

单分子混合物：聚合物科学中的一个新概念

唐宇^{1*}

¹ 中国科学院上海有机化学研究所 生命过程小分子调控全国重点实验室

(E-mail: tangyu@sioc.ac.cn; ORCID: 0000-0002-4272-2234)

摘要：从理论上讲，物质存在两种极端形式：纯态和单分子混合态。后者包含结构各异的分子混合物。受“化学空间”概念的启发，本文结合模型构建与数学分析，对单分子混合态进行了研究，并获得了一些有趣的结果。这些结果提供了理论依据，表明单分子混合态确实可能存在于现实中的合成或天然聚合物体系中。

关键词：单分子混合物；聚合物

本文的英文版已于 2024 年 7 月发表于 *International Journal of Molecular Sciences* 杂志, Tang, Y. Single-Molecule Mixture: A Concept in Polymer Science. *Int. J. Mol. Sci.* **2024**, 25, 7571. 原文链接: <https://doi.org/10.3390/ijms25147571>

1. 引言

现实中的物质总是由多种分子混合而成的，理论上，物质存在两种极端形式——纯态和单分子混合态，如图 1 所示。前者仅含一种分子，而后者则由具有不同分子结构的混合物组成。据我所知，这种难以捉摸的分子形态在现有文献中尚未被探讨。在探讨“单分子混合物”时，有两个有趣的问题浮现出来：(1) “单分子混合物”状态的分子是否存在于自然界或合成体系中？(2) 是否有任何可能的方法来合成“单分子混合物”状态的分子？为探索这些问题，本文尝试建立一种结合模型构建与数学分析的理论方法，以研究这种分子状态，并提供理论依据支持以下结论：此类混合物可能确实广泛存在于自然界或合成聚合物体系中。此外，本文还提出了一条合成“单分子混合物”状态分子的可行路线。

Two extreme states of molecules

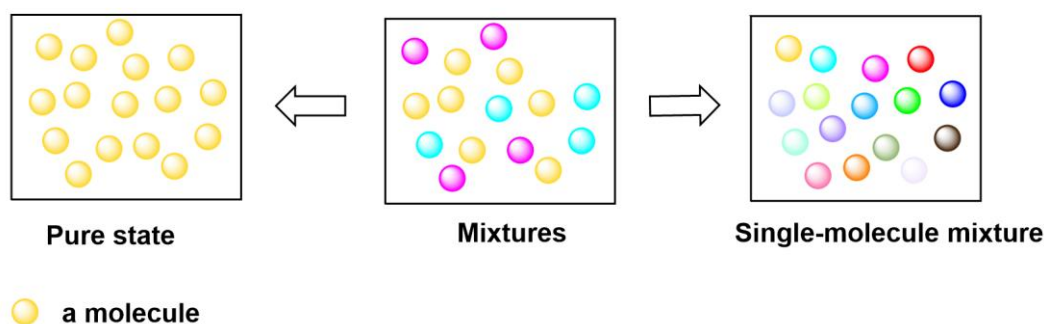


图 1. 物质的两种极端状态：纯态和单分子混合态

自 21 世纪初以来，“化学空间”这一概念逐渐引起了科学界的关注¹⁻⁴。化学

空间——涵盖所有可能的小分子有机化合物，包括生物系统中存在的那些——极其广阔。例如，Dobson¹估算指出，鉴于存在 20 种不同的氨基酸类型，且天然蛋白质的平均残基数约为 300 个，因此可能的异构体总数高达 20^{300} ，甚至超过 10^{390} ；若要合成每种多肽各一分子，其总质量将远超已知宇宙的质量。受这些惊人结果的启发，我设想：如果先构建一个包含极大量可能异构体的模型聚合物空间，然后（随机）合成仅包含该空间中极小部分结构的分子，那么获得处于“单分子混合物”状态的分子的概率将非常高。这一设想构成了本研究的基本设计原则。

2. 结果与讨论

为了开展这项研究，首先构建了一个模型聚合物分子系统，如图 2 所示。其中，可取代位点的总数为 a ，每个位点随机被两种可能的基团之一 (R^1 或 R^2) 取代。因此，该聚合物空间中的结构异构体总数为 2^a ，且随着 a 的增大而呈指数级增长。例如，若 $a=200$ ，则该空间中的异构体总数为 2^{200} ，即 $1.60693804 \times 10^{60}$ 。

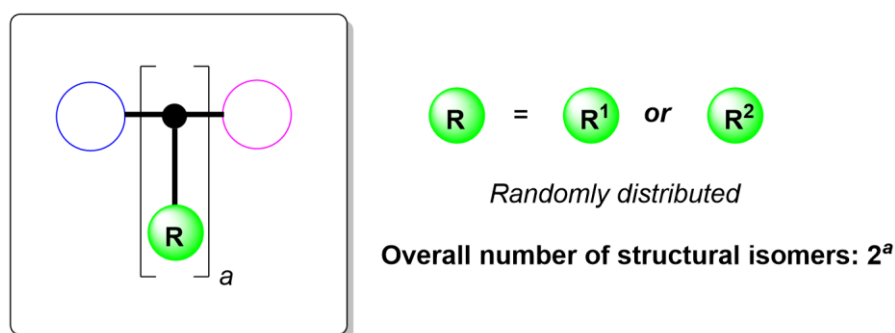


图 2. 一个典型的聚合物分子体系

由此产生了一个非常有趣的问题：如果我从上述结构空间（总结构数为 m ）中单独且随机地选取一定数量 (n) 的结构（即同一结构可能被选取多次），该如何计算得到单分子混合物（其中每个结构各不相同）的概率 (P)？

这个问题可以通过计算概率来解决。

$$P = \frac{C_{m-n}^n}{C_m^n} = \frac{(m-n)(m-n-1)(m-n-2)\cdots(m-n-n)}{m(m-1)(m-2)\cdots(m-n)}$$

$$= \left(1 - \frac{n}{m}\right) \left(1 - \frac{n}{m-1}\right) \left(1 - \frac{n}{m-2}\right) \cdots \left(1 - \frac{n}{m-n}\right)$$

$$\text{若 } m > 2n, \text{ 则, } P > \left(1 + \frac{n}{n-m}\right)^n \text{ 且 } \frac{n}{n-m} > -1 \quad (\text{Eq.1})$$

然后，可以利用伯努利不等式将 $\left(1 + \frac{n}{n-m}\right)^n$ 近似为：

$$\left(1 + \frac{n}{n-m}\right)^n > 1 + n \times \frac{n}{n-m}$$

在此情况下，我选择 $n = N_A (6.02 \times 10^{23})$ 作为从结构空间中选取的分子数，

$a = 200$ 作为聚合物模型系统中的取代位点数 (因此, $m = 2^{200} = 1.60693804 \times 10^{60}$)。因此, 获得单分子混合物的概率 (P) 如下:

$$P > \left(1 + \frac{n}{n-m}\right)^n > 1 + n \times \frac{n}{n-m} \approx 1 - 2.26 \times 10^{-13}$$

这一结果表明, 如果我们从该模型结构空间 ($a = 200, m = 1.60693804 \times 10^{60}$) 中单独且随机地选取 $n = N_A$ (6.02×10^{23}) 个分子, 则获得单分子混合物的概率 (P) 将接近 1。

根据方程 (1), P 与 m 之间的关系为

$$m < n + n^2 \times \frac{1}{1-P} \quad (\text{Eq. 2})$$

由式(2)可推算, 对于给定的 P 值, 模型结构空间中结构异构体的最小数量 m 应大于 $n + n^2 \times \frac{1}{1-P}$ 。

已有研究指出, 在许多实际的合成聚合物体系和天然生物聚合物体系中, 母体分子常发生部分随机取代, 例如聚合物溴化过程^{5,6} (氢原子被溴原子随机部分取代)、许多其他聚合后修饰过程⁷⁻¹⁵, 以及 DNA¹⁶ 和蛋白质的甲基化过程 (氢原子被甲基随机部分取代)^{17,18}, 如图 3 和图 S1 所示。这种取代过程会产生大量概率相等的异构体。为了精确计算潜在异构体的数量, 我构建了两个额外的模型体系, 如图 4 所示。

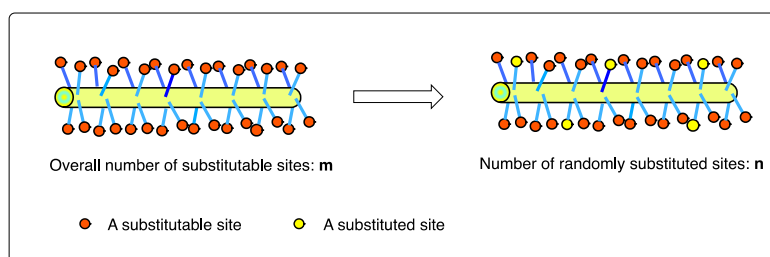


图 3. 含有 m 个可取代位点的母体分子被 n 个取代基取代的部分随机取代过程

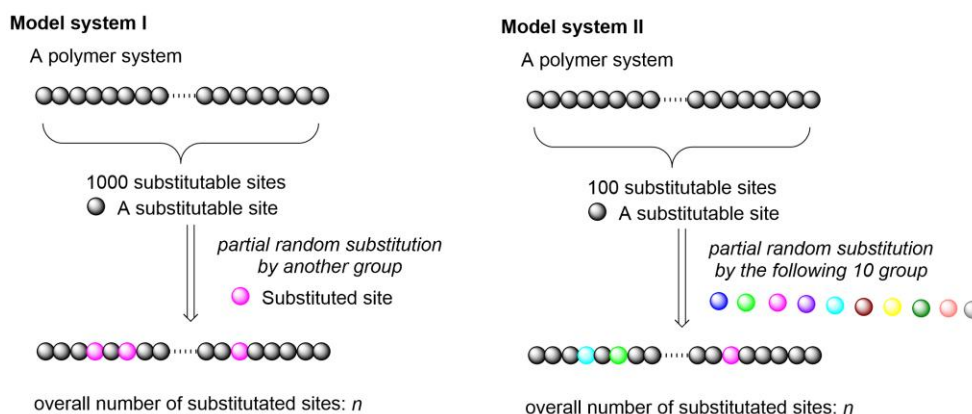


图 4. 展示模型系统 I 和 II 的示意图

如果母体分子中可取代位点的总数为 m, 被取代位点的总数为 n, 且每个随

机被取代位点中可能的取代基数为 s ，那么该结构空间 r 中潜在异构体的总数可按以下方式计算：

$$r = C_m^n \times s^n \quad (\text{Eq. 3})$$

在模型系统 I 中，我选择了 $m = 1000$ 和 $s = 1$ ，这意味着母体分子含有 1000 个可取代位点。如果其中几个位点被另一个基团随机取代 ($R_0 \rightarrow R_1$)，则潜在异构体的总数可按以下方式计算：

$$r = C_{1000}^n \quad (\text{Eq. 4})$$

在模型系统 II 中，我选择了 $m = 100$ 和 $s = 10$ ，这意味着母体分子含有 100 个可取代位点。如果其中几个位点被十种可能的 R 基团之一随机取代 [$R_0 \rightarrow (R_1 \text{ 至 } 10)$]，则潜在异构体的总数可按以下方式计算：

$$r = C_{100}^n \times 10^n \quad (\text{Eq. 5})$$

表 S1 和表 S2 列出了计算得到的潜在异构体(r)的总数 ($n = 0 - 25$)，图 5 展示了 $\lg r$ 与 n 之间的关系。

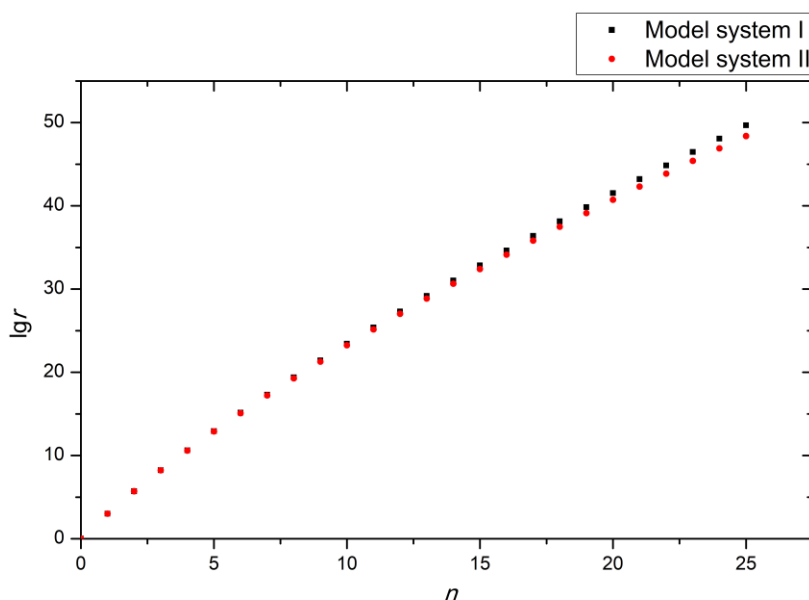


图 5. $\lg r$ 与 n 之间的关系

从图 5 可以看出，对于这两个模型体系，潜在异构体的总数会随着取代位点的数量呈指数级增长。若 $n = 25$ ，则模型体系 I 和 II 的潜在异构体总数分别为 4.76×10^{49} 和 2.43×10^{48} 。根据式(1)，当从这两个结构空间中取 1 mmol (6.02×10^{20}) 分子时，获得单分子混合物的概率（模型系统 I 和 II 分别记为 P_1 和 P_2 ）如下：

$$P_1 > \left(1 + \frac{n}{n-m}\right)^n > 1 + n \times \frac{n}{n-m} \approx 1 - 7.61 \times 10^{-9}$$

$$\text{和 } P_2 > \left(1 + \frac{n}{n-m}\right)^n > 1 + n \times \frac{n}{n-m} \approx 1 - 1.43 \times 10^{-7}$$

这些计算结果表明，这两个概率值足够大，足以使单分子混合物在毫摩尔量

级上存在。

根据方程 (2), 我们可以估算: 如果从结构空间中取 1 毫摩尔 (6.02×10^{20}) 分子, 获得单分子混合态的概率将超过 0.999; 因此, 模型结构空间中的结构异构体最小数目 r 应超过 $n + n^2 \times \frac{1}{1-p} \approx 3.62404 \times 10^{44}$; 该数值远小于上述两个空间的实际异构体数量。

因此, 我们可以得出结论: 如果一个聚合物分子含有 1000 个可取代位点, 且其中 2.5% 的位点被另一基团随机取代, 则潜在异构体数超过 4.76×10^{49} ; 如果该取代过程以毫摩尔量级进行, 则产物最可能以单分子混合态存在。

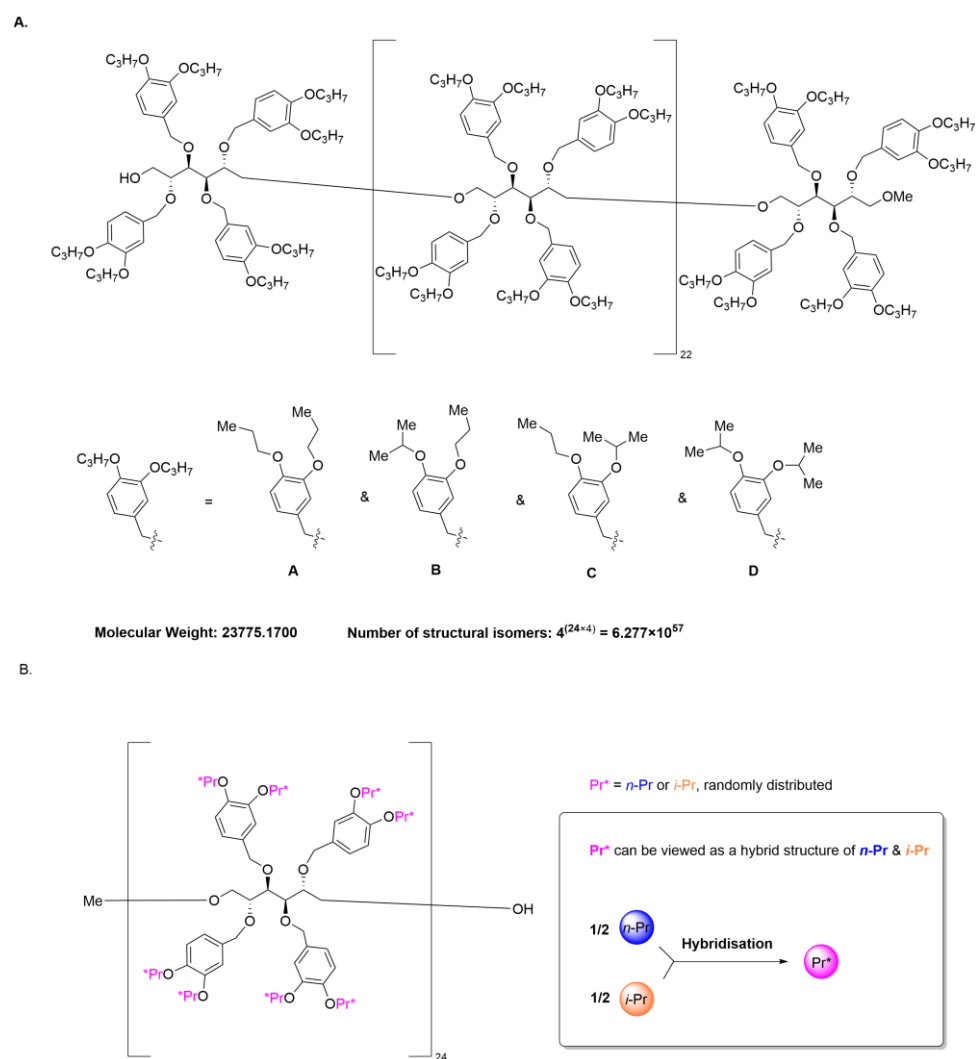


图 6. (A) 一个由 24 个单体组成的 O-丙基取代 D-甘露醇模型体系 (B) 单分子混合模型体系的混合分子结构

最后, 为了更好地阐明单分子混合态聚合物的结构特征, 构建了一个由 24 个单体组成的 O-丙基取代 D-甘露醇模型体系, 如图 6A 所示 (该 24 聚体的可能合成路线见补充信息, 参见图 S2)。该体系含有 192 个酚羟基, 这些羟基随机被正丙基或异丙基取代。该体系的分子量为 23,775.1700 Da, 结构异构体数量为 2192

种——即 6.277×10^{57} 种。根据方程(1)，即使我们以 100 摩尔规模（吨级）制备该聚合物，获得单一分子混合物的概率(P)也将如下：

$$P > \left(1 + \frac{n}{n-m}\right)^n > 1 + n \times \frac{n}{n-m} \approx 1 - 5.77 \times 10^{-7}$$

在此模型体系中，尽管每个分子结构各不相同（因各取代位点上的 R 基团不同），但它们均随机源自同一结构空间，这一构想源自轨道杂化理论¹⁹⁻²²。如果我们将 Pr* 定义为 *n*-Pr 和 *i*-Pr 的杂化结构，那么该聚合物体系可被视为一种“纯物质体系”，因为每个酚羟基都会被一种前所未有的、杂化的 Pr* 基团取代，如图 6B 所示。因此，在某种程度上，该聚合物体系的某些物理性质预计会与纯物质体系相似。然而，单分子状态对天然及合成聚合物物理、化学和生物性质的影响，仍是需要进一步研究的课题。

3. 材料与方法

本文所述工作是一项结合了模型构建、数学分析和思想实验的理论研究。模型构建使用 ChemOffice 软件(PerkinElmer ChemOffice 2020 - Version 20.0)完成，概率近似采用伯努利不等式，详细计算则使用免费的在线科学计算器在普通家用计算机上完成（Desmos, <https://www.desmos.com/scientific?lang=zh-CN>, 访问于 2024 年 7 月 8 日）

4. 结论

综上所述，通过结合模型构建与数学分析，对与“绝对纯净”这一极端分子状态相对的单分子混合物进行了理论研究，并获得了一些有趣的结果。这些结果支持了以下结论：单分子混合物状态可能广泛存在于现实中的合成或天然聚合物体系中。

化学微不均一性——即在微观层面呈现不均一性的特性——是所有生命系统以及大多数软材料和晶体材料的共同特征^{23,24}。根据 Gentili 等人²³最近的分类，天然和合成聚合物系统是五类主要微不均一性系统之一。尽管人们早已认识到微不均一性对糖蛋白等生物聚合物功能具有重大影响²⁵，但“单分子混合物”这一概念以及本研究的成果，将为在分子层面理解“微不均一性”提供一种独特途径。希望本研究能激发对这一引人入胜领域进行进一步探索。

补充材料

补充材料附在文后。

致谢

本文中的工作受到中国科学院青年创新促进会会员专项基金（基金号：

2021251) 的资助, 在此表示感谢。

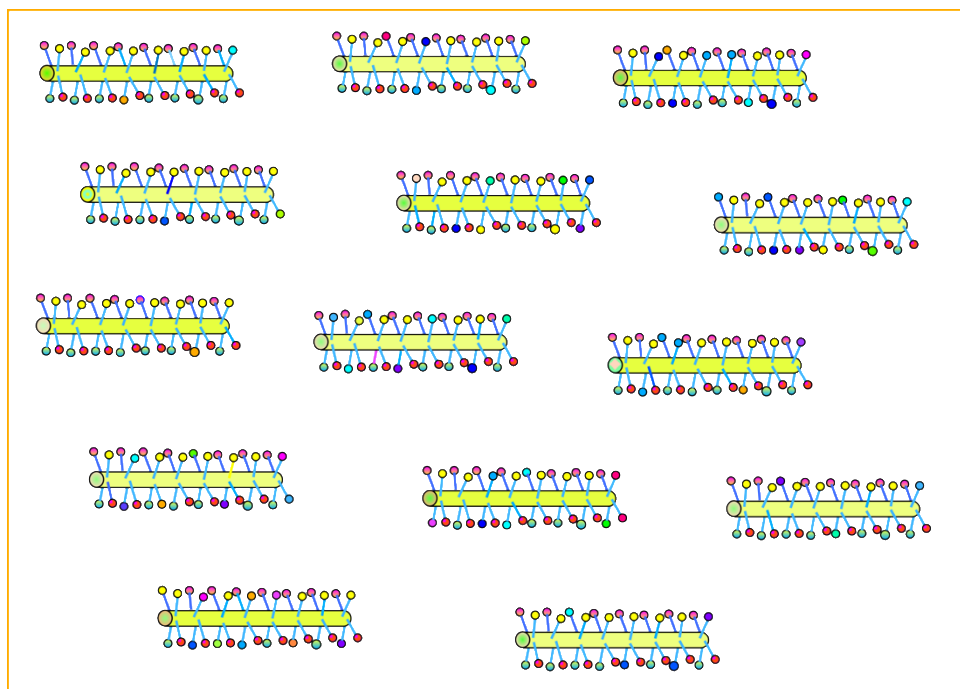
参考文献:

1. C. M. Dobson, Chemical space and biology. *Nature* **432**, 824 (2004)
2. J-L. Reymond, The Chemical Space Project. *Acc. Chem. Res.* **48**, 722 (2015)
3. E. J. Llanos, W. Leal *et al.* Exploration of the chemical space and its three historical regimes. *PNAS*. **116**, 12660 (2019)
4. Y. Liu, C. Mathis *et al.* Exploring and mapping chemical space with molecular assembly trees. *Sci. Adv.* **7**, eabj2465 (2021)
5. J. J. Semler, Y. K. Jhon *et al.* Facile Method of Controlling Monomer Sequence Distributions in Random Copolymers. *Adv. Mater.* **19**, 2877 (2007).
6. K. F. Noble, A. M. Noble, *et al.* Blocky bromination of syndiotactic polystyrene via post-polymerization functionalization in the heterogeneous gel state. *Polym. Chem.* **9**, 5095 (2018).
7. N. K. Boasen, M. A. Hillmyer. Post-Polymerization Functionalization of Polyolefins. *Chem. Soc. Rev.* **34**, 267 (2005).
8. M. A. Gauthier, M. I. Gibson *et al.* Synthesis of Functional Polymers by Post-Polymerization Modification. *Angew. Chem., Int. Ed.* **48**, 48 (2009).
9. P. Zinck, F. Bonnet *et al.* Functionalization of Syndiotactic Polystyrene. *Prog. Polym. Sci.* **34**, 369 (2009).
10. T. S. Jo, S. H. Kim *et al.* Highly Efficient Incorporation of Functional Groups into Aromatic Main-Chain Polymer Using Iridium-Catalyzed C–H Activation and Suzuki–Miyaura Reaction. *J. Am. Chem. Soc.* **131**, 1656 (2009).
11. C. Dizman, M. A. Tasdelen *et al.* Recent Advances in the Preparation of Functionalized Polysulfones. *Polym. Int.* **62**, 991 (2013).
12. J. B. Williamson, S. E. Lewis *et al.* C-H Functionalization of Commodity Polymers. *Angew. Chem., Int. Ed.* **58**, 8654 (2019).
13. G. P. Abatti, I. P. Gross *et al.* Tuning the Thermal and Mechanical Properties of PSU by Post-Polymerization Friedel-Crafts Acylation. *Eur. Polym. J.* **142**, 110111 (2021).
14. E. R. King, S. B. Hunt *et al.* Gold-Catalyzed Post-Polymerization Modification of Commodity Aromatic Polymers. *JACS Au.* **1**, 1342 (2021).
15. T. Kouzarides, Chromatin Modifications and their Function. *Cell.* **128**, 693 (2007).
16. F. Lyko, The DNA Methyltransferase Family: a Versatile Toolkit for Epigenetic Regulation. *Nat. Rev. Genet.* **19**, 81 (2017).
17. M. Luo, Chemical and Biochemical Perspectives of Protein Lysine Methylation. *Chem. Rev.* **118**, 6656 (2018).

18. S. Han,; A. Brunet, Histone Methylation Makes its Mark on Longevity. *Trends Cell Biol.* **22**, 42 (2012)..
19. A possible synthetic route toward this 24 mer was provided in Supporting Information.
20. L. Pauling, The Nature of the Chemical Bond. Application of Results Obtained from the Quantum Mechanics and from a Theory of Paramagnetic Susceptibility to the Structure of Molecules. *J. Am. Chem. Soc.* **53**, 1367 (1931).
21. J. C. Slater, Directed Valence in Polyatomic Molecules. *Phys. Rev.* **37**, 481 (1931).
22. C. Barbier, G. Berthier, Half A Century of Hybridization. *Adv. Quantum Chem.* **36**, 1 (2000).
23. W. A. Bingel, W. Lüttke, Hybrid Orbitals and Their Applications in Structural Chemistry. *Angew. Chem., Int. Ed.* **20**, 899 (1981).
24. Gentili PL and Perez-Mercader J, Quantitative estimation of chemical microheterogeneity through the determination of fuzzy entropy. *Front. Chem.* 10:950769 (2022).
25. Q. Chen Beyond snowflakes: Heterogeneity in nanomaterials. *Nano Lett.* **22**, 3. (2022).
26. L. F. Zacchi, B. L. Schulz, N-glycoprotein macroheterogeneity: biological implications and proteomic characterization. *Glycoconj J* **33**, 359 (2016).

补充材料

图 S1. 取代聚合物体系单分子混合物的示意图



ChinaXiv:202605.00087v1

表 S1. 模型体系 I 的潜在异构体总数

<i>n</i>	overall substitute rate (%)	Overall number of potential isomers (<i>r</i>)	lgr
0	0	1	0
1	0.1	1000	3
2	0.2	4.995×10^5	5.70
3	0.3	1.66×10^8	8.22
4	0.4	4.14×10^{10}	10.6
5	0.5	8.25×10^{12}	12.9
6	0.6	1.37×10^{15}	15.1
7	0.7	1.94×10^{17}	17.3
8	0.8	2.41×10^{19}	19.4
9	0.9	2.66×10^{21}	21.4
10	1.0	2.63×10^{23}	23.4
11	1.1	2.37×10^{25}	25.4
12	1.2	1.95×10^{27}	27.3
13	1.3	1.48×10^{29}	29.2
14	1.4	1.05×10^{31}	31.0
15	1.5	6.88×10^{32}	32.8
16	1.6	4.24×10^{34}	34.6
17	1.7	2.45×10^{36}	36.4
18	1.8	1.34×10^{38}	38.1
19	1.9	6.92×10^{39}	39.8
20	2.0	3.39×10^{41}	41.5
21	2.1	1.58×10^{43}	43.2
22	2.2	7.05×10^{44}	44.8
23	2.3	3.00×10^{46}	46.5
24	2.4	1.22×10^{48}	48.1
25	2.5	4.76×10^{49}	49.7

ChinaXiv:202605.00087v1

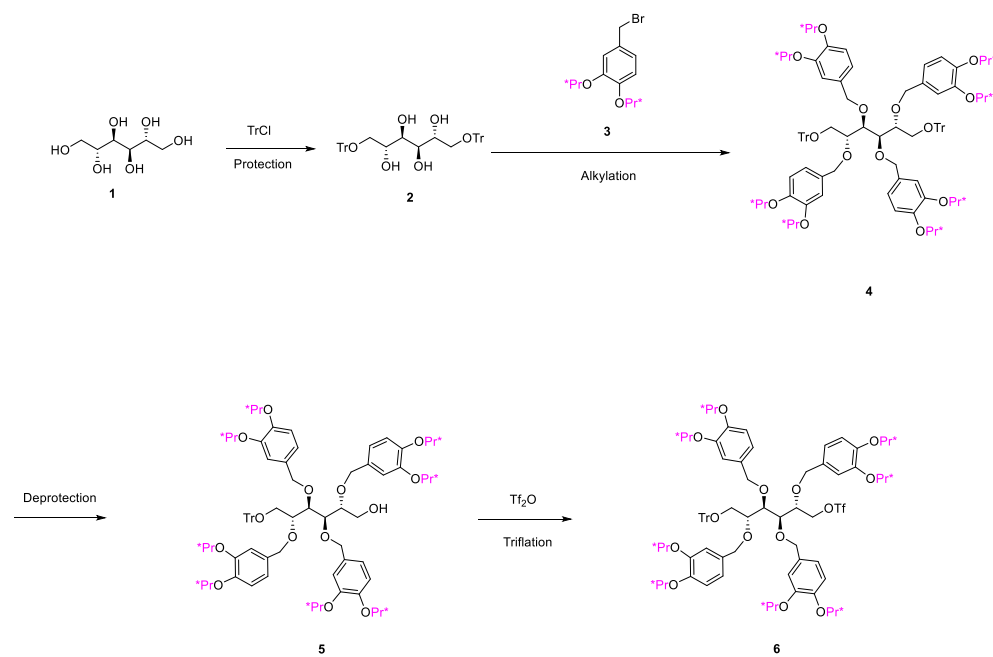
表 S2. 模型体系 II 的潜在异构体总数

<i>n</i>	overall substitute rate (%)	Overall number of potential isomers (<i>r</i>)	<i>lgr</i>
0	0	1	0
1	1	1000	3
2	2	4.95×10^5	5.69
3	3	1.62×10^8	8.21
4	4	3.92×10^{10}	10.6
5	5	7.53×10^{12}	12.9
6	6	1.19×10^{15}	15.1
7	7	1.60×10^{17}	17.2
8	8	1.86×10^{19}	19.3
9	9	1.90×10^{21}	21.3
10	10	1.73×10^{23}	23.2
11	11	1.41×10^{25}	25.1
12	12	1.05×10^{27}	27.0
13	13	7.11×10^{28}	28.9
14	14	4.42×10^{30}	30.6
15	15	2.53×10^{32}	32.4
16	16	1.35×10^{34}	34.1
17	17	6.65×10^{35}	35.8
18	18	3.07×10^{37}	37.5
19	19	1.32×10^{39}	39.1
20	20	5.36×10^{40}	40.7
21	21	2.04×10^{42}	42.3
22	22	7.33×10^{43}	43.9
23	23	2.49×10^{45}	45.4
24	24	7.98×10^{46}	46.9
25	25	2.43×10^{48}	48.4

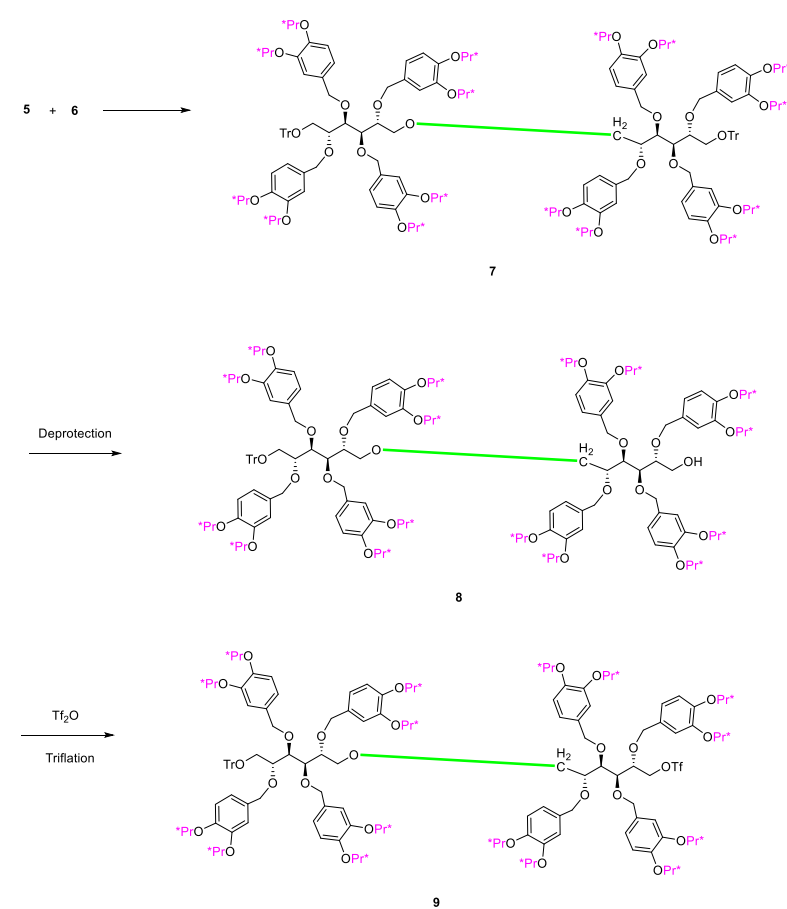
ChinaXiv:202605.00087v1

图 S2. 制备 *O*-丙基取代 D-甘露醇模型体系 24 聚糖的可能合成路线

A.

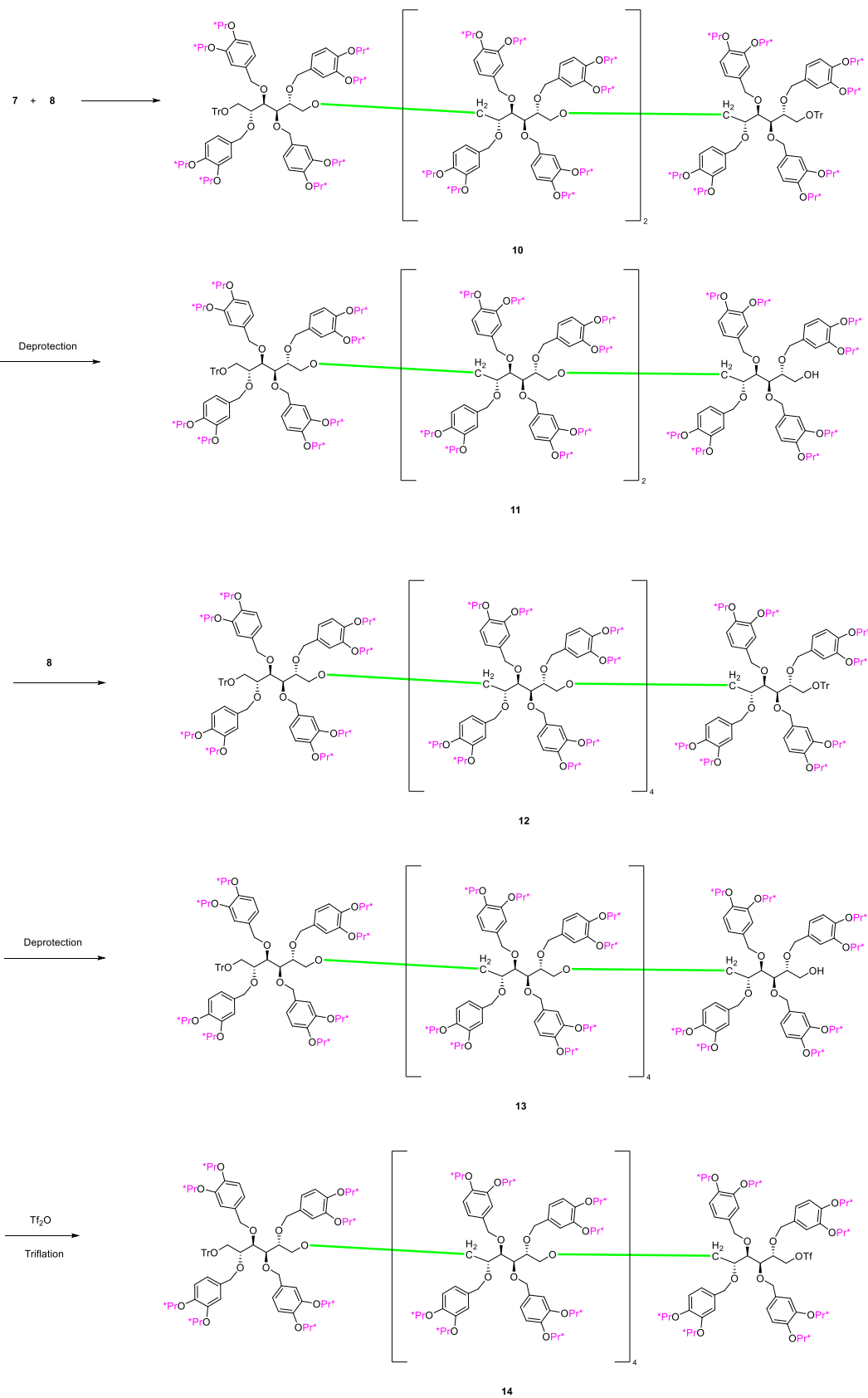


B.



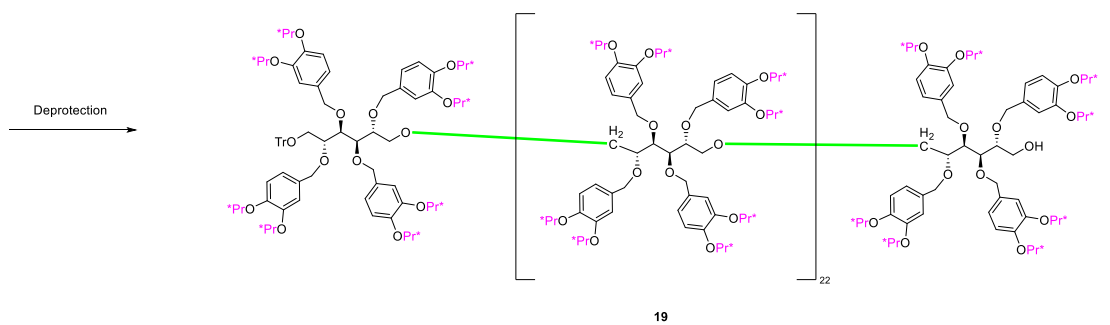
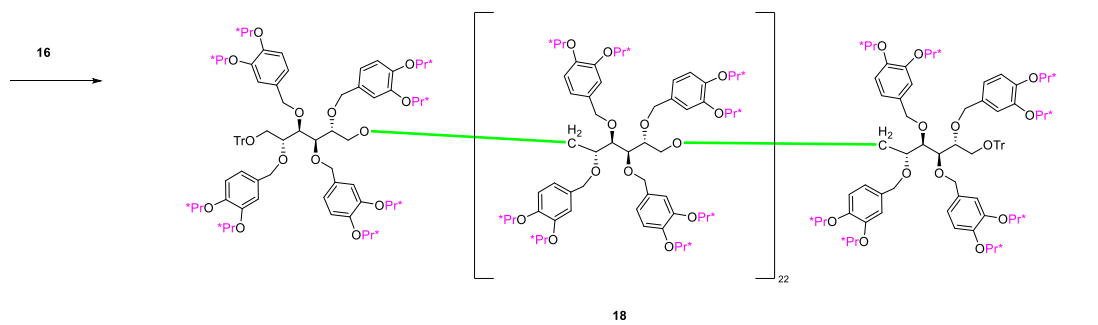
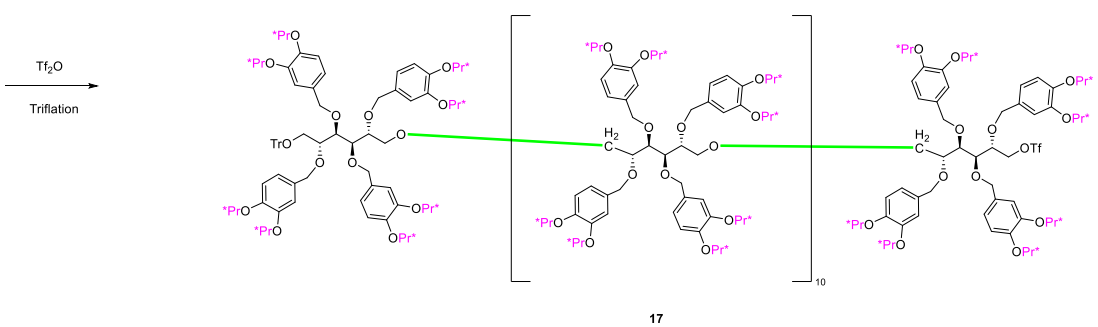
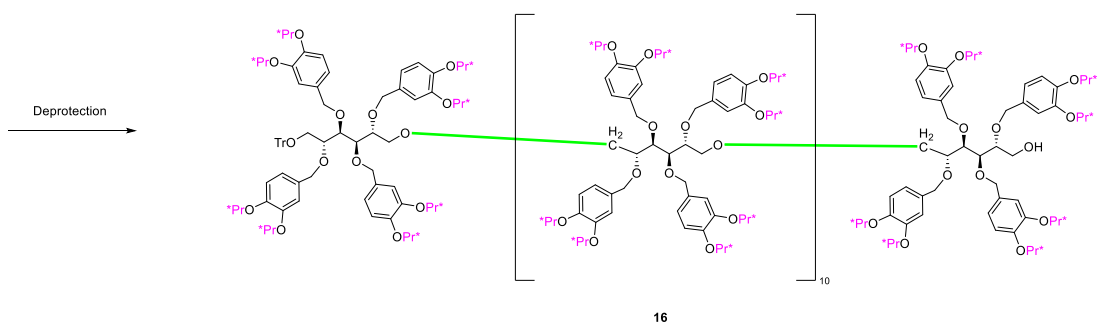
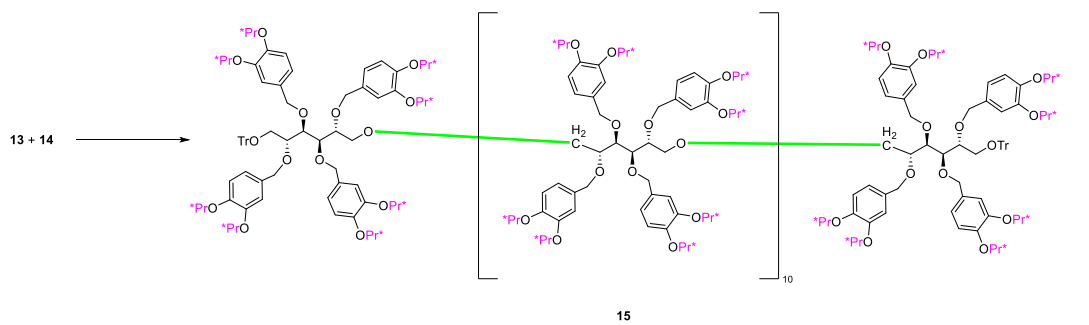
ChinaXiv:202605.00087v1

C.



ChinaXiv:202605.00087v1

D.



ChinaXiv:202605.00087v1

图 S2 展示了单分子混合态 24 聚体的合成路线。在此路线中，使用了等量的取代溴苄 3，预计这些化合物具有基本相同的反应活性和极性。因此，其合成和纯化程序预计与合成纯单异构体产物¹类似。

参考文献

1. Y. Huang, Novel Crosslinking Reagents, Macromolecules, Therapeutic Conjugates, And Synthetic Methods Thereof. WO2013012961A2, 2013.1.24