

生物大分子从头合成和设计的研究进展*

徐欣东 齐鹏翔 蓝尉冰 陈玉颖 陈 山**

(广西大学轻工与食品工程学院, 蔗糖产业省部共建协同创新中心, 南宁 530004)

摘要 生物大分子指生物体内存在的DNA、蛋白质、多糖等物质, 其对生物体正常生命活动至关重要. 从头合成和设计技术在生物大分子的合成和结构设计上具有自由度高、前体简单等特点, 能够按照特定研究目的对生物大分子进行全新设计和高效合成. 近年来, 从头合成与设计技术在人造基因组合成、新型蛋白质类药物设计、糖缀合物合成等领域已开始受到重视. 基于生物大分子从头合成和设计技术, 可以定向制备全新设计的DNA或全新的基因表达产物, 以及具有识别功能的糖链或糖缀合物, 将大大推进诸如细胞因子模拟物、基因治疗递送载体等生物活性物质的开发, 为人工生物系统的构建、罕见疾病的治疗等提供新的解决方法. 本文就DNA、蛋白质和多糖的从头合成和设计进行了综述, 阐述了相关方法及应用, 最后概括分析了三者之间的关系.

关键词 从头合成, 从头设计, DNA, 蛋白质, 多糖, 方法, 应用

中图分类号 Q81, Q5

DOI: 10.16476/j.pibb.2019.0094

相比于系统生物学“从大到小”对生物的模拟和改造, 合成生物学则是“从小到大”地对生物系统进行根本性的创新设计, 旨在创造全新的人造生物系统. 利用合成生物学方法和理论可以对生物系统进行有目的地设计、改造甚至重新合成, 在解释生物体内复杂合成过程机理、解决关系民生的重大生物技术问题等方面具有重要战略意义.

在生物体中生物大分子是指DNA、蛋白质、多糖等, 这些物质对生命体的正常生理活动具有不可或缺的作用. 在上述生物大分子的组成结构中都存在一种或多种的重复单体, 且单体间以磷酸二酯键、肽键、糖苷键等进行连接. 这类物质所具有的一种或多种单体的简单前体特点, 尤其是DNA和蛋白质所具有的模板合成特征, 为实现以小分子体系合成生物大分子以及分子结构的设计提供了理论可能.

从头合成(*de novo synthesis*)指由简单小分子(例如糖或氨基酸)合成复杂大分子, 而非大分

子部分降解后再生的过程. 而从头设计(*de novo design*)则是在分子的结构及组成均未知的情况下, 辅以计算机模拟等技术对分子的结构组成进行全新设计的过程. 在实际研究中两者的关系密不可分, 前者为后者的研究提供了合成基础, 例如从头设计的蛋白质需要以特定DNA进行转录表达; 而后者研究出的先进成果则可为前者的定向研究提供扎实参考. 基于目前对利用从头合成与设计技术定向制备生物活性物质的热门研究, 本文对DNA、蛋白质、多糖从头设计与合成的相关进展进行了归纳综述, 对相关合成和设计方法及应用进行总结及展望(图1), 为生物大分子的人工合成创新提供理论依据.

* 国家自然科学基金地区科学基金(21768001)资助项目.

** 通讯联系人.

Tel: 0771-3231590, E-mail: chenshan@gxu.edu.cn

收稿日期: 2019-05-01, 接受日期: 2019-06-28

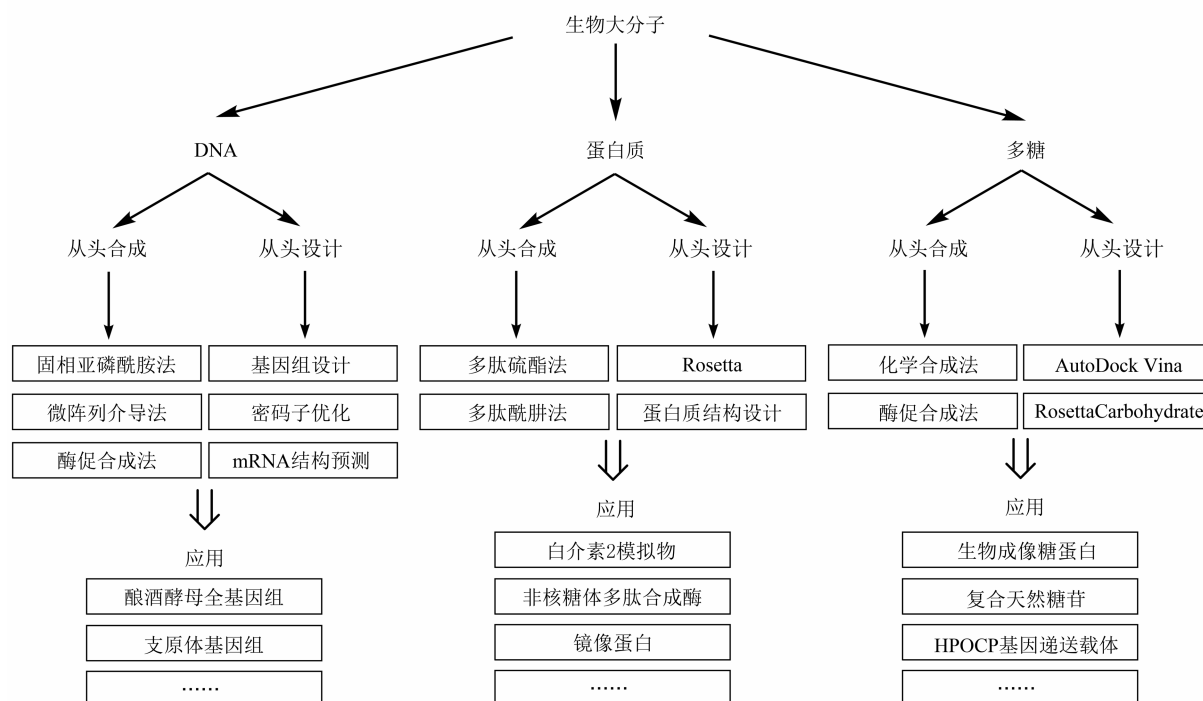


Fig. 1 Schematic diagram of the advances in *de novo* synthesis and *de novo* design of biomacromolecules

图1 生物大分子从头合成和设计的进展示意图

1 从头设计和合成DNA

DNA作为生物体的遗传信息载体,决定着生物体特异的生理特征.对DNA的结构设计及修饰可以定向地改变基因表达产物的特征,有利于构建符合相关研究或生产需求的人工生物分子及系统.然而研究一个具有全新功能的DNA需要对碱基序列进行大规模的修改和重排,因此从头合成和设计技术在研究合成全新DNA上具有一定的优越性,在新型蛋白质开发^[1]、基因编辑工具类似物开发^[2]、人工全基因组合成^[3-4]等热点研究领域具有巨大潜力.

1.1 从头合成DNA

在DNA的从头合成体系中,主要分为寡核苷酸合成和基因合成两个阶段,分别负责寡核苷酸序列合成和寡核苷酸片段之间的组装.

1.1.1 寡核苷酸的合成

寡核苷酸合成是DNA从头合成过程中的基础,旨在通过一定方法对零散的脱氧核苷酸进行按序合成,合成方法的成本、准确性、合成长度、合成数量等都是考察的重要指标.目前寡核苷酸的从头合成方法有固相亚磷酸胺合成法、微阵列介导合成

法、酶促合成法等.

a. 固相亚磷酸胺合成法

固相亚磷酸胺合成法最早由Beaucage等^[5]提出,图2a为该法的机理图.该法的具体步骤为:(1)5'羟基脱保护.使用酸催化法使固定在固相上的核苷酸的二甲氧基三苯甲基(DMT)脱离,使5'羟基具有反应活性.(2)添加核苷酸.待添加的核苷酸的3'羟基经亚磷酸胺活化后与四唑活化剂混合,得到的核苷-亚磷酸活化体再与之前5'羟基活化的核苷酸进行缩合.(3)“带帽”反应.少数未参与缩合反应的5'羟基活化核苷酸通过乙酰化以防止其继续参与反应.(4)磷酸基团氧化.采用碘溶液将三价磷酸三酯氧化为五价磷酸三酯^[6].

相关学者还对固相亚磷酸胺合成法的方法及产物进行了优化改进.Hayakawa等^[7]研究了咪唑衍生物的酸式盐对亚磷酸胺法合成低聚DNA的促进作用,其通过比较发现N-苯基咪唑三氟甲磺酸盐、N-对乙酰基苯基咪唑三氟甲磺酸盐、苯并咪唑三氟甲磺酸盐及N-甲基苯并咪唑三氟甲磺酸盐可作为固相亚磷酸胺法的启动及促进因子,这些促进因子还可减少方法中亚磷酸胺的使用量及反应剩余的核苷酸含量.Satheesh等^[8]根据亚磷酸胺法的机理

开发出一种基于DMT和乙酰丙酸保护基团、可聚合在DNA链上的支化功能物质,并运用该支化因子成功合成出具有3个支化臂的梳状单链DNA,相较于非共价连接及树枝状的支化DNA更适合应用于生物传感器、多色成像等领域.

固相亚磷酸胺合成法在DNA的从头合成中是研究和使用较早的一种方法,且广泛应用于目前的自动化寡核苷酸合成仪器,其合成的DNA聚合度约为100~400,但在合成成本上较高.此外,该法在循环合成过程中会产生部分未参与反应的5'羟基活化核苷酸,从而导致最终合成的全长DNA占比偏低,且合成产物中可能还会出现一些结构错误,因此该法还需进一步改进.

b. 微阵列介导合成法

微阵列介导合成法是一种基于亚磷酸胺法的改进方法,其与固相亚磷酸胺合成法的主要不同在于保护基团的脱去方式.目前,脱保护方法中存在光控法、电化学法、“喷墨”法等.

光控法运用了基于掩模的光刻技术选择性地让光不稳定的核苷-亚磷酸胺脱去保护而进行循环过程,即运用光线对合成过程进行控制^[9].电化学法采用紧邻的阴阳两极产生具有去保护活性物质,进而耦合到进行合成反应的玻璃板上以使DMT基团脱去保护而控制寡核苷酸的合成^[10].如图2b所示,Levie等^[11]提出使用金电极作为亚磷酸胺法的固相载体,同时基于4-氨基苯乙醇的电化学还原将羟基引入金电极表面,使这些羟基可作为寡核苷酸合成序列的起始点,相较于传统电化学法去除了合成玻璃板的使用,提升了电化学微阵列的效率.此外,相关学者还采用在微阵列上集成“喷墨”技术(递送亚磷酸胺相关物质)进行寡核苷酸的合成,同时还利用酶法对合成过程进行错误校正^[12-13].

微阵列介导合成法与固相亚磷酸胺合成法相比,虽然合成错误率比后者略高,但在单位长度寡核苷酸的生产成本上要大大低于后者,是目前工业上寡核苷酸批量生产的主要方法.

c. 酶促合成法

酶促合成法相较于传统化学法拥有高效性、专一性、副产物少等特点,因而在近几年成为生物大分子合成领域的研究热点,受到研究人员的广泛关注.

如图2c所示,Palluk等^[14]利用聚合酶-核苷酸缀合物进行寡核苷酸的从头合成,其合成步骤主要分为两步:(1)序列扩增.首先一个末端脱氧核糖

酸转移酶(terminal deoxynucleotidyl transferase, TdT)与单个脱氧核糖核苷三磷酸(deoxyribonucleoside triphosphate, dNTP)进行缀合,而后将缀合物加入到寡核苷酸底物中,通过5'端和3'端的缩合使底物长度增加一个核苷酸单位.(2)产物脱离.剩余的TdT-dNTP缀合物被灭活或清除,此外通过设定脱离环境(二硫苏糖醇或 β -巯基乙醇,365~405 nm光照射,肽酶)使寡核苷酸链与TdT分离,以便于后续扩展使用.该法添加一个核苷酸的时间经测定为10~20 s,且可通过迭代过程进行核苷酸序列的编码,为酶法寡核苷酸合成仪的研发提供了理论基础.

Ramandan等^[15]研究了人源DNA聚合酶 λ 、DNA聚合酶 μ 以及TdT在酶法从头合成DNA中的效果,发现这3种聚合酶均可在没有模板的情况下体外合成寡核苷酸链,其中DNA聚合酶 λ 在寡核苷酸的合成中对 Mn^{2+} 存在依赖性,且其结构上的Phe506氨基酸残基对合成活性至关重要.Veneziano等^[16]采用不对称式聚合酶链反应,结合LongAmp Taq聚合酶合成了长度为15 kb以上的单链DNA,提升了酶法合成DNA的最大长度.目前寡核苷酸的酶法从头合成相较于微阵列介导合成法,在经济成本及可控程度还未得到很好地控制,解决这些问题将有望成为新一代寡核苷酸合成技术.

1.1.2 基因组装

由于寡核苷酸在长度上存在限制,远远达不到一般DNA基因长度的要求,因此采用多种方法对多个寡核苷酸片段进行组装合成则显得尤为重要.类似于寡核苷酸的合成,目前的基因组装方法大多基于酶促组装,但由于组装过程将会合成出具有正确及错误序列的基因片段混合物,因此对于组装后序列的错误校正以及降低组装错误率也显得格外重要.

Jin等^[17]设计一个名为DATEL的DNA片段组装系统,该系统基于Taq、PfuDNA聚合酶和TaqDNA连接酶,采用变性-退火-裂解-连接的循环组装策略,最终的准确率为74%~100%,可在1~2 h内组装2~10个DNA片段(图3a).Potapov等^[18]采用Golden Gate组装体系(IIs型限制性内切酶法)对24个不同DNA片段进行成功组装;在错误检测中,研究者对四碱基对的连接效率进行量化,同时在典型条件下评估了T4和T7 DNA连接酶可能出现的错配组合,基于这些数据对Golden Gate组装体系的准确性进行系统评估,最终得到的

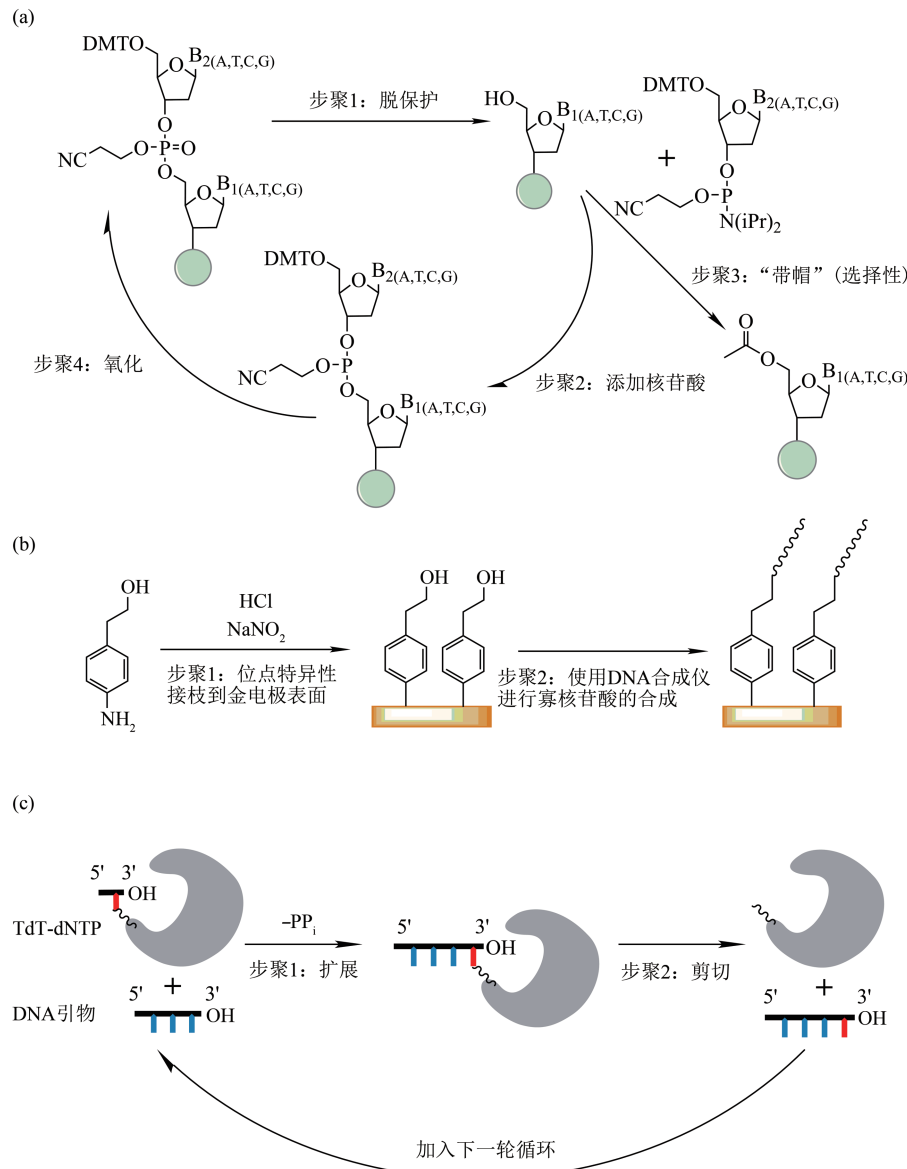
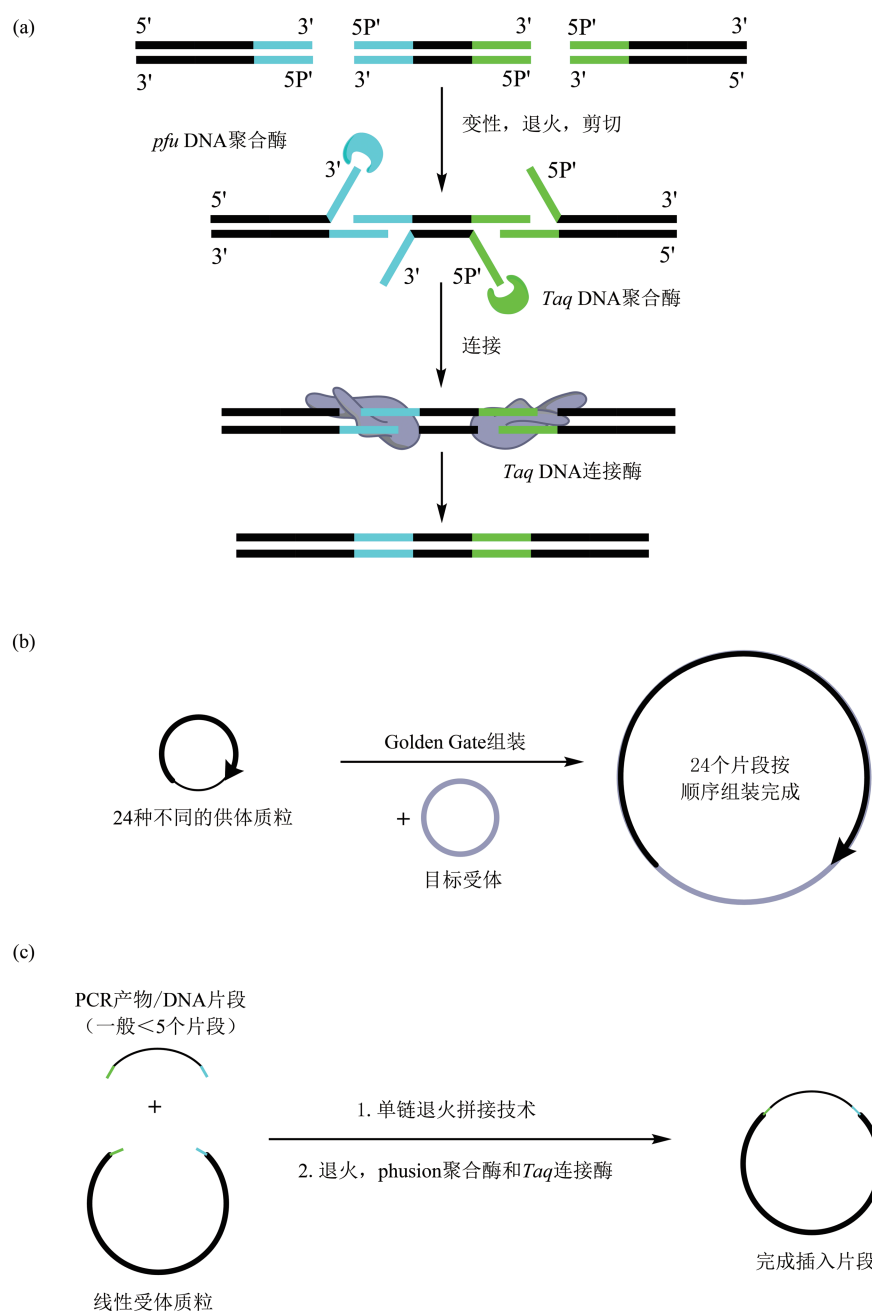


Fig. 2 Three synthesis methods of oligonucleotides^[5,11,14]
图2 三种寡核苷酸合成方法^[5,11,14]

(a) 固相亚磷酸胺合成法，DMT：4, 4'-双甲氧基三苯甲基；(b) 电化学微阵列-金相亚磷酸胺合成法；(c) 基于TdT-dNTP缀合物的酶促合成法，TdT：末端脱氧核苷酸转移酶，TdT-dNTP：末端脱氧核苷酸转移酶-脱氧核糖核苷三磷酸缀合物，PP_i：焦磷酸。

准确率为 (84±5) %，略低于预测的 91% (图 3b)。Andreou 等^[19] 基于 Golden Gate 体系开发了名为 Mobius 组装的通用 DNA 组装方法，其与 Golden Gate 体系基本兼容，且利用发色蛋白对组装的准确性进行评估，简化了组装步骤。与 Golden Gate 体系相比，Gibson 组装法^[20] 具有更低的组装难度，其可直接使用 PCR 扩增后的片段进行多段组装，最大组装片段也可达到 80 kb 以上 (图 3c)，但由

于没有限制性位点，所以在合成片段的载体转移及复杂体组装上不如 Golden Gate 体系；近年来相关研究者针对 Gibson 法在高 GC 碱基对含量基因片段的组装能力及降低错误率等方面作了改进优化^[21-22]。由此可见，Gibson 组装法、Golden Gate 体系以及其他基于酶促反应的基因组组装方法各有优势，而与自动化技术的结合将有助于开发出基于电脑编码和合成的 DNA 合成仪。

Fig. 3 Three gene assembly methods^[17-18,20]图3 3种基因组组装方法^[17-18,20]

(a) DATEL组装法; (b) Golden Gate组装法; (c) Gibson组装法.

1.2 从头设计DNA

1.2.1 基因组设计

从头设计DNA在基因组合成设计上具有天然优势,但是由于基因组设计的工作量大,因此如何高效协作地完成基因组的设计工作是亟待解决的关键问题.

其中, Richardson等^[23]描述了酿酒酵母的基

因组设计方法,其通过使用BioStudio软件来对基因组进行大规模计算机辅助设计,大大提升了基因组的设计效率.同时,该软件提供了通用的组装策略,并使用共享语言来对设计序列进行协调修改,为团队协作设计酵母基因组提供了扎实的工作框架.可见,以此为借鉴开发一类允许团队合作和统一框架策略的基因组设计软件将可以大大提高基因

组设计的效率与全面性.

1.2.2 mRNA结构预测

mRNA 作为翻译阶段的核心大分子, 其二级结构的复杂度、稳定性直接影响翻译过程的进行^[24-25]. 因此在 DNA 的从头设计中, 预测对应 mRNA 的二级结构将会有助于提升所设计 DNA 的合理性.

Bellaousov 等^[26] 介绍了名为 RNAstructure 的 RNA 二级结构预测软件, 该软件最早开发于 1998 年^[27], 目前结合 web 服务器可对自由能最小化、最大期望精度结构、双分子二级结构等方面进行预测, 是 mRNA 结构预测领域功能齐全、及时更新的软件. 此外, 近年来相关研究学者对 mRNA 结构预测的算法和模型作了进一步优化和完善, 提高了 mRNA 结构预测的准确性, 这将为 DNA 的从头设计提供充分可信的设计参考^[28-29].

1.2.3 密码子优化

密码子被认为是异源蛋白质表达水平的关键因素, 密码子优化可使蛋白质进行高效大量表达, 对相关工程基因的异源表达意义明显^[30]. 因此, 对密码子进行优化, 包括消除稀有密码子、密码子配对优化等, 也是 DNA 从头设计过程中重要的研究方向之一. Chin 等^[31] 开发了密码子优化在线工具 (codon optimization online, COOL), 其特点在于可提供多个目标的密码子优化方案, 并通过密码子适应指数、单体密码子使用频率、密码子上下文配对、隐藏终止密码子等多个方面基因序列进行优化, 具有较好的优化水平. Papamichail 等^[32] 对密码子上下文配对优化进行了统计分析, 从统计学对优化的复杂程度进行考察, 并开发了一项具有较好精确性和有效性的密码子上下文评估算法, 用于评估基因序列的合理程度. 可见, 基因序列中存在复杂多样的密码子组合, 因此基于计算机算法和经验模型对密码子进行优化是受到认可的研究手段.

1.3 从头合成和设计 DNA 的应用

DNA 的从头合成和设计在构建人工生物系统、研究基因表达差异等方面具有根本性优势, 可通过特定的基因组设计合成工具和统一的染色体间连接策略来实现对生物体基因组的定制设计和多团队协作合成, 还可针对工程基因的表达效率进行基因序列从头设计优化, 这对人造生命合成等前沿研究领域意义重大.

Gibson 等^[4] 对含有 582 970 bp 的生殖支原体基因组进行了完整化学合成, 该工作的成功完成对

构建原核生物基因组具有重要意义, 也为后续更高等级生物的基因组合成提供了方法上的参考. 此外, 近年来在 DNA 从头设计领域取得的又一代表性成果为真核生物酿酒酵母 (*Saccharomyces cerevisiae*) 全基因组项目 (Sc2.0), Sc2.0 首次实现了人工设计和合成真核生物全基因组, 在基因组设计与合成方面取得重要进展^[33-35]. 其中, Xie 等^[36] 对酿酒酵母的 5 号染色体进行了从头设计和合成, 研究者以天然酿酒酵母 5 号染色体 (wtV) 的核苷酸序列为参考, 从头设计并组装合成了人工酿酒酵母 5 号染色体 (synV), 且运用 CRISPR-Cas9 基因编辑工具对合成过程中产生的突变和错位进行校正, 最后运用全基因组测序、RNA 测序和培养实验验证了 synV 的序列准确性、表达差异性和菌体适应性, 同时 synV 还能被环化, 可作为模型用于人体环状染色体紊乱机制的研究. 由此可见, 原核及真核生物基因组的设计合成已得到初步实现, 而以此为模型研究人体基因组的设计合成将能大大提高研究的准确性.

2 从头合成和设计蛋白质

蛋白质作为生物体内基因表达的主要产物之一, 其通过酶、抗体、膜蛋白等形式深刻影响着生物体的生理系统及活动. 到目前为止, 对于蛋白质结构的研究大多与天然蛋白质的修饰有关, 这虽然可以从一定程度上定向增强或减弱蛋白质的某些功能和理化特性, 但不能从根本上对蛋白质进行优化. 在蛋白质从头设计中, 骨架结构和对应序列都是未知的, 因此基于这种情况的设计是相当困难的, 但同时也拥有最高的自由度, 这为天然蛋白质功能及人工蛋白质的机理及开发研究提供了先进方法, 在人工物质的设计合成上具有里程碑的意义^[37].

由于蛋白质的合成需要 DNA 作为编码设计模板, 定向地改变或重设氨基酸序列则需要对 DNA 进行对应修饰和重设, 因此 DNA 从头合成和设计技术的日渐成熟和完善大大提升了蛋白质从头设计和合成的可能性, 同时也降低了设计成本和难度.

2.1 从头合成蛋白质

近年来, 蛋白质的从头合成也得到了快速发展, 这对制备蛋白质或多肽类药物等生物活性物质至关重要. 类似于 DNA 中的固相亚磷酰胺合成法的原理, 在短肽的合成方面, 目前广泛使用的是基于叔丁氧羰基 (Boc) 和 9-芴甲氧羰基 (Fmoc) 保

护策略的多肽固相合成法 (solid phase peptide synthesis, SPPS), 且具有较完善的合成体系^[38]. 而在类似于基因组装的多肽间连接方面, 目前进展较为显著.

2.1.1 多肽硫酯法

多肽硫酯法是由 Johnson 等^[39-40] 提出的天然化学连接法 (图 4a), 主要合成原理为具有 C 端硫酯的多肽与具有 N 端半胱氨酸残基的多肽进行反应, 以形成天然肽键, 并发现 4- 巯基苯基乙酸 (MPAA) 可以使合成反应的速度提升一个数量级. 该法的成功发现在多肽的合成领域具有里程碑式的意义.

2.1.2 多肽酰肼法

Fang 等^[41-42] 提出了使用多肽酰肼法进行蛋白质的从头化学合成 (图 4b), 该方法首先采用 Boc/Fmoc-SPPS 法合成多肽酰肼, 之后在 NaNO_2 氧化剂作用下产生叠氮化多肽, 然后在 MPAA 作用下形成多肽硫酯, 最后再与半胱氨酸残基端多肽进行连接以合成更长的多肽或蛋白质; 此外, 还可采用脱硫方法使半胱氨酸残基转变为丙氨酸残基 (图 4c)^[43]. 与多肽硫酯法相比, 该法可制备酸敏感性修饰的蛋白质片段, 也避免了高腐蚀性氟化氢的使用, 良好地替代了多肽硫酯法, 被目前相关领域研究人员所广泛使用.

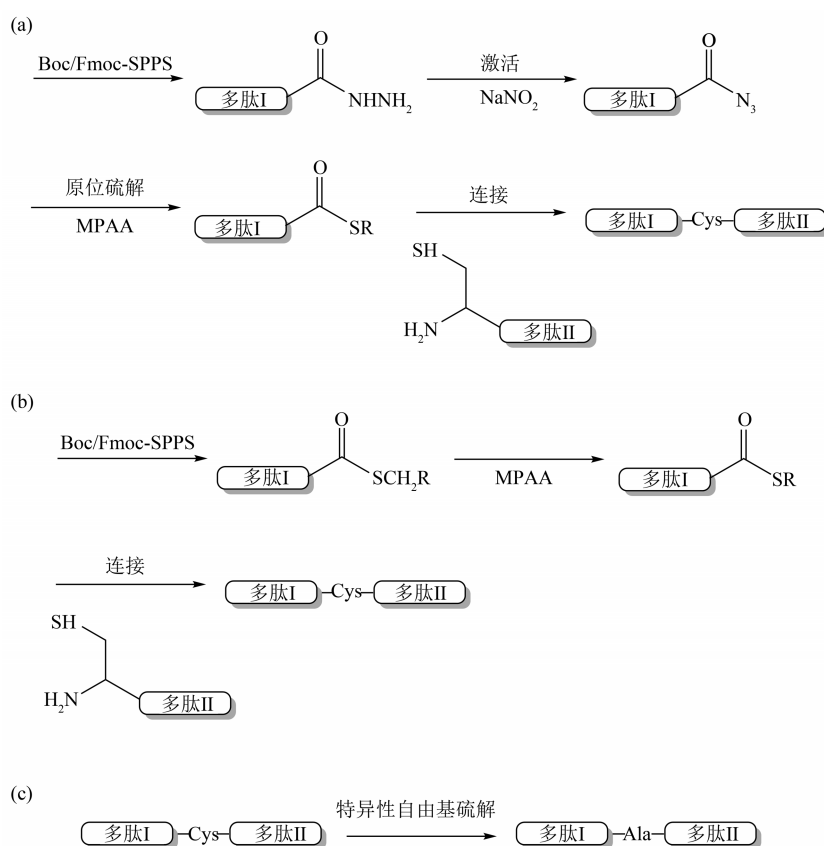


Fig. 4 Two methods of native chemical ligation of peptides and the free-radical-based specific desulfurization method for cysteine^[39-43]

图4 两种多肽天然化学连接法和基于自由基的半胱氨酸特异性脱硫法^[39-43]

(a) 多肽硫酯法; (b) 多肽酰肼法; (c) 基于自由基的半胱氨酸特异性脱硫法. Boc/Fmoc-SPPS: 叔丁氧羰基保护策略/ 9-芴甲氧羰基保护策略-多肽固相合成法; MPAA: 4-巯基苯基乙酸; Cys: 半胱氨酸残基; Ala: 丙氨酸残基.

2.2 从头设计蛋白质

2.2.1 基于Rosetta软件的设计方法

一般来说, 蛋白质的空间结构决定着其有效活性及稳定性. 在蛋白质的从头设计中, 蛋白质的整

体骨架结构取决于设计理念, 而设计合成后的蛋白质的表征结构往往和预测结构存在一定差异, 因此如何高效地预测蛋白质的空间结构则是决定蛋白质有效与否的重要前提.

由于分子结构模拟的复杂性, 蛋白质的结构设计序列预测基本采用计算机模拟的方法, 其主要依靠初始设定的骨架结构和启发式算法来产生大量可能结果, 并对这些结果进行比对筛选. 其中, Rosetta 是一套用于蛋白质结构计算建模和分析的软件, 其基本工作流程如下: a. 定义初始的骨架结构; b. 计算具有接近前者骨架结构的序列集; c. 根据包装质量、氢键模式、可折叠性、能量大小等指标对序列集进行缩小; d. 从头合成所需的蛋白质序列并表征结构和效力^[44]. 在近些年的研究中, 该软件已广泛应用于蛋白质的从头设计, 在细菌蛋白^[45]、跨膜蛋白^[46]、酶^[47-48]、多肽^[49]、蛋白质寡聚体^[50]等生物分子的设计合成工作中起到重要作用.

此外, 相关学者还将此软件与其他理论方法及模型相结合, 强化了结构设计与序列预测. André等^[51]采用 Rosetta 能量函数结合 fold-dock 法^[52]对异二聚体卷曲螺旋结构进行了准确预测, 且通过模拟发现由卷曲螺旋引起的折叠能量图景具有几种等能量状态, 其中独立的卷曲螺旋结构的能量图景更加收敛于目标结构的能量, 这两种方法的结合可以有效地分析折叠图景及结构特异性. Ludwiczak等^[53]将 Rosetta 与分子动力学相结合以强化蛋白质的结构设计与对应序列筛选, 该组合方法比现有方法产生的序列多 20%~30%, 提升了序列预测的多样性和与天然序列的相似性.

2.2.2 功能结构的设计

天然存在的蛋白质往往存在一些特殊的结构特征, 这些结构对某些特定的蛋白质功能具有决定性作用. 因此如何从头设计和合成出具有这些特殊结构特征的蛋白质, 对拓展从头设计蛋白质的应用范围影响重大.

对此, 相关学者采用计算模拟方法对蛋白质中拓扑结构的定向设计和合成作了研究. Huang等^[54]从头设计了具有原子级精度的四重对称、耐高温且可逆折叠的 TIM 桶状蛋白质, 经从头设计的模型结构与天然结构的相似度较高, 同时通过表征验证其与其他天然存在该结构的蛋白质不属于同一家族, 该结构的成功合成为酶的定制化提供了新的路径. Doyle等^[55]对串联重复蛋白的从头设计进行了开发和验证, 其通过全新定义重复单元间的几何关系成功地设计出具有闭合结构的左旋 α -螺旋桶状串联重复蛋白, 并通过了 X 射线晶体结构验证. Thomas等^[56]以开发亲脂性生物活性分子和带电分

子的受体为目的, 从头设计了 α -螺旋桶状蛋白(α HB), 该蛋白质的优势在于可通过调节疏水通道的尺寸以适应特定靶分子的结合, 同时由于其结构中存在热稳定及低复杂的重复序列, 因而该蛋白质在其疏水通道中能耐受较多的突变, 可见其可作为小分子识别结合、储存释放的潜力载体. Noguchi等^[57]对 β -螺旋桨家族蛋白的结构进行了计算设计, 研究人员首先制作了 Tako8 的完美对称同源物, 之后采用不同的计算方法对 Tako8 重新设计以产生 Ika8 四重对称蛋白, 该蛋白质能够由各个重复单元自组装而成, 其在生物纳米技术及 WD40 蛋白折叠及进化模型的研究上具有应用潜力. 可见, 相关蛋白质超二级结构的存在会使蛋白质具有载体和催化活性等优良特性, 因此对功能结构的从头设计也是一个重要研究方向.

2.3 从头合成和设计蛋白质的应用

由于从头设计手段可以在一定程度上自由定义蛋白质的骨架结构和序列, 同时蛋白质又是调节和响应生物体内生理活动的主要物质形式, 因此研究日渐深入的从头设计和合成方法将有希望开发出具有减毒增效或保效、靶向捕获或释放等优良功能的人工蛋白, 其在生物医药等领域拥有巨大应用潜力.

2.3.1 药物开发领域

肿瘤及癌症治疗一直是医学领域的热门研究方向, 近年来对相关蛋白质抗肿瘤、抗癌等功能的机理研究以及新型蛋白质的设计方法与合成过程研究受到研究者的广泛关注^[58].

Gallardo等^[59]从头设计了一种具有生物活性的淀粉样蛋白生成肽 VASCIN, 研究者通过小鼠皮下黑色素瘤表征实验表明 VASCIN 会抑制依赖血管内皮生长因子受体 2 (VEGFR2) 的肿瘤细胞生长, 但对其他组织无毒性, 即具有靶向释放的能力. VASCIN 毒性的主要机理为, VASCIN 导致 VEGFR2 的功能丧失, 进而导致细胞生长受到影响. Silva等^[60]对细胞因子白介素 2 (IL-2) 和白介素 15 (IL-15) 的模拟物 (Neo-2/15) 进行了设计, 其设计出一种全新结构来保留模拟物中细胞因子的功能, 但完全消除了与 IL-2R α 链 (CD25) 及 IL-15R α 链 (CD215) 的相互作用, 且这 2 种模拟物在 95℃ 下煮沸 1 h 后仍能有效驱使 T 细胞存活, 可见其稳定性十分优秀; 这类细胞因子模拟物的成功设计和合成对癌症药物的减毒增效具有十分重要的意义. Chevalier等^[61]以靶向治疗为设计理念, 设计

并表征了 22 600 个平均含有 40 个残基的微蛋白, 以靶向作用于甲型 H1 流感血凝素及肉毒杆菌神经毒素 B; 经筛选后的微蛋白对靶物质作用明显, 在 80℃ 下处理 1 h 后未检出活性损失, 且得益于小尺寸, 这些微蛋白几乎不会引起免疫反应. 可见, 人工药用蛋白模拟物已经开始逐步从理论研究向应用研究靠近, 同时从头合成和设计技术也能赋予产物靶向释放、减毒增效、提升环境耐受性等优良特性, 这对高效蛋白质类药物的开发具有重要意义.

2.3.2 生物工程领域

生物材料及生物分子合成研究已在人造器官、人造抗生素等方面取得较大的进展. 从头设计蛋白质可以人为地改变蛋白质的功能及应用范围, 这将进一步推进生物工程产品的研究与开发.

Zheng 等^[62] 对具有高韧性、生物降解性等材料特性的蛛丝蛋白 (spider silk proteins, SSP) 进行分析后, 基于基因工程及蛋白质从头设计技术对 SSP 进行重组设计以获得重组蛛丝蛋白 (recombinant spider silk proteins, RSSP), 其中包括蛛丝蛋白激发肽、蛛丝蛋白嵌段共聚物、全长蛋白质等; 由于 RSSP 的序列可预测性、其免疫原性可得到较好地控制, 且还可用于设计具有多尺度自组装能力的 RSSP, 在仿生材料的开发上具有一定潜力. Rapson 等^[63] 对金属蛋白的模拟作了研究, 其通过金属中心、有机环、蛋白骨架、材料格式的设计顺序成功设计出重组丝蛋白 (AmelF3), 其可结合血红素及相关金属元素, 可应用于光捕获复合物、催化、传感器等领域. Bozhüyük 等^[64] 提出一种新的针对非核糖体多肽合成酶 (nonribosomal peptide synthetase, NRPS) 的修饰策略, 即用重新定义的交换单元作为功能单元, 最终成功设计出可合成出多种新型非核糖体多肽的新型 NRPS, 经进一步优化后有望通过非核糖体的手段合成一些具有特定功能的生物分子.

此外, 基于多肽酰肼合成法, 相关学者对合成镜像蛋白及蛋白质晶体学进行了研究. Weinstock 等^[65] 合成了具有 312 个氨基酸残基的镜像分子伴侣蛋白 (DapA), 提升了从头化学合成蛋白质的长度. Wang 等^[66] 以合成镜像 DNA 聚合酶为目的, 利用该法成功合成了具有 174 个氨基酸残基的镜像非洲猪瘟病毒聚合酶 X (ASFV pol X), 这在一定程度上建立了手性的生物分子系统, 对镜像生物大分子的研究具有重要意义. Pan 等^[67] 利用该法合成出 K27 位多聚泛素化二联蛋白质 (diUb) 和三联蛋

白质 (triUb), 其中 triUb 具有 228 个氨基酸残基, 并得到了 2.1 Å 分辨率的 X 射线衍射结构, 为生物结构的研究提供了参考.

由此可见, 蛋白质从头合成和设计技术推进了相关仿生材料、功能酶、镜像蛋白等内容的相关研究, 可合成一般基因重组技术所无法合成的相关蛋白质或多肽, 在开发功能性蛋白质材料、探索镜像生物系统等方面具有巨大潜力.

3 从头合成和设计多糖

不同于 DNA 与蛋白质, 多糖在合成过程中不存在严格意义上的模板, 因此在从头合成和设计的可控性上远不及前两者. 然而多糖对于生物体的正常生理活动而言也必不可少, 具体可以表现为储存能量的肝糖原、支撑生物骨架的壳多糖和纤维素等. 此外, 糖基化反应作为糖蛋白、脂多糖等生物大分子的生成手段, 与疾病的产生机理密切相关, 受到研究者的广泛关注^[68]. 因此, 开发适宜的从头合成与设计方法对于开发具有优良性质和功能的多糖及糖缀合物影响深远.

3.1 从头合成多糖

从头合成作为以小分子体系构建大分子体系的方法, 在多糖的合成领域同样受到广泛关注. 目前对于多糖的合成方法主要为酶促合成法和化学合成法.

3.1.1 酶促合成法

在自然界中, 酶促合成法在许多微生物产生天然产物的方法, 同时合成酶在众多领域内应用广泛, 例如用于合成寡核苷酸的 TdT、用于酶促糖基化反应的 Leloir 型转糖基酶等^[69]. 利用酶优异的催化效率及专一性, 辅以定向调控优化, 可以为天然多糖的从头合成提供一条高效路径, 同时也成为合成生物学的一个重要研究方向^[70].

Park 等^[71] 对来自奈瑟氏球菌的新型淀粉蔗糖酶进行了克隆和表征, 该酶可利用蔗糖作为简单前体从头合成直链淀粉, 其产率最高达到 24%, 为直链淀粉的生产提供参考. Song 等^[72] 考察了来自柠檬明串珠菌的葡聚糖合成酶的合成特性, 从表征结果中发现合成产物具有 58% 的 α -1, 6 糖苷键和 42% 的 α -1, 3 糖苷键, 同时还发现该酶可发生基于麦芽糖的受体反应生成聚合度为 3~5 的葡寡糖. Grimaud 等^[73] 利用蔗糖作为底物, 分别利用葡聚糖蔗糖酶及交替糖合成酶进行酶促合成, 得到了聚合度范围为 29~170 的三嵌段葡聚糖共聚物, 这为

交替多糖的开发应用提供了方法基础 (图 5a) . Wen 等^[74] 对酶法从头合成低聚糖的自动化进展进行了阐述, 指出酶法合成的自动化研究将有助于人造高尔基体、自动化固相酶法合成系统的开发, 最终达到基于多糖酶法合成仪的高效生产水平 (图 5b) . 可见目前研究者在多糖的酶促合成中趋向于研究不同菌种来源的酶的合成特性、多糖结构及分子量定向控制等方面, 以达到多糖的选择性高效

生产. **3.1.2 化学合成法** 类似于寡核苷酸合成中的亚磷酰胺法, 多糖的化学合成主要针对单体间的糖苷键进行方法研究. 尽管化学合成法在合成产物的分子质量上远不及酶促合成法, 但可在一定程度上对多糖链结构进行人为调控^[75], 且不会像酶促合成法受限于酶种类的影响.

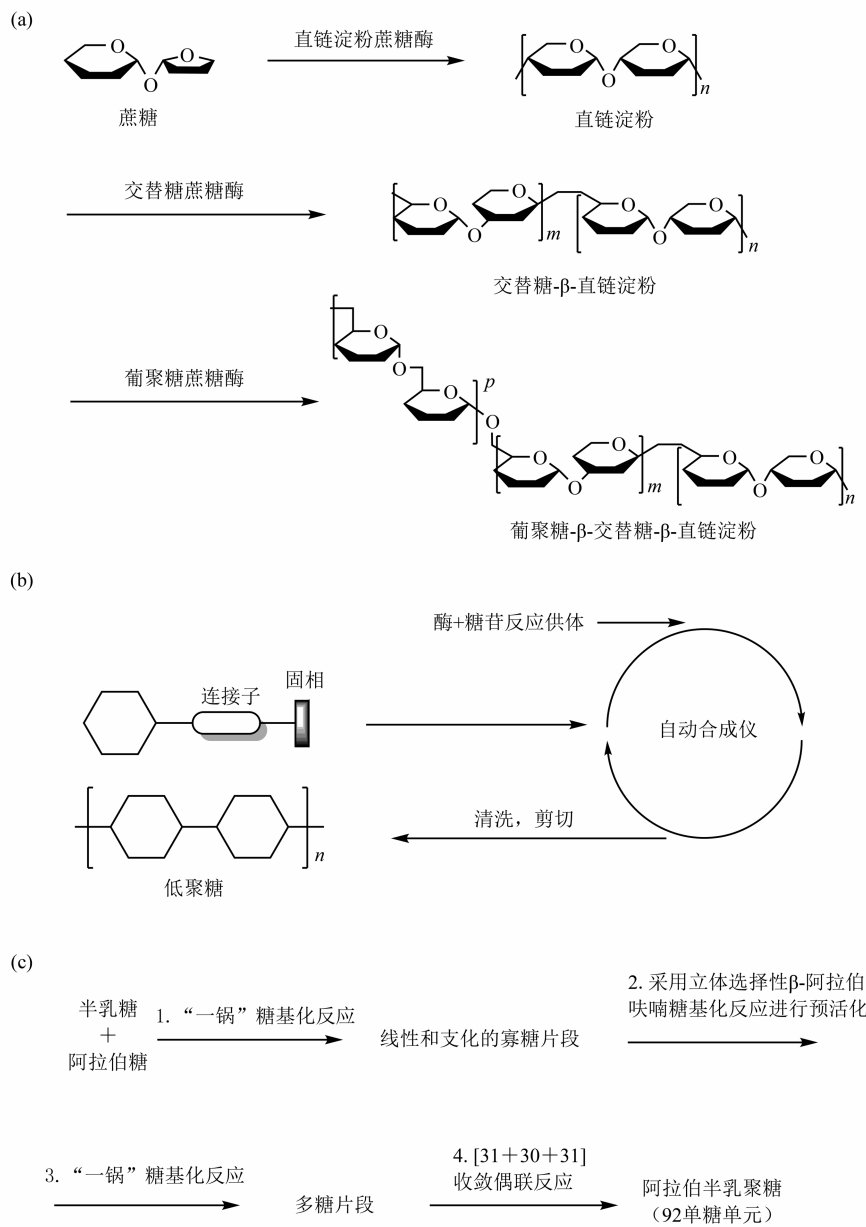


Fig. 5 Three methods of *de novo* glycan synthesis^[73-74, 76]

图5 三种从头合成聚糖的方法^[73-74, 76]

(a) 三嵌段葡聚糖酶促合成法; (b) 低聚糖酶促自动合成法; (c) 阿拉伯半乳糖的化学合成法.

Wu等^[76]首次合成出含有92个糖单体且高度支化的阿拉伯半乳糖，研究者首先采用一锅糖基化策略^[77]合成出主链、分支所需的寡糖或多糖片段，之后进行立体选择性 β -糖基化，最后对A链及B链进行链内组装及链间偶联得到完整多糖（图5c）。该研究在以化学合成法合成人工定义结构的多糖方面是一项重要进展。Wang等^[78]采用磷酸和微波分别作为催化剂和辅助手段，以D-甘露糖作为底物从头合成出重均分子质量为2.457 ku的聚甘露糖，实现了甘露糖的低分子质量聚合。Kanazawa等^[79]同样以磷酸作为催化剂，在110℃氮气氛围下以天然戊醛和6-脱氧己糖作为底物，通过固态缩聚反应得到重均分子质量范围为2.7~12.0 ku、高度支化的多糖，底物转化率为47%~81%；该反应整个过程中混合物均成粉末状，为多糖的固态合成提供了实际参考。可见，化学合成法能够在物理场作用以及温度变化较大的反应条件下进行，且在多糖结构上具有一定的选择性。

3.2 从头设计多糖

在多糖或糖缀合物的设计中，由于存在大量可能出现的单体糖残基和缀合物配体种类，因此从头设计方法的构建便变得较为困难。

Trott等^[80]开发了名为AutoDock Vina的设计模拟软件，较AutoDock 4在计算速度、预测准确性上有显著提高，该软件也可用于部分多糖或糖缀合物结构的设计预测，但在糖蛋白的设计上没有有效的解决办法。Labonte等^[81]采用RosettaCarbohydrate软件对多糖及糖缀合物进行结构预测，软件以糖残基为中心，对糖苷键、侧链构象、环形式等参数进行分析建模，并具有一定的模拟糖基化功能。可见，和蛋白质从头设计中广泛使用的Rosetta软件相比，构建多糖和糖缀合物的从头设计方法具有更高的复杂性和难度，仍需研究者进行进一步的计算开发。

3.3 从头合成和设计多糖的应用

纯粹的多糖由于受自身基团的限制，较难在功能上实现多样化，而结合其他大分子将大大拓展多糖的应用范围。其中，糖缀合物指糖链以共价键的形式与蛋白质、多肽、脂质、核酸及抗体等生物分子连接而成的一类物质，其反应机理为糖基化反应。对于生命体而言，糖链或糖缀合物具有细胞识别、转运物质等功能，其对生命科学中病理和药理学研究至关重要。因此，对目前糖缀合物研究近况的认识将有助于利用从头合成与设计技术定向制备

具有更优特性的糖链或糖缀合物，同时推进多糖从头合成和设计技术的发展。

Yang等^[82]描述了33种复合天然糖苷从头合成的O-糖基化反应过程，详细分析了每种反应的糖基化条件及结果，为目前糖缀合物的从头合成和设计提供了十分重要的机理参考。Li等^[83]对金催化的糖基化反应机理进行了阐述，并对基于相关糖缀合物合成验证了该方法的可行性，为从头合成和设计糖缀合物提供了催化条件。Hjuler等^[84]介绍了一种由脱保护的碳水化合物形成糖缀合物的三段法，具体为将具有缀合位点及保护作用的化合物连接到低聚糖的还原端，其次通过脱保护以暴露缀合位点，最后根据需要发生缀合反应。由于缀合位点拥有较好的反应自由性，该糖缀合物可应用在生物成像、糖-蛋白质相互作用研究等方面。Kong等^[85]开发出一种以D-甘露糖作为简单前体合成四糖，进而与白喉毒素突变体载体蛋白进行缀合。该缀合物在体外实验中能与荚膜多糖转运蛋白抑制剂产生协同作用，结合细菌表面多糖，进而诱导免疫系统产生抗体用于杀灭细菌，推进了糖缀合物的抗菌疫苗开发。Nam等^[86]对具有优异基因转染效率的聚乙烯亚胺（polyethyleneimine, PEI）进行了改造，将其低分子质量的PEI和靶向配体分别与羧甲基壳聚糖的氨基和羧甲基进行缀合以获得HPOCP共聚物；对该物质进行表征和细胞实验后表明，基因-HPOCP复合物拥有较好的结合力、保护与释放能力以及较低的细胞毒性，具有作为基因治疗中基因递送载体的潜力。由此可见，糖缀合物在生物材料、疫苗、基因治疗等领域具有潜力，而对多糖定向从头合成的研究将有助于合成出具有更优载体能力、抗菌活性等特点的糖缀合物。

4 总结与展望

随着从头合成和设计技术的不断发展，将可以根据自身设计理念合成出全新的DNA或基因组，或者使其进行表达生成全新的表达产物。DNA从头合成和设计的成熟研究在一定程度上加速了蛋白质的从头合成和设计，为氨基酸序列设计的可行性提供了重要保障，进而可以开发出诸如白介素2模拟物、镜像蛋白等对医药、生物领域影响重大的人工产物。此外，多糖尤其是具有识别功能的糖链或糖缀合物是生命科学领域中病理和药理学研究的核心内容，结合从头合成和设计技术将可以批量定向制备具有优良生物活性的糖链或糖缀合物，有利于推

动诸如HPOCP等基因治疗药物递送载体及其他功能物质的开发. 从长远角度看, 从头合成和设计技术未来将与自动化技术甚至人工智能技术深度融合, 有望以“一键合成”的简单操作合成复杂的生物大分子, 而深度融合的成果将使自然科学研究向前迈进一大步.

参 考 文 献

- [1] Polizzi N F, Wu Y, Lemmin T, *et al.* *De novo* design of a hyperstable non-natural protein – ligand complex with sub-Å accuracy. *Nature Chemistry*, 2017, **9**(12): 1157-1164
- [2] Barrangou R, Fremaux C, Deveau H, *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 2007, **315**(5819): 1709-1712
- [3] Gibson D G. Programming biological operating systems: genome design, assembly and activation. *Nature Methods*, 2014, **11**(5): 521-526
- [4] Gibson D G, Benders G A, Andrews-Pfannkoch C, *et al.* Complete chemical synthesis, assembly, and cloning of a mycoplasma genitalium genome. *Science*, 2008, **319**(5867): 1215-1220
- [5] Beaucage S L, Caruthers M H. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters*, 1981, **22**(20): 1859-1862
- [6] Kosuri S, Church G M. Large-scale *de novo* DNA synthesis: technologies and applications. *Nature Methods*, 2014, **11**(5): 499-507
- [7] Hayakawa Y, Kawai R, Hirata A, *et al.* Acid/azole complexes as highly effective promoters in the synthesis of DNA and RNA oligomers *via* the phosphoramidite method. *Journal of the American Chemical Society*, 2001, **123**(34): 8165-8176
- [8] Satheesh E, Phillips J D, Heemstra J M. Synthesis of comb-shaped DNA using a non-nucleosidic branching phosphoramidite. *Organic & Biomolecular Chemistry*, 2018, **16**(25): 4659-4664
- [9] Gao X L, LeProust E, Zhang H, *et al.* A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. *Nucleic Acids Research*, 2001, **29**(22): 4744-4750
- [10] Egeland R D, Marken F, Southern E M. An electrochemical redox couple activated by microelectrodes for confined chemical patterning of surfaces. *Analytical Chemistry*, 2002, **74**(7): 1590-1596
- [11] Levrie K, Jans K, Schepers G, *et al.* Direct on-chip DNA synthesis using electrochemically modified gold electrodes as solid support. *Japanese Journal of Applied Physics*, 2018, **57**(4S): 04FM01
- [12] Saaem I, Ma K S, Marchi A N, *et al.* *In situ* synthesis of DNA microarray on functionalized cyclic olefin copolymer substrate. *ACS Applied Materials & Interfaces*, 2010, **2**(2): 491-497
- [13] Quan J, Saaem I, Tang N, *et al.* Parallel on-chip gene synthesis and application to optimization of protein expression. *Nature Biotechnology*, 2011, **29**(5): 449-452
- [14] Palluk S, Arlow D H, Rond T D, *et al.* *De novo* DNA synthesis using polymerase-nucleotide conjugates. *Nature Biotechnology*, 2018, **36**(7): 645-650
- [15] Ramadan K, Shevelev I V, Maga G, *et al.* *De novo* DNA synthesis by human DNA polymerase λ , DNA polymerase μ and terminal deoxyribonucleotidyl transferase. *Journal of Molecular Biology*, 2004, **339**(2): 395-404
- [16] Veneziano R, Shepherd T R, Ratanalert S, *et al.* *In vitro* synthesis of gene-length single-stranded DNA. *Scientific Reports*, 2018, **8**(1): 6548
- [17] Jin P, Ding W, Du G, *et al.* DATEL: A scarless and sequence-independent DNA assembly method using thermostable exonucleases and ligase. *ACS Synthetic Biology*, 2016, **5**(9): 1028-1032
- [18] Potapov V, Ong J L, Kucera R B, *et al.* Comprehensive profiling of four base overhang ligation fidelity by T4 DNA Ligase and application to DNA assembly. *ACS Synthetic Biology*, 2018, **7**(11): 2665-2674
- [19] Andreou A I, Nakayama N. Mobius assembly: a versatile Golden-Gate framework towards universal DNA assembly. *Plos One*, 2018, **13**(1): e0189892
- [20] Gibson D G, Young L, Chuang R Y, *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 2009, **6**(5): 343 – 345
- [21] Li L, Jiang W, Lu Y. *Synthetic metabolic pathways*. New York: Humana Press, 2018: 203-209
- [22] Yuliya K, Weyman P D, Glass J I, *et al.* DNA assembly with error correction on a droplet digital microfluidics platform. *BMC Biotechnology*, 2018, **18**(1): 37
- [23] Richardson S M, Mitchell L A, Stracquandano G, *et al.* Design of a synthetic yeast genome. *Science*, 2017, **355**(6329): 1040-1044
- [24] Li B, Vilardell J, Warner J R. An RNA structure involved in feedback regulation of splicing and of translation is critical for biological fitness. *Proc Natl Acad Sci USA*, 1996, **93**(4): 1596-1600
- [25] Kozak M. Regulation of translation *via* mRNA structure in prokaryotes and eukaryotes. *Gene*, 2005, **361**: 13-37
- [26] Bellaousov S, Reuter J S, Seetin M G, *et al.* RNAstructure: web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Research*, 2013, **41**(W1): W471-W474
- [27] Mathews D H, Andre T C, Kim J, *et al.* An updated recursive algorithm for RNA secondary structure prediction with improved thermodynamic parameters// Leontis N B, SantaLucia J. *ACS Symposium Series: 682*. Washington DC: American Chemical Society, 1997: 246-259
- [28] Legendre A, Angel E, Tahi F. Bi-objective integer programming for RNA secondary structure prediction with pseudoknots. *BMC Bioinformatics*, 2018, **19**(1): 13
- [29] Zhu Y, Xie Z, Li Y, *et al.* Research on folding diversity in statistical learning methods for RNA secondary structure prediction. *International Journal of Biological Sciences*, 2018, **14**(8): 872-882.

- [30] Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. *Trends in Biotechnology*, 2004, **22**(7): 346-353
- [31] Chin J X, Chung K S, Lee D Y. Codon optimization online (COOL): a web-based multi-objective optimization platform for synthetic gene design. *Bioinformatics*, 2014, **30**(15): 2210-2212
- [32] Papamichail D, Liu H, Machado V, *et al.* Codon context optimization in synthetic gene design. *IEEE/ACM Transactions on Computational Biology & Bioinformatics*, 2018, **15**(2): 452-459
- [33] Mitchell L A, Wang A, Stracquadanio G, *et al.* Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science*, 2017, **355**(6329): eaaf4831
- [34] Wu Y, Li B Z, Zhao M, *et al.* Bug mapping and fitness testing of chemically synthesized chromosome X. *Science*, 2017, **355**(6329): eaaf4706
- [35] Mercy G, Mozziconacci J, Scolari V F, *et al.* 3D organization of synthetic and scrambled chromosomes. *Science*, 2017, **355**(6329): eaaf4597
- [36] Xie Z X, Li B Z, Mitchell L A, *et al.* "Perfect" designer chromosome V and behavior of a ring derivative. *Science*, 2017, **355**(6329): eaaf4704
- [37] Rosenbaum D M, Rasmussen S G, Kobilka B K. The structure and function of G-protein-coupled receptors. *Nature*, 2009, **459**(7245): 356-363
- [38] Miranda L P, Alewood P F. Challenges for protein chemical synthesis in the 21st century: bridging genomics and proteomics. *Biopolymers*, 2000, **55**(3): 217-226
- [39] Johnson E C, Kent S B. Insights into the mechanism and catalysis of the native chemical ligation reaction. *Journal of the American Chemical Society*, 2006, **128**(20): 6640-6646
- [40] Dawson P E, Muir T W, Clark-Lewis I, *et al.* Synthesis of proteins by native chemical ligation. *Science*, 1994, **266**(5186): 776-779
- [41] Fang G M, Li Y M, Shen F, *et al.* Protein chemical synthesis by ligation of peptide hydrazides. *Angewandte Chemie International Edition*, 2011, **50**(33): 7645-7649
- [42] Zheng J S, Tang S, Qi Y K, *et al.* Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nature Protocols*, 2013, **8**(12): 2483-2495
- [43] Wan Q, Danishefsky S J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angewandte Chemie International Edition*, 2007, **46**(48): 9248-9252
- [44] Leaver-Fay A, Tyka M, Lewis S M, *et al.* Chapter nineteen rosetta3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in Enzymology*, 2011, **487**: 545-574
- [45] Ovchinnikov S, Kinch L, Park H, *et al.* Large-scale determination of previously unsolved protein structures using evolutionary information. *eLife*, 2015, **4**: e09248
- [46] Lu P, Min D, Dimaio F, *et al.* Accurate computational design of multipass transmembrane proteins. *Science*, 2018, **359**(6379): 1042-1046
- [47] Bertolani S J, Carlin D A, Siegel J B. Computational design of ligand binding proteins. New York: Humana Press, 2016: 213-231
- [48] Zanghellini A. *De novo* computational enzyme design. *Current Opinion in Biotechnology*, 2014, **29**(1): 132-138
- [49] Marvic M, Hu H, Lu Z, *et al.* *De novo* designed transmembrane peptides activating the $\alpha 5 \beta 1$ integrin. *Protein Engineering, Design and Selection*, 2018, **31**(5): 181-190
- [50] Boyken S E, Chen Z, Groves B, *et al.* *De novo* design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science*, 2016, **352**(6286): 680-687
- [51] André I, Bjelic S. Computational assessment of folding energy landscapes in heterodimeric coiled coils. *Proteins: Structure, Function, and Bioinformatics*, 2018, **86**(7): 790-801
- [52] Das R, André I, Shen Y, *et al.* Simultaneous prediction of protein folding and docking at high resolution. *Proc Natl Acad Sci USA*, 2009, **106**(45): 18978-18983
- [53] Ludwiczak J, Jarmula A, Dunin-Horkawicz S. Combining rosetta with molecular dynamics (MD): a benchmark of the MD-based ensemble protein design. *Journal of Structural Biology*, 2018, **203**(1): 54-61
- [54] Huang P S, Feldmeier K, Parmeggiani F, *et al.* *De novo* design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chemical Biology*, 2016, **12**(1): 29-34
- [55] Doyle L, Hallinan J, Bolduc J, *et al.* Rational design of α -helical tandem repeat proteins with closed architectures. *Nature*, 2015, **528**(7583): 585-588
- [56] Thomas F, Dawson W M, Lang E J M, *et al.* *De novo* designed α -helical barrels as receptors for small molecules. *ACS Synthetic Biology*, 2018, **7**(7): 1808-1816
- [57] Noguchi H, Addy C, Simoncini D, *et al.* Computational design of symmetrical eight-bladed β -propeller proteins. *IUCrJ*. 2019, **6**(Pt1): 46-55
- [58] Sockolosky J T, Trotta E, Parisi G, *et al.* Selective targeting of engineered T cells using orthogonal IL-2 cytokine-receptor complexes. *Science*, 2018, **359**(6379): 1037-1042
- [59] Gallardo R, Ramakers M, De Smet F, *et al.* *De novo* design of a biologically active amyloid. *Science*, 2016, **354**(6313): aah4949
- [60] Silva D A, Yu S, Ulge U Y, *et al.* *De novo* design of potent and selective mimics of IL-2 and IL-15. *Nature*, 2019, **565**(7738): 186-191
- [61] Chevalier A, Silva D A, Rocklin G J, *et al.* Massively parallel *de novo* protein design for targeted therapeutics. *Nature*, 2017, **550**(7674): 74-79
- [62] Zheng K, Ling S. *De novo* design of recombinant spider silk proteins for material applications. *Biotechnology Journal*, 2018, **14**(1): e1700753
- [63] Rapson T D, Sutherland T D, Church J S, *et al.* *De novo* engineering of solid-state metalloproteins using recombinant coiled-coil silk. *ACS Biomaterials Science & Engineering*, 2015, **1**(11): 1114-1120
- [64] Bozhüyük K A J, Fleischhacker F, Linck A, *et al.* *De novo* design and engineering of non-ribosomal peptide synthetases. *Nature*

- Chemistry, 2018, **10**(3): 275-281
- [65] Weinstock M T, Jacobsen M T, Kay M S. Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity. *Proc Natl Acad Sci USA*, 2014, **111**(32): 11679-11684
- [66] Wang Z, Xu W, Liu L, *et al.* A synthetic molecular system capable of mirror-image genetic replication and transcription. *Nature Chemistry*, 2016, **8**(7): 698-704
- [67] Pan M, Gao S, Zheng Y, *et al.* Quasi-racemic X-ray structures of K27-linked ubiquitin chains prepared by total chemical synthesis. *Journal of the American Chemical Society*, 2016, **138**(23): 7429-7435
- [68] Ohtsubo K, Marth J D. Glycosylation in cellular mechanisms of health and disease. *Cell*, 2006, **126**(5): 855-867
- [69] Zhao G, Guan W, Cai L, *et al.* Enzymatic route to preparative-scale synthesis of UDP - GlcNAc/GalNAc, their analogues and GDP - fucose. *Nature Protocols*, 2010, **5**(4): 636-646
- [70] Badenhorst C P S, Bornscheuer U T. Getting momentum: from biocatalysis to advanced synthetic biology. *Trends in Biochemical Sciences*, 2018, **43**(3): 180-198
- [71] Park M O, Chandrasekaran M, Yoo S H. Expression, purification, and characterization of a novel amylsucrase from *Neisseria subflava*. *International Journal of Biological Macromolecules*, 2018, **109**: 160-166
- [72] Song L, Miao M, Jiang B, *et al.* *Leuconostoc citreum* SK24.002 glucansucrase: biochemical characterisation and *de novo* synthesis of α -glucan. *International Journal of Biological Macromolecules*, 2016, **91**: 123-131
- [73] Grimaud F, Faucard P, Tarquis L, *et al.* Enzymatic synthesis of polysaccharide-based copolymers. *Green Chemistry*, 2018, **20**(17): 4012-4022
- [74] Wen L, Edmunds G, Gibbons C, *et al.* Toward automated enzymatic synthesis of oligosaccharides. *Chemical Reviews*, 2018, **118**(17): 8151-8187
- [75] Xiao R, Grinstaff M W. Chemical synthesis of polysaccharides and polysaccharide mimetics. *Progress in Polymer Science*, 2017, **74**: 78-116
- [76] Wu Y, Xiong D C, Chen S C, *et al.* Total synthesis of mycobacterial arabinogalactan containing 92 monosaccharide units. *Nature Communications*, 2017, **8**: 14851
- [77] Chao C S, Yen Y F, Hung W C, *et al.* Solvent participation in a one-pot glycosylation strategy (SPOG). *Advanced Synthesis & Catalysis*, 2011, **353**(6): 879-884
- [78] Wang H, Cheng X, Shi Y, *et al.* Preparation and structural characterization of poly-mannose synthesized by phosphoric acid catalyzation under microwave irradiation. *Carbohydrate Polymers*, 2015, **121**: 355-361
- [79] Kanazawa A, Suzuki M. Solid-state polycondensation of natural aldopentoses and 6-deoxyaldohexoses. Facile preparation of highly branched polysaccharide. *Polymer*, 2006, **47**(1): 176-183
- [80] Trott O, Olson A J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 2010, **31**(2): 455-461
- [81] Labonte J W, Adolf-Bryfogle J, Schief W R, *et al.* Residue-centric modeling and design of saccharide and glycoconjugate structures. *Journal of Computational Chemistry*, 2016, **38**(5): 276-287
- [82] Yang Y, Zhang X, Yu B. O-glycosylation methods in the total synthesis of complex natural glycosides. *Natural Product Reports*, 2015, **32**(9): 1331-1355
- [83] Li W, Yu B. Gold-catalyzed glycosylation in the synthesis of complex carbohydrate-containing natural products. *Chemical Society Reviews*, 2018, **47**(21): 7954-7984
- [84] Hjuler C T, Maolanon N N, Sauer J, *et al.* Preparation of glycoconjugates from unprotected carbohydrates for protein-binding studies. *Nature Protocols*, 2017, **12**(11): 2411-2422
- [85] Kong L, Vijayakrishnan B, Kowarik M, *et al.* An antibacterial vaccination strategy based on a glycoconjugate containing the core lipopolysaccharide tetrasaccharide Hep2Kdo2. *Nature Chemistry*, 2016, **8**(3): 242-249
- [86] Nam J P, Nah J W. Target gene delivery from targeting ligand conjugated chitosan-PEI copolymer for cancer therapy. *Carbohydrate Polymers*, 2016, **135**: 153-161

Advances in *De Novo* Synthesis and *De Novo* Design of Biomacromolecules*

XU Xin-Dong, QI Peng-Xiang, LAN Wei-Bing, CHEN Yu-Ying, CHEN Shan**

(Light Industry and Food Engineering College, Collaborative Innovation Center of the Sucrose Industry Co-Sponsored by Province and Ministry,
Guangxi University, Nanning 530004, China)

Abstract Biomacromolecules refer to the substances such as DNA, proteins, polysaccharides, *etc.* in living organisms, which are essential for the normal life activities of organisms. *De novo* synthesis and design techniques have high degree of freedom and simple precursors in the synthesis and structural design of biomacromolecules, enabling new design and efficient synthesis of biomacromolecules for specific research purposes. Recently, *de novo* synthesis and design techniques have begun to receive attention in the fields of artificial gene combination, novel protein drug design, glycoconjugate synthesis, *etc.* Based on the *de novo* synthesis and design of biomacromolecules, the newly designed DNA or novel gene expression products, as well as glycosylation or glycoconjugates with recognition functions, can be targeted prepared. Simultaneously, the techniques will greatly advance biologically active substances such as cytokine mimics, gene therapy delivery vehicles, providing a new solution for the construction of artificial biological systems and the treatment of rare diseases. This paper reviews the *de novo* synthesis and design of DNA, protein and polysaccharides, expounds the related methods and applications, and finally summarizes the relationship between three kinds of substances.

Key words *de novo* synthesis, *de novo* design, DNA, protein, polysaccharide, method, application

DOI: 10.16476/j.pibb.2019.0094

* This work was supported by a grant from The Regional Science Foundation of National Natural Science Foundation of China (21768001).

** Corresponding author.

Tel: 86-771-3231590, E-mail: chenshan@gxu.edu.cn

Received: May 1, 2019; Accepted: June 28, 2019