

# 真核生物 DNA 连接酶 III 的功能演化

靳春艳<sup>1,2</sup>, 盛自章<sup>2</sup>, 黄京飞<sup>2,3</sup>

(1. 中国科学技术大学生命科学学院, 安徽合肥 230027;

2. 中国科学院昆明动物研究所 遗传资源与进化国家重点实验室, 云南昆明 650223;

3. 中国科学院昆明动物研究所-香港中文大学生物资源与疾病分子机理联合实验室, 云南昆明 650223)

**摘要:** DNA 连接酶 III 被认为只存在于脊椎动物, 并在细胞核 DNA 的修复和线粒体 DNA 的复制和修复过程中发挥功能. 虽然近来有关于无脊椎动物中存在着 DNA 连接酶 III 的报道, 但其功能演化及在无脊椎动物中的分布仍不清楚. 为进一步探讨 DNA 连接酶 III 的功能演化, 进行了数据库搜索、线粒体定位信号 (MLS) 预测和功能位点保守性分析等. 研究结果显示: DNA 连接酶 III 在变形虫、动物界和领鞭毛虫中广泛存在, 但其在真菌界等发生整个蛋白或部分结构域的丢失; 很多物种的 DNA 连接酶 III 不含线粒体定位信号, 因此, 它们不太可能在线粒体中发挥作用, 而参与细胞核 DNA 的修复是 DNA 连接酶 III 较为古老和保守的功能.

**关键词:** DNA 连接酶 III; 线粒体定位信号; ZnF; BRCT; 功能演化

**中图分类号:** Q349+.53

**文献标识码:** A

doi:10.3969/j.issn.0253-2778.2012.04.007

**引用格式:** Jin Chunyan, Sheng Zizhang, Huang Jingfei. Functional evolution of Eukaryote DNA ligase III [J].

Journal of University of Science and Technology of China, 2012, 42(4):302-310.

靳春艳, 盛自章, 黄京飞. 真核生物 DNA 连接酶 III 的功能演化[J]. 中国科学技术大学学报, 2012, 42(4):302-310.

## Functional evolution of Eukaryote DNA ligase III

JIN Chunyan<sup>1,2</sup>, SHENG Zizhang<sup>2</sup>, HUANG Jingfei<sup>2,3</sup>

(1. School of Life Sciences, University of Science and Technology of China, Hefei 230027, China;

2. State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China;

3. Kunming Institute of Zoology-Chinese University of Hongkong Joint Research Center for Bio-resources and Human Disease Mechanisms, Kunming 650223, China)

**Abstract:** Previous studies revealed that DNA ligase III was restricted to vertebrates and functioned in nucleus DNA repair and mitochondria DNA replication and repair. Although recent researches have reported that DNA ligase III is also found in non-vertebrates, little attention has been devoted to the distribution and functional evolution of DNA ligase III. To probe the functional evolution of DNA ligase III, database searches, mitochondrial localization signal prediction (MLS) and functional conservation analysis were performed. The results show that, DNA ligase III can be observed in amoebozoa, metazoa and choanoflagellates, but the whole protein or some domains are lost in some species including fungi. The

收稿日期:2011-12-12;修回日期:2012-04-11

基金项目:国家重点基础研究发展(973)计划(2009CB941300),国家自然科学基金(31123005)资助.

作者简介:靳春艳,女,1984年生,硕士生.研究方向:蛋白质结构和功能的进化. E-mail: chunyj@mail.ustc.edu.cn

通讯作者:黄京飞,研究员. E-mail: huangjf@mail.kiz.ac.cn

MLS prediction analysis suggests that, the DNA ligase III in many species can not function in mitochondria, and is consequently less likely to play a role for DNA ligase III in mitochondria. The conservation analyses of functional site demonstrate that nucleus DNA repair is an ancient and conserved function of DNA ligase III.

**Key words:** DNA ligase III; mitochondria localization signal; ZnF; BRCT; functional evolution

## 0 引言

DNA 连接酶(DNA ligase)能够催化 DNA 末端的 3'-羟基和 5'-磷酸基形成 3',5'-磷酸二酯键<sup>[1]</sup>, 是 DNA 复制、重组和修复过程的关键蛋白<sup>[2]</sup>. 根据催化反应所需能量来源的不同, DNA 连接酶可以分为腺苷三磷酸(ATP)依赖型和烟酰胺腺嘌呤二核苷酸(NAD<sup>+</sup>)依赖型. ATP 依赖型 DNA 连接酶广泛分布于真核生物、古细菌、真细菌和病毒中; NAD<sup>+</sup> 依赖型 DNA 连接酶则分布于古细菌、真细菌和病毒中<sup>[2]</sup>. 真核生物存在 3 种 ATP 依赖型 DNA 连接酶: DNA 连接酶 I, DNA 连接酶 III 和 DNA 连接酶 IV. 研究显示, DNA 连接酶 I 和 DNA 连接酶 IV 广泛分布于真核生物中, 如植物界和动物界, DNA 连接酶 III 则主要分布于脊椎动物中<sup>[2]</sup>. 这 3 个 DNA 连接酶有相同的催化核心, 即 DBD (DNA binding domain) 和 OBD (OB-fold domain) 结构域环绕并稳定 DNA, NTase (nucleotidyl transferase domain) 结构域催化连接 DNA 末端<sup>[2]</sup>. 然而这 3 个酶具有不同的底物特异性并倾向于在不同的 DNA 复制和修复途径中发挥作用, 这与 3 个连接酶含有的特异的 N-端和 C-端 motif 或结构域有关<sup>[2]</sup>. DNA 连接酶 I 主要在细胞核 DNA 复制和修复过程行使功能<sup>[3]</sup>, 在真菌等中, DNA 连接酶 I 的 N-端含有线粒体定位信号 (mitochondrial localization signal, MLS), 使 DNA 连接酶 I 可以参与线粒体 DNA 的复制和修复<sup>[4]</sup>. DNA 连接酶 IV 在非同源重组的 DNA 双链断裂修复和免疫相关基因的 V(D)J 重组中发挥作用<sup>[5-6]</sup>, 其 C-端含有两个 BRCT (breast cancer susceptibility protein 1 C-terminal domain) 结构域, 可以与 XRCC4 (X-ray repair cross-complementing group 4) 相互作用, 该相互作用对于维持 DNA 连接酶 IV 的稳定和催化功能有重要作用<sup>[7]</sup>.

DNA 连接酶 III 在核苷酸切除修复 (nucleotide excision repair)、碱基切除修复 (base excision repair) 和单链断裂修复 (single strand break repair)

等过程发挥功能<sup>[8-11]</sup>. 在哺乳动物和两栖动物中, DNA 连接酶 III 可以形成  $\alpha$  和  $\beta$  两种选择性剪切体,  $\alpha$  型剪切体的 C-端含有一个 BRCT 结构域,  $\beta$  型的 C-端则是由 17 到 18 个氨基酸残基组成的细胞核定位信号 (nuclear localization signal, NLS)<sup>[12]</sup>. DNA 连接酶 III  $\alpha$  和  $\beta$  都可以通过选择性翻译形成 N-端含有 MLS 和不含 MLS 的两种形式<sup>[12-13]</sup>. 但目前只报道 DNA 连接酶 III  $\alpha$  在线粒体中有功能, 而 DNA 连接酶 III  $\beta$  主要在生殖细胞中表达<sup>[12]</sup>. DNA 连接酶 III 的 BRCT 结构域可以与 XRCC1 (X-ray repair cross-complementing group 1) C-端的 BRCT 结构域形成异二聚体, 该相互作用可以维持 DNA 连接酶 III 的稳定和将其定位至 DNA 损伤位点<sup>[9]</sup>. DNA 连接酶 III 的 N-端含有能探测 DNA 末端的 ZnF (zinc finger) 结构域, 其与 DBD 结构域相互作用, 可以提高 DNA 连接酶 III 催化连接 DNA 单链断裂和平齐末端的效率<sup>[14]</sup>.

最近研究显示, DNA 连接酶 III 还在无脊椎动物中存在, 在系统演化关系上, DNA 连接酶 III 与 DNA 连接酶 IV 首先聚在一起, 它们与 DNA 连接酶 I 的距离较远. 可以推测出, 与 DNA 连接酶 I 相比, DNA 连接酶 III 与 DNA 连接酶 IV 的演化关系更为密切<sup>[15]</sup>. 但其功能演化及在无脊椎动物中的分布还不清楚. 本研究通过数据库搜索、线粒体定位信号 (MLS) 预测结合功能位点保守性分析, 更进一步地对 DNA 连接酶 III 物种分布及其功能演化进行了探讨.

## 1 材料与方法

DNA 连接酶 III 的序列从 NCBI non-redundant (nr), DOE joint genome institute (JGI) 和 Ensemble 数据库获取. 序列搜索使用 Blastp<sup>[16]</sup>, 最大搜索序列数设为 1000, 搜索序列使用人的 DNA 连接酶 III (Accession number: P49916. 2). 直系同源关系的鉴定使用双向 Blastp 最佳匹配方法 (reciprocal blast best hit)<sup>[17]</sup>. 结构域预测使用 Pfam 数据库<sup>[18]</sup>. 多序列比对使用 MAFFT 软件和

genafpair 策略<sup>[19]</sup>. MLS 的预测使用 Mitoprot 和 TargetP<sup>[20-21]</sup>.

结构域的保守性计算:首先从蛋白质大分子数据库 PDB 中获取了 DNA 连接酶 III 各结构域的 PDB 结构. ZnF 结构域 PDB ID 为 1UW0; DBD, Ntase 和 OBD 结构域 PDB ID 为 3L2P; BRCT 结构域 PDB ID 为 1IMO. 然后对数据库搜索得到的 DNA 连接酶 III 序列的每个结构域的序列分别使用 MAFFT 进行比对. 最后将每个结构域的多序列比对和结构提交到 ConSurf 网站, 提取序列的保守性信息并将

其绘制到结构上, 保守性计算采用最大似然法<sup>[22]</sup>.

## 2 结果

### 2.1 DNA 连接酶 III 的物种分布和结构域组成

通过搜索 NCBI nr 蛋白数据库和 Ensemble 基因组数据库, 我们获取了 58 个物种的 DNA 连接酶 III 序列(表 1). 这些物种属于变形虫(Amoebozoa)、动物界(Metazoa)和领鞭毛虫(Choanoflagellates), 而在真菌界(Fungi)则没有找到 DNA 连接酶 III, 说明其在真菌界中发生了丢失.

表 1 DNA 连接酶 III 的物种分布及结构域排列

Tab. 1 The distribution and domain composition of DNA ligase III

species	length	accession number	domain	targetP	MLS	MitoProt
<i>Monosiga brevicollis</i> MX1	1 132	NCBI XP_001745548	ZnF DBD NTase OBD BRCT	0.796	M	0.899 3*
<i>Dictyostelium discoideum</i> AX4	1 175	NCBI XP_638847.1	NTase OBD BRCT BRCT	0.108	_	0.060 1
<i>Polysphondylium pallidum</i> PN500	1 017	NCBI EFA78161	NTase OBD BRCT BRCT	0.121	_	0.033 5
<i>Trichoplax adhaerens</i>	823	NCBI XP_002113418.1	ZnF DBD NTase OBD	0.103	_	0.131 7
<i>Nematostella vectensis</i>	864	NCBI XP_001630058.1	ZnF DBD NTase OBD	0.058	_	0.139 4
<i>Brugia malayi</i>	607	NCBI XP_001896024.1	DBD NTase OBD	0.039	_	0.027 8
<i>Caenorhabditis remanei</i>	491	NCBI XP_003113685.1	ZnF DBD	0.261	_	0.553 8
<i>Caenorhabditis elegans</i>	493	NCBI NP_505659.1	ZnF DBD	0.325	_	0.696 8
<i>Caenorhabditis briggsae</i>	467	NCBI XP_002646686.1	ZnF DBD	0.276	_	0.848 2
<i>Helobdella robusta</i>	1 013	JGI 113751	ZnF DBD NTase OBD	0.117	_	0.133 8
<i>Capitella capitata</i>	841	JGI 213234	ZnF DBD NTase OBD BRCT	0.083	_	0.589 8
<i>Lottia gigantea</i>	903	JGI 228755	ZnF DBD NTase OBD BRCT	0.092	_	0.422 2
<i>Drosophila ananassae</i>	788	NCBI XP_001953809.1	DBD NTase OBD BRCT	0.818	M	0.949 2*
<i>Drosophila willistoni</i>	796	NCBI XP_002070126.1	DBD NTase OBD BRCT	0.049	_	0.263 4
<i>Drosophila melanogaster</i>	806	NCBI NP_650187.2	DBD NTase OBD BRCT	0.562	M	0.963 5*
<i>Drosophila mojavensis</i>	788	NCBI XP_001999197.1	DBD NTase OBD BRCT	0.067	_	0.104 3
<i>Drosophila virilis</i>	818	NCBI XP_002054326.1	DBD NTase OBD BRCT	0.781	M	0.984 1*
<i>Nasonia vitripennis</i>	941	NCBI XP_001601405.1	ZnF DBD NTase OBD	0.04	_	0.012 3
<i>Apis mellifera</i>	1 009	NCBI XP_396537.3	ZnF DBD NTase OBD	0.06	_	0.033 1
<i>Pediculus humanus corporis</i>	927	NCBI XP_002425250.1	ZnF DBD NTase OBD	0.089	_	0.011 9
<i>Acyrtosiphon pisum</i>	875	NCBI XP_001943689.1	ZnF DBD NTase OBD	0.048	_	0.050 3
<i>Tribolium castaneum</i>	853	NCBI XP_967954.1	ZnF DBD NTase OBD	0.05	_	0.646 6
<i>Ixodes scapularis</i>	851	NCBI XP_002406581.1	ZnF DBD NTase OBD BRCT	0.075	_	0.351 3
<i>Strongylocentrotus purpuratus</i>	875	NCBI XP_001188625.1	ZnF DBD NTase OBD	0.07	_	0.066 4
<i>Ciona savignyi</i>	854	EN ENSCSAVP00000011468	ZnF DBD NTase OBD BRCT	0.118	_	0.706 4
<i>Ciona intestinalis</i>	854	NCBI XP_002129234.1	ZnF DBD NTase OBD BRCT	0.118	_	0.706 4
<i>Branchiostoma floridae</i>	928	NCBI XP_002613830	ZnF DBD NTase OBD BRCT	0.341	_	0.086 2
<i>Taki fugu rubripes</i>	936	EN ENSTRUP00000027351	ZnF DBD NTase OBD BRCT	0.092	_	0.072 8
<i>Gasterosteus aculeatus</i>	984	EN ENSGACP00000027561	ZnF DBD NTase OBD BRCT	0.816	M	0.976 7*
<i>Danio rerio</i>	751	NCBI NP_001025345.1	ZnF DBD NTase OBD	0.071	_	0.252 8
<i>Tetraodon nigroviridis</i>	873	NCBI CAG00890.1	ZnF DBD NTase OBD	0.082	_	0.340 9
<i>Xenopus tropicalis</i>	986	EN ENSXETG00000019634	ZnF DBD NTase OBD BRCT	0.888	M	0.989 7*
<i>Xenopus (Silurana) tropicalis</i>	893	NCBI NP_001016351.2	ZnF DBD NTase OBD	0.888	M	0.990 3*
<i>Xenopus laevis</i>	988	NCBI NP_001082183.1	ZnF DBD NTase OBD BRCT	0.543	M	0.976 6*
<i>Gallus gallus</i>	900	NCBI NP_001006215.1	ZnF DBD NTase OBD BRCT	0.073	_	0.186 8

表 1(续)

species	length	accession number	domain	targetP	MLS	MitoProt
<i>Taeniopygia guttata</i>	929	NCBI XP_002192251.1	ZnF DBD NTase OBD BRCT	0.073	_	0.154
<i>Meleagris gallopavo</i>	897	NCBI XP_003211777.1	ZnF DBD NTase OBD BRCT	0.073	_	0.069 7
<i>Microcebus murinus</i>	1 011	EN ENSMICG00000006814	ZnF DBD NTase OBD BRCT	0.469	M	0.869 9*
<i>Pteropus vampyrus</i>	1 012	EN ENSPVAG000000013427	ZnF DBD NTase OBD BRCT	0.328	S	0.866 6
<i>Equus caballus</i>	1 013	NCBI XP_001501551	ZnF DBD NTase OBD BRCT	0.692	M	0.930 8*
<i>Macaca mulatta</i>	1 009	NCBI XP_001113800.1	ZnF DBD NTase OBD BRCT	0.718	M	0.815 1*
<i>Gorilla gorilla</i>	1 009	NCBI ENSGGOG00000025637	ZnF DBD NTase OBD BRCT	0.75	M	0.859 9*
<i>Pongo abelii</i>	1 009	NCBI XP_002827301.1	ZnF DBD NTase OBD BRCT	0.608	M	0.861*
<i>Pan troglodytes</i>	1 009	NCBI XP_511409.2	ZnF DBD NTase OBD BRCT	0.804	M	0.882 6*
<i>Homo sapiens</i>	1 009	NCBI P49916.2	ZnF DBD NTase OBD BRCT	0.763	M	0.880 9*
<i>Callithrix jacchus</i>	1 016	NCBI XP_002748407.1	ZnF DBD NTase OBD BRCT	0.286	_	0.722 5
<i>Tursiops truncatus</i>	1 000	EN ENSTTRG00000015470	ZnF DBD NTase OBD BRCT	0.112	_	0.558 6
<i>Ailuropoda melanoleuca</i>	1 014	NCBI XP_002923661.1	ZnF DBD NTase OBD BRCT	0.687	M	0.862 1*
<i>Oryctolagus cuniculus</i>	1 011	NCBI XP_002719278.1	ZnF DBD NTase OBD BRCT	0.195	_	0.590 2
<i>Canis familiaris</i>	1 012	NCBI XP_548265.2	ZnF DBD NTase OBD BRCT	0.647	M	0.834*
<i>Cavia porcellus</i>	1 006	EN ENSCPOP00000011732	ZnF DBD NTase OBD BRCT	0.193	_	0.751 4
<i>Dipodomys ordii</i>	1 011	EN ENSDORG00000013422	ZnF DBD NTase OBD BRCT	0.556	M	0.643 3*
<i>Ochotona princeps</i>	1 007	EN ENSOPRG00000011746	ZnF DBD NTase OBD BRCT	0.816	M	0.726*
<i>Loxodonta africana</i>	1 017	EN ENSLAFG00000001574	ZnF DBD NTase OBD BRCT	0.822	M	0.823 1*
<i>Mus musculus</i>	1 015	NCBI P97386.2	ZnF DBD NTase OBD BRCT	0.199	_	0.642 2
<i>Rattus norvegicus</i>	1 018	NCBI EDM05447.1	ZnF DBD NTase OBD BRCT	0.291	_	0.778 6
<i>Bos taurus</i>	944	NCBI NP_001033196.1	ZnF DBD NTase OBD	0.657	M	0.617 9*
<i>Monodelphis domestica</i>	1 011	NCBI XP_001373823.2	ZnF DBD NTase OBD BRCT	0.74	M	0.971 6*

【注】 Accession numbers are from NCBI(<http://www.ncbi.nlm.nih.gov/protein/>), Ensemble(<http://www.ensembl.org/>) and JGI Gonome portal (<http://genome.jgi-psf.org/>) databases, referred to as “NCBI|”, “EN|” and “JGI|” respectively. TargetP and Mitoprot are used to predict MLS (“M”: proteins in mitochondria; “S”: proteins in cytoplasm, “\_”: the protein location is unclear). “\*” means the signal sequences can be predicted.

序列分析发现, DNA 连接酶 III 的氨基酸序列长度从 467 至 1172 不等, 多数序列长度在 1000 氨基酸左右. 我们通过 Pfam 数据库预测了这些 DNA 连接酶 III 序列的结构域组成(表 1)<sup>[18]</sup>. 结果显示, 多数物种的 DNA 连接酶 III 都含有 ZnF 结构域、DBD 结构域、NTase 结构域、OBD 结构域和 BRCT 结构域. 粘液菌(*Polysphondylium pallidum* PN500)和盘基网柄菌(*Dictyostelium discoideum* AX4)中的 DNA 连接酶 III 不含 ZnF 结构域, 但 C-端含有两个 BRCT 结构域(*Dictyostelium discoideum* AX4 中 BRCT1:886-963 和 BRCT2:1070-1161), 而 BRCT2 与其他物种 DNA 连接酶 III 中的 BRCT 结构域具有较高的序列一致性. 在线虫动物门(Nematoda)线虫属(*Caenorhabditis*)的秀丽隐杆线虫(*Caenorhabditis elegans*), *Caenorhabditis remanei* 和 *Caenorhabditis briggsae* 中, DNA 连接酶 III 序列较短(*C. elegans*: 493 氨基酸), 只含有 ZnF 结构域和 DBD 结构域; 而布鲁格氏丝虫属(*Brugia*)中的

马来丝虫(*Brugia malayi*)含有 DBD 结构域、NTase 结构域和 OBD 结构域, 但缺失了 ZnF 结构域和 BRCT 结构域. 此外, 果蝇属物种不含有 ZnF 结构域, 而丝盘虫(*Trichoplax adhaerens*)、海葵(*Nematostella vectensis*)、斑马鱼(*Danio rerio*)和绿河豚(*Tetraodon nigroviridis*)中的 DNA 连接酶 III 则不含 BRCT 结构域.

## 2.2 DNA 连接酶 III 的功能演化

为了探讨 DNA 连接酶 III 的功能演化, 我们分析了 DNA 连接酶 III 各结构域和 motif 的功能位点的保守性.

### 2.2.1 N-端 MLS 的保守性

序列比对结果显示, DNA 连接酶 III 的 MLS 区序列一致性较差, 这与之前研究发现直系同源蛋白的 MLS 保守性较差相一致<sup>[23]</sup>. 因此, 我们不能通过序列比对得到除哺乳纲和两栖纲动物外其他物种中 DNA 连接酶 III 入线粒体的信息. 本研究选用了应用较多和预测准确性较高的 MitoProt 和 TargetP 来

预测 DNA 连接酶 III 是否含有 MLS(表 1)<sup>[24]</sup>.

结果显示,非洲爪蟾属和大部分哺乳纲动物的 DNA 连接酶 III 被预测含有 MLS,这与之前报道是一致的<sup>[12-13]</sup>.然而,在珍珠鸟(*Zebra Finch*)、鸡(*Gallus gallus*)和火鸡(*Meleagris gallopavo*)中的 DNA 连接酶 III 没有发现 MLS;鱼纲除三刺鱼(*Gasterosteus aculeatus*)外,也没有发现 MLS.在无脊椎动物中,只有领鞭毛虫(*Monosiga brevicollis* MX1)和果蝇属中的嗜凤梨果蝇(*Drosophila ananassae*),黑腹果蝇(*Drosophila melanogaster*)和黑果蝇(*Drosophila virilis*)的 DNA 连接酶 III 含有 MLS 序列(表 1,图 1(a)).这些结果表明,只有部分物种中的 DNA 连接酶 III 可以入线粒体.

### 2.2.2 ZnF 结构域的保守性

序列比对发现,ZnF 结构域结合锌离子的 4 个氨基酸残基<sup>[25]</sup>:人 DNA 连接酶 III  $\alpha$  中的 Cys18, Cys21, His52 和 Cys55 都是非常保守的. ZnF 结构域结合 DNA 的区域位于  $\beta$  折叠与  $\alpha 1$  螺旋形成的 V-形裂隙处,包括 Lys19, Lys20, Arg31, Lys48, Trp50, Lys59, Arg62, Arg64, Lys68 和 Lys69 等<sup>[25]</sup>.除 Trp50 在少数物种中被 Tyr 替代和 Arg64 在一些物种中被疏水性残基替代外,其他残基都较为保守(图 1(a),图 2(a)),表明 ZnF 结构域结合 DNA 的功能是保守的.

### 2.2.3 催化核心结构域的保守性

研究显示,DBD 结构域有两个重要的功能位点<sup>[14]</sup>:一个是由两个螺旋-发卡-螺旋结构及一段伸展的环区组成的 DNA 结合区(人 DNA 连接酶 III  $\alpha$ :191-194,230-236,301-304,332-338),可以结合含切口的 DNA 的主链和小沟;另一个是由  $\alpha 1$  螺旋(人 DNA 连接酶 III  $\alpha$ :175-189)和  $\alpha 9$  螺旋(人 DNA 连接酶 III  $\alpha$ :318-330)组成的带正电势的沟槽区,可能与 ZnF 结构域相互作用.序列比对分析显示,DNA 结合区保守性较差,这些序列差异可能会导致结合 DNA 亲和力的变化<sup>[14]</sup>.在不含 ZnF 结构域的 DNA 连接酶 III 中, $\alpha 1$  螺旋和  $\alpha 9$  螺旋的保守性较差,如粘液菌(*Polysphondylium pallidum* PN500),盘基网柄菌(*Dictyostelium discoideum* AX4)和果蝇属等(图 1(a)),因此,ZnF 结构域和这两个螺旋间可能存在共进化,该结果提示,两个螺旋可能与 ZnF 结构域相互作用,并且这一相互作用的缺失可能会导致这些物种的 DNA 连接酶 III 催化效率降低.

NTase 和 OBD 结构域中的残基尤其是环绕

DNA 的碱性氨基酸残基都较保守.除线虫属外,NTase 结构域的活性位点 Lys421<sup>[2]</sup>(人 DNA 连接酶 III  $\alpha$ )在其他物种都是保守的.此外,绿河豚(*Tetraodon nigroviridis*)中的 DBD,NTase 和 OBD 结构域区域含有 5 个片段插入,火鸡(*Meleagris gallopavo*)中的 NTase 和 OBD 结构域则含有小片段缺失(图 1(a),图 3),但这些插入和缺失都位于环区,因而可能不会影响结构域的稳定,但对结构域功能的影响还有待进一步研究.

### 2.2.4 BRCT 结构域的保守性

DNA 连接酶 III 的 BRCT 结构域以  $\alpha 1$  至  $\beta 2$  区结合 XRCC1,相互作用残基包括 Asp849,Arg867,Arg869,Arg870,Asp876 和 Asp878(人 DNA 连接酶 III  $\alpha$ )<sup>[9]</sup>.序列分析显示(图 1(a),图 4(a)):除了 DNA 连接酶 III 的 BRCT 结构域在一些物种丢失外,该相互作用位点还在昆虫纲丢失;盘基网柄菌(*Dictyostelium discoideum* AX4)中的 BRCT2 与其他物种中的 BRCT 结构域序列相似性较高,但 Asp849 被 Thr1070 替代,Arg867 被 Thr1088 替代,Asp876 被 Gly1097 替代,该 BRCT 能否结合 XRCC1 C-端的 BRCT 结构域还有待进一步研究;而 BRCT1 结构域的功能不详.此外,DNA 连接酶 III 和 XRCC1 中的相互作用位点在领鞭毛虫(*Monosiga brevicollis* MX1)中都保守(XRCC1 中的相互作用位点数据未显示).参照 Tree of life project (<http://tolweb.org/tree/>)中动物界物种的演化关系,与盘基网柄菌(*Dictyostelium discoideum* AX4)相比,领鞭毛虫(*Monosiga brevicollis* MX1)与动物界和真菌界演化关系较近.这些分析说明,DNA 连接酶 III 与 XRCC1 的相互作用可能现于动物界和领鞭毛虫分化之前,但在随后的演化中发生了一些物种特异的丢失.

## 3 讨论

本文通过物种分布、结构域组成和功能位点保守性的分析,探讨了真核生物 DNA 连接酶 III 的演化.研究结果进一步表明:DNA 连接酶 III 可能起源于 DNA 连接酶 IV 的复制<sup>[26]</sup>,因而较古老的 DNA 连接酶 III 的 N-端不含 ZnF 结构域,但 C-端含两个 BRCT 结构域,如粘液菌(*Polysphondylium pallidum* PN500)和盘基网柄菌(*Dictyostelium discoideum* AX4)中的 DNA 连接酶 III;在进化过程中,DNA 连接酶 III 丢掉了 BRCT1,并且其 N-端

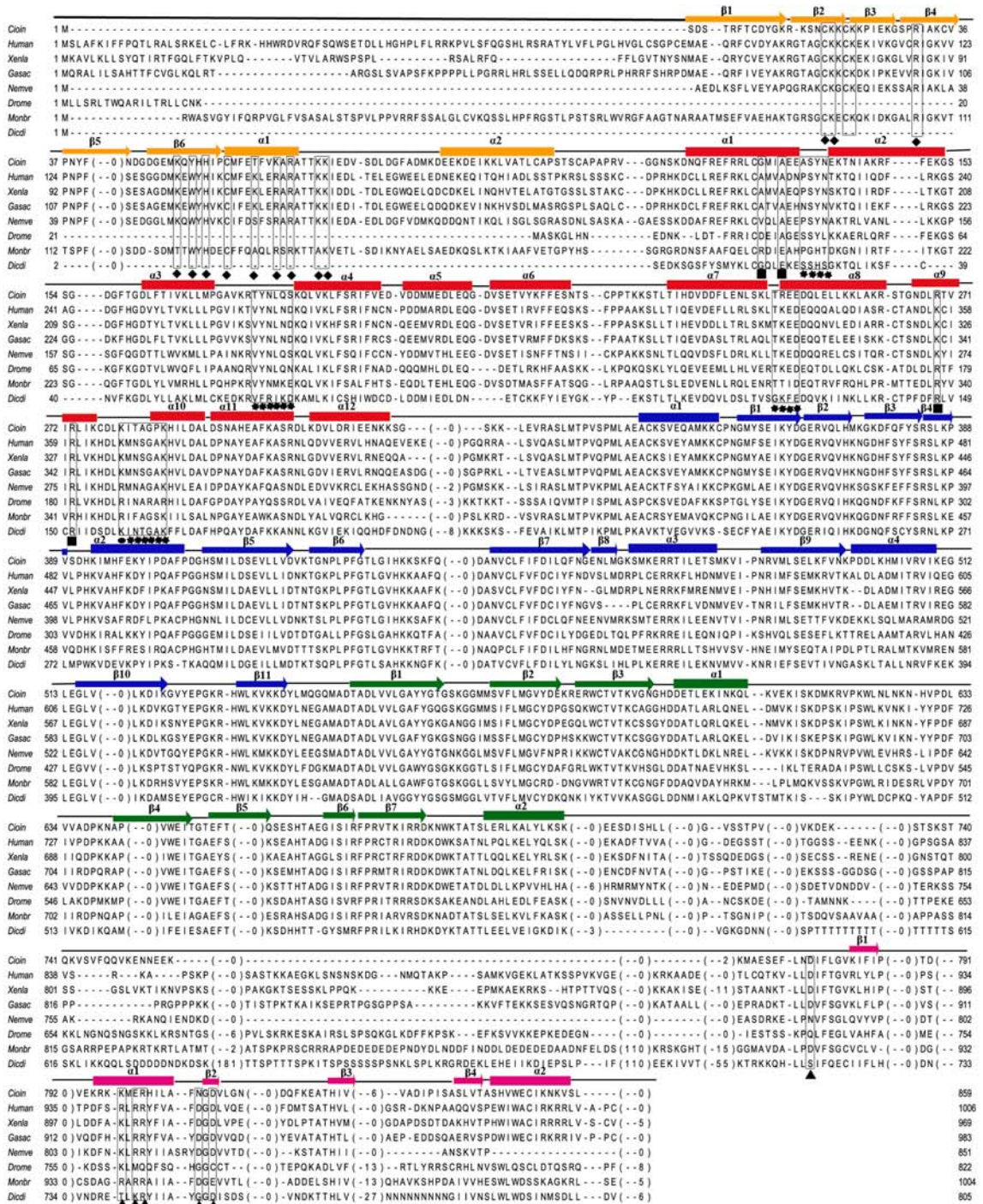


图 1 真核生物中的 DNA 连接酶 III  $\alpha$  多序列对比图

Fig. 1 The sequence alignments and domain arrangement of DNA Ligase III

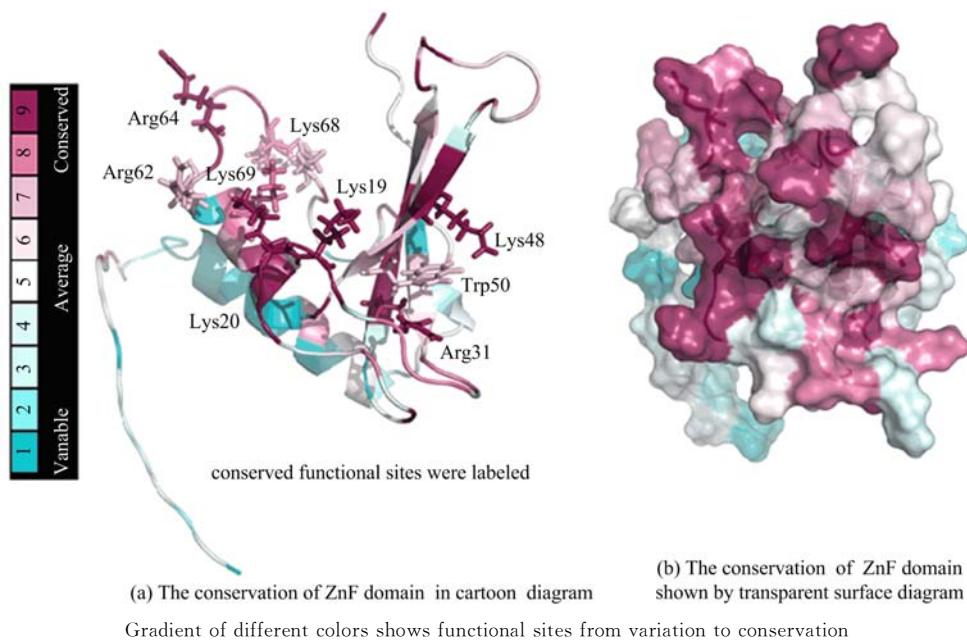


图 2 ZnF 结构域结构保守性

Fig. 2 The conservation of ZnF domain

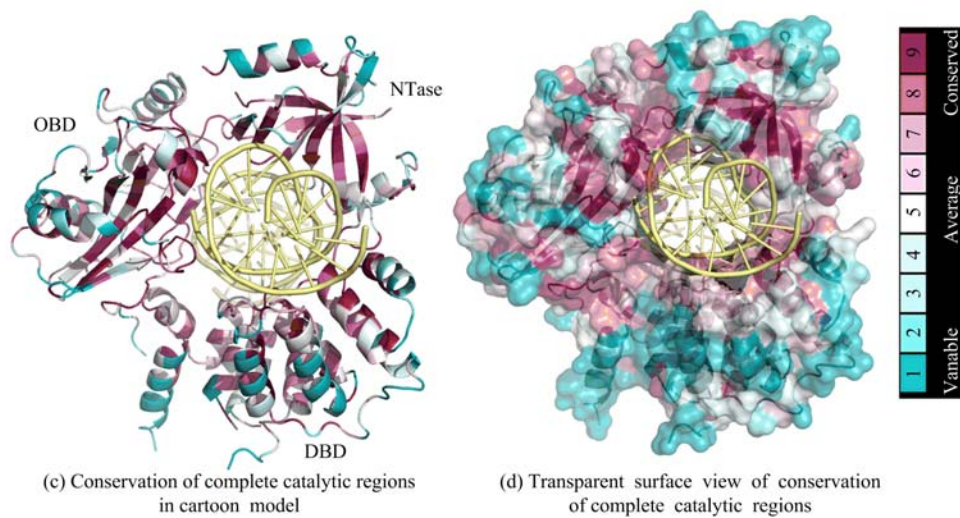
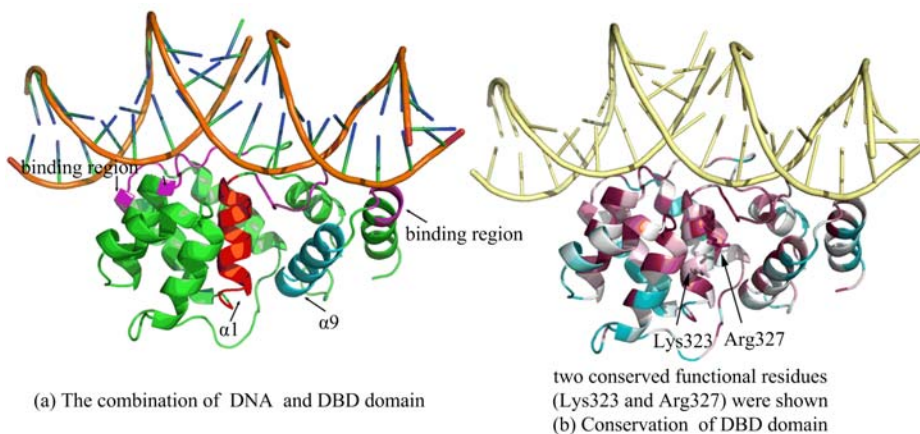


图 3 DBD, NTase 和 OBD 结构域的结构保守性

Fig. 3 The conservation of DBD, NTase and OBD domain

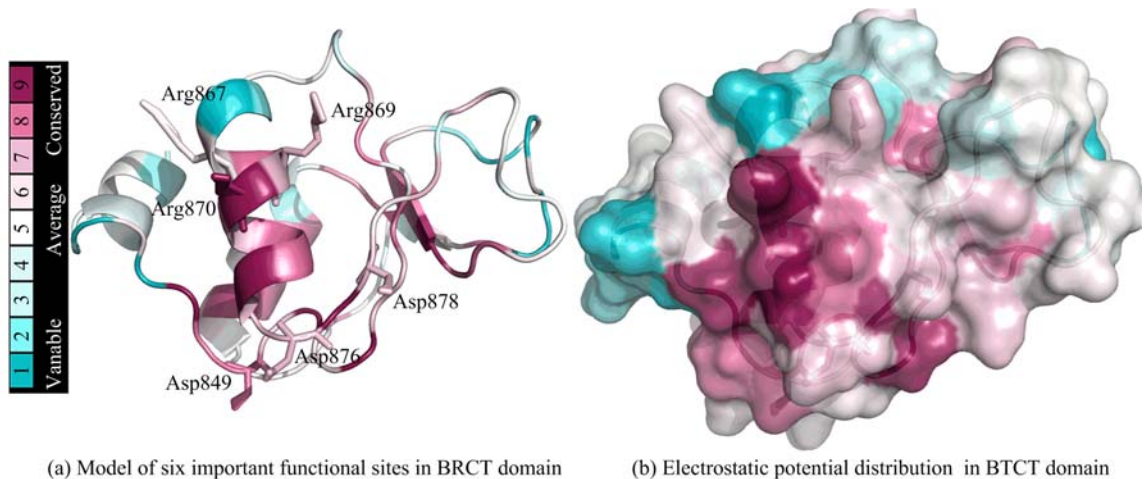


图 4 卡通模式和表面模式显示的 BRCT 结构域结构保守性

Fig. 4 Transparent surface and cartoon diagram of BRCT domain

演化出了 ZnF 结构域;在动物界中,ZnF 结构域和 BRCT 结构域发生了多次种系特异的丢失.在两栖纲和哺乳纲动物中,DNA 连接酶 III 还演化出不同数量的选择性剪接体.

真核生物的 3 个 DNA 连接酶的功能演化密切相关.在真核生物演化出 DNA 连接酶 III 之前,真核生物 DNA 复制、修复和重组中 DNA 末端的连接由 DNA 连接酶 I 和 DNA 连接酶 IV 完成,而 DNA 连接酶 III 功能的演化一方面取代了这两个连接酶的部分功能<sup>[2]</sup>,另一方面形成了这两个连接酶功能过程的备用过程<sup>[27]</sup>.在 DNA 连接酶 III 的功能演化过程中,MLS,ZnF 结构域和 BRCT 结构域起着关键作用.MLS 的演化使得 DNA 连接酶 III 可以参与线粒体 DNA 的复制和修复过程.最近研究认为,DNA 连接酶 III 的主要功能是线粒体 DNA 的复制和修复,而细胞核 DNA 的修复是其次要的功能<sup>[28-29]</sup>.但本研究发现,很多物种的 DNA 连接酶 III 不含 MLS,因此,它们不太可能在线粒体中发挥作用,而 DNA 连接酶 III 的主要功能是线粒体 DNA 的复制和修复的观点,可能只适用于哺乳纲、两栖纲、三刺鱼、果蝇属、领鞭毛虫等 DNA 连接酶 III 含有 MLS 且其 DNA 连接酶 I 和 IV 不含 MLS 的物种.ZnF 结构域和 BRCT 结构域可能只与 DNA 连接酶 III 在细胞核的功能演化有关,因为它们并不参与 DNA 连接酶 III 在线粒体的功能<sup>[28]</sup>.ZnF 结构域可以提高 DNA 连接酶 III 探测和连接 DNA 末端的效率,这种高效性对于基因组复杂度较高和 DNA 修复效率要求较高的物种尤为重要<sup>[30]</sup>,这可以解释为什么 ZnF 结构域的丢失只在基因组复杂度较低的物种发生.BRCT

结构域与 XRCC1 的相互作用使得 DNA 连接酶 III 可以定位到细胞核 DNA 的碱基切除修复和单链断裂修复等过程,从而部分取代了 DNA 连接酶 I 在这些修复过程中的功能.然而 BRCT 结构域在一些物种的 DNA 连接酶 III 中丢失或与 XRCC1 相互作用的位点丢失,说明 DNA 连接酶 III 可能还有 XRCC1 功能之外的其他功能,而与 XRCC1 相互作用可能阻碍其发挥作用,因为 DNA 连接酶 III 和 XRCC1 总是以复合物的形式存在<sup>[31]</sup>.

综上所述,本研究对 DNA 连接酶 III 的分布和功能演化进行了探讨,对进一步研究无脊椎动物 DNA 连接酶 III 的功能提供了依据.本研究还提示,在以无脊椎动物作为模式动物研究线粒体和细胞核 DNA 的复制和修复机制时,还应该考虑 DNA 连接酶 III 对实验结果的影响.

#### 参考文献 (References)

- [1] Lehman I R. DNA ligase: Structure, mechanism, and function[J]. *Science*, 1974, 186(4166): 790-797.
- [2] Ellenberger T, Tomkinson A E. Eukaryotic DNA ligases: Structural and functional insights[J]. *Annu Rev Biochem*, 2008, 77: 313-338.
- [3] Hartwell L H. *Saccharomyces-cerevisiae* cell-cycle[J]. *Bacteriological Reviews*, 1974, 38(2): 164-198.
- [4] Willer M, Rainey M, Pullen T, et al. The yeast CDC9 gene encodes both a nuclear and a mitochondrial form of DNA ligase I[J]. *Current Biology*, 1999, 9(19): 1 085-1 094.
- [5] Frank K M, Sekiguchi J M, Seidl K J, et al. Late embryonic lethality and impaired V(D)J recombination



- in mice lacking DNA ligase IV [J]. *Nature*, 1998, 396(6707): 173-177.
- [6] Chen L, Trujillo K, Sung P, et al. Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase [J]. *J Biol Chem*, 2000, 275(34): 26 196-26 205.
- [7] Bryans M, Valenzano M C, Stamato T D. Absence of DNA ligase IV protein in XR-1 cells: Evidence for stabilization by XRCC4 [J]. *Mutat Res*, 1999, 433(1): 53-58.
- [8] Frosina G, Fortini P, Rossi O, et al. Two pathways for base excision repair in mammalian cells [J]. *J Biol Chem*, 1996, 271(16): 9 573-9 578.
- [9] Dulic A, Bates P A, Zhang X, et al. BRCT domain interactions in the heterodimeric DNA repair protein XRCC1-DNA ligase III [J]. *Biochemistry*, 2001, 40(20): 5 906-5 913.
- [10] Okano S, Lan L, Caldecott K W, et al. Spatial and temporal cellular responses to single-strand breaks in human cells [J]. *Mol Cell Biol*, 2003, 23(11): 3 974-3 981.
- [11] Okano S, Lan L, Tomkinson A E, et al. Translocation of XRCC1 and DNA ligase III  $\alpha$  from centrosomes to chromosomes in response to DNA damage in mitotic human cells [J]. *Nucleic Acids Res*, 2005, 33(1): 422-429.
- [12] Lakshminpathy U, Campbell C. The human DNA ligase III gene encodes nuclear and mitochondrial proteins [J]. *Mol Cell Biol*, 1999, 19(5): 3 869-3 876.
- [13] Perez-Jannotti R M, Klein S M, Bogenhagen D F. Two forms of mitochondrial DNA ligase III are produced in *Xenopus laevis* oocytes [J]. *J Biol Chem*, 2001, 276(52): 48 978-48 987.
- [14] Cotner-Gohara E, Kim I K, Hammel M, et al. Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states [J]. *Biochemistry*, 2010, 49(29): 6 165-6 176.
- [15] Yutin N, Koonin E V. Evolution of DNA ligases of nucleocytoplasmic large DNA viruses of eukaryotes: A case of hidden complexity [J]. *Biol Direct*, 2009, 4: 51.
- [16] Altschul S F, Madden T L, Schaffer A A, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs [J]. *Nucleic Acids Res*, 1997, 25(17): 3 389-3 402.
- [17] Tatusov R L, Koonin E V, Lipman D J. A genomic perspective on protein families [J]. *Science*, 1997, 278(5338): 631-637.
- [18] Finn R D, Mistry J, Tate J, et al. The Pfam protein families database [J]. *Nucleic Acids Res*, 2010, 38 (suppl 1): D211-222; doi: 10.1093/nar/gkp985.
- [19] Katoh K, Misawa K, Kuma K, et al. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform [J]. *Nucleic Acids Res*, 2002, 30(14): 3 059-3 066.
- [20] Claros M G, Vincens P. Computational method to predict mitochondrially imported proteins and their targeting sequences [J]. *Eur J Biochem*, 1996, 241(3): 779-786.
- [21] Nielsen H, Engelbrecht J, Brunak S, et al. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites [J]. *Protein Engineering*, 1997, 10(1): 1-6.
- [22] Landau M, Mayrose I, Rosenberg Y, et al. ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures [J]. *Nucleic Acids Res*, 2005, 33(suppl 2): W299-302; doi: 10.1093/nar/gki370.
- [23] Gaston D, Tsaousis A D, Roger A J. Predicting proteomes of mitochondria and related organelles from genomic and expressed sequence tag data [J]. *Methods Enzymol*, 2009, 457: 21-47.
- [24] Claros M G, Vincens P. Computational method to predict mitochondrially imported proteins and their targeting sequences [J]. *Eur J Biochem*, 1996, 241(3): 779-786.
- [25] Kulczyk A W, Yang J C, Neuhaus D. Solution structure and DNA binding of the zinc-finger domain from DNA ligase III  $\alpha$  [J]. *J Mol Biol*, 2004, 341(3): 723-738.
- [26] Yutin N, Wolf Y I, Raouf D, et al. Eukaryotic large nucleocytoplasmic DNA viruses: Clusters of orthologous genes and reconstruction of viral genome evolution [J]. *Virology*, 2009, 6: 223.
- [27] Wang H C, Rosidi B, Perrault R, et al. DNA ligase III as a candidate component of backup pathways of nonhomologous end joining [J]. *Cancer Research*, 2005, 65(10): 4 020-4 030.
- [28] Simsek D, Furda A, Gao Y, et al. Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair [J]. *Nature*, 2011, 471 (7337): 245-248.
- [29] Gao Y K, Katyal S, Lee Y S, et al. DNA ligase III is critical for mtDNA integrity but not Xrcc1-mediated nuclear DNA repair [J]. *Nature*, 2011, 471 (7337): 240-244.
- [30] Aravind L, Walker D R, Koonin E V. Conserved domains in DNA repair proteins and evolution of repair systems [J]. *Nucleic Acids Res*, 1999, 27(5): 1 223-1 242.
- [31] Luo H, Chan D W, Yang T, et al. A new XRCC1-Containing complex and its role in cellular survival of methyl methanesulfonate treatment [J]. *Molecular and Cellular Biology*, 2004, 24(19): 8 356-8 365.