大肠杆菌和志贺氏菌 O-抗原糖基转移酶与聚合酶的生物信息学研究

陶 江[1] 刘 斌[1] 王 荃[1] 郭宏杰[1,2,3] 冯 露[1,2,3*]

([1] 南开大学生命科学学院 天津 300071)

([2] 南开大学泰达生物技术学院 天津 300457)

([3] 天津市微生物功能基因组学重点实验室 天津 300457)

摘 要 利用生物信息学手段对大肠杆菌和志贺氏菌的 110 个 O-抗原糖基转移酶与 39 个 O-抗原聚合酶的序列进行分析,探讨这两种酶的序列和结构特点。统计了其序列一致性,密码子使用和(G + C)% 含量的特点;讨论了 O-抗原糖基转移酶和聚合酶对底物的特异性;推测了 6 组糖基转移酶的功能;通过对蛋白拓扑结构的预测,发现 O-抗原聚合酶中广泛存在一个位于细胞周质中的亲水环(Loop),是可能的功能区域;通过对蛋白高级结构的预测,发现 O-抗原糖基转移酶属于两个不同的蛋白超家族。

关键词 大肠杆菌,志贺氏菌,O-抗原,糖基转移酶,聚合酶,生物信息学

中图分类号:Q93 **文献标识码**:A **文章编号**:0001-6209(2004)03-0345-06

大肠杆菌又名大肠埃希氏菌(*Escherichia coli*),一些特定的 O-抗原血清型可以引起人类的腹泻与尿道感染,严重时导致肾衰或死亡。志贺氏菌(*Shigella*)是随着人类进化而发展起来的重要致病菌,引起人的细菌性痢疾。埃希氏菌属与志贺氏菌属在进化上的地位十分接近,近年来发现二者有许多共同的生化和分子遗传学特征,因此被建议重新划分成一个属[1~4]。

O-抗原是革兰氏阴性细菌最主要的表面抗原。由于长期受到宿主的选择压力,大肠杆菌有 166 种不同的 O-抗原,志贺氏菌有 33 种不同的 O-抗原,其中 13 种二者共有[2,5,6]。合成 O-抗原的基因是研究分子进化的最好的材料之一。

O-抗原的合成过程如下:首先由糖基转移酶(Glycosyltransferase,GT)将对应的核苷二磷酸单糖依次转移到一个固定于细胞内膜的脂分子上,在胞质内合成寡糖单位,然后通过转运酶(O-unit flippase,Wzx)转移到内膜外侧的细胞周质中,在聚合酶(O-antigen polymerase,Wzy)的作用下形成寡糖单位间的二糖连接键,聚合成由相同寡糖单位组成的多糖,最后 O-抗原连接到一个糖脂分子上成为脂多糖[7]。

与 O-抗原合成有关的基因在染色体上成簇排列,O-抗原基因簇含有 3 类基因:单糖合成酶基因,糖基转移酶基因和寡糖单位处理酶基因(包括转运酶基因和聚合酶基因)[7]。O-抗原基因簇中的基因决定了 O-抗原的组成和多样性。在参与 O-抗原合成的 3 类酶中,单糖合成酶较保守,可以通过序列比对预测其功能。而糖基转移酶和寡糖单位处理酶有特异的供体、受体和连接键,序列间的一致性很低,仅通过序列比对不能准确预测其功能,且这类酶的生物学鉴定非常困难。转运酶有其特有的基序和蛋白拓扑结构[8,9];聚合酶蛋白拓扑结构也有其特点,但没有发现基序[10];糖基转移酶可以分成不同家族,但不能进一步区分其功能(http://afmb.cnrs-mrs.fr/CAZY)。

本实验室最新测定了大量大肠杆菌和志贺氏菌 O-抗原基因簇序列,本文运用生物信息学手段,以其中的糖基转移酶和聚合酶基因序列为材料,并加入 GenBank 中已有的数据,从多角度对这些序列的信息进行分析,为阐明其生物学功能提供基础。

1 材料和方法

1.1 研究用的基因序列

使用本实验室测定的和 GenBank 上已有的共 39

基金项目:国家“863 计划”(2002AA2Z2051);国家杰出青年科学基金资助项目(30125001)

* 通讯作者。Tel:86-22-66229592;Fax:86-22-66229596;E-mail:lfeng@tedamail.nankai.edu.cn

作者简介:陶 江(1976 -),男,江苏人,博士生,主要从事细菌表面抗原的分子遗传学方面的研究。E-mail:friendtaojiang@sina.com.cn

其他作者:王 磊

收稿日期:2003-08-08,修回日期:2003-11-19

© 中国科学院微生物研究所期刊联合编辑部 http://journals.im.ac.cn

种血清型的大肠杆菌和志贺氏菌 O 抗原基因簇序列中的 110 个糖基转移酶和 39 个聚合酶的基因序列，以及与大肠杆菌有相同 O 抗原的 3 种沙门氏菌（*Salmonella enterica*）的聚合酶基因序列（表 1）。

表 1 研究用的基因序列

Table 1 Genes involved in this study

O serotype	GenBank No. or source of gene cluster	Genes involved in this study
Shigella sonnei	AF285971	*wzy*
Shigella flexneri 2a	X71970	*rfbE*, *rfbF*, *rfbG*, *wzy*
Shigella dysenteriae O1	L07293	*wbbQ*, *wbbR*, *wzy*
S. dysenteriae O8	Unpublished data**	*orf*1, *orf*4, *orf*7, *wzy*
S. dysenteriae O7	Unpublished data	*wbnI*, *wbnK*, *wbnL*, *wzy*
S. dysenteriae O3	Unpublished data	*orf*4, *orf*5, *orf*7, *orf*9, *wzy*
S. dysenteriae O2	Unpublished data	*orf*3, *orf*4, *orf*6, *orf*7, *wzy*
/*S. boydii* O15		
/*E. coli* O112*		
S. dysenteriae O12	Unpublished data	*wzy*
Shigella boydii O8	Unpublished data	*orf*3, *orf*6, *orf*7, *orf*11, *wzy*
S. boydii O11	Unpublished data	*orf*5, *orf*7, *orf*8, *orf*9, *orf*10, *wzy*
S. boydii O9	AF402315	*wbgQ*, *wbgR*, *wbgS*, *wzy*
S. boydii O7	Unpublished data	*orf*8, *orf*10, *orf*11, *orf*12, *wzy*
S. boydii O6	AF402314	*wbaS*, *wbaT*, *wbaX*, *wbaY*, *wzy*
S. boydii O5	AF402313	*wbdT*, *wbdX*, *wbdU*, *wbdV*, *wbdW*, *wzy*
S. boydii O4	AF402312	*wbdE*, *wbdF*, *wbdG*, *wbdS*, *wzy*
S. boydii O13	Unpublished data	*wbwE*, *wbwF*, *wbwG*, *wbwH*, *wzy*
S. boydii O1	Unpublished data	*orf*8, *orf*9, *wzy*
Escherichia coli O55	AF461121	*wbgM*, *wbgN*, *wbgO*, *wbgP*, *wzy*
E. coli O96	Unpublished data	*wzy*
E. coli O91	AY035396	*wbsA*, *wbsE*, *wbsF*, *wbsG*, *wzy*
E. coli O77 / O73 / O44*	O77 and O73 are unpublished data. O44 is from sanger center (http://www.sanger.ac.uk/)	*orf*1, *orf*2, *wzy*
E. coli O7	AF125322	*wbbA*, *wbbB*, *wbbC*, *wbbD*, *wzy*
E. coli O6	AJ426045	*orf*3, *orf*4, *orf*5, *orf*7, *wzy*
E. coli O5	Unpublished data of Peter Reeves	*wzy*
E. coli O26	AF529080	*wbuA*, *wbuB*, *wzy*
E. coli O172	Unpublished data	*orf*3, *orf*4, *orf*5, *orf*9, *wzy*
E. coli O157	AF061251	*wbdN*, *wbdO*, *wbdP*, *wzy*
E. coli O150	Unpublished data	*wzy*
E. coli O127	http://www.sanger.ac.uk/	*orf*2, *orf*12, *orf*13, *wzy*
E. coli O113	AF172324	*wbnA*, *wbnB*, *wbnD*, *wbnE*, *wzy*
E. coli O111	AF078736	*wbdH*, *wbdL*, *wbdM*, *wzy*
E. coli O110	Unpublished data	*wzy*
E. coli O107	Unpublished data	*wzy*
E. coli O104	AF361371	*wbwA*, *wbwB*, *wbwC*, *wzy*
E. coli K12 (O16)	U09876	*wbbI*, *wbbJ*, *wbbK*, *wbbL*, *wzy*
Salmonella enterica O30	Unpublished data	*wzy*
S. enterica O50	Unpublished data	*wzy*
S. enterica O35	AF285969	*wzy*

* Three O-antigen forms are identical; ** Unpublished data were obtained in our laboratory.

© 中国科学院微生物研究所期刊联合编辑部 http://journals.im.ac.cn

1.2 生物信息学分析方法

用 Artemis 进行基因注释和分析基因组成(http://www.sanger.ac.uk/)。使用 Blast(http://www.ncbi.nlm.nih.gov/BLAST/)在 Genbank 中进行 DNA 和氨基酸序列的同源性搜索。用 ClustalW (http://www.ebi.ac.uk/clustalw/)进行序列比对,比对结果在 Genedoc(http://www.cris.com/~Ketchup/genedoc.shtml)中展示。

通过序列一致性和糖结构的比较来推测糖基转移酶基因可能的功能。对糖基转移酶蛋白的氨基酸序列在 Pfam 蛋白基序数据库(http://www.sanger.ac.uk/Software/Pfam/search.shtml)中确定糖基转移酶所属的家族。

蛋白的疏水性分析和拓扑结构的预测使用 TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/),HMMTOP 2.0 (http://www.enzim.hu/~tusi/hmmtop/html/submit.html),SOSUI (http://sosui.proteome.bio.tuat.ac.jp)和 TMpred (http://www.ch.embnet.org/software/TMPRED-form.html)。蛋白高级结构的预测使用 3D-PSSM(http://www.sbg.bio.ic.ac.uk/~3dpssm/)。

2 结果和讨论

2.1 O-抗原糖基转移酶与聚合酶的序列一致性、密码子使用和(G+C)%含量的特点

我们对 110 个 O-抗原糖基转移酶,39 个 O-抗原聚合酶的序列一致性,氨基酸组成,密码子使用和(G+C)%含量进行了统计,并统计了 17 个 O-抗原单糖合成酶的数据作为参照。

除 *E. coli* O172 的 *orf*9 与 *E. coli* O26 的 *wbuC* 因具有共同的祖先(本室,未发表)而氨基酸序列一致性为 94%以外,其它 O-抗原糖基转移酶的氨基酸序列一致性为 9%~57%;除 *E. coli* O172 与 *S. boydii* O13 的 O-抗原聚合酶因具有共同的祖先(本室,未发表)而氨基酸序列一致性为 42%以外,O-抗原聚合酶间的氨基酸序列一致性为 8%~19%;O-抗原单糖合成酶的氨基酸序列一致性为 58%~99%。O-抗原糖基转移酶与聚合酶的序列一致性显著低于单糖合成酶。

大肠杆菌和志贺氏菌的 O-抗原受到来自宿主免疫系统很强的选择压力,不断适应环境的过程造成了 O-抗原血清型的高度多样性。O-抗原中单糖的种类有限,但通过单糖的不同组合形成的糖链则可以表现出丰富的多样性,糖基转移酶基因和聚合酶基因正是这种多样性在基因水平上的反映,糖基转移酶和聚合酶在 O-抗原的合成中直接决定着 O-抗原的多样性。O-抗原糖基转移酶与聚合酶基因序列一致性低,所以是血清型特异的基因,是通过 PCR 或杂交方法快速区分不同 O-抗原血清型的很好的分子靶点[11]。

O-抗原糖基转移酶的平均稀有密码子含量为 11.18%,O-抗原聚合酶的平均稀有密码子含量为 11.06%,单糖合成酶的平均稀有密码子含量为 6.33%,糖基转移酶与聚合酶的稀有密码子含量显著高于单糖合成酶。糖基转移酶与聚合酶中高含量的稀有密码子可能是二者在实验中表达量低的重要原因。在细菌生长的对数中期之后,某些同功 tRNA 的含量会降低,从而限制翻译,这可能是细菌对 O-抗原基因表达的调控途径之一,使表达水平随细菌生长率的降低而降低。

O-抗原糖基转移酶的 GC 平均百分含量为 31.97%,O-抗原聚合酶的 GC 平均百分含量为 28.47%,单糖合成酶的 GC 平均百分含量为 40.14%,糖基转移酶与聚合酶的(G+C)%含量显著低于合成酶。大肠杆菌和志贺氏菌 O-抗原基因簇的(G+C)%含量为 35%,显著低于基因组的平均(G+C)%含量(50%),说明了 O-抗原基因簇的外源性[7]。

在 O-抗原基因簇中,血清型特异基因编码的酶:糖基转移酶和聚合酶的(G+C)%含量显著低于合成酶,稀有密码子含量显著高于合成酶,这说明血清型特异基因与保守的合成酶基因可能有着不同的来源与进化过程。目前普遍认为,外源基因的(G+C)%含量和稀有密码子含量会随着时间的推移而逐步被宿主同化[12],因此糖基转移酶和聚合酶基因可能是在更近的时期从其它低 GC 含量的细菌种属中转移到大肠杆菌和志贺氏菌 O-抗原基因簇中的。通过糖基转移酶和聚合酶基因在不同种属细菌表面多糖抗原基因簇间的横向转移产生新的 O-抗原是细菌 O-抗原进化可能的重要途径之一。

2.2 聚合酶二级结构的特征

O-抗原聚合酶中 3 种疏水氨基酸:亮氨酸(Leu),异亮氨酸(Ile)和苯丙氨酸(Phe)的含量很高,三者的总含量平均达到 31.71%。通过对蛋白疏水性的预测我们发现聚合酶普遍存在 8~13 个穿膜区。聚合酶在细胞内膜和外膜之间的细胞周质中行使功能,多个穿膜区的存在说明 O-抗原聚合酶是典型的膜蛋白。O-抗原聚合酶被推测同时具有透性酶

© 中国科学院微生物研究所期刊联合编辑部 http://journals.im.ac.cn

的功能，其多个穿膜区可能构成一个内膜通道，通过电化学势驱动脂载体回到胞质中，使脂载体可以循环用于寡糖单位的合成[10]。

通过对疏水区与亲水区拓扑结构的分析，我们发现 O-抗原聚合酶在不同位置都存在一个长度超过 30 个氨基酸的位于细胞周质中的主要由亲水氨基酸组成的环(loop)，这一结果与已经鉴定功能的 2 个大肠杆菌和志贺氏菌 O-抗原聚合酶的蛋白拓扑结构情况相一致[10, 13]，我们的结果说明这样一个亲水环在 O-抗原聚合酶中是广泛存在的。大的亲水环的普遍存在使聚合酶显著区别于转运酶，转运酶的多个穿膜区在整个蛋白中均匀分布[8, 9]。我们比较了具有相同 O-抗原的大肠杆菌和沙门氏菌的聚合酶的氨基酸序列，结果显示亲水环有着比整个聚合酶更高的保守性(表 2)。O-抗原的聚合过程发生在细胞周质中，聚合酶在细胞周质中的亲水环对 O-抗原的聚合可能起着重要作用，是可能的底物结合位点和聚合活性位点所在区域。

表 2 具有相同 O-抗原结构的大肠杆菌和沙门氏菌聚合酶氨基酸序列的比较

Table 2 Comparison of Wzy proteins of *E. coli* and *S. enterica* with identical O-antigen

O serotype	Amino acid identity of the Wzy protein/%	Amino acid identity of the periplasmic loop/%
E. coli O157 and *S. enterica* O30	59	75
E. coli O55 and *S. enterica* O50	55	61
E. coli O111 and *S. enterica* O35	74	84

2.3 聚合酶对寡糖链的特异性

聚合酶催化形成的键为寡糖单位第一个单糖和前一个寡糖单位最后一个单糖之间的二糖连接键，但我们对寡糖单位间二糖键相同而寡糖单位组成不同的 4 个 O-抗原(*S. boydii* O4，*S. boydii* O9，*S. dysenteriae* O1，*E. coli* O26)的聚合酶基因的比较发现，其蛋白序列一致性只有 11%至 14%，说明聚合酶可能是以整个寡糖单位作为底物，对不同的寡糖单位具有特异性，O-抗原聚合酶较低的序列一致性可能源于寡糖单位组成的多样性。

通过对相同 O-抗原寡糖单位被温和型噬菌体在不同位点特异性糖基化修饰产生的 3 种不同的大肠杆菌 O-抗原血清型 O44，O73 和 O77 的聚合酶比较发现，其氨基酸序列的一致性为 99.8%～100%，说明 O-抗原寡糖单位的糖基化修饰对聚合酶行使功能没有影响。温和型噬菌体对大肠杆菌和志贺氏菌 O-抗原的特异性糖基化修饰也发生在细胞周质中[14]，不同糖基化修饰的 O-抗原由基本上完全相同的聚合酶聚合显示糖基化修饰过程是发生在聚合过程之后的。

2.4 通过序列比对推测糖基转移酶可能的功能

我们通过比较 O-抗原基因簇中糖基转移酶基因的一致性和寻找不同 O-抗原中相同的二糖键预测了 4 组糖基转移酶可能的功能(表 3)。我们进一步比较了催化形成 D-GlcA(β1→3)D-GalNAc 和 D-GlcA(β1→3)D-GlcNAc 的糖基转移酶的蛋白序列一致性，为 46%～53%。大肠杆菌和志贺氏菌以 GlcNAc 或 GalNAc 作为寡糖链的第一个糖，第一个糖的合成和转移由 O-抗原基因簇以外的基因决定[15]，较高的蛋白序列一致性表明 O-抗原糖基转移酶对受体是 GlcNAc 还是 GalNAc 的特异性较低。根据这一点，我们预测了另外 2 组供体相同而受体分别为 GlcNAc 和 GalNAc 的糖基转移酶的功能(表 3)。

利用糖基转移酶来合成重要的寡糖分子是刚刚起步的一个领域，有巨大的应用前景。在已知序列的约 2000 种可能编码此类酶的基因中，人们只对大约 30 种的基因功能进行了研究。本文推测了 6 组 12 个糖基转移酶催化的二糖键，可以在此基础上进行生物学实验鉴定其功能。

表 3 推测的 6 组糖基转移酶的功能

Table 3 Proposed function of six group glycosyltransferases

Gene 1/O serotype	Gene 2/O serotype	Disaccharide linkage catalyzed by glycosyltransferases	Amino acid identity(%)/Similarity(%)
*orf*7/*S. dysenteriae* O8	*orf*11/*S. boydii* O8	D-GlcA(β1→3)D-GalNAc	57/75
wbdX/*S. boydii* O5	*wbgS*/*S. boydii* O9	D-GlcA(β1→3)D-GlcNAc	52/66
*orf*9/*S. boydii* O1	*wbdS*/*S. boydii* O4	L-Rhap(β1→4)D-GlcNAc	51/68
*orf*9/*E. coli* O172	*wbuB*/*E. coli* O26	L-FucNAc(α1→3)D-GlcNAc	94/96
*orf*9/*S. dysenteriae* O3	*orf*12/*S. boydii* O7	D-Gal(β1→3) D-GlcNAc/GalNAc	43/62
*orf*7/*E. coli* O6	*orf*2/*E. coli* O73/O77/O44	D-Man(β1→3)D-GlcNAc/GalNAc	52/70

© 中国科学院微生物研究所期刊联合编辑部 http://journals.im.ac.cn

2.5 O-抗原的糖基转移酶属于两个不同的糖基转移酶超家族

糖基转移酶在生物体中广泛存在,按底物的结构,反应的立体化学特征和序列一致性分为 65 个家族(糖基转移酶家族 1-65,http://afmb.cnrs-mrs.fr/CAZY)。近年来,来源于原核和真核的 10 个不同家族的糖基转移酶晶体结构的研究显示,虽然糖基转移酶的氨基酸序列具有高度多样性,却只存在两个糖基转移酶超家族 糖基转移酶超家族 A 和糖基转移酶超家族 B。超家族 A 的糖基转移酶的结构为两个紧密相连的结构域共同组成一个圆锥体,在其顶部形成一个结合底物的口袋状区域。超家族 A 的糖基转移酶都含有一个结合二价金属离子(多为 Mn^{2+})的"DXD"基序,"DXD"基序是一个序列为 DXD、DDX 或 XDD 的天冬氨酸富含区[16],其中 X 代表任意氨基酸。超家族 B 的糖基转移酶由两个拓扑结构相似的结构域组成,两个结构域之间有一个容纳底物的缝隙[16]。

通过对 Pfam 蛋白基序数据库的搜索,我们发现 110 个 O-抗原糖基转移酶中大部分是糖基转移酶家族 1 和家族 2 的成员,其中 36 个属于糖基转移酶家族 2(E 值 $= 1.4 \times e^{-4} \sim 3.0 \times e^{-45}$),37 个属于糖基转移酶家族 1($E$ 值 $= 6.0 \times e^{-2} \sim 1.8 \times e^{-45}$)。在已知晶体结构的糖基转移酶中,*Bacillus subtilis* 中的 SpsA 是糖基转移酶家族 2 的成员,属于糖基转移酶超家族 A;*Amycolatopsis orientalis* 中的 GtfB 是糖基转移酶家族 1 的成员,属于糖基转移酶超家族 B。我们通过进一步分析发现,36 个属于家族 2 的糖基转移酶在蛋白的氨基端一侧都含有两个天冬氨酸(D)富含区,其中第二个天冬氨酸富含区为'DXD'基序,这是超家族 A 成员的典型特征,因此这 36 个糖基转移酶家族 2 的成员属于糖基转移酶超家族 A。由于糖基转移酶超家族 B 的成员在一级序列上没有明显的特征,我们对 37 个糖基转移酶家族 1 的成员用 3D-PSSM 进行了高级结构的预测,它们都被预测为糖基转移酶超家族 B 的成员(E 值 $= 2.53 \times e^{-2} \sim 9.11 \times e^{-6}$)。O-抗原糖基转移酶虽然功能相近,作用区域相同,但其高级结构却分别属于两个有明显区别的蛋白超家族,这说明了 O-抗原糖基转移酶来源的多样性,也暗示了宿主的此类基因通过基因漂移进入细菌表面抗原基因簇的可能。

参 考 文 献

[1] Ewing W H, Lindberg A A. Serology of *Shigella*. In: Bergan T. ed. Methods Microbiol. London: Academic Press, 1984, 113 – 142.

[2] Ewing W H. Edwards and Ewing's identification of the *Enterobacteriaceae*. 4th ed. The Netherlands: Elsevier Science Publishers, 1986, 273 – 274.

[3] Pupo G M, Lan R, Reeves P R. Multiple independent origins of *Shigella clones* of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci USA*, 2000, **97**: 10567 – 10572.

[4] Wang L, Qu W, Reeves P R. Sequence analysis of four *Shigella boydii* O – antigen loci: implication for *Escherichia coli* and Shigella relationships. *Infect Immun*, 2001, **69**: 6923 – 6930.

[5] Lior H. Classification of *Escherichia coli*. In: Gyles C L. ed. *Escherichia coli* in domestic animals and humans. UK: CAB International, 1994, 31 – 72.

[6] Brenner D J. Enterobacteriaceae. In: Krieg N R, *et al*. ed. Bergey's Manual of Systematic Bacteriology. Baltimore: Williams and Wilkins, 1984, 408 – 420.

[7] Reeves P R, Wang L. Genomic organization of LPS – specific loci. *Curr Top Microbiol Immunol*, 2002, **264**: 109 – 135.

[8] Liu D, Cole R, Reeves P R. An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. *J Bacteriol*, 1996, **178**: 2102 – 2107.

[9] Feldman M F, Marolda C L, Monteiro M A, *et al*. The activity of a putative polyisoprenol-linked sugar translocase (Wzx) involved in *Escherichia coli* O antigen assembly is independent of the chemical structure of the O repeat. *J Biol Chem*, 1999, **274**: 35129 – 35138.

[10] Daniels C, Vindurampulle C, Morona R. Overexpression and topology of the *Shigella flexneri* O-antigen polymerase (Rfc/Wzy). *Mol Microbiol*, 1998, **28**: 1211 – 1222.

[11] Wang L, Curd H, Qu W, *et al*. Sequencing of *Escherichia coli* O111 O antigen gene cluster and identification of O111 specific genes. *J Clin Microbiol*, 1998, **36**: 3182 – 3187.

[12] Nomura M, Sor F, Yamagishi Y, *et al*. Heterogeneity of GC content within a single bacterial genome and its implications for evolution. *Cold Spring Harb Symp Quant Biol*, 1987, **52**: 658 – 663.

[13] Lukomski S, Hull R A, Hull A I. Identification of the O antigen polymerase (*rfc*) gene in *Escherichia coli* O4 by insertional mutagenesis using a nonpolar chloramphenicol resistance cassette. *J Bacteriol*, 1996, **178**: 240 – 247.

[14] Allison G E, Verma N K. Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends Microbiol*, 2000, **8**: 17 – 23.

[15] Alexander D C, Valvano M A. Role of the *rfe* gene in the biosynthesis of the *Escherichia coli* O7-specific lipopolysaccharide and other O – specific polysaccharides containing N-acetylglucosamine. *J Bacteriol*, 1994, **176**: 7079 – 7084.

[16] Breton C, Imberty A. Structure/function studies of glycosyltransferases. *Current Opinion in Structural Biology*, 1999, **9**: 563 – 571.

© 中国科学院微生物研究所期刊联合编辑部 http://journals.im.ac.cn

Computing Analysis of Glycosyltransferases and Polymerases Involved in O – antigen Biosynthesis of *Escherichia coli* and *Shigella*

TAO Jiang[1] LIU Bin[1] WANG Quan[1] GUO Hong-Jie[1,2,3] FENG Lu[1,2,3 *]

([1] *College of Life Sciences*, *Nankai University*, *Tianjin* 300071, *China*)

([2] *TEDA School of Biological Sciences and Biotechnology*, *Nankai University*, *Tianjin* 300457, *China*)

([3] *Tianjin State Laboratory of Microbial Functional Genomics*, *TEDA College*, *Nankai University*, *Tianjin* 300457, *China*)

Abstract: The sequential and structural characters of 110 glycosyltransferases and 39 polymerases (Wzy) of *Escherichia coli* and *Shigella* was studied using bioinformatic methods. The statistic data of sequence similarity, codon usage and (G + C) % content are shown. The substrate specificity of O-antigen glycosyltransferases and polymerases are proposed. Analysing amino acid sequences of all glycosyltansferases indicates fuctions of six different groups. It is found that a predicted periplasmic loop is present in every polymerase and is the putative functional region of polymerase. Based on their predicted 3D-structures, glycosyltransferases can be grouped into two protein super families.

Key words: *Escherichia coli*, *Shigella*, O-antigen, Glycosyltransferase, Polymerase, Bioinformatics

---

Foundation item: Chinese National Programs for High Technology Research and Development (2002AA2Z2051); Chinese National Science Fund for Distinguished Young Scholars (30125001)

* Corresponding author. Tel: 86-22-66229592; Fax: 86-22-66229596; E-mail: lfeng@tedamail.nankai.edu.cn

Other author: WANG Lei

Received date: 08-08-2003

© 中国科学院微生物研究所期刊联合编辑部 http://journals.im.ac.cn