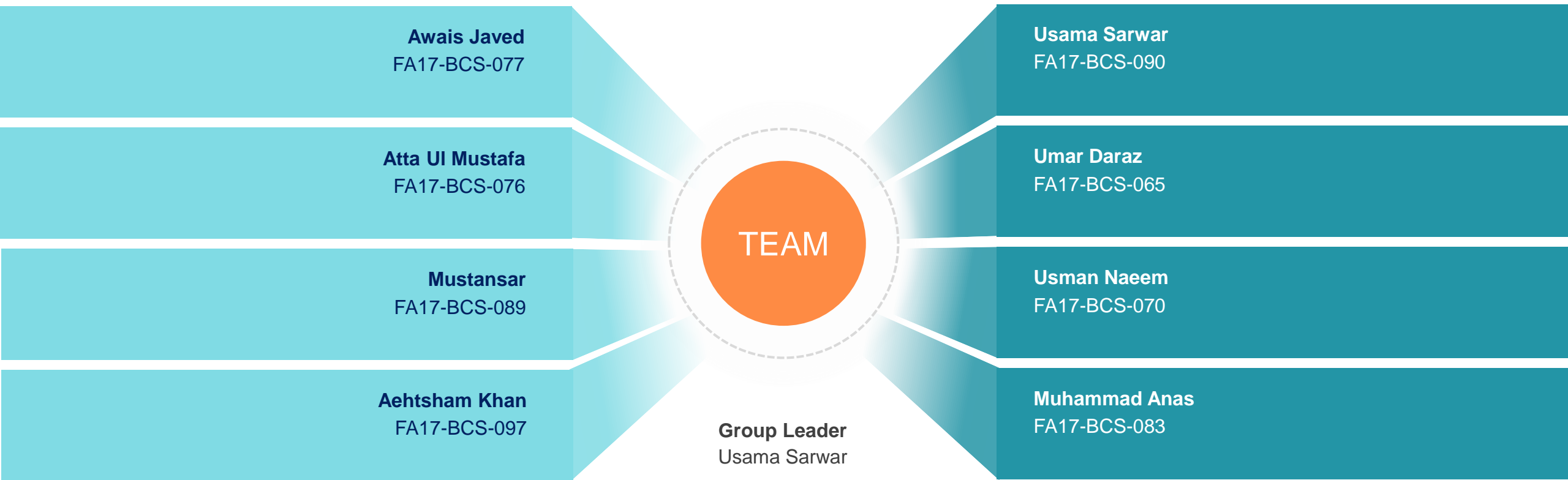


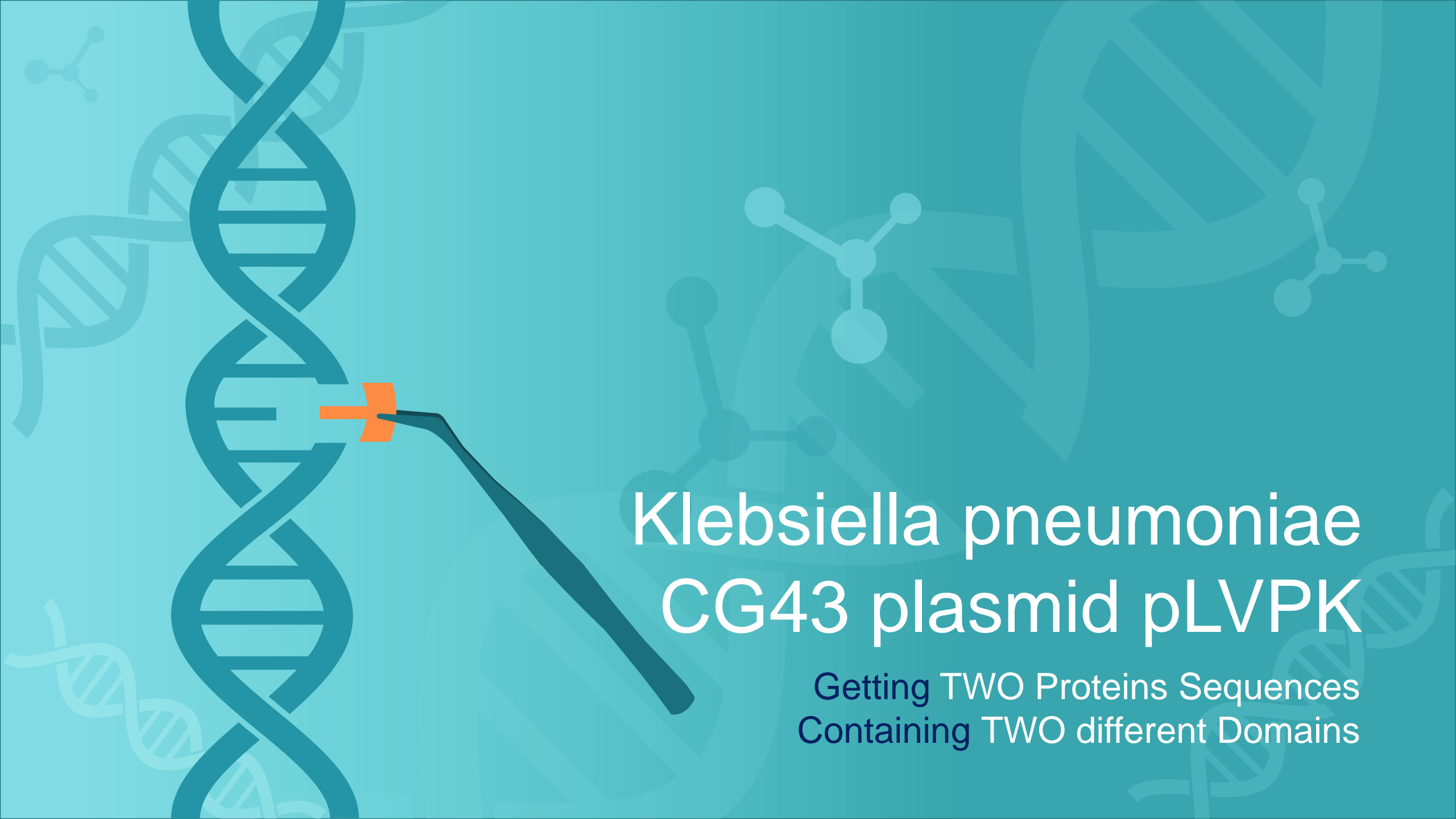


# Bioinformatics

COMSATS University Islamabad, Sahiwal Campus

# GROUP 10





# Klebsiella pneumoniae CG43 plasmid pLVPK

Getting TWO Proteins Sequences  
Containing TWO different Domains



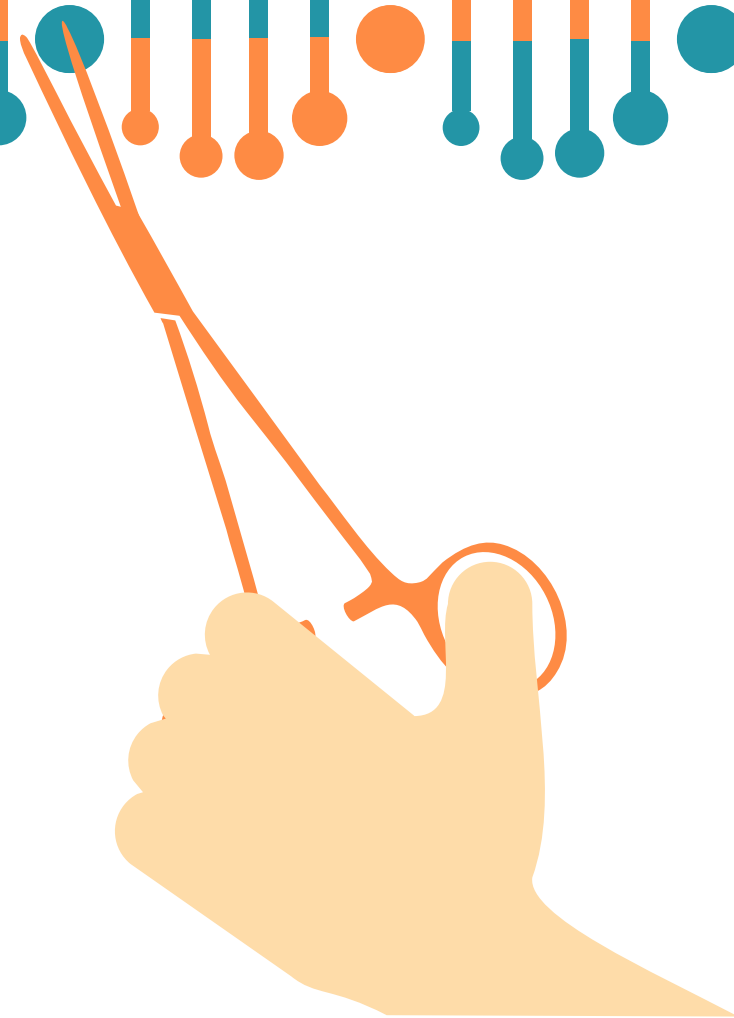
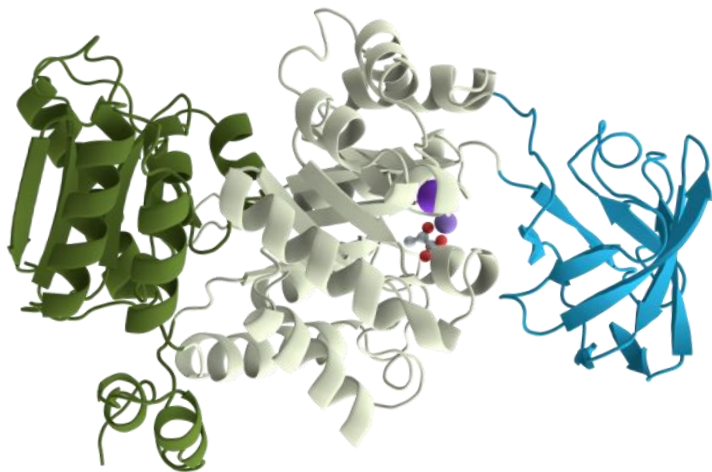
# Proteins

Proteins are large **biomolecules**, or macromolecules, consisting of one or more **long chains of amino acid** residues. Proteins **perform a vast array of functions within organisms**, including catalyzing metabolic reactions, DNA replication, responding to stimuli, providing structure to cells, and organisms, and transporting molecules from one location to another. Proteins differ from one another primarily in their **sequence of amino acids**, which is **dictated by the nucleotide sequence of their genes**, and which usually results in protein folding into a specific 3D structure that determines its activity.

# Protein Domains



Many **proteins are composed of several protein domains**, i.e. segments of a protein that fold into distinct structural units. **Domains usually also have specific functions**, such as enzymatic activities (e.g. kinase) or they serve as binding modules.







# Prerequisites

## 01 UniProt

UniProt is a freely accessible database of protein sequence and functional information, many entries being derived from genome sequencing projects.

## 02 FASTA Sequence

In bioinformatics and biochemistry, the FASTA format is a text-based format for representing either nucleotide sequences or amino acid sequences, in which nucleotides or amino acids are represented using single-letter codes.

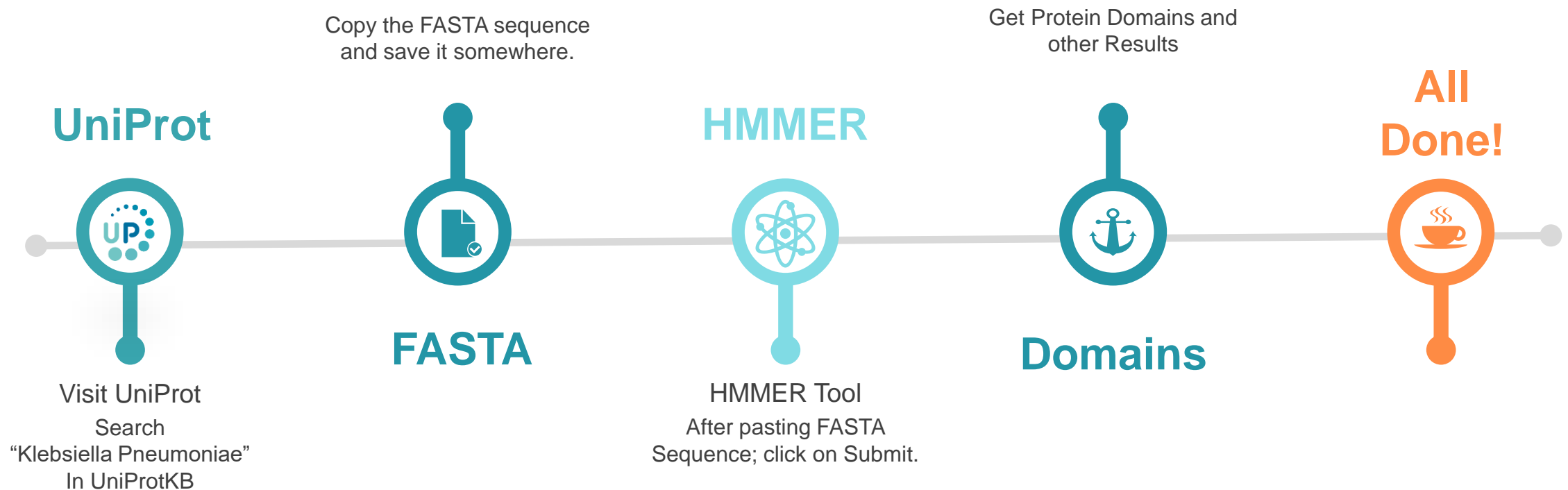
## 03 HMMER Tool

Biosequence analysis using profile hidden Markov Models  
Protein sequence VS Profile-HMM database

## 04 Pfam

Pfam is a database of protein families that includes their annotations and multiple sequence alignments generated using hidden Markov models.

# Steps to find Protein Domain







**TUTORIAL**

UniProtKB

BLAST Align Retrieve/ID mapping Peptide search SPARQL Help Contact

## UniProtKB results

Filter by:

1 to 25 of 489,812 Show 25

Reviewed (1,707)

Unreviewed (488,105)

### Popular organisms

Mouse (4)

E. coli K12 (3)

Human (1)

KLEPN (265,530)

KLEP7 (5,127)

Other organisms

Search terms

Quote terms: "klebsiella pneumoniae"

<input type="checkbox"/>	Entry	Entry name		Protein names	Gene names	Organism	Length
<input type="checkbox"/>	A6T923	HPXO_KLEP7		<b>FAD-dependent urate hydroxylase</b>	hpxO KPN78578_16330, KPN_01663	Klebsiella pneumoniae subsp. pneumoniae (strain ATCC 700721 / MGH 78578)	384
<input type="checkbox"/>	A6T7D6	PYRC_KLEP7		<b>Dihydroorotase</b>	pyrC KPN78578_10460, KPN_01074	Klebsiella pneumoniae subsp. pneumoniae (strain ATCC 700721 / MGH 78578)	348
<input type="checkbox"/>	C7C422	BLAN1_KLEPN		<b>Metallo-beta-lactamase type 2</b>	blaNDM-1	Klebsiella pneumoniae	270
<input type="checkbox"/>	Q9AGA6	AGLB_KLEPN		<b>6-phospho-alpha-glucosidase</b>	agIB	Klebsiella pneumoniae	440
<input type="checkbox"/>	Q93LQ8	BGLK_KLEPN		<b>Beta-glucoside kinase</b>	bglK	Klebsiella pneumoniae	297
<input type="checkbox"/>	O85746	TYRB_KLEPN		<b>Tyrosine aminotransferase</b>	tyrB	Klebsiella pneumoniae	397

hpxO - FAD-dependent urate hyc

uniprot.org/uniprot/A6T923

Display

Sequence<sup>i</sup>

Entry

Publications

Feature viewer

Feature table

Sequence status<sup>i</sup>: Complete.

A6T923-1 [UniParc] [FASTA](#) [Add to basket](#)

« Hide

Length: 384

Mass (Da): 42,167

Last modified: August 21, 2007 - v1

Checksum: 4B656520269C3535

BLAST [GO](#)

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequence

Similar proteins

Cross-references

Entry information

Miscellaneous

10	20	30	40	50
MKAIVIGAGI	GGLSAAVALK	QSGIDCDVYE	AVKEIKPVGA	AISVWPNGVK
60	70	80	90	100
CMAHLGMGDI	METFGGPLRR	MAYRDFRSGE	NMTQFSLAPL	IERTGSRPCP
110	120	130	140	150
VSRAELQREM	LDYWGRDSVQ	FGKRVTRCEE	DADGVTWFT	DGSSASGDLL
160	170	180	190	200
IAADGSHSAL	RPWVLGFTPQ	RRYAGYVNW	GLVEIDEALA	PGDQWTTFGV
210	220	230	240	250
EGKRVSLMPV	SAGRFYFFFD	VPLPAGLAED	RDTLRADLSR	YFAGWAPPVQ
260	270	280	290	300
KLIAALDPQT	TNRIEIHDI	PFSRLVRGRV	ALLGDAGHST	TPDIGQGGCA
310	320	330	340	350
AMEDAVVLGA	VFRQTRDIAA	ALREYEAQRC	DRVRDLVLKA	RKRCIDITHGK
360	370	380		
DMQLTEAWYQ	ELREETGERI	INGMCDTILS	GPLG	

Sequence databases

Select the link [CP000647 Genomic DNA Translation: ABR77094.1](#)

>sp|A6T923|HPX0 KLEP7 FAD-dependent urate hydroxylase OS=Klebsiella pneumoniae subsp. pneumoniae (strain ATCC 700721 / MGH 78578) OX=272620 GN=hpx0 PE=1 SV=1  
MKAIVIGAGIGGLSAAVALKQSGIDCDVYEAVKEIKPVGAAISVWPNGVKMAHLGMGDI  
METFGGPLRRMAYRDFRSGENMTQFSLAPLIERTGSRPCPVSRaelQREMLDYWGRDSVQ  
FGKRVTRCEEDADGVTWFTDGSSASGDLLIAADGSHSALRPWVLGFTPQRRYAGYVWNW  
GLVEIDEALAPGDQWTTFVGEGKRVSLMPVSAGRFFFDVPLPAGLAEDRDTLRADLSR  
YFAGWAPPVQKLI AALDPQTTNRIEIHDI EPFSRLVRGRVALLGDAGHSTTPDIGQGGCA  
AMEDAVVLGAVFRQTRDIAAALREYEAQRCDRVRDLVLKARKRCDITHGKDMQLTEAWYQ  
ELREETGERIINGMCDTILSGPLG

4



hmmsearch | HMMER

ebi.ac.uk/Tools/hmmer/search/hmmsearch

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

# HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

phmmer **hmmsearch** hmmscan jackhmmer

## protein sequence vs profile-HMM database

Paste a Sequence | Upload a File | Accession Search

Paste in your sequence or use the [example](#)

```
MKAIVIGAGIGLSAAVALKQSGIDCDVYEAVKEIKPVGAAISVWPNGVKMAHLGMGDI
METFGGPLRRMAYRDFRSGENMTQFSLAPLIERTGSRPCPVSRALQREMLDYWGRDSVQ
FGKRVTRCEEDADGVTWFTDGSSASGDLLIAADGSHSALRPWWLGFTPQRRYAGYVNW
GLVEIDEALAPGDQWTFVGEGRVSLMPVSAGRFFFDVPLPAGLAEDRDTLRADLSR
YFAGWAPPVQKLIAALDPQTTNRIEIHDIPEFSRLVGRVALLGDAGHSTTPDIGGGCA
AMEDAVLGAVFRQTRDIAAALREYEAQRCDRVRDLVLKARKRCDITHGKDMQLTEAWYQ
ELREETGERIINGMCDTILSGPLG
```

Submit Reset

5

HMM Database



# HMMER

Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

## HMMSCAN Results

Search Again

Score

Download

## Sequence Matches and Features



✓ disorder ✓ coiled-coil ✓ tm & signal peptide

## Pfam Matches

Advanced

Family		Clan	Description	Cross-references	Start	End	Domain E-values	
Id	Accession						Ind.	Cond.
> FAD_binding_3	PF01494.19	CL0063	FAD binding domain		1	340	3.7e-21	6.1e-25
> NAD_binding_8	PF13450.6	CL0063	NAD(P)-binding Rossmann-like domain		5	70	2.0e-05	3.3e-09

Your search took: 0.04 secs

UniProtKB  Advanced Search

BLAST Align Retrieve/ID mapping Peptide search SPARQL Help Contact

# UniProtKB results

[About UniProtKB](#) Basket

Filter by: BLAST Align Download Add to basket Columns  1 to 25 of 489,812 Show 25

Reviewed (1,707)

Unreviewed (488,105)

### Popular organisms

Mouse (4)

E. coli K12 (3)

Human (1)

KLEPN (265,530)

KLEP7 (5,127)

Other organisms  Go

Quote terms: "klebsiella pneumoniae"






<input type="checkbox"/>	Entry	Entry name		Protein names	Gene names	Organism	Length	
<input type="checkbox"/>	A6T923	HPXO_KLEP7		<b>FAD-dependent urate hydroxylase</b>	hpxO KPN78578_16330, KPN_01663	Klebsiella pneumoniae subsp. pneumoniae (strain ATCC 700721 / MGH 78578)	384	
<input type="checkbox"/>	A6T7D6	PYRC_KLEP7		<b>Dihydroorotase</b>	pyrC KPN78578_10460, KPN_01074	Klebsiella pneumoniae subsp. pneumoniae (strain ATCC 700721 / MGH 78578)	348	
<input type="checkbox"/>	C7C422	BLAN1_KLEPN		<b>Metallo-beta-lactamase type 2</b>	blaNDM-1	Klebsiella pneumoniae	270	
<input type="checkbox"/>	Q9AGA6	AGLB_KLEPN		<b>6-phospho-alpha-glucosidase</b>	aglB	Klebsiella pneumoniae	440	
<input type="checkbox"/>	Q93LQ8	BGLK_KLEPN		<b>Beta-glucoside kinase</b>	bglK	Klebsiella pneumoniae	297	
<input type="checkbox"/>	O85746	TYRB_KLEPN		<b>Tyrosine aminotransferase</b>	tyrB	Klebsiella pneumoniae	397	











# Comparison

## Protein A

	Family		Clan	Description	Cross-references	Start	End	Domain E-values	
	Id	Accession						Ind.	Cond.
>	<a href="#">FAD_binding_3</a>	<a href="#">PF01494.19</a>	<a href="#">CL0063</a>	FAD binding domain	  	1	340	3.7e-21	6.1e-25
>	<a href="#">NAD_binding_8</a>	<a href="#">PF13450.6</a>	<a href="#">CL0063</a> 	NAD(P)-binding Rossmann-like domain	 	5	70	2.0e-05	3.3e-09

## Protein B

	Family		Clan	Description	Cross-references	Start	End	Domain E-values	
	Id	Accession						Ind.	Cond.
>	<a href="#">Glyco_hydro_4</a>	<a href="#">PF02056.16</a>	<a href="#">CL0063</a>	Family 4 glycosyl hydrolase	  	5	185	2.0e-65	2.2e-69
>	<a href="#">Glyco_hydro_4C</a>	<a href="#">PF11975.8</a>	<a href="#">CL0341</a>	Family 4 glycosyl hydrolase C-terminal domain	  	195	415	3.6e-51	4.0e-55



Thank You