3D modelling and functional annotation of the Transglutaminase from Arabidopsis thaliana

Matteo Bolner

February 4, 2019

Abstract

Q9FGY9 is a gene that codifies for a protein of the plant *Arabidopsis thaliana* which, after manual annotation of the transcript, is believed to be a Peptide-N(4)-(N-acetyl-beta-glucosaminyl)asparagine amidase (PNGase). However, upon further investigation it appears that the annotation might be wrong: in this project, i would like to verify whether the annotation as a PNGase can be disproved in favor of an annotation as a Transglutaminase. The conservation of the function allowed me to also infer structural similarity of the catalytic domain.

1 Introduction

Transglutaminases are proteins that catalyze the crosslinking of substrate proteins through the formation of gamma-glutamyl-epsilon-lysyl amide bonds, while PNGases are involved in the degradation of misfolded N-linked glycoproteins, through the cleaving between the innermost GlcNAc and asparagine residues of N-linked glycoproteins. Both proteins contain a highly conserved catalytic triad (Cys-His-Asp) in the active site; however, the spatial organization of the surface of the proteins, and the clefts in which the substrates can interact with the active site differ significantly.
Mammalian transglutaminases are calcium-dependent proteins, indispensable for the formation of stable structures; many maladies, such as celiac disease and predisposition to hemorrhage are linked to the deficiency of particular transglutaminases. Transglutaminases are also found in plants: after the discovery of transglutaminase activity in plant cells, a gene encoding for the protein (AtPng1p)was found, and its TGase activity experimentally confirmed [1].
In UniProtKB/Swiss-Prot Q9FGY9 is manually annotated but not endowed with a 3D structure. In order to verify the possibility of annotating it as a Transglutaminase, I verified the conservation of the critical residues involved in the activity of the protein by comparison with a human transglutaminase, referring to a review article in which the same problem is faced [2].

2 Methods

2.1 Data bases

The Q9FGY9 target sequence is obtained from the UniProt database (release 2018_11)[3] ; the 1KV3 template sequence and structure are obtained from the PDB file in RCSB PDB (18.12.18 release)[4]. The GO terms are obtained from the UniProt page associated to 1KV3. Information in relation to the protein domains and secondary structure is derived from InterPro (v.72, 17.01.19)[5].

2.2 Computational methods

The target sequence (Q9FGY9) is locally aligned with the UNIPROTKB_PDB database with BLASTP (2.7.1+)[6], using the parameters highlighted in Table 1.

Table 1: Parameters used to retrieve the template with BLAST

Program	BLASTP 2.7.1+
Target database	uniprotkb_pdb
E-Threshold	10
Matrix	Blosum62(Auto)
Filtering	None
Gapped	Yes
Hits	50

In order to compare the target and template sequences, I run Clustal Omega (1.2.4)[7] for multiple sequence alignment with the default parameters.

Modeller (release 9.20, r11208) is downloaded from https://salilab.org/modeller [8] and then used to generate five model structures of the target. Pymol is adapted as the molecular visualizer (Version 2.2.0)[9]. The online version of PROCHECK 3.5 [10] analyzes the stereochemical quality of the model. The computed target structure is aligned with the template structure with jCE (V2.11 2014 March 17)[11] installed locally.

3 Template selection

The procedure of building by homology requires the identification of templates within the family. In order to locate the protein family, the first operation is to browse the target sequence against the UniprotKB_PDB with BLASTP.

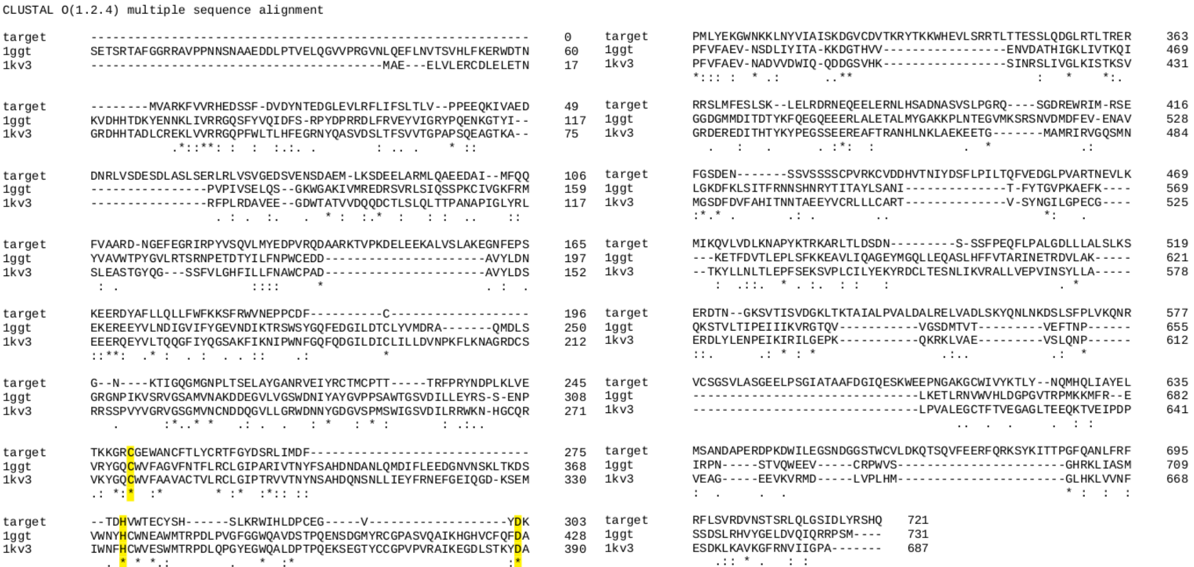
The results of BLASTP are conflicting: there are no entries with a high sequence identity, and the only manually reviewed sequence (Q9FGY9) puts the protein in the PNGase family, while the automatically annotated sequence A5PHD1, with the same sequence, puts it in the transglutaminase family. The highest sequence identities lie with the PNGase proteins, but a brief consultation of the available literature [1] confirms the annotation as a transglutaminase. Since there are no experimentally determined structures for plant transglutaminases or PNGases, and especially for the target sequence, it is not possible to confirm the hypothesis with experimental data. However, it can be implied that the evolution of the protein meant a major change in sequence identity between plant and animal transglutaminases, while the structure and function were conserved.

Since in article [1] the functionality of AtPng1P as a transglutaminase is experimentally demonstrated, I start from the assumption that the sequence belongs to a transglutaminase. Therefore, I must choose a well documented transglutaminase structure with an experimentally determined electron density as template. After a thorough search on UniProt for transglutaminase proteins with an experimentally determined structure, I find the human tissue transglutaminase protein (P21980). The sequence of Human tissue transglutaminase is long enough (687 aa) to accomodate almost the whole length of the target sequence (721 aa). The chosen PDB structure for this protein is 1KV3, obtained with a 2.8 Å resolution. Since it covers most of the sequence and in particular the whole domain containing the active site, it is deemed a good template structure.

In order to highlight the possible conservation of the most important residues in the target sequence, another transglutaminase is chosen to build a multiple sequence alignment; with the support of the literature available I choose the human coagulation factor XIII (P00488). Since the 1GGT structure is well documented, and the sequence length of 721 aa covers most of the target sequence, I decide to use it for the sequence alignment.

4 Sequence alignment

Figure 1 shows the multiple sequence alignment of the target sequence with two templates. The sequence identity is 17%; the residues of the catalytic triad (Cys-277, His-335 and Asp-358)are conserved.



```
CLUSTAL O(1.2.4) multiple sequence alignment

target          ------------------------------------------------------------      0
1ggt            SETSRTAFGGRRAVPPNNSNAAEDDLPTVELQGVVPRGVNLQEFLNVTSVHLFKERWDTN      60
1kv3            ----------------------------------------MAE---ELVLERCDLELETN      17


target          --------MVARKFVVRHEDSSF-DVDYNTEDGLEVLRFLIFSLTLV--PPEEQKIVAED      49
1ggt            KVDHHTDKYENNKLIVRRGQSFYVQIDFS-RPYDPRRDLFRVEYVIGRYPQENKGTYI--      117
1kv3            GRDHHTADLCREKLVVRRGQPFWLTLHFEGRNYQASVDSLTFSVVTGPAPSQEAGTKA--      75
                           .*::**: :  :  :.:. .        : .. .    * ::

target          DNRLVSDESDLASLSERLRLVSVGEDSVENSDAEM-LKSDEELARMLQAEEDAI--MFQQ      106
1ggt            ----------------PVPIVSELQS--GKWGAKIVMREDRSVRLSIQSSPKCIVGKFRM      159
1kv3            ----------------RFPLRDAVEE--GDWTATVVDQQDCTLSLQLTTPANAPIGLYRL      117
                                 . : .  :.   .  * :  :.*   :   : :  ..    ::

target          FVAARD-NGEFEGRIRPYVSQVLMYEDPVRQDAARKTVPKDELEEKALVSLAKEGNFEPS      165
1ggt            YVAVWTPYGVLRTSRNPETDTYILFNPWCEDD---------------------AVYLDN      197
1kv3            SLEASTGYQG---SSFVLGHFILLFNAWCPAD---------------------AVYLDS      152
                 : .                  ::::     *                      . :  .

target          KEERDYAFLLQLLFWFKKSFRWVNEPPCDF----------C-------------------      196
1ggt            EKEREEYVLNDIGVIFYGEVNDIKTRSWSYGQFEDGILDTCLYVMDRA-------QMDLS      250
1kv3            EEERQEYVLTQQGFIYQGSAKFIKNIPWNFGQFQDGILDICLILLDVNPKFLKNAGRDCS      212
                ::**:  .* :  . :  . . ::    .:          *

target          G--N----KTIGQGMGNPLTSELAYGANRVEIYRCTMCPTT-----TRFPRYNDPLKLVE      245
1ggt            GRGNPIKVSRVGSAMVNAKDDEGVLVGSWDNIYAYGVPPSAWTGSVDILLEYRS-S-ENP      308
1kv3            RRSSPVYVGRVGSGMVNCNDDQGVLLGRWDNNYGDGVSPMSWIGSVDILRRWKN-HGCQR      271
                   .      :*..* *   .: .  .   : *   : * :       : .:..

target          TKKGRCGEWANCFTLYCRTFGYDSRLIMDF------------------------------      275
1ggt            VRYGQCWVFAGVFNTFLRCLGIPARIVTNYFSAHDNDANLQMDIFLEEDGNVNSKLTKDS      368
1kv3            VKYGQCWVFAAVACTVLRCLGIPTRVVTNYNSAHDQNSNLLIEYFRNEFGEIQGD-KSEM      330
                .: *:*  :*       * :*  :*:: ::

target          --TDHVWTECYSH------SLKRWIHLDPCEG-----V------------------YDK      303
1ggt            VWNYHCWNEAWMTRPDLPVGFGGWQAVDSTPQENSDGMYRCGPASVQAIKHGHVCFQFDA      428
1kv3            IWNFHCWVESWMTRPDLQPGYEGWQALDPTPQEKSEGTYCCGPVPVRAIKEGDLSTKYDA      390
                  . * * *.:        .   *  :*                             :*

target          PMLYEKGWNKKLNYVIAISKDGVCDVTKRYTKKWHEVLSRRTLTTESSLQDGLRTLTRER      363
1ggt            PFVFAEV-NSDLIYITA-KKDGTHVV-----------------ENVDATHIGKLIVTKQI      469
1kv3            PFVFAEV-NADVVDWIQ-QDDGSVHK-----------------SINRSLIVGLKISTKSV      431
                *::: :  *  .:      ..**                         :   *    *:.

target          RRSLMFESLSK--LELRDRNEQEELERNLHSADNASVSLPGRQ----SGDREWRIM-RSE      416
1ggt            GGDGMMDITDTYKFQEGQEEERLALETALMYGAKKPLNTEGVMKSRSNVDMDFEV-ENAV      528
1kv3            GRDEREDITHTYKYPEGSSEEREAFTRANHLNKLAEKEETG-------MAMRIRVGQSMN      484
                  .   :   .      . :*: :            .  *            .:

target          FGSDEN-------SSVSSSSCPVRKCVDDHVTNIYDSFLPILTQFVEDGLPVARTNEVLK      469
1ggt            LGKDFKLSITFRNNSHNRYTITAYLSANI-------------T-FYTGVPKAEFK----      569
1kv3            MGSDFDVFAHITNNTAEEYVCRLLLCART-------------V-SYNGILGPECG----      525
                :*.* .       .: .        ..                     *:    .

target          MIKQVLVDLKNAPYKTRKARLTLDSDN---------S-SSFPEQFLPALGDLLLALSLKS      519
1ggt            ---KETFDVTLEPLSFKKEAVLIQAGEYMGQLLEQASLHFFVTARINETRDVLAK-----      621
1kv3            --TKYLLNLTLEPFSEKSVPLCILYEKYRDCLTESNLIKVRALLVEPVINSYLLA-----      578
                   :  .::.  * . :.  : :   :                       . *

target          ERDTN--GKSVTISVDGKLTKTAIALPVALDALRELVADLSKYQNLNKDSLSFPLVKQNR      577
1ggt            QKSTVLTIPEIIIKVRGTQV-----------VGSDMTVT---------VEFTNP------      655
1kv3            ERDLYLENPEIKIRILGEPK-----------QKRKLVAE---------VSLQNP------      612
                ::.      .: * : *                 .:..           .:  *

target          VCSGSVLASGEELPSGIATAAFDGIQESKWEEPNGAKGCWIVYKTLY--NQMHQLIAYEL      635
1ggt            ------------------------------LKETLRNVWVHLDGPGVTRPMKKMFR--E      682
1kv3            ------------------------------LPVALEGCTFTVEGAGLTEEQKTVEIPDP      641
                                                    ..  .  .     .  : :

target          MSANDAPERDPKDWILEGSNDGGSTWCVLDKQTSQVFEERFQRKSYKITTPGFQANLFRF      695
1ggt            IRPN-----STVQWEEV-----CRPWVS----------------------GHRKLIASM      709
1kv3            VEAG-----EEVKVRMD-----LVPLHM----------------------GLHKLVVNF      668
                :  .     .  .                                      * :  :  :

target          RFLSVRDVNSTSRLQLGSIDLYRSHQ    721
1ggt            SSDSLRHVYGELDVQIQRRPSM----    731
1kv3            ESDKLKAVKGFRNVIIGPA-------    687
                   .:: * .   : :
```

Figure 1: Alignment of the target and template sequences with Clustal Omega; in yellow are highlighted the conserved residues of the catalytic triad

5 Modeller at work

5.1 Input preparation

From the multiple sequence alignment, the 1GGT template was removed, and the remaining alignment was converted into .pir format by following Modeller's instructions. The number of models to produce was set to 5 in the Modeller script.

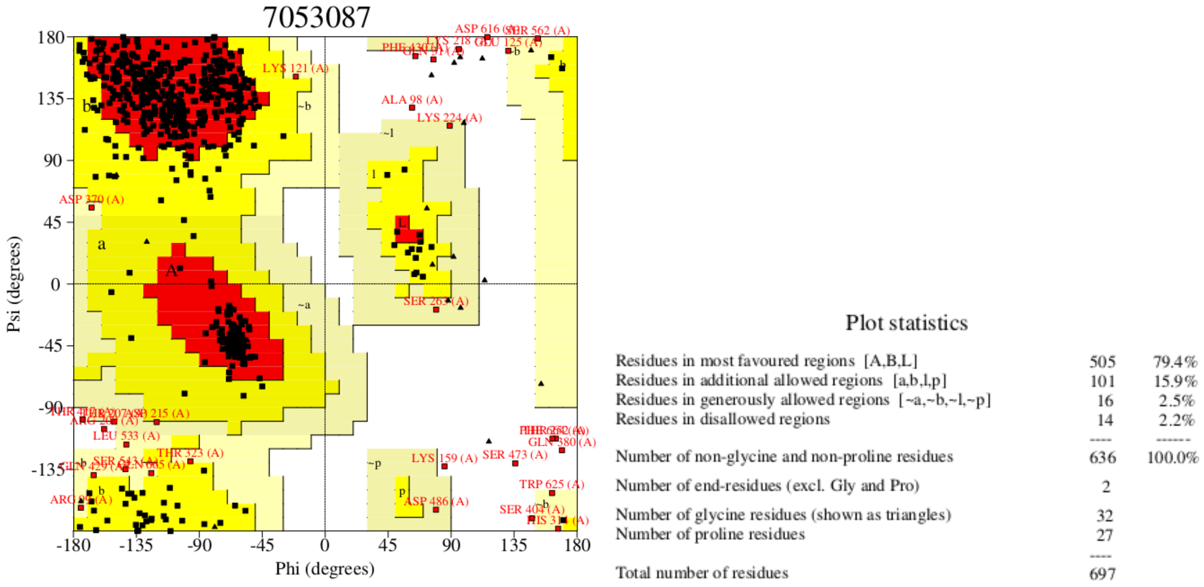
5.2 Modeller output

Table 2 lists the scores of the five different models obtained: molpdf (molecular probability density function), DOPE (discrete optimized protein energy) and GA341 (score that uses the percentage sequence identity between the template and the model as a parameter) . The scores are computed according to Modeller's internal validation procedure. TARGET.B99990003, having the lowest molpdf score and the second highest DOPE score is adopted as the final result for modelling the structure of the target. TARGET.B99990003.pdb was rejected by modeller and is therefore not included in the final output.

Table 2: Modeller output table

Filename	molpdf	DOPE score	GA341 score
TARGET.B99990001.pdb	9116.55469	-47185.06641	0.28854
TARGET.B99990003.pdb	6837.61426	-48852.23828	0.11289
TARGET.B99990004.pdb	7227.91602	-48115.46094	0.18775
TARGET.B99990005.pdb	7916.51855	-49156.42969	0.12151

The model stability was also evaluated with PROCHECK. According to PROCHECK, based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions. As shown in the ramachandran plot (Figure 2) the model should be mostly stable, with 79.4% of the residues having the proper torsion angles and being accordingly in the most favoured regions of the main chain backbone, and only 2.2% of the residues in the disallowed regions. Since this model was built on a very low sequence identity alignment, these results are to be expected.



Plot statistics

Residues in most favoured regions [A,B,L]	505	79.4%
Residues in additional allowed regions [a,b,l,p]	101	15.9%
Residues in generously allowed regions [~a,~b,~l,~p]	16	2.5%
Residues in disallowed regions	14	2.2%
	----	------
Number of non-glycine and non-proline residues	636	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	32	
Number of proline residues	27	

Total number of residues	697	

Figure 2: Ramachandran plot and statistics of model 2

6 Target annotation and discussion

The template and model pdb structures were aligned with jCE; the resulting pdb file was analyzed using Pymol.

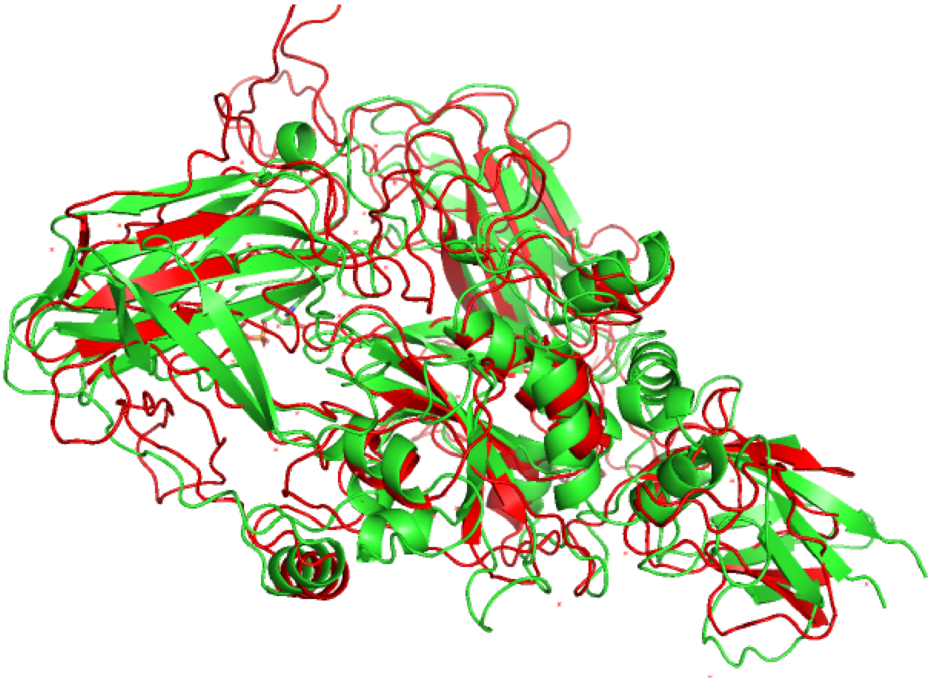
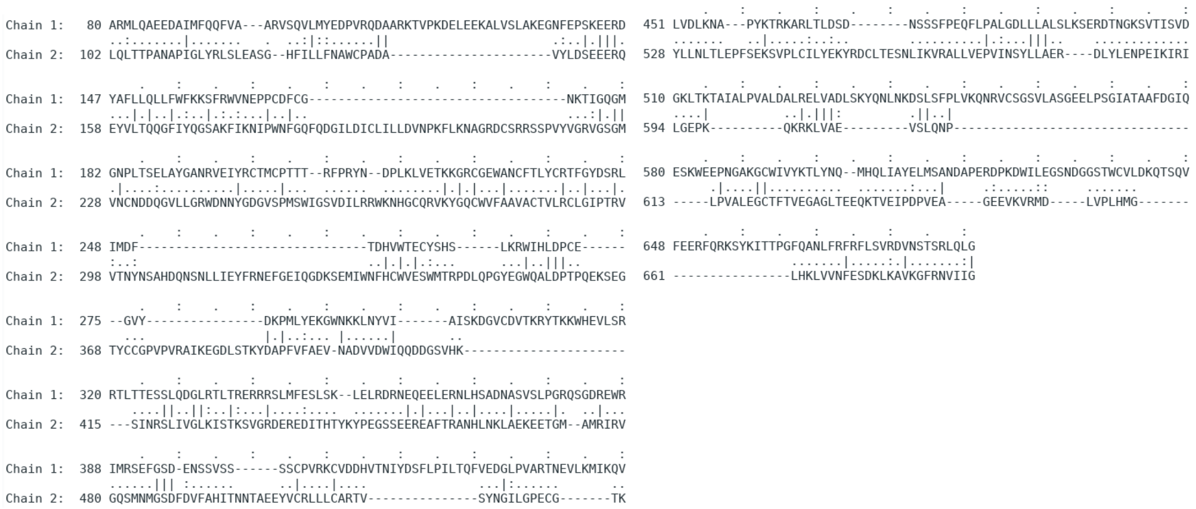


Figure 3: Structural comparison of template(red) and model(green)

Figure 3 shows the comparison of the template structure with the target model; the figure indicates that some secondary structural motives are conserved, and the general topology is mostly similar. The RMSD calculated by JCE is 2.75 Å. Generally speaking, a sequence identity of less than 30% is not enough to obtain a reliable model through building by homology; threading procedures are more adapt to predict secondary structures for low sequence identities. However, having confirmed the function of the target sequence, I may use building by homology to derive a structure for the target using a known template: if the function is conserved, I can infer that the structure of the transglutaminase domain is also conserved.



```
                                                                                           .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:   80 ARMLQAEEDAIMFQQFVA---ARVSQVLMYEDPVRQDAARKTVPKDELEEKALVSLAKEGNFEPSKEERD   451 LVDLKNA---PYKTRKARLTLDSD--------NSSSFPEQFLPALGDLLLALSLKSERDTNGKSVTISVD
              ..:.......|.......   .  ..:|::......||                      .:..|.|||.       .......   ..|.....:..:..        ..........|.:...|||..    .............
Chain 2:  102 LQLTTPANAPIGLYRLSLEASG--HFILLFNAWCPADA----------------------VYLDSEEERQ   528 YLLNLTLEPFSEKSVPLCILYEKYRDCLTESNLIKVRALLVEPVINSYLLAER----DLYLENPEIKIRI

                  .    :    .    :    .    :    .    :    .    :    .    :    .    :           .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:  147 YAFLLQLLFWFKKSFRWVNEPPCDFCG-----------------------------------NKTIGQGM   510 GKLTKTAIALPVALDALRELVADLSKYQNLNKDSLSFPLVKQNRVCSGSVLASGEELPSGIATAAFDGIQ
              ...|.|..|.:..|.:.:...|..|..                                   ...:|.||       ....|          ..|.|||:         .||..|
Chain 2:  158 EYVLTQQGFIYQGSAKFIKNIPWNFGQFQDGILDICLILLDVNPKFLKNAGRDCSRRSSPVYVGRVGSGM   594 LGEPK----------QKRKLVAE---------VSLQNP--------------------------------

                  .    :    .    :    .    :    .    :    .    :    .    :    .    :           .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:  182 GNPLTSELAYGANRVEIYRCTMCPTTT--RFPRYN--DPLKLVETKKGRCGEWANCFTLYCRTFGYDSRL   580 ESKWEEPNGAKGCWIVYKTLYNQ--MHQLIAYELMSANDAPERDPKDWILEGSNDGGSTWCVLDKQTSQV
              .|...:...........|.....|...  ......  ........|.|.|...|.......|..|...|.            .|....||..........  .......:...|     .:.....::     .......
Chain 2:  228 VNCNDDQGVLLGRWDNNYGDGVSPMSWIGSVDILRRWKNHGCQRVKYGQCWVFAAVACTVLRCLGIPTRV   613 -----LPVALEGCTFTVEGAGLTEEQKTVEIPDPVEA-----GEEVKVRMD-----LVPLHMG-------

                  .    :    .    :    .    :    .    :    .    :    .    :    .    :           .    :    .    :    .    :    .    :
Chain 1:  248 IMDF-------------------------------TDHVWTECYSHS------LKRWIHLDPCE------   648 FEERFQRKSYKITTPGFQANLFRFRFLSVRDVNSTSRLQLG
              :..:                               ..|.|.|.:...      ...|..|||..                            .......|.....:.|.......:|
Chain 2:  298 VTNYNSAHDQNSNLLIEYFRNEFGEIQGDKSEMIWNFHCWVESWMTRPDLQPGYEGWQALDPTPQEKSEG   661 ----------------LHKLVVNFESDKLKAVKGFRNVIIG

                  .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:  275 --GVY----------------DKPMLYEKGWNKKLNYVI-------AISKDGVCDVTKRYTKKWHEVLSR
                ...                |.|..:... |......|       ..
Chain 2:  368 TYCCGPVPVRAIKEGDLSTKYDAPFVFAEV-NADVVDWIQQDDGSVHK----------------------

                  .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:  320 RTLTTESSLQDGLRTLTRERRRSLMFESLSK--LELRDRNEQEELERNLHSADNASVSLPGRQSGDREWR
                 ....||..||:..|:...|....:....  .......|.|...|..|....|.....|.  ..|...
Chain 2:  415 ---SINRSLIVGLKISTKSVGRDEREDITHTYKYPEGSSEEREAFTRANHLNKLAEKEETGM--AMRIRV

                  .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:  388 IMRSEFGSD-ENSSVSS------SSCPVRKCVDDHVTNIYDSFLPILTQFVEDGLPVARTNEVLKMIKQV
              ......||| :......      ..|....|....               ...|:......       ..
Chain 2:  480 GQSMNMGSDFDVFAHITNNTAEEYVCRLLLCARTV---------------SYNGILGPECG-------TK
```

Figure 4: Sequence alignment of template (chain 1) and model (chain 2) derived from structural superimposition by jCE

Figure 4 shows the sequence alignment derived with jCE from the structural superimposition. The structure of the template roughly compares with the structure of the model: the sequence identity as derived after structural alignment is 10.39% with a 18.9% similarity. In order to verify whether the active site described in the reference article is conserved, the superimposition is zoomed to the point to which it is visible and comparable. [Figure 5]

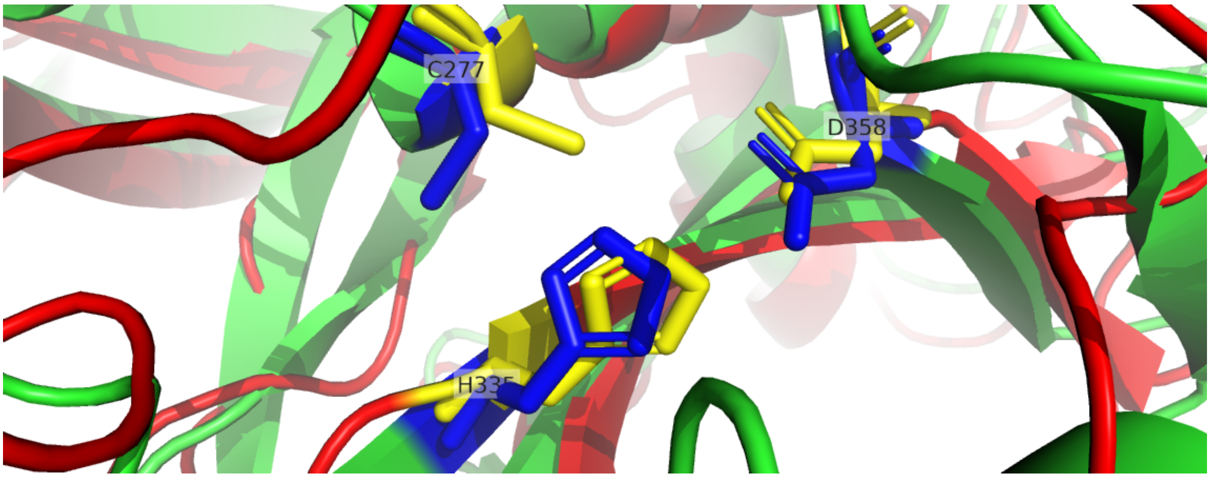


Figure 5: comparison of the catalytic triad of target(green and blue) and template(red and yellow)

The distances between the c-alpha atoms of the residues in the catalytic triad are measured and reported in table 3.

Table 3: Distance between c-alpha atoms in the catalytic triad of template and target

[H]

Residues	Distance in Template (Å)	Distance in Target (Å)
C277-D358	8.5	8.6
D358-H335	9.6	9.8
H335-C277	6.6	6.7

As is seen in the table, the distances between the residues are mostly conserved; in Figure 5, C277 is the only residue whose spatial orientation differs significantly between template and target; however, since the protein is not an immutable entity but rather a flexible one, I could say that the residue can still reach a conformation that supports the function of the catalytic triad. I determine that the catalytic triad is conserved both in residues and spatial organization, as is most of the secondary structure of the core domain; therefore I may conclude that the function of the core domain is conserved.
Having confirmed the conservation of the active site residues, and having previously obtained the confirmation that the model is a mostly stable protein, I can confirm that the model protein shares the function of the template protein; therefore, I have all the necessary information to prove that the structure of the core domain of template and target proteins is conserved. Considering the GO terms associated to the core transglutaminase domain, I may transfer the "Protein-glutamine gamma-glutamyltransferase activity" term to my target.

7 Swiss-model

In order to verify whether manual annotation of the target sequence can be replaced by automatic methods, I submit the target to the Swiss-model[12]; however, its first choice of template is 1X3W, a PNGase. Having just demonstrated how Q9FGY9 is not a PNGase, I can conclude that in the case of this protein, automatic methods are still not perfected enough to replace the manual procedure.

8 References

1 "AtPng1p. The First Plant Transglutaminase1" M.Della Mea, D.Caparros-Ruiz, I.Claparols, D.Serafini-Fracassini, J.Rigau; Plant Physiology, August 2004, Vol. 135, pp. 2046–2054

2 "Plant and animal transglutaminases: do similar functions imply similar structures?" Donatella Serafini-Fracassini, Massimiliano Della Mea, Gianluca Tasco, Rita Casadio, Stefano Del Duca; Amino Acids (2009) 36:643–657

3 https://www.uniprot.org/uniprot/Q9FGY9

- The UniProt Consortium "UniProt: the universal protein knowledgebase" Nucleic Acids Res. 46: 2699 (2018)

4 https://www.rcsb.org

- "The Protein Data Bank" H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne (2000) Nucleic Acids Research, 28: 235-242. doi:10.1093/nar/28.1.235

5 https://www.ebi.ac.uk/interpro/

- Alex L Mitchell, Teresa K Attwood, Patricia C Babbitt, Matthias Blum, Peer Bork, Alan Bridge, Shoshana D Brown, Hsin-Yu Chang, Sara El-Gebali, Matthew I Fraser, Julian Gough, David R Haft, Hongzhan Huang, Ivica Letunic, Rodrigo Lopez, Aurélien Luciani, Fabio Madeira, Aron Marchler-Bauer, Huaiyu Mi, Darren A Natale, Marco Necci, Gift Nuka, Christine Orengo, Arun P Pandurangan, Typhaine Paysan-Lafosse, Sebastien Pesseat, Simon C Potter, Matloob A Qureshi, Neil D Rawlings, Nicole Redaschi, Lorna J Richardson, Catherine Rivoire, Gustavo A Salazar, Amaia Sangrador-Vegas, Christian J A Sigrist, Ian Sillitoe, Granger G Sutton, Narmada Thanki, Paul D Thomas, Silvio C E Tosatto, Siew-Yit Yong and Robert D Finn (2019). "InterPro in 2019: improving coverage, classification and access to protein sequence annotations." Nucleic Acids Research, Jan 2019; doi: 10.1093/nar/gky1100

6 https://www.uniprot.org/blast

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410. PubMed

7 https://www.ebi.ac.uk/Tools/msa/clustalo/

- Sievers F, Higgins DG (2014-01-01). Russell DJ, ed. Multiple Sequence Alignment Methods. Methods in Molecular Biology (Clifton, N.j.). Methods in Molecular Biology. 1079. Humana Press. pp. 105–116.

8 https://salilab.org/modeller/

- A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993

9 https://pymol.org

- DeLano, W. L. (2002). "Pymol: An open-source molecular graphics tool." CCP4 Newsletter On Protein Crystallography, 40, 82-92.

10 http://servicesn.mbi.ucla.edu/PROCHECK

- Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). "PROCHECK - a program to check the stereochemical quality of protein structures." J. App. Cryst., 26, 283-291.

- Laskowski R A, Rullmannn J A, MacArthur M W, Kaptein R, Thornton J M (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR". J Biomol NMR, 8, 477-486. [PubMed id: 9008363]

11 source.rcsb.org/jfatcatserver/

- Shindyalov, I.N. & Zhuang, Pelion. (1998). "Protein Structure Alignment by Incremental Combinatorial Extension (CE) of the Optimal Path." Protein engineering. 11. 739-47

- "Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity" Shenping Liu, Richard A. Cerione, and Jon Clardy; PNAS March 5, 2002 vol. 99 no. 5 2743–2747

12 https://www.swissmodel.expasy.org

- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. "SWISS-MODEL: homology modelling of protein structures and complexes." Nucleic Acids Res. 46(W1), W296-W303 (2018).

- https://www.rcsb.org/structure/1kv3

- https://www.rcsb.org/structure/1ggt