

论著·基础研究

3种灵长类动物 SAGE1 与 INTS3 相互作用保守性的验证

李晓畅, 李明月, 雷鸣, 武健

上海交通大学医学院附属第九人民医院上海精准医学研究院, 上海 200125

[摘要] **目的**·研究恒河猴 (Rhesus)、狨猴 (Marmoset)、倭狐猴 (Mouse Lemur) 这3种代表性灵长类动物的癌-睾丸抗原 (cancer-testis antigen, CTA) 家族成员之一的肉瘤抗原1 (sarcoma antigen 1, SAGE1) 的进化, 及其与整合体复合物 (integrator complex, INT) 亚基 INTS3 之间的相互作用的保守性。**方法**·首先在 HEK293T 细胞中利用免疫共沉淀 (co-immunoprecipitation, Co-IP) 的方法验证3种灵长类动物的 SAGE1 和 INTS3 相互作用。随后在互作蛋白截短的片段上分别添加 His-SUMO 和 GST 标签, 并在大肠埃希菌中进行表达, 经过系列纯化步骤得到目标蛋白。通过表面等离子体共振 (surface plasmon resonance, SPR) 实验测定目标蛋白之间的结合强弱, 借助广角激光散射凝胶排阻层析 (size exclusion chromatography with multi-angle light scattering, SEC-MALS) 技术检测复合物的结合比例, 并利用分子对接技术预测 SAGE1 和 INTS3 截短体的结合模型。分析人类 INTS3 和 SAGE1 互作界面, 挑选在灵长类中也相对保守的关键互作氨基酸位点, 将其突变后再次使用 Co-IP 分析 INTS3 和 SAGE1 的结合情况。**结果**·通过 Co-IP 方法, 确定3种灵长类动物中 INTS3 和 SAGE1 全长蛋白之间存在相互作用。截短体蛋白经过外源诱导表达和多步纯化, 获得高纯度蛋白样品; 经过 SPR 对其进行结合和解离检测分析, 得到了它们的解离常数在微摩尔级别; 并且 SEC-MALS 结果显示 INTS3 与 SAGE1 的结合比例为 2:1。使用分子对接预测 SAGE1 和 INTS3 截短体结合模型, 发现该复合物具有典型的“蝴蝶型结构”, 由 INTS3 二聚体组成“蝴蝶身体”, 由 SAGE1 组成“蝴蝶头部”, 结合界面中疏水相互作用构成了主要的互作模式。分析突变后的 SAGE1 与 INTS3 全长蛋白相互作用, 它们之间的结合明显减弱。**结论**·恒河猴、狨猴、倭狐猴中 INTS3 与 SAGE1 存在相互作用, 并且相互作用具有高度的保守性。

[关键词] 肉瘤抗原1; 整合体复合物亚基3; 相互作用**[DOI]** 10.3969/j.issn.1674-8115.2024.04.004 **[中图分类号]** R34 **[文献标志码]** A

Conservation of protein interaction between SAGE1 and INTS3 identified in 3 different types of primates

LI Xiaochang, LI Mingyue, LEI Ming, WU Jian

Shanghai Institute of Precision Medicine, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200125, China

[Abstract] **Objective**·To investigate the evolution of sarcoma antigen 1 (SAGE1), a member of the cancer-testis antigen (CTA) family, in three representative primate species—*Macaca mulatta* (Rhesus), *Callithrix jacchus* (Marmoset), and *Microcebus murinus* (Mouse Lemur), as well as the conservation of its interaction with INTS3, a subunit of the integrator complex. **Methods**·The interaction between INTS3 and SAGE1 in these primates and the interaction between INTS3 and human SAGE1 mutants were first validated in HEK293T cells by using co-immunoprecipitation (Co-IP). Truncated proteins were then tagged with His-SUMO and GST, respectively, and expressed in *Escherichia coli*, followed by a series of purification steps to obtain the target proteins. Surface plasmon resonance (SPR) was used to measure the binding affinity of the target proteins, and size exclusion chromatography with multi-angle light scattering (SEC-MALS) was used to determine the stoichiometry of the complex. Additionally, molecular docking was employed to predict the binding model of the truncated SAGE1 and INTS3. Mutations were performed on human SAGE1 key interface residues to analyze their binding to INTS3 by Co-IP. **Results**·The interaction between the full-length human-derived INTS3 and the full-length SAGE1 from the three primate species was confirmed by Co-IP. Truncated proteins were purified through multiple steps of tandom chromatography. The dissociation constants of the three pairs of truncated INTS3-SAGE1 were measured via SPR, and the SEC-MALS results demonstrated that the binding ratio of INTS3-SAGE1 was 2 : 1. Furthermore, molecular

[基金项目] 国家自然科学基金 (32071189)。**[作者简介]** 李晓畅 (1997—), 男, 硕士生; 电子信箱: lxc2020@sjtu.edu.cn。**[通信作者]** 武健, 电子信箱: wujian@shsmu.edu.cn。**[Funding Information]** National Natural Science Foundation of China (32071189)。**[Corresponding Author]** WU Jian, E-mail: wujian@shsmu.edu.cn。

docking predicted a structural model of the truncated INTS3-SAGE1 complex and the binding interface was extensively constituted with hydrophobic contacts assisted by some hydrophilic interactions. The interaction between the mutant SAGE1 and INTS3 full-length protein was significantly weakened. **Conclusion** There is a conserved interaction between INTS3 and SAGE1 across Rhesus, Marmoset, and Mouse lemur.

[Key words] sarcoma antigen 1 (SAGE1); integrator complex subunit 3 (INTS3); interaction

癌-睾丸抗原 (cancer-testis antigen, CTA) 是一类在恶性肿瘤和睾丸组织中高表达, 而在其他正常组织中不表达的特定蛋白, 通常被视为肿瘤生物标志物和免疫治疗靶点^[1]。迄今已有数百个 CTA 被鉴定报道, 并且大部分 CTA 以扩增子 (amplicon) 的形式分布于多个染色体的不同区域。肉瘤抗原 1 (sarcoma antigen 1, SAGE1) 是 CTA 的重要组成部分之一, 位于 X 染色体长臂 (Xq26.3) 区域, 与紧邻的另一 CTA 组分——CT45A (CTA family 45 member A) 家族串联基因阵列一起, 共同构成了 CT45A-SAGE1 扩增子结构。该扩增子结构的两侧被 2 个单拷贝基因 INTS6L 和 MMGT1 所包围隔离。从进化角度而言, CT45A-SAGE1 扩增子的祖先基因肇始于胎盘类哺乳动物, 然而 SAGE1 基因被发现仅仅存在于灵长类动物的基因组中, 说明 CT45A-SAGE1 扩增子在生物进化上属于一个非常新的基因家族。从表达模式而言, 生理条件下 SAGE1 仅在睾丸和垂体等少数组织中高表达, 在其他组织器官中不表达或者痕量表达, 然而在诸如肺癌、膀胱癌以及头颈癌等多种恶性肿瘤中 SAGE1 均被发现过量表达^[2]。值得注意的是, 尽管有报道明确表明 SAGE1 的异常激活与肿瘤进展、侵袭性和不良预后之间存在密切关联, 但是目前对于 SAGE1 的生理功能、结构基础、相互作用网络等等分子机制方面的认知甚少, 有待进一步深入的挖掘和系统性的研究^[3]。

我们使用 InterProScan (<http://www.ebi.ac.uk/interpro/>) 对人源 SAGE1 的结构域进行分析, 发现 SAGE1 的氨基末端结构域 (SAGE1_{NTD}) 含有 14 个短串联重复序列, 每个重复单元大约包含 47 个氨基酸, 该串联重复序列的功能迄今未见报道; SAGE1 的羧基末端结构域 (SAGE1_{CTD}) 是由 4 个 α -螺旋构成的 4 螺旋束结构, 具有潜在的核酸结合功能和 RNA 解旋酶活性^[4]。有研究^[5]表明部分 RNA 解旋酶如 DDX21 等能够与 RNA 聚合酶 II (RNA polymerase II, Pol II) 直接相互作用调控基因转录。然而 SAGE1 是否通过与 Pol II 或其他转录因子相互作用,

参与特定基因的转录调控还需要进一步明确。

整合体复合物 (integrator complex, INT) 是一种由 14 个亚基组成的大分子复合物, 能够结合 Pol II, 参与 mRNA 剪接、3'端加工和转录终止等过程^[6-7]。有报道^[8]明确表明, 该复合物中 INTS6 亚基的羧基末端结构域 (INTS6_{CTD}) 与 INTS3 亚基的羧基末端结构域 (INTS3_{CTD}) 之间存在直接相互作用。同时也有研究^[9-10]报道, INTS6 可以和 INTS3、单链 DNA 结合蛋白 1 (human single stranded DNA binding protein 1, hSSB1) 相互结合, 形成 INTS6-INTS3-hSSB1 亚复合物, 参与 DNA 的损伤修复。有趣的是, 我们发现 INTS6_{CTD} 与 SAGE1_{CTD} 之间具有较高的序列同源性, 基于 INTS6_{CTD} 与 INTS3_{CTD} 之间的结合模式信息, 我们合理地推测 SAGE1 与 INTS3 之间极有可能也存在直接相互作用, 并且该相互作用在灵长类动物中具有普遍性。

相较于 INT 复合物这种在多细胞真核生物中古老又保守的基因家族而言, SAGE1 是灵长类动物特有的、具有生物进化价值的一类特殊基因。在癌症研究中, 灵长类动物由于其与人类在进化上的相似性^[11-13], 能够更为准确地模拟人类疾病的生物学特性^[14-15]。因此通过对不同灵长类动物 SAGE1 蛋白的研究, 可以进一步了解 SAGE1 的进化及其生理病理意义。进化上灵长类动物大致可以分为: 猿类 (人和黑猩猩等为代表)、旧世界猴类 [恒河猴 (Rhesus) 和食蟹猴等为代表]、新世界猴类 [狨猴 (Marmoset) 和卷尾猴等为代表] 以及原猴类 (狐猴和眼镜猴等为代表) 四大类别。在我们实验室对人源 SAGE1 前期工作探索的基础上, 我们选择恒河猴、狨猴和倭狐猴 (Mouse Lemur, 下文简称 Lemur) 这 3 种代表性灵长类动物的 SAGE1 作为研究对象展开相关研究工作^[15-17]。通过构建哺乳动物细胞体外过表达质粒, 采用免疫共沉淀的形式确认了 3 种灵长类动物的 SAGE1 与 INTS3 具有相互作用。通过大肠埃希菌 (*Escherichia coli*, *E. coli*) 外源蛋白表达体系和蛋白纯化体系, 分别获得了 SAGE1 羧基末端截短体

蛋白 (SAGE1_{CTD})、INTS3 羧基末端截短体蛋白 (INTS3_{CTD}) 的生物样品, 然后借助表面等离子体共振技术分析了它们的结合动力学特性, 并使用广角激光散射测定了 INTS3_{CTD} 与 SAGE1_{CTD} 的结合比例。此外, 我们还利用分子对接工具, 预测了 INTS3_{CTD}-SAGE1_{CTD} 复合物的三维结构模型, 并将其与 INTS6 羧基末端截短体蛋白 (INTS6_{CTD}) 与 INTS3_{CTD} 形成的 INTS3_{CTD}-INTS6_{CTD} 复合物进行比较。我们的研究结果证明了灵长类动物中 INTS3 与 SAGE1 之间的相互作用具有极高的保守性, 提供了对 SAGE1 在灵长类动物中出现和进化的初步理解, 为后续 SAGE1 相关机制研究、完整的 INTS3 与 SAGE1 复合物的高分辨率结构的解析等奠定了坚实的理论基础。

1 材料与方法

1.1 实验材料

1.1.1 表达载体、菌株与细胞 pGEX-6P1 载体、pET-28a 载体、pcDNA3.4-3×FLAG-N 载体、pcDNA3.4-STREP-N 载体、*E. coli* Rosetta (DE3) 菌株、*E. coli* DH5 α 菌株、HEK293T 细胞均源自本实验室保存。

1.1.2 实验试剂与主要仪器 限制性核酸内切酶、T4 DNA 连接酶 (Thermo Fisher, 美国), 无缝克隆重组酶 (ClonExpress II, 南京诺唯赞), 位点特异性突变试剂盒 (QuikChange II, Agilent; 美国), 三羟甲基氨基甲烷-盐酸 (Tris-HCl)、氯化钠 (NaCl)、甘油、乙二胺四乙酸 (EDTA)、乙二醇双四乙酸 (EGTA)、磷酸缓冲液 (PBS)、吐温-20 (Tween-20)、聚乙二醇辛基苯基醚 (Triton-X-100)、二硫苏糖醇 (DTT)、4-羟乙基哌嗪乙磺酸 (HEPES)、苯甲基磺酰氟 (PMSF)、亮氨酸、苯甲脒、异丙基- β -D-硫代半乳糖苷 (IPTG)、氨基青霉素、卡那霉素、咪唑 [生工生物工程 (上海) 股份有限公司], Cocktail 蛋白酶抑制剂 (Roche, 瑞士), DMEM 培养基 (Thermo Fisher, 美国)、胎牛血清 (FBS, Gibco; 美国), 脂质体转染试剂 (Lipofectamine 3000, Thermo Fisher; 美国), 青霉素-链霉素双抗溶液 (Thermo Fisher, 美国), FLAG M2 beads (M8823, Sigma; 美国), STREP Tactin XT resin (IBA, 德国), LB 培养基 (青岛生工生物科技有限公司), 考马斯亮蓝 G250

(P0017, 碧云天), 镍离子金属螯合亲和层析介质 (Ni-NTA, Qiagen; 德国), 谷胱甘肽琼脂糖凝胶 (glutathione sepharose 4B, Cytiva; 美国); 而重组酵母 SUMO 蛋白酶 1 (ULP1-His) 以及 GST-3C 酶 (PreScission Protease) 均由本实验室表达、纯化、冻存。

冷冻离心机 (5814R, Eppendorf, 德国), 电泳仪与电泳槽 (Mini-PROTEAN, Bio-Rad; 美国), 超声波细胞破碎仪 (VCX750, SONICS, 美国), 高速离心机 (Avanti J-26XP, Beckman Coulter; 美国), 超声波清洗器 (SB-5200DTN, 宁波新芝生物), 控温振荡培养箱 (YR101, 上海西雷生物), 快速纯化液相色谱 (fast protein liquid chromatography, FPLC) 系统 (ÄKTATM pure)、层析柱 (SuperdexTM)、高通量分子互作分析系统 (Biacore 8K) (Cytiva, 美国), 静态广角激光散射仪 (DAWN HELEOS II, Wyatt; 美国) 为实验室保有。

1.2 实验方法

1.2.1 引物、DNA 合成和序列测定 均由生工生物工程 (上海) 股份有限公司完成。

1.2.2 质粒构建 委托生工生物工程 (上海) 股份有限公司进行 DNA 合成, 获得了恒河猴、狨猴、倭狐猴和人类 SAGE1 的全长基因编码序列。将这些序列分别插入 pcDNA3.4-3×FLAG-N 质粒, 得到全长 SAGE1 表达质粒, 使用位点突变试剂盒将人类 SAGE1 的关键作用位点 (M832、F838、Q840、R872、F873、K874) 突变为丙氨酸, 得到突变体 SAGE1 表达质粒; 再将恒河猴、狨猴、倭狐猴和人类 INTS3 的全长基因编码序列, 分别插入 pcDNA3.4-STREP-N 质粒, 得到 3 种灵长类动物的转染用全长 INTS3 表达质粒。随后, 利用 PCR 技术获得了 3 种灵长类动物 SAGE1 的 CTD, 并将 SAGE1 的 CTD 基因序列插入 pGEX-6P1 载体, 这使得 SAGE1 的氨基端带上谷胱甘肽巯基转移酶 (GST) 纯化标签, 得到恒河猴、狨猴和倭狐猴的氨基端带有谷胱甘肽巯基转移酶 (GST) 纯化标签的 SAGE1_{CTD} 表达质粒。借助 PCR 获得了 INTS3 的 CTD, 并将其插入 pET-28a 载体, 使得 INTS3 的氨基端带上 His-SUMO 纯化标签, 获得氨基端带有 His-SUMO 纯化标签的 INTS3_{CTD} 表达载体。

1.2.3 免疫共沉淀实验 本实验使用 HEK293T 作为

哺乳动物细胞表达的细胞系,将状态良好的细胞按25%的密度铺培养皿培养,37℃培养6~12 h之后观察细胞状态良好即可进行转染。转染前再次观察细胞状态(密度控制在70%~80%),使用Lipo3000进行质粒转染。48 h后收集转染好的细胞,加入400 μL细胞裂解液(20 mmol/L Tris-HCl pH 7.4, 150 mmol/L NaCl, 10%甘油, 0.5% Triton-X-100, 1 mmol/L EDTA, 1 mmol/L EGTA, 4 μL Cocktail)轻柔重悬细胞,使用SONICS超声波细胞破碎仪以30%的功率超声处理。超声处理后的样品,在4℃下,16 900 ×g离心15 min,取40 μL上清液保存,剩余上清液加入到预平衡好的FLAG M2 beads/Strep resin中,4℃转盘上孵育过夜。使用冷冻离心机300 ×g离心1 min,吸弃上清液,用裂解液重悬清洗3次,每次离心1 min,离心弃上清液。加入40 μL 2×SDS-PAGE loading buffer(125 mmol/L Tris-HCl pH 6.8, 10% SDS, 0.2%溴酚蓝, 25%甘油, 2.5% DTT),99℃加热5 min变性,样品冻存于-20℃。样品通过蛋白质印迹法(Western blotting)进行鉴定。

1.2.4 蛋白表达与纯化 将含表达目的蛋白的Rosetta菌液用3 200 ×g离心15 min,弃上清液,用裂解液(50 mmol/L HEPES, 400 mmol/L NaCl, 10%甘油, 0.1 mmol/L亮肽素, 1 mmol/L PMSF, 1 mmol/L苯甲脒, cocktail, pH 8.0)重悬菌体后,采用超声法裂解菌体。菌体裂解后经39 000×g离心40 min,取上清液。表达His标签的蛋白用Ni-NTA介质在4℃低速旋转混合1 h,表达GST标签的蛋白用GST琼脂糖凝胶珠混合3 h,分别用含有咪唑的洗涤缓冲液(50 mmol/L HEPES, 400 mmol/L NaCl, 25 mmol/L咪唑, 10%甘油, pH 8.0)和普通洗涤缓冲液(50 mmol/L HEPES, 400 mmol/L NaCl, 10%甘油, pH 8.0)洗去杂蛋白。再分别用含有咪唑的洗脱缓冲液(50 mmol/L HEPES, 150 mmol/L NaCl, 250 mmol/L咪唑, pH 8.0)和含有GSH的洗脱缓冲液(50 mmol/L HEPES, 150 mmol/L NaCl, 300 mmol/L GSH, pH 8.0)洗脱介质得到目的蛋白。将目的蛋白浓缩至10 mL,使用离子交换层析和凝胶排阻层析分开收集流出组分,经十二烷基磺酸钠-聚丙烯酰胺凝胶电泳(SDS-PAGE)分析鉴定目的蛋白质。

1.2.5 表面等离子体共振 首先,该实验通过氨基羧基偶联的方式将SAGE1_{CTD}蛋白偶联到仪器CM5芯

片上。将SAGE1_{CTD}蛋白溶解于仪器配套的不同pH的乙酸钠溶液中,进行pH筛选,以确定偶联最佳pH,确保最佳偶联效果。将SAGE1_{CTD}蛋白用流动相缓冲液(50 mmol/L HEPES, 150 mmol/L NaCl, pH 8.0)稀释10个梯度,使用配套甘氨酸作为再生液。选取数据最佳的几个梯度作为动力学分析,实验方法模板采用动力学法进行。动力学参数设定为:分析物流动相与固定相的接触时间为120 s,分离时间设定为180 s,流动相流速设置为30 μL/min。同时将梯度浓度输入到执行的实验循环信息中,实验结果将根据这些信息自动拟合结合解离曲线。该实验主要使用高通量分子互作分析系统Biacore 8K仪器来进行,数据分析采用Biacore insight evaluation软件。

1.2.6 广角激光散射 广角激光散射凝胶排阻层析(size-exclusion chromatography with multi-angle light scattering, SEC-MALS)是一种用于测量生物大分子的关键信息的技术,可以表征生物分子质量^[18]。制备INTS3_{CTD}-SAGE1_{CTD}复合物样品2 mg/mL,使用上样针将样品上样到机器中,以SEC-MALS缓冲液(50 mmol/L HEPES, 400 mmol/L NaCl, pH 8.0)作为流动相,样品在室温下以0.5 mL/min的流速在流动相中运行。使用ASTRA软件对SEC-MALS数据进行分析。

1.2.7 复合物结构预测与分析 RoseTTAFold Colab^[19]是基于Colab在线服务器搭建的蛋白质结构预测平台,虽然是云端计算平台但其预测的蛋白质结构依然能够达到原子水平的准确度^[20]。通过将MMseqs2的快速同源性搜索与RoseTTAFold相结合,提供了对蛋白质结构和复合物的加速预测^[19-21]。通过准备好云端环境和GPU等计算资源,提交INTS3_{CTD}、3种灵长类动物和人类的SAGE1_{CTD}氨基酸序列,即可获得相应的蛋白质三维结构预测数据。

ZDOCK server^[22]是一个在线网络交互式服务器,用于执行蛋白质-蛋白质分子对接,模拟蛋白质与蛋白质之间的相互作用,并生成可能的复合物结构。这有助于理解蛋白质复合物的构建方式,以及预测潜在的相互作用界面。上传已有RoseTTAFold预测出的SAGE1_{CTD}蛋白质结构文件与生物数据库已有的INTS3_{CTD}晶体结构文件(PDB id: 6WLG),设置好默认参数,ZDOCK会搜索和分析三维空间中INTS3_{CTD}和SAGE1_{CTD}之间所有可能的结合姿态,进行对接模拟。

2 结果

2.1 SAGE1 的进化及其与 INTS3 相互作用的保守性

借助同源搜索, 我们定义到 INTS6 是与 SAGE1 最接近的一个蛋白。为了探究 SAGE1 的进化, 我们选择了多种代表性物种并构建了 INTS6 的进化谱系树。分析结果表明, *INTS6* 是一个古老的基因, 它的起源可以追溯到多细胞真核生物的早期进化阶段。从七鳃鳗到人类, 脊椎动物的所有谱系都能观察到 INTS6 的存在 (图 1A)。我们得到一个线索, SAGE1 或许是灵长类动物 INTS6 演化的一个分支。提出这一假设是因为我们在 3 种不同的灵长类动物的 SAGE1 中都识别出了其独特的 CTD (图 1B), 而这一结构域在 INTS6 中也有所体现。氨基酸序列比对表明人类的 SAGE1_{CTD} 和 INTS6_{CTD} 在序列上呈现出部分保守性 (图 1C), 并且这一发现不仅只局限于人类, 进一步的序列比对揭示了人类与灵长类动物的 SAGE1_{CTD} 也具有相当高的保守性 (图 1D)。与此同时, 我们观察到 INTS3_{CTD} 在灵长类动物中也展现了高度的保守性 (图 1E)。已有的研究显示了 INTS6_{CTD} 与 INTS3_{CTD} 之间的相互作用能力。基于这些发现, 我们推测 SAGE1 极有可能与 INTS3 结合, 形成相应的蛋白质复合物。

2.2 3 种灵长类 SAGE1 与 INTS3 存在相互作用且作用力强度相似

为了验证灵长类动物 SAGE1 与 INTS3 之间存在潜在相互作用, 我们进行了免疫共沉淀分析。研究结果显示, 恒河猴的 INTS3 与其 SAGE1 存在相互作用 (图 2A)。该结果同样出现在狨猴 (图 2B) 和倭狐猴 (图 2C) 免疫共沉淀样品组中。为了研究不同灵长类动物中 INTS3_{CTD} 和 SAGE1_{CTD} 之间的相互作用强度以及结合或解离速度的差异, 我们纯化得到了 INTS3_{CTD} 蛋白质 (图 2D) 和 3 种灵长类动物的 SAGE1_{CTD} 蛋白质 (图 2E~G), 并采用了表面等离子体共振技术来评估这 2 种蛋白在体外相互作用的结合强度。通过动力学实验, 我们获得了 3 种灵长类动物 SAGE1_{CTD} 与 INTS3_{CTD} 之间的结合解离曲线 (图 2H~J), 根据动力学分析, 观察到不同浓度下的表面等离子体共振信号响应在一段时间内缓慢上升, 达到最大值后缓慢下降。这表明 SAGE1_{CTD} 与 INTS3_{CTD} 之间的相互作用呈现出慢结合和慢解离的趋势。为了进行比较, 我们选

择了基线稳定的空白组作为对照, 并选取了 5 个响应值较合理的浓度梯度进行分析。通过分析软件计算, 得到了恒河猴 SAGE1_{CTD} 与 INTS3_{CTD} 的解离常数为 2.26 $\mu\text{mol/L}$, 狨猴为 8.44 $\mu\text{mol/L}$, 倭狐猴为 1.95 $\mu\text{mol/L}$ 。3 种灵长类动物的平衡解离常数都在微摩尔数量级浮动, 表明它们之间的结合强度非常相似。

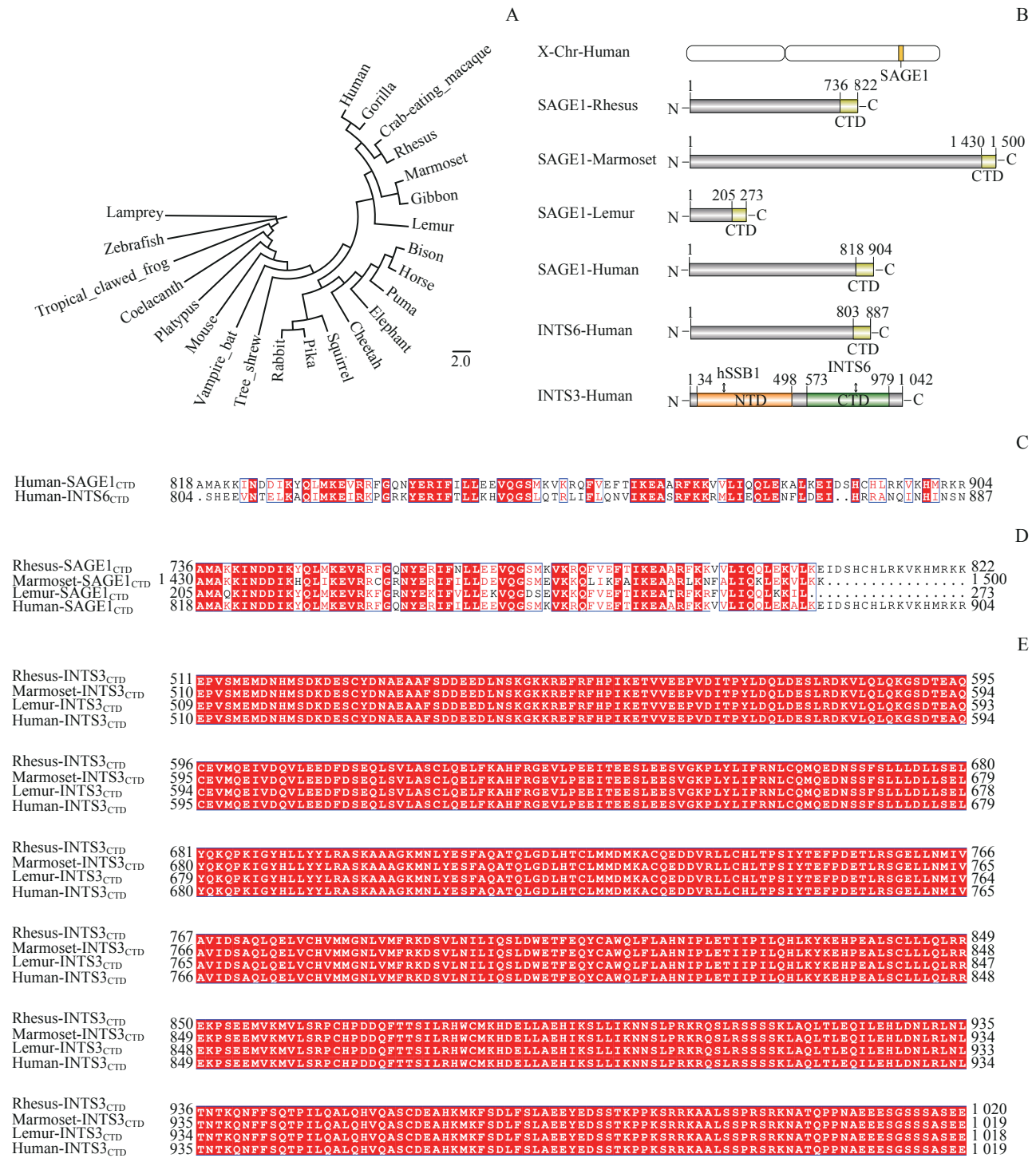
2.3 INTS3 与 SAGE1 的结合比例

为了探索 INTS3 与 SAGE1 的结合比例, 我们将纯化好的 INTS3_{CTD} 蛋白和 3 种灵长类动物的 SAGE1_{CTD} 混合孵育, 在体外组装成复合物。将复合物样品浓缩并使用凝胶排阻层析, 得到 2 个吸收峰 (图 3A~C 左)。SDS-PAGE 分析第一个吸收峰收集物 (图 3A~C 右) 结果显示, 在预期的位置出现了 INTS3_{CTD} 与对应灵长类动物的 SAGE1_{CTD} 的条带, 说明成功在体外组装得到复合物, 进一步证实了不同 SAGE1_{CTD} 与 INTS3_{CTD} 之间的相互作用。

通过 SEC-MALS 分析纯化的复合物 (图 3D~F), 结果显示 3 种灵长类动物均表现出清晰的信号, 并在 8 mL 流动相体积处产生吸收峰。同时, 紫外线检测显示了单峰信号的特征, 再次验证了 INTS3_{CTD}-SAGE1_{CTD} 复合物的形成。具体而言, 恒河猴 INTS3_{CTD}-SAGE1_{CTD} 复合物的分子质量测定为 103.0 kDa (单个 INTS3_{CTD} 理论分子质量为 47.1 kDa, 单个恒河猴 SAGE1_{CTD} 理论分子质量为 10.5 kDa); 而狨猴和倭狐猴的 INTS3_{CTD}-SAGE1_{CTD} 复合物分子质量都测定为 101.0 kDa (单个狨猴、倭狐猴 SAGE1_{CTD} 理论分子质量为 8.3 kDa), 所有检测结果均在可接受的误差范围内。该结果提示, INTS3 与 SAGE1 以 2 : 1 的比例形成复合物, 和 INTS3 与 INTS6 的作用模式相同。

2.4 INTS3_{CTD}-SAGE1_{CTD} 复合物初步结构分析

根据公共数据库中已有的 INTS3_{CTD}-INTS6_{CTD} 结构 (PDB id: 7BV7), 我们推测 3 种灵长类动物和人类的 INTS3_{CTD}-SAGE1_{CTD} 复合物也具有相似的结构。为了获取它们的三维结构信息, 我们借助蛋白质结构预测工具 RoseTTAFold2, 将 SAGE1_{CTD} 的氨基酸序列输入进云端结构模拟程序, 然后与 PDB 数据库中的 INTS3_{CTD} (PDB id: 6WLG) 进行分子对接。结果显示, 该复合物无一例外呈现出典型的蝴蝶形结构 (图 4),



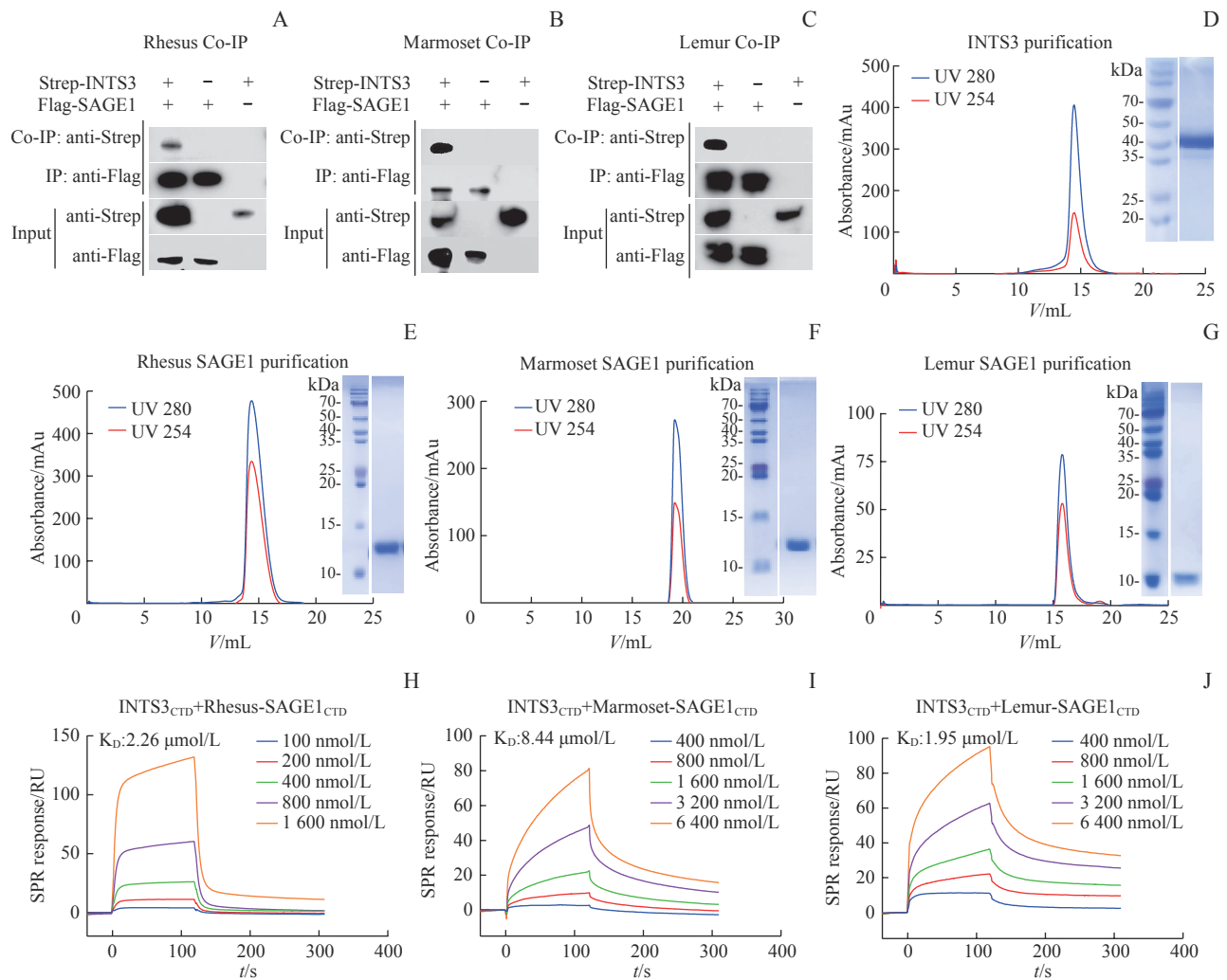
Note: A. Evolution tree of INTS6. B. The SAGE1 full-length sequence and C-terminal sequence domain pattern map of four species of human, rhesus, marmoset and mouse lemur, and the full-length domain pattern map of human INTS3 and INTS6. C. Sequence alignment between human SAGE1_{CTD} and INTS6_{CTD}. D. Sequence alignment between human SAGE1_{CTD} and three different primates SAGE1_{CTD}. E. Sequence alignment between human INTS3_{CTD} and three different primates INTS3_{CTD}.

图1 SAGE1 与 INTS6 在进化上的相似性

Fig 1 Evolutionary similarity between SAGE1 and INTS6

其中“蝴蝶”的头部由 SAGE1_{CTD} 的 4 个 α-螺旋组成，而“身体”则由 2 个 INTS3_{CTD} 蛋白相互交叉形成二聚体。同时，我们将其与 INTS3_{CTD}-INTS6_{CTD} 结构进行

了比较，发现对接结果与该结构高度相似（图 4A），SAGE1 与 INTS6 和 INTS3 的相互作用界面有一定重叠性。



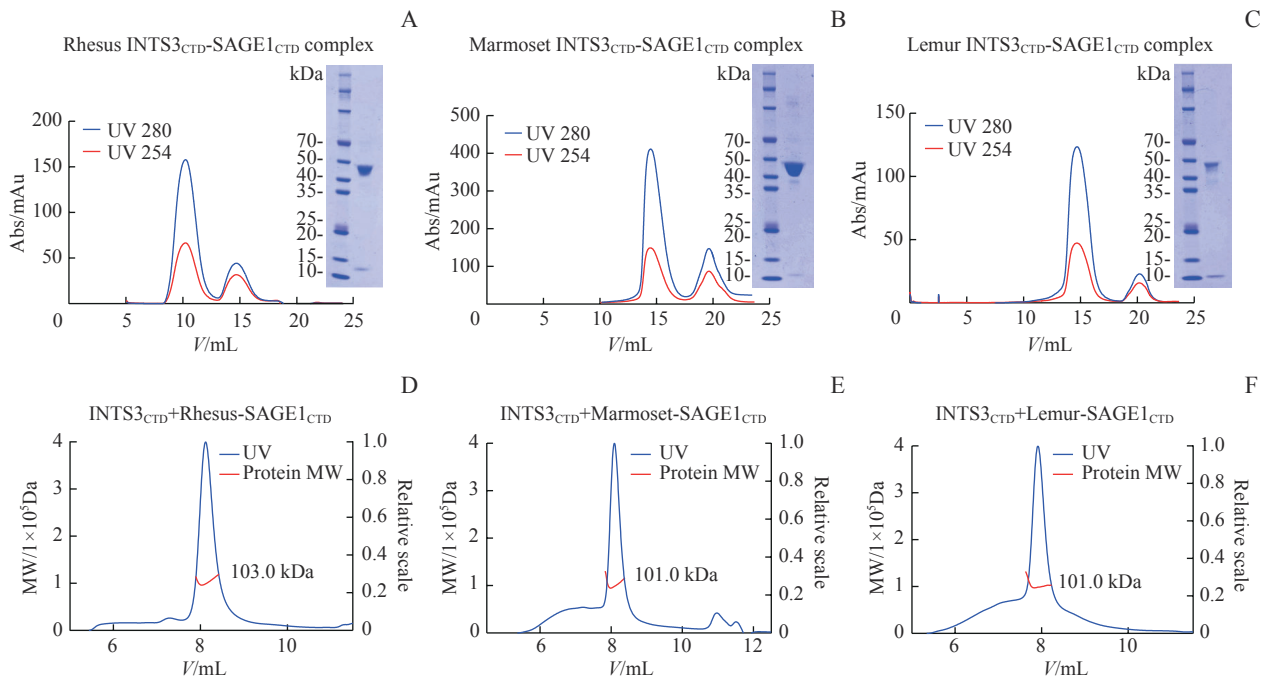
Note: A. Immunoprecipitation of full-length INTS3 and full-length SAGE1 in Rhesus. B. Immunoprecipitation of full-length INTS3 and full-length SAGE1 in Marmoset. C. Immunoprecipitation of full-length INTS3 and full-length SAGE1 in Lemur. D. Purification of human INTS3_{CTD}. E. Purification of Rhesus SAGE1_{CTD}. Gel filtration chromatography results shown on the left; SDS-PAGE analysis of fractions across the major peak shown on the right. F. Purification of Marmoset SAGE1_{CTD}. G. Purification of Lemur SAGE1_{CTD}. H. Rhesus SAGE1_{CTD} binding to INTS3_{CTD} in a concentration-dependent manner with a slow kinetics. I. Marmoset SAGE1_{CTD} binding to INTS3_{CTD} in a concentration-dependent manner with a slow kinetics. J. Lemur SAGE1_{CTD} binding to INTS3_{CTD} in a concentration-dependent manner with a slow kinetics. Co-IP—Co-immunoprecipitation; SPR—surface plasmon resonance.

图2 SAGE1与INTS3的免疫共沉淀实验、蛋白纯化及解离常数测定

Fig 2 Co-immunoprecipitation experiment, protein purification and dissociation constant determination of SAGE1 and INTS3

通过对人类INTS3_{CTD}-SAGE1_{CTD}复合物模拟结构分析,我们发现其相互作用主要由疏水相互作用介导,并伴有亲水相互作用辅助(图4B右)。随后,我们挑选了相互作用界面上6个关键互作位点(M832、F838、Q840、R872、F873、K874),将其全部突变为丙氨酸后,利用免疫共沉淀方法检测INTS3与突变体SAGE1的互作情况(图4C),结果显示,突变体SAGE1与INTS3的互作明显下降,这也验证了对接结果的可靠性。并且进一步观察分析后,我们发现这些位点大多在灵长类中具有保守性。3种灵长类动物INTS3_{CTD}和SAGE1_{CTD}的作用界面也主要由疏水相互作用介导,即恒河猴(图4D右)SAGE1_{CTD}的F791、狨猴(图4E右)

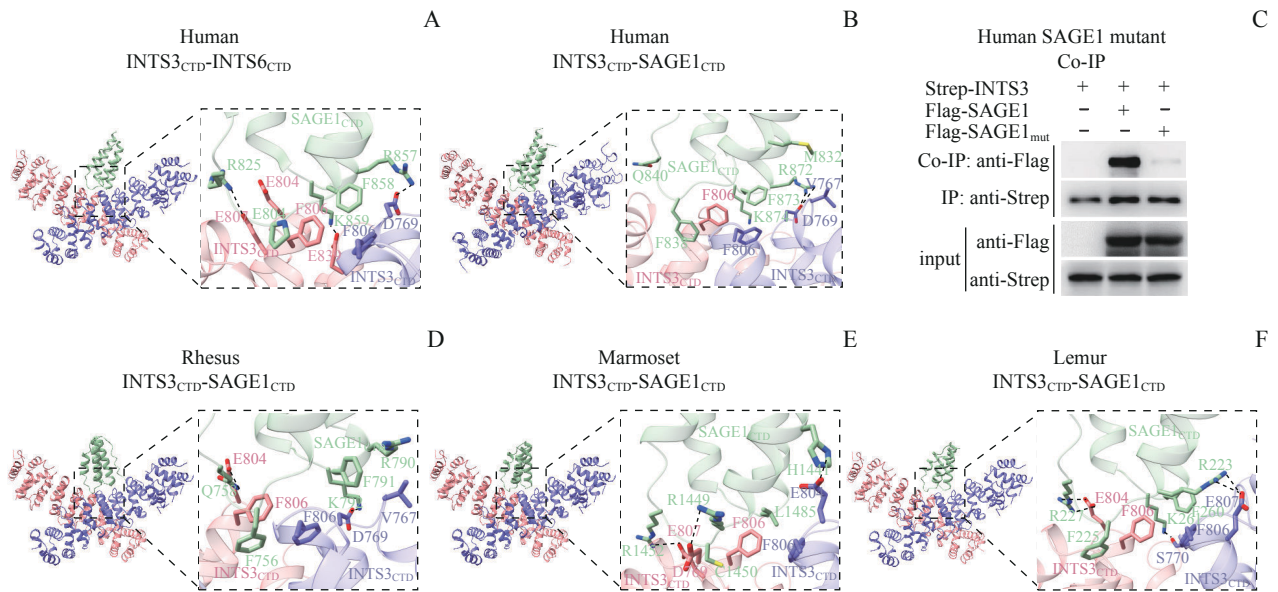
SAGE1_{CTD}的L1485和倭狐猴(图4F右)SAGE1_{CTD}的F260与2个INTS3_{CTD}的F806参与形成疏水作用。此外,还有一些氨基酸残基形成了离子键,如位于恒河猴SAGE1_{CTD}中的K792与INTS3_{CTD}的D769,位于狨猴SAGE1_{CTD}中的H1441与INTS3_{CTD}的E804,位于倭狐猴SAGE1_{CTD}中的R227与INTS3_{CTD}的E804,它们之间的相互作用被氢键进一步加强。在3种灵长类动物以及人类INTS3_{CTD}-SAGE1_{CTD}复合物结构上具有一定保守性,都呈现一种“蝴蝶”状结构,以疏水相互作用为主,辅助亲水相互作用,同时SAGE1与INTS6和INTS3的相互作用界面有一定重叠性,暗示了可能存在竞争性,需要未来进一步验证和探索。



Note: A. Rhesus INTS3_{CTD}-SAGE1_{CTD} complex assemble *in vitro*. B. Marmoset INTS3_{CTD}-SAGE1_{CTD} complex assemble *in vitro*. C. Lemur INTS3_{CTD}-SAGE1_{CTD} complex assemble *in vitro*. D. SEC-MALS analysis of purified Rhesus INTS3_{CTD}-SAGE1_{CTD} complex. E. SEC-MALS analysis of purified Marmoset INTS3_{CTD}-SAGE1_{CTD} complex. F. SEC-MALS analysis of purified Lemur INTS3_{CTD}-SAGE1_{CTD} complex. MW—molecular weight.

图3 INTS3_{CTD}-SAGE1_{CTD} 复合物体外组装与 SEC-MALS 分析

Fig 3 *In vitro* INTS3_{CTD}-SAGE1_{CTD} assembly and SEC-MALS analysis



Note: A. The Human INTS3_{CTD}-INTS6_{CTD} structure published on PDB (id: 7BV7) and its interaction interface. H-bond is marked by black dash line. B. The Human INTS3_{CTD}-SAGE1_{CTD} model created by molecular docking and its interaction interface. C. Immunoprecipitation of full-length INTS3 and full-length SAGE1 mutant in Human. D. The Rhesus INTS3_{CTD}-SAGE1_{CTD} model created by molecular docking and its interaction interface. E. The Marmoset INTS3_{CTD}-SAGE1_{CTD} model created by molecular docking and its interaction interface. F. The Lemur INTS3_{CTD}-SAGE1_{CTD} model created by molecular docking and its interaction interface.

图4 INTS3_{CTD}-SAGE1_{CTD} 结构预测与相互作用界面分析、INTS3 与人类突变体 SAGE1 免疫共沉淀实验

Fig 4 Structure prediction and interaction interface analysis of INTS3_{CTD}-SAGE1_{CTD}, and co-immunoprecipitation experiment of INTS3 with human mutant SAGE1

3 讨论

通过比较多种胎盘类哺乳动物的基因组数据, 研究人员证实了X染色体的保守性^[23-25]。但在X染色体的一些区域, 特别是在灵长类中, 一些主要在睾丸中表达的扩增子正在快速进化, *SAGE1*就是X染色体上的快速进化而来的一个基因^[25-27]。是什么进化事件促成了这些灵长类动物基因的快速扩增, 并最终形成我们现在所观察到的*SAGE1*, 很大程度上依然存在众多的未知因素。

本研究通过验证3种代表性灵长类动物(恒河猴、狨猴、倭狐猴)的*SAGE1*与*INTS3*的相互作用保守性, 来探讨*SAGE1*在灵长类动物的进化。*SAGE1*作为CTA家族的一部分, 已在多种癌症中显示出重要性^[28]。*SAGE1*与*INTS6*在C末端具有一定的序列同源性, 引发了我们的关注和猜想。前人已经证实, *INTS6*能与*INTS3*形成稳定复合物而发挥功能^[8]。我们推测*SAGE1*也可能与*INTS3*形成复合物, 但详细具体的机制并不清楚。

我们使用了几种蛋白互作方法来验证*SAGE1*与*INTS3*之间的相互作用。首先, 通过免疫共沉淀和蛋白质印迹的方法, 我们证实了全长*INTS3*与3种灵长类动物全长*SAGE1*之间存在明确的相互作用, 暗示了*INTS3*和*SAGE1*在基因表达和调控中的协同作用, 并且这种作用在进化过程中是高度保守的。通过使用表面等离子体共振技术, 得到*INTS3-SAGE1*的解离常数, 并且其处在微摩尔级别, 显示出较高的亲和性。这种级别的相互作用力可能在生物学上有其特定重要性, 需要进一步的实验来揭示。根据SEC-MALS的结果, *INTS3-SAGE1*复合物的组成比例为2:1。此外, 通过分子对接预测, 我们揭示了截短体*INTS3_{CTD}-SAGE1_{CTD}*复合物的蛋白质结构模型。复合物的“蝴蝶型结构”提供了一个全新的视角, 进一步解释了*INTS3*和*SAGE1*是如何进行结合的。此外, *INTS3*的二聚体形式可能对*SAGE1*的结合方式和其功能有重要的影响。

研究表明, *INTS3*具备与*SAGE1*和*INTS6*的相互作用能力, 并且它们的相互作用界面存在一定的重叠, 这暗示了*SAGE1*与*INTS6*之间可能存在潜在的

竞争关系。*INTS3*作为整合体复合物的组成部分, 在RNA的剪接、加工等生物过程中发挥关键作用。*SAGE1*与*INTS6*竞争性与*INTS3*结合可能影响转录功能的正常执行, 进而调节相关基因的表达。这一发现对未来的转录调控分子机制研究具有重要的意义。

针对*SAGE1*的探讨显示, 从低等灵长类到高等灵长类, 其C端具有保守性, 可能与*INTS6*竞争性结合*INTS3*, 而N端呈现出显著的变异, 通过模式图展示, 倭狐猴、狨猴、恒河猴、人类的*SAGE1*N端分别为204、1429、735、805个氨基酸。在漫长的进化过程中, N端长度一直在不断地“优化”, 最终塑造了人类*SAGE1*目前的N端长度, 从而更为有效地执行其功能。

综上所述, 本研究通过探索*SAGE1*与*INTS3*之间的相互作用, 提供了对*SAGE1*起源的新见解, 并为未来更深入地探讨它们的生物学功能打下了基础。为了更好地理解这两个蛋白质如何协同工作并参与到具体的生物过程中, 还需要更多的实验验证和分析。进一步的结构生物学研究, 可能为我们提供更多关于这两个蛋白质如何结合并发挥其功能的信息。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors disclose no relevant conflict of interests.

作者贡献/Authors' Contributions

武健设计并指导了全部课题的研究, 李晓畅完成了主要实验过程的实施、实验数据处理与文章撰写, 李明月和雷鸣负责实验方案指导和实验结果分析指导以及整篇论文的修改。所有作者均阅读并同意最终稿件的提交。

WU Jian designed and guided the whole research project. LI Xiaochang mainly completed the work of experiment execution, data analysis and processing and article writing. LI Mingyue and LEI Ming were mainly responsible for the guidance of experimental scheme and data analysis, as well as the revision of the whole paper. All authors have read the final manuscript and approved the submission.

- Received: 2023-12-26
- Accepted: 2024-02-09
- Published online: 2024-04-28

参·考·文·献

- [1] SALMANINEJAD A, ZAMANI M R, POURVAHEDI M, et al. Cancer/testis antigens: expression, regulation, tumor invasion, and use in immunotherapy of cancers[J]. *Immunol Invest*, 2016, 45(7): 619-640.
- [2] MARTELANGE V, DE SMET C, DE PLAEN E, et al. Identification on a human sarcoma of two new genes with tumor-specific expression[J]. *Cancer Res*, 2000, 60(14): 3848-3855.
- [3] SCANLAN M J, GURE A O, JUNGBLUTH A A, et al. Cancer/testis antigens: an expanding family of targets for cancer immunotherapy: Scanlan et al. Cancer/testis antigens[J]. *Immunol Rev*, 2002, 188(1): 22-32.
- [4] SHANG B, GAO A, PAN Y, et al. CT45A1 acts as a new proto-oncogene to trigger tumorigenesis and cancer metastasis[J]. *Cell Death Dis*, 2014, 5(6): e1285-e1285.
- [5] CALO E, FLYNN R A, MARTIN L, et al. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing[J]. *Nature*, 2015, 518(7538): 249-253.
- [6] HUANG K L, JEE D, STEIN C B, et al. Integrator recruits protein phosphatase 2A to prevent pause release and facilitate transcription termination[J]. *Mol Cell*, 2020, 80(2): 345-358. e9.
- [7] ZHENG H, QI Y L, HU S B, et al. Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase[J]. *Science*, 2020, 370(6520): eabb5872.
- [8] WELSH S A, GARDINI A. Genomic regulation of transcription and RNA processing by the multitasking Integrator complex[J]. *Nat Rev Mol Cell Biol*, 2023, 24(3): 204-220.
- [9] ZHANG F, MA T, YU X C. A core hSSB1-INTS complex participates in the DNA damage response[J]. *J Cell Sci*, 2013, 126 (Pt 21): 4850-4855.
- [10] KAPP L D, ABRAMS E W, MARLOW F L, et al. The integrator complex subunit 6 (Ints6) confines the dorsal organizer in vertebrate embryogenesis[J]. *PLoS Genet*, 2013, 9(10): e1003822.
- [11] GUO Y E, MANTEIGA J C, HENNINGER J E, et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates[J]. *Nature*, 2019, 572(7770): 543-548.
- [12] OZIRMAK LERMI N, GRAY S B, BOWEN C M, et al. Comparative molecular genomic analyses of a spontaneous rhesus macaque model of mismatch repair-deficient colorectal cancer[J]. *PLoS Genet*, 2022, 18(4): e1010163.
- [13] XIA H J, CHEN C S. Progress of non-human primate animal models of cancers[J]. *Dongwuxue Yanjiu*, 2011, 32(1): 70-80.
- [14] TARDIF S D, MANSFIELD K G, RATNAM R, et al. The marmoset as a model of aging and age-related diseases[J]. *ILAR J*, 2011, 52(1): 54-65.
- [15] PIFFERI F, EPELBAUM J, AUJARD F. Strengths and weaknesses of the gray mouse lemur (*Microcebus murinus*) as a model for the behavioral and psychological symptoms and neuropsychiatric symptoms of dementia[J]. *Front Pharmacol*, 2019, 10: 1291.
- [16] CARRION R Jr, PATTERSON J L. An animal model that reflects human disease: the common marmoset (*Callithrix jacchus*)[J]. *Curr Opin Virol*, 2012, 2(3): 357-362.
- [17] ESTES J D, WONG S W, BRENCHELEY J M. Nonhuman primate models of human viral infections[J]. *Nat Rev Immunol*, 2018, 18(6): 390-404.
- [18] SOME D, AMARTELY H, TSADOK A, et al. Characterization of proteins by size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS)[J]. *J Vis Exp*, 2019(148). DOI: 10.3791/59615.
- [19] BAEK M, DIMAIO F, ANISHCHENKO I, et al. Accurate prediction of protein structures and interactions using a three-track neural network[J]. *Science*, 2021, 373(6557): 871-876.
- [20] MIRDITA M, SCHÜTZE K, MORIWAKI Y, et al. ColabFold: making protein folding accessible to all[J]. *Nat Methods*, 2022, 19(6): 679-682.
- [21] AMOOZADEH S, JOHNSTON J, MEISRIMLER C N. exploiting structural modelling tools to explore host-translocated effector proteins[J]. *Int J Mol Sci*, 2021, 22(23): 12962.
- [22] 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.
- [23] LAHN B T, PAGE D C. Four evolutionary strata on the human X chromosome[J]. *Science*, 1999, 286(5441): 964-967.
- [24] GRAVES J A. Evolution of vertebrate sex chromosomes and dosage compensation[J]. *Nat Rev Genet*, 2016, 17(1): 33-46.
- [25] MUELLER J L, MAHADEVAIAH S K, PARK P J, et al. The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression[J]. *Nat Genet*, 2008, 40(6): 794-799.
- [26] MUELLER J L, SKALETSKY H, BROWN L G, et al. Independent specialization of the human and mouse X chromosomes for the male germ line[J]. *Nat Genet*, 2013, 45(9): 1083-1087.
- [27] BRASHEAR W A, BREDEMEYER K R, MURPHY W J. Genomic architecture constrained placental mammal X Chromosome evolution[J]. *Genome Res*, 2021, 31(8): 1353-1365.
- [28] ZHANG Y J, YU X H, LIU Q P, et al. SAGE1: a potential target antigen for lung cancer T-cell immunotherapy[J]. *Mol Cancer Ther*, 2021, 20(11): 2302-2313.

[本文编辑] 徐敏

