	Phe	TTC
Leu	CTA,CTT,TTA,TTG	His	CAC
Ala	GCA,GCT		

*Total number of codons in the gene cluster

Generally, the threefold,fourfold and sixfold synonymous codons have less incidence of rare codons in *Pr. mirabilis* so that Ala, Gly, Leu, Thr,Val have no rarely used codons in their synonymous codons as compared with those in *K. pneumoniae*.The frequency of rare codons is higher in rarely transcribed genes and this is often ascribed to adaptive pressures modulating gene expression e.g the low level of expression of *dnaG*, which is cotranscribed with the highly expressed *rpsU* and *rpoD* genes, was attributed to its higher frequency of rare codons (9). This is supported by the finding that genes with a high proportion of rare codons are translated at lower elongation rate than those with a low proportion.Insertion of tandem rare codons into homologous genes has also been shown to increase the translation time and to lower expression levels(24).Chen and Inouye(25) proposed that the rate of translation at these codons decreases as result of limited availability of charged tRNAs. Genes that are highly expressed generally tend to have codons recognized by relatively abundant tRNA substituting rare codons with synonymous codons reduces the total time of translating a mRNA sequence(26). In *E. coli* there are about 30 low usage codons but only 20 of them are determined to be the bacterial rare codons including 7(AGG,AGA,CGA,CUA,AUA,CCC and CGG) used at frequency less than 0.5% and 13 (ACA, CCU, UCA, GGA, AGU, UCG, CCA, UCC, GGG, CUC, CUU, UCU and UUA) used at frequency higher than 0.5% (27).

Conclusion

The relative synonymous codon usage, P_2 index and the rare codons differences between the organisms may explain the relative higher rate of

urea hydrolysis (highly expressed gene cluster) of *Pr. mirabilis* which takes about 4 hours as compared with the 24 hours in *K. pneumoniae*(28).

References

1) Mobley H.L.T. and Hausinger R.P., Microbiol. Rev., 53(1):85-108 (1989).

2) Lee M.H., Mulrooney S.B., Renner M.J., Markowicz Y. and Hausinger R.P., J.Bacteriol., 174(13):4324-4330(1992).

3) Mulrooney S.B., Ward S.K. and Hausinger R.P., J.Bacteriol., 187 (10): 3581-3585(2005).

4) Island M.D. and Mobley H.L.T., J. Bacteriol., 177(3):1027-1033(1995).

5) Walz S.E., Wray S.K., Hull Sh.I. and Hull R.A., J. Bacteriol., 170 (3):1027-1033(1988).

6) D'Orazio S.E.F., Collins C.M., J. Bacteriol., 175(6):1860-1864 (1993).

7) Mulrooney S.B. and Hausinger R.P., J.Bacteriol., 172(10):5837-5843 (1990).

8) Chaudhuri R.R., Loman N.J., Synder L.A.S., Bailey C.M., Stekel D.J. and Pallen, M.J., Nucleic Acids Res., 36:D543-D546 (2007).

9) Peden J.F., Ph.D thesis, Department of Genetics, University of Nottingham, Nottingham, UK (1999).

10) Gupta S.K., Bhattacharyya T.K., and Ghosh T.C., J. Biomol. Str. Dynamics, 21(4):1-9(2004).

11) Grosjean H. and Fiers W., Gene, 18(3); 199-209(1982).

12) Thompson J.D., Higgins D.G.; and Gibson T.J., Nucleic Acids Res., 22 (22), 4673-4680(1994).

13) Rosano G.L. and Ceccarelli E.A., Microbial cell Factories, 8(41):1-9 (2009).

14) Ermolaeva M.D., Curr. Issues. Mol. Biol. 3(4):91-97(2001).

15) Sharp P.M. and Li W-H., Nucleic acids Res., 14(19):7737-7749 (1986).

16) Gu W, Zhou T., Ma J., Sun X. and Lu Z., Biosystems, 37:89-97 (2004).

17) Grocock R.J. and Sharp P.M., Gene, 289:131-139 (2002).

18) Shields D.C. and Sharp P.M., Nucleic Acids Res., 15(19):8023-8040.

19) Sharp P.M., Tuohy Th.M.F. and Mosurski K.R., Nucleic Acids Res., 14(3):5125-5143 (1986).

20) Sau K., Gupta S.K., Sau S., Mandal,S.Ch. and Ghosh T.Ch., J. Microbiol., 45(1):58-63(2007).

21) Gustafsson C., Govindarajan S. and Minshull J., Trends Biotechnol., 22:346-353 (2004)

22) Lynn D.J., Singer G.A.C. and Hickey D.A., Nucleic Acids Res., 30(19):4272-4277 (2002).

23) Jones B.D. and Mobley H.L.T., J. Bacteriol., 171(12):6414-6422 (1989).

24) Makoff A.J., Oxer M.D., Romanos M.A., Fairweather N.F. and Ballantine S., Nucleic Acids Res., 17(24):10191-10202(1989).

25) Chen G.T. and Inouye M. Genes and Development, 8:2641-2652 (1994).

26) Ponnala L., Bioinformation, 4(2):63-65(2009).

27) Chen D. and Texada D.F., Gene therapy Mol.Biol.,10:1-2(2006).

28) Koneman, E.W.; Allen, S.D.; Janda, W.M., Schreckenberger, P.C. and Winn, W.C. "Colour Atlas and Textbook of Diagnostic Microbiology". 5thed., Lippincott-Raven Publishers, Philadelphia, USA, pp.212-228 (1997).