

doi: 10.3969/j.issn.0490-6756.2017.03.039

同源建模方法预测蛋白质突变结构的适用性分析

刘继龙, 肖智雄, 曹洋

(四川大学生命科学学院生物资源与生态环境教育部重点实验室, 生长代谢衰老研究中心, 成都 610064)

摘要: 通过从 Protein Data Bank(PDB) 结构数据库中提取单氨基酸突变的晶体结构, 构建了一组无冗余的测试数据集, 对目前应用最广泛的两款同源建模预测软件(SWISS-MODEL 和 MODELLER)进行了测试分析, 发现它们对蛋白质的整体结构预测效果良好, 均方根偏差小于 0.5 埃($RMSD < 0.5 \text{ \AA}$), 但在突变导致结构显著变化($RMSD > 1.5 \text{ \AA}$)的情况下却均不能得到准确结果. 分类统计显示, 发生在蛋白质结构内部和极性氨基酸之间的突变结构变化小, 两款软件预测效果较好($RMSD < 1.0 \text{ \AA}$). 突变导致结构显著变化的可能性不高($< 5\%$), 但它对蛋白质功能的影响不可忽视, 因此应用同源建模方法对于蛋白质突变的模拟并不完全适用, 还需要开发新方法来提高准确性.

关键词: 生物信息; 突变; 同源建模; 结构预测

中图分类号: Q811.4 **文献标识码:** A **文章编号:** 0490-6756(2017)03-0658-07

The applicability of predicting protein structures with residue substitutions using homology modeling

LIU Ji-Long, XIAO Zhi-Xiong, CAO Yang

(Center of Growth, Metabolism and Aging, Key Lab of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610064, China)

Abstract: Single residue substitution crystal structures were extracted from Protein Data Bank (PDB) to construct a non-redundant test set, which was used to assess two homology modeling structure prediction tools, SWISS-MODEL and MODELLER. The results indicate that the two methods perform well in overall structure prediction ($RMSD < 0.5 \text{ \AA}$), however, they fail to predict the mutants that own significant structural changes ($A_RMSD > 1.5 \text{ \AA}$) upon residue substitutions. The two methods perform better in mutations of embedded residues and mutations between polar residues. Although it is a small fraction of residue substitutions ($< 5\%$) that lead to significant conformation changes, they may play important roles in loss or gain protein functions. Therefore traditional homologous modeling methods are not appropriate for predicting such protein structures. There is urgent need to developing more powerful prediction methods.

Keywords: Bioinformatics; Mutation; Homology Modeling; Structure Prediction

收稿日期: 2016-01-27

基金项目: 国家自然科学基金(31401130)

作者简介: 刘继龙(1988-), 男, 河北张家口人, 硕士研究生, 研究方向为生物信息学. E-mail: liujilong@yeah.net

通讯作者: 曹洋. E-mail: cao@scu.edu.cn

1 引言

蛋白质是生命的物质基础, 是生命活动的主要承担者. 根据中心法则^[1], 遗传信息由 DNA 编码, 经转录翻译后折叠成具有四级结构的蛋白质发挥生物学功能. 蛋白质的空间构象决定生物学功能^[2], DNA 的点突变会引起蛋白质氨基酸残基突变^[3], 导致蛋白质三维结构发生变化, 并最终影响蛋白质稳定性和催化活性. 一个典型的例子是 xyl-d 二型乙醇脱氢酶(SADH)发生点突变第 198 位氨基酸残基由甘氨酸(Gly, G)置换为精氨酸(Arg, R), 总体催化活性相对于野生型有所下降, 其中突变蛋白对 NAD/NADH 的亲合性分别下降了约 8 倍和 6 倍, 而且对 NADPH 的亲合性下降也非常明显^[4]. 另一项研究中, 研究者将脱毛碱性蛋白酶(DHAP) 248 位天冬氨酸(Asp, D)在体外突变为苯丙氨酸(Phe, F)、丝氨酸(Ser, S), 酪蛋白水解活性均有提高^[5]. 因此, 研究蛋白质单点突变造成的生物学功能变化有着重要的意义.

分析蛋白质单氨基酸残基突变带来的结构变化是研究其功能变化的基础. 由于解析晶体结构难度大, 耗费高, 三维结构预测成为探索残基突变引起的结构变化的重要方法. 蛋白质三维结构预测软件根据算法分为两类^[6, 7]. 第一类为 de novo 从头预测, 输入氨基酸一级序列, 程序根据能量最小、减少碰撞等原则调整每个原子的三维空间位置, 进而计算确定整个蛋白质的三维结构, 如 Rosetta^[8]、I-TASSER^[9]等; 第二类基于同源建模算法, 选定输入一级序列的同源蛋白结构, 然后以同源蛋白的三维结构为模板利用理论计算方法进行优化, 如 SWISS-MODEL^[10]和 MODELLER^[11]. 由于同源模板的存在, 第二类方法预测结果与真实结构符合性更好, 但这一方法对突变蛋白质的三维结构预测是否有效还缺少系统的研究.

在本研究中, 我们对 Protein Data Bank (PDB)^[12]数据库中晶体结构进行分类, 提取出只有单个氨基酸残基不同的晶体结构对, 构建标准单氨基酸残基结构预测软件测试数据集, 对基于同源建模的蛋白质三维结构预测软件 MODELLER 和在线服务 SWISS-MODEL 进行测试. 分别应用两款软件预测发生单氨基酸突变的蛋白质结构, 根据预测结果和晶体结构之间的差异来判断其对发生残基突变的蛋白质结构的预测准确性.

2 材料和方法

2.1 蛋白质晶体结构差异的量化

RMSD(Root-Mean-Square Deviation, 均方根偏差)^[13]是结果构象和目标构象间的偏差统计, 是量化蛋白三维结构差异程度的最常用标准. *RMSD* 计算前首先要将模拟结果构型和目标构型(一般为初始结构)进行重叠, 使两者重合或部分重合,

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

其中 δ_i 为对应原子间的距离, i 为原子数目. 在我们的工作中, 为排除突变位点周边的氨基酸残基对计算的影响, 对突变位点前后各 10 个氨基酸残基中离突变位点较远的 8 个共 16 个氨基酸残基的 N、CA、C、O 四个主链原子进行空间叠加. 为了更好的描述突变位点和突变位点周边的结构变化, 我们进一步定义了 *M_RMSD*(突变位点的均方根偏差)和 *A_RMSD*(突变位点序列所处环境的均方根偏差). *M_RMSD* 计算突变位点及前后共三个氨基酸残基主链原子 N、CA、C、O 的均方根偏差, 用以描述突变位点的结构变化; *A_RMSD* 计算上边提到的 16 个空间叠加的氨基酸残基在突变发生前后的主链原子 N、CA、C、O 的均方根偏差, 用以描述突变位点所处环境的晶体结构变化.

2.2 构建测试数据集

PDB 是目前全球最大的实验获取的生物大分子结构数据库, 本研究从 PDB 中筛选分辨率小于 3 Å 的 X 射线衍射晶体结构, 拆分为蛋白质单链, 以 PDB ID 加链 ID 进行命名(如 1a2cA 为蛋白 1a2c 的 A 链)^[14], 然后剔除存在断裂的蛋白链, 最后对来源于同一蛋白链的晶体结构归一化处理, 形成无冗余单链蛋白结构数据库(图 1). 对该数据库中的蛋白链 FASTA 序列进行比对, 找出包含单个氨基酸残基不同的晶体结构组合. 链末端的残基热运动更强烈, 热扰动因子 B-factor 体现晶体结构中原子构象状态的模糊度, *A_RMSD* 描述突变位点所处环境的晶体结构变化, 定义

$$B_{norm} = \frac{(B_{res} - B_{ave})}{B_{ave}}$$

其中 B_{res} 为氨基酸残基的热扰动因子, B_{ave} 为该链氨基酸残基的热扰动因子均值^[15]. 使用突变位点到末端距离小于 10、 B_{norm} 小于 0.3 和 *A_RMSD* 小于 1.0 Å 进行过滤后生成点突变结构对数据库(图 1). *M_RMSD* 标识了突变位点的空间构象变化,

$M_RMSD < 1.0 \text{ \AA}$ 的突变对结构变化小,为阴性集, $M_RMSD \geq 1.5 \text{ \AA}$ 的突变对结构变化大,为阳性集.为得到普遍性结论,限定突变对分辨率差值小于 0.2 \AA ,按突变对的分辨率均值进行排序,在保证 20 种残基的分布尽量平均的前提下,分别提取出 100 个突变对作为测试集.

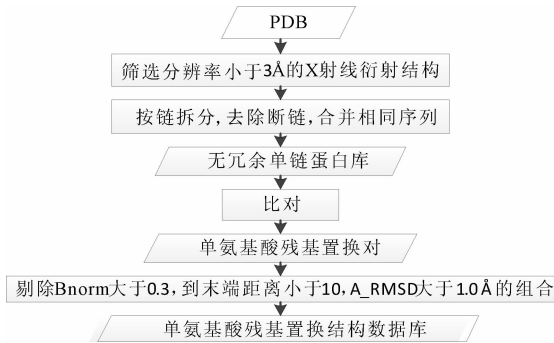


图 1 无冗余单链蛋白数据库和点突变数据库的构建方法和步骤

Fig. 1 The flowchart of constructing non-redundant protein database and protein point mutation database

为了进一步探索不同氨基酸性质对结构预测准确性的影响,我们将测试数据分别按照突变氨基酸的包埋情况、二级结构和氨基酸带电情况进行了划分.我们 1)应用分子溶剂可及表面积计算软件 NACCESS^[16] 计算了每个突变氨基酸暴露于蛋白质溶剂可及表面的比例,以 40% 为分界线,其下为包埋氨基酸,其上为暴露氨基酸;2)使用 DSSP^[17] 计算突变氨基酸二级结构,分成有稳定二级结构(即 α -螺旋和 β 片层)和无规卷曲两类;3)根据氨基酸侧链基团极性,把突变氨基酸分为非极性氨基酸(A, V, L, I, F, W, M, P)、极性不带电氨基酸(G, S, T, C, Y, E, Q)和极性带电氨基酸(H, K, R, D, E)三类,而它们在突变中共有 $3 \times 3 = 9$ 种突变组合.所有的计算结果将依据这三大分类来分析氨基酸突变结构预测的准确性.

2.3 SWISS-MODEL 和 MODELLER 测试

对测试数据集中每一对残基突变蛋白结构,设定任意一个蛋白链为野生型,另一个即为突变型,以野生型为同源模板预测突变型序列的三维结构,然后计算预测结构和突变晶体结构的 M_RMSD 、 A_RMSD 以评估预测软件对单氨基酸突变的预测准确性.

SWISS-MODEL 是欧洲瑞士生物信息中心开发的蛋白质同源建模程序,提供全自动的计算流程,用户通过登陆网站 <http://swissmodel.expasy.org>,在 Provide a protein sequence 输入框中

输入突变型的 FASTA 序列,点击 Submit Modeling Request 提交即可.我们的计算测试需要在 Advanced options 中指定野生型蛋白为模板链,其它采用默认参数,待服务器计算完成后,下载返回的 PDB 格式的蛋白质三维结构文件(附图 1).MODELLER 是美国加州大学洛杉矶分校开发的同源建模程序,是当前蛋白质同源建模方法的标杆之一.相比 SWISS-MODEL 使用更复杂,需要用户编写 python 脚本来运行.我们编写了计算脚本,并配置野生型模板结构和野生型、突变型的 FASTA 序列,执行比对脚本后取得结构预测结果 PDB 格式的蛋白质三维结构文件(附图 2).

表 1 测试数据集

Tab. 1 Test set

分类标准	突变氨基酸分类	样本量	A_RMSD / \AA	M_RMSD / \AA
测试数据集	阴性突变	100	0.195	0.273
	阳性突变	100	0.404	1.945
氨基酸位置	包埋氨基酸突变	122	0.197	0.919
	暴露氨基酸突变	78	0.307	1.438
二级结构	稳定二级结构突变	135	0.292	1.130
	无规卷曲突变	65	0.315	1.064
氨基酸侧链极性	非极性 \rightarrow 非极性	47	0.300	1.331
	非极性 \rightarrow 中性	26	0.319	1.114
	非极性 \rightarrow 极性	14	0.301	1.410
	中性 \rightarrow 非极性	24	0.337	1.191
	中性 \rightarrow 中性	16	0.208	1.015
	中性 \rightarrow 极性	25	0.270	0.789
	极性 \rightarrow 非极性	16	0.274	1.266
	极性 \rightarrow 中性	21	0.303	0.908
	极性 \rightarrow 极性	11	0.392	0.600

3 结果

3.1 测试数据集

从 PDB 下载晶体结构数据,加工处理后我们得到有 32154 条非重复序列无冗余蛋白结构数据库,比对处理后得到 6385 个单氨基酸突变对,分别取出 100 对晶体结构组成阳性测试集(附表 1)和阴性测试集(附表 2).以氢原子半径(约 1 \AA)为结构变化大的界限,蛋白结构整体(A_RMSD)变化都很小,点突变位点结构(M_RMSD)变化在不同分类中分布并不均匀.测试集总共包含 1)78 个暴露氨基酸突变,点突变结构变化大,122 个包埋氨基酸突变,点突变结构变化小,原因可能是包埋氨基酸比暴露在溶剂中的氨基酸有更好的结构稳定性;2)135 个稳定二级结构氨基酸突变,65 个无规卷曲氨基酸突变,结构变化都较大;3)氨基酸带电

性质分类的 9 种组合中,非极性氨基酸参与的点突变结构变化较大,而极性带电氨基酸到极性带电氨基酸的点突变结构变化小,其原因可能是极性带电氨基酸残基受静电作用约束,不易移动,非极性氨基酸在蛋白质中主要受非特异的范德华作用约束,易发生空间位置变动(表 1). 接下来我们首先用阳性集与阴性集进行测试分析,然后对所有测试数据按突变位点的空间位置、二级结构、氨基酸带电类型进行分类研究,探索哪些因素影响同源建模软件对点突变预测的准确性.

3.2 SWISS-MODEL 测试结果

应用 SWISS-MODEL 预测的突变蛋白结构与晶体结构的比较结果见图 2. 在阴性集合中,应用 SWISS-MODEL 对蛋白质点突变做模拟预测的结

果和晶体结构的 M_RMSD 平均值为 0.370 \AA (图 2B),且 95% 的值分布于区间 $(0 \text{ \AA}, 1 \text{ \AA}]$ (图 2A 黑色虚线),这个距离小于氢原子的半径(约 1 \AA),表明它对突变位点本身的预测很准确, A_RMSD 的平均值为 0.322 \AA ,95% 以上值分布于区间 $(0 \text{ \AA}, 1 \text{ \AA}]$ (图 2A 黑色实线),表明它对突变位点周边的预测也很准确. 在阳性集合中,SWISS-MODEL 的 A_RMSD 依然很小,平均值为 0.412 \AA ,全部值分布于区间 $(0 \text{ \AA}, 1 \text{ \AA}]$ (图 2A 灰色实线),说明对突变位点周围的情况的预测依然准确,但是它对突变位点本身的预测 M_RMSD 相差很大,平均值为 1.938 \AA (图 2B),分布于区间 $(1 \text{ \AA}, 4 \text{ \AA}]$ (图 2B 灰色虚线),远大于氢原子的半径(约 1 \AA),说明预测结构与真实结构偏差很大.

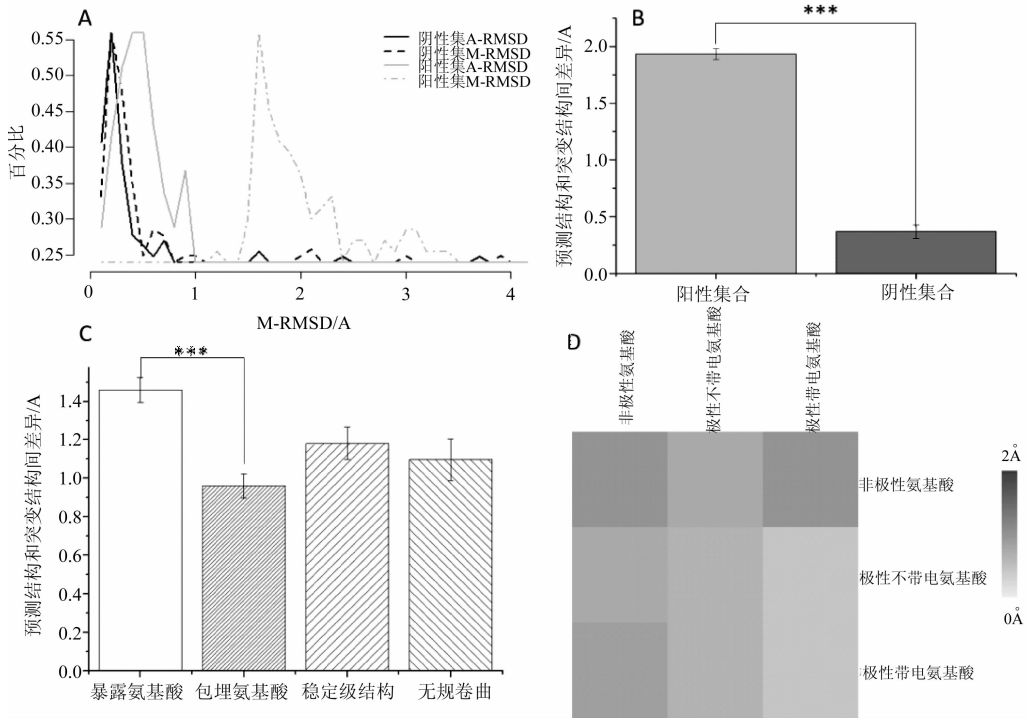


图 2 SWISSMODEL 测试结果.

A: A_RMSD 分布和 M_RMSD 分布. B: M_RMSD 在阳性集合和阴性集合中的差异. C: M_RMSD 在突变位点位置/二级结构分类中的差异. D: M_RMSD 在突变氨基酸带电性质上的差异热图,颜色越深表明 M_RMSD 数值越大. *** 表示 $P\text{-value} < 0.005$.

Fig. 2 Assessment of SWISSMODEL.

A: A_RMSD and M_RMSD distribution. B: M_RMSD between positive set and negative set. C: M_RMSD of different types of amino acids. D: Heat map of M_RMSD on pairs of amino-acid classifications. The darker the color, the larger the value of M_RMSD . *** indicates $P\text{-value} < 0.005$.

上面的结果显示 SWISS-MODEL 对阳性集合数据的预测准确性显著低于阴性数据,那么不同类型氨基酸和预测准确性是否存在关联? 我们对数据进行了如下分类:1)以突变位点所在三维结构中的位置分类,发现包埋氨基酸的预测结果和晶体结构的 M_RMSD 为 0.960 \AA ,而暴露氨基酸的 M_RMSD 1.459 \AA (图 2C),两者具有显著差异,说明

氨基酸的空间位置对预测准确性有影响,发生在蛋白质内部的突变预测准确性更好;2)以突变位点所在二级结构分类,预测结果的 M_RMSD 分别为 1.181 \AA 和 1.097 \AA ,两者无显著差异(图 2C),说明二级结构对于预测准确性的相关程度不高;3)以突变位点氨基酸带电类型分类,非极性氨基酸参与的突变结构变化大, M_RMSD 均大于氢原子半

径,极性带电氨基酸间的突变结构变化小,其 M_RMSD 仅为 0.775\AA (图 2D),表明 SWISS-MODEL 对中性 and 极性带电氨基酸突变的预测准确性相对较高。综上所述,包埋在蛋白质内部,中性或亲水氨基酸的突变前后结构变化小,SWISS-MODEL 预测的平均 M_RMSD 小于 1.0\AA ,较为准确,而暴露在蛋白质表面,非极性氨基酸的突变易于发生主链结构变化,SWISS-MODEL 突变预测结果与真实结构偏差较大。

3.3 MODELLER 测试结果

应用 MODELLER 预测的突变蛋白结构与晶体结构的比较结果见图 3,在阴性集合中,用 MODELLER 给出的预测蛋白结构和晶体结构的 M_RMSD 的平均值为 0.399\AA (图 3B),且 $>95\%$ 均分布于区间 $(0\text{\AA}, 1\text{\AA}]$ (图 3A 黑色虚线),说明它对突变位点做出的预测准确, A_RMSD 的平均值为 0.333\AA , $>95\%$ 分布于区间 $(0\text{\AA}, 1\text{\AA}]$ (图 3A 黑色实线),说明它对突变位点周边预测同样准确。在阳性集合中,MODELLER 给出的 A_RMSD 依然很小,

平均值为 0.432\AA , 95% 以上的值分布于区间 $(0\text{\AA}, 1\text{\AA}]$ (图 3A 灰色实线),说明对突变位点周围的情况的预测准确,但它对突变位点本身的预测却误差很大,两者间的 M_RMSD 平均值达到 1.891\AA (图 3B),分布于区间 $(0\text{\AA}, 5\text{\AA}]$,且 95% 以上的值分布在区间 $[1\text{\AA}, 3\text{\AA}]$ (图 3B 灰色虚线),这个偏差说明 MODELLER 预测结构与真实结构偏差较大。

与 SWISS-MODEL 的结果相似,MODELLER 对阳性集合数据的预测准确性也显著低于阴性数据,我们同样进行了分类研究:1)以突变位点所处位置分类,包埋氨基酸的预测结果和晶体结构的 M_RMSD 为 0.958\AA ,暴露氨基酸的 M_RMSD 为 1.437\AA (图 3C),两者具有显著差异,表明氨基酸的空间位置对预测准确性有影响,发生在蛋白质内部的突变预测准确性更好;2)以突变位点所在二级结构分类,稳定二级结构和无规卷曲的结构变化均大于 1\AA , M_RMSD 分别为 1.170\AA 和 1.091\AA ,两者无显著差异(图 3C),同样说明二级结构对于预测准确性的相关程度不高;3)以突变位

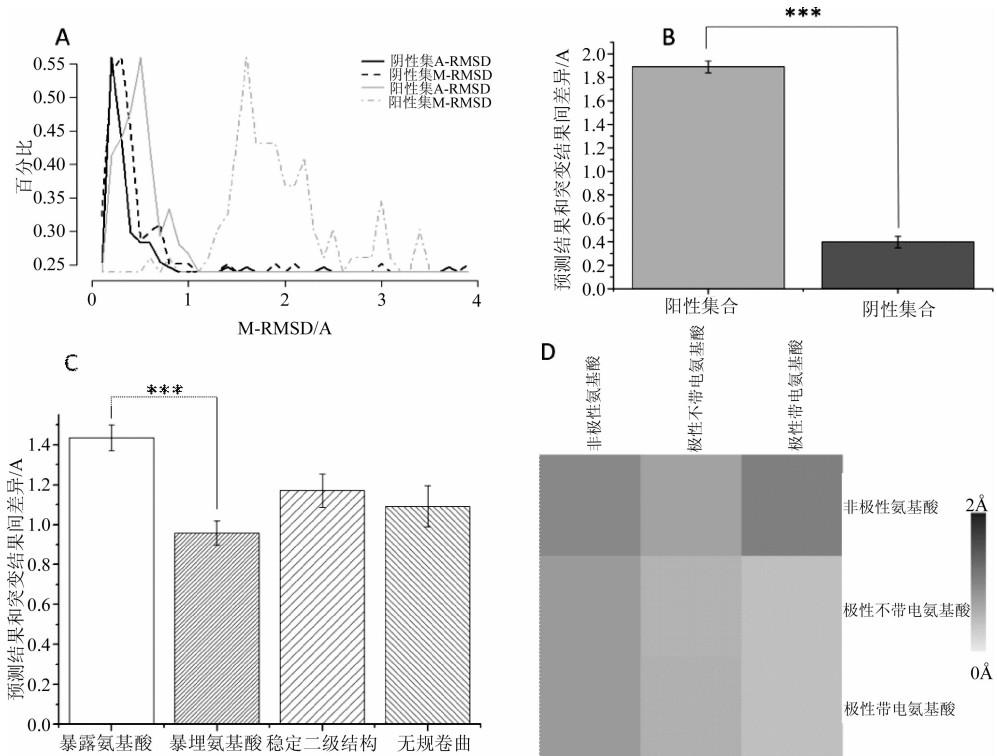


图 3 MODELLER 测试结果

A: A_RMSD 分布和 M_RMSD 分布. B: M_RMSD 在阳性集合和阴性集合中的差异. C: M_RMSD 在突变位点位置/二级结构分类中的差异. D: M_RMSD 在突变氨基酸带电性质的差异热图,颜色越深表明 M_RMSD 数值越大. *** 表示 $P\text{-value} < 0.005$.

Fig. 3 Assessment of MODELLER

A: A_RMSD and M_RMSD distribution. B: M_RMSD of positive set and negative set. C: M_RMSD of different types of amino acids. D: Heat map of M_RMSD on pairs of amino-acid classifications. The darker the color, the larger the value of M_RMSD . *** indicates $P\text{-value} < 0.005$.

点氨基酸类型分类,非极性氨基酸参与的突变结构变化大, M_RMSD 均大于氢原子半径,极性带电氨基酸间的突变结构变化小,其 M_RMSD 为 0.783 \AA (图3D),表明预测软件对极性不带电和极性带电氨基酸突变的预测准确性相对较高。结果表明,MODELLER的结果与SWISS-MODEL有相同的趋势,包埋在蛋白质内部,中性或亲水氨基酸的突变前后结构变化小,平均 M_RMSD 小于 1.0 \AA ,较为准确。

当然,这两款软件仍有一些区别,在阳性测试

数据中,我们注意到MODELLER有两个预测较为准确的情况.黄素氧还蛋白57位Thr(1fldA)到Gly的突变(5ullA,图4A),58位的Gly(1czhA)到Asn的突变(1ofvA,图4B)中,MODELLER的预测结果与晶体结构偏差很小,分别为 0.785 \AA 和 0.503 \AA .这说明当氨基酸的侧链体积发生较大变化时,MODELLER的主链自动调整流程对预测结构的主链做出了有益调整,这在SWISS-MODEL中是没有的.不过多数阳性数据中(98%),其结果仍不能与真实结构相符合。

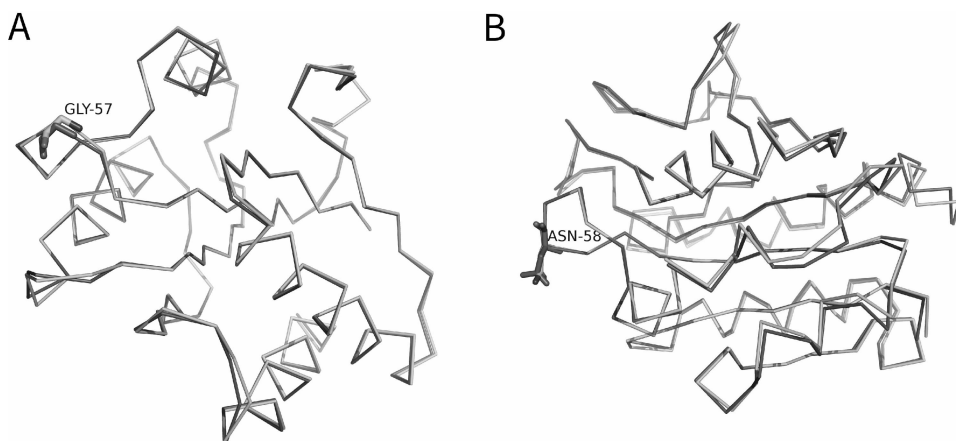


图4 预测结构(浅灰色)与晶体结构(深灰色)主链构象比较.棍状表示突变位点

A: 1fldA. pdb到5ullA. pdb 57位Thr到Gly突变, $M_RMSD=0.785 \text{ \AA}$. B: 1czhA. pdb到1ofvA. pdb 58位的Gly到Asn, $M_RMSD=0.503 \text{ \AA}$.

Fig. 4 Comparison of the predicted structure (Light grey) and the crystal structure (dark grey). Sticks indicate the mutation site. A: Mutation from 1fldA. pdb 57 Thr to 5ullA. pdb 57 Gly, $M_RMSD=0.785 \text{ \AA}$. B. The mutation of 1czhA. pdb 58 Gly to 1ofvA. pdb 58 Asn, $M_RMSD=0.503 \text{ \AA}$.

4 讨论

同源建模结构预测方法基于两条假设:蛋白质的结构由其氨基酸序列唯一的决定;结构在进化中更稳定,变化比序列层面的变化要缓慢许多.预测过程中软件通过简单拷贝模板蛋白主链坐标来设计目标蛋白主链,然后对能量、位置有冲突的原子进行计算调整,生成预测结果. MODELLER和SWISS-MODEL是目前应用最广泛的同源结构预测方法,具有很强的权威性,但是它们都没有对蛋白突变引起结构的变化情况进行过系统分析测试.我们在单氨基酸突变结构预测分析中发现,对于非突变位点区域,MODELLER和SWISS-MODEL预测非常准确,对于突变位点区域,其预测结果均是蛋白质主链结构基本不会改变,而不能

对残基突变导致结构较大变化的情况出准确预测.其原因可能是同源建模算法优先遵循同源模板结构,只在能量、位置有冲突时才考虑调整主链原子坐标.相比之下,MODELLER能在一定程度上做出较好的主链调整.

包埋在蛋白结构内部氨基酸空间位阻更大,极性带电氨基酸在蛋白结构中受到静电作用的约束,因此这两类氨基酸的点突变对结构的影响小,发生突变后本身主链结构变化并不大,突变结构和野生型之间的 M_RMSD 很小,SWISS-MODEL和MODELLER的预测结果与突变晶体结构也最接近.综上,同源建模方法并不完全适用于单氨基酸残基突变的结构预测,用户可计算突变位点的包埋程度等特征来评估预测可靠性,在关键位点差异显著的情况下谨慎使用。

参考文献:

- [1] Crick F. Central dogma of molecular biology[J]. *Nature*, 1970, 227(5258): 561.
- [2] 苏智端, 林旭, 曾嵘, 等. 人群变异的分子基础: 从单核苷酸多态性到单氨基酸多态性[J]. *中国科学: 生命科学: 中文版*, 2013, 43(1): 16.
- [3] 殷志祥. 蛋白质结构预测方法的研究进展[J]. *计算机工程与应用*, 2004, 40(20): 54.
- [4] 韩俊, 刘志斌, 马诗淳, 等. xyl-d 二型乙醇脱氢酶 (SADH) G198R 突变体的研究[J]. *四川大学学报: 自然科学版*, 2015, 52(1): 187.
- [5] 黄蓉, 冯红. 脱毛碱性蛋白酶 (DHAP) 的饱和突变 [J]. *四川大学学报: 自然科学版*, 2015, 52(5): 1177.
- [6] 孙侠, 殷志祥. 蛋白质结构预测的理论方法及阶段 [J]. *生物学杂志*, 2007, 24(1): 15.
- [7] 席玉生. 蛋白质折叠及二级结构预测[J]. *广州化工*, 2008, 35(6): 16.
- [8] Simons K T, Kooperberg C, Huang E, *et al.* Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions[J]. *J Mol Biol*, 1997, 268(1): 209.
- [9] Zhang Y. I-TASSER server for protein 3D structure prediction [J]. *BMC Bioinformatics*, 2008, 9(1): 40.
- [10] Arnold K, Bordoli L, Kopp J, *et al.* The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling[J]. *Bioinformatics*, 2006, 22(2): 195.
- [11] Eswar N, Webb B, Marti-Renom M A, *et al.* Comparative protein structure modeling using Modeller [J]. *Curr protoc Bioinformatics*, 2006, 5(6): 1.
- [12] Berman H M, Westbrook J, Feng Z, *et al.* The protein data bank[J]. *Nucleic Acids Res*, 2000, 28(1): 235.
- [13] Kirchmair J, Markt P, Distinto S, *et al.* Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection-What can we learn from earlier mistakes[J]. *J Comput Aid Mol Desi*, 2008, 22(3-4): 213.
- [14] Eyal E, Najmanovich R, Sobolev V, *et al.* MutaP-rot: a web interface for structural analysis of point mutations[J]. *Bioinformatics*, 2001, 17(4): 381.
- [15] Bordner A J, Abagyan R A. Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations[J]. *Proteins*, 2004, 57(2): 400.
- [16] Hubbard S J, Thornton J M, Campbell S F. Substrate recognition by proteinases[J]. *Faraday Discuss*, 1992, 93: 13.
- [17] Wouter T, Coos B, Jon B, *et al.* A series of PDB related databases for everyday needs[J]. *Nucleic Acids Res*, 2015; 43(Database issue): D364.