

计算机模拟在优化核酸适体筛选分析流程中的应用研究

王明华¹, 李杜娟^{2*}

(1.忻州师范学院生物系, 山西 忻州 034000;

2.杭州电子科技大学生命信息与仪器工程学院, 浙江 杭州 310018)

摘要:通过分析列举成功筛选应用实例及研究新趋势,对计算机模拟辅助虚拟筛选核酸适体方法的适用性、程序性、必要性进行综述和分析,为计算机虚拟筛选技术推动核酸适体的开发和应用提供新的思路。

关键词: 适体; 分子模拟; 结构预测; 虚拟筛选

中图分类号: O6-04; O629.7; Q811.4

文献标志码: A

文章编号: 0253-4320(2018)11-0216-05

DOI: 10.16606/j.cnki.issn 0253-4320.2018.11.047

Application of computer simulation in optimization of nucleic acid aptamer screening and analysis

WANG Ming-Hua¹, LI Du-Juan^{2*}

(1. Biology Department, Xinzhou Teachers University, Xinzhou 034000, China; 2. College of Life Information Science and Instrument Engineering, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: This paper reviews and analyzes the applicability, procedure and necessity of the method for virtual screening nucleic acid aptamer assisted by computer simulation, based on listing the successful application examples in screening and studying new trends. It aims to provide a new idea for computer virtual screening technology to promote the development and application of nucleic acid aptamers.

Key words: aptamer; molecular simulation; structure prediction; virtual screening

核酸适体是 20 世纪 90 年代发展的一类体外筛选的可与目标分子高效、高特异性亲合的单链 RNA、双链 DNA 或单链 DNA 寡核苷酸片段^[1], 具有相对于抗体的诸多优势, 在医药、生物传感器等领域有广泛的研究和应用。适体的筛选是通过“指数富集配体系统进化体外筛选技术”(systematic evolution of ligands by exponential enrichment, SELEX), 在众多随机序列中通过多次扩增-亲合筛选-分离的循环获得。循环过程中关键步骤之一是使用高效的分析检定、分离技术和仪器(如生物传感器等)将与目标物具有高亲和性的“适体”分离, 因而随机影响多, 耗费较大, 制约其应用的发展。

核酸适体的筛选和应用基础是在单分子水平上与目标分子之间的相互作用^[2], 主要机理与分子的空间结构特征密切相关。生物大分子空间结构研究的实验技术是基于对所获取纯净分子结晶进行 X 射线衍射或核磁共振成像来解析其空间结构, 对于

科研条件的需求较高, 研究周期也比较长。其研究成果则形成了便于利用的各种生物信息学数据库资源, 满足科研的需求。

与此同时, 计算机和信息技术的快速发展, 为科学家提供了便利。分子模拟研究^[3]是依赖于量子力学、分子力学, 在分子或原子水平上利用计算机来构造、实现、分析和存储复杂的分子模型, 计算微观粒子之间的相互作用的方法, 可提供直观的分子立体图像, 核酸适体筛选研究中也涉及。本文中将对计算机分子模拟相关技术、方法在核酸适体筛选、应用方面的主要方式及相关结果进行分析, 提出应用计算机技术改进核酸适体筛选、应用的思路和建议。

1 核酸适体空间结构的计算机模拟

核酸适体识别目标分子高亲合性的基础是其特异性的空间结构, 二者在接触面上能形成原子或分

收稿日期: 2018-04-12; 修回日期: 2018-09-05

基金项目: 国家自然科学基金项目(31201367); 浙江省自然科学基金项目(LY15H200003); 忻州师范学院博士启动基金(2015)

作者简介: 王明华(1982-), 女, 博士, 副教授, 研究方向为生物大分子结构与功能, wmh820312@126.com; 李杜娟(1983-), 女, 博士, 副研究员, 研究方向为生物传感器, 通讯联系人, dujuanli@hdu.edu.cn。

子间的相互作用。寡核苷酸的不同碱基间的相互作用即可形成具有不同特异性空间结构和状态的空间结构模体,对其进行预测有助于了解核酸适体的功能。Szczesniak等^[4]对非编码RNA-RNA相互作用预测研究中考虑RNA二级结构时增进了预测准确性;Fiannaca等^[5]以13类RNA特征二级结构进行非编码RNA结构预测分类,获得了74%的准确率;表明RNA的空间结构受到其二级结构特征的影响。研究报道中常见模体有G-四聚体、发夹、茎环等^[6],相关序列和空间结构特征被越来越多地收录到数据库中^[7]。借助计算机的运算能力和生物信息学数据的积累,而对寡核苷酸序列结构及碱基排列顺序的规律性深入研究,有助于揭示和预测核酸适体的空间结构。

1.1 RNA 适体空间结构预测

RNA在天然状态下大多为单核苷酸链,mRNA启动识别区、tRNA等由相近碱基间相互作用所形成的特定空间结构实现其生物功能。深入研究RNA天然结构的形成方式,有助于开发RNA适体空间结构预测的相关方法和程序。早期RNA序列二级结构的预测方法多依赖碱基排列顺序形成的分子自由能、原子间相互作用的距离和稳定性,如mfold^[8]预测方法等。为了提高预测的准确度,Reuter等^[9]提出了基于已知小片段的二级结构,同时考虑热力学参数和RNA 2'-羟基酰化实验数据限制的RNA structure方法,用来预测单链RNA的二级结构。近年来,随着对核酸结构模体种类和作用机理研究的深入,预测程序和方法报道增长迅速,并且方法中较多地参考已知的各种RNA结构模体、空间形状特征。RNAscClust^[10]方法以RNA结构模体作为预测分析的基础。JNSViewer^[11]内置能够获得一致性预测结果的多种DNA/RNA预测方法以提高预测准确性,并提供二级结构的图形。这些方法促进了RNA二级结构预测的实用性。

预测RNA适体的空间结构是以预测的二级结构为基础的。Biesiada等^[12]提出RNAComposer程序来预测RNA空间结构,其输入为RNA序列,输出为预测分子能量最低或最多聚类的RNA空间结构三维图形和数据,便于进行分子对接的虚拟筛选。此外,Das等^[13]提出了天然RNA空间结构的从头预测方法FARNA,通过全原子模型的应用来预测一般非典型区域的RNA空间结构。

综上所述,虽然对于RNA适体空间结构的预测

程序种类较多,所依据的原理、采用的算法、对结果输出的评价体系和方法也有所差异,但总体趋势是更加注重运算速度和已知RNA结构模体类型对适体空间结构的影响。因而,在进行以针对蛋白质类目标物的RNA适体设计筛选时,其空间结构的预测应当依据RNA适体序列长度、结构模体特征、作用对象性质、结构评价、输出文件方式、计算机性能等,选取适宜的预测方法。

1.2 DNA 适体空间结构预测

与RNA不同,天然状态的DNA主要为碱基互补配对原则所决定的稳定双螺旋结构,因而形成的高级结构也较为相近。其空间结构的预测和重建亦以双链稳定结构的形成为基础,偏向大量DNA所形成的纳米量级特殊结构和功能的预测。预测程序主要有3D-DART^[14]、oxDNA^[15]等。对于适体适用的单链短DNA,其二级结构预测方法报道较少,且主要关注PCR过程中的单链构象多态性研究,适用于适体的预测方法一般近似RNA,空间结构预测的专用方法和程序鲜有报道。RNA适体由于其核糖上所具有的2'-羟基未参与形成有稳定氢键的碱基配对,故而基团的反应活性较大,从而影响功能稳定性。因此对DNA适体空间结构的研究有利于了解DNA适体的作用机理,而对空间结构模体认识的积累促进了新的应用。2017年,Jeddi等^[16]对单链DNA序列做二级结构预测后,等效转换单链RNA模型,再将单链RNA模型转换成单链DNA空间结构,最后精炼得到单链DNA空间结构,为DNA适体的虚拟筛选提供了发展的思路。

2 适体-蛋白相互作用分子模拟预测

在适体筛选中,对能够发生相互作用的核酸适体与目标物形成的复合物进行收集、分离、鉴别是筛选的关键步骤,并且这一过程受到初始筛选序列集大小、筛选方法本身效率及具体的操作条件、分离鉴别方法的有效性等诸多因素影响。此外,这一过程还需要不断重复,耗费大量时间和研究精力,因此通常会影响到适体的成功筛选。蛋白质与RNA或DNA相互作用(如DNA复制、RNA转录等)是生命活动的基础,许多特异性的相互作用模式和规律已经被解析出来,成为理解和预测适体-配体相互作用的基础,并用于相关的高通量虚拟研究。由于适体目标种类繁多,如小分子、离子、细胞等,因此分子模拟研究方法需要具有较强的针对性,本文中主要关注

适体-蛋白相互作用分子模拟预测。

对于能够发生潜在相互作用的 RNA 碱基位点区域,以及蛋白质氨基残基位点区域进行预测的方法有 RsiteDB^[17]、PRIdictor^[18]等。RsiteDB 是用于预测能与 RNA 碱基发生相互作用的蛋白质作用位点的数据库。PRIdictor 是一种直接输入生物大分子序列的考虑蛋白-RNA 相互作用模式,在氨基残基-核苷水平预测蛋白质、RNA 各自的相互作用位点的方法。

对于适体-蛋白质相互作用复合物空间结构的分子对接预测方法有 ZDOCK^[19]、RosettaDock^[20]、NPDOCK^[21]等。ZDOCK 是从相互作用分子的三维形状角度,在相关量子力学理论下进行计算、搜寻分子间结合模式的刚性对接方法,适用于多种生物大分子间相互作用模拟;RosettaDock 考虑了一定的分子柔性,是预测蛋白-蛋白、蛋白-适体的相互作用的普遍性方法。2015 年报道的 NPDock 是针对核酸-蛋白质相互作用的在线预测方法,因而在程序设定和变量选择上有一定的针对性,最佳构象也经聚类分析筛选。已经报道的适体虚拟筛选成果中,用到比较多的是 ZDOCK 方法^[22-23]。

综上所述,适体-蛋白质相互作用的预测方法以 RNA 适体为主,这可能与现阶段研究者对于非编码 RNA 所具有的基因表达调控等生物功能关注及研究的深入有关,因此其预测方法的发展也逐渐增加了针对性,关注空间结构外观及不同生物大分子间相互作用主要作用力之间的差异。

3 计算机辅助的适体虚拟筛选实例

随着人们对于适体结构和功能的深入研究,以及生物信息学成果的累积,采用分子模拟研究对适体筛选技术进行改进的需求和可行性不断增加,现有的创新应用主要体现在适体设计和筛选范围的紧缩 2 方面。

3.1 适体设计

计算机辅助设计适体的方法实现是以 RNA、DNA 高级结构预测、核酸-蛋白质相互作用模拟预测的具体程序和方法为基础,从而实现使用计算机辅助筛选适体的创新应用。

2009 年,Chushak 等^[24]在提出相应评价标准的基础上,利用计算程序在 RNA 二级结构的基础上设计适体筛选的起始序列集,进而预测三维结构,并与选定的目标小分子进行高通量对接的虚拟筛选,与

已知实验筛选的高亲和适体相比,虚拟筛选适体同样对目标具有高亲和力。2013 年,Savory 等^[25]利用计算机产生变异序列筛选适体,在 2 个循环后获得了高于亲和力大肠杆菌 36% 的变形杆菌特异性适体。2014 年,Savory 等^[26]又采用定量 PCR 控制的 cell-SELEX 结合指定适体二级构象,筛选 DNA 适体检测尿路致病性大肠杆菌,并提出了计算机辅助筛选方法的一般流程。

单链 DNA 适体筛选方面,2015 年,Shcherbinin 等^[22]在计算机上首先确定蛋白质适体相互作用位点,然后设计适体高级结构,随之对二者的相互作用通过分子对接程序进行评估,设计筛选了细胞色素 P450 单链 DNA 适体,并通过 SPR 生物传感器验证该适体对目标的高特异性。2016 年,Heiat 等^[23]在 discovery studio 药物发现与生物大分子计算模拟平台上将预测的 RNA 三维结构转换成代换碱基的 DNA 三维结构,并用 ZDOCK 程序模拟分离出的单链 DNA 适体与目标蛋白的相互作用,并将结果与实验研究结果印证,成功筛选出了血管紧张素 II 的 DNA 适体。2017 年,Yokoyama 等^[27]采用 G-4 聚体启动子引导的适体筛选方法(G4 promoter-derived aptamer selection,G4PAS)和计算机扩展序列对接分析方法(图 1 左图)成功筛选了肝细胞生长因子的特异性 DNA 适体。这些研究表明,随着 RNA、DNA 结构功能、生物信息学研究的积累和创新,计算机辅助的 RNA、DNA 高级结构预测方法在设计适体过程中作用突出,有利于实现蛋白类目标物特异性适体的高效成功筛选。

3.2 适体筛选范围紧缩

在适体的筛选过程中,首先需要形成一个数量巨大的序列库,其中的每一条序列与其靶标物相互作用都有一定的亲和力,都应当进行与目标物相互作用的研究,因此筛选范围对于适体筛选的效率存在较大的影响。2015 年,Fukaya 等^[28]通过计算机对三聚化适体的关键突环序列进行 5 轮复选后,获得了具有高亲和力的血管内皮生长因子的特异性 DNA 适体。2015 年,Zhou 等^[29]报道了一种应用计算机虚拟搜索适体序列范围的方法,包含 2 个步骤(图 1 右图)。首先以期望获得的二级结构模体为目标,获取其对应的序列,形成筛选集,然后以基于分子动力学模拟的方法对序列进行虚拟筛选。从 413 条 RNA 序列中筛选出的 6 个茶碱特异性适体,在验证性实验中均具有相对较高的亲和力。2016 年,

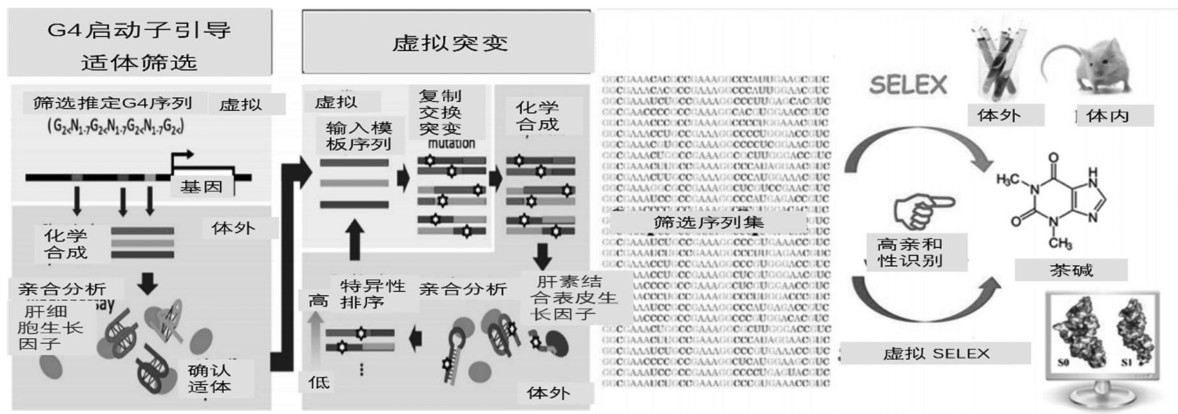


图1 适体虚拟筛选流程

(左,肝细胞生长因子特异性适体^[27];右,茶碱特异性适体^[28])

Ahirwar 等^[30]应用分子模拟虚拟筛选雌激素 α 受体高亲和 RNA 适体,并经过了试验验证。由此表明分子模拟研究技术的应用,对于提高适体筛选效率具有较为明显的效果和应用潜力。

4 结论与展望

随着对非编码 RNA 结构和功能的进一步深入认识和研究,进行作用目标明确的核酸适体的筛选和设计,正逐渐成为既有研究基础和资源又有广阔推广应用潜力和市场的新领域。适体与目标分子的相互作用以各自形成的特异性空间结构为基础,生物大分子的空间结构源于其一级结构,在一定程度上可被预测。RNA 适体结构预测主要是紧缩筛选的范围,预测程序发展方向为对预测获得的 RNA 适体空间结构的约束方式扩展、输出文件格式的普适通用性;DNA 虽然在结构稳定性、合成特性等方面相对于 RNA 有优势,便于商业化应用,但短链 DNA 的局部空间结构预测方法有一定局限,影响筛选效率,因而 DNA 适体虚拟筛选的需求也较为紧迫;用于适体功能可视化的分子对接模拟预测方面,涌现了一些以核酸-蛋白相互作用特有性质为基础的针对性程序方法。

应当指出,相对于适体应用的广阔前景,分子模拟研究在优化其筛选和分析流程方面还没有形成具有一定指导性的技术流程,对于已经获得的适体结构特性的应用也存在不足,各种预测方法之间的比较也较少,针对各预测步骤的筛选验证实验仪器与方法也还需要优化和比较,这需要交叉学科研究的深入。综上,以生物信息学研究成果积累为素材,采用分子模拟技术逐步实现核酸

适体、作用目标蛋白质的空间结构模型预测和构建,并采用高通量的虚拟筛选技术,对于体外筛选核酸适体技术具有较好的改进和推动作用,也有利于适体应用的进一步市场化,具有良好的应用前景!

参考文献

- [1] Bock L C, Griffin L C, Latham J A, *et al.* Selection of single-stranded DNA molecules that bind and inhibit human thrombin[J]. *Nature*, 1992, 355(6360): 564-566.
- [2] Draper D E. Themes in RNA-protein recognition[J]. *Journal of Molecular Biology*, 1999, 293(2): 255-270.
- [3] Brooks B R, Ma Jr, Nilsson L, *et al.* CHARMM: The biomolecular simulation program[J]. *Journal of Computational Chemistry*, 2009, 30(10): 1545-1614.
- [4] Szczesniak M W, Makalowska I. lncRNA-RNA interactions across the human transcriptome[J]. *PLoS One*, 2016, 11: e0150353.
- [5] Fiannaca A, Rosa M L, Paglia L L, *et al.* nRC: Non-coding RNA classifier based on structural features[J]. *Biodata Mining*, 2017, 10(8): e27.
- [6] Collie G W, Parkinson G N. The application of DNA and RNA G-quadruplexes to therapeutic medicines[J]. *Chemical Society Reviews*, 2012, 40(12): 5867-5892.
- [7] Nawrocki E P, Burge S W, Bateman A, *et al.* Rfam 12.0: Updates to the RNA families database[J]. *Nucleic Acids Research*, 2015, 43(D1): D130-D137.
- [8] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction[J]. *Nucleic Acids Research*, 2003, 31(13): 3406-3415.
- [9] Reuter J S, Mathews D H. RNA structure: Software for RNA secondary structure prediction and analysis[J]. *BMC Bioinformatics*, 2010, 11(1): e129.

- [10] Miladi M, Junge A, Costa F, *et al.* RNAscClust: Clustering RNA sequences using structure conservation and graph based motifs [J]. *Bioinformatics*, 2017, 33 (14): 2089–2096.
- [11] Shi J, Li X, Dong M, *et al.* JNSViewer-A java script-based nucleotide sequence viewer for DNA/RNA secondary structures [J]. *Plos One*, 2017, 12 (6): e0179040.
- [12] Biesiada M, Purzycka K J, Szachniuk M, *et al.* Automated RNA 3D structure prediction with RNA composer [J]. *Methods in Molecular Biology*, 2016, 1490 (6): 199–215.
- [13] Das R, Baker D. Automated de novo prediction of native-like RNA tertiary structures [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104: 14664–14669.
- [14] Dijk M V, Bonvin A M J J. 3D-DART: A DNA structure modelling server [J]. *Nucleic Acids Research*, 2009, 37 (7): W235–W239.
- [15] Snodin B E, Romano F, Rovigatti L, *et al.* Direct simulation of the self-assembly of a small DNA origami [J]. *Acs Nano*, 2016, 10 (2): 1724–1737.
- [16] Jeddi I, Saiz L. Three-dimensional modeling of single stranded DNA hairpins for aptamer-based biosensors [J]. *Scientific Reports*, 2017, 7 (4): e1178.
- [17] Alexandra S P, Ruth N, Wolfson H J. RsiteDB: A database of protein binding pockets that interact with RNA nucleotide bases [J]. *Nucleic Acids Research*, 2009, 37 (1): D369–D373.
- [18] Tuysinjalgal N, Lee W, Park B, *et al.* PRIdictor: Protein-RNA interaction predictor [J]. *Biosystems*, 2016, 139 (1): 17–22.
- [19] Pierce B G, Wiehe K, Hwang H, *et al.* ZDOCK Server: Interactive docking prediction of protein-protein complexes and symmetric multimers [J]. *Bioinformatics*, 2014, 30 (12): 1771–1773.
- [20] Gray J J, Moughon S, Wang C, *et al.* Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations [J]. *Journal of Molecular Biology*, 2003, 331 (1): 281–299.
- [21] Tuszyńska I, Magnus M, Jonak K, *et al.* NPdock: A web server for protein-nucleic acid docking [J]. *Nucleic Acids Research*, 2015, 43 (W1): W425–W430.
- [22] Shcherbinin D S, Gnedenko O V, Khmeleva S A, *et al.* Computer-aided design of aptamers for cytochrome p450 [J]. *Journal of Structural Biology*, 2015, 191 (2): 112–119.
- [23] Heiat M, Najafi A, Ranjbar R, *et al.* Computational approach to analyze isolated ssDNA aptamers against angiotensin II [J]. *Journal of Biotechnology*, 2016, 230: 34–39.
- [24] Chushak Y, Stone M O. In silico selection of RNA aptamers [J]. *Nucleic Acids Research*, 2009, 37 (12): e87.
- [25] Savory N, Lednor D, Tsukakoshi K, *et al.* In silico maturation of binding-specificity of DNA aptamers against proteus mirabilis [J]. *Biotechnology and Bioengineering*, 2013, 110 (10): 2573–2580.
- [26] Savory N, Nzakizwanayo J, Abe K, *et al.* Selection of DNA aptamers against uropathogenic Escherichia coli NSM59 by quantitative PCR controlled Cell-SELEX [J]. *Journal of Microbiological Methods*, 2014, 104 (9): 94–100.
- [27] Yokoyama T, Tsukakoshi K, Yoshida W, *et al.* Development of HGF-binding aptamers with the combination of G4 promoter-derived aptamer selection and in silico maturation [J]. *Biotechnology and Bioengineering*, 2017, 114 (10): 2196–2203.
- [28] Fukaya T, Abe K, Savory N, *et al.* Improvement of the VEGF binding ability of DNA aptamers through in silico maturation and multimerization strategy [J]. *Journal of Biotechnology*, 2015, 212 (6863): 99–105.
- [29] Zhou Q, Xia X, Luo Z. Searching the sequence space for potent aptamers using SELEX in silico [J]. *Journal of Chemical Theory and Computation*, 2015, 11 (12): 5939–5946.
- [30] Ahirwar R, Nahar S, Aggarwal S, *et al.* In silico selection of an aptamer to estrogen receptor alpha using computational docking employing [J]. *Scientific reports*, 2016, 6 (2): e21285. ■

空气产品公司全球首个煤气化项目在山西投产

11月2日空气产品公司宣布,其位于山西省长治市的潞安煤气化项目已经全面投产,向山西潞安煤基清洁能源有限责任公司的合成气制油项目供应合成气和其他工业气体,这是空气产品公司全球首个煤气化项目。

这一世界规模的气化项目包括4套大型空分装置、4座气化炉和2套合成气净化系统。该项目的气化炉采用壳牌成熟的气化技术,是全球最大的壳牌(粉煤)气化炉,在投煤量、产气量、碳转化率和运营效率上均表现优异。

空气产品公司于2013年和山西潞安矿业(集团)有限公司签订一项协议,为其子公司潞安清洁能源在长治建造并运行4套日产能为10000多吨氧气、6000多吨氮气以及700多吨仪表空气的大型空分装置。2017年9月,空气产品公司和潞安清洁能源组建了一家合资公司并持有60%的

股份,双方共同拥有并运营长治工厂的这些空分装置以及气化和合成气净化设施。

空气产品公司在过去两年一直积极实施其气化战略。2018年8月,该公司获得了一份长期现场供气合同,为久泰新材料有限公司在内蒙古呼和浩特投资数十亿美元的乙二醇项目供应合成气,这是空气产品公司首个全资拥有的气化工厂;当月,空气产品公司还和沙特阿美及ACWA电力公司签署了一份投资意向书,将在沙特阿拉伯的吉赞经济城组建总投资额超过80亿美元的气化/电力合资公司;2017年11月,公司宣布和兖矿集团组成合资企业,将在陕西省榆林兴建总投资额为35亿美元的煤制合成气生产设施;2018年早些时候,该公司还收购了壳牌的气化技术和专利。(中化新网)