

文章编号: 1674-8115(2010)04-0408-07

## 人巨细胞病毒耐药相关蛋白 PUL97 局部结构模拟和突变分析

方凤琴<sup>1</sup>, 谢琼<sup>2</sup>, 张玥<sup>1</sup>, 毛客自<sup>1</sup>, 章莉<sup>1</sup>, 陆怡德<sup>1</sup>, 华丽<sup>1</sup>, 曹国君<sup>1</sup>, 季育华<sup>1</sup>

(1. 上海交通大学医学院瑞金医院检验科, 上海 200025; 2. 复旦大学药学院, 上海 200032)

**摘要:** **目的** 初步探索人巨细胞病毒耐药相关蛋白 PUL97 局部序列(329~572aa)的三维模拟结构并进行突变的耐药分析。**方法** 采用巢式 PCR 方法扩增 UL97 基因 776 bp 片段并测序, 以此发现非同义突变; 然后利用 SWISS-MODEL 结构模拟综合服务器和其他相关软件比对模式对野生型 PUL97 蛋白片段进行同源模建并优化, 通过各种指标或服务器来评估模拟结构的优劣并作筛选。将筛选所得的最佳模拟结构用于非同义新突变型的初步耐药分析。**结果** 最终筛选出的模拟结构是以 1YDTE 为模板, 以 L1.FASTA 为比对文件所得的 Model A。整个分子的能量最低, ProQ 的预测分数最高。分子对接分析也发现了潜在的 ATP 和 GCV 结合口袋。利用已知的阴阳性耐药点突变验证了该结构的合理性。同时根据突变前后结构信息的改变, 提示 C518Y 单点突变以及联合突变(E517G 和 C518Y) 呈现高度可疑的耐药特性。**结论** PUL97 野生型蛋白计算机结构模拟和突变分析法是初步评估新突变型耐药特性的又一方便快捷的方法, 可用于实验验证前的筛选; 但模拟结构本身的准确性是成功运用该方法的前提。

**关键词:** 人巨细胞病毒; PUL97 蛋白; 三维结构模拟; 突变分析

**中图分类号:** Q816; TP311.11

**文献标志码:** A

## Modelling of partial structure of human cytomegalovirus PUL97 protein and drug-resistant analysis of new non-synonymous mutations

FANG Feng-qin<sup>1</sup>, XIE Qiong<sup>2</sup>, ZHANG Yue<sup>1</sup>, MAO Ke-zi<sup>1</sup>, ZHANG Li<sup>1</sup>, LU Yi-de<sup>1</sup>, HUA Li<sup>1</sup>, CAO Guo-jun<sup>1</sup>,  
JI Yu-hua<sup>1</sup>

(1. Department of Clinical Laboratory, Ruijin Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai 200025, China; 2. College of Pharmacy, Fudan University, Shanghai 200032, China)

**Abstract: Objective** To preliminarily explore the 3-dimensional modelling of partial structure (329-572aa) of cytomegalovirus PUL97 protein and elementarily analyse the possibility of drug-resistance of new non-synonymous mutations. **Methods** Nested PCR was employed to amplify 776 bp fragment of UL97 gene. And some non-synonymous mutations in it were found by sequencing. After that, the online server of SWISS-MODEL and other related softwares were used for modelling and optimizing the partial structure of wildtype PUL97 protein based on the alignment mode. Then, the modelled structures were evaluated and screened by various classical indicators or online servers. Finally, the best structure was obtained and employed for preliminarily analysis of potential drug-resistant possibility of non-synonymous mutations. **Results** The finally-screened model was Model A based upon the template of 1YDTE and the alignment file of L1.FASTA. The whole energy of the structure was the lowest, while the prediction result of ProQ was on the top compared with the rest. Using the method of molecular docking, the potential binding sites of ATP and GCV were found in it. And most important, the mutation analysis based upon the modelled structure and positive/negative drug-resistant mutations confirmed that this structure was reasonable. Finally, the newly-found mutations (namely, point mutation C518Y and combined mutation E517G/C518Y) showed some evidences of having the possibility of drug-resistance under the benefit of the above-established system. Such evidences were the significant differences in structural information before and after mutagenesis. **Conclusion** The above-described system of PUL97. structure modelling and mutation analysis is the valuable and alternative

**基金项目:** 上海交通大学医学院 2007 级博士点基金(BXJ0813); 上海市科委重大基础研究项目(074119521)(2007 Doctorial Grant of Shanghai Jiaotong University School of Medicine, BXJ0813; Major Fundamental Research Program of Shanghai Science and Technology Committee, 074119521)。

**作者简介:** 方凤琴(1982—), 女, 博士生; 电子邮箱: fengqin999@126.com。

**通讯作者:** 季育华, 电子邮箱: rjjyh@yahoo.com.cn。

method for preliminarily predicting the possibility of drug-resistance of new mutations. It is more convenient and quick, and can be regarded as the screener before experimental verification. However, the accuracy of modelled structure itself is the big prerequisite for successful application of this system.

**Key words:** human cytomegalovirus; PUL97 protein; 3-dimensional structure modelling; mutation analysis

随着现代移植医学的迅猛发展和新型免疫抑制剂的不断推出,移植已经成为许多终末器官疾病以及血液病重要的治疗手段。感染和排斥一直是移植医学面临的两大问题,其中人巨细胞病毒(human cytomegalovirus, HCMV)内源性潜伏一激活感染的控制是重点之一。中国人群 HCMV 的感染率高达 90% 以上,健康带毒者无碍其生命及生活质量。但是,在免疫低下的群体中(如婴幼儿及越来越多的移植患者),特别是受到同种异体移植物的刺激时,病毒极易呈现激活感染,引发全身各脏器的病变以及移植物失功,甚至危及生命<sup>[1]</sup>。

考虑到移植群体持续的免疫抑制以及 HCMV 病所带来的严重后果,预防措施显得至关重要。目前,临床上主要采用两种方法:预防性用药和干预性用药<sup>[2]</sup>。不管是哪种措施,抗 HCMV 药物的长期暴露似乎不可避免。但由于这些药物的高昂价格和对骨髓抑制的不良反应,往往导致临床医师在用药时剂量偏小或者疗程不足。长期低剂量、间断性的暴露和机体持续免疫力低下使耐药株得以筛选和复制,最终演变成进行性或致死性 HCMV 病。据国外资料统计,在所有移植群体中, HCMV 耐药发生率高达 10% 左右<sup>[3]</sup>。

UL97 为 HCMV 最主要的耐药相关基因,其编码的蛋白 PUL97 为一种激酶,底物广泛,包括病毒或宿主编码的蛋白以及小分子药物,如更昔洛韦(ganciclovir, GCV)等。GCV 必须先经过 PUL97 蛋白的磷酸化,才能继续被细胞激酶磷酸化,最后才能抑制 DNA 聚合酶。若发生 GCV 耐药,90% 是 UL97 突变所致<sup>[3]</sup>。因此,UL97 基因是研究 GCV 耐药的重要靶点<sup>[3]</sup>。

中国人群 HCMV 感染有其自身特点,如潜伏感染率高,导致移植类型以 D + / R + 为主,缺乏耐药极其重要的易感因素(D + / R -)<sup>[2, 4-5]</sup>。尽管国际公认的有效抗 HCMV 药物种类有限,但是我国临床医师的用药剂量、疗程、时段均有自身的特点。根据目前的部分研究结果,我国移植群体 UL97 非同义突变率虽然与国外无异(排除已经证实的多态性位点),但其中近一半位点不同于国外报道,尤其是 C518Y 单突变及 E517G/C518Y 联合突变在人群中占一定比例。

验证新突变位点与耐药的相关性,最经典的方法就是分子转移病毒重组技术<sup>[6-7]</sup>。即利用病毒核

酸易发生同源重组的性质,将克隆的突变分子引入 HCMV AD169 株,并通过药物筛选平台,获得突变的重组 HCMV 并验证与耐药的相关性。事实上, HCMV 是人类最大的病毒之一(230 kb),顺利完成分子转移实验并获得预想的突变株,在技术或成本等方面都要求较高,而且费时费力。若用计算机技术对 HCMV 耐药相关蛋白 PUL97 三维结构进行模拟,并利用得到的模拟结构对非同义新突变前后的结构或功能改变进行分析,有助于对新突变型的耐药性作出初步预测。对预测结果提示有耐药可能性的突变型,再通过病毒重组实验验证,这样可节约大量的时间和成本,并给临床治疗一个早期提示。

蛋白质结构预测在医学或生物学各个领域有着成熟的应用经验<sup>[8]</sup>。2006 年,有研究<sup>[9]</sup>已经对 HCMV DNA 聚合酶(PUL54)三维结构进行了预测。因为在 PDB 数据库已经存有大量的聚合酶晶体结构,很容易就能找到与 HCMV DNA 聚合酶序列同源性高的模板;此时利用 SWISS MODEL 中 automated mode 模式就能预测出可靠的三维模拟结构,用于 UL54 基因新突变点的耐药性预测。而对于 PUL97 蛋白,虽然目前在 PDB 库中还没有同源性很高的模板,但是从 2003 年至今已经有人陆续注册 PUL97 局部的或粗略的模拟结构<sup>[10]</sup>。本研究就该野生型蛋白 244 个氨基酸(329 ~ 572aa)的三维结构作一初步探索,并以此为基础分析非同义新突变的耐药特性。

## 1 材料与方法

### 1.1 标本来源及处理

**1.1.1 标本来源:**为外周抗凝全血标本,参与 UL97 PCR 扩增的标本近 1 000 份,而能得到条带并测序的标本只有近 200 份。这些标本主要来自上海交通大学医学院附属瑞金医院造血干细胞移植或肾移植患者,少部分来自肝移植或心脏移植患者。这些患者多数采用系列跟踪方式,采血时间间隔 1 周~1 个月。干细胞输注前准备一般采用清髓方式,输注前 10 d 进行静脉注射 GCV 预防(250 mg, Bid),然后口服阿昔洛韦(aciclovir, ACV),若出现血浆 HCMV DNA 阳性,则改用缙阿昔洛韦(VACV)或静脉注射 GCV。免疫抑制方案一般采用环孢素(cyclosporin A, CsA) + 霉酚酸酯(mycophenolate mofetil, MMF; 商品

名为骁悉),若出现Ⅱ度及以上移植物抗宿主病(graft versus host disease, GVHD),主要用甲基强的松龙(methylprednisolone, MP)或强的松(prednisone, Pred)治疗。肾移植一般采用三联免疫抑制方案,即CsA + MMF + Pred或者他克莫司(FK506;商品名为普乐可复) + MMF + Pred,手术后一般采用ACV预防HCMV,若出现HCMV DNA阳性,则改用VACV或静脉滴注GCV治疗。

**1.1.2 标本处理:**全血标本量为2~3 mL,首先利用低速原理将0.8%氯化铵(NH<sub>4</sub>Cl)按(5~10):1比例与全血混合,彻底裂解红细胞,通过反复离心和PBS洗涤获得白细胞团。然后用传统的苯酚-氯仿抽提法抽提白细胞中的核酸(包括细胞和病毒基因组)备用。

**1.2 UL97基因扩增和测序** 针对含有大量白细胞的基因组背景以及该基因G+C比例偏高导致PCR相对难扩的特点,本实验室建立了巢式PCR方法。外围前引物1038F:5'-CTGCACAACGTCACGGTACATC-3',外围后引物1038R:5'-CTCCTCATCGTCGTCGTAGTC-3';内围前引物776F:5'-GTTGGCCGACGCTATCAAAT-3',内围后引物776R:5'-GGTCCTCCTCGCAGATTA TG-3'。两轮的PCR反应程序均为:95℃ 10 min;55℃ 30 s,72℃ 30 s,72℃ 70 s,共35个循环;最后72℃延伸7 min。第二轮PCR总体积为25 μL,其中3 μL模板来自第一轮PCR的反应产物。两轮后所得PCR片段长度为776 bp,在UL97基因中位置是1399~2068 nt(核苷酸),即467~689 aa(氨基酸)。该段序列涵盖了UL97基因的高突变区(包括有义突变和无义突变)<sup>[3]</sup>。测序由英骏公司(Invitrogen)完成,所用仪器为ABI 3730。每个标本正向(前引物)和反向(后引物)各测1次,确保测序无差错。同时每个标本均进行2次独立扩增,确保无污染。

**1.3 蛋白三维结构模拟的相关软件或在线服务器**

**1.3.1 Clustal X 1.83:**是一个Windows界面的多重序列比对软件(包括核苷酸序列和氨基酸序列)。但是导入的序列格式必须是FAS格式,而不能是SEQ或PRO格式。在DNA STAR中可进行格式间的转换。

**1.3.2 SWISS-MODEL<sup>[11]</sup>:**是三维结构模拟软件,同时也是一个免费的综合在线服务器,操作网址为<http://swissmodel.expasy.org/>,由瑞士生物信息学研究所研发。它主要根据序列决定构象、构象决定功能、同源序列有相似构象的原理进行同源建模。建模方法主要有三种:自动模式(automated mode)、比对模式(alignment mode)以及组合模式(project mode)。

自动模式只要提交待模拟序列即可,但是靶序列与模板序列之间的同源度要求最高。比对模式需要人为寻找合适的模板(可以为多个),然后建立靶序列与多个模板之间的比对文件,最后将比对文件、靶序列以及最优的一条模板序列共同提交给服务器;比对模式所要求的同源度可相对较低,但至少应有15%~20%。组合模式需要提交一条靶序列与一个或多个模板的重叠三维结构以及靶序列与模板序列之间的比对文件,主要用于结构建模后的优化。本研究主要采用的是比对模式。

**1.3.3 PDB数据库:**为蛋白质晶体结构数据库,任何模拟结构都不被收入。同源模板就是在PDB库中寻找。而SWISS-MODEL已经整合了自身的PDB库(ExPDB库),方便其直接寻找同源模板。

**1.3.4 模拟结构评估系统:**SWISS-MODEL本身带有三个模拟结构区段评估指标,即Anolea、Gromos和Verify3D。Anolea主要用于评估三维结构的折叠质量,y轴代表每个氨基酸的折叠能量,负值表示处在有利的能量环境中,正值表示处在不利的能量环境中。Gromos主要用于评估经验力场能,y轴代表蛋白链中每个氨基酸的经验力场能,负值和正值分别代表该氨基酸处于有利的和不利的能量环境中。Verify3D代表每个原子的三维结构与其所在的氨基酸序列(一维结构)之间的兼容性,分值为-1~+1,兼容性由差到好。此外,还有一个ProQ结构质量预测服务器,操作网址<http://www.sbc.su.se/~bjornw/ProQ/ProQ.html>,主要计算两个参数,即LGscore(结构正确>1.5;结构良好>3;结构优秀>5)和MaxSub(结构正确>0.1;结构良好>0.5;结构优秀>0.8),优势在于尽量找到合理且正确的结构,而不是尽量去找该蛋白的结构原形。

**1.3.5 DeepView(Swiss-PdbViewer 3.7):**是一种蛋白三维模拟结构的查看、调整和优化软件。由SWISS-MODEL网站免费提供下载。

**1.3.6 Autodock:**是一种最权威以及使用最广泛的分子对接软件,主要用于小分子(如药物)与大分子(如蛋白质)之间的对接,具有优化模拟结构的功能。该软件操作、分析由复旦大学药学院协助完成。

## 2 结果

**2.1 UL97基因序列中存在非同义新突变** 在PUL97蛋白520号氨基酸附近发现有义突变:518号位置的氨基酸从半胱氨酸(C)变为酪氨酸(Y),所占比例为1%~2%。此外,还发现517和518号位置氨

氨基酸有联合突变,前者从谷氨酸(E)变为甘氨酸(G),后者从半胱氨酸变为酪氨酸(即 517aa E-G和

518aa C-Y);所占比例约 1% (2/200)。但从未发现 517 位氨基酸单独突变。阳性测序结果见图 1。

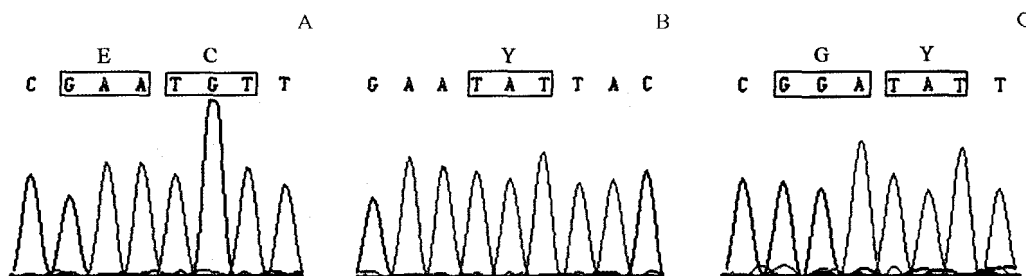


图 1 UL97 基因 517 和 518 号位置密码子的非同义突变

Fig 1 Non-synonymous mutations found in UL97 gene at the position of 517aa and 518aa

A: 野生型序列; B: UL97 基因 518 号位置密码子从 TGT(C 为半胱氨酸)变为 TAT(Y 为酪氨酸); C: 517 和 518 号位置密码子同时发生突变,前者从 GAA(E 为谷氨酸)变为 GGA(G 为甘氨酸),后者从 TGT 变为 TAT。

## 2.2 选择待模拟区段及寻找同源性较高的模板

### 2.2.1 分析并选择待模拟的区段:

通过对实验株 PUL97 蛋白的保守结构功能域<sup>[12]</sup>以及 PDB 数据库中现有的三维结构资源的深入分析,选用 329 ~ 572 位氨基酸序列作为模拟靶序列。这一区域不仅是 PUL97 蛋白重要的功能区域(该段序列涵盖了 PUL97 蛋白 I ~ IX 亚域)<sup>[12]</sup>,同时一些已验证的高频率、高耐药的突变型(如 M460V、H520Q)也落在此区域。此外,一些新发现的非同义点突变也落在该区域,位于 VII/VIII 亚域,该保守骨架在细胞激酶中缺失,被认为是 GCV 结合的可能部位<sup>[12]</sup>。

### 2.2.2 寻找同源性较高的模板:

直接利用 SWISS-MODEL 链接的 PDB 模板库,找到了一些同源性较好的模板。因为没有同源程度特别突出的模板,为了尽量不遗漏 PDB 库中已有的结构信息,以及最大程度地反映该蛋白真实结构,选择同源性较好的不同类型的多个模板,预备在结构建模时采用比对模式。根据软件给出的 e 值(e 值越小,同源性越好),各种模板蛋白的功能类别以及二级结构同源度等信息综合分析,最终选择了以下 12 条模板:2c47A(2c47 为 PDB 编号,A 为蛋白链编号)、2chlA、2cmwA、1ckiA、1vywC、1J91A、1z57A、2eu9A、1wakA、1YDTE、1TKIA 和 1TQIA。同源 e 值均在 e-09 至 e-11 之间,序列同源度为 10% ~ 20%。在具体比对时,可以只选择其中一部分,以最终得出的模拟结构优劣为标准。

## 2.3 同源比对及建模

利用 Clustal X 1.83 软件将靶序列与多个模板序列进行同源比对,然后将比对文件递交给 SWISS-MODEL 进行比对模式的同源建模。

### 2.3.1 同源比对是关键:

同源建模的准确性很大程度上依赖于比对结果,特别是当靶序列与模板序列

的同源度并不很高时。一般来说,最好的比对结果就是能找到最多的同源点。

### 2.3.2 比对模式的同源建模:

将所得的比对文件提交给 SWISS-MODEL 服务器,同时从中选择一条模板序列。该服务器会自行演算并优化,得出模拟结构。同一模板,不同的比对文件会得到不同的模拟结构;同一比对文件,不同的模板也会得到不同的结构。该服务器在给出“.pdb”结构数据的同时,还会给出 Anolea、Gromos 以及 Verify3D 等评估数据。初步筛选后,ProQ 服务器可作进一步的预测和淘汰。图 2 是经过筛选后得到的 3 个最优结构,模拟残基数为 244 个氨基酸,筛选体系为 SWISS-MODEL 自带的评估指标以及 ProQ 服务器。Model A 和 B 是根据同一个模板和不同比对文件得出的不同结构。这三个模拟结构的 ProQ 预测档位基本一致(LGscore 位于“良好”档,MaxSub 位于“正确”档),均属于可接受范围。

## 2.4 PUL97 模拟结构 Model A 与小分子 ATP/GCV 的对接

利用 Autodock 软件先优化上述三个模拟结构,然后采用分子对接技术寻找该蛋白结构中理论上应存在的 ATP 或者 GCV 结合口袋,作进一步筛选。已经证实 PUL97 为一种激酶蛋白,它不但能使各种病毒或宿主蛋白底物磷酸化,还能使本身以及 GCV 磷酸化。该蛋白一直被认为存在 ATP 以及 GCV 的结合口袋,而且这些结合区位于该模拟区段内的可能性很大<sup>[13]</sup>。对三个模拟结构优化后进行分子对接,发现只有 Model A 结构具有结合 ATP 或 GCV 小分子的较大可能性,其对接分数分别为 75 和 48(100 为满分)(图 3),而其余结构能同时具备两种口袋的可能性很小。对接能将口袋进行初步定位,但定位结果与理论上推测的结合位置有一定偏差。

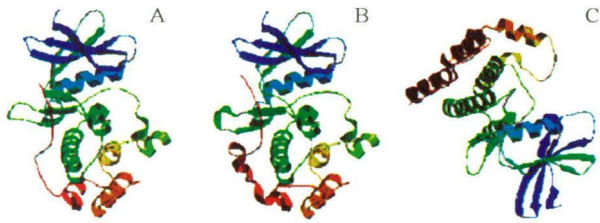


图2 从不同比对文件和不同模板所建模的一组结构中筛选出的3个最优结构

Fig 2 Three optimal modelling structures selected from a group of structures modelled on different alignments and templates

A: 以L1.FASTA为比对文件,以1YDTE为模板的建模结构(Model A);结构命名:329-572-UL97-1YDTE-L1.FASTA;ProQ预测结果:LGscore 4.197,MaxSub 0.344;B:以J3.FASTA为比对文件,以1YDTE为模板的建模结构(Model B);结构命名:329-572-UL97-1YDTE-J3.FASTA;ProQ预测结果:LGscore 4.152,MaxSub 0.329;C:以N1.FASTA为比对文件,以1ckiA为模板的建模结构(Model C);结构命名:329-572-UL97-1ckiA-n1.FASTA;ProQ预测结果:LGscore 3.960,MaxSub 0.254

2.5 利用筛选出的最优模拟结构对新发现的非同义点突变进行初步耐药分析 为了验证该模拟结构的可靠性,利用已知的耐药相关点突变(M460V以及H520Q<sup>[14]</sup>)和耐药无关点突变(V354A<sup>[12]</sup>)作为阴性和阳性对照,分析其氨基酸突变前后的结构信息(包括结构框架、范德华力大小及作用范围、氢键等方面)是否发生有意义的改变(图4)。从图4A/B中可以看出354位氨基酸从缬氨酸(V)变为丙氨酸(A)时,上述三个方面的结构信息都没有发生显著改变,与其突变前后对药物敏感性不变吻合。而之后

的一系列图(包括阳性耐药点突变以及新发现的非同义点突变和联合突变)显示,上述三个方面的结构信息都发生了或多或少的显著性改变。这不仅证实了阳性点突变的结构信息改变吻合了其耐药特性,也提示新发现的非同义点突变或联合突变产生耐药的可能性很大,但并没有找出一个量化的指标从而给出一个界值来判定是否发生质变。如果对接效果理想,可以根据点突变对分子对接造成的影响进行量化;但对接效果不够理想。另外,本实验室已经利用经典的病毒重组实验证实了C518Y点突变的耐药特性。

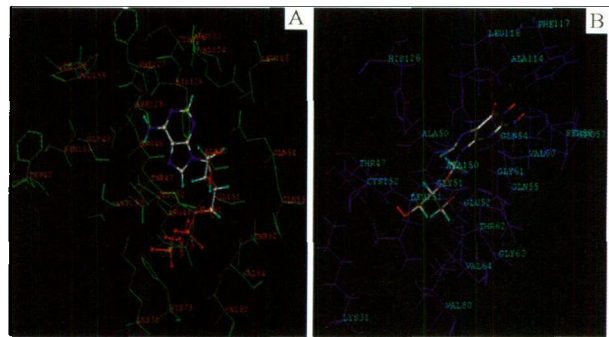


图3 PUL97蛋白局部模拟结构Model A与小分子ATP/GCV的对接

Fig 3 Docking results of PUL97 local modelling structure Model A with small molecules ATP or GCV

A: Model A与小分子ATP的对接图,对接分数为75分;B: Model A与小分子GCV的对接图,对接分数为48分。绿色或蓝色的细线条为局部蛋白主链;中间线条较粗的球棒结构为GCV或ATP分子;桔黄色或蓝色的英文字母及数字为氨基酸缩写名及其在链中的位置(全长为244个氨基酸)。

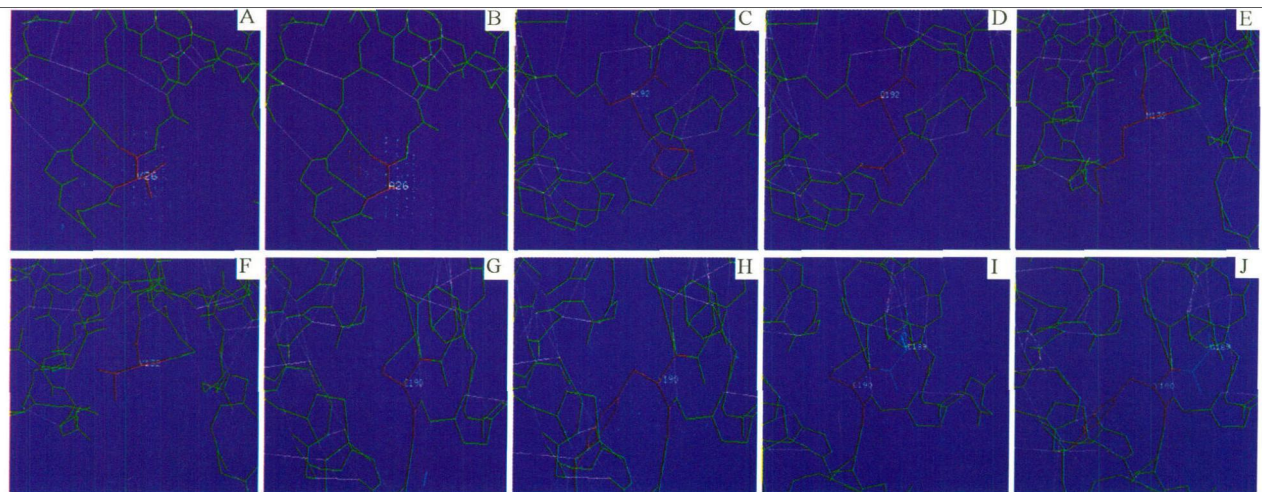


图4 对新发现的非同义点突变或联合突变进行耐药分析

Fig 4 Analysis of drug-resistant possibility of newly-found mutations using finally-screened modelling structure

A、B为耐药无关点突变(V354A)的突变前后结构信息对比;C、D和E、F分别为耐药相关点突变H520Q和M460V突变前后的结构信息对比;G、H和I、J分别为新发现的非同义点突变C518Y和联合突变(E517G,C518Y)的突变前后结构信息对比。绿色的线条为蛋白主链;粉色的线条为氢键;红色的线条或蓝色的线条(见I、J)为靶氨基酸(即待分析的氨基酸)的主链和支链;点状图案为靶氨基酸所具有的范德华力的大小和作用范围;图中的字母和数字为靶氨基酸的缩写及其在链中的位置。

### 3 讨论

HCMV 潜伏再激活和 HCMV 病处于移植后病毒感染的首要地位。临床上, GCV 等抗 HCMV 药物显示了良好的治疗和预防作用,但这些药物的广泛及非规范化使用导致耐药株不断出现。若不及时控制耐药株感染,与移植排斥形成恶性循环,严重威胁到移植以及患者的存活。

但至目前为止,国内外仍没有建立良好的耐药监测方法。表型检测(主要是蚀斑试验)耗时长,由于培养过程的不确定性,并存在较强的主观因素及缺乏标准化,使其在临床上的实用价值较小。基因型方法(主要是全基因测序)对耐药基因的检测较成熟,结果客观且耗时较短,是一种有临床实用价值的监测方法。但是,只能根据序列对已验证的耐药相关突变作报告从而指导临床用药。当出现新的突变位点或种类时,无法对其耐药性进行评估;而且目前较明确的 UL97 基因耐药谱基本是根据国外的 HCMV 流行株得出的(括号内为突变型的发生频率和 IC50 药物浓度相对于敏感株的倍数): A594V (30% ~ 34.5%, 10.7)、L595S (20% ~ 24%, 4.9 ~ 11.5)、M460V (11.5% ~ 14.5%, 7)、H520Q (5% ~ 11.5%, 10)、C592G 和 C603W 等<sup>[3]</sup>。根据前期的研究,国内 HCMV 流行株的 UL97 基因非同义突变近 50% 不同于国外报道,而这些新的突变点主要集中在该模拟区段内。对于不断发现的新突变点,需要花费大量的时间和精力通过生物学实验加以验证。但是,临床对数据的迫切需求促使我们利用已有的数据建立一种快速的计算机预测方法,对新突变点作出相对可靠的预测,从而及时指导临床调整治疗方案。而且,若所得模拟结构准确性较高,可用于计算机新药筛选,有着巨大的商业价值。目前,最大的好处是,如果计算机在野生型蛋白模拟结构的基础上对新突变点的耐药潜能作出较可靠的预测,就可以大大降低实验验证的盲目性以及减少工作量。

结构模拟是一种过渡方法,这种方法的存在有其强大的知识体系支撑。如序列同源导致三维结构相似和功能相似;即使序列同源度较差,若二级结构同源,也有足够的理由导致相似的三维结构及相似的功能。关键在于二级结构的折叠类型非常有限,再加上分子对接技术的发展(基于蛋白发挥某种功能往往会与其他物质发生亲密接触的特性),PDB 库中模板不断增多以及各种生物信息学强大的建模和

评估方法使得蛋白结构模拟可靠性越来越强,可替代真实结构解决的问题越来越多。

但是, PUL97 蛋白模拟系统最大的不足是, PDB 库中的模板与 PUL97 蛋白序列同源度不是很高,只有 10% ~ 20%。但是,我们可以利用各种现有的方法来部分弥补这一缺陷对结构准确性所带来的影响<sup>[15-16]</sup>。① SWISS-MODEL 实际上也部分整合了二级结构同源的建模原理;② 多个模板序列参与的比对模式实际上部分起到了分段模拟的效果;③ SWISS-MODEL 本身对建模后的结构也有很好的优化系统,同时 AUTODOCK 软件也能对模拟结构进一步优化;④ 各种评估服务器以及 SWISS-MODEL 自带的评估指标,特别是与小分子药物的对接分析等都为模拟结构的筛选设定了标准。因此,经过不断地优化分析以及重重地筛选,所得到的模拟结构理论上基本具备了对新突变型进行初步耐药分析的条件。但是,对于结构信息改变不敏感的单个氨基酸突变,分析时需谨慎。我们相信,随着 PDB 库中模板信息的不断增加及模板与 PUL97 序列同源度的不断提高,计算机的耐药分析方法会变得愈加正确和主流。

另外, PUL97 蛋白的功能盲点也为其结构模拟增加了一定的难度。PUL97 是 HCMV 的功能蛋白,具有磷酸转移酶的活性,其总长度是 707 个氨基酸。N 端(大致是 1 ~ 328aa)主要包括细胞核内定位信号;中间区(329 ~ Xaa,具体划分到哪里不明确,有可能至 459aa)主要是蛋白激酶域,缺失这部分区域对任何底物的磷酸化作用都将丧失。因此,据推测,该区域理论上存在 ATP 结合部位。C 端具体从哪里开始不是很明确,各种实验现象都表明该区域主要与 GCV 的磷酸化有关。但是,是不是存在 GCV 的结合区域,如果存在这一区域,主要位于 C 端的哪个部位,这些问题目前仍然处于猜测阶段。目前,主流的观点还是认为 C 端存在 GCV 结合区。我们模拟的蛋白区域为 329 ~ 572aa(I ~ IX 亚域),很明显包含了假定的 ATP 结合域,而且根据 Michel 等<sup>[12]</sup>的观点, PUL97 蛋白 VII/VIII 亚域为细胞激酶的“缺失区”,推测很可能是 GCV 的结合区域。因此,从现有的数据资料以及理论推测上来看,模拟的蛋白区域很可能包含了 ATP 和 GCV 的结合位点(尽管目前还没有办法用数据证实这一点)。而本研究恰恰利用了上述的可能性作为筛选模拟结构的一个辅助条件,这是该系统的亮点,也存在一定的风险。其实 Model B 和

C 与 ATP 的结合分数也不高,若不考虑 GCV 结合分数,选择 Model A 也是理所当然。另外,分子对接软件实际上可以对小分子结合部位进行定位,但是图 3 显示可能与 GCV 相互作用的氨基酸只有少部分位于 C 端,而且这些氨基酸与已验证的耐药位点不是很相符。因此,PUL97 突变型的模拟结构与小分子的对接也没有进行下去。但在缺少上述对接数据的情况下,直接从模拟结构的突变前后结构信息改变上发现了一些可喜的结果。阴、阳性点突变均出现了与耐药特性不吻合或吻合的结构信息改变,而且这些经临床资料提示具有耐药可能的新突变或联合突变(C518Y;E517G + C518Y)也出现了显著的结构信息改变。最重要的是,本实验室也已经用病毒重组实验证实了 C518Y 的耐药特性。

综上所述,该模型尽管有很多局限性,但事实证明了他的可靠性和有效性,整个平台为临床检验领域带来了新思路和潜在的应用价值。

#### 参考文献:

- [ 1 ] Reeves M, Sinclair J. Aspects of human cytomegalovirus latency and reactivation[J]. *Curr Top Microbiol Immunol*, 2008, 325: 297 - 313.
- [ 2 ] Boeckh M, Nichols WG, Papanicolaou G, et al. Cytomegalovirus in hematopoietic stem cell transplant recipients: Current status, known challenges, and future strategies[J]. *Biol Blood Marrow Transplant*, 2003, 9(9): 543 - 558.
- [ 3 ] Gilbert C, Boivin G. Human cytomegalovirus resistance to antiviral drugs[J]. *Antimicrob Agents Chemother*, 2005, 49(3): 8738 - 8783.
- [ 4 ] Chou S. Cytomegalovirus UL97 mutations in the era of ganciclovir and maribavir[J]. *Rev Med Virol*, 2008, 18(4): 233 - 246.
- [ 5 ] Avery RK. Ganciclovir-resistant cytomegalovirus disease in heart transplant recipients: the dilemma of donor-positive/recipient-negative serostatus[J]. *Clin Infect Dis*, 2007, 45(4): 448 - 449.
- [ 6 ] Chou S, Erice A, Jordan MC, et al. Analysis of the UL97 phosphotransferase coding sequence in clinical cytomegalovirus isolates and identification of mutations conferring ganciclovir resistance[J]. *J Infect Dis*, 1995, 171(3): 576 - 583.
- [ 7 ] Chou S, Van Wechel LC, Lichy HM, et al. Phenotyping of cytomegalovirus drug resistance mutations by using recombinant viruses incorporating a reporter gene[J]. *Antimicrob Agents Chemother*, 2005, 49(7): 2710 - 2715.
- [ 8 ] Jones DT. Protein structure prediction in the postgenomic era[J]. *Curr Opin Struct Biol*, 2000, 10(3): 371 - 379.
- [ 9 ] Shi R, Azzi A, Gilbert C, et al Three-dimensional modeling of cytomegalovirus DNA polymerase and preliminary analysis of drug resistance[J]. *Proteins*, 2006, 64(2): 301 - 307.
- [ 10 ] Romaker D, Schregel V, Maurer K, et al. Analysis of the structure-activity relationship of four herpesviral UL97 subfamily protein kinases reveals partial but not full functional conservation[J]. *J Med Chem*, 2006, 49(24): 7044 - 7053.
- [ 11 ] Schwede T, Kopp J, Guex N, et al. SWISS-MODEL: An automated protein homology-modeling server[J]. *Nucleic Acids Res*, 2003, 31(13): 3381 - 3385.
- [ 12 ] Michel D, Kramer S, Höhn S, et al. Amino acids of conserved kinase motifs of cytomegalovirus protein UL97 are essential for autophosphorylation[J]. *J Virol*, 1999, 73(10): 8898 - 8901.
- [ 13 ] Michel D, Schaarschmidt P, Wunderlich K, et al. Functional regions of the human cytomegalovirus protein pUL97 involved in nuclear localization and phosphorylation of ganciclovir and pUL97 itself[J]. *J Gen Virol*, 1998, 79 (Pt 9): 2105 - 2112.
- [ 14 ] Baldanti F, Michel D, Simoncini L, et al. Mutations in the UL97 ORF of ganciclovir-resistant clinical cytomegalovirus isolates differentially affect GCV phosphorylation as determined in a recombinant vaccinia virus system[J]. *Antiviral Res*, 2002, 54(1): 59 - 67.
- [ 15 ] Arnold K, Bordoli L, Kopp J, et al. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling[J]. *Bioinformatics*, 2006, 22(2): 195 - 201.
- [ 16 ] Kiefer F, Arnold K, Künzli M, et al. The SWISS-MODEL repository and associated resources[J]. *Nucleic Acids Res*, 2009, 37 (Database issue): D387 - D392.

收稿日期: 2009-10-27

本文编辑: 周殊凤