

传统观点认为，蛋白质具有绝对的功能特异性，同时也只具有单一、固定的结构，但事实并非如此。蛋白质具有很强的适应性，可以“进化”出新的功能与结构。

蛋白质发挥功能是非常准确的、充满特异性的，同时也是非常娴熟的、有效率的。这些特性都是由于蛋白质本身缺乏多能性造成的，不过蛋白质也有很强的获取新功能和结构的能力。实际上，能够证明蛋白质进化适应性的例子随处可见，例如自然界中存在的、数量众多的起源于共同祖先的各种蛋白质，又比如最近几十年间由于滥用药物导致的耐药性机制，即各种耐药性酶（蛋白质）的产生等等进化事件。

那么是什么让蛋白质进化的呢？进化的过程就是将已经存在的多样性不断聚集的过程。如果真的像过去那种认为蛋白质只有一种功能、一种结构的传统观点所界定的那样，那么它就不可能适应环境，对新出现的选择压力做出适当的反应了。不过现在，一种所谓“天然态（native state）”的新观点开始流行。该观点由R. L. Baldwin 和K. A. Dill最先提出。他们认为蛋白质是由各种可变底物（alternative substructure）或构象异构体（conformer）组合起来的整体。起初，R. L. Baldwin等人只是用这种观点来解释蛋白质折叠现象，后来又将该观点推广到蛋白质是各种天然状态物质的大集合。这种新观点更符合蛋白质进化的现象。现在，它已经扩展成为一个有关蛋白质动态性和进化论的新观点。

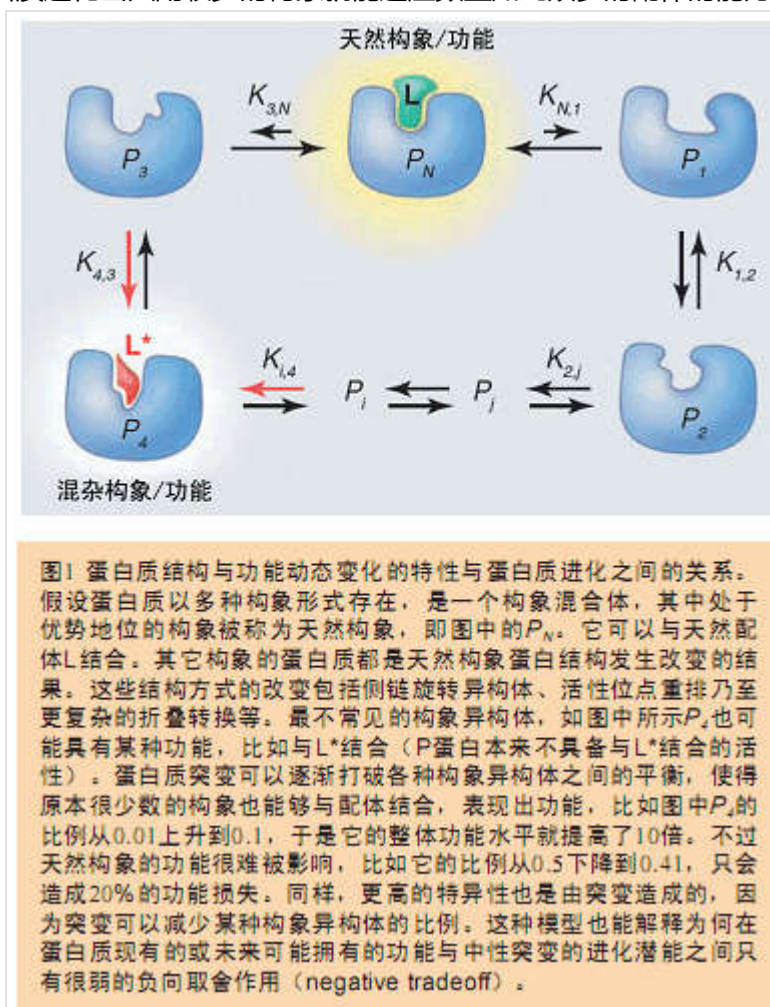
构象可变性（conformational variability）或者称为动态性（dynamism）是任何一条多肽链固有的特性。我们在蛋白质中发现的构象多样性可以见于各种情况，从侧链的变动和活性位点的运动到二级结构的交换再到整个蛋白质折叠的重排等等，不一而足。各种构象异构体可以调节蛋白质的折叠与功能。这种蛋白质结构与功能的多样性就是蛋白质进化的基础，也使得蛋白质能够快速适应新出现的情况，利用已有的折叠模式发展出新功能，或者进化出新的折叠模式。

蛋白质出现功能混杂，看来似乎是进化出新功能的首要条件。蛋白质突变可以打破它在结构与功能之间形成的平衡状态，形成一个可供环境因素选择的原材料，让蛋白质能够发展出更多的结构与功能。在本文中我们将讨论蛋白质积累大量突变以帮助它们适应环境的机制。我们还将借助与RNA与蛋白质折叠中间体对比的方法向读者介绍最古老的蛋白质可能的进化历程，我们相信，这些知识有可能会帮助我们了解原始蛋白质的形成过程。由于我们只有很少的现代蛋白质进化的例子（比如耐药性等）可借鉴，而远古的例子又太难搜寻，因此只能给出可能的进化途径和促使蛋白质进化出新功能、新结构和新折叠方式的“推动力”。不过所有这些都只处于理论层面，还需要实验加以验证，我们会在后面对此进行专门讨论。

1. 局部活性位点的灵活性促使蛋白质功能混杂现象和进化的发生

蛋白质中的活性位点是具有高度灵活性的，这种在不同时间以不同强度表现出来的灵活性与蛋白质的催化活性和调节功能有关。传统的酶化学认为，蛋白质发挥催化活性时，活性位点是那些刚性的、结构固定的位点，但是其它步骤，例如产物释放步骤等则决定了酶的更新速率，构象重排可促进这一现象的出现。这种酶在催化过程中表现出来的灵活性同样也赋予了酶多种功能，以及进化出更多功能的可能性。很多蛋白都具有多种细胞功能，各种酶也进化出了比它们最初功能多得多的功能。蛋白质的这种混杂性（promiscuity）或者叫多重特异性（multispecificity）应该归功于自然界中存在各种不同的底物或配体。正是因为这些底物

或配体的存在使得蛋白质进化出只用较少的构象就能适应数量如此众多的配体的能力（图1）。



我们对于多重特异性并不陌生，抗体其实就是一个非常好的例子。在与配体（抗原）结合以前，它们实际上是有好几种构象的，这些构象异构体之间处于一种动态平衡状态当中。抗体能与两种毫无关联的配体结合，每一种配体都会打破抗体之间的平衡而偏向另一种构象。最近，一项对泛素蛋白开展的核磁共振成像研究发现，泛素蛋白居然有46种不同的构象。蛋白质分子固有的这种灵活性使得泛素蛋白能够结合多种底物，而且它在与每一种配体结合时都有对应的单独构象。蛋白质分子虽然具有如此强大的构象可塑性，但仍然能特异地与底物分子结合，这种奇特的现象也可见于T细胞受体和“模糊复合体”（这些分子存在多种构象，甚至大部分时间处于构象混杂的状态）。类似这种结构上灵活可变的酶数不胜数，例如细胞色素P450就具有多种活性构象，能够与多种不同的底物结合（图2A）。蛋白结构具有灵活性还能让铁调节蛋白1（iron regulatory protein 1）的结构域复位，即一种构象异构体与mRNA结合，抑制其翻译或降解该mRNA，另一种构象异构体则可以与铁硫簇结合，变成顺乌头酸酶。

在蛋白质构象多样性与功能多样性之间似乎存在某种关联。以细胞色素P450为例，相对缺乏构象灵活性的CYP2A6蛋白的底物特异性就非常低，而CYP3A4蛋白是所有细胞色素蛋白中构象最多变的，因此也就表现出了非常高的底物灵活性。再看看抗体蛋白分子，如果抗体分子与配体的亲和力增高，相应地就会降低结合位点的灵活性。我们通过定向进化试验（directed evolution experiment）在蛋白质灵活活性位点

(flexible active-site loop) 中引入突变, 结果获得了新的酶学特异性蛋白。引入的这些突变位点通常都是不稳定的, 这说明如果增加了蛋白质立体构型的熵和活性位点的灵活性, 也会引入新的“特异性”, 即带来新的功能。同样, 蛋白质分子通过自然选择获得的突变通常也能诱导出更广泛的特异性, 使其能够“接纳”更多其它未被选择的底物, 这其中的机制可能就是因为蛋白质活性位点自由度的改变(图1)而造成的。虽然上述所有例子都表明, 蛋白质活性位点的灵活性越高, 就意味着它具有更多样的功能, 同时也意味着它更适宜进化, 但是蛋白质分子活性位点的灵活性与蛋白质分子的可进化性之间并没有确切的联系。同样, 蛋白质与各种配体或底物结合时所扮演的角色, 即在反应过程中扮演的角色是我们研究基因功能的起点, 不过目前这方面的研究还比较少, 还需要我们进行更多、更深入的工作。

2. 蛋白质整体构象多样性与新构象的进化

除了前文所述的可变侧链旋转异构体和loop的构象等这些蛋白质局部结构灵活性(多样性)之外, 还存在蛋白质整体构象重排和折叠转换现象。比如淋巴细胞趋化因子(lymphotactin)有两种不同的折叠形式(图2B); Mad2蛋白是一个同源二聚体(homodimer), 它也有两种不同的 β 折叠形式。

有科研人员通过对序列折叠空间网络进行图谱研究的方法来研究折叠转换问题。最初用该方法准确预测出了RNA的二级结构, 随后又预测出了蛋白质的晶体结构。每一个折叠都参与形成一个网络。组成该网络的这些序列都具有相同的结构(这也就意味着它们具有相同的功能)。它们彼此之间依靠单个位点突变来进行“联系”。因为每一个网络都可以通过特定的转换位点, 即序列, 与其它网络相连, 因此, 只需要对该位点进行改变(即突变), 就可以顺利地将一种结构转变成另一种结构。我们已经发现了两种核酶。这两种酶的构象和功能可以通过点突变来相互转换, 而且还发现了这两种构象互换过程中存在的中间物, 从而证明了上述网络观点的正确性。体外进化试验也证明, 某些位点的突变能够改变核酶的功能, 而这正是因为突变后核酶形成了新的折叠的缘故。不过我们很少发现这种蛋白质分子折叠与功能的转变, 因为蛋白质分子的折叠情况要复杂得多。有一个由28个氨基酸组成的富含半胱氨酸的蛋白质表现出了和核酶一样的特性, 即该蛋白质分子中某个位点突变后会令它改变构象, 形成另一种自然界中原本存在的构象异构体, 而再引入突变后又会抑制这种构象转变(图2C)。

自然界中另一个蛋白质构象转换的例子是朊病毒(prion)。它既可以以溶解状态存在, 也可以聚集形成淀粉样蛋白(amyloid), 还可以形成无数的各种低聚状态的中间分子。固有无序蛋白(IDP)则是另一种完全不同的蛋白质。它们并没有固定的结构, 不同于上述那些能够改变构象的“大型”蛋白质分子。通常来说, 只有当配体或底物分子与蛋白质分子结合之后我们才能观察到紧密有序的蛋白质折叠结构, 但这也只限于与底物结合的区域, 其它蛋白质区域仍然处于无序状态。我们认为IDP是一大类完全不同的蛋白, 它们具有特殊的序列组成特点, 这些序列特点也决定了它们具有特殊的功能。不过无论如何, 将蛋白质折叠过程和它与底物结合过程联系起来考虑我们发现, 这两个过程都在蛋白质进化过程中起到了非常重要的作用, 而且可能都在主要蛋白质折叠转换的过程中起到了关键性作用。值得注意的是, 蛋白质的有序和无序状态并不是一个全或无的选择。我们经常可以在蛋白质分子中看到, 大部分序列都处于有序状态的同时, 仍然有一小段序列会处于无序状态。无序状态相对较高的蛋白质似乎也是某些蛋白质特有的“标志”, 比如病毒蛋白质就是如此。实际上, 正是部分区域有序这个特性才使得蛋白质分子具有改变序列的能力, 也赋予了蛋白质进化的能力。

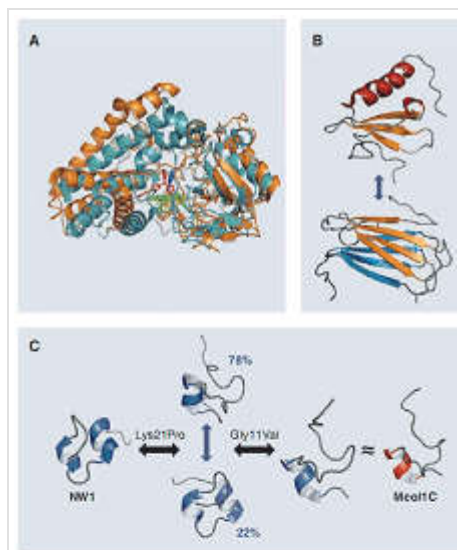


图2 蛋白质构象多样性示意图。A：酶局部构象发生改变使得它能与更广泛的底物相结合。图中P450-CYP2B4蛋白在与较大的红色底物联苯苄唑（bifonazole）结合时呈现的是开放型构象，如图中橙色所示。而与较小的深蓝色底物4-4氯苯基咪唑（4-(4-chlorophenyl)imidazole）结合时呈现的是闭合型构象，如图中浅蓝色所示；B：变形的蛋白质。淋巴细胞趋化因子（Lymfotactin）具有两种构象，即图中上部所示的beta-alpha单体结构和底部所示的全beta二聚体结构；C：折叠转换。在蛋白质中有两个富含半胱氨酸的结构域NW1和Mco11C，它们除了富含半胱氨酸之外在序列上没有任何相似之处，以这两个结构域为基础，通过3个二硫键的转换，可以形成两种构象。蛋白质第21位的赖氨酸突变成脯氨酸（K21P）就能将蛋白质从一种构象转换成中间体，然后第11位的甘氨酸突变成缬氨酸（G11V）就可以将中间体转变成第二种构象。

3. 蛋白质的进化能力与突变带来的影响

所谓进化，实际上就是将突变或序列改变的信息给固定、保留下来，因为这些改变通常都伴随新功能的出现，也使物种更能够适应周围的环境。不过生物也可能会将毫无用处的突变信息保留下来，就好像Kimura在他的中性学说（文后小词典1）理论中描述的那样。与此相关的一个问题就是进化本身就是由彼此矛盾的两个部分组成的。生物物种、基因以及基因所编码的蛋白质每时每刻都有突变的可能，而蛋白质的中性特质或“坚强”特质（即抵抗突变对其结构与功能造成影响的能力）是有限的，因此就会造成物种适应能力的下降。但是，通常来说，每次只会积累一个突变。为了适应环境的需要，蛋白质的结构与功能应该对突变做出适当的反应。中性学说认为突变对物种不会造成影响，然而蛋白可塑性又要求突变会对蛋白的结构与功能造成很大的影响，那么这两种理论之间到底谁对谁错呢，它们能够达成一致吗？下面，我们就来讨论一下这个问题。

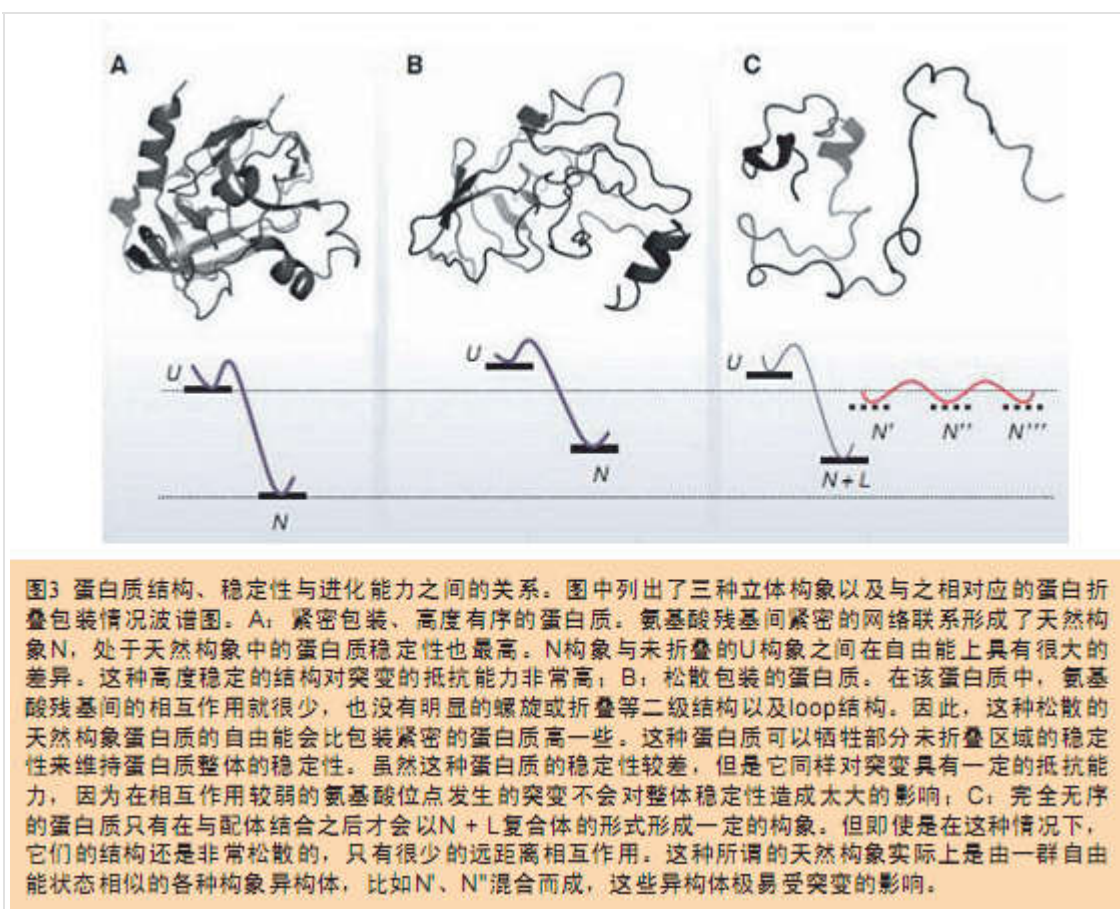
与上述观点相反，我们通过比较蛋白质的折叠情况之后发现，更多的中性折叠（即序列多样性更高的折叠）意味着蛋白质具有更大的功能多样性。为什么会这样呢？一种可能是，基因在复制时从基因组众多的拷贝中积累了大量的突变信息，久而久之，形成了新的功能，但大量缺失突变会造成基因功能丧失的事实似乎并不支持这种理论。另一种解释认为，突变能够影响蛋白质功能，但是只会造成很小的影响。同样，在某个环境中，突变可能只会起到一个中性作用，但是如果环境发生改变，该突变可能就会起到积极作用了（图1）。蛋白质的各种构象和功能也可以看做是表型变异。就好像调控网络一样，比如，适应环境变化的生理性适应机制可能与机体改变对突变的反应，即进化适应有关。对生物体这种在蛋白质的现有功能与进化功能之间不断取舍的过程，即在相关生理功能与进化适

应功能间取舍的机制还有待进一步研究。除了我们在本文中介绍的最简单的模型之外，还存在着许多其它的机制。

另一个与“中性 - 可塑性二元理论（neutrality-plasticity dichotomy）”相关的现象是具有高度进化能力的蛋白质的结构特征。按照传统观点，中性意味着在热力学上高度稳定，超过80%的缺失突变都会造成蛋白质稳定性降低，同时蛋白质的水溶性也会降低，这就意味着会影响到蛋白质发挥功能。因此，高稳定性，尤其是高度的热稳定性需要蛋白质具有高度稳定的构象，即包装折叠正确、结构高度紧凑，并且能够“抵御”突变造成的影响（图3）。以免疫球蛋白为例，正是因为具有高度包装折叠的构象，具有高度稳定性的骨架结构，所以免疫球蛋白才具有如此重要的功能，同时又可以具有高度的序列多样性。免疫球蛋白的这种多样

性是通过改变蛋白表面的loop序列，即抗体互补决定簇区域来实现的。这种高度稳定的骨架结构搭配高度灵活的活性位点结构方式在其它蛋白质中也非常常见，比如TIM barrels酶家族等。这种“稳定+灵活”的结构模式既赋予了蛋白中性的特质（稳定的蛋白骨架结构），又赋予了蛋白可塑性的特质（表面的loop结构）。

不过大部分的蛋白质稳定性都有限，尤其是处于高度突变环境下更是如此。RNA病毒蛋白质的突变速率要比细菌或真核生物蛋白质的突变速率高出100万倍，因此RNA病毒蛋白质的稳定性非常低，这也就意味着这些蛋白质的结构都比较松散，折叠、包装不够，即紧密性不够，非常容易出现局部结构紊乱的现象。不过这种结构特性也具有一些优势，比如RNA病毒蛋白对于突变的耐受性就非常好，因为单独的突变对氨基酸残基间相互作用力的影响很小，所以对蛋白质整体稳定性的影响不是很大。还有其它证据也能支持这种蛋白质二元结构理论。比如蛋白质序列多样性越大的折叠结构就越容易出现结构紊乱。很多固有无序蛋白（IDP）都具有非常高的序列多样性。RNA分子所具备的极强的进化能力也是因为它们缺乏远程三级接触（long-range tertiary contacts）作用。因此，RNA病毒蛋白质所具备的高度的进化能力可能也与它们折叠程度不高（即三级结构上的相互作用不够）、紧凑度不高以及局部或总体结构紊乱有关。虽然有报道称，蛋白质分子中存在中性程度更高的结构紊乱区域，但这些区域是否具有快速进化出新功能或新结构的能力还有待研究。蛋白质结构稳定性低导致的必然结果就是牺牲功能或者让未折叠区域更加不稳定（即采取负性调节的办法），而不会采取提高已折叠区域的稳定性，即正向调节的办法（图3）。



4. 对早期蛋白质折叠结构进化的思辨以及该领域未来的发展方向

正如前文所述，我们可以用蛋白质构象和功能多样性的概念来诠释蛋白质的进化过程。这个过程是一个渐进的、通过蛋白质氨基酸序列突变而发生的过程。整个进化过程非常平缓，是一个需要借助中间功能产物过渡的过程。我们也可以把这些概念应用到原始蛋白前体研究领域。不过原始蛋白的进化问题目前还只是处于猜测和假说阶段。

生物大分子聚合物的序列空间是非常惊人的。我们以 nL 表示序列空间这个概念，其中 n 代表单体类型的数目， L 表示生物大分子的长度。对于一个由100个氨基酸组成的多肽链来说，具有 $20^{100} \sim 10^{130}$ 种排列方式。这就使得测序工作几乎成了一件不可能完成的任务。虽然存在冗余现象（即很多序列都具有相同的结构和功能），但面对如此众多种排列组合的可能性，我们还是显得无能为力。RNA分子相比蛋白质分子来说进化能力更强，这是因为RNA分子的序列空间很小，RNA分子的 n 只有4，而蛋白质分子的 n 等于20。不过，其实只需要9个氨基酸就能获得一个具备功能的蛋白质。长度小于30个氨基酸的短肽可以相互聚集形成同源或异源多聚体，通过这种方式也能形成一定的功能。这种寡聚化作用或者是基于有序 β 折叠的聚集作用可以把暴露在外的疏水性基团侧链折叠到蛋白质内部，这样也能提高蛋白质的稳定性和水溶性。这种低聚化的作用界面构成了第一个结合位点或活性位点，就好像现在蛋白质中常见的位点一样。该过程如此往复下去就能形成更大的单体蛋白质。半数已知的蛋白质折叠都具有的内部对称性以及我们假设的高度对称的寡聚蛋白前体结构，比如 β 螺旋等都符合这种假说。

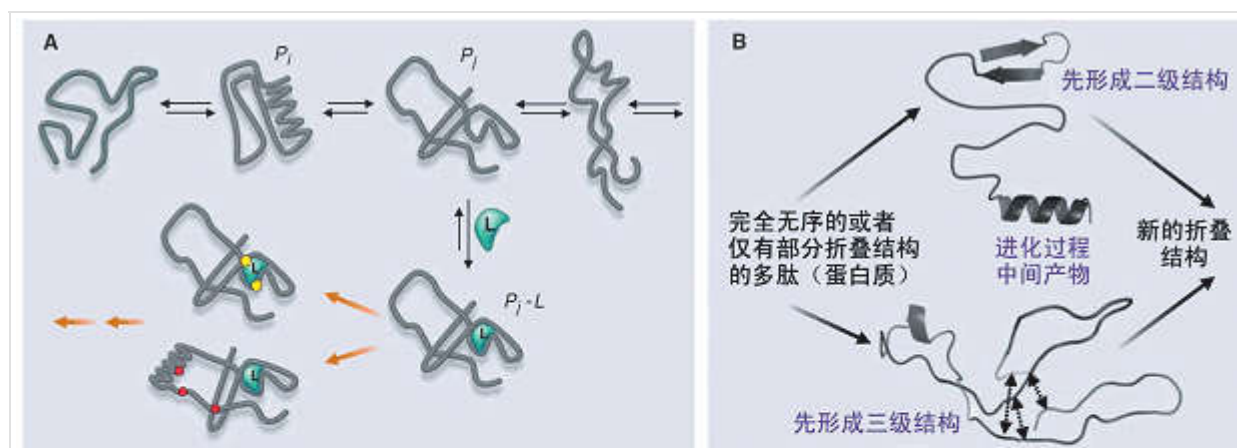


图4 原始蛋白质进化过程假想示意图。A：从 P_i 、 P_j 等所有蛋白质中，蛋白质的构象与功能经过构象筛选过程得到了共进化。蛋白质与配体 L 或底物结合之后就会打破构象平衡状态，形成更多的构象异构体 P_j 。随后的突变（图中橘色箭头所示）通过形成图中红色圆点所示的新折叠结构的同时改变配体结合位点的方式（如图中黄色圆点所示），从而将这种有功能的构象给“固定”了下来；B：进化中间体理论。从最初的完全无序或仅有部分折叠结构的蛋白质进化成高度折叠的蛋白质可以经由两条途径：二级结构途径和三级结构途径。在二级结构途径理论里，蛋白质首先形成二级结构，出现远距离作用，最后形成折叠完好的蛋白质；在三级结构途径理论里，蛋白质首先形成疏水核心，该疏水核心由20~30个氨基酸组成的闭合loop结构组成，然后再形成折叠完好的蛋白质。

在研究蛋白质，尤其是酶蛋白的进化问题时还有一大难题，那就是蛋白质的功能都必须依赖于一定的天然结构，但实际上蛋白质结构本身并不能提供任何的选择优势。有一种情况是某一种功能先被选择、固定下来，然后进行结构共进化过程（图4A）。假设是结构部分有序的蛋白质形成了第一个进化中间产物，那么就可

能出现如图4B中所示的两种情况。与赫克尔 (Haeckel) 关于个体发育的重演说 (ontogeny recapitulates phylogeny) 相似, 中间折叠产物可能也反映了进化中间产物的本质特征。蛋白质可以通过多种方式折叠。某些情况下, 某个二级结构的核心 (该区域可能还处于部分未折叠的状态) 会随即与其它部分发生远距离三级结构作用 (tertiary long-range contact)。同样, 折叠可能是通过远距离相互作用起始于一系列的疏水核心基团 (hydrophobic core), 然后再形成二级结构。对于上述这些推测既有支持意见, 也有反对意见。处于熔球 (文化小词典2) 状态的蛋白质也具有酶学活性。 α 螺旋、 β 折叠这些二级结构似乎是极性氨基酸与非极性氨基酸简单变化而来的。形成 β 折叠的序列都非常小。RNA进化能力的关键似乎是来源于稳定的、可简单交换二级结构的能力, 同时缺乏远距离三级结构相互作用力所致。另一方面, 对于蛋白质来说, 三级结构相互作用力又非常关键, 因为构成三级结构的都是经由疏水作用形成的由20~30个氨基酸肽段组成的loop结构, 这些结构是形成蛋白质折叠的基础。

综上所述, 按照文中所述的各种机制, 蛋白质结构与功能的动态变化特性为蛋白质的进化 (不论是依靠已有结构表现出新功能, 还是进化出全新的结构与功能) 打下了基础。不过还需要对蛋白质进化过程进行更深入的研究。早期蛋白是如何进化的? 蛋白质折叠转换是如何发生的? 蛋白质整体或局部的结构松散性 (紊乱程度) 是否提供了更大的进化能力 (如图4所示)? 反过来, 在稳定的蛋白骨架结构上搭配具有丰富多样性的局部活性位点, 是否能增加蛋白质的功能? 如果出现大量高度灵活的构象多样性生物可能会造成蛋白质的活性降低, 那么蛋白质是如何在进化与保证功能之间做出取舍的呢? 比如病毒酶相比结构高度有序的、折叠紧凑的同源蛋白, 是否具有更强的进化能力, 同时活性更低呢? 那么蛋白质的进化能力是一种与生俱来的能力吗? 这种功能上的多样性与构象上的灵活多变性是天生就有、无需选择的吗? 其它的性状, 比如对突变的耐受能力可能是因为环境因素造成的, 而非遗传因素造成的。不过纵观整个蛋白质进化史, 有一些酶就是要比其它蛋白质进化得更多、更快。次级代谢 (文后小词典3) 通常都与环境因素相适应, 但核心代谢 (core metabolism) 则一般不受环境因素影响。那么是不是参与次级代谢的酶更多变, 更富有进化能力呢? 它们是不是活性不高呢? 可能更接近实际的蛋白质结构而不是晶体模型 (lattice model) 更能解答上述问题。具体的实验, 例如能重建进化过程的体外进化实验也会有一定的帮助。在体外进化实验中可以用完全随机的肽段作为最初的“原材料”, 但是如果在多个生物信息学信息, 比如种系发生信息和蛋白超家族等信息的引导下设计序列效果会更好。如何在体外进化实验中通过简短的肽段构建原始蛋白质, 以及如何在整个试验过程中跟踪研究中间产物, 这是未来研究中最主要的问题。

