

通过断裂内含肽介导的反式剪接合成大的蛋白*

张静 刘环 周晶 刘建华**

(上海交通大学生命科学技术学院 上海 200240)

摘要 大肠杆菌难以表达大的蛋白,毒性蛋白以及膜蛋白,“Npu DnaE 内含肽表达系统”使这些蛋白的表达成为可能。该系统的基本原理是:在特定位点处将目标基因(编码 T7 RNA 聚合酶的基因)断裂成两部分,然后分别与 Npu DnaE 内含肽的 N 端,C 端片段融合,两种融合基因分别表达纯化,在体外将两种融合蛋白等摩尔比混合即可产生有功能的 T7 RNA 聚合酶。理论上,该体系也可用于合成其他大的蛋白,毒性蛋白或膜蛋白。

关键词 大的蛋白 毒性蛋白 膜蛋白 Npu DnaE 内含肽 T7 RNA 聚合酶

中图分类号 Q819

蛋白质剪接是一种翻译后成熟的过程,在此过程中,一种命名为“内含肽”的插入序列能进行自我剪接,以肽键连接两侧外蛋白子形成成熟蛋白质^[1]。有些编码内含肽的基因在特定位点处断裂成 N 端片段(简称 I^N)和 C 端片段(简称 I^C),这两部分单独存在时没有活性,但当断裂内含肽的 I^N 和 I^C 混合时,两者相互识别,重建催化活性中心,介导蛋白质反式剪接^[2-5]。蛋白质反式剪接在生物学上已经得到了广泛应用,包括蛋白半合成^[6],蛋白环化^[7-9],对蛋白部分片段进行同位素标记^[5],蛋白开关^[10-12],以及在转基因植物中激活某些基因^[13]。蛋白反式剪接已成功地应用于腺病毒介导的体内基因治疗中^[14]。他们将 6.3kb 长的 dys cDNA 断裂成两部分,分别与断裂内含肽的 N 端和 C 端片段连接。这两种融合基因分别克隆到 VV1 载体中。同时将这两种载体导入 mdx 鼠中,通过上述两种融合蛋白有效地剪接产生全长的 Becker 型肌萎缩蛋白。这项研究成果给难表达蛋白的合成带来了曙光。

目前 Ssp DnaE 内含肽是用得最广泛的断裂内含肽。而 Iwai 等报道的 *Nostoc punctiforme* PCC73102 (Npu)DnaE 内含肽虽与 Ssp DnaE 内含肽高度同源,但

前者具有更高的剪接活性。Npu DnaE 内含肽融合蛋白有 55%~90% 转化成剪接产物,并且仅有少量的 N 端和 C 端切割产物^[15]。综合考虑,我们决定利用 Npu DnaE 内含肽来合成大的蛋白。

我们将 Npu DnaE 断裂内含肽与大肠杆菌 T7 蛋白表达系统相结合,构建了一套体外 T7 表达系统。该策略将 T7 RNA 聚合酶在特定位点处断裂成 N 端和 C 端两部分,然后分别与 Npu DnaE 内含肽的 N 端片段(DnaE_N)和 C 端片段(DnaE_C)相融合。这两种融合蛋白分别表达,纯化,然后在体外混合。最终通过 Npu DnaE 内含肽介导的反式剪接产生有功能的 T7 RNA 聚合酶。

1 材料与方法

1.1 材料

大肠杆菌 ER2566 购自美国 NEB 公司;大肠杆菌 DH5 α 和 Ni-NTA 树脂购自美国 Novagen 公司;质粒 pLIC1, pLIC2 为本实验室构建;质粒小量抽提试剂盒、PCR 产物纯化试剂盒和核酸凝胶回收试剂盒,购自北京天根生化科技有限公司;细菌基因组抽提试剂盒购自上海生物工程技术有限公司;限制性内切酶、T4 DNA 连接酶、Taq DNA 聚合酶、核糖核酸酶抑制剂购自大连宝生生物技术公司;各种寡核苷酸片段于上海英骏生物

收稿日期:2009-09-09

* 国家“973”计划子课题(2007CB914500)资助项目

** 通讯作者,电子信箱:jianhualiu@sjtu.edu.cn

技术有限公司合成; [α - 32 P]ATP 购自北京亚辉生物科技有限公司。

1.2 方法

1.2.1 PCR 扩增 T7 RNA 聚合酶的 N 端(1~491 位残基)、C 端(492~883 位残基) T7 RNA 聚合酶的 N 端片段的扩增是以 ER2566 基因组为模板, T7N-F/R (T7N-F; 5' ATATACATCTTAAGATGAACACGATTAACA TCGC; T7N-R; 5' TTTCATAACTTAAGCAAGCCATGATG TTCTCGTGG) 为引物。获得的 PCR 片段有 LIC 特定的末端(正向: ATATACATCTTAA; 反向: TTTCATAACTTAAGCA)。

以 ER2566 基因组为模板, T7C-F/R (T7C-F; 5' AA AAATGGGTTTATAGGCTCTAATTGCTGCGCTAAGTCTC CACTG; T7C-R; 5' AGGTTGTGTATAAAAAGGCTTAATA AAGGCCACTCTTGCGA) 为引物, PCR 扩增 T7 RNA 聚合酶的 C 端片段, 其两端含有 LIC 特定的末端(正向: AAAAATGGGTTTATAGGCTCTAATTGC; 反向: AGGT TGTGTATAAAAAGGC)。

1.2.2 pLIC-N, pLIC-C 表达载体的构建 在用 T4 DNA 聚合酶处理之前, 载体 pLIC1 及 pLIC2 需分别经限制性内切酶 *Afl* II、*Stu* I 单酶切成线性片段。线性化的载体片段以及 PCR 片段经电泳检测后, 纯化回收, 然后用 T4 DNA 聚合酶消化。50 μ l 反应体系中含 1 个单位的 T4 DNA 聚合酶, 0.5 mmol/L dGTP 或 dCTP, 1 \times T4 DNA 聚合酶缓冲液。37 $^{\circ}$ C 反应 0.5 h 后 65 $^{\circ}$ C 热处理 10 min, 使 T4 DNA 聚合酶失活。消化后的载体及 PCR 片段用 PCR 产物纯化试剂盒回收, 最终得到的载体及 PCR 片段的浓度分别约为 3 ng/ μ l 和 6 ng/ μ l。

3 μ l T4 DNA 聚合酶处理的 pLIC1/pLIC2 (9 ng) 和 4 μ l T4 DNA 聚合酶处理的 PCR 片段 (24 ng) 混合, 冰浴 30 min 后, 混合物化学转化 DH5 α , 铺 LBA 平板, 37 $^{\circ}$ C 过夜培养; 挑取克隆并提取质粒, 测序鉴定, 阳性质粒分别命名为 pLIC-N, pLIC-C。

1.2.3 融合蛋白的制备 表达质粒 pLIC-N 和 pLIC-C 分别转化大肠杆菌 ER2566, 转化菌株涂布于含 100 μ g/ml Ampicilin 的 LB 固体培养基, 37 $^{\circ}$ C 过夜培养。转化单菌落接种于 25 ml 含 100 μ g/ml Ampicilin 的 LB 液体培养基, 37 $^{\circ}$ C, 220 r/min, 培养过夜。过夜培养物以 1:50 的比例接种于新鲜的 LB 液体培养基 (含 100 μ g/ml Ampicilin), 培养至 OD₆₀₀ = 0.6 (37 $^{\circ}$ C, 220 r/min), 加入 IPTG 至终浓度为 0.5 mmol/L, 37 $^{\circ}$ C 诱导重组蛋白表达 3 h。

将 1 L 细胞培养液所产生菌体悬浮于 30 ml 裂解缓冲液 (6 mol/L 盐酸胍, 100 mmol/L NaH₂PO₄, 10 mmol/L Tris-HCl, pH 8.0), 温和混合直至溶液变清。室温, 12000 r/min 离心 60 min, 取上清。取上清加样到裂解缓冲液彻底平衡过的 1 ml Ni-NTA 树脂上。用 50 ml 洗涤液 (10 mmol/L Tris-HCl pH 6.3, 100 mmol/L NaH₂PO₄, 8 mol/L urea) 以 15 ml/h 速度洗柱后, 用 5 ml 洗脱液 (100 mmol/L NaH₂PO₄, 10 mmol/L Tris-HCl, pH 8.0, 8 mol/L urea, 250 mmol/L 咪唑) 以 8 ml/h 的速度洗脱; 分部收集洗脱液; 每组取等体积的蛋白溶液用 15% SDS-PAGE 分析, 并用 Bradford 方法测定蛋白含量。

变性 Ni-NTA 亲和纯化回收的蛋白必须复性才有活性。将含有表达蛋白的洗脱液合并, 在 4 $^{\circ}$ C 依次在含不同浓度尿素 (4 mol/L; 2 mol/L; 1 mol/L; 0.5 mol/L; 0 mol/L) 的储存液 (20 mmol/L Tris-HCl, 30 mmol/L NaCl pH 7.5, 5% glycerol, 5 mmol/L β -ME) 中透析, 最后在蛋白储存液 (20 mmol/L Tris-HCl, 30 mmol/L NaCl pH 7.5, 50% glycerol, 5 mmol/L β -ME) 中透析。获得的蛋白保藏于 -20 $^{\circ}$ C 备用。

1.2.4 体外剪接 纯化的融合蛋白于 4 $^{\circ}$ C 在剪接缓冲液 (20 mmol/L Tris-HCl, 30 mmol/L NaCl pH 7.5, 5% glycerol, 0.1% Triton X-100, 5 mmol/L β -ME) 中等浓度 (10~20 μ mol/L) 混合。在不同的时间段分别取出一定量的反应物, 并加入 SDS-PAGE 上样缓冲液 (8% SDS, 20% 巯基乙醇) 终止反应。剪接产物通过 SDS-PAGE 定量。

1.2.5 剪接产物体外活性测定 配置反应液, 含有 1 \times T7 RNA 聚合酶反应缓冲液, 0.1 ng/ μ l pET28a, 1 U/ μ l RNasin, 2 mmol/L UTP, 2 mmol/L CTP, 2 mmol/L GTP, 0.02 mmol/L ATP, 1 μ Ci/ μ l [α - 32 P]ATP; 在反应液中分别加入相同体积的细胞抽提物或商品 T7 RNA 聚合酶作为标准样, 于 37 $^{\circ}$ C 反应 60 min 后加入上样缓冲液 (50 mmol/L EDTA, 8 mol/L urea, 0.1% bromophenol blue, and 0.1% xylene cyanol) 终止反应。产物经含 8 mol/L 尿素的 20% 聚丙烯酰胺凝胶电泳检测。电泳结束后, 用磷屏成像系统 (Fuji, FLA5000) 显示凝胶电泳的荧光图谱。

2 结果

2.1 体外剪接系统的建立

尽管 npu DnaE 内含肽介导的剪接反应完全是自发的, 但是与内含肽 C 末端直接相连的外蛋白子残基

会影响其剪接。我们选择在 Ala491 和 Cys492 之间将 T7 RNA 聚合酶断裂成 N 端和 C 端两部分,如此一来,与 DnaE_C 的 C 末端毗邻的氨基酸残基为 Cys,而与 DnaE_N 的 N 端相邻接的残基为 Ala,这些都为 Npu DnaE 内含肽能有效地介导蛋白反式剪接提供了必要条件。

N 端前体蛋白包含 T7 RNA 聚合酶的 N 端部分(1~491 位残基),DnaE_N 以及与之直接相连的组氨酸标签(His₆)(图 1 a);C 端前体蛋白包含 DnaE_C 以及与之直接相连的 His₆ 标签, T7 RNA 聚合酶的 C 端部分(492~883 位残基)(图 1 b)。两者都是在 T7 启动子的控制下,加入诱导剂(IPTG)后表达。这两种前体蛋白分别表达,纯化,4℃下在剪接缓冲液中等量混合。借助 DnaE_N 和 DnaE_C 之间强的结合力,T7 RNA 聚合酶的 N 端和 C 端部分再结合形成完整的聚合酶(图 1 c)。

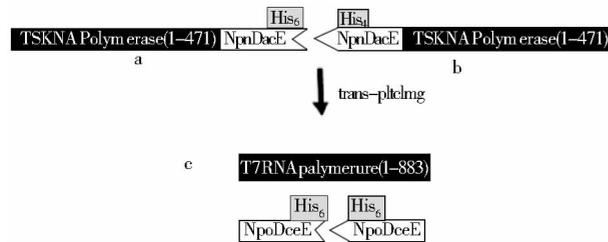


图 1 反式剪接过程图示

Fig. 1 Schematic drawing of the trans-splicing process

a: N-precursor protein; b: C-precursor protein;

c: Spliced T7 RNA polymerase

2.2 N 端和 C 端融合蛋白的有效表达

重组蛋白的可溶性预测 (<http://biotech.ou.edu/>)显示,N 端及 C 端融合蛋白主要以包涵体形式存在,所以我们选择在 8mol/L 尿素的变性条件下对其进行纯化。经 IPTG 诱导表达、Ni-NTA 树脂变性纯化、梯度透析复性,我们得到了电泳纯的可溶性重组蛋白,所获得的 C 端和 N 端融合蛋白的电泳纯度能达到 90% 以上(图 2,泳道 1,2)。SDS-PAGE 凝胶电泳所显示的 N 端和 C 端融合蛋白的表观分子量分别为 66kDa 和 47kDa,与根据基因序列所推测的分子量是一致的。

2.3 T7 RNA 聚合酶 N 端和 C 端片段的体外连接

首先,我们将 N 端和 C 端融合蛋白分别纯化,透析复性,浓缩。待两者的浓度达到 10~20μmol/L 时,在剪接缓冲液中等量混合。两种前体蛋白混合 30min 时即可在 97kDa 处检测到剪接产物的存在(图 2,泳道 3)。从 1L 菌液中纯化得到 2mg 的剪接产物—T7 RNA

聚合酶。

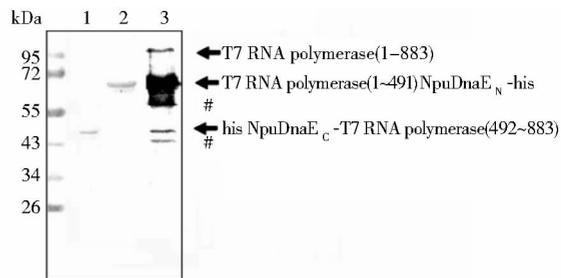


图 2 SDS-PAGE 检验前体蛋白的表达及反式剪接产物

Fig. 2 SDS-PAGE Examination of the expression of the precursor proteins containing split *Npu DnaE* and the trans-splicing reaction of the *Npu DnaE* intein

The gel was stained with Coomassie brilliant blue after electrophoresis. Lane1: his-NpuDnaE_C-T7 RNA polymerase (492~883) purified with denaturing Ni-NTA; Lane 2: T7 RNA polymerase (1~491)-NpuDnaE_N-his purified with denaturing Ni-NTA; Lane3: Reaction between T7 RNA polymerase (1~491)-NpuDnaE_N-his and his-NpuDnaE_C-T7 RNA polymerase (492~883) performed at 4℃. “#”

denotes protein impurity

2.4 连接产物的活性测定

截短的 T7 RNA 聚合酶检测不到活性(图 3,泳道 2,3),然而,剪接产生的完整的 T7 RNA 聚合酶具有类似于野生型聚合酶的活性(图 3,泳道 4)。我们通过体外转录反应来测定通过该反式剪接系统产生的 T7 RNA 聚合酶是否具有活性,同时以商品 T7 RNA 聚合酶(购自 Fermentas)作为阳性对照。

3 讨论

用传统的分子生物学的方法来合成大的蛋白、毒性蛋白或膜蛋白仍是个巨大的挑战。虽然之前有研究表明,某些毒性蛋白可以利用特殊的方法合成,但这些方法往往具有局限性,不适用于其他蛋白。在本项研究中,我们建立了一套系统,即“利用内含肽介导的反式剪接来合成大的蛋白”,并验证了该系统的可行性和有效性,从一定程度上解决了上述难题。

天然或人工断裂内含肽介导的反式剪接系统给毒性蛋白的合成带来了希望。然而,人工断裂内含肽如 Mtu RecA 内含肽,其两个片段再结合成有功能的内含肽需要经历变性/复性的过程^[16]。虽然天然断裂内含肽 *Synechocystis DnaE* 内含肽避免了变性/复性的过程,

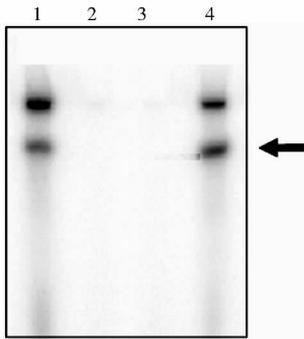


图3 连接的 T7 RNA 聚合酶的活性测定

Fig. 3 Activity assay of ligated T7 RNA polymerase

The reaction was performed at 37°C for 60 minutes in 20 μ l reaction solution contained 1 \times transcription buffer, 60 ng linearized pBS, 40U RNase inhibitor, 0.5mmol/L UTP, 0.5mmol/L CTP, 0.5mmol/L GTP, 0.05mmol/L ATP, 10 μ Ci [α -³²P] ATP, and certain amounts of spliced product. Analysis was carried out by electrophoresis on 20% polyacrylamide gels (8mol/L urea). Lane 1: Linearized pBS incubated with commercial T7 RNA polymerase; Lane 2: Linearized pBS incubated with N-part of T7 RNA polymerase (1 ~ 491)-NpuDnaE_N; Lane 3: Linearized pBS incubated with NpuDnaE_C-C-part of T7 RNA polymerase (492 ~ 883); Lane 4: Linearized pBS incubated with ligated T7 RNA polymerase. The transcript is highlighted by an arrow

但是对与其相邻接的外蛋白子残基的特定要求局限了它的应用。近期的研究表明, Npu DnaE 内含肽具有更大的潜力应用于蛋白质和细胞化学中。该内含肽在高浓度尿素存在时以及低温下仍具高的剪接活性, 这一特性使它更有优势, 尤其当要连接的目标蛋白可溶性低, 不稳定, 或对温度敏感。基于上述考虑, 我们选择 Npu DnaE 内含肽介导蛋白反式剪接。

当 T7 RNA 聚合酶在细胞内过量表达时会对细胞产生毒性, 另外, 它本身属于较大的蛋白, 分子量高达 97kDa, 所以, 它可以作为一种模式蛋白来检测我们的蛋白纯化系统是否可行。使用该系统时首先要解决的问题: 在哪个位点断裂 T7 RNA 聚合酶? 最佳断裂位点应不仅能使蛋白失活, 而且能产生最高的体外/体内剪接水平比。在本研究中, 我们在 491 位的 Ala 和 492 位的 Cys 之间断裂 T7 RNA 聚合酶, 因为这样一来, 与内含肽的 C 末端相连的外蛋白子残基为 Cys, 而与内含肽 N 末端相连的外蛋白子残基为 Ala, 这两点都保证了内含肽的高剪接活性。这一断裂位点在本实验室其他人员的研究中也被证实可行。

我们对 N 端及 C 端融合蛋白的可溶性进行了分

析, 发现约 80% 以包涵体的形式存在, 因此须在高浓度液盐中纯化这两种内含肽融合蛋白。当 N 端及 C 端前体蛋白的浓度均达到 10 ~ 20 μ mol/L, 并于 4°C 在剪接缓冲液中等比例混合时, 大小约 97kDa 的剪接产物——T7 RNA 聚合酶在 30min 后即可产生。从 1L 培养物中能获得 2mg 有功能的 T7 RNA 聚合酶。

本研究中建立的“利用断裂内含肽介导的反式剪接来合成大的蛋白”的体系, 理论上也可以应用于其他大蛋白, 毒性蛋白或膜蛋白的合成。这一体系不仅使以前难以表达的蛋白的合成变得简便可行, 而且有望使其合成产业化、规模化。

参考文献

- [1] Perler F B. InBase, the Intein Database. *Nucleic Acids Res.*, 2000, 28: 344 ~ 345
- [2] Southworth M W, Adam E, Panne D, et al. Control of protein splicing by intein fragment reassembly. *EMBO J.*, 1998, 17: 918 ~ 926
- [3] Mills K V, Lew B M, Jiang S Q, et al. Protein splicing in trans by purified N- and C-terminal fragments of the *Mycobacterium tuberculosis* RecA intein. *Proc Natl Acad Sci USA*, 1998, 95: 3543 ~ 3548
- [4] Wu H, Xu M Q, Liu X Q. Protein trans-splicing and functional mini-inteins of a cyanobacterial DnaB intein. *Biochim Biophys Acta*, 1998, 1387: 422 ~ 432
- [5] Yamazaki T, Otomo T, Oda N, et al. Segmental isotope labeling for protein NMR using peptide splicing. *J Am Chem Soc.*, 1998, 120: 5591 ~ 5592
- [6] Lew B W, Mills K V, Paulus H. Protein splicing *in vitro* with a semisynthetic two-component minimal intein. *J Biol Chem.*, 1998, 273: 15887 ~ 15890
- [7] Scott C P, Abel-Santos E, Wall M, et al. Production of cyclic peptides and proteins *in vivo*. *Proc Natl Acad Sci USA*, 1999, 96: 13638 ~ 13643
- [8] Evans T C, Martin D, Kolly R, et al. Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of *Synechocystis* species PCC6803. *J Biol Chem.*, 2000, 275: 9091 ~ 9094
- [9] Iwai H, Lingel A, Plückthun A. Cyclic green fluorescent protein produced *in vivo* using an artificially split PI-PfuI intein from *Pyrococcus furiosus*. *J Biol Chem.*, 2001, 276: 16548 ~ 16554
- [10] Mootz H D, Tyszkiewicz A B, Muir T W. Conditional protein splicing: a new tool to control protein structure and function *in vitro* and *in vivo*. *J Am Chem Soc.*, 2003, 125: 10561 ~ 10569
- [11] Buskirk A R, Ong Y C, Gartner Z J, et al. Directed evolution of

- ligand dependence: small-molecule-activated protein splicing. Proc Natl Acad Sci USA, 2004, 101:10505 ~ 10510
- [12] Ozawa T, Kaihara A, Sato M, et al. Split luciferase as an optical probe for detecting protein - protein interactions in mammalian cells based on protein splicing. Anal Chem, 2001, 73: 2516 ~ 2521
- [13] Chin H G, Kim G D, Marin I, et al. Protein trans-splicing in transgenic plant chloroplast: reconstruction of herbicide resistance from split genes. Proc Natl Acad Sci USA, 2003, 100: 4510 ~ 4515
- [14] Li J. Protein trans-splicing as a means for viral vector-mediated *in vivo* gene therapy. Human Gene Therapy, 2008, 19: 958 ~ 964
- [15] Zettler J, Schütz V. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. FEBS Letters, 2009, 583: 909 ~ 914
- [16] Mills K V, Lew B W, Jiang S, et al. Protein splicing in trans by purified N- and C-terminal fragments of the *Mycobacterium tuberculosis RecA* intein. Proc Natl Acad Sci USA, 1998, 95: 3543 ~ 3548

Large Protein Production Through Split Intein-Mediated Trans-Splicing

ZHANG Jing LIU Huan ZHOU Jing LIU Jian-hua

(College of Life Science & Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract *Npu DnaE* intein was used to produce some large proteins, which were difficult to obtain through conventional expression systems. A T7 expression system was described, by which the gene of T7 RNA polymerase is split into two pieces, and each piece fuses with *Npu DnaE* N- and C- terminal sequences respectively. Functional T7 RNA polymerase is created by mixing the two kinds of fusion constructs *in vitro*. The approach of split intein-mediated production of large proteins, in theory, readily generalizable to the purification of other large, cytotoxic or membrane proteins.

Key words Large cytotoxic or membrane proteins *Npu DnaE* intein T7 RNA polymerase

《〈自然〉百年科学经典》首发式在北京举行

全球迄今最大规模的《自然》杂志论文选集——《〈自然〉百年科学经典》12月3日在中国科学院国家科学图书馆首发。全国人大常委会副委员长、中国科学院院长路甬祥, 国家新闻出版总署副署长邝书林, 《自然》主编菲利普·坎贝尔, 麦克米伦出版(中国)有限公司出版董事朗礼, 丛书的中方编委中国科学院院士许智宏、赵忠贤、滕吉文以及审稿专家、译者等200余人参加首发式。

英汉双语对照版《〈自然〉百年科学经典》(Nature: the Living Record of Science)是由外语教学与研究出版社联合麦克米伦出版集团和自然出版集团共同策划编辑的科学主题丛书。丛书共10卷, 收录并翻译了《自然》杂志自1869年创刊以来近150年间发表过的840余篇经典文献。

该丛书所选文章涵盖物理、化学、天文、地理和生物等基础学科及众多交叉学科, 从狭义相对论的提出到量子理论的日趋成熟, 从同位素的发现到纳米管的诞生, 从进化论之争到人类基因组测序完成……再现了一个多世纪以来人类在自然科学领域艰辛跋涉、不断探索的历史足迹, 堪称一部鲜活的近代科学史诗。

许智宏在首发式上的讲话中指出, 目前, 我国正在建设自主创新型的社会, 这从根本上取决于对青年科技人才的培养和使用。不拘一格, 海纳百川, 让各种创新思维竞相迸发, 这是我国科技政策的核心, 也是我们未来努力的方向, 期望这套《〈自然〉百年科学经典》的出版能够对此有所裨益。

此次发布的是这套丛书的前两卷, 其余8卷预计于2010年陆续出版。