

# 大肠杆菌 O138 O-抗原基因簇的破译及 Gne 的生物信息学鉴定

孔庆科<sup>1</sup> 程剑松<sup>1</sup> 赵 广<sup>1</sup> 王 磊<sup>1,2</sup> 郭宏杰<sup>1,2\*</sup>

(<sup>1</sup> 南开大学生命科学学院 天津 300071)

(<sup>2</sup> 南开大学泰达学院功能基因组学研究中心 天津 300457)

**摘 要** 利用鸟枪法对大肠杆菌 *E. coli* O138 O-抗原基因簇进行测序,序列全长 14139bp,用生物信息学的方法进行序列分析,共发现 11 个基因,分别为鼠李糖合成酶基因(*rmlB*, *rmlD*, *rmlA*, *rmlC*)、UDP-GalNAcA 合成酶基因(*gne*, *gna*)、糖基转移酶基因(3 个)、O-抗原转运酶基因(*uzzx*)和 O-抗原聚合酶基因(*uzy*)。发现一种稀有单糖 UDP-GalNAcA 的合成途径,对合成该糖的第一种酶 Gne 进行了生物信息学鉴定,另外用 PCR 方法筛选出了针对大肠杆菌 O138 的特异基因。

**关键词** 大肠杆菌 O138, O-抗原, 特异基因, UDP-GlcNAc C4 异构酶

中图分类号: Q93 文献标识码: A 文章编号: 0001-6209(2004)06-0756-05

大肠杆菌是动物和人类肠道中占主导地位的革兰氏阴性菌。部分大肠杆菌是条件致病菌,按致病机制分为肠致病性大肠杆菌(Enteropathogenic *Escherichia coli*, EPEC)、肠侵袭性大肠杆菌(Enteroinvasive *E. coli*, EIEC)、肠产毒性大肠杆菌(Enterotoxigenic *E. coli*, ETEC)、肠出血性大肠杆菌(Enterohaemorrhagic *E. coli*, EHEC)和肠聚集性大肠杆菌(Enterocaggregative *E. coli*, EaggEC)等<sup>[1]</sup>。大肠杆菌 O138 是一株非常重要的 ETEC 致病菌,既能产生热稳定毒素 ST,又可以产生细胞性毒素 2e(VT2e),引起家畜的腹泻、皮下组织水肿和神经系统病,在畜牧业生产上危害很大<sup>[2,3]</sup>。

细菌表面多糖与细菌的致病性关系密切,脂多糖存在于革兰氏阴性细菌表面,是表面多糖的一种。脂多糖一般由 3 部分组成,从内到外分别是脂质 A、核心寡糖和 O-特异性多糖即 O-抗原。O-抗原一般由 3~8 个单糖重复单位组成,O-抗原在细菌外膜的外侧,是抵御环境变化(如抗生素,胁迫)的第一道屏障<sup>[4,5]</sup>。而且,在许多生物互作中,O-抗原在致病性、信号识别、黏附、免疫逃避等过程中发挥重要作用。在进化压力的选择下,大肠杆菌产生了数量众多的 O-抗原血清型,到目前为止,已经有超过 160 种的 O-抗原被鉴定。O-抗原是自然界中最具多样性的分子之一,选取大肠杆菌 O-抗原作为材料进行遗传多样性和进化机理的研究,最具代表性<sup>[6]</sup>。

本实验破译了大肠杆菌 O138 的 O-抗原基因簇全序列,筛选出了对大肠杆菌 O138 高度特异性的引物,可以用于对该菌的快速检测。另外对其进化机理进行了分析,发现了一种稀有单糖 GalNAcA 的合成路径。

## 1 材料和方法

### 1.1 材料

**1.1.1 菌种**: 大肠杆菌 O138 标准菌株购自中国生物制品鉴定中心,受体菌株(大肠杆菌 *E. coli* DH5 $\alpha$ )由本室保存。

**1.1.2 载体质粒**: pGEM-T-easy Vector 购自 Promega 公司。

**1.1.3 主要酶和试剂**: Expand Long Template PCR 酶购自 Boehringer Mannheim 公司,PCR 纯化试剂盒、T4 连接酶、Wizard PCR Preps 纯化试剂盒购自 Promega 公司, EcoRI、DNase I、X-Gal、IPTG、氨苄青霉素购自上海 Sangon 公司。

### 1.2 基因组 DNA 的提取

采用 Bastin 法<sup>[7]</sup>提取细菌染色体 DNA。

### 1.3 O-抗原基因簇的获得

参照郭宏杰<sup>[8]</sup>的实验方法。

### 1.4 O-抗原基因文库的构建

用 DNase I 消化 PCR 纯化产物,提取大小集中在 1~3kb 之间的片段,经过补平,尾部再加上一个

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/Fax: 86-22-66229591, E-mail: guohj@nankai.edu.cn

作者简介: 孔庆科(1976-),男,山东济宁人,博士研究生,主要从事细菌表面抗原的分子遗传学研究。E-mail: kongqingke@eyou.com

其他作者: 郭 玺

收稿日期: 2004-02-24, 修回日期: 2004-05-13

碱基 A,与 pGEM-T-easy 载体在 T4 连接酶作用下 16℃连接 24 h,取连接产物 2 $\mu$ L 电转化感受态大肠杆菌 DH5 $\alpha$ ,涂布在含有氨苄青霉素、X-Gal 和 IPTG 的 LB 固体培养基上,37℃过夜培养,筛选白色克隆,以碱裂解法用 96 孔平板提取质粒,*Eco*R I 酶切鉴定插入片段大小,挑选插入片段在 1kb 以上的克隆。

利用双脱氧终止法,采用 T7-sp6 公用引物,用 ABI3730 型 DNA 自动测序仪对 96 个克隆中的插入片段进行测序,用英国剑桥 MRC(Medical Research Council)分子生物学实验室出版的 Staden package 软件包的 Pregap4 和 Gap4 软件拼接和编辑所有序列。

### 1.5 生物信息学方法进行序列分析

用美国国家生物技术信息学中心(The National Center for Biotechnology Information,NCBI)的 ORFfinder 发现基因;用 BLAST 系列软件与 GenBank 中的基因比较以发现这些开放阅读框的功能并确定基因的功能;

用英国 Sanger 中心的 Artemis 软件完成基因注释;用 Clustal W 软件做 DNA 和蛋白质序列间的精确比对;用 HMMTOP(<http://www.enzim.hu/hmmtop/>)和 SMART(<http://smart.embl-heidelberg.de/>)进行蛋白质跨膜片段的分析;用 MEGA 软件产生系统进化树;用 <http://cn.expasy.org/> 中的软件 Jpred(<http://www.compbio.dundee.ac.uk/~www-jpred/>)作蛋白二级结构预测。

### 1.6 特异基因的筛选

用 PCR 方法筛选针对大肠杆菌的特异基因。针对大肠杆菌 O138 抗原基因簇中的 *wzx* 和 *wzy* 基因,在基因内部各设计两对引物(表 1),分别以 150 株大肠杆菌和 42 株志贺氏菌的基因组为模板进行 PCR。PCR 扩增条件:94℃ 2min;94℃ 15s,退火温度因引物的不同而不同(表 1),退火时间 50s,72℃ 2min,共 30 个循环;72℃ 10min,反应体系是 25 $\mu$ L。

表 1 *E. coli* O138 的 O-抗原基因簇中寡糖单位处理基因及其中的引物和 PCR 数据

Gene	Base positions of genes	Base positions of forward/Reverse primers	Length of PCR product/bp	Annealing T of PCR/℃
<i>wzx</i>	4632 ~ 5909	5388 ~ 5407/5827 ~ 5847	459	55
		4690 ~ 4708/5393 ~ 5411	721	55
<i>wzy</i>	11469 ~ 12480	11536 ~ 11554/12023 ~ 12041	505	55
		12024 ~ 12042/12502 ~ 12520	496	55

## 2 结果和分析

### 2.1 大肠杆菌 O138 O-抗原基因簇的序列分析

将获得的测序结果,用 Pregap4 软件拼接后,留下的 gap 用 PCR 方法补完,得到大肠杆菌 O138 的 O-抗原基因簇的全序列,序列全长 14139bp,专利申请号为(200310117855.4)。用 ORFfinder 找到 11 个开放阅读框,所有开放阅读框的 5'→3'方向为从 *galF*(1~765)基因到 *gnd*(12800~14139)基因,图 1 示 O138 O-抗原基因簇的排列顺序和(G+C)%含

量。*galF* 和 *gnd* 两个基因并不负责 O-抗原基因簇的合成,仅用来扩增 O-抗原基因簇。根据基因的相似性,所预测的 11 个开放阅读框包含有合成鼠李糖合成酶基因 *rmIBDAC*、一种稀有单糖 GalNAcA 合成酶基因 *gna* 和 *gne*,并且预测该单糖的合成途径为  $UDP-GlcNAc \xrightarrow{Gne} UDP-GalNAc \xrightarrow{Gna} UDP-GalNAcA$ ,以及 3 个糖基转移酶基因 *orf6*,*orf7* 和 *orf10*。两个寡糖单位处理基因 *wzx* 和 *wzy*,其特征为多次跨膜的膜蛋白,并且 *Wzy* 含有一个约 60 个氨基酸的亲水环(表 2)。

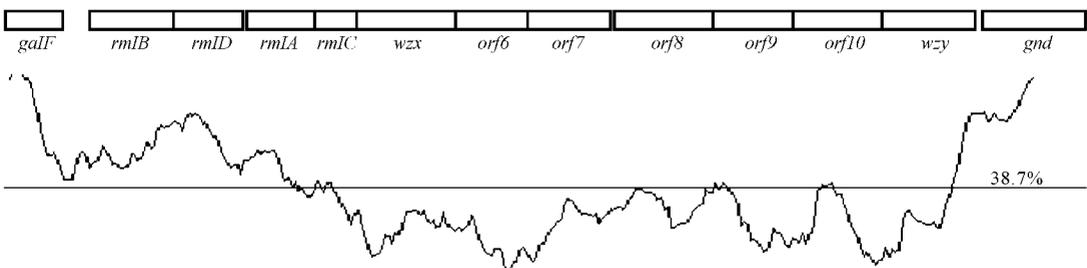


图 1 *E. coli* O138 的 O-抗原基因簇结构及(G+C)%含量

Fig.1 O-antigen gene cluster and (G+C)% content of *E. coli* O138

表 2 大肠杆菌 O138 O-抗原基因簇的基因功能预测

Table 2 Characteristics of the ORFs located in the *Escherichia coli* O138 O-antigen gene cluster

Gene	Location in sequence	No. of aa	G + C content/%	Similar proteins Accession number	Identical Similar aa (aa overlap) %	Putative function of protein
<i>rmlB</i>	1137..2222	361	43.2	RmlB, <i>Shigella boydii</i> (AAL27322)	98/99 (361)	dTDP-glucose 4- $\beta$ -dehydratase
<i>rmlD</i>	2222..3121	299	47.3	RmlD, <i>Escherichia coli</i> (BAA15882)	97/99 (299)	TDP-6-deoxy-L-mannose-dehydrogenase
<i>rmlA</i>	3179..4057	292	42.7	RmlA, <i>Shigella boydii</i> (AAL27312)	98/99 (292)	glucose-1-phosphate thymidyltransferase
<i>rmlC</i>	4062..4598	178	38.2	RmlC, <i>Shigella boydii</i> (AAL27313)	84/91 (179)	dTDP-6-deoxy-D-glucose-3,5-epimerase
<i>wzx</i>	4632..5909	425	33.9	Wzx, <i>Shigella flexneri</i> 2a str. 301 (AAN43639)	47/69 (406)	O-antigen flippase
<i>orf6</i>	5899..6855	318	31.2	ORF_9, similar to Glycosyl transferase, <i>Pseudomonas aeruginosa</i> (AAM27629)	42/63 (302)	glycosyl transferase
<i>orf7</i>	6848..7924	358	32.7	ORF_10, similar to Glycosyl transferases group 1 <i>Pseudomonas aeruginosa</i> (AAM27630)	44/61 (345)	glycosyl transferase
<i>gna</i>	7985..9262	425	36.5	UDP-N-acetyl-D-galactosamine dehydrogenase, <i>Vibrio vulnificus</i> (AAQ17507)	58/75 (415)	UDP-N-acetyl-D-galactosamine dehydrogenase
<i>gne</i>	9264..10289	341	34.9	wbpP, <i>Pseudomonas aeruginosa</i> (AAM27817)	66/77 (337)	epimerase
<i>orf10</i>	10333..11484	383	34.3	glycosyl transferase, group 1 family protein <i>Shewanella oneidensis</i> MR-1 (AAN56174)	52/70 (364)	glycosyl transferase
<i>wzy</i>	11469..12680	403	31.2	O-antigen polymerase <i>Shigella boydii</i> (AAL27318)	22/43 (397)	O-antigen polymerase

## 2.2 大肠杆菌 O138 特异基因的鉴定

O-抗原基因簇中的寡糖单位处理基因(*wzx*、*wzy*)为不同菌的特异基因,可用于对该菌的检测。根据这一原理,设计针对寡糖单位处理基因的 PCR 引物,表 1 示引物序列,退火温度和引物在基因中的位置,以大肠杆菌 O138 菌株作为正对照,对 150 株大肠杆菌和 42 株志贺氏菌进行 PCR 反应。不能产生正反应产物的基因将极有可能是大肠杆菌 O138

的特异基因。结果显示,在所有待测的细菌中,仅在 O138 中产生了正确的大小片段,在其它细菌中都没有扩增出正确的片段,说明我们根据 *wzx* 和 *wzy* 基因设计的引物对 *E. coli* O138 是高度特异的,*wzx* 和 *wzy* 基因在 *E. coli* O138 中也是特异基因。

## 2.3 *gne* 基因的生物信息学鉴定

O138 O-抗原基因簇中的 *orf9* 命名为 *gne*,是一种 UDP-GlcNAc C4 异构酶基因,催化 UDP-GlcNAc

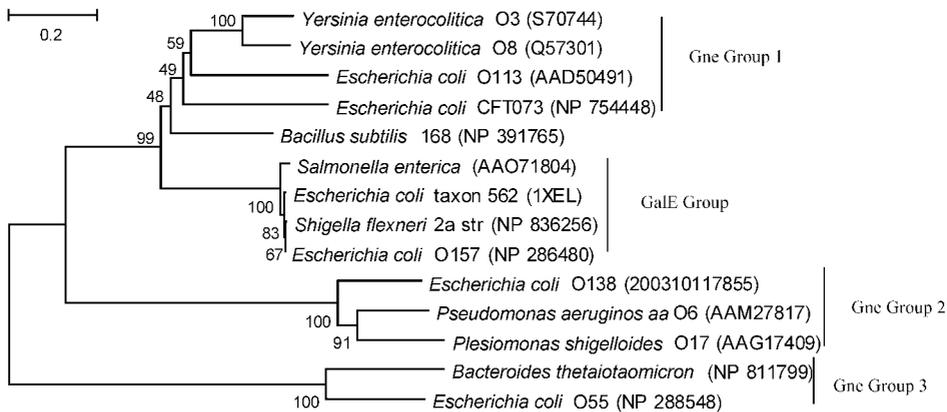


图 2 依据相似蛋白的序列构建的 Gne 和 GalE 的系统进化树

Fig. 2 Phylogenetic trees based on Gne and GalE proteins sequences showing the relationship between

*E. coli* O138 Gne and other representative Gne or GalE proteins

Numbers in parentheses represent the sequences' accession number in GenBank, number of O138 in parentheses represents the Chinese patent number.

The scale bar indicates 0.2 substitutions per amino acid position, the number at each node represents the bootstrap value. © 中国科学院微生物研究所期刊联合编辑部 http://journals.im.ac.cn

形成 UDP-GalNAc。我们通过 BLAST 检索 找到了与 Orf9 同源性较高的蛋白,其中包括 GalE。利用这些基因产生的蛋白作系统进化树(图 2),发现 O138 Gne 蛋白位于一类较特殊的 Gne 蛋白序列中。

选取晶体结构已经解析的 *E. coli* O157 GalE 和活性位点已知的 *Yersinia enterocolitica* O8 的 Gne 与 O138 的 Gne 进行比对,寻找其共同的二级结构。共同的二级结构 N-端有 6 个  $\beta$ -sheet,C-端有 4 个  $\beta$ -sheet。C 端结合 UDP,N 端结合单糖<sup>[9-11]</sup>,由于这类酶催化需要辅助因子 NADP<sup>+</sup>,而这 3 种酶都有 GXXGXXG 位点结合辅助因子。同时我们根据已知酶的活性位点也找到了 O138 Gne 的活性位点 SYK,该催化位点位于两个区域中间。136 和 297 附近位点的氨基酸大小决定底物的专一性,如果两个位点的氨基酸体积较小,就可以容纳体积较大的底物,催化范围较广,Holden<sup>[9]</sup>将 GalE 中 299 位的 Tyr 用 Cys 代替后,GalE 的催化活性发生变化,能同时催化 UDP-Glucose 和 UDP-GlcNAc。我们发现 O138 的 Gne 在 136 和 297 位附近的氨基酸体积较小,我们推测该酶不仅有催化 UDP-GlcNAc 的功能,可能也有催化 UDP-Glucose 的功能。

### 3 讨论

以表面多糖为目标的血清学免疫反应自上世纪 30 年代以来一直被用于对细菌的分型和鉴定,是鉴定致病菌的重要手段。这种诊断方法需要大量的抗血清,而抗血清一般种类不全,数量不足,大量的抗血清在制备和储存中也存在一些困难。另一方面此法耗时长、灵敏度低、漏检率高、准确性差,所以,现在普遍认为这种传统的血清学检测方法将为现代分子生物学方法取代。我们根据大肠杆菌 O138 的 O-抗原序列,所设计的 4 对引物,仅在大肠杆菌 O138 中扩增出大小与预测结果一致的片段,在其它供试细菌中都没有扩增出片段。因此可以利用筛选出的特异 DNA 片段用于基于基因芯片或 PCR 方法对大肠杆菌的快速检测。

大肠杆菌、致贺氏菌和沙门氏菌 O-抗原基因簇位于 *galF* 和 *gnd* 之间,为低 GC 含量的 DNA 片段,这一特征说明它们起源于低 GC 含量的细菌,近才通过横向转移进入大肠杆菌中。O-抗原中的糖合成酶基因,两端的持家基因 *galF*、*gnd* 及 O-抗原的调控序列等都可以介导新的 O-抗原的形成。

### 参 考 文 献

- [ 1 ] Morris J A , Sojka W J. *Escherichia coli* as a pathogen in animals. In : Sussman M. ed. *The Virulence of Escherichia coli : Reviews and Methods*. London : Academic Press , 1985 , 47 - 77 .
- [ 2 ] Frydendahl K , Strodl Andersen J , Fredholm M , et al . Association between the porcine *Escherichia coli* F18 receptor genotype and phenotype and susceptibility to colonisation and postweaning diarrhoea caused by *E. coli* O138 :F18. *Vet Microbiol* , 2003 , **93** ( 1 ) : 39 - 51 .
- [ 3 ] Frydendahl K . Prevalence of serogroups and virulence genes in *Escherichia coli* associated with postweaning diarrhoea and edema disease in pigs and a comparison of diagnostic approaches. *Vet Microbiol* , 2002 , **85** ( 2 ) : 169 - 182 .
- [ 4 ] Lerouge I , Vanderleyden J . O-antigen structural variation : mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiology Reviews* , 2002 , **26** ( 1 ) : 17 - 47 .
- [ 5 ] Reeves P R , Hobbs M , Valvano M , et al . Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol* , 1996 , **4** : 495 - 503 .
- [ 6 ] Bettelheim K A . Biochemical characteristics of *Escherichia coli* . In : Gyles C L . ed. *Escherichia coli* in domestic animals and humans . Oxon , U . K : CAB International , 1994 , 31 - 72 .
- [ 7 ] Bastin D A , Romana L K , Reeves P R . Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain. *Mol Microbiol* , 1991 , **5** : 2223 - 2231 .
- [ 8 ] 郭宏杰,冯 露,张 淳,等. 大肠杆菌 O150 O-抗原基因簇的破译和 dTDP-鼠李糖合成酶基因的鉴定. *微生物学报* , 2004 , **44** ( 1 ) : 34 - 40 .
- [ 9 ] Thoden J B , Fridovich-Keil J L , Holden H M . Structural analysis of the Y299C mutant of *Escherichia coli* UDP-galactose 4-epimerase : Teaching an old dog new tricks. *J Biol Chem* , 2001 , **276** ( 23 ) : 20617 - 20623 .
- [ 10 ] Thoden J B , Wesenberg G , Chapeau M C , et al . Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from *Escherichia coli* . *Biochemistry* , 1997 , **36** ( 21 ) : 6294 - 6304 .
- [ 11 ] Thoden J B . Dramatic differences in the binding of UDP-galactose and UDP-glucose to UDP-galactose 4-epimerase from *Escherichia coli* . *Biochemistry* , 1998 , **37** ( 33 ) : 11469 - 11477 .

## Sequence of *Escherichia coli* O138 O-antigen Gene Cluster and Gne Identification by Bioinformatics

KONG Qing-Ke CHENG Jian-Song ZHAO Guang WANG Lei GUO Hong-Jie\*

( College of Life Science , Nankai University , Tianjin 300071 , China )

( Center for Functional Genomics Research , TEDA College , Tianjin 300457 , China )

**Abstract :** *E. coli* O138 is one of the enterotoxigenic *Escherichia coli* , causing the postweaning diarrhea and edema disease of weaned pigs. The O-antigen gene cluster of *E. coli* O138 was sequenced and found to contain the genes *rmlB-DAC* and *gne* , *gna* for the biosynthesis of nucleotide sugars dTDP-rhamnose and UDP-GalNAcA , respectively , genes encoding for O unit flippase( *wzx* ) , O-antigen polymerase( *wzy* ) and 3 potential transferase genes. The possible biosynthesis pathway for rare UDP-GalNAcA was proposed. Two genes specific to *E. coli* O138 were identified. This work provides the basis for a sensitive test by PCR for the rapid detection of *E. coli* O138. Phylogenetic tree for Gne and GalE proteins was generated and comparisons were made among different strains , and results revealed that these proteins are similar in the second structure , and Gne of *E. coli* O138 was identified by bioinformatics.

**Key words :** *Escherichia coli* O138 , O-antigen , Specific gene , Gne

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

\* Corresponding author. Tel : 86-22-66229591 ; Fax : 86-22-66229596 ; E-mail : guohj@nankai.edu.cn

Received date : 02-24-2004

## 生活垃圾综合处理原理与技术( 专著 )

郭长军 , 许泽森 , 李泽泉 编著 2004 年 11 月黑龙江人民出版社出版 ; RSDNT-207-00678-3/X57  
字数 25 万字 ; 开本 大 32 开 ; 印数 3000 册 ; 定价 28.00 元

### 读者对象

大专院校相关专业师生、市容环境卫生管理部门、从事城市垃圾处理的相关单位 , 以及所有关心城市垃圾处理工作的社会各界读者。

### 内容简介

环境保护是我国的基本国策 , 而城市垃圾处理是城市环境保护的重要内容 , 也是衡量城市文明程度和城市管理水平的重要标志之一。近几年来 , 我国的大部分城市都正在积极的进行城市生活垃圾处理工程的探索和实践 , 但由于国内无系统的垃圾处理理论和少有成功的实例 , 而从国外引进的卫生填埋、堆肥、焚烧发电三大处理方式均不合中国国情。照搬国外垃圾处理方式建设的处理厂 , 败多成少 , 造成不小的浪费。至今 , 全国城市垃圾无公害化处理率尚不足 20% , 在全球重视环保的今天 , 城市垃圾处理工作任务十分迫切 , 任重而道远。

为探索适合中国城市生活垃圾处理的科学方案 , 在省市有关领导支持和参与下 , 黑龙江高维企业集团于今年 6 月召开了全国首次“城市生活垃圾处理方案研讨会” 。来自中国科学院、清华大学、哈尔滨工业大学、国家发改委、机械部、中石化的专家进行了认真热烈的研讨 , 确定综合处理方式是适合中国国情的最佳方式 , 是与卫生填埋、堆肥、焚烧发电三大方式并列的第四种新方式。

本书将依据研讨会学术成果 , 系统论述生活垃圾综合处理方式的基本原理和使用技术 , 明确阐述其内涵和外延。全书共分七章 30 节。