

物质相互作用的方程大为简化, 伴随而来的就是有关空间模式的形成、对称性的自发破缺、相位奇异点等有趣现象也就从讨论中消失。直到最近才有文章^[19-24]论及光学模式的横向弛豫以及时空混沌现象。这比单一时间混沌现象要复杂得多、要推广麦克斯韦-布洛赫方程, 使之包括有限光束截面引起的衍射损耗, 在传播过程中波面曲率的变化等因素并能解释模式跳变等现象。从利用、控制混沌现象来说, 对光学时空混沌的深入研究, 无疑有其积极意义。

- [1] A. G. Fox and T. Li, *Bell System Tech. J.*, **40** (1961), 453.
- [2] T. H. Maiman, *Phys. Rev. Lett.*, **4**(1960), 564.
- [3] A. Yariv, *Proc. IEEE*, **51**(1963), 1723.
- [4] W. E. Lamb Jr, *Phys. Rev.*, **134** (1964), A1429.
- [5] W. R. Bennett Jr., *Phys. Rev.*, **126**(1962), 580.
- [6] H. Statz and C. L. Tang, *J. Appl. Phys.*, **36** (1965), 1816.
- [7] W. R. Bennett Jr., In Grivet, P. and N. Bloembergen eds., *Proceedings of Third International Congress on Quantum Electronics*, Columbia Univ. Press, New York, 441-458.
- [8] H. Haken, *Handbuch der Physik*, Vol. XXV/2C, eds. by L. Genzel, Springer-Verlag, Berlin, (1969).
- [9] M. O. Scully and W. E. Lamb Jr., *Phys. Rev.*, **159**(1967), 208.
- [10] 谭维翰, *物理*, **10**(1981), 193.
- [11] 固体激光导论编写组, *固体激光导论*, 上海人民出版社, (1974), 150, 476, 463.
- [12] H. Statz and G. deMars, in *Quantum Electronics*, C. H. Townes ed., Columbi Press, New York, (1960), 530.
- [13] E. N. Lorenz, *J. Atmos. Sci.*, **20**(1963), 130.
- [14] Heing G. Schuster, *Deterministic Chaos*, Weinheim Physik-Verlag, (1984), 92, 172.
- [15] H. Poincare', *Les Me' thodes Nouvelles de la Me' chanique Celeste*, Gauthier-Villars, Paris, (1892).
- [16] H. Haken, *Phys. Lett. A*, **53**(1975), 77.
- [17] P. W. Milonni et al., *Chaos in Laser-Matter Interactions*, World Scientific Publishing Co. Pte Ltd, (1987), 117.
- [18] C. O. Weiss and J. Brock, *Phys. Rev. Lett.*, **57** (1986), 2804.
- [19] L. A. Lugiato et al., *J. Opt. Soc. Am. B*, **7** (1990), 1019.
- [20] L. A. Lugiato et al., *Opt. Comm.*, **69**(1989), 387.
- [21] M. Bramlilla et al., *Phys. Rev. A*, **43** (1991), 5090.
- [22] P. Conllet et al., *Opt. Comm.*, **73**(1989), 409.
- [23] F. T. Arecchi et al., *Phys. Rev. Lett.*, **65**(1990), 2531.
- [24] K. Ikeda et al., *Prog. Theor. Phys. Suppl.*, **99** (1989), 295; K. Otsuka, *Phys. Rev. Lett.*, **65** (1990), 329.

DNA 与 分 形

肖 奕

(华中理工大学物理系, 武汉 430074)

生命体中脱氧核糖核酸(DNA)分子的碱基序列记载着它生长发育的全部指令, 是一部人类至今还没有读懂的“天书”。它是一种表面上看起来不规则但实际上蕴含着有序结构的符号序列。对物理学来说, 这是一个典型的复杂系统。最近的研究发现, DNA 碱基序列也具有自相似性, 存在高度的长程关联。介绍近年来在利用分形理论对 DNA 碱基序列的研究中, 发展起来的二种定量分析符号序列的方法, 即度量空间的 DNA 行走和 Voss 的分解子序列方法。

生物大分子(DNA 和蛋白质)是自然界中典型的复杂系统, 已引起越来越多的物理学家的兴趣。最近 Peng 等^[1], Voss^[2] 和 Li 等^[3] 分别在“*Nature*”, “*Physical Review Letters*” 和 “*Europhysics Letters*” 上报道, DNA 碱基

序列中存在达数千个碱基位置的长程关联。这引起了人们的极大关注。早在 1988 年, 我国内蒙古大学的罗辽复和蔡禄^[4] 就计算了一些有机体的 DNA 的分维, 发现分维与进化水平保持很好的相关性, 而且也指出了核酸序列的长程

关联。本文简要地介绍在这些工作中发展起来的处理不规则符号序列的一些方法以及揭示出的 DNA 碱基序列的新性质。为此,我们首先介绍 DNA 结构方面的一些知识。

一、DNA 和 RNA 的结构

生命运动千姿百态、绚丽多彩。然而描绘出它的语言却是异常地简单,仅由四种化学字母组成。像人类语言一样,生命信息也是通过这些字母的线性排列来表达的。当然它不是写在纸上,而是记录在核酸分子中的化学链上。生物体中的核酸主要有 DNA 和核糖核酸 (RNA)。DNA 是基因的物质基础,记录了生物体携带的所有生命信息。RNA 则是指导蛋白质的合成。

DNA 是一种很长的大分子链。它由四种基本单位(核苷酸)组成。每种核苷酸包含一个磷酸基(P),一个脱氧核糖(S)和一个碱基。四种核苷酸中的磷酸基和脱氧核糖都相同,只是碱基不同。这四种碱基分别是: A(腺嘌呤), G(鸟嘌呤), C(胞嘧啶)和 T(胸腺嘧啶)。它们是核苷酸的核心,也就是生命语言中的四种

GCTTCTAACTTTACTCAGTTCGTTCTCGT
CGACAATGGCGGAAGTGGCGACGTGACT
GTCGCCCCAAGCAACTTCGCTAACGGGGT
CGCTGAATGGATCAGCTCTAACTCGCGTT
CACAGGCTTACAAAGTAACCTGTAGCGTT
CGTCAGAGCTCTGCGCAGAATCGCAAATA
CACCATCAAAGTTCGAGGTGCCTAAAGTGG
CAACCCAGACTGTTGGTGGTGTAGAGCTT
CCTGTAGCCGCATGGCGTTTCGTACTIONAAA
TATGGAATAACCATTTCCAATTTTCGCTA
CGAATTCGACTGCGAGCTTATTGTTAA
GGCACTGCAAGGTCTCCTAAAAGATGGA
GGCCC GATTCCCTCAGCAATCGCAGCAA
ACTCCHHCATCTAC

图 1 一种噬菌体的外壳蛋白质合成基因的碱基排列顺序
物理

字母。DNA 中的生命信息就是通过这些碱基的序列表达出来的。图 1 是一种噬菌体的外壳蛋白质合成基因的碱基排列顺序。

DNA 分子由两条核苷酸链组成。每条链中一个核苷酸的磷酸基与另一个核苷酸的糖相互结合形成糖-磷酸骨架,碱基则从骨架垂直伸出。两条链中伸出的碱基通过氢键相互配对形成像梯子一样的结构(图 2),且扭成著名的双螺旋形状。但这种配对是有规则的,即 A 只能与 T 配对, G 只能和 C 配对。

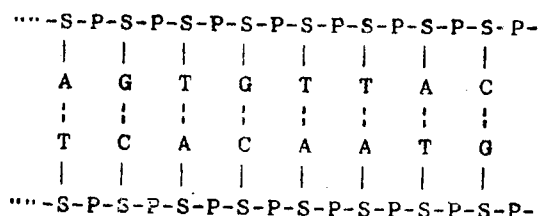


图 2 DNA 结构的示意图

生物是通过各种各样的蛋白质来完成它的各种生命活动的。这些蛋白质的结构由 DNA 决定,但 DNA 并不直接参与蛋白质的合成。这是因为 DNA 是在细胞核内,而蛋白质的合成在核外。DNA 就像建筑蓝图,具体的施工者则是 RNA。我们通常说的基因就是 DNA 中确定一个 RNA 分子的一段碱基序列。一个 DNA 分子可以确定许多 RNA 分子。这些 RNA 分子中有些用于确定蛋白质中的氨基酸排列顺序(mRNA),有些用做合成蛋白质的工具(如 tRNA)。RNA 的结构与 DNA 相似,只是用核糖取代了脱氧核糖,用 U(尿嘧啶)取代了 T,且只有一根单链。mRNA 中的碱基序列是以 DNA 中的一条链为模板根据配对规则转录过来的,并通过所谓的三联密码(即三个一组的碱基)来指导氨基酸按一定顺序合成特定的蛋白质。mRNA 中三个三个一组碱基的顺序就被翻译成蛋白质中氨基酸的顺序。但有的基因中存

在一段或数段称为内含子的碱基序列。内含子并不被翻译成氨基酸序列，它们在转录成 mRNA 后被除去。因此在有内含子的基因中，决定蛋白质的信息被内含子分隔成了一些不连续的段(称为外显子)。

目前生物学家已经测出了大量的 DNA 碱基序列。正在进行的人类基因组计划，将在十几年内确定人类基因的所有碱基序列，从而得到人体的整套指令。然而面对这部生命之书，我们却知之甚少，我们还不知道它的语法规则。从图 1 中的碱基序列中，我们看不出碱基的排列有什么规律，但我们相信它不是无穷的，不然就不可以产生如此有序的生命运动。研究表明，DNA 碱基序列中一个或数千碱基的改变，就可能使合成的蛋白质的功能发生改变，甚至完全丧失。这是因为蛋白质的功能决定于它的空间构像，而空间构像又由它的氨基酸序列决定。改变 DNA 碱基序列中的碱基，必定改变蛋白质中的氨基酸。正是由于 DNA 中碱基排列顺序的重要性，人们已做了大量的研究试图揭示其中的有序结构和它产生的模式。近年来随着分形理论的发展，为我们处理不规则符号序列提供了一种新的有效的方法，并揭示出了 DNA 碱基序列的一些重要性质。这方面的工作才刚刚开始，值得我们进一步去研究。

二、DNA 行走

不知你有没有听说过 DNA 音乐。有人曾把碱基与音符对应起来，根据不同的 DNA 碱基序列谱成了一曲曲美妙的音乐。比如病毒的 DNA 音乐听起来非常低沉，而另外一些生物的 DNA 音乐听起来则欢快明亮。在把 DNA 碱基序列转变成一首乐曲时，必须确定一套碱基与音符之间的对应规则。我们在利用分形理论分析 DNA 碱基序列时也遇到类似的问题。DNA 碱基序列是四种符号组成的符号序列，直接处理符号序列比较困难。因此，一种很自然的方法是把它映射为度量空间的一种不规则行走，也即把每个碱基映射为度量空间的一个矢

量，并按照 DNA 中的排列顺序加起来。这种不规则行走就称为 DNA 行走。根据映射规则的不同，我们可以得到不同维数空间的 DNA 行走。

1. 一维 DNA 行走

从化学上讲，DNA 中四种碱基可以分为两类：嘌呤(A 和 G)和嘧啶(C 和 T)。在这种意义上，我们可以认为 DNA 碱基序列只有嘌呤和嘧啶两种类型的碱基组成。这样就可以把 DNA 碱基序列映射为一维的不规则行走。在通常的一维随机行走中，向正方向和负方向走一步(单位长度)的几率相等。一维的 DNA 行走则定义为：沿 DNA 链的第 i 个位置如果是嘧啶，则向正方向走一步；如果是嘌呤，则向负方向走一步。图 3 分别是一种富内含子基因，它的 cDNA (去除内含子的 mRNA 的互补 DNA) 和一种较少内含子基因的一维 DNA 行走表示。

这种 DNA 行走为每个基因提供了一种简单的直观表示，而且使我们能直接分析 DNA 碱基序列中的关联。Peng 等发现 DNA 中的

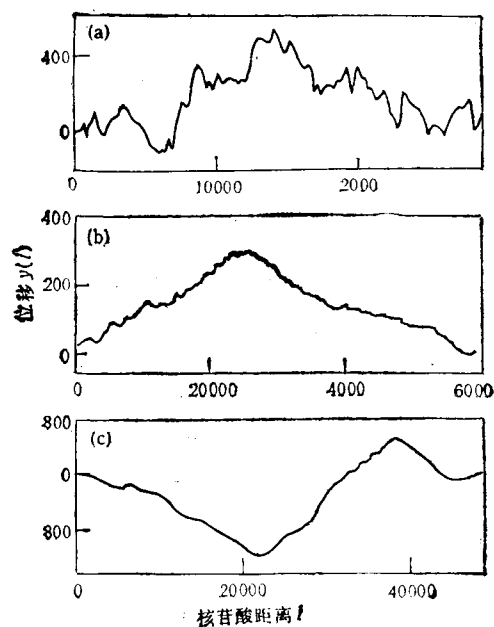


图 3 富内含子的人 β -心肌球蛋白重链基因序 (a)，它的 cDNA (b) 和较少内含子的噬菌体 λ DNA 序列 (c) 的一维 DNA 行走表示

长程关联正是采用的一维 DNA 行走^[1]。定量表示关联的方法是计算行走 l 步后的净位移 $y(l)$ 。再进一步可计算关于位移平均值的均方差涨落 $F(l)$ 。 $F(l)$ 是描述任意行走中关联的一个重要统计量。利用计算出的 $F(l)$ 可以区分三种不同的行为：(1) 如果碱基序列是完全随机的，那么 $F(l) \sim l^{1/2}$ ；(2) 如果在某特征长度内存在局域关联(像 Markov 链那样)，它的渐近行为仍然和纯随机情况一样，即 $F(l) \sim l^{1/2}$ ；(3) 如果不存在特征长度(即关联是“无限长”)，那么涨落 $F(l)$ 与 l 的关系为幂次形式， $F(l) \sim l^\alpha$ ，其中 $\alpha \approx \frac{1}{2}$ 。这种情况表示 DNA 行走具有自相似(分形)性。

Peng 等分析了一些基因的碱基序列。他们发现富含内含子基因中的碱基间存在长程关联，而且更惊奇地是这种长程关联可达数千个碱基位置。到目前为止，还没有人能解释这种现象。与此相反，在 cDNA 和较少内含子的基因中，他们得到 $\alpha \approx \frac{1}{2}$ ，即它们中不存在长程关联。Peng 等认为，通过计算 α ，可以为区分富含内含子与较少内含子的基因提供一种新的定量的方法。

然而也有不少学者提出不同的看法^[5-8]。一些人认为 $F(l)$ 与 l 的幂次关系 ($\alpha > \frac{1}{2}$) 也可以有另外解释，例如是由 AT 和 CG 分布不均匀造成的，不一定意味着长程关联。一些人认为 Peng 等的方法并不能一般地区分富含内含子和较少内含子或无内含子的基因。

2. 二维 DNA 行走

Peng 等的方法虽然十分简单，但它忽略了 A 和 G, C 和 T 之间的差别，这显然是个缺点。二维 DNA 行走则区分了四种不同的碱基。它是由 Gates 最先提出来的^[9]。他利用了 A 和 T, G 和 C 互补的性质(配对规则)，把 DNA 碱基序列映射成二维空间的不规则行走，其中 T 和 A 分别映射为 X 轴正、负方向的单位矢量，G 和 C 分别映射为 Y 轴正、负方向的单位矢量。也即，如果某步对应的碱基为 T(C)，则向 X 轴正

(负)方向走一步(单位长度)；如果为 G(C)，则向 Y 轴的正(负)方向走一步。把这些矢量按 DNA 碱基的排列顺序加起来，就形成了二维 DNA 行走。图 4 是两种 DNA 碱基序列的二维 DNA 行走表示。

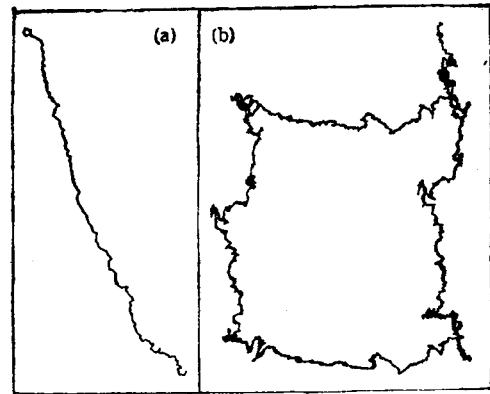


图 4 (a) 艾滋病病毒和 (b) 玉米的一 DNA 碱基序列的二维 DNA 行走表示

从图 4 可以看出，二维 DNA 行走比一维 DNA 行走能更直观地显示出 DNA 碱基序列中的结构。利用二维 DNA 行走，我国学者罗辽复等^[4]研究了 14 种有机体的分维与它们进化复杂性间的关系。他们计算了二维 DNA 行走中相邻 m 个碱基的平均端一端距离随 m 的变化，得到了它们的分维。他们发现从线粒体的分维 1.2 到哺乳类的分维 1.6，分维与进化间保持一定的统计相关性。他们的结果也说明了 DNA 中存在长程关联。

3. 三维和四维 DNA 行走

Li 和 Kaneko^[3]在分析 DNA 序列时采用的是三维 DNA 行走。他们把四个碱基分别映射为三维空间的单位矢量，每秒碱基的分量定义为

$$\begin{aligned} A: & \{ 1, 0, 0 \} \\ C: & \{ -0.333333, 0, 0.942809 \} \\ G: & \{ -0.333333, -0.816497, -0.471405 \} \\ T: & \{ -0.333333, 0.816497, -0.471405 \} \end{aligned}$$

这里他们利用了四种碱基的对称性，把它们分别与正四面体的四个顶角对应起来。他们的结果也证明了富含内含子基因中碱基存在长程关

联。而且认为可能是序列中存在重复结构产生的。

美国 Utah 大学的 Berthelsen 等则采用了更复杂的四维 DNA 行走^[10]。他们考虑了相邻碱基的关联，把碱基对分别映射成四维空间的单位矢量或零矢量，他们的这种方法可以推广到考虑任意近邻碱基间的关联，但计算变得很复杂。

三、Voss 的分解子序列方法

DNA 行走非常直观，但对由四种字母组成的碱基序列，小于四维的 DNA 行走都会人为地引入一些关联。例如，在一维 DNA 行走中把两种碱基看成相同，Berthelsen 等虽然采用了四维行走，但表示的是相邻碱基的关联。为了在分析符号序列的关联时避免引入人为的关联，Voss 发现了一种非常简单的方法。我们称它为分解子序列方法。

Voss 的方法是把 DNA 碱基序列分解为 A, C, G 和 T 四个子序列。每个序列表示一种碱基在原 DNA 序列中的位置。例如，A 子序列是在原 DNA 碱基序列中用 1 代替 A，用 0 代替其他三种碱基得到的 1 和 0 的数字序列（比如 AGCATACTG 序列的 A 子序列为 100101000）。其他三个子序列也用相同的方法得到。图 5 是人的一 DNA 碱基序列的四个子序列。

对于不规则过程，通常采用关联函数和谱密度来描述它的关联程度。Voss 方法的奇特之处在于可以证明原 DNA 碱基序列的关联函数是分解成的四个子序列的关联函数之和。同

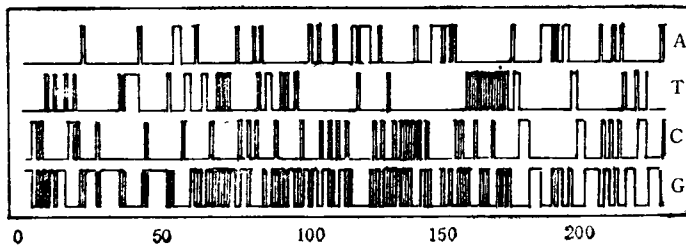


图 5 人基因中的一段 DNA 碱基序列的四个子序列的一部分

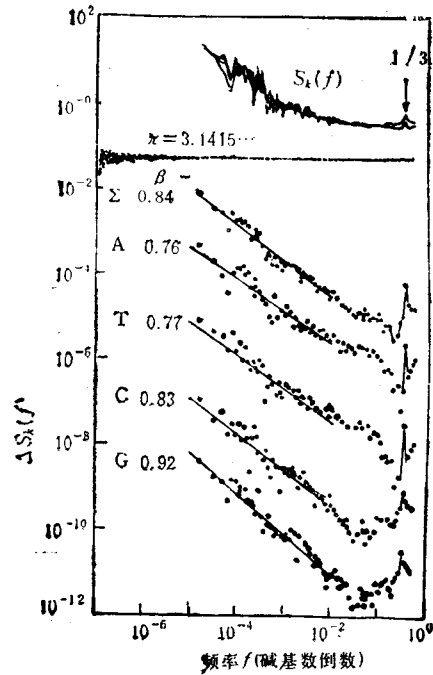


图 6 人的一 DNA 碱基序列的谱密度

样的结论对谱密度也成立。这样通过求子序列的关联函数和谱密度，来得到原 DNA 序列的关联函数和谱密度，而不用假定四种碱基间有任何关系。

Voss 用这种方法研究了 25000 多 DNA 序列，证明了确实存在长程关联。图 6 是图 5 所示的碱基序列的谱密度。图中最上面的曲线是四个子序列的谱密度 $S_k(f)$ ，可以看出它们的变化趋势基本相同。在 f 比较大时， $S_k(f)$ 主要是白噪声，而且在 $f = 1/3$ 时有一尖峰，这可能对应于三联密码。在 f 较小的区域， $S_k(f)$ 表现出类似于 $1/f$ 噪声的幂次形式 $S_k(f) \sim 1/f^\beta$ 。为了比较，图中还给出了 π 的前 1.13×10^9 个

数的十个子序列的 $S_k(f)$ 。在许多情况下，我们可以假定 $1/f^\beta$ 噪声和高 f 白噪声是独立的，这样可从 $S_k(f)$ 中减去高 f 白噪声， $\Delta S_k(f) = S_k(f) - S(\infty)$ ，而更清楚地显示出低 f 区域中的标度不变性。图 6 中下面的几条曲线就是四个子序列的 $\Delta S_k(f)$ 和它们

的和。从中我们可以清楚地看到 $\Delta S(f) \propto 1/f^\beta$ 的标度区从 100 到 100000 个碱基不等。四个子序列的谱指数 β 值也不同。

从上面我们看到, DNA 行走比较直观,除了能定量研究关联外,还能直接显示碱基序列中的有序结构,所以生物学家对它比较偏爱。Voss 的方法可以避免人为引入的关联,而且还能推广到含任意种符号的符号序列,因此在定量分析方面将会有广泛地应用。

用分形理论分析 DNA 碱基序列在我国开展得也比较早。除内蒙古大学的罗辽复等外,还有中国科学院生物物理研究所的陈润生等。最近陈润生等还利用密码学来研究 DNA 碱基序列,已经取得了一些很有意义的结果。DNA 分子需要从多角度进行研究。物理学,特别是

复杂系统物理学,不但在这里有用武之地,而且还能从这里吸取自然之精华,使其自身得到丰富和发展。

- [1] C. -K. Peng et al., *Nature*(London), **356**(1992), 168.
- [2] R. F. Voss, *Phys. Rev. Lett.*, **68** (1992), 3805.
- [3] W. Li and K. Kaneko, *Europhys. Lett.*, **17**(1992), 655.
- [4] L. Luo and L. Tsai, *Chin. Phys. Lett.*, **5**(1988), 421.
- [5] S. Nee, *Nature (London)*, **357**(1992), 450.
- [6] V. V. Prabhu and J. -M. Claverie, *Nature (London)*, **359**(1992), 782.
- [7] C. A. Chtzidimitrou-Dreismann and D. Larhammar, *Nature (London)*, **361**(1993) 212.
- [8] S. Karlin and V. Brendel, *Science*, **259** (1993), 677.
- [9] M. A. Gates, *J. Theor. Biol.*, **119**(1986), 319.
- [10] C. L. Berthelsen et al., *Phys. Rev. A*, **45**(1992), 8902.

硅气凝胶和它的分形结构¹⁾

沈军 王珏 吴翔

(同济大学玻耳固体物理研究所,上海 200092)

硅气凝胶是从溶胶-凝胶工艺衍生出来的非晶态纳米多孔性物质。它是研究分形结构最理想的材料。研究分形结构最有效的实验手段是小角散射。小角散射实验表明,硅气凝胶由尺寸为 a ($a \sim 1\text{nm}$) 的颗粒组成。这些颗粒组成分形网络,分形结构持续到尺度 ξ ($\xi \sim 100\text{nm}$)。在 ξ 以上的长度范围,材料可视为均匀的。介绍了硅气凝胶的制备、结构及结构测量。

关键词 硅气凝胶,小角散射,分形结构

Abstract

Silica aerogels are sol-gel derived solid materials with extraordinarily good properties resulting from their nanoporosity. They are the best materials to investigate the fractal structures. The most powerful and non-destructive methods for probing the structures are small angle scattering techniques. The experimental results show that the silica aerogels are consisting of a network of particles which have the size on the order of a ($a \sim 1\text{nm}$), and the network has self-similarity or dilation symmetry up to the length of ξ ($\xi \sim 100\text{nm}$); above this length scale, the material is continuous and homogeneous. This paper describes the preparation of silica aerogels and the investigation of their structures.

Key words silica aerogels, small angle scattering, fractal structure

1) 国家自然科学基金资助项目。