www.biodicon.com Biological Diversity and Conservation

ISSN 1308-8084 Online ISSN 1308-5301 Print

Research article
DOI: 10.46309/biodicon.2026.1904444
19/2 (2026) 188-202

Detection and structural mapping of potential fucosyltransferases motif sequences in some insect species

Sadiye OYLUM[1], Ceren ERSOY[1], Savaş İZZETOĞLU[2], Gamze TURGAY İZZETOĞLU [*] [1]
ORCID: 0009-0003-4198-1984; 0000-0002-6408-2987; 0000-0002-1546-1083; 0000-0001-9828-2402

[1] Ege University, Institute of Science, Biology Department, Bornova, İzmir, Türkiye
[2] Ege University, Faculty of Science, Biology Department, Bornova, İzmir, Türkiye

Abstract
Purpose: Fucosyltransferases (FucTs) catalyze the transfer (fucosylation) of α-L-fucose from GDP-Fuc onto N- and O-linked glycans, free oligosaccharides, or lipids or directly onto proteins. In mammals, 13 fucosyltransferase gene families (*FUT*) have been identified, including *FUT* 1-11, *POFUT1* (Protein O-fucosyltransferase 1) and *POFUT2* (Protein O-fucosyltransferase 2). Despite the existing studies in the literature, the diversity of many fucosyltransferases in organisms remains unknown. In insects, there are few studies on fucosyltransferase and *FUT* types are not characterized by species.
Method: Fucosyltransferase motif sequences were determined from the transcriptomes (all mRNA sequences) obtained from the databases of 5 insects (*Bombyx mori*, *Callosobruchus maculatus*, *Galleria mellonella*, *Manduca sexta*, *Myzus persicae*) with known genomes through databases, and putative fucosyltransferase sequences in insects were determined with computational analysis (Clustal Omega, Prosite ve MEME-Suite/FIMO). Then, the similarities between the three-dimensional protein structures of the mammalian fucosyltransferase enzymes were revealed using Swiss-Model, AlphaFold and PyMOL databases.
Findings: Based on motif and three-dimensional structure comparison regions, putative homologs corresponding to all 13 fucosyltransferases were predicted in *Callosobruchus maculatus*, FucT1 and FucT9 in *Galleria mellonella*, FucT1, FucT4, FucT5, FucT8, and FucT11 in *Manduca sexta*, FucT2, FucT3, FucT5, FucT6, FucT7, FucT9, FucT10, and FucT11 in *Myzus persicae*, and FucT1, FucT3, FucT5, FucT6, and FucT9 motifs were characterized in *Bombyx mori*.
Conclusion: This information provides guidance for continuing studies on fucosyltransferases in invertebrates. New sequences for the fucosyltransferase enzyme will be revealed as the genome of determined insects is studied. While the similarities of fucosyltransferase motifs in insects and mammals indicate that they may be orthologous genes in terms of evolution, it is also an indication that new developmental and evolutionary studies will be conducted on this subject.
†

Keywords: motif scanning, 3D protein model, fucosyltransferases, insect

---------- * ----------

Bazı böcek türlerinde olası fukoziltransferaz motif dizilerinin tespiti ve yapısal haritalanması

Özet
Amaç: Fukoziltransferazlar (FucTs), GDP-Fukoz (GDP-Fuc)'tan α-L-fukozun N- ve O-bağlı glikanlara, serbest oligosakkaritlere veya lipitlere ya da doğrudan proteinlere aktarılmasını (fukozilasyon) katalizler. Memelilerde, *FUT*1–11, *POFUT1* (Protein O-fukoziltransferaz 1) ve *POFUT2* (Protein O-fukoziltransferaz 2) dahil olmak üzere 13 fukoziltransferaz gen ailesi (*FUT*) tanımlanmıştır. Literatürde mevcut çalışmalar bulunmasına rağmen, birçok organizmada fukoziltransferazların çeşitliliği hâlâ tam olarak bilinmemektedir. Böceklerde ise fukoziltransferazlar üzerine yapılan çalışmalar oldukça sınırlıdır ve *FUT* tipleri tür bazında henüz karakterize edilmemiştir.
Metod: Genomları bilinen 5 böceğin (*Bombyx mori, Callosobruchus maculatus, Galleria mellonella, Manduca sexta, Myzus persicae*) veri tabanlarından elde edilen transkriptomlarından (tüm mRNA dizileri) fukoziltransferaz motif

[*] *Corresponding author*: Tel.: +90232 311 1791; Fax: +90232 311 1791; E-mail: gamze.turgay@ege.edu.tr
 Received: 10.03.2026; Published: 15.06.2026 BioDiCon. 1230-100326

dizileri veri tabanları (Clustal Omega, Prosite ve MEME-Suite/FIMO) aracılığıyla tespit edilerek böceklerdeki olası fukoziltransferazlar belirlenmiştir. Daha sonra memelilerdeki fukoziltransferaz enzimlerinin üç boyutlu protein yapıları arasındaki benzerlikler Swiss-Model, AlphaFold ve pyMOL veri tabanları kullanılarak ortaya konulmuştur.
Bulgular: Motif ve üç boyutlu yapı karşılaştırma bölgelerine göre, *Callosobruchus maculatus* türünde 13 fukoziltransferazın tümüne karşılık gelen olası homologlar bulunurken , *Galleria mellonella* türünde FucT1 ve FucT9, *Manduca sexta* türünde FucT1, FucT4, FucT5, FucT8 ve FucT11, *Myzus persicae* türünde FucT2, FucT3, FucT5, FucT6, FucT7, FucT9, FucT10 ve FucT11, *Bombyx mori* türünde FucT1, FucT3, FucT5, FucT6 ve FucT9 motifleri karakterize edilmiştir.
Tartışma: Bu çalışma, omurgasızlarda fukoziltransferazlar üzerine yapılacak ilerleyen çalışmalar için yol gösterici özellik taşımaktadır. Genomu belirlenen böceklerle çalışıldıkça fukoziltransferaz enzimi için yeni diziler ortaya çıkarılacaktır. Böcekler ve memelilerdeki fukoziltransferaz motiflerinin benzerlikleri, bu genlerin evrimsel açıdan ortolog genler olabileceğine işaret ederken, ayrıca bu konuda gelişimsel ve evrimsel yeni çalışmaların yapılacağının göstergesidir.

Anahtar kelimeler: motif, 3D protein model, fukoziltransferazlar, böcek

1. Introduction

Fucosylation is a biochemical process that refers to a glycosylation reaction occurring in the cells of organisms. Fucosylation, carried out by fucosyltransferases, catalyzes the transfer of α-L-fucose from GDP-Fuc to N- and O-linked glycans, free oligosaccharides, or lipids, or directly to proteins [1-4]. The fucosyltransferase gene family (*FUT*) encodes a group of proteins (FucT) that exhibit a complex cell and tissue specific expression pattern [5]. Similarities and differences in the enzymatic activities of FucTs enable distinctions among various enzyme activities based on substrate specificities. Fucosyltransferases (FucTs) consist of a small N-terminal cytoplasmic tail, a membrane region, and a catalytic domain directed toward the Golgi lumen. Between the membrane region and the catalytic domain is a region known as the body. The N-terminal region (cytoplasmic, transmembrane, and body regions) has the highest sequence heterogeneity. Soluble forms of FucTs are produced by naturally occurring proteases and have been found in serum, milk, amniotic fluid, semen, and other body fluids. This suggests that part of the N-terminal of FucTs is not required for enzyme activity, while it remains unclear which specific portion of the amino acid sequence is essential for enzymatic activity. Studies on protein cleavage have shown that catalytic activity is unaffected by the removal of a region of 50-70 amino acids from the N-terminal region, depending on the enzyme form, while the removal of a few amino acids from the C-terminal leads to a complete loss of enzyme activity. The catalytic domain of FucT3 consists of 62-361 amino acids, while FucT4 consists of 405, FucT5 of 374, FucT6 of 359, FucT7 of 342, and FucT9 of 359 amino acids [5,6]. In the α1,3-fucosyltransferase family (FucT3–FucT7 and FucT9-11), two conserved peptide motifs have been primarily identified, referred to as I and II. Subsequently, a third conserved motif consisting of six amino acids has been discovered, named the acceptor motif [(I/V/F)HH(R/W)(D/E)(I/V/L)] [3]. As research progressed, a motif located at the NH2 terminal end of the sequence, preceding the acceptor motif, was discovered and named motif III [7,8]. The order of these four conserved motifs is arranged as III, acceptor motif, I, and II. In a later study, a fifth conserved motif was found between the acceptor motif and motif I. To avoid confusion, the five peptide motifs were renamed from I to V according to their order in the protein sequence [3]. Alignment of fucosyltransferase sequences across various species of invertebrates and vertebrates shows that motif regions are highly conserved in enzymes ranging from insects to mammals, indicating that homologous sequences originated from a common ancestor [9,10].

The study examined motif sequences of vertebrate FucTs from databases and compared them with insect protein sequences. Detection and visualization of potential fucosyltransferases in insects were conducted, and their motif regions were observed to intersect with the three-dimensional protein structures of vertebrates.

2. Materials and methods

2.1 Sequence data collection

Protein sequences from 5 insect species were analyzed. Transcriptome and proteome data for the remaining insect species were downloaded from the National Center for Biotechnology Information (NCBI) database (URL: https://www.ncbi.nlm.nih.gov/) and have been made available in the supplementary materials. In total, the protein databases searched included over 79704 sequences, with individual proteome sizes ranging from 13604 to 21246 sequences (Table 1).

2.2 Motif sequence profiles and motif scanning

Multiple sequence alignment of the motifs was performed with Clustal Omega (URL: https://www.ebi.ac.uk/Tools/msa/clustalo/). Regular expressions representing each motif were generated using PRATT (URL: https://web.expasy.org/pratt/) and compiled into sequence profiles specific to each fucosyltransferase subfamily. These profiles were scanned against insect protein sequences using FIMO from the MEME suite to identify significant matches (p<0.00001) (URL: https://meme-suite.org/meme/tools/fimo). The 3D protein structures of putative insect fucosyltransferases were predicted with Swiss Model (URL: https://swissmodel.expasy.org/). Then, known three-dimensional structures of vertebrate fucosyltransferases were obtained using the AlphaFold database (URL: https://alphafold.ebi.ac.uk/), and the identified fucosyltransferase protein sequences in insects were aligned with possible fucosyltransferases in PyMOL (URL: https://www.pymol.org/). Motif regions were compared by examining their three-dimensional structures.

Table 1. Protein sequence length of the insect species studied

Species	Reference genome accession number	Protein sequence length
Bombyx mori	GCF_014905235.1	13881
Callosobruchus maculatus	GCA_900659725.1	21246
Galleria mellonella	GCF_026898425.1	13604
Manduca sexta	GCF_014839805.1	15967
Myzus persicae	GCF_001856785.1	15006

3. Results

The motif models created in the Pratt application from mammalian fucosyltransferase amino acid sequences were searched in the protein sequences of 5 different insect species. Fucosyltransferase motif sequences were identified in 20-30 vertebrates for all thirteen characterized fucosyltransferases. Regular expression profiles representing each of the four motifs were generated using PRATT for each sialyltransferase subfamily (Table 2).

Table 2. Models created in PRATT for all fucosyltransferase motifs

Fucosyltransferases	Motif	Motif sequences
FUT1	I	A-T-L-[FL]-A-L-A-[QR]-L
	II	W-K-G-V-V-[AG]-[DN]-x-A-Y-L-[QR]-x-A-M-D-W-F-R
FUT2	I	A-T-L-[FY]-A-L-A-[KR]
	II	W-[KR]-G-V-V-A-[DG]-[QR]-x-Y-L-[EQR]-x-A-L-[DGS]-x-F
FUT3	I	I-L-L-W-T-W-P-F
	II	V-I-V-H-H-x-[DE]-I
	III	Q-R-W-[IM]-W-F-[NS]-[LM]-E-[PS]-P-[PS]
	IV	F-N-L-T-M-S-Y-R-S-D-S-D-[IV]-F-T-[LP]-Y-G
	V	Y-K-F-Y-L-A-F-E-N-S-x(3)-D-Y-I-T-E-K
FUT4	I	L-L-x(0,2)-E
	II	A-V-L-F-H-H-R-D-[FL]
	III	Q-x-W-V-W-M-N-F-E-S-P-[ST]
	IV	F-N-W-T-L-S-Y-R-[ATV]-D-S-D-[IV]-F-V-P-Y-G
	V	Y-K-F-Y-L-A-F-E-N-S-Q-H-x-D-Y-I-T-E-K
FUT5	I	I-L-L-W-T-W-P-F
	II	A-V-I-V-H-H-x-[DE]-[IV]
	III	Q-R-W-[IV]-W-[FL]-S-[LM]-E-S-P-S
	IV	F-N-L-T-M-S-Y-R-S-D-S-D-I-F-T-P-Y-G
	V	Y-K-F-Y-L-A-F-E-N-S-x-H-[LP]-D-Y-I-T-E-K
FUT6	I	L-[LV]-W-T-W-P-F
	II	A-V-[IL]-[IMV]-H-H-x-[DE]-[IV]
	III	W-[ILV]-W-x-[NS]-[LM]-E-S-P-S
	IV	F-N-L-T-M-[ST]-Y-[HR]-x-D-S-D-[IV]-F-x-P-Y-G
	V	Y-K-F-x-L-A-F-E-N-S-x-H-x-D-Y-I-T-E-[KR]
FUT7	I	L-[IV]-W-x-W-P-F
	II	A-V-V-F-H-H-R-E-L
	III	Q-[LP]-W-[IV]-W-A-S-[LM]-E-S-P-S
	IV	F-N-x-V-L-S-Y-R-[RS]-D-[AS]-D-[IV]-x-[MV]-P-Y-G
	V	Y-x-F-Y-L-[AS]-[FL]-E-N-S-[EQ]-H-x-D-Y-[IV]-T-E-K
FUT8	I	N-W-R-Y-A-T-G-G-W
	II	D-P-[AST]-L-L-x-E-A-K-T-K-Y-[PS]-x-Y-E-F-I-S
FUT9	I	L-[IV]-W-[LV]-W-P-F
	II	A-V-L-I-H-H-R-D-I
	III	Q-K-W-I-W-M-N-L-E-S-P-T
	IV	F-N-L-T-L-T-Y-R-R-D-S-D-I-Q-V-P-Y-G

Table 2. Continued

	V	C-K-F-Y-L-S-F-E-N-S-I-H-[KR]-D-Y-I-T-E-K
FUT10	I	W-W-S-P-L
	II	A-[FLV]-L-F-Y-G-T-D-F
	III	H-[DE]-W-A-[LV]-F-H-E-E-S-P-K
	IV	F-N-[HY]-T-A-T-F-S-R-H-S-[DH]-L-P-L-T-T-Q
	V	Y-K-F-I-L-A-F-E-N-A-[IV]-C-x(0,1)-D-x(0,1)-Y-[IV]-T-E-K
FUT11	I	V-L-L-W-W-S-P-G
	II	A-L-L-F-Y-G-T-D-F
	II	Q-[NST]-W-A-L-L-H-E-E-S-P-L
	IV	F-N-L-T-[AS]-T-F-S-R-[HY]-S-D-Y-P-L-[APS]-L-Q
	V	Y-K-F-H-L-A-L-E-x-A-I-C-[DNS]-D-Y-M-T-E-K
POFUT1	I	Y-H-R-[IV]-[IV]-S-L
	II	L-K-D-G-T-A-[DG]-[AS]-H-F-M-A-S-P-Q-C-V-G-Y
POFUT2	I	N-[IV]-P-V-[IMV]-E-[HY]
	II	L-E-L-[FY]-[KR]-D-G-G-[AV]-A-I-[IV]-D-Q-W-[IV]-C-[AS]-H

Potential fucosyltransferase sequences were identified in all 5 insect species studied. In the *C. maculatus*, all 13 fucosyltransferases, in the *G. mellonella*, FucT1 and FucT9, in the *M. sexta*, FucT1, FucT4, FucT5, FucT8, and FucT11, in the *M. persicae*, FucT2, FucT3, FucT5, FucT6, FucT7, FucT9, FucT10, and FucT11, and in the *B. mori*, FucT1, FucT3, FucT5, FucT6, and FucT9 were predicted to be present. In *C. maculatus*, the VEN42375.1 and VEN42374.1 uncharacterized proteins sequences (Table 3) were compared to the *H. sapiens* FucT1 protein sequence, and similarity was observed only with motif II (Figure 1) out of the two motifs.

The *C. maculatus* VEN51608.1 uncharacterized protein sequence (Table 4) was compared to the *H. sapiens* FucT2 protein sequence (Figure 2), it was determined that only motif I was conserved due to the overlapping of similar motifs with each other, while motif II was not conserved.

Table 3. High probability data of motif II of *C. maculatus* FucT1

Sequence name	Start	End	p-value	q-value	Matching sequences
VEN42375.1	249	267	1.6e-06	1	WGEDRGEKGYLLMTTDWFK
VEN42374.1	394	412	1.6e-06	1	WGEDRGEKGYLLMTTDWFK

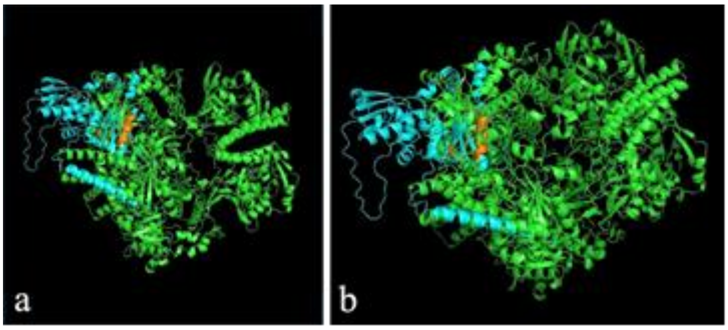

Figure 1. Comparison of the 3D structures of *C. maculatus* and *Homo sapiens* FucT1. a) Green represents *C. maculatus* (VEN42375.1), blue represents *H. sapiens* (AXX39110.1), and orange represents the shared motif II b) Green represents *C. maculatus* (VEN42374.1), blue represents *H. sapiens* (AXX39110.1), and orange represents the shared motif II.

Table 4. High probability data of motif I of *C. maculatus* FucT2 according to FIMO results

Sequence name	Start	End	p-value	q-value	Matching sequences
VEN51608.1	396	413	5.81e-06	1	LKTIKADREDLEDALADK

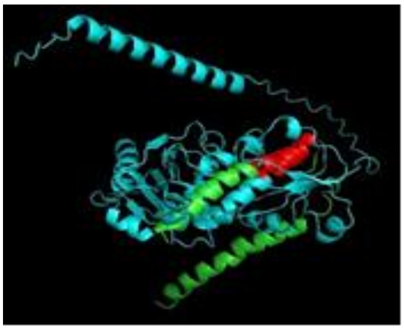

Figure 2. Comparison of the 3D structures of *C. maculatus* (VEN51608.1) and *H. sapiens* (NP_000502.4) FucT2. Green represents *C. maculatus*, blue represents *H. sapiens*, and red represents the shared motif I

It is observed that in the *C. maculatus* VEN34891.1 uncharacterized protein sequence, three motif regions that match with human FucT3 protein sequence were conserved (Figure 3a), while in the VEN34890.1 protein sequence, only one motif region was conserved (Figure 3b), and in the VEN58850.1 protein sequence, four motif regions were conserved (Figure 3c).

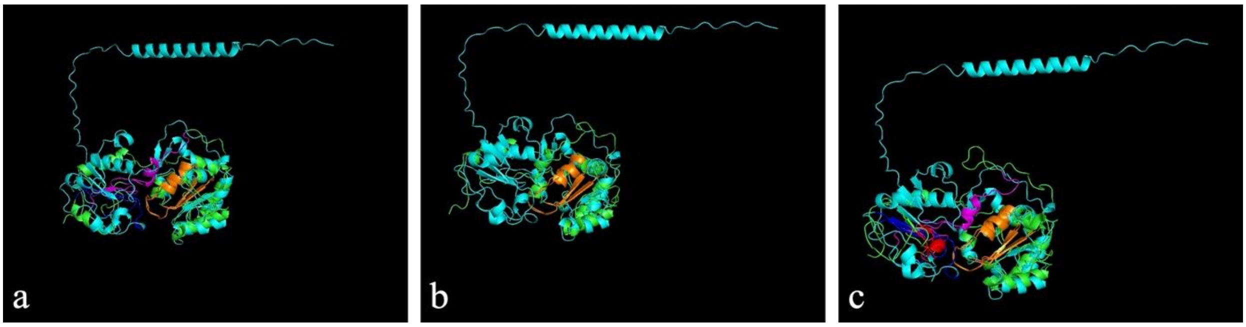

Figure 3. Comparison of the 3D structures of *C. maculatus* (VEN34891.1) putative and *H. sapiens* (NP_001369679.1) FucT3 sequence. a) Green represents *C. Maculatus putative FucT3 sequence* (VEN34890.1), blue represents *H. Sapiens FucT3 sequence* (NP_001369679.1), navy represents motif III, purple represents motif IV, and orange represents motif V. b) Green represents *C. maculatus putative FucT3 sequence*, blue represents *H. Sapiens FucT3 sequence*, and orange represents motif V c) Green represents *C. maculatus putative FucT3 sequence* (VEN58850.1), blue represents *H. sapiens FucT3 sequence* (NP_001369679.1), red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V

The *C. maculatus* VEN34891.1 uncharacterized protein sequence exhibited three similar motifs when compared to the *M. musculus* FucT4 protein sequence, specifically in motifs III-V regions (Figure 4a). In the VEN34890.1 uncharacterized protein sequence, only motif V region is conserved (Figure 4b). The VEN58850.1 uncharacterized protein sequence showed four similar motifs with *M. musculus*, specifically in motifs II-V regions (Figure 4c). Motifs IV and V were identified in VEN50351.1 (Figure 4d) and VEN50349.1 (Figure 4e) protein sequences, whereas only motif V region was observed in the VEN53775.1 protein sequence (Figure 4f).

C. maculatus VEN34891.1 uncharacterized protein sequence showed similarity in motif II-V regions with *H. sapiens* FucT6 protein (Figure 6a). Compared with the VEN34889.1 uncharacterized protein showed similarity in motifs II-V regions (Figure 6b), in VEN34890.1 uncharacterized protein only motif V region was similar (Figure 6c), and in VEN58850.1 uncharacterized protein motifs II-V regions showed similarity (Figure 6d). Two additional sequences were found in *C. maculatus* that are similar to FucT6. The first one is the VEN40379.1 uncharacterized protein, which shows similarity with *H. sapiens* only in motif V region (Figure 6e), and the other sequence, VEN40380.1 uncharacterized protein, shows similarity only in motif IV region (Figure 6f).

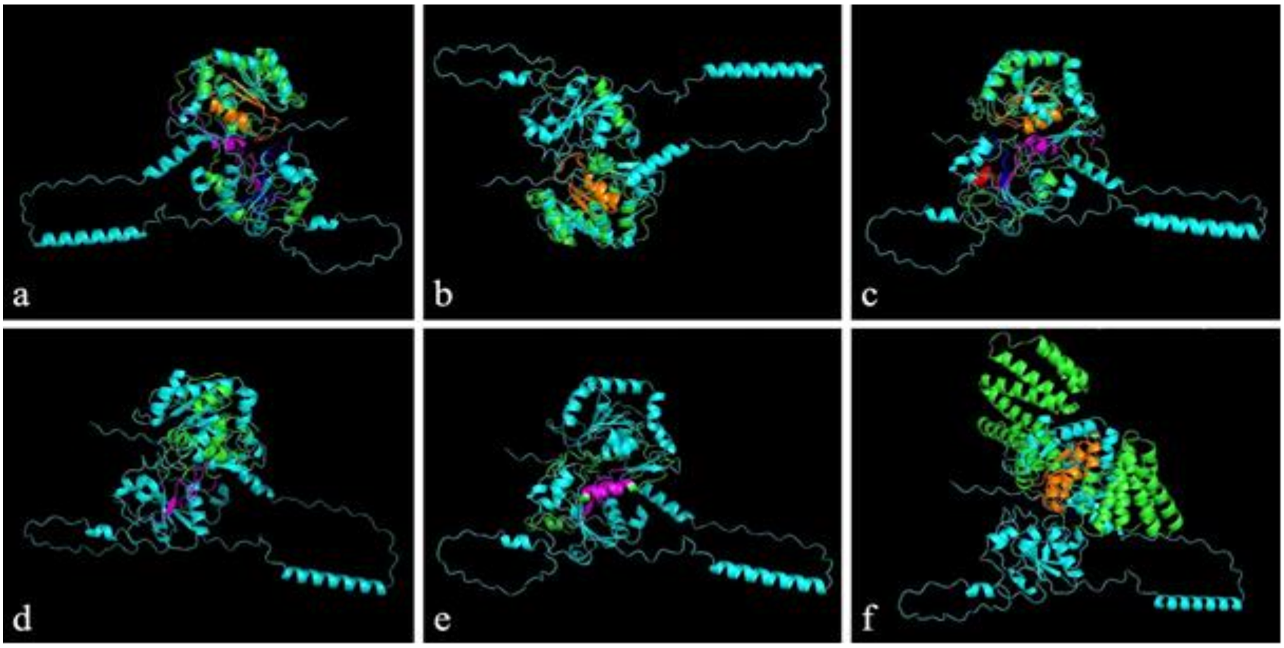

Figure 4. Comparison of the 3D structures of *C. maculatus* and *M. musculus* FucT4. a)Green represents *C. maculatus (*VEN34891.1), blue represents *M. musculus (*AAI37589.1), navy represents motif III, purple represents motif IV, and orange represents motif V b) Green represents *C. maculatus (*VEN34890.1), blue represents *M. musculus (*AAI37589.1) and orange represents motif V c) Green represents *C. maculatus (*VEN58850.1), blue represents *M. musculus (*AAI37589.1), red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V d) Green represents *C. maculatus* (VEN50351.1), blue represents *M. musculus (*AAI37589.1) purple represents motif IV e) Green represents *C. maculatus (*VEN50349.1), blue represents *M. musculus (*AAI37589.1) purple represents motif IV f) Green represents *C. maculatus* (VEN53775.1), blue represents *M. musculus (*AAI37589.1) and orange represents motif V

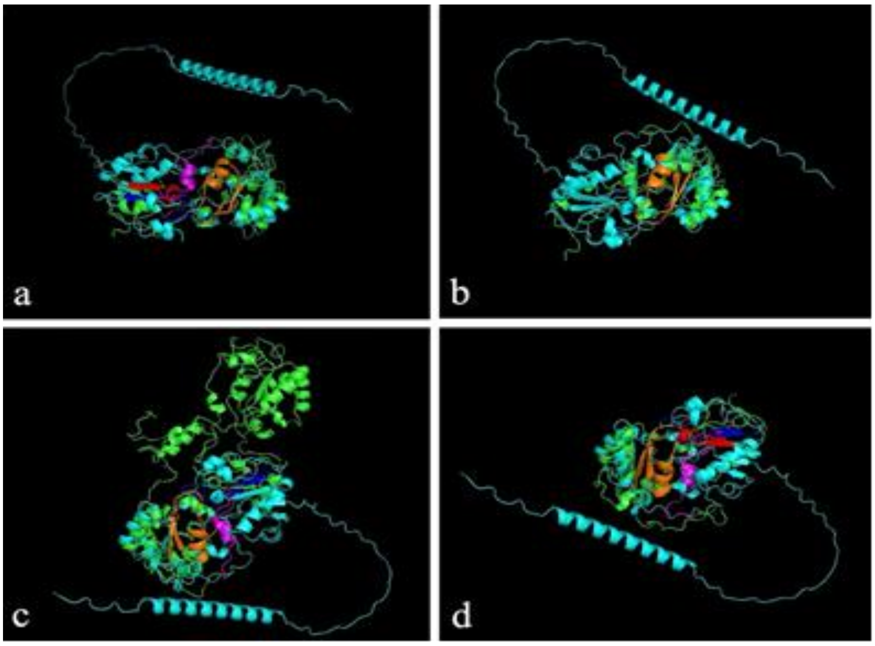

Figure 5. Comparison of the 3D structures of *C. maculatus* and *H. sapiens* FucT5. a) Green represents *C. maculatus* (VEN34891.1), blue represents *H. sapiens* (NP_002025.2), red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V b) Green represents *C. maculatus* (VEN34890.1), blue represents *H. sapiens* (NP_002025.2) orange represents motif V c) Green represents *C. maculatus* (VEN34889.1), blue represents *H. sapiens* (NP_002025.2), navy represents motif III, purple represents motif IV, and orange represents motif V, which are shared between both. d) Green represents *C. maculatus* (VEN58850.1), blue represents *H. sapiens* (NP_002025.2), red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V

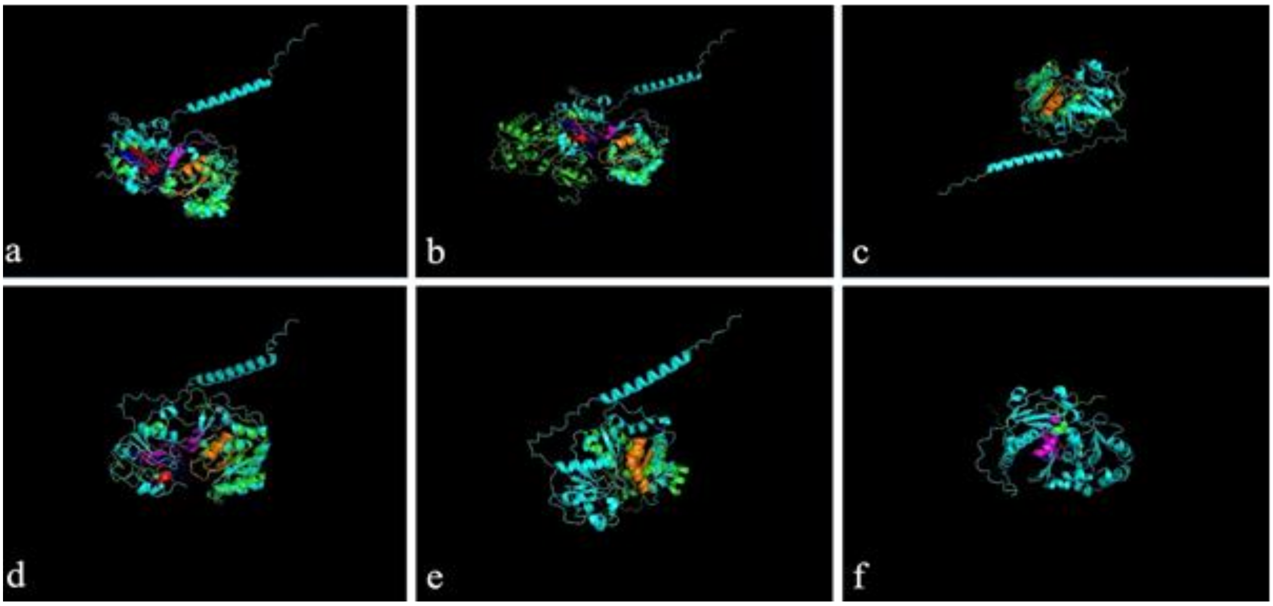

Figure 6. Comparison of the 3D structures of *C. maculatus* and *H. sapiens* (AAC50190.1) FucT6. a) Green represents *C. maculatus* (VEN34891.1), blue represents *H. sapiens*, red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V. b) Green represents *C. maculatus* (VEN34889.1), blue represents *H. sapiens*, red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V c) Green represents *C. maculatus* (VEN34890.1), blue represents *H. sapiens* and orange represents motif V d) Green represents *C. maculatus* (VEN58850.1), blue represents *H. sapiens*, red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V e) Green represents *C. maculatus* (VEN40379.1), blue represents *H. sapiens*, and orange represents motif V f) Green represents *C. maculatus* (VEN40380.1), blue represents *H. sapiens* purple represents motif IV

Aligned with the *H. sapiens* FucT7 protein, similarity in motifs II-V regions was found in the *C. maculatus* VEN34889.1 uncharacterized protein (Figure 7a). Comparison with the VEN34890.1 uncharacterized protein showed similarity only in motif V region (Figure 7b), in VEN34891.1 uncharacterized protein motifs II-V regions showed similarity (Figure 7c), and in VEN58850.1 uncharacterized protein motifs II-V regions showed similarity as well (Figure 7d)

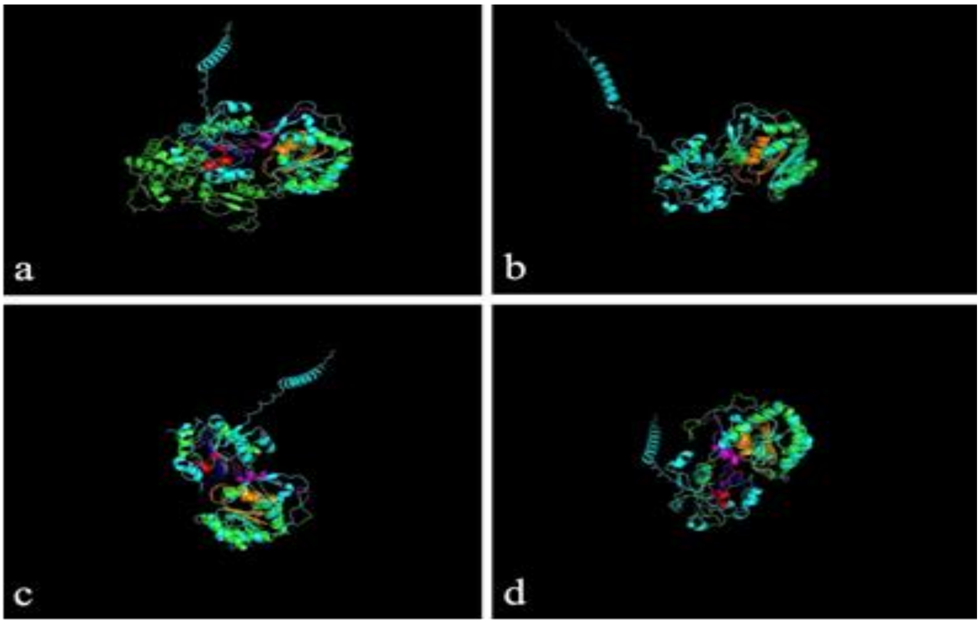

Figure 7. Comparison of the 3D structures of *C. maculatus* and *H. sapiens* (NP_004470.1) FucT7. a) Green represents *C. maculatus* (VEN34889.1), blue represents *H. sapiens*, red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V b) Green represents *C. maculatus* (VEN34890.1), blue represents *H. sapiens* and orange represents motif V c) Green represents *C. maculatus* (VEN34891.1), blue represents *H. sapiens*, red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V d) Green represents *C. maculatus* (VEN58850.1), blue represents *H. sapiens*, red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V

Comparing the 3D protein structure of the *H. sapiens* FucT8 protein with the *C. maculatus* VEN62377.1 uncharacterized protein sequence, it was found that both motif I and motif II showed similarity with humans (Figure 8)

Figure 8. Comparison of the 3D structures of *C. maculatus* (VEN62377.1) and *H. sapiens* (CAA76988.1) FucT8. Green represents *C. maculatus*, blue represents H. sapiens, red represents motif I, and orange represents motif II, which are shared between both

Aligned with the *H. sapiens* FucT9 protein, similarity in motifs II-V regions was observed in the *C. maculatus* VEN34889.1 uncharacterized protein sequence (Figure 9a). Compared with the VEN34890.1 uncharacterized protein showed similarity only in motif V region (Figure 9b), in VEN34891.1 uncharacterized protein motifs III-V regions showed similarity (Figure 9c), and in VEN58850.1 uncharacterized protein motifs III-V regions showed similarity as well (Figure 9d). Another sequence was found in *C. maculatus* that is similar only to FucT9. The VEN58429.1 uncharacterized protein sequence showed similarity with *H. sapiens* only in motif V region (Figure 9e).

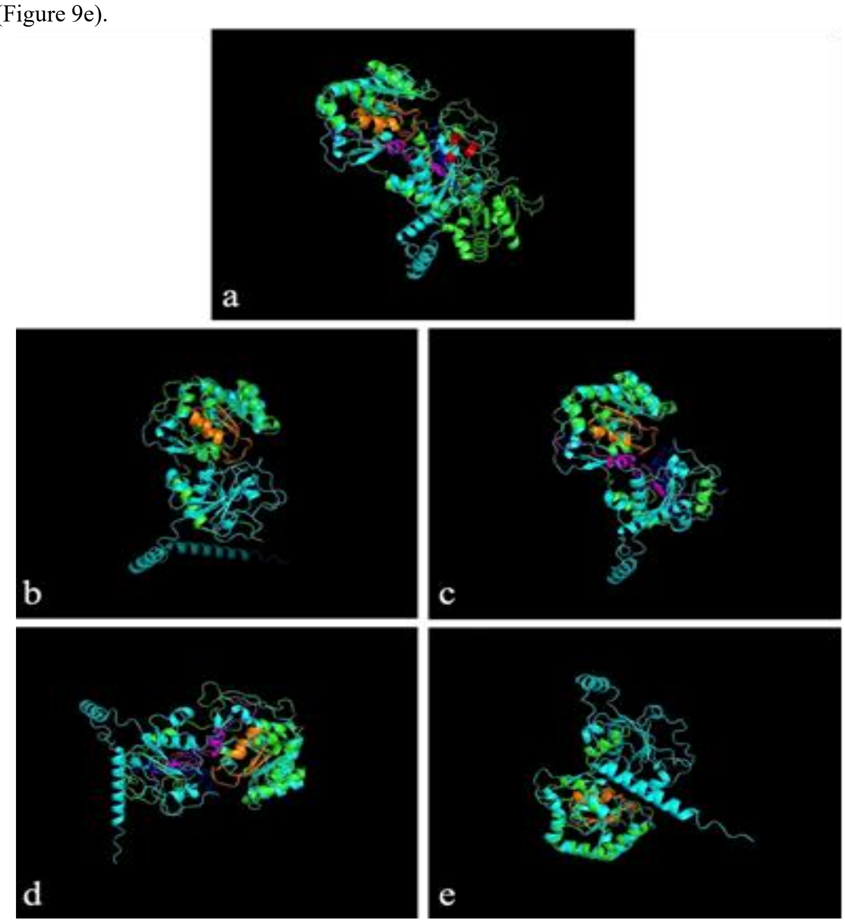

Figure 9. Comparison of the 3D structures of *C. maculatus* and *H. sapiens* (NP_006572.2) FucT9. a) Green represents *C. maculatus* (VEN34889.1), blue represents *H. sapiens*, red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V b) Green represents *C. maculatus* (VEN34890.1), blue represents *H. sapiens*, and orange represents motif V c) Green represents *C. maculatus* (VEN34891.1), blue represents *H. sapiens*, navy represents motif III, purple represents motif IV, and orange represents motif V d) Green represents *C. maculates* (VEN58850.1), blue represents *H. sapiens*, navy represents motif III, purple represents motif IV, and orange represents motif V e) Green represents *C. maculatus* (VEN58429.1), blue represents *H. sapiens* and orange represents motif V

When the protein of *C. maculatus* VEN34889.1 uncharacterized protein sequence was aligned with the *B. taurus* FucT10 protein, similarities in motif III-V regions were noted (Figure 10a). In comparison with the VEN34890.1 uncharacterized protein, only the motif V region showed similarity (Figure 10b), while in the VEN34891.1 uncharacterized protein, motif III-V regions exhibited similarities (Figure 10c). In the case of the VEN58850.1 uncharacterized protein, similarities were found in motif II, IV, and V regions (Figure 10d).

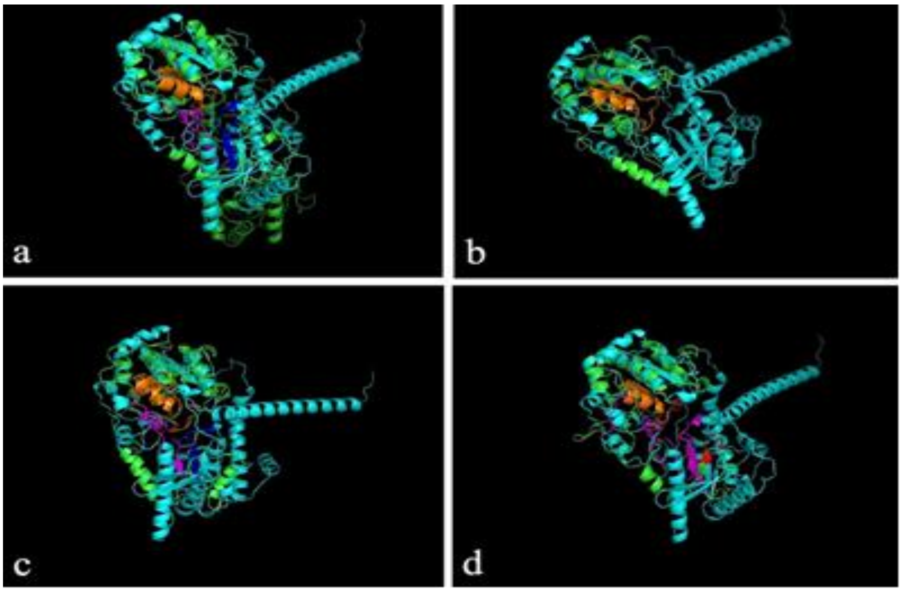

Figure 10. The comparison of the 3D structures of *C. maculatus* and *B. taurus* FucT10. a) Green represents *C. maculatus* (VEN34889.1), blue represents *B. taurus (*CAH05071.1) and navy represents motif III, lilac represents motif IV, and orange represents motif V b) Green represents *C. maculatus* (VEN34890.1), blue represents *B. taurus* (CAH05071.1) and orange represents motif V c) Green represents *C. maculatus* (VEN34891.1), blue represents *B. taurus (*CAH05071.1) and navy represents motif III, lilac represents motif IV, and orange represents motif V d) Green represents *C. maculatus*(VEN58850.1), blue represents *B. taurus (*CAH05071.1) and red represents motif II, lilac represents motif IV, and orange represents motif V

Aligned with the uncharacterized protein sequence of *C. maculatus* VEN34889.1, similarities were observed in motif III-V regions with the *M. musculus* FucT11 protein (Figure 11a). In comparison with the VEN34890.1 uncharacterized protein sequence, only the motif V region showed similarity (Figure 11b), while in the VEN34891.1 uncharacterized protein sequence, motif III-V regions exhibited similarities (Figure 11c). Similarly, in the VEN58850.1 uncharacterized protein sequence, similarities were found in motif II-V regions (Figure 11d). Additionally, another sequence resembling only FucT11 according to FIMO results was identified in *C. maculatus*. The VEN53344.1 uncharacterized protein sequence showed similarities with motif II-V regions in *M. musculus* (Figure 11e).

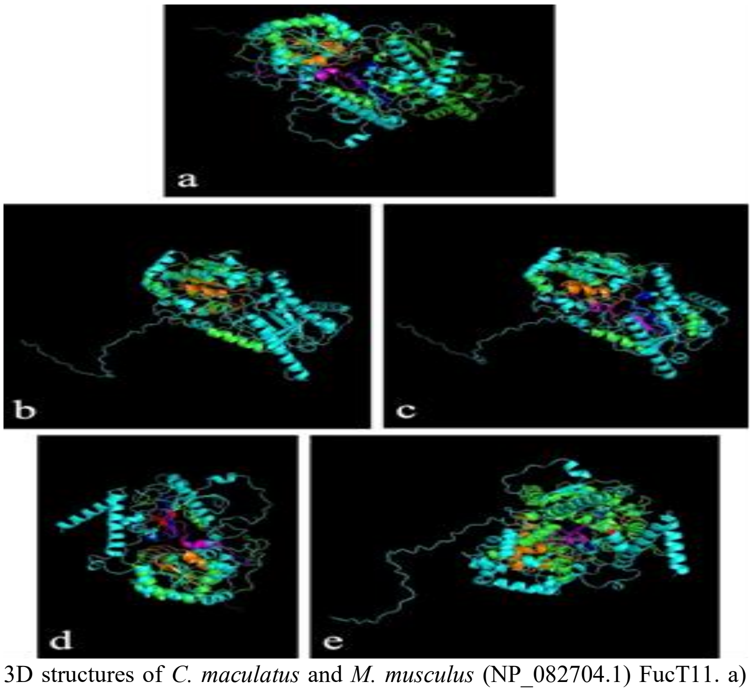

Figure 11. The comparison of the 3D structures of *C. maculatus* and *M. musculus* (NP_082704.1) FucT11. a) Green represents *C. maculatus* (VEN34889.1), blue represents *M. musculus* (NP_082704.1), navy represents motif III, lilac represents motif IV, and orange represents motif V b) Green represents *C. maculatus* (VEN34890.1), blue represents *M. musculus* (NP_082704.1) and orange represents motif V c) Green represents *C. maculatus* (VEN34891.1), blue represents *M. musculus* (NP_082704.1),navy represents motif III, lilac represents motif IV, and orange represents motif V d) Green represents *C. maculatus* (VEN58850.1), blue represents *M. musculus* (NP_082704.1), red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V e) Green represents *C. maculatus* (VEN53344.1), blue represents *M. musculus* (NP_082704.1), red represents motif II, navy represents motif III, purple represents motif IV, and orange represents motif V

In *B. mori*, the XP_021207495.1 uncharacterized protein sequence has been compared with the *H. sapiens* (NP_002025.2) FucT5 protein, and similarities have been identified with motif III and motif V out of five motifs (Figure 12).

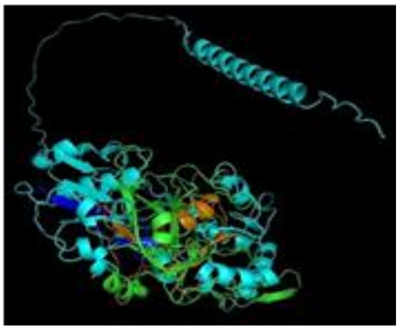

Figure 12. Comparison of the 3D structures of *B. mori* (XP_021207495.1) and *H. sapiens* (NP_002025.2) FucT5. Green represents *B. mori*, blue represents *H. sapiens*, navy blue represents motif III, and orange represents motif V, indicating the overlapping regions between the two.

B. mori putative FucT6 protein sequence (XP_021207495.1) has been compared with the *H. sapiens* (AAC50190.1) FucT6 protein, and similarity has been found with motif V out of five motifs (Figure 13).

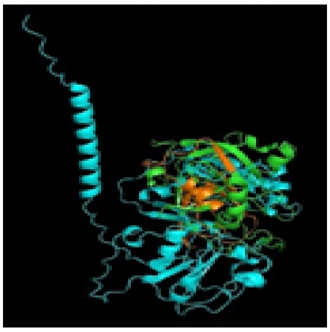

Figure 13. Comparison of the 3D structures of *B. mori* (XP_021207495.1) and *H. sapiens* (AAC50190.1) FucT6. Green represents *B. mori*, blue represents *H. sapiens*, and orange represents motif V, indicating the overlapping regions between the two.

G. mellonella uncharacterized protein (XP_031767290.1) has been aligned with the *H. sapiens* (AXX39110.1) FucT1 protein using the PyMOL program, and the motif regions of the sequences have been compared. According to the results, similarity has been determined only in motif II region (Figure 14).

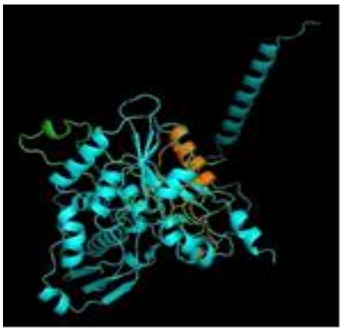

Figure 14. Comparison of the 3D structures of *G. mellonella* (XP_031767290.1) and *H. sapiens* (AXX39110.1) FucT1. Green represents *G. mellonella*, blue represents *H. sapiens*, and orange represents the shared motif II region between the two.

G. mellonella XP_026757326.1 uncharacterized protein sequence has been aligned with the *H. sapiens* FucT9 protein sequence (NP_006572.2), and the motif regions have been compared. Unlike the other compared sequences, a region similar to motif I has been found for the first time in the *G. mellonella* XP_026757326.1 protein sequence (Figure 15). Additionally, similarity with motif III and motif V regions is indicated by the match of motif sequences (Figure 15).

Figure 15. Comparison of the 3D structures of *G. mellonella* (XP_026757326.1) and *H. sapiens* (NP_006572.2) FucT9. Green represents *G. mellonella*, blue represents *H. sapiens*, yellow represents motif I, navy blue represents motif III, and orange represents motif V, indicating the overlapping regions between the two.

M. sexta XP_030033298.2 uncharacterized protein sequence has been compared with the *M. musculus* (AAI37589.1) FucT4 protein sequence, and despite the short length of the XP_030033298.2 protein sequence, a match has been observed in motif V region (Figure 16).

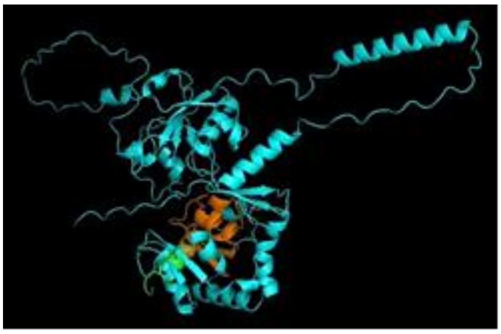

Figure 16. Comparison of the 3D structures of *M. sexta* (XP_030033298.2) and *M. musculus* (AAI37589.1) FucT4. Green represents *M. sexta*, blue represents *M. musculus*, and orange represents motif V, indicating the overlapping region between the two.

The uncharacterized protein sequence of *M. sexta* (XP_037296652.1) was aligned with the protein sequence of *M. musculus* (AAH10666.1) FucT8, similarity was found in the motif V region (Figure 17).

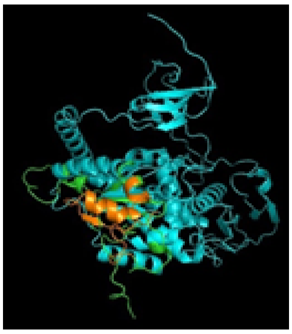

Figure 17. Comparison of the 3D structures of *M. sexta* (XP_037296652.1) and *M. musculus* (AAH10666.1) FucT8. Green represents *M. sexta*, blue represents *M. musculus*, and orange represents motif V, indicating the overlapping region between the two.

The uncharacterized protein sequences XP_022182156.1 (Figure 18a) and XP_022182157.1 (Figure 18b) of *M. persicae* were aligned with the protein sequence of *P. troglodytes* (NP_001009149.1) FucT3, it was observed that motif V region is similar in both sequences.

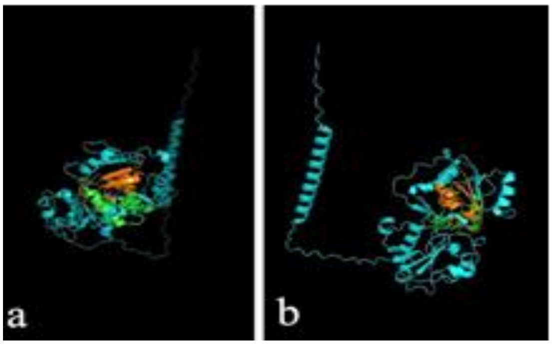

Figure 18. Comparison of the 3D structures of *M. persicae* and *P. troglodytes* (NP_001009149.1) FucT3. a) Green represents *M. persicae (*XP_022182156.1), blue represents *P. troglodytes* (NP_001009149.1), and orange represents the shared motif V b) Green represents *M. persicae* (XP_022182157.1), blue represents *P. troglodytes* (NP_001009149.1), and orange represents the shared motif V.

The protein sequence XP_022180514.1 of *M. persicae* (Table 5) and the protein sequence of *P. troglodytes* (CAA74361.1) FucT5 were aligned with each other, and their motif sequences were compared. As seen in Figure 19, the motif V region is common in both sequences.

Table 5. Presents the data with high probability of motif V of *M. persicae* FucT5

Sequence name	**Start**	**End**	**p-value**	**q-value**	**Matching sequences**
XP_022180514.1	259	277	8.22e-06	0.615	YGIYIAKDCAGFPTYKLEK

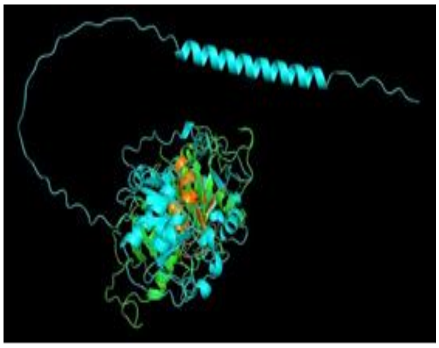

Figure 19. Comparison of the 3D structures of *M. persicae* (XP_022180514.1) and *P. troglodytes* (CAA74361.1) FucT5. Green represents *M. persicae*, blue represents *P. troglodytes* (CAA74361.1), and orange represents the shared motif V.

The uncharacterized protein sequence of *M. persicae* (XP_022160661.1) and the protein sequence of *H. sapiens* (NP_006572.2) FucT9 have been aligned, and their motif sequences have been compared. It has been determined that the motif V regions are shared (Figure 20).

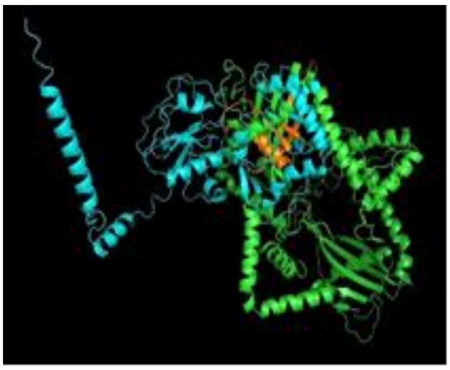

Figure 20. Comparison of 3D structures of *M. persicae* (XP_022160661.1) and *H. sapiens* (NP_006572.2) FucT9. Green represents *M. persicae*, blue represents *H. sapiens*, and orange represents the region common to both, which corresponds to motif V.

4. Conclusions and discussion

Despite the existing studies in the literature, the diversity of many fucosyltransferases in organisms remains unknown. In the model nematode *C. elegans*, a wide variety of fucosylated glycans, including N-linked oligosaccharides, are found. These N-glycans may contain up to three fucose residues in the standard N, N-diacetylchitobiose unit, but only two of them have been characterized previously as fucosyltransferases responsible for their transfer (FucT1 and FucT8) [11]. It has been demonstrated that N-glycan proximal N-acetylglucosamine-linked α1,3-linked fucose is found in plant and insect glycoproteins, including pollen, food, and venom allergens [12]. In a study conducted with the mosquito *Anopheles gambiae*, both new and known fucosylated N-glycans were identified in insects compared to other glycomic studies. Enzymatic activities and properties for all α1,3 and α1,6-fucosyltransferase homologs were demonstrated in this insect species using MALDI-TOF MS-based methods. In the *A. gambiae* genome, fucosyltransferases FucT6, FucTA, and FucTC were identified, cloned, and successfully expressed recombinantly in *Pichia pastoris*. Additionally, an α1,3-fucosyltransferase (FucTA) was also identified in *D. melanogaster* [12]. It has been revealed that the α1,6 fucosyltransferase found in invertebrates is homologous to the FucT8 in vertebrates and the α1,3-fucosyltransferase in *B. mori* is homologous to the FucTA gene in *D. melanogaster* [10].

Using bioinformatics approaches in the *D. melanogaster* genome, two orthologous genes encoding the fucosyltransferases of the insect and human FucT10 and FucT11 have been identified to describe the proteome involved in α-L-fucosylated glycan metabolism [13]. Due to the presence of two conserved key motifs, they were predicted to be α1,3-fucosyltransferases [14], however, their activities have not yet been experimentally demonstrated in any species, including humans [13], mice [15], flies [9, 16], or bees [17]. Firstly, four α1,3-fucosyltransferases (Fuc-TA, Fuc-TB, Fuc-TC, and Fuc-TD) were identified in *D. melanogaster* [9,16]. It was determined that Fuc-TA is an α1,3-fucosyltransferase [16], Fuc-TB is orthologous to human FucT10 and FucT11, Fuc-TC is likely involved in the synthesis of the Lex antigen [17], but no function has yet been identified for Fuc-TD [3].

In our study, in addition to α1,6-fucosyltransferases that can be aligned between invertebrates and vertebrates, α1,2-, α1,3-, α1,4-, and protein-O-fucosyltransferases were aligned among mammals and species including *B. mori*, *C. maculatus*, *G. mellonella*, *M. sexta*, and *M. persicae* using Clustal Omega. The presence of at least 1 and up to 4 conserved motif regions indicates that the catalytic region is highly conserved in enzymes ranging from insects to mammals.

Approximately 30% sequence divergence is observed between the same fucosyltransferase genes in birds and mammals (with the difference estimated to have occurred around 120 million years ago), while around 20% sequence difference is found among the same fucosyltransferase genes in different mammalian species. The difference between human and chimpanzee fucosyltransferase genes is approximately 2% (with divergence estimated to have occurred around 5 million years ago). There is approximately a 10% difference between the *FUT3, FUT5*, and *FUT6* genes that diverged over 10 million years ago [18]. The high similarity between fucosyltransferase sequences explains the common presence of VEN34891.1, VEN34890.1, VEN58850.1, VEN34889.1 protein sequences in all α1,3 fucosyltransferases in the *C. maculatus* species. When compared to mammals, similarities are observed in conserved motif regions in these protein sequences.

The α1,3-fucosyltransferase family possesses five conserved motif regions. The two main conserved motifs (IV and V) are present in all α1,3-fucosyltransferases [3]. In the research conducted within the scope of the thesis, when the three-dimensional structure was examined, the two main motifs, mainly motif V, were observed, confirming the literature. These two motifs facilitate the recognition and binding of GDP-fucose. The first three motifs, on the other hand, are involved in recognizing the acceptor substrate and are positioned towards the NH_2 terminus of the sequence [3].

The equivalent ratio of specific positions in the insect FucT10/11 enzymes (FucT10 and FucT11) being found at a rate of 50.9% indicates that they have the same genetic lengths and hence could share a common ancestor. The same concept applies when evaluated by Petit et al. [19] for specific positions from FucT10 or FucT11, with a percentage of 50.14, for classical monoexonic α1,3-fucosyltransferases (FucT3-7 and FucT9) located at the same genetic distance [3].

Experimental 3D structures determined for FucTs are not widely available in the literature. Furthermore, in the Protein Data Bank (PDB), there is a lack of structures with sufficient sequence similarity to FucTs for direct homology modeling. However, sequence similarities observed when comparing insect protein sequences generated using Swiss-Model with mammalian fucosyltransferases are promising for homology studies.

Literature sources [12] report the presence of different numbers of FucT enzymes in various insect species compared to those obtained from NCBI. Specifically, in the protein sequences obtained from NCBI, 4 types of FucTs are identified in *B. mori*, 5 types in *G. mellonella*, 6 types in *M. sexta*, and 3 types in *M. persicae*, while no identified enzyme is found in *C. maculatus*. However, contrary to these findings, in this study, the motifs of FucT1, FucT3, FucT5, FucT6, and FucT9 found in mammals are characterized in *B. mori* species, FucT1 and FucT9 are characterized in *G. mellonella*, FucT1, FucT4, FucT5, FucT8, and FucT11 are characterized in *M. sexta*, FucT2, FucT3, FucT5, FucT6, FucT7, FucT9, FucT10, and FucT11 are characterized in *M. persicae*, and all 13 fucosyltransferases known in mammals are characterized in *C. maculatus*.

The studies provide guidance for continuing studies on fucosyltransferases in invertebrates. New sequences for the fucosyltransferase enzyme will be revealed as the genome of determined insects is studied. While the similarities of fucosyltransferase motifs in insects and mammals indicate that they may be orthologous genes in terms of evolution, it is also an indication that new developmental and evolutionary studies will be conducted on this subject.

Conflicts of interest: No conflict of interest.

Funding: This study was funded by the Ege University Scientific Research Fund (FYL-2023-27378).

Ethical statement: This study does not require ethical approval.

Author contributions: Conception: Savaş İZZETOĞLU, Gamze TURGAY İZZETOĞLU; Design: All authors; Supervision: Gamze TURGAY İZZETOĞLU; Materials: Gamze TURGAY İZZETOĞLU; Data Collection or Processing: Sadiye OYLUM, Ceren ERSOY; Analysis or Interpretation: Sadiye OYLUM, Ceren ERSOY, Gamze TURGAY İZZETOĞLU; Literature review: All authors; Critical review: All authors.

References

[1] Oriol, R., Mollicone, R., Cailleau, A., Balanzino, L. and Breton, C. (1999). Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria, *Glycobiology* 9: 323–334. https://doi.org/10.1093/glycob/9.4.323

[2] Wang, Y., Shao, L., Shi, S., Harris, R. J., Spellman, M. W., Stanley, P., & Haltiwanger, S. (2001). Modification of Epidermal Growth Factor-like Repeats with O-Fucose. *Journal of Biological Chemistry,* 276, 40338–40345. https://doi.org/10.1074/jbc.M107849200

[3] Mollicone, R., Moore, S., Bovin, N., Rosasco, M., Candelier, J., Duncker, I., & Oriol, R. (2009). Activity, splice variants, conserved peptide motifs, and phylogeny of two new α1,3-fucosyltransferase families (*FUT10* and *FUT11*). *Journal of Biological Chemistry,* 284, 4723–4738. https://doi.org/10.1074/jbc.M809312200

[4] Shan, M., Yang, D., Dou, H., & Zhang, L. (2019). Fucosylation in cancer biology and its clinical applications. *Progress in Molecular Biology and Translational Science,* 162, 1877-1173. https://doi.org/10.1016/bs.pmbts.2019.01.002

[5] Vries, T., Knegtel, R., Holmes, E., & Macher, B. (2001). Fucosyltransferases: Structure/function studies. *Glycobiology,* 11, 119-128. https://doi.org/10.1093/glycob/11.10.119R

[6] Xu, Z., Vo, L., & Macher, B. A. (1996). Structure-Function Analysis of Human α1,3-Fucosyltransferase. *Journal of Biological Chemistry,* 271, 8818–8823. https://doi.org/10.1074/jbc.271.15.8818

[7] Dupuy, F., Germot, A. M., Oriol, R., Blancher, A., Julien, R., & Maftah, A. (2002). α1,4-Fucosyltransferase activity: A significant function in the primate lineage has appeared twice independently. *Molecular Biology Evolution,* 19, 815–824. https://doi.org/10.1093/oxfordjournals.molbev.a004138

[8] Dupuy, F., Germot, A., Julien, R., & Maftah, A. (2004). Structure/function study of Lewis α3- and a3/4-fucosyltransferases: The a1,4 fucosylation requires an aromatic residue in the acceptor-binding domain. *Glycobiology,* 14, 347–356. https://doi.org/10.1093/glycob/cwh053

[9] Paschinger, K., Staudacher, E., Stemmer, U., Fabini, G., & Wilson, I. B. H. (2005). Fucosyltransferase substrate specificity and the order of fucosylation in invertebrates. *Glycobiology,* 15, 463–474. https://doi.org/10.1093/glycob/cwi028

[10] Ihara, H., Okada, T., & Ikeda, Y. (2014). Cloning, expression and characterization of *Bombyx mori* α1,6-fucosyltransferase. *Biochemical and Biophysical Research Communications,* 450, 953–960. https://doi.org/10.1016/j.bbrc.2014.06.087

[11] Yan, S., Serna, S., Reichardt, N.C., Paschinger, K., Wilson, I. B. H. (2013). Array-assisted characterization of a fucosyltransferase required for the biosynthesis of complex core modifications of nematode N-Glycans. *Journal of Biological Chemistry,* 288, 21015–21028. https://doi.org/10.1074/jbc.M113.479147

[12] Kurz, S., King, J.G., Dinglasan, R. R., Paschinger, K., Wilson, I. B. H. (2016). The fucomic potential of mosquitoes: Fucosylated N-glycan epitopes and their cognate fucosyltransferases. *Insect Biochemistry and Molecular Biology,* 68, 52-63. https://doi.org/10.1016/j.ibmb.2015.11.001

[13] Roos, C., Kolmer, M., Mattila, P., & Renkonen, R. (2002). Composition of *Drosophila melanogaster* proteome involved in fucosylated glycan metabolism. *The Journal of Biological Chemistry,* 277, 3168–3175. https://doi.org/10.1074/jbc.M107927200

[14] Breton, C., Oriol, R. & Imberty, A. (1998). Conserved structural features in eukaryotic and prokaryotic fucosyltransferases, *Glycobiology*, 8, 87–94. https://doi.org/10.1093/glycob/8.1.87

[15] Baboval, T., & Smith, F. I. (2002). Comparison of human and mouse Fuc-TX and Fuc-TXI genes, and expression studies in the mouse. *Mammalian Genome,* 13, 538–541. https://doi.org/10.1007/s00335-001-2152-5.

[16] Fabini, G., Freilinger, A., Altmann, F., & Wilson, I. B. H. (2001). Identification of core α1,3-fucosylated glycans and cloning of the requisite fucosyltransferase cDNA from *Drosophila melanogaster*. *Journal of Biological Chemistry,* 276, 28058–28067. https://doi.org/10.1074/jbc.M100573200

[17] Rendic, D., Klaudiny, J., Stemmer, U., Schmidt, J., Paschinger, K., & Wilson, I. B. H. (2007). Towards abolition of immunogenic structures in insect cells: Characterization of a honey-bee (*Apis mellifera*) multi-gene family reveals both an allergy-related core α1,3-fucosyltransferase and the first insect Lewis-histo-blood-group-related antigen-synthesizing enzyme. *Biochemistry Journal,* 402, 105–115. https://doi.org/10.1042/BJ20060964

[18] Costache, M., Cailleau, A., Fernandez-Mateos, P., Oriol, R., & Mollicone, R. (1997). Advances in molecular genetics of and (x-3/4-fucosyltransferases). *Transfusion Clinique et Biologique,* 4, 367–382. https://doi.org/10.1016/S1246-7820(97)80042-0

[19] Petit, D., Maftah, A., Julien, R., & Petit, J. M. (2006). En bloc duplications, mutation rates, and densities of amino acid changes clarify the evolution of vertebrate α-1,3/4-fucosyltransferases. *Journal of Molecular Evolution, 63,* 353–364. https://doi.org/10.1007/s00239-005-0189-x