

Hamza Akmal

24100232

Assignment 3

BIO 101

TA: AJ

UniProtKB - P05205 (HP1_DROME)

Display [Help video](#) [BLAST](#) [Align](#) [Format](#) [Add to basket](#) [History](#) [Add a publication](#) [Feedback](#)

Entry **Protein** **Heterochromatin protein 1**
Gene **Su(var)205**
Organism *Drosophila melanogaster* (Fruit fly)
Status Reviewed - Annotation score: ●●●●● - Experimental evidence at protein levelⁱ

Functionⁱ
Structural component of heterochromatin, involved in gene repression and the modification of position-effect-variegation. Recognizes and binds histone H3 tails methylated at 'Lys-9', leading to epigenetic repression.
[2 Publications](#)

Sites

Feature key	Position(s)	Description	Actions	Graphical view	Length
Site ¹	24	Histone H3K9me2 binding			1
Site ¹	45	Histone H3K9me2 binding			1
Site ¹	48	Histone H3K9me2 binding			1

GO - Molecular functionⁱ
■ chromatin binding [Source: FlyBase](#)
■ histone binding [Source: FlyBase](#)

Answer 1:

a) 618 bp

b) 5 exons, 4 introns

All orientations in 5' to 3'

E1- CGC...TAG

E2- TAT...AAG

E3- GTG...GAG

E4- GAG...CAG

E5- CGC...TTC

I1- GTG...TTC

I2- GTG...CAG

I3- GTA...TAG

I4- GTA...AAG

- c) 2. 992 bp and 1109 bp
- d) It is a key structural part of heterochromatin and is pivotal in the repression of genes and modification of position-effect-variegation. It binds with the Histone H3 tail methylated at Lys9 and causes epigenetic repression.

There are chromo 1 and Chromo 2 domains of length 59 present in it.

- e) Homo Sapiens blast:

The screenshot displays the NCBI BLAST search results for the query **sp|P05205|HP1_DROME Heterochromatin protein...**. The search is limited to records that include Homo sapiens (taxid:9606). The results are filtered to show the top 20 sequences. The table below lists the sequences producing significant alignments, including their descriptions, scientific names, and key statistics.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
CRIS [Homo sapiens]	Homo sapiens	149	149	88%	1e-44	43.72%	191	CAG33699.1
chromobox protein homolog 5 [Homo sapiens]	Homo sapiens	146	146	88%	2e-43	43.17%	191	NP_061120793.1
HP1s-alpha [Homo sapiens]	Homo sapiens	146	146	88%	3e-43	43.17%	190	AAC50553.1
chromobox homolog 5, HP1 alpha homolog, Drosophila, isoform CRA_a [Homo sapiens]	Homo sapiens	143	143	88%	3e-42	43.72%	190	EAIV96758.1
heterochromatin-like protein 1 [Homo sapiens]	Homo sapiens	138	138	100%	2e-40	43.20%	183	AAF62376.1
chromobox protein homolog 3 [Homo sapiens]	Homo sapiens	137	137	100%	6e-40	39.81%	183	NP_059207.2
HP1s-gamma [Homo sapiens]	Homo sapiens	135	135	100%	2e-39	39.81%	173	AAB40101.1
chromobox protein homolog 1 [Homo sapiens]	Homo sapiens	130	130	83%	3e-37	43.35%	185	NP_061120760.1
heterochromatin-specific nohistone protein [Homo sapiens]	Homo sapiens	119	119	81%	3e-33	40.12%	153	AAD21972.1
Crystal structure of human chromobox homolog 1 [CRX1] [Homo sapiens]	Homo sapiens	80.9	80.9	25%	2e-19	66.04%	55	3F2U_A

We have a high score, very low e-value and decently high query coverage for the top few hits which tells us that the match of the protein is not due to chance and that the sequences are alike.

Arabidopsis Thaliana Blast:

Job Title sp|P05205|HP1_DROME Heterochromatin protein 1 OS=Drosophila melanogaster
RID E3VRFN00013 Search expires on 07-06 01:25 am [Download All](#)
Program BLASTP [Citations](#)
Database nr [See details](#)
Query ID kjQuery_13466
Description sp|P05205|HP1_DROME Heterochromatin protein 1 OS=Drosophila melanogaster
Molecule type amino acid
Query Length 206
Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)

Filter Results
Organism only top 20 will appear ☐ exclude
 Type common name, binomial, taxid or group name
[+ Add organism](#)
Percent Identity to
E value to
Query Coverage to
[Filter](#) [Reset](#)

Sequences producing significant alignments
[Download](#) [Select columns](#) [Show](#) 100

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Pos. Ident.	Acc. Len	Accession
<input checked="" type="checkbox"/> like heterochromatin protein LHP1 [Arabidopsis thaliana]	Arabidopsis thaliana	66.6	66.6	63%	3e-12	32.88%	445	AAU44363.1
<input checked="" type="checkbox"/> like heterochromatin protein LHP1 [Arabidopsis thaliana]	Arabidopsis thaliana	66.6	66.6	63%	3e-12	32.88%	435	NP_021330617.1
<input checked="" type="checkbox"/> like heterochromatin protein LHP1 [Arabidopsis thaliana]	Arabidopsis thaliana	66.6	66.6	63%	3e-12	32.88%	445	NP_191271.1
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]	Arabidopsis thaliana	66.6	66.6	63%	3e-12	32.88%	421	CA0533384.1
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]	Arabidopsis thaliana	66.6	66.6	63%	3e-12	32.88%	445	CA0543372.1
<input checked="" type="checkbox"/> TPL2 [Arabidopsis thaliana]	Arabidopsis thaliana	54.7	54.7	63%	3e-08	30.61%	452	CA051354.1
<input checked="" type="checkbox"/> like heterochromatin protein LHP1 [Arabidopsis thaliana]	Arabidopsis thaliana	54.7	54.7	63%	3e-08	30.61%	452	NP_021330615.1
<input checked="" type="checkbox"/> chromomethylene 2 [Arabidopsis thaliana]	Arabidopsis thaliana	41.2	41.2	25%	0.001	37.29%	1295	NP_193637.2
<input checked="" type="checkbox"/> chromomethylene 2 [Arabidopsis thaliana]	Arabidopsis thaliana	41.2	41.2	25%	0.001	37.29%	1247	NP_021330724.1
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]	Arabidopsis thaliana	38.1	38.1	24%	0.014	35.99%	1295	CA05255105.1
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]	Arabidopsis thaliana	38.1	38.1	24%	0.014	35.99%	1291	Y0532007.1
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]	Arabidopsis thaliana	38.1	38.1	24%	0.014	35.99%	1295	CA05338402.1
<input checked="" type="checkbox"/> chromomethylene CMT2 [Arabidopsis thaliana]	Arabidopsis thaliana	37.7	37.7	24%	0.016	35.99%	1214	AA050757.1
<input checked="" type="checkbox"/> CMT2 [Arabidopsis thaliana]	Arabidopsis thaliana	37.7	37.7	24%	0.016	35.99%	1295	CA053784.1

We have a low score and very low query coverage for the top few hits which tells us that the match of the protein is incomplete and we are probably not looking at the same sequence. There may be a lot of mutation in this protein sequence.

Musca Domestica:

Job Title: sp|P06205|HP1_DROME Heterochromatin protein...
RID: F3VRXSUF013
Program: BLASTP
Database: nr
Query ID: lc|Query_31766
Description: sp|P06205|HP1_DROME Heterochromatin protein 1 OS=
Molecule type: amino acid
Query Length: 206
Other reports: Distance tree of results Multiple alignment MSA viewer

Filter Results
Organism: only top 20 will appear
 Type common name, binomial, taxid or group name
 + Add organism
Percent Identity: to to
E value: to to
Query Coverage: to to
 Filter Reset

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
PREDICTED: heterochromatin protein 1 (Musca domestica)	Musca domestica	218	218	96%	3e-72	57.69%	208	XP_012925599.1
PREDICTED: chromodex protein homolog 1 (Musca domestica)	Musca domestica	158	158	89%	3e-49	44.57%	189	XP_005177738.1
PREDICTED: heterochromatin protein 1 like (Musca domestica)	Musca domestica	143	143	92%	3e-43	41.12%	201	XP_019894582.1
PREDICTED: heterochromatin protein 1 like (Musca domestica)	Musca domestica	102	102	33%	1e-28	68.67%	99	XP_005183726.2
PREDICTED: chromodex protein homolog 2 (Musca domestica)	Musca domestica	105	105	87%	4e-28	32.69%	227	XP_005190971.1
PREDICTED: polycomb group protein P2 (Musca domestica)	Musca domestica	62.8	62.8	25%	1e-11	53.85%	441	XP_005180443.1
PREDICTED: ABC transporter F family member 4 isoform X1 (Musca domestica)	Musca domestica	47.0	90.1	25%	3e-06	40.91%	472	XP_005184059.1
PREDICTED: chromodomain-helicase-DNA-binding protein 1 isoform X2 (Musca domestica)	Musca domestica	47.0	90.1	25%	3e-06	40.91%	471	XP_005184059.1
PREDICTED: M-phase phosphoprotein 8 isoform X4 (Musca domestica)	Musca domestica	46.6	89.3	21%	3e-06	44.44%	384	XP_019892492.1
PREDICTED: DNA (cytosine-5)-methyltransferase 1 isoform X1 (Musca domestica)	Musca domestica	46.6	89.3	21%	3e-06	44.44%	391	XP_005184060.1
PREDICTED: histone lysine N-methyltransferase Su(Hw)3 isoform X1 (Musca domestica)	Musca domestica	37.4	37.4	29%	0.004	35.71%	642	XP_005192782.1
PREDICTED: AFAFM2 family member 1 (Musca domestica)	Musca domestica	35.8	35.8	28%	0.014	36.67%	1037	XP_005179411.1
PREDICTED: chromodomain-helicase-DNA-binding protein 1 (Musca domestica)	Musca domestica	35.8	35.8	31%	0.014	32.89%	2134	XP_019893961.1

There is a match with only 3-4 hits that have a high score, high query coverage and low e-value which tells us that the protein is possibly matching the sequences in those protein sequences to a greater degree. In the other hits, the protein has a very low score and even lesser query coverage which tells us that they are probably not similar proteins.

f)

Display Entry Structure

STRNG: 7227.FBpp0079251

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Domain

Family & Domains

Sequence

Similar proteins

Cross-reference

Entry information

Miscellaneous

Secondary structure

1

Legend: Helix Turn Beta strand PDB Structure known for this area

Show more details

3D structure databases

SMR: P05205

ModBase: Search...

PDB-KB: Search...

PDB Entry	Method	Resolution	Chain	Positions	Links
1KNA	X-ray	2.10 Å	A	17-76	PDB RCSB PDB PDBj PDBsum
1KNE	X-ray	2.40 Å	A	17-76	PDB RCSB PDB PDBj PDBsum
1Q3L	X-ray	1.64 Å	A	17-76	PDB RCSB PDB PDBj PDBsum
3P7J	X-ray	2.30 Å	A/B	131-206	PDB RCSB PDB PDBj PDBsum
6ASZ	X-ray	1.52 Å	A	17-76	PDB RCSB PDB PDBj PDBsum

We used X-Ray crystallography to get the structure of this protein. Its assembly composition is heterodimer.

Answer 2:

a) HENRYCAVILL
HARRYSTYLES
8-1057-10-1-2-3-2

Score: 14

b) CHRISPRATT
CHRISEVANS
98544-1-3401

Score: 31

Citations:

For Blast results:

Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

For function and domain results:

"Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3."
Jacobs S.A., Taverna S.D., Zhang Y., Briggs S.D., Li J., Eissenberg J.C., Allis C.D., Khorasanizadeh S.

EMBO J. 20:5232-5241(2001) [PubMed] [Europe PMC] [Abstract]

Cited for: FUNCTION, MUTAGENESIS OF VAL-26.

"Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail."

Jacobs S.A., Khorasanizadeh S.

Science 295:2080-2083(2002) [PubMed] [Europe PMC] [Abstract]

Cited for: X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 17-76, FUNCTION.

For protein structure:

<https://www.ebi.ac.uk/pdbe/entry/pdb/1KNA>