

研究报告

Research Report

基于多个模板进行 *in-silico* 分析预测菠菜(*Spinacia olearacea*)光系统 II 的 D2 蛋白的最佳模型

Pranati Swain ✉

奥里萨邦农业技术大学, 印度

✉ 通讯作者: rosylora20@gmail.com

计算分子生物学, 2014 年, 第 3 卷, 第 7 篇 doi: 10.5376/cmb.cn.2015.03.0007

本文首次发表在 *Computational Molecular Biology* 上。现依据版权所有人授权的许可协议, 采用 Creative Commons Attribution License 协议对其进行授权, 再次发表与传播。只要对原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

建议最佳引用格式:

Swain, 2014, In-silico analysis predicting the best model for photosystemII D2 Protein of *Spinaciaolearacea* using multiple templates, *Computational Molecular Biology*, Vol.4, No.13, 1-6 (doi: 10.5376/cmb.2014.04.0013)

摘要 菠菜 (*Spinacia olearacea*) 是一种天然的抗糖尿病、前列腺癌、哮喘、便秘、高血压的药物。菠菜具有抗炎、抗增殖、抗氧化的作用。在这项研究中的考虑用 *in-silico* 分析光系统 II 的 D2 蛋白。使用 1I ZLD, 3A0B, 3WU2, 4IL6 模板生成 D2 蛋白的模型。从 UniProt 检索序列, 通过 blastp 工具预测模板, 使用 prot-param 工具分析蛋白质的理化性质, 使用 CFFSP 服务器进行二级结构预测, 分析其螺旋、旋转和折叠, 使用 modeller 9.12 工具生成同源模型, 使用 Rampage 服务器进行骨架确认, 最后利用 3A0B 模板建立最好的模型。3A0B 模板与 *Thermosynechococcus vulcanus* 光系统 Q (B) 蛋白有 91% 的疑似覆盖率, 即有 94.0% 残基落在允许区内, 2.0% 的残基落在非允许区, 以及有 95% 的序列同一性。

关键词 光系统 II 的 D2 蛋白; 模板预测; 同源建模; 模型验证; 最佳预测模型

In-silico analysis predicting the best model for photosystem II D2 Protein of *Spinaciaolearacea* using multiple templates

Pranati Swain ✉

Orissa university of agriculture and technology, India

✉ Corresponding author, rosylora20@gmail.com

Abstract Spinach is a natural medicine against diabetes, prostate cancer, asthma, constipation, high blood pressure. Spinach acts as anti-inflammatory, antiproliferative, antioxidative. In this study the PHOTOSYSTEM II D2 protein has considered for *in-silico* analysis. Models of the protein were generated using 1I ZLD, 3A0B, 3WU2, 4IL6 templates. The sequence retrieved from uniprot, templates were predicted by using blastP tool, physico-chemical analysis showed the properties of protein using prot-param tool, secondary structure prediction showed helices, turns and sheets using CFFSP server, homologous models were generated using modeller9.12 tool, backbone confirmation was performed by using Rampage server and finally the best model generated by using 3A0B template having 94.0% of residues lying in favored region, 2.0% residues lying in outlier region, with 91% of query coverage and 95% of identity with photosystemQ (B) protein of *Thermosynechococcus vulcanus*.

Keywords PHOTOSYSTEM II D2 protein; Template prediction; Homology modeling; Model validation; Best model prediction

菠菜 *Spinacia olearacea* 俗称菠菜, 属藜科, 是能量的重要来源。这种绿色的叶子常用来作为食物。研究证明, 菠菜富含维生素 C, 这有助于保护菠菜叶中所有氧敏感的植物营养素, 使得菠菜叶看起来

充满生机和活力。人们发现菠菜中支持健康的营养主要是甘油酯类。天然的菠菜能够抗炎(Lomnitski et al., 2000), 抗增殖(Bergman et al., 2011), 抗氧化(Sani et al., 2004)。

在一个花椰菜, 菠菜, 花椰菜, 卷心菜, 芥菜, 羽衣甘蓝和无头甘蓝的研究中发现, 菠菜有显著的抗男性前列腺癌发生的作用。菠菜具有得天独厚的天然抗癌类胡萝卜素即环氧叶黄素。它含有胡萝卜

收稿日期: 2014 年 10 月 27 日

接受日期: 2014 年 10 月 27 日

发表日期: 2014 年 10 月 27 日

素, 即 β -胡萝卜素, 叶黄素, 玉米黄质, 以及抗氧化剂, 如黄酮。菠菜是很健康的食物, 因为它富含维生素 K、维生素 A、维生素 B1、维生素 B3、维生素 B2、维生素 E、维生素 B6、铁、铜、叶酸、锰、钙、纤维、钾、锌、蛋白、胆碱、omega-3 脂肪酸、硒、泛酸等。维生素 K1 和维生素 K2 帮助激活骨质生成的骨钙蛋白。

然而, 菠菜含有过多的草酸盐, 如果食用过多危害健康。菠菜含有天然的嘌呤, 如果食用过多会导致肾结石和痛风病。通常意义上, 菠菜有助于对抗骨弱、高血压、糖尿病、哮喘、男性前列腺癌、便秘、胰腺癌细胞(Lomnitski et al., 2000)。菠菜也有助于能量代谢, 维持肌肉和神经功能, 心脏节律, 健康的免疫系统及维持血压。

在这项研究中, 我们考虑了菠菜光系统 II 的表 1 光系统 II D2 蛋白的信息

Table 1 Information about Photosystem II D2 protein

UniprotID	基因	功能	催化活性	亚细胞定位	氨基酸长度
	Gene	Function	Catalytic activity	Subcellular location	Amino acid length
P06005	psbD	ATP 形式 ATP formation	$2 \text{H}_2\text{O} + 2 \text{质体醌} + 4 \text{光} = \text{O}_2 + 2 \text{质体醌}$ $2 \text{H}_2\text{O} + 2 \text{plastoquinone} + 4 \text{light} = \text{O}_2 + 2 \text{plastoquinol}$	Transmembrane 横跨膜的 41-61 125-141 153-166 208-228 279-295	353

1.2 光系统 II 的 D2 蛋白的理化特性

利用 ProtParam 工具研究蛋白质的理化性质。分析了蛋白质的理论 PI、分子量、脂肪指数、消光

D2 蛋白。D2 蛋白是一种质体醌氧化还原酶, 利用光能从 H_2O 中转移电子, 产生氧气和质子梯度以产生 ATP。光系统 II 的 D2 是一种膜蛋白。PS II 是由一套 PsaA、PsaB、PsaC、PsaD、PsaE、PsaF、PsaH、PsaI、PsaJ、PsaK、PsaL、PsaM、PsaT、PsaX、PsaY、PsaZ、Ycf12 的膜蛋白基因复制组成。在研究中, 蛋白质的理化分析已随着同源建模完成, 利用多个模板进行模型验证和优化成为最佳预测模型。

1 材料和方法

1.1 序列检索

用 FASTA 格式从 UniProt 中检索光系统 II 的 D2 蛋白的氨基酸序列。详细信息在表 1 中给

系数、氨基酸数量, 带正负电荷的残基的总数量, 原子位置, 化学公式, 不稳定指数和 GRAVY (总平均亲水性)。详细信息是在表 2 和表 3 中给出的。

表 2 光系统 II D2 蛋白的理化性质

Table 2 Physico-chemical properties of Photosystem II D2 protein

分子量	理论 pI	带负电荷的残基的总数	带正电荷的残基的总数	原子总数	脂肪族氨基酸指数	总平均亲水性
Molecular weight	Theoretical pI	Total number of negatively charged residues	Total number of positively charged residues	Total number of atoms	Aliphatic index	Grand average of hydropathicity (GRAVY)
39507.4	5.46	28	20	5514	88.78	0.358

1.3 模板预测

一般利用 BLAST 工具进行序列同一性搜索, 而蛋白质之间的序列同一性搜索是通过 blastP 工具

进行。因此, 检索得到的氨基酸序列在 PDB 中进行 blastP。从结果中获得能够进一步研究的合适的模板。选定的建模模板见表 4。

表 3 氨基酸组成

Table 3 Amino acid composition

氨基酸	标志	残基数量	比例	氨基酸	标志	残基数量	比例
Amino acids	Symbols	Number of residues	In percentage	Amino acids	Symbols	Number of residues	In percentage
Ala	A	39	11.0%	Lys	K	5	1.4%
Arg	R	15	4.2%	Met	M	9	2.5%
Asn	N	15	4.2%	Phe	F	39	11.0%
Asp	D	12	3.4%	Pro	P	14	4.0%
Cys	C	4	1.1%	Ser	S	18	5.1%
Gln	Q	11	3.1%	Thr	T	20	5.7%
Glu	E	16	4.5%	Trp	W	13	3.7%
Gly	G	31	8.8%	Tyr	Y	8	2.3%
His	H	8	2.3%	Val	V	22	6.0%
Ile	I	13	3.7%	Pyl	O	0	0.0%
Leu	L	41	11.6%	Sec	U	0	0.0%

表 4 用于同源建模的 4 个模板列表

Table 4 List of four templates used for homology modelling

模板	链	序列同一性	疑似覆盖率	E 值	微生物	分子
Templates	Chain	Identity	Query cover	E-value	Organism	Molecule
1IZL	D	89%	100%	0.0	金黄色葡萄球菌 <i>Thermosynechococcus vulcanus</i>	光系统 II psbA 亚基 Photosystem II subunit psb A
4IL6	D	90%	96%	0.0	金黄色葡萄球菌 <i>Thermosynechococcus vulcanus</i>	光系统 Q(B)蛋白 Photosystem Q (B) protein
3A0B	D	95%	91%	0.0	金黄色葡萄球菌 <i>Thermosynechococcus vulcanus</i>	光系统 Q(B)蛋白 Photosystem Q (B) protein
3WU2	D	96%	90%	0.0	金黄色葡萄球菌 <i>Thermosynechococcus vulcanus</i>	光系统 Q(B)蛋白 Photosystem Q(B) protein

1.4 蛋白质的二级结构预测

通过 CFFSP 服务器预测蛋白质的二级结构, 分析其螺旋、折叠和旋转的比例 (图 1)。详细信息在表 5 中给出。

表 5 螺旋、旋转和折叠的组成

Table 5 Composition of helices, sheets, turns

模板	螺旋	旋转	折叠
Templates	Helices	Sheets	Turns
1IZL	36.8%	0.0	5.7%
4IL6	51.3%	0.0	3.4%
3A0B	55.2%	0.0	1.1%
3WU2	54.1%	0.0	6.2%

1.5 同源建模

光系统 II 的 D2 蛋白的三维结构是利用同源建模概念并选择四种不同的模板构建模型。利用 modeller 9.12 工具构建模型 (Bilal et al., 2013, Singh et al., 2009)。设定目标、模板和需要构建的模型数量, 运行 align2d.py, model-single.py 和 evaluate-model.py 文件的 Python 脚本。每个模板生成 5 个模型, 并根据 DOPE 得分最低选择最佳的模型。结构可视化后发现的属性在表 6 和表 7 中给出。蛋白质的三级结构如图 2、图 3、图 4、图 5 所示。

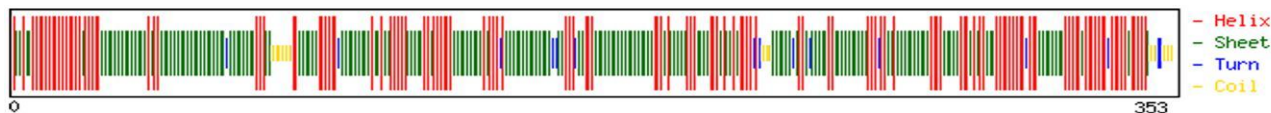


图 1 光系统 II D2 蛋白的二级结构

Figure 1 Secondary structure of Photosystem II D2 protein

表 6 从 yasara 工具得到的结果

Table 6 Result obtained from yasara tool

模板	β 系数	对象稳定性	最小能量	范德华半径
Templates	Beta factor	Stability of object	Minimized energy	VDW radius
1IZL	140.0	930.7kcal/mol	-0.737849	60.267\AA^0
4IL6	110.9	723.92 kcal/mol	-2.8532	53.888\AA^0
3A0B	105.4	607.54 kcal/mol	-2.21035	53.484\AA^0
3WU2	99.8	591.03 kcal/mol	-1.38159	52.549\AA^0

表 7 Frompymol 工具获得的结果

Table 7 Result obtained frompymol tool

模板	原子数	形式电荷的总和	分子表面积	溶剂可比表面积
Templates	Atom count	Formal Charge sum	Molecular surface area	Solvent accessible surface area
1IZL	2804	-9.0	34263.043\AA^0	21350.953\AA^0
4IL6	2804	-9.0	34034.191\AA^0	21938.918\AA^0
3A0B	2804	-9.0	34332.879\AA^0	21257.391\AA^0
3WU2	2804	-9.0	34461.117\AA^0	21316.729\AA^0

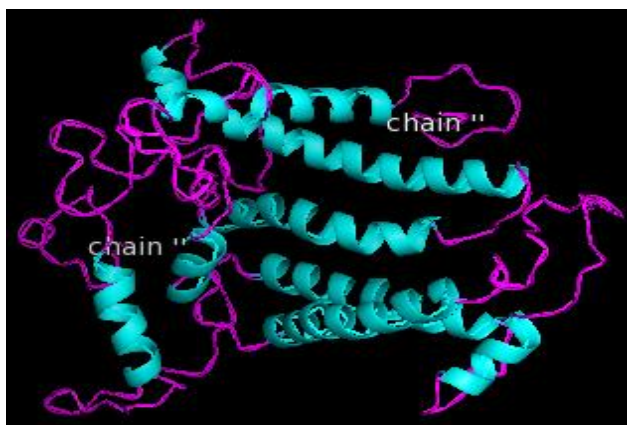


图 2 使用 1IZLD 生成的模型

Figure 2 model generated using 1IZLD



图 4 使用 3WU2 生成的模型

Figure 4 Model generated using 3WU2

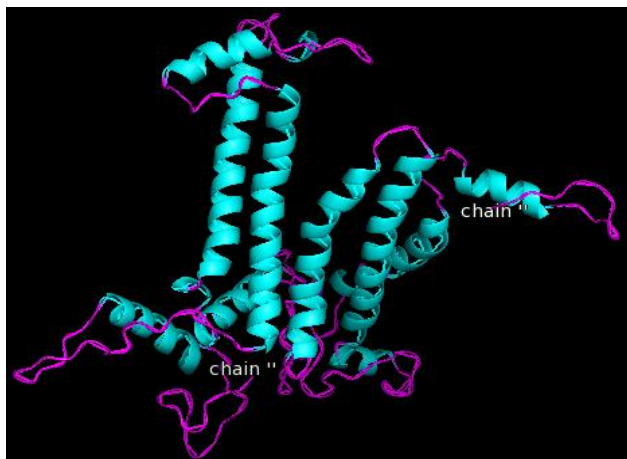


图 3 使用 3A0B 生成的模型

Figure 3 Model generated using 3A0B

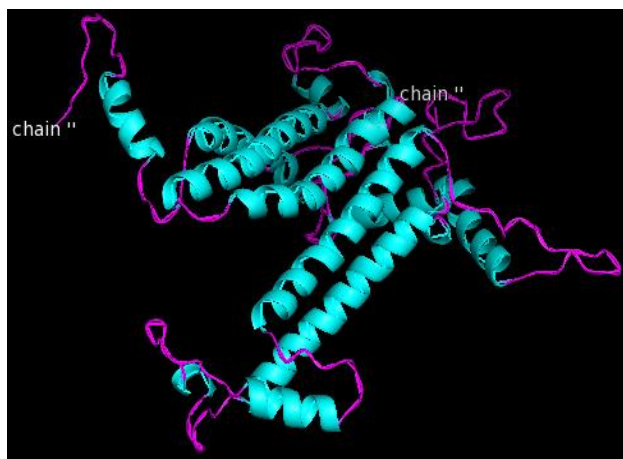


图 5 使用 4IL6 生成的模型

Figure 5 Model generated using 4IL6

1.6 模型验证

利用 Rampage 服务器进一步进行最终模型的蛋白质骨架确认分析。每个建成的模型的骨架确认显示了落在允许域，最适区和非允许区的残留物的数量。根据这些特征，选择最佳的模型。ANOLEA 服务器是用来找出 Z 值和 Q-mean 得分。Z 值最小表示最好的模型。已验证的模型信息和骨架确认在表 8 和表 9 中给出。蛋白质模型的骨架确认如图 6、图 7、图 8、图 9。

表 8 ANOLEA-SWISS SERVER 获得的结果

Table 8 Result obtained from ANOLEA-SWISS SERVER

模型	QMEAN 值	Z-值
Templates	QMEAN score	Z-score
1IZL	0.156	-7.194
4IL6	0.314	-5.356
3A0B	0.311	-5.382
3WU2	0.366	-5.439

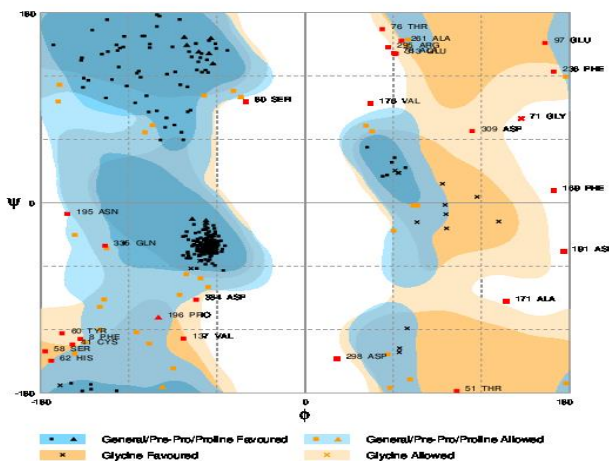


图 6 用 1IZL 模板的模型
Figure 6 Model with 1IZL template

表 9 模型的主干确认

Table 9 Backbone confirmation of models

模型	最适区的残基	允许区的残基	非允许区的残基
Templates	Residues in favoured region	Residues in allowed region	Residues in outlier region
1IZL	290 (82.6%)	35 (10.0%)	26 (7.4%)
4IL6	330 (94.0%)	14 (4.0%)	7 (2.0%)
3A0B	333 (94.9%)	12 (3.4%)	6 (1.7%)
3WU2	330 (94.0%)	20 (5.7%)	1 (0.3%)

1.7 选择最佳模型，

然而，最好的模型是依据序列同一性，疑似覆盖率，Z 值和 Q-mean 值，E 期望值等选择。

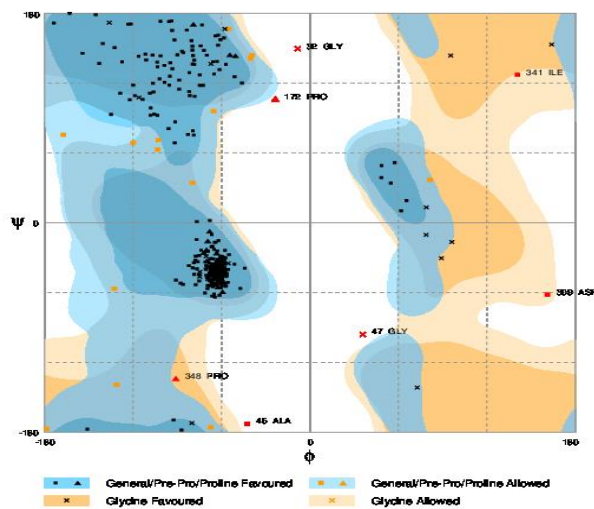


图 7 用 3A0B 模板的模型
Figure 7 Model with 3A0B template

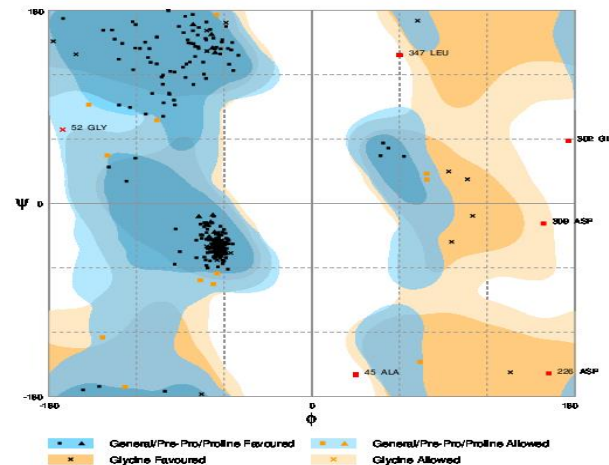


图 8 用 3WU2 模板的模型
Figure 8 Model with 3WU2 template

2 结果

2.1 序列检索结果

从 UniProt 上检索的光系统 II 的 D2 蛋白序列 ID 是 P06005。下面给出了功能、位置和催化活性。

2.2 理化性质分析结果

分析显示蛋白质的氨基酸组成、理论 PI、带正负电荷的残基的总数量, GRAVY 以及脂肪指数。

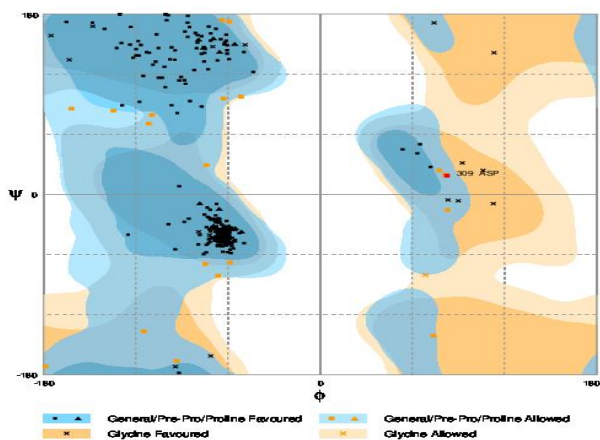


图 9 用 4IL6 模板的模型

Figure 9 Model with 4IL6 template

2.3 选择的模板

进行 blastP 后, 选择 1IZL, 4IL6, 3A0B, 3WU2 这四个引物作为有以下性质的蛋白质的模板。为了找到一个合适的模板的蛋白质的最佳模型, 我们选择了更多的模板。

2.4 二级结构预测结果

利用 CFFSP 服务器产生的蛋白质的二级结构显示了以下结果。

2.5 同源建模结果

利用 PyMol 可视化工具使最后生成的模型可视化。分别用天蓝色表示螺旋, 紫色表示循环。利用 PyMol 分析模型的原子数的总和, 形式电荷, 分子表面积和溶剂可比表面积, 用 Yasara 工具进行 β 因子, 模型稳定性, VDW 半径, 最小化能量分析。

2.6 模型验证分析结果

最后生成的模型在 Rampage 服务器中寻找最

佳的蛋白质。最佳的蛋白质是根据落在允许区内的氨基酸残基进行预测。

3 讨论

从以上分析发现, 光系统 II 的 D2 蛋白的最佳模型是以模板 3A0B 构建。3A0B 模板与 *Thermosynechococcus vulcanus* 光系统 Q (B) 蛋白有 91% 的疑似覆盖率, 即有 94.0% 残基落在允许区内, 2.0% 的残基落在非允许区, 以及有 95% 的序列同一性。

参考文献

- Anugolu et al., 2011, Homology modeling and ligand interaction of Cytochrome b protein, *Journal of Medical and Allied Science*, 1(2): 79-83
- Bansal et al., 2014, Computational characterization of antifreeze proteins of *Typhula ishikariensis*, *Gray Snow Mould*, JPPR9012014d
- Bergman et al., 2011, The anti-oxidant activity of aqueous spinach extract: chemical identification of active fractions, *Phytochemistry journal*, 58(1): 143-152
- Bilal et al., 2013, Generation of a 3D model for human cereblon using comparative modeling, *Journal of Bioinformatics and Sequence Analysis*, 5(1): 10-15
- Lomnitski et al., 2000, Effects of apocynin and natural antioxidant from spinach on inducible nitric oxide synthase and cyclooxygenase-2 induction in lipo polysaccharide-induced hepatic injury in rat. *Pharmacology & toxicology journal*, 87(1): 18-25b
- Panda et al., 2014, In silico predictive studies of mAHR congener binding using homology modelling and molecular docking, *Toxicol Ind Health*, 30(8): 765-766
- Sani et al., 2004, Potential anticancer effect of red spinach (*Amaranthus gangeticus*) extract, *Asia Pac J Clin Nutr journal*, 13(4): 396-400
- Singh et al., 2009, Comparative modeling and analysis of 3-D structure of Hsp 70, in *Cancer irroratus An International Journal*, 1(2): 1-4