

综述与专论

蛋白质全新设计的现状和展望*

曹傲能 来鲁华 唐有祺

(北京大学物理化学研究所, 北京 100871)

摘要 对蛋白质全新设计的方法、设计原则、迄今为止取得的成就和存在的问题及目前面临的困难进行了综述。

关键词 蛋白质, 全新设计, 预测

学科分类号 Q51

1 简介

利用定点突变的方法对天然蛋白质进行改造已有很多成功的实例。但点突变方法只对天然蛋白质中少数的氨基酸残基进行替换, 蛋白质的高级结构基本保持不变, 因而对蛋白质功能的改造极为有限。我们知道, 蛋白质的空间结构受其氨基酸序列控制, 而功能又与结构密切相关。基因工程的发展为蛋白质序列的表达提供了有力的手段, 如果能够找到构造蛋白质结构的方法, 我们就能得到具有任意结构和功能的全新蛋白质, 这就是蛋白质全新设计 (protein de novo design) 问题。

达, 经结构检测确定是否与目标相符, 并根据检测结果指导进一步的设计。蛋白质设计一般都要经过反复多次设计→合成→检测→再设计的过程。

2 设计目标的选择

蛋白质全新设计可分为功能设计和结构设计两个方向^[1,2]。目前的重点和难点是结构设计。如同所有的探索工作, 结构设计也是从最简单的二级结构开始, 以摸索蛋白质结构稳定的规律。在超二级结构和三级结构设计中, 一般选择天然蛋白质结构中一些比较稳定的模块作为设计目标, 如四螺旋束和锌指结构等。在蛋白质功能设计方向主要进行天然蛋白质功能的模拟, 如金属结合蛋白和离子通道等。最近, Rose等^[3]给蛋白质全新设计提出一个新目标——Paracelsus Challenge, 他们悬赏1 000美元给第一位通过改变少于50%的序列, 将蛋白质从一种折叠类型成功地改变到另一种类型的设计者。

3 设计方法

最早的设计方法是序列最简化方法 (minimalist approach), 其特点是尽量使设计序列

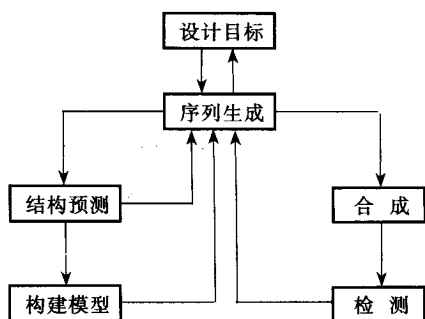


图1 蛋白质全新设计过程

蛋白质全新设计的一般过程可用图1表示。确定设计目标后, 先根据一定的规则生成初始序列, 经过结构预测和构建模型对序列进行初步的修改, 然后进行多肽合成或基因表

*攀登项目“生命过程中重要化学问题”及国家杰出青年科学基金 (29525306) 资助。

收稿日期: 1997-01-29, 修回日期: 1997-04-17

的复杂性最小, 一般仅用很少几种氨基酸, 设计的序列往往具有一定的对称性或周期性^[4]. 这种方法使设计的复杂性减少, 并能检验一些蛋白质的折叠规律如 HP 模型. 现在很多设计仍然采用这一方法.

1988 年, Mutter^[5] 首先提出模板组装合成法 (template assembled synthetic protein approach, TASP). 其思想是将各种二级结构片段通过共价键连接到一个刚性的模板分子上, 形成一定的三级结构. 模板组装合成法绕过了蛋白质三级结构设计的难关, 通过改变二级结构中的氨基酸残基来研究蛋白质中远程作用力, 是研究蛋白质折叠规律和进行蛋白质全新设计规律探索的有效手段.

为了提高设计的速度和效率, 现在已经发展了很多种自动设计方法, 如 Jones^[6] 运用蛋白质反向折叠方法结合遗传算法发展的自动设计方法和来鲁华等^[7] 运用三维剖面 (3D-profile) 结合遗传算法发展的自动设计方法.

4 结构检测

设计的蛋白质序列只有合成并经结构检测后才能判断设计是否与预想结构符合. 一般从三方面来检测: 设计的蛋白质是否为多聚体, 二级结构含量是否与目标符合, 是否具有预定的三级结构.

测定分子体积大小的方法 (如体积排阻色谱法) 可以判断分子以几聚体形式存在, 同时可以初步判断蛋白质结构是无规卷曲还是有一定的三级结构. 检测蛋白质浓度对圆二色谱 (CD) 和 NMR 谱的影响也可以判断蛋白质是否以单分子形式存在.

CD 是检测设计蛋白质二级结构最常用的方法, 根据远紫外 CD 谱可以计算蛋白质中各二级结构的大致含量. 三级结构测定目前主要依靠 NMR 技术和荧光分析.

5 蛋白质结构的全新设计

根据蛋白质折叠的研究和全新设计的探索, 人们发现一些蛋白质全新设计的原则和经

验, 如由于半胱氨酸形成二硫键的配对无法预测, 一般都尽量少用甚至不用, 特别是在自动设计方法中. 但当序列能够折叠成预定的二级结构后, 常常可以引入二硫键来稳定蛋白质的三级结构. 色氨酸的吲哚环具有生色性, 且与其所处环境有关. 天然态的蛋白质中色氨酸常常因埋藏在蛋白质内部, 其荧光频率相对失活态发生蓝移. 因而在蛋白质全新设计中常引入色氨酸作为荧光探针以检验设计蛋白质的三级结构. 以下分别概述各级结构设计的原则.

5.1 二级结构的设计

α 螺旋的设计研究较早而且研究得较多. 这是因为 α 螺旋结构简单, 在溶液中比较稳定. 以往的设计经验表明, α 螺旋设计一般遵循以下指导原则^[1, 2, 6, 8, 9]:

(1) 选择形成 α 螺旋倾向性比较大的氨基酸残基, 如 Leu、Glu、Ala 和 Met 等.

(2) α 螺旋每圈平均有 3.6 个氨基酸残基, 因此在设计两亲性的 α 螺旋时, 为使结构中形成一个亲水面和一个疏水面, 疏水性氨基酸残基应按 3 或 4 的间隔排列.

(3) α 螺旋中所有的氢键指向同一方向, 沿螺旋轴积累总的效果形成一个由 N 端指向 C 端的偶极矩. 因而设计 α 螺旋时, 常使带正电荷的氨基酸残基靠近 C 端, 带负电荷的氨基酸残基靠近 N 端.

(4) 由于 α 螺旋的两端各有 3 个残基的氢键能力未能得到全部满足, 为稳定 α 螺旋, 常常需要在其两端各加一个 N 帽和 C 帽. 形成 N 帽的氨基酸残基有 Gly、Asn、Ser 和 Met 等, 形成 C 帽的有 Gly、Ser、Arg 和 Gln 等.

β 折叠片结构中氢键在不同的 β 折叠股间形成, 也就是说, β 折叠片的形成与序列的“上下文”的关系更密切, 另外单个氨基酸残基平均形成的氢键数较 α 螺旋中少, 因而结构稳定性较差, 所以 β 折叠片的设计比 α 螺旋困难. 单个 β 折叠股结构不能稳定存在, 但可以将其固定在一个模板上来研究其形成规律. 根据模型肽和全 β 结构与 α/β 结构蛋白质设计的经验, β 折叠片的设计原则主要是选择形成 β

折叠片倾向性较大的氨基酸残基 (如 Val, Ile, Tyr) 和使亲水性残基和疏水性残基相间排列^[9].

转角是连接蛋白质中不同二级结构的常见单元, 对维持各种二级结构的相对位置和稳定蛋白质的三级结构起着重要的作用. 转角可分很多类, 不同类的转角对每个位置的氨基酸残基的二面角有一定的要求^[2], 而这些二面角决定了连接的两个二级结构的相对空间关系, 所以转角设计的关键是选择合适的转角类型.

某些氨基酸残基对蛋白质的二级结构有终止作用. 如 Pro 和 Gly 是 α 螺旋的中断者, Glu 是 β 折叠的中断者^[2]. 因而设计时可利用这些氨基酸残基来终止和分隔不同的二级结构.

5.2 超二级结构和三级结构的设计

超二级结构和三级结构的设计一般首先根据氨基酸形成二级结构的倾向性选择各二级结构片段的氨基酸残基, 此外最重要的一点就是考虑疏水中心的形成^[9].

α 螺旋的 coiled-coil 结构是很多跨膜蛋白中一个重要的结构特征, 也是研究得比较多的超二级结构. 其序列有一个明显的特征, 即存在一个周期性重复的七肽, 称为 heptad repeat, 序列可以用 abcdefg 来表示, 其中 a 和 d 位置处于螺旋间的界面, 因而设计时应选择疏水性氨基酸残基, 而其他位置为暴露的表面, 应选择亲水的氨基酸残基^[8].

β 发卡是另一种常见的超二级结构. β 发卡的形成被认为在 β 折叠片形成中起着形核的作用. 因此, 如能合成 β 发卡的模型肽, 就可以研究 β 折叠片中各种结构稳定因素. 一般来说, 单独的 β 发卡在溶液中不能稳定存在, 常常因多肽聚集而形成分子间 β 折叠片. 但最近发现, G 蛋白 B1 结构域中的一个片段在水溶液中能够折叠成单体的天然 β 发卡结构. 在这一发现的鼓励下, Ramirez-Alvarado 等^[10] 设计了由 8 个氨基酸残基组成的具有 β 发卡结构的模型肽. 无论是多肽, 还是蛋白质结构中一个结构单元, β 发卡设计的一个关键部分是其中 β 转角的设计, 特别是选择合适的转角类型.

三级结构的设计同样也是模仿天然蛋白质中较稳定的结构模块. 现在已经有多个比较成功的设计实例, 如 Richardson 夫妇^[11] 设计的全 β 蛋白质 betabellin, Goraj 等^[11] 设计的 octarellin 即 $(\alpha/\beta)_8$. 由于三级结构的设计极为困难, 为使结构更为稳定, 很多设计引入二硫键或金属离子. 最近 Struthers 等^[12] 根据锌指结构成功地设计了一个不含二硫键也无需任何金属离子而形成 $\beta\beta\alpha$ 三级结构的 23 肽, 其结构通过 NMR 检测得到证实, 这是迄今为止最为成功的三级结构设计实例. 此外 Fedorov 等^[13] 还进行自然界不存在的全新结构设计的尝试.

目前大多数设计近似熔球态或介入熔球态和天然态之间. 一般认为, 如果由 CD 谱计算的二级结构含量与目标基本相符, 则设计就算成功.

6 天然蛋白质功能的模拟

很多纤维蛋白的序列由几乎完全相同的片段重复而形成卷曲的螺旋或 coiled-coil 螺旋结构, 如胶原蛋白序列为 $(\text{Gly-Pro-X})_n$. Bella 等^[14] 模拟胶原蛋白设计了一个 30 肽, 其晶体结构已用 X 射线衍射方法解出, 三个肽链形成 coiled-coil 结构.

金属结合蛋白是在生命过程中起着广泛作用的一类蛋白质, 如酶的亲核催化, 电子传递及蛋白质结构的稳定等. 由于金属离子是最简单的配体, 设计时只需考虑 2~4 个金属结合位点的空间关系, 而且可以用很多光谱的方法来检测, 所以很多人进行了金属结合蛋白质的全新设计. 如 DeGrado 等^[4] 设计的 α_4 , Pessi 等^[15] 设计的 Minibody.

α 螺旋 coiled-coil 结构可以用来模拟离子通道蛋白. DeGrado 等^[8] 设计了一个“翻转”的 Rop 蛋白结构, 即将 Rop 蛋白中疏水残基变成亲水残基, 将其亲水残基换成疏水残基, 形成疏水的外表面和亲水的内核. 为使设计的结构具有离子通道功能, 内核不是设计成一般的紧密堆积形式, 而是为水分子和离子留下足够的空间. 为此, 在 a 和 d 位置他们选择了侧

链最小的亲水性残基丝氨酸, 合成了 21 个氨基酸残基组成的 α 螺旋 $H_2N(LSLLLSL)_3CONH_2$, 并由这样三股螺旋组成一个 H^+ 通道模型. 他们还设计了可以容纳更大离子的四螺旋离子通道模型.

其他例子还有 Michael 等^[16]用最简化序列法设计的一个 26 肽来模拟和研究锌指结构规律, Sasaki 和 Kaiser^[17]用卟啉环作模板组装四螺旋束来模拟血红素蛋白等很多实例. 总的来看, 蛋白质功能的全新设计距实用还很远, 当前主要进行天然蛋白质功能的模拟.

7 蛋白质全新设计的现状和前景

蛋白质全新设计不仅使我们有可能得到自然界不存在的具有全新结构和功能的蛋白质, 并且已经成为检验蛋白质折叠理论和研究蛋白质折叠规律的重要手段. 由于我们对蛋白质全新设计的理论基础即蛋白质折叠规律的认识还不够, 蛋白质全新设计还处在探索阶段. 在设计思想上, 目前往往偏重考虑某一蛋白质结构稳定因素, 而不是平衡考虑各种因素, 如在超二级结构和三级结构的设计中, 常常是力求使各二级结构片段都具有最大的稳定性, 这与天然蛋白质中三级结构的形成是二级结构形成和稳定的重要因素刚好相反. 从设计的结果看, 目前还只能设计较小的蛋白质, 其水溶性也较差, 而且大多不具备确定的三级结构.

对是否及何时能够从蛋白质的氨基酸序列准确地预测其三级结构存在不同的看法, 最乐观的预计^[3]认为这个问题将在本世纪末解决. 果真如此, 到那时我们就可以设计具有期望结构的蛋白质了. 但即使无法准确地从蛋白质序列预测蛋白质结构, 对蛋白质折叠规律的不断了解及蛋白质设计经验的不断积累也将使蛋白质全新设计的成功率不断提高. 另外, 随着新实验技术的发展, 蛋白质全新设计的速度和效率必将得到极大的提高. 如组合化学方法应用到蛋白质全新设计中必然能够大大地缩短设计的周期, 并将彻底改变蛋白质全新设计的面貌.

参 考 文 献

- Richardson J S, Richardson D C. The de novo design of protein structures. *TIBS*, 1989, **14** (July): 304~ 309
- 来鲁华. 蛋白质的结构预测与分子设计. 北京: 北京大学出版社, 1993. 100~ 105
- Rose G D, Creamer T P. Protein folding: predicting predicting. *Proteins: Struct Funct Genet*, 1994, **19** (1): 1~ 3
- DeGrado W F, Wasserman Z R, Lear J D. Protein design, a minimalist approach. *Science*, 1989, **243** (4891/4892): 622~ 628
- Mutter M. Nature's rules and chemist's tools: a way for creating novel proteins. *Trends Biochem Sci*, 1988, **13** (7): 260~ 265
- Jones D T. De novo protein design using pairwise potentials and a genetic algorithm. *Protein Sci*, 1994, **3** (4): 567~ 574
- Lai Luhua, Guo Feng, Xu Xiaojie *et al.* Automatic protein de novo design by genetic algorithm and 3D profile. *J Biomol Struct Dyn*, 1995, **12** (6): a128
- Betz S F, Raleigh D P, DeGrado W F. De novo protein design: from molten globules to native-like states. *Curr Opin Struct Biol*, 1993, **3** (4): 601~ 610
- Cordes M H J, Davidson A R, Saner R T. Sequence space, folding and protein design. *Curr Opin Struct Biol*, 1996, **6** (1): 3~ 10
- Ramírez-Alvarado M, Blanco F J, Serrano L. De novo design and structural analysis of a model β -hairpin peptide system. *Nature Struct Biol*, 1996, **3** (7): 604~ 612
- Goraj K, Renard A, Martial J A. Synthesis, purification and initial structural characterization of octarellin, a de novo polypeptide modelled on the α/β -barrel proteins. *Protein Eng*, 1990, **3** (4): 259~ 266
- Struthers M D, Cheng R P, Imperiali P. Design of a monomeric 23-residue polypeptide with defined tertiary structure. *Science*, 1996, **271** (5247): 342~ 345
- Fedorov A N, Dolgikh D A, Chemeris V V *et al.* De novo protein design, synthesis and study of Albetetin, a polypeptide with a predetermined three-dimensional structure. *J Mol Biol*, 1992, **225** (4): 927~ 931
- Bella J, Eaton M, Brodsky B *et al.* Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science*, 1994, **266** (5182): 75~ 81
- Pessi A, Bianchi E, Creamer A *et al.* A designed metal-binding protein with a novel fold. *Nature*, 1993, **362** (6418): 367~ 369
- Michael S F, Kilfoil V J, Schmidt M H *et al.* Metal binding and folding properties of a minimalist Cys₂His₂ zinc finger peptide. *Proc Natl Acad Sci USA*, 1992, **89** (11): 4796~ 4800
- Sasaki T, Kaiser E T. Synthesis and structural stability of helichrome as an artificial hemeproteins. *Biopolymers*, 1990, **29** (1): 79~ 88

The Current State and Prospect of de novo Protein Design. CAO Aoren, LAI Luhua, TANG Youqi (*Institute of Physical*

Chemistry, Peking University, Beijing 100871, China).

Abstract The methods, principles, achievements as well as obstacles of de novo protein

design were reviewed.

Key words protein, de novo design, prediction

细胞因子受体超家族与信号转导

贺平 汤钊猷

(上海医科大学中山医院肝癌研究所, 上海 200032)

摘要 细胞因子受体超家族及其介导信号转导的新途径已引起人们的广泛注意. 主要包括受体结构特征, Janus 激酶 (JAKs) 的作用, 信号转导及转录激活因子 (STAT) 和 Ras 蛋白的激活, 以及细胞因子信号转导的异常与临床疾病的关系等. 有助于人们正确认识细胞因子的生物学效应.

关键词 细胞因子, 细胞因子受体, 信号转导, 酪氨酸激酶

学科分类号 Q73

细胞因子 (cytokine) 是一类具有广泛生物学活性的可溶性多肽, 它与受体结合后可通过多种信号转导途径及激活不同的转录因子, 调节细胞的生长、分化和功能活性. 最近研究发现, 绝大多数细胞因子与其受体结合后, 通过一条共同的信号转导途径调控基因的表达, 即激活 Janus 酪氨酸激酶系统 (JAKs), 活化信号蛋白及转录因子^[1].

1 细胞因子受体超家族

细胞因子受体是一类跨膜的细胞表面受体, 它本身缺乏激酶结构域, 但它与另一类激酶 (JAKs) 密切联系. 细胞因子受体超家族虽然由一个共同的祖先进化而来, 但由于长期的演变, 各成员间同源性并不高. 它们的一个显著特征是绝大多数受体都由几个亚基组成 (EPOR、GHR 等例外), 至少含一条配体结合链和一条信号转导链. 前者又称“私有链”, 可特异地与配体结合; 后者又称“公有链”, 它不与配体结合, 而参与多种受体的信号转导, 是共同的信号传导组分.

根据受体结构特征, 一般将细胞因子受体

分成两大类.

第一类受体至少包括 20 多种, 其配体分别是 IL-2 ~ IL-7、IL-9、IL-11、IL-12、IL-15、EPO、PRL、GH、G-CSF、GM-CSF、LIF (白细胞抑制因子)、CNTF (睫状神经营养因子) 等. 这些受体的共同特征是在受体细胞外区特定位置上有四个保守的半胱氨酸残基和 Trp-Ser-Xxx-Trp-Ser (WSXWS) 结构模式 (X 代表任一氨基酸残基), 形成一个与配体结合的结构域. 在受体胞浆区近胞膜处含有 BOX1 和 BOX2 结构域, BOX1 由一簇疏水氨基酸残基和 Pro-Xxx-Pro 序列组成, BOX2 由几个疏水氨基酸、一个带负电的氨基酸和一至二个带正电的氨基酸残基组成. BOX1 和 BOX2 是受体发挥有丝分裂活性 (mitogenic activity) 所必需. 这类受体根据其亚基的组成形式、信号转导链的特征及与之关联的 JAK 激酶的不同, 又可分为不同的亚类. 如 EPOR 亚类、IL-3R 亚类、IL-6R 亚类及 IL-2R 亚类等.

第二类受体数量较少, 其配体分别是