中山大学学报（自然科学版）
第28卷 ACTA SCIENTIARUM NATURALIUM Vol.28
第1期，1989年 UNIVERSITATIS SUNYATSENI №.1(1989)

DNA物理图谱构建的数学模型和位置网络图*

蔡 勇 陈继承
（计算机科学系）

摘 要

把交叉酶切和双酶切实验的DNA限制性内切酶物理图谱的构建，归入一类特殊的带线性约束的0－1二次规划问题．提出一种DNA物理图谱的交叉位置网络图（简称交叉位置图），并通过实例说明它在构建物理图谱和分析酶切实验中的应用意义．

关键词 DNA酶切物理图谱，数学规划，交叉位置网络图

利用计算机构建DNA物理图谱的研究工作国外从1978年已见开展，如文献〔1—4〕等，国内亦有研究[5,6]．过去的方法基本上都建立在片段的组合排列基础上，并无明确的构建DNA物理图谱的数学模型．本文的研究表明，建立DNA物理图谱构建的数学模型，对于明确双酶切与交叉酶切模型的关系、利用已有的数学方法和应用计算机研究构建DNA物理图谱的算法都有着重大的意义．

1 一类特殊的数学规划问题

所谓DNA限制性内切酶的物理图谱的构建，就是利用双酶切或交叉酶切实验得到片段的大小信息，找出DNA分子（或分子片段）的子片段排列的顺序，即DNA物理图谱．设有两种不同的限制性内切酶α和β．所谓双酶切实验，是指用α和β两种酶共同作酶切实验，记作$\alpha\oplus\beta$酶切；而交叉酶切实验，是指对α（或β）酶切的各个片段分别单独用β（或α）内切酶再作酶切实验，记作$\alpha\otimes\beta$酶切（或$\beta\otimes\alpha$酶切）．

为了讨论方便，先作一些定义

n,m,k——分别为$\alpha,\beta,\alpha\oplus\beta$酶切的段数．对于环状DNA，$n+m=k$；线状DNA，$n+m=k+1$．

本文1988年2月1日收到
* 中山大学高等学术研究中心基金会资助项目

A,B,C——分别为$\alpha,\beta,\alpha\oplus\beta$酶切子片段的长度集合，$A=\{\alpha_1,\alpha_2,\cdots,\alpha_n\}$，$B=\{\beta_1,\beta_2,\cdots,\beta_m\}$，$C=\{\nu_1,\nu_2,\cdots,\nu_k\}$。由背景意义知，成立 $\sum_{i=1}^{n}\alpha_i=\sum_{i=1}^{m}\beta_i=\sum_{i=1}^{k}\nu_i$。

现在定义一种“分派”问题：对上面的集合C作两种分划，求使所得到的两个子集族 $\{C_j^A\}_1^n$和$\{C_j^B\}_1^m$满足

$$\sum_{\nu_i\in C_j^A}\nu_i=\alpha_j,\quad j=1,2,\cdots,n,\qquad \sum_{\nu_i\in C_j^B}\nu_i=\beta_j,\quad j=1,2,\cdots,m$$

其中 $C_i^A\cap C_j^A=\phi$（当$i\neq j$），$C_i^B\cap C_j^B=\phi$（当$i\neq j$），ϕ表示空集合。

这样，DNA酶切物理图谱构建便可归入这一“分派”问题。$\alpha\otimes\beta$交叉酶切模型是已知C的一组分划子集族 $\{C_j^B\}_1^m$；而对于$\alpha\oplus\beta$双酶切模型，C的两组分划子集族皆未知。实际上，由下一节的讨论可知，当确定了集合C的两个分划子集族后便可得到对应的DNA酶切物理图谱。

对于集合C对A的分划可以用矩阵形式表示（见表1）。

表1

	α_1	α_2	$\cdots$	α_n
ν_1	X_{11}	X_{12}	$\cdots$	X_{1n}
ν_2	X_{21}	X_{22}	$\cdots$	X_{2n}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
ν_k	X_{k1}	X_{k2}	$\cdots$	X_{kn}

其中 $X_{ij}=\begin{cases}0 & \nu_i\bar{\in}C_j^A \quad i=1,2,\cdots,k\\ 1 & \nu_i\in C_j^A \quad j=1,2,\cdots,n\end{cases}$ 且满足 $\sum_{j=1}^{n}X_{ij}=1,(i=1,2,\cdots,k)$，

$$\sum_{i=1}^{k}\nu_i\cdot X_{ij}=\alpha_j,(j=1,2,\cdots,n)。$$

$$(\mathrm{PMP})\quad \min\sum_{j=1}^{n}\left(\alpha_j-\sum_{i=1}^{k}\nu_i\cdot X_{ij}\right)^2$$

$$s.t.\begin{cases}\sum_{j=1}^{n}X_{ij}=1, & i=1,2,\cdots,k\\ X_{ij}=0,1 & i=1,2,\cdots,k,\quad j=1,2,\cdots,n\end{cases}$$

对于集合C对B的分划的数学模型可同样建立。(PMP)是一个带线性约束的 0—1 二次规划问题，并不属于传统的分派问题。它的求解比较困难，有关的求解算法等问题将不在本文讨论。

由于实验中难免出现误差，$\sum_{i=1}^{k} v_i \cdot X_{ij} = \alpha_j (j=1,2,\cdots,n)$亦难严格成立，故考虑规划模型(PMP)是合理的。

交叉酶切或双酶切的DNA物理图谱构建模型实际上是一个PMP规划问题或由两个PMP规划问题构成的复合问题。对于交叉酶切模型，我们已找到一种较有效的算法——有向图构建法[6]。

2　DNA酶切物理图谱的交叉位置图

用以描述DNA酶切物理图谱的网络图——交叉位置图，它适用于正、反交叉酶切实验。

首先定义所谓的交叉位置图，假定已知前面讨论的集合C对A和B的两个分划$\{C_j^A\}_1^n$、$\{C_j^B\}_1^m$，那么可建立一张n行、m列的图表，使当$v_l \in C_i^A$和$v_l \in C_j^B (l=1,2,\cdots,k)$，则$v_l$置在相应的$\alpha_i$行、$\beta_j$列的位置上，因此图表上有且仅有$k$个非空元素，它们是构成网络图的结点。然后，确定交叉位置图的路径，首先把同行或同列的结点用无向路径连结起来，然后修订网络的路，把有横、竖交叉连结路径的结点作为行（或列）的有向路径上的边缘结点，并确定路径的方向使之构成单向路。

例　设$A=\{\alpha_1,\alpha_2,\alpha_3\}$、$B=\{\beta_1,\beta_2,\cdots,\beta_6\}$，$C=\{v_1,v_2,\cdots,v_9\}$。

C分划为：

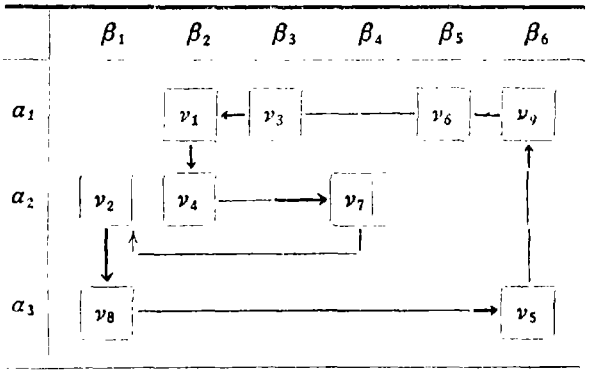
$C_1^A=\{v_1,v_3,v_6,v_9\}$，$C_2^A=\{v_2,v_4,v_7\}$，$C_3^A=\{v_5,v_8\}$和$C_1^B=\{v_2,v_8\}$，

$C_2^B=\{v_1,v_4\}$，$C_3^B=\{v_3\}$，$C_4^B=\{v_7\}$，$C_5^B=\{v_8\}$，$C_6^B=\{v_5,v_9\}$。

它们的交叉位置图如表2。

实际上，根据表2的有向路图便可得到相应的环状DNA酶切物理图谱（图1）。

表2



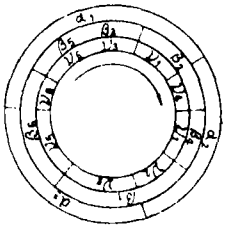


图　1

由实验模型的背景决定了交叉位置应遵循下面几项原则：

（1）　它应是全连通的简单有向路图，对于环状DNA是闭路，线性DNA是开路。

（2）　在n行m列的图表上每行(列)至少有一个结点，且每行(列)必有且仅有两个边缘结点。

（3） 若某行(列)中有多于一个边缘结点，则它们之间的路径不标方向，表示对应的那些片段未能判定相互先后的排序。

（4） 假若得到的交叉位置图不是全连通，则必遗缺了一些片段，只能得到相应于若干连通分支的片段物理图谱。

引入交叉位置图对构建和分析DNA酶切物理图谱起着显著的作用。

（1） 只要知道酶切片段的交叉位置,便可以得到DNA物理图谱，只是在确定图距时才依靠片段的分子量，这样对于实验的误差要求可以放宽，迁移率与分子量的数值换算误差不会影响图谱的构建。

（2） 对于分子量很小的片段，在凝胶上无法捕捉到，这往往造成一些片段的丢失，从而不能构成全连通的交叉位置图，但利用交叉位置图的构成原则可以估计出丢失片段的数目、位置和大小。

（3） 交叉位置图可用来检验实验的结果，提供了一种DNA物理图谱的精度评估的方式。

3 实例

选用武汉大学马延高等[7]的蓖麻核型多角体病毒，作为说明应用交叉位置图构建DNA酶切物理图谱的实例，并改进了文[7]分析得到的物理图谱。

蓖麻核型多角体病毒 ArNPV-DNA 是环状的 DNA。

已知交叉酶切的实验结果见表 3 — 6 。

表 3 Bgl Ⅱ与EcoR Ⅰ交叉酶切的结果

Tab.3 The restriction results of Bgl Ⅱ and EcoR Ⅰ

Bgl Ⅱ片段	用EcoR Ⅰ酶切后在双酶切电泳图上位置	EcoR Ⅰ片段	用Bgl Ⅱ酶切后在双酶切电泳图上的位置
B1	BE1,BE2,BE4	E1	BE1,BE10,BE11
B2	BE6,BE7,BE8,BE12	E2	BE2,BE12
B3	BE3,BE10,BE13	E3	BE3
B4	BE5,BE9	E4	BE5,BE13
B5	BE11	E5	BE4
		E6	BE6
		E7	BE7
		E8	BE8
		E9	BE9

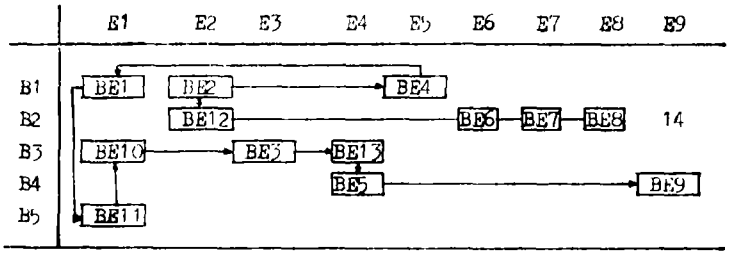
表4 Kpn I 与EcoR I 交叉酶切的结果

Tab.4 The restriction results of Kpn I and EocR I

Kpn I 片段	用EcoR I 酶切后在双酶切电泳图上的位置	EcoR I 片段	用Kpn I 酶切后在双酶切电泳图上的位置
K1	EK1,EK5	E1	EK1,EK3,EK10
K2	EK4,EK8	E2	EK8,EK12,EK13,EK14
K3	EK2	E3	EK5,EK9
K4	EK7,EK12	E4	EK2
K5	EK3	E5	EK4,EK17
K6	EK6	E6	EK6
K7	EK9	E7	EK7,EK18
K8	EK15,EK16	E8	EK11
K9	EK10	E9	EK15,EK16
K10	EK11		
K11	EK13		
K12	EK14		
K13	EK18		
K14	EK17		

表5 Bgl Ⅱ、EcoR Ⅰ酶切片段的交叉位置图

Tab.5 The cross network graph of Bgl Ⅱ,EcoR Ⅰ restriction fragments



由于这是一个全连通有向路图，便可以得到线状的DNA物理图谱：

E9　E4　E3　E1　E5　E2　(E6　E7　E8)

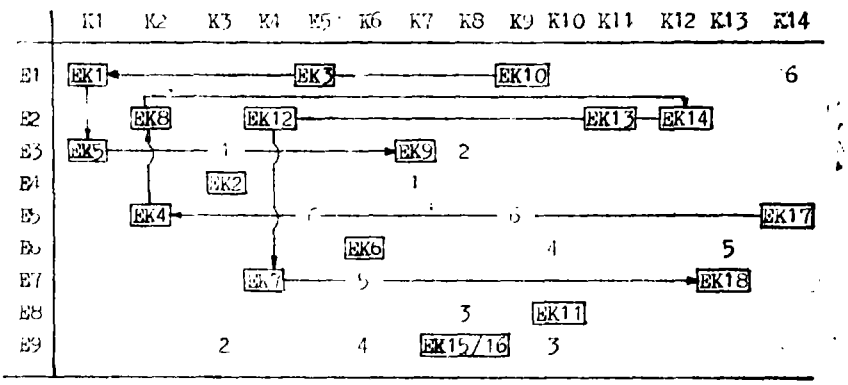
--- BE9 BE5 BE13 BE3 BE10 BE1 BE11 BE4 BE2 BE12(BE6 BE7 BE8)---

E4　B3　B5　B1　B2

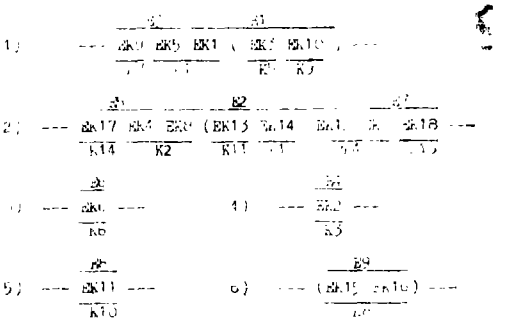
注意到环状DNA对应的交叉位置图应是一个全连通的闭路图，由于*BE*的片段数应是5 + 9 = 14，故知缺失一小片段BE14，可在表5上推知其位置。

表6 作Kpn I与EcoR I切段的交叉位置图

Tab. 6 The cross network graph of Kpn I, EcoR I restriction fragments



上图是由六个独立的连通分支构成的，它们分别对应六组片段的排序(见下)：



但由前面 Bgl Ⅱ、EcoR Ⅰ交叉酶切的DNA物理图谱知EcoR Ⅰ切段的排序，从而，可推得Kpn Ⅰ、EcoR Ⅰ交叉酶切的DNA物理图谱。

归纳上述结果可得 ArNPV-DNA的Bgl Ⅱ，EcoR Ⅰ，Kpn Ⅰ酶切的物理图谱，图谱为环状(见图2)，以完整 ArNPV 基因组为100%，以其1%为一个图谱单位。事实上，只是在确定图距时才依靠片段的大小。

这一结果使文[7]中给出的物理图谱得到了改进，进一步明确了E7，BE7，EK18和K13四个子片段的位置。

表6的交叉位置图不能环状全连通，原因是丢失了 (14 + 9) − 17 = 6 个 EK 子片段，由交叉位置图的构成原则可以估计出它们在表6个的位置，用数字标记。此处我们将EK15、EK16理解为同一个子片段。

图2中的“−”符号表示隐含有丢失子片段。

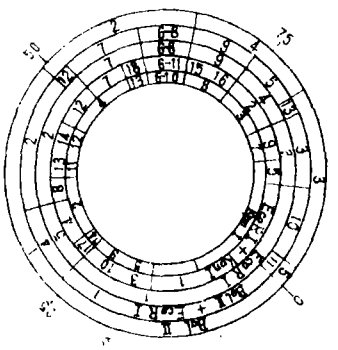


图2 ArNPV-DNA 的物理图谱

Fig. 2 ArNPV-DNA physicalapm

4 总 结

本文把DNA限制性内切酶物理图谱的构建归入一类特殊的带线性约束的0－1二次规划问题，这是进一步研究构建算法的基础，并且由于其形式和应用的特殊，相信它的算法和理论的深入研究，对其它属于这一类的应用问题也会有一定的意义。

文内引入了交叉位置图作为分析DNA物理图谱的工具，使构建DNA物理图谱的逻辑结构能简便地在网络图中表达出来，这对于分析DNA物理图谱及其实验，特别是运用计算机来辅助构建，都有着明显的意义。

交叉位置网络图方法的计算程序已由国家机械委员会广州电器科研所蔡耀编出。

参 考 文 献

〔1〕 Stefik,M., *Artificial Intelligence*,1978,11,85—114

〔2〕 Pearson,W. R., *Nucleic Acids Research*,1982,10,217—227

〔3〕 Water M. Fitch et al., *Gene*,1983,22,19—29

〔4〕 Claude V. Maina et al., *Nucleic Acids Research*,1984,12,717—729

〔5〕 王静怡等，华中师范大学学报，1986，20，321—330

〔6〕 Cai Yong,Chen Jicheng,*A Directed Graph Approach in Constructing DNA Restriction Maps*, International Conference of Bio—Mathematics, Xi'an,1988

〔7〕 马延高等，生物化学与生物物理进展，1985，2，39—43

The Mathematical Model and the Network Graph for the Construction of DNA Physical Maps

Cai Yong　　*Chen Jicheng**

Abstract

There are two main results. Firstly, we set up a (0,1)quadratic programming with linear contraints to simulate the construction of DNA restriction physical maps based on the experiements of cutting the DNA of larger viruses into many fragments. Secondly, we have designed cross network graph so as to construct and analyse the DNA restriction physical maps. Examples ars given to show the advantages of the approach.

Keyword DNA restriction physical map, mathematical programming, cross network graph

* Department of Computer Science