**实验一：海洋生物功能基因生物信息学分析**

**1. 基因全长序列分析：**通过NCBI的Open Reading Frame Finder分析可知，给定的Grx基因序列全长为270bp，并能编码一个包含89个氨基酸的蛋白质(见图1-1)。



图1-1 Grx基因的全长序列及推测的氨基酸序列

**2. 蛋白理化性质预测与分析：**根据在线服务系统ExPASy中的ProtParam工具和Proscale工具进一步对Grx基因蛋白氨基酸序列的基本理化性质进行综合预测分析。预测结果显示：Grx蛋白的分子量是9.8kDa(9837.1)；理论等电点pI 5.15；原子组成是C433H677N119O133S5，总原子数为1367；……；且Grx为亲水性蛋白（见图1-2）。

Number of amino acids: 89

Molecular weight: 9838.1

Theoretical pI: 5.15

Amino acid composition:

Ala (A) 8 9.0%

Arg (R) 4 4.5%

Asn (N) 5 5.6%

Asp (D) 6 6.7%

Cys (C) 2 2.2%

Gln (Q) 6 6.7%

Glu (E) 4 4.5%

Gly (G) 7 7.9%

His (H) 1 1.1%

Ile (I) 4 4.5%

Leu (L) 8 9.0%

Lys (K) 4 4.5%

Met (M) 3 3.4%

Phe (F) 4 4.5%

Pro (P) 4 4.5%

Ser (S) 6 6.7%

Thr (T) 3 3.4%

Trp (W) 1 1.1%

Tyr (Y) 3 3.4%

Val (V) 6 6.7%

Pyl (O) 0 0.0%

Sec (U) 0 0.0%

Total number of negatively charged residues (Asp + Glu): 10

Total number of positively charged residues (Arg + Lys): 8

Formula: C433H677N119O133S5

Total number of atoms: 1367

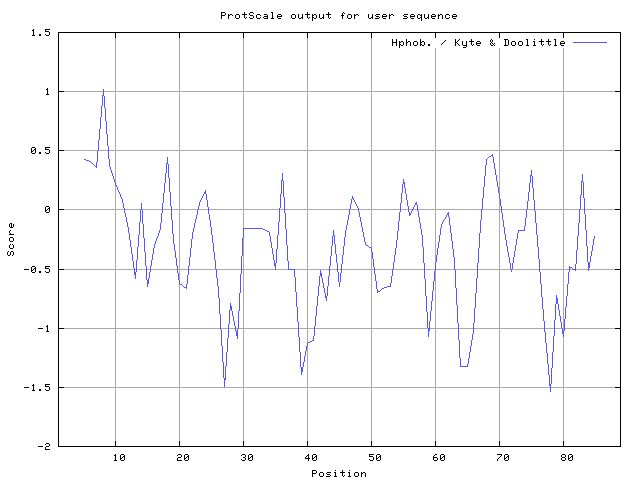


图1-2 Grx蛋白亲疏水特性分析

**3. 磷酸化位点预测与分析：**应用在线服务NetPhos 2.0 Serve对Grx基因的磷酸化位点进行预测与分析。对蛋白序列中的Ser、Thr和Tyr三种氨基酸残基可能成为的磷酸化位点作出预测，分析位点见表1-1、1-2、1-3结合图1-3可知道，整个多肽链中分值在0.5以上的氨基酸位点有1个，即29位的Tyr位点可能成为蛋白激酶磷酸化位点。

表1-1 Serine predictions

|  |  |  |  |
| --- | --- | --- | --- |
| Pos | Context | Score | Pred |
| 2 | ---MSNVVL | 0.002 | - |
| 8 | VVLYSKAYC | 0.003 | - |
| 24 | ALLDSKGVQ | 0.374 | - |
| 51 | KAGGSRTVP | 0.072 | - |
| 70 | FWDDSIAVE | 0.007 | - |
| 80 | QGQLSKPLN | 0.142 | - |

表1-2 Threonine predictions

|  |  |  |  |
| --- | --- | --- | --- |
| Pos | Context | Score | Pred |
| 30 | GVQYTNFDI | 0.022 | - |
| 53 | GGSRTVPQM | 0.018 | - |
| 86 | PLNATLNA- | 0.010 | - |

表1-3 Tyrosine predictions

|  |  |  |  |
| --- | --- | --- | --- |
| Pos | Context | Score | Pred |
| 7 | NVVLYSKAY | 0.072 | - |
| 11 | YSKAYCPFC | 0.065 | - |
| 29 | KGVQYTNFD | 0.753 | \*Y\* |

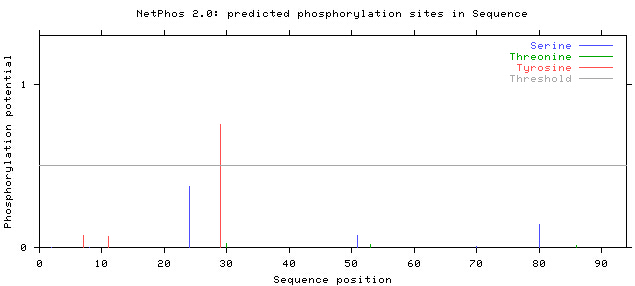


图1-3 Grx氨基酸序列翻译后的磷酸化修饰预

**4. 蛋白卷曲螺旋结构的预测与分析：**利用在线服务Coils分析工具对Grx基因蛋白序列形成卷曲螺旋的倾向性进行预测，以window=14，21和28为实验参数，按照几率>50%就可形成螺旋的规则，进行分析。得出结论：在Grx蛋白氨基酸序列不可能形成卷曲螺旋(图1-4）。

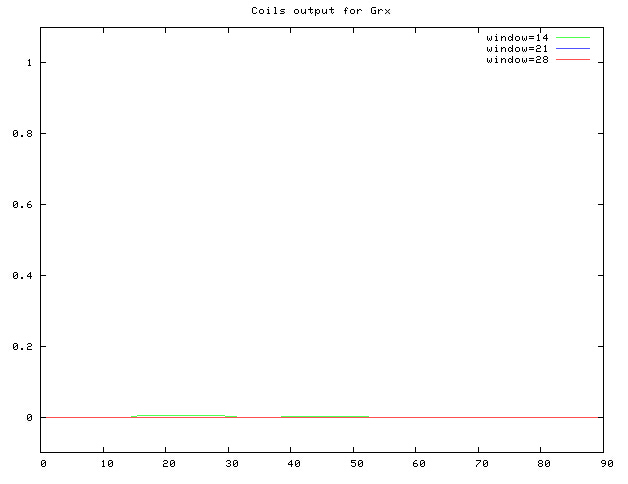


图1-4 Grx 蛋白氨基酸序列卷曲螺旋的预测

**5. 蛋白质二级结构预测与分析：**应用JPred和PredictProtein预测服务器对Grx基因蛋白进行二级结构预测。结果如图1-5所示，Grx是由4个α-螺旋和4个β-折叠组成。按组分分，Grx蛋白由46%的α螺旋、20%的延伸链和34%的不规则卷曲组成。纵观蛋白的整体结构，α螺旋和不规则卷曲是Grx蛋白最大量的结构元件，散布于整个蛋白质中。从中也能看出，Grx氨基酸序列中有7个GSH结合位点(G-site)(Lys9, Cys12, Pro13, Phe14, Gly50, Thr53,和Val54)和2个催化残基(Cys12和Cys15)。

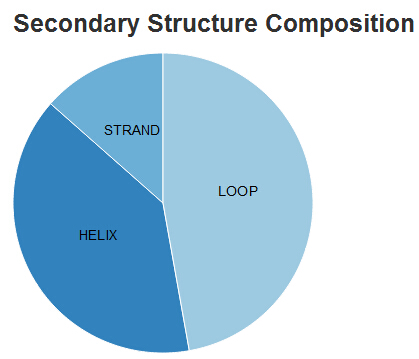
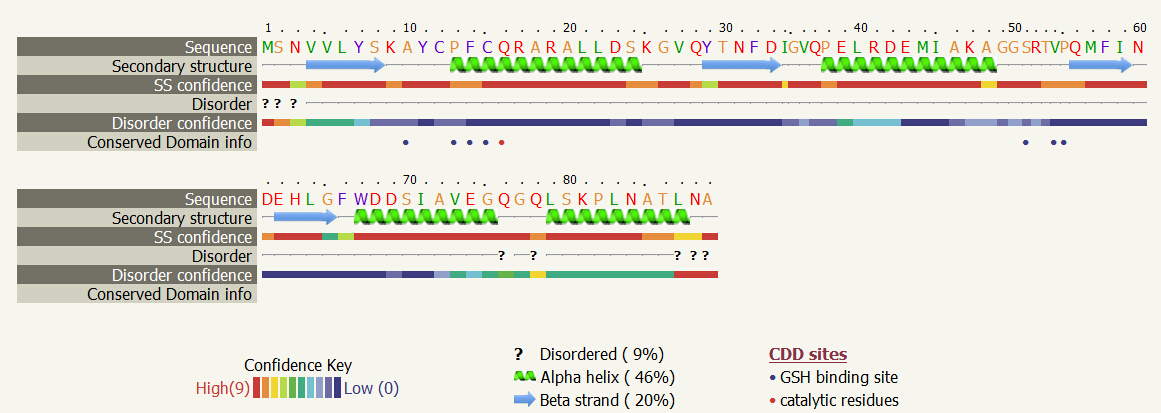
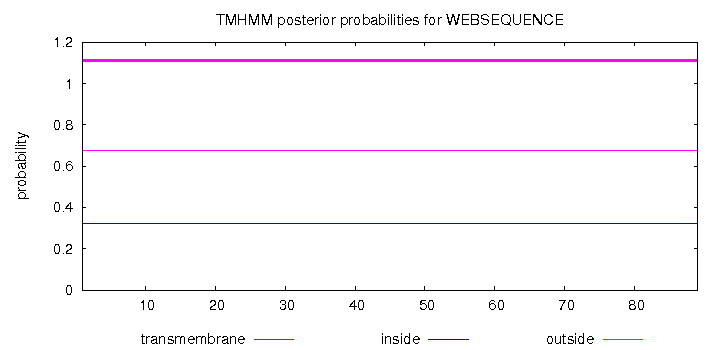


图1-5 Grx蛋白序列二级结构预测结果

**6 蛋白的跨膜结构预测与分析：**应用在线服务TMHMM和Tmpred对目的基因蛋白的跨膜结构进行预测及分析。红色表示跨膜区，蓝色即在膜内部，相反紫色细线表示在膜外的概率。可知Grx无明显跨膜区，不可能是膜上的受体或定位于膜上，属非跨膜蛋白，不发生跨膜运动，完全在细胞内起作用。



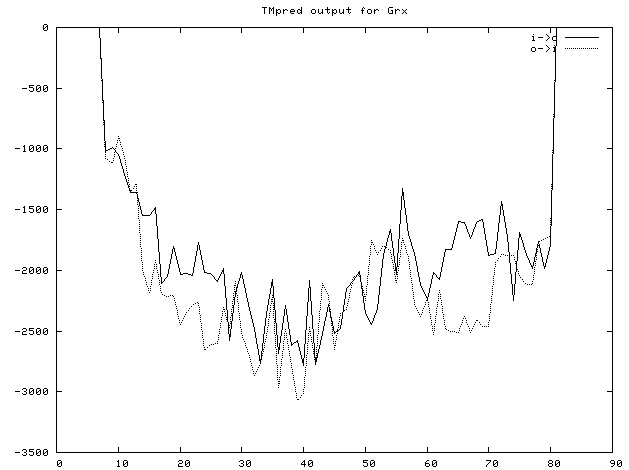


图1-6 Grx蛋白跨膜区的预测

**7. 信号肽的预测和分析：**应用在线服务SignaIP 4.1 Server对目的基因的信号肽进行预测与分析。得到的数值如表1-4所示，结合图1-7的我们可以发现，Grx没有明显的信号肽，这说明这一蛋白是分泌性蛋白的可能性并不大，是一胞质酶，在细胞质中直接作用于底物。

表1-4 Grx氨基酸序列信号肽分析

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Measure | Position | Value | Cut off | signal peptide? |
| max. C | 21 | 0.154 |  |  |
| max. Y | 21 | 0.176 |  |  |
| max. S | 1 | 0.376 |  |  |
| mean S | 1-20 | 0.247 |  |  |
| D | 1-20 | 0.209 | 0.570 | NO |
| SP='NO' D=0.209 D-cutoff=0.570 Networks=SignalP-noTM | | | | |

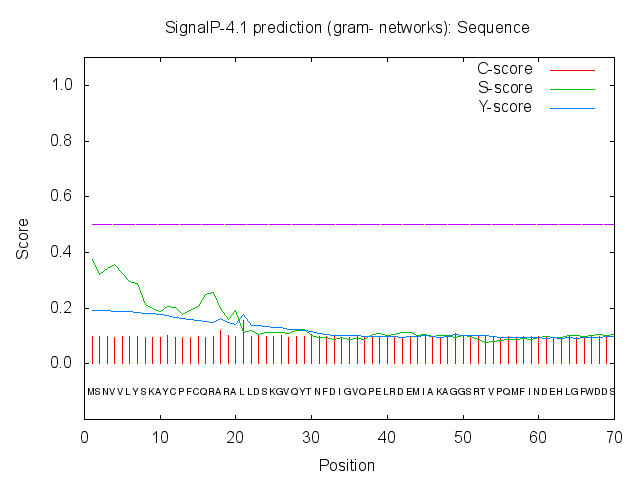


图1-7 Grx蛋白的信号肽预测结果

**8. 亚细胞定位预测与分析：**应用PSORT软件对目的基因蛋白的亚细胞内定位进行预测。预测结果为：

----- Final Results -----

bacterial cytoplasm --- Certainty= 0.287(Affirmative)

bacterial periplasmic space --- Certainty= 0.000(Not Clear)

bacterial outer membrane --- Certainty= 0.000(Not Clear)

bacterial inner membrane --- Certainty= 0.000(Not Clear)

----- The End -----

可见，Grx在细胞质基质中的可能性最大，因此可以认为Grx蛋白是一种细胞质基质蛋白，主要定位在细胞质基质中。

**9. 三维结构的预测和分析：**利用Phyre在线工具对目的基因氨基酸序列进行蛋白质三维结构预测。结果如图1-8所示，可以发现Grx三维结构主要由α螺旋和不规则卷曲组成，其中有4个α-螺旋和4个β-折叠，在结构中间有一个较大的凹陷，这可能是此Grx的催化活性部位，通过这个中心，蛋白与底物接触。

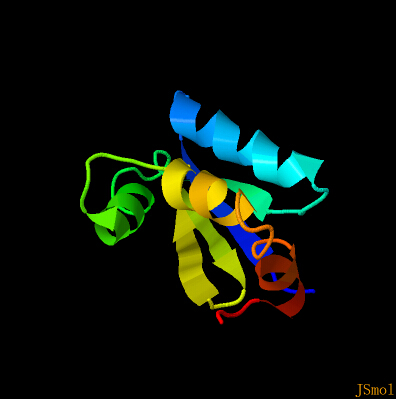
 

图1-8 Grx蛋白预测的三级结构预测

**10. 保守结构域与功能域分析：**根据NCBI-CDS对目的基因蛋白的保守结构域与功能域进行在线分析。在线分析表明：Grx氨基酸序列中有7个GSH结合位点(G-site)(Lys9, Cys12, Pro13, Phe14, Gly50, Thr53,和Val54)和2个催化残基(Cys12和Cys15)，Grx可能是一种属于类Trx超家族的新的Grx基因（见图1-9）。

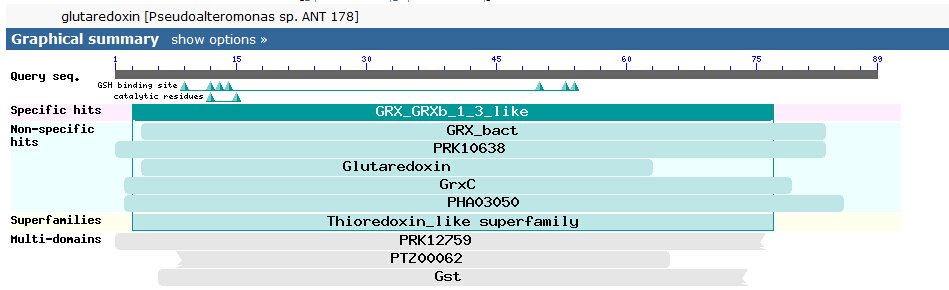
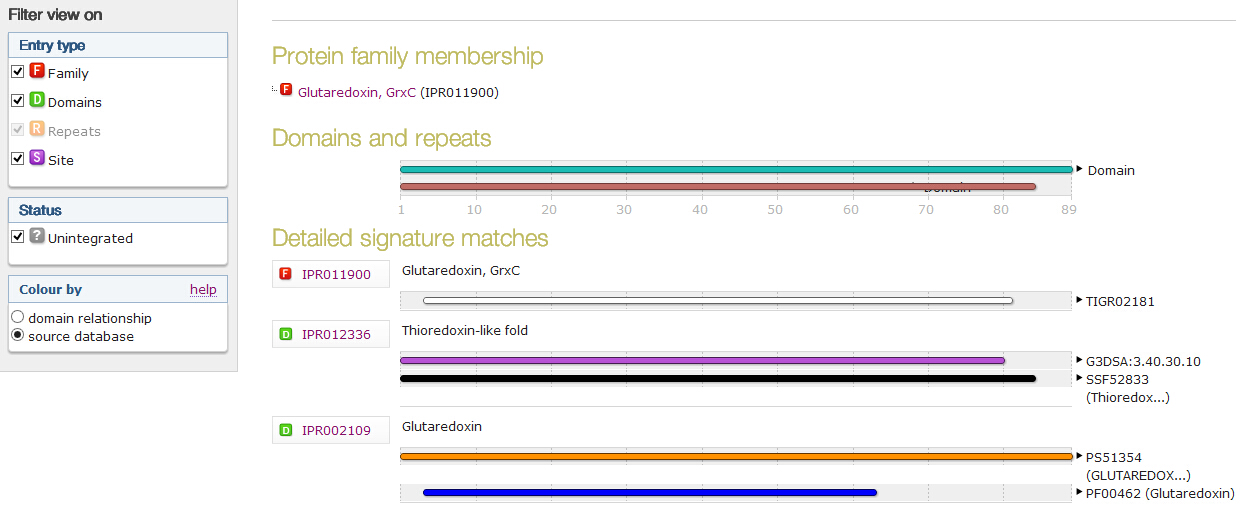


图1-9 NCBI上Grx蛋白氨基酸序列的保守结构域检索

**11. 同源蛋白质家族比较分析**：应用InterPro数据库进行同源蛋白质分析比较。发现Grx基因的开放阅读框编码的89个氨基酸残基的蛋白质是硫氧还蛋白折叠家族（Thioredoxin-like fold）的成员。



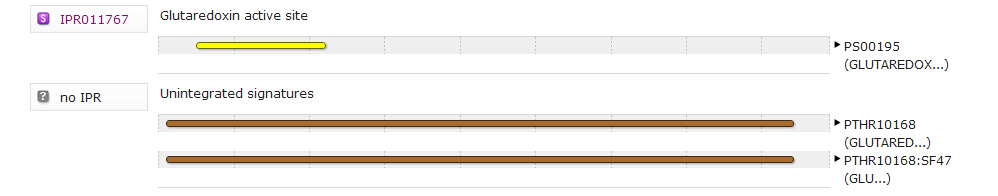


图1-10 Grx编码的蛋白质 Interpro 数据库搜索结果

**12.** **分子系统发育预测与分析：**对Grx的氨基酸序列进行BLAST分析，发现它和Trx超家族具有极高的相似度。与*P. haloplanktis* TAC 125 Grx (YP\_338909), *P. agarivorans* Grx (WP\_004588615), *P. haloplanktis* Grx (WP\_016708488), *P. undina* Grx (WP\_010391834), *R. nanhaiensis* Grx (WP\_008217797)氨基酸序列的相似度分别为85.6%, 74.4%, 74.4%, 72.2%和50.0%（见图1-11）。

*P.sp.*ANT178

*P.haloplanktis*

*P.agarivorans*

*P.haloplanktis*

*P.undina*

*R.nanhaiensis*

*P.sp.*ANT178

*P.haloplanktis*

*P.agarivorans*

*P.haloplanktis*

*P.undina*

*P.nanhaiensis*

50

50

50

50

50

50

89

89

85

85

85

85

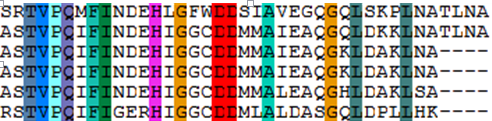


图1-11 Grx氨基酸序列与其他Grxs氨基酸序列的多重序列对齐分析

注：*P. haloplanktis* TAC 125 Grx (YP\_338909), *P. agarivorans* Grx (WP\_004588615), *P. haloplanktis* Grx (WP\_016708488), *P. undina* Grx (WP\_010391834), *R. nanhaiensis* Grx (WP\_008217797), ▲G位点; \*底物结合位点。

用MEGA5.2软件构建了系统进化树（图1-12）。其中与*P. haloplanktis.*TAC125的*Glutaminase*同源性最高，最大相似度高达99%，其次与*P.* *haloplanktis*的*Glutaminase*氨基酸序列最大相似度为92%，与*Pseudoalteromonas arctica* A 37-1-2的*Glutaminase*氨基酸序列最大相似度为91%，与*Serratia plymuthica* 4Rx13的亲缘关系最远，氨基酸序列相似度为57%。可以初步证实*Glutaminase*.ANT178属于假单胞交替菌属。

发育树2

图1-12 Glutminase与其它Gluts的系统进化关系

**13.** **基因功能的预测与分析**：应用在线软件InterPro对目的蛋白参与的生物过程、分子功能进行预测与分析。从图1-13可以看出，Grx蛋白参与的生物过程为细胞氧化还原反应的平衡；分子功能为：电子载体活性；蛋白质二硫氧化还原酶活性。

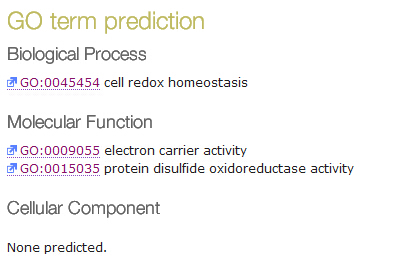


图1-13 Grx蛋白功能预测