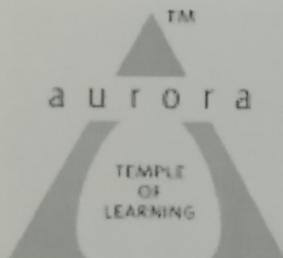


A Systematic Study of HXK1 gene with the help of bioinformatic databases

By

Thammi Meharsai
(1051-21-467-001)



Department of Biotechnology,
Aurora's Degree & PG College, Chikkadpally
(Accredited by NACC with "B++" grade)
2022-2023



**FACULTY OF SCIENCE
DEPARTMENT OF GENETICS
AURORA's DEGREE & PG COLLEGE**
(Accredited by NACC with "B++" grade)

CERTIFICATE

This is to certify that the dissertation report entitled, "**A Systematic Study of HXK1 gene with the help of bioinformatic databases**" is the result of project work carried out by **Thammi Meharsai (1051-21-467-001)** at the Department of Biotechnology, Aurora's Degree & PG College, Hyderabad, 500020 under my supervision and guidance during the academic year 2022-2023.

Guided By:

G. Rahul

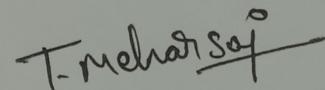
Department of Biotechnology
Aurora's Degree & PG College

Head
Department of Biotechnology
Aurora's Degree & PG College
Chikkadpally, Hyderabad-20.

DECLARATION

I hereby, declare that the project work entitled "**A Systematic Study of HXK1 gene with the help of bioinformatic databases**" has been carried out by **Thammi Meharsai (1051-21-467-001)** an original work done and submitted by in **Aurora's Degree & Pg College, Hyderabad** is a project work done by me under the guidance of **G. Rahul**, Head, Department of Biotechnology, Aurora's Degree & PG College, Hyderabad, during the course of year 2022-2023.

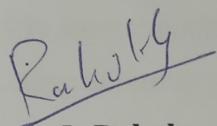
DATE : 26/04/23
PLACE: Hyderabad



(Thammi Meharsai)

CERTIFICATE

This is to certify that the project entitled, "**A Systematic Study of HXK1 gene with the help of bioinformatic databases**" is a bona fide work carried out by Thammi Meharsai(1051-21-467-001), student of B.Sc.- MSCS, Aurora's Degree & PG College for a period of 3 months from 17th November 2022 to 17th February , 2023 . This work has been carried out under my guidance and supervision and the results embodied in the project work have not been submitted to any other University or Institution for the award of any degree or diploma.



G. Rahul

Head,

Department of Biotechnology
Aurora's Degree & PG College

Head

**Department of Biotechnology
Aurora's Degree & PG College
Chikkadpally, Hyderabad-20.**

ACKNOWLEDGMENT

- My deepest thanks to my mentor G. Rahul, Head of Biotechnology, Aurora's Degree and PG College, for giving me an opportunity to work under him and for the guidance, hard work, encouragement, and motivation during the session of my project work.
- My profound thanks to the Principal, Dr Vishwanatham Bulusu, without whom this project could not have been complete.
- My heartfelt thanks to the B. Rajini, Head of department, Computer Science, faculty of computer science and lab assistants for their support in using computer facilities and solving the queries on various programming processes
- I would like to heartfully thank my fellow group members for their enlightening ideas and constant patience throughout the project
- My sincere gratitude to my parents, family members, and friends for their courtesy, companionship, and support.

2023/04/27 10:42

Contents

	page no
1. Introduction to biology	1
1.1 cell,tissue,organ,organ system and organism	3
1.2 types of biological organisms	5
1.3 cell	6
1.4 type of cells	9
2. Molecular biology	14
2.1 gene and genome	15
2.2 central dogma	18
2.3 Transcription, translation, and genetic code	19
3. Introduction to bioinformatics	22
3.1 primary and secondary database	23
3.2 ncbi	24
3.3 uniprot	25
3.4 RCSB	25
3.5 RasMol	25
4. Methodology	26
5. Result/output	29
5.1 HXK1	29
5.2 about the HXK1	30
5.3 gene id file	31
5.4a nucleotide id file	34
5.4b nucleotide id fasta file	37
5.5a protein id file	39
5.5b protein id fasta file	42
5.6 uniprot id file	44
5.7a Alpha fold	51
5.7b Rasmol	55
6. Conclusion	57

1. Introduction to biology

Biology is the scientific study of life. First, we will consider the question, “What is life?” We all have an intuitive sense of what life is. If we see a rabbit on a rock, we know that the rabbit is alive and the rock is not. But it is difficult to state just what makes the rabbit alive. Likewise, in the instant, after an individual dies, we may wonder what invisible essence has transformed the living into the dead. One way to define life is to list its basic components. The cell is the basic unit of life; every organism, or living individual, consists of one or more cells. Every cell has an outer membrane that separates it from its surroundings. This membrane encloses the water and other chemicals that carry out the cell’s functions. One of those biochemicals, deoxyribonucleic acid (DNA), is the informational molecule of life (figure 1.1).

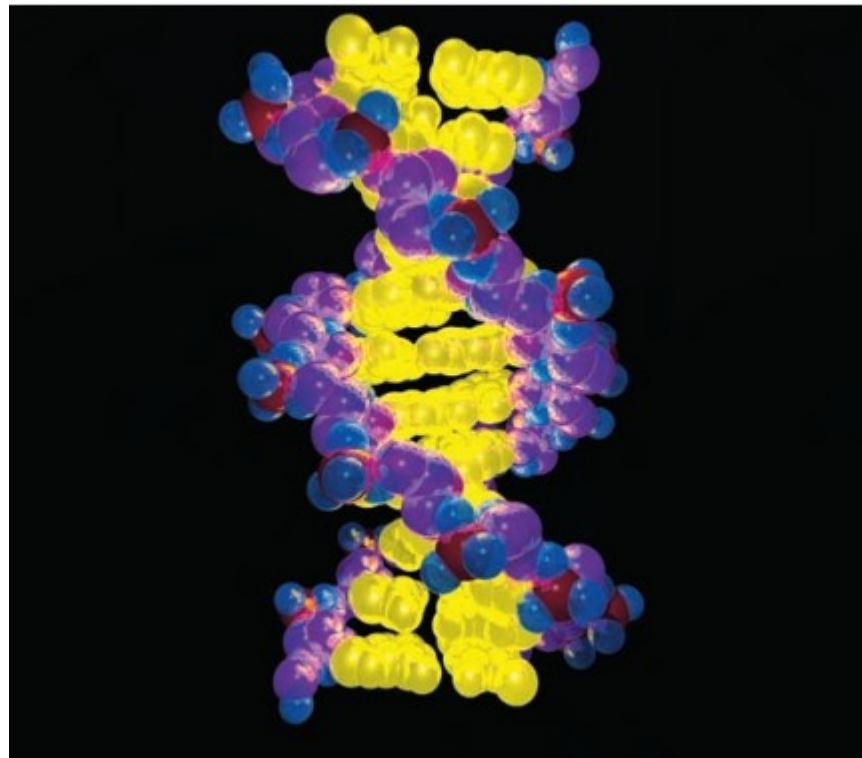


Figure 1.1 Informational Molecule of Life. All cells contain DNA, a series of “recipes” for proteins that each cell can make

Cells use genetic instructions—as encoded in DNA—to produce proteins, which enable cells to carry out specialized functions in tissues, organs, and organ systems. A list of life’s biochemicals, however, provides an unsatisfying definition of life. After all, placing DNA, water, proteins, and a membrane in a test tube does not create artificial life. And a crushed insect still contains all of the biochemicals that it had immediately before it died. In the absence of a concise definition, scientists have settled on five qualities that, in combination, constitute life (table 1.1).

Table 1.1 Characteristics of Life

Characteristic	Example
Organization	Atoms make up molecules, which make up cells, which make up tissues, and so on.
Energy use	A kitten uses the energy from its mother's milk to fuel its own growth
Maintenance of internal constancy	The kidneys regulate the body's water balance by adjusting the concentration of urine
Reproduction, growth, and development	An acorn germinates, develops into an oak seedling, and, at maturity, reproduces sexually to produce its own acorns.
Evolution	Increasing numbers of bacteria survive treatment with antibiotic drugs.

An organism is a collection of structures that function together and exhibit all of these qualities. Note, however, that each of the traits listed in table 1.1 may also occur in nonliving objects. A rock crystal is highly organized, but it is not alive. A fork placed in a pot of boiling water absorbs heat energy and passes it to the hand that grabs it, but this does not make the fork alive. A fire can “reproduce” and grow very rapidly, but it lacks most of the other characteristics of life.

1.1 cell,tissue,organ,organ system and organism

Just as the city where you live belong to a county, state, and nation, living matter also consists of parts organized in a hierarchical pattern (figure 1.2). At the smallest scale, all living structures are composed of particles called atoms, which bond together to form molecules. These molecules form organelles, which are compartments that carry out specialized functions in cells (note that not all cells contain organelles). Many organisms consist of single cells. In multicellular organisms such as the tree illustrated in figure 1.2, however, the cells are organized into specialized tissues that make up organs such as leaves. Multiple organs are linked into an individual’s organ systems. Organization in the living world extends beyond the level of the individual organism. A population includes members of the same species living in the same place at the same time. A community includes the populations of different species in a region, and an ecosystem includes both the living and nonliving components of an area. Finally, the biosphere

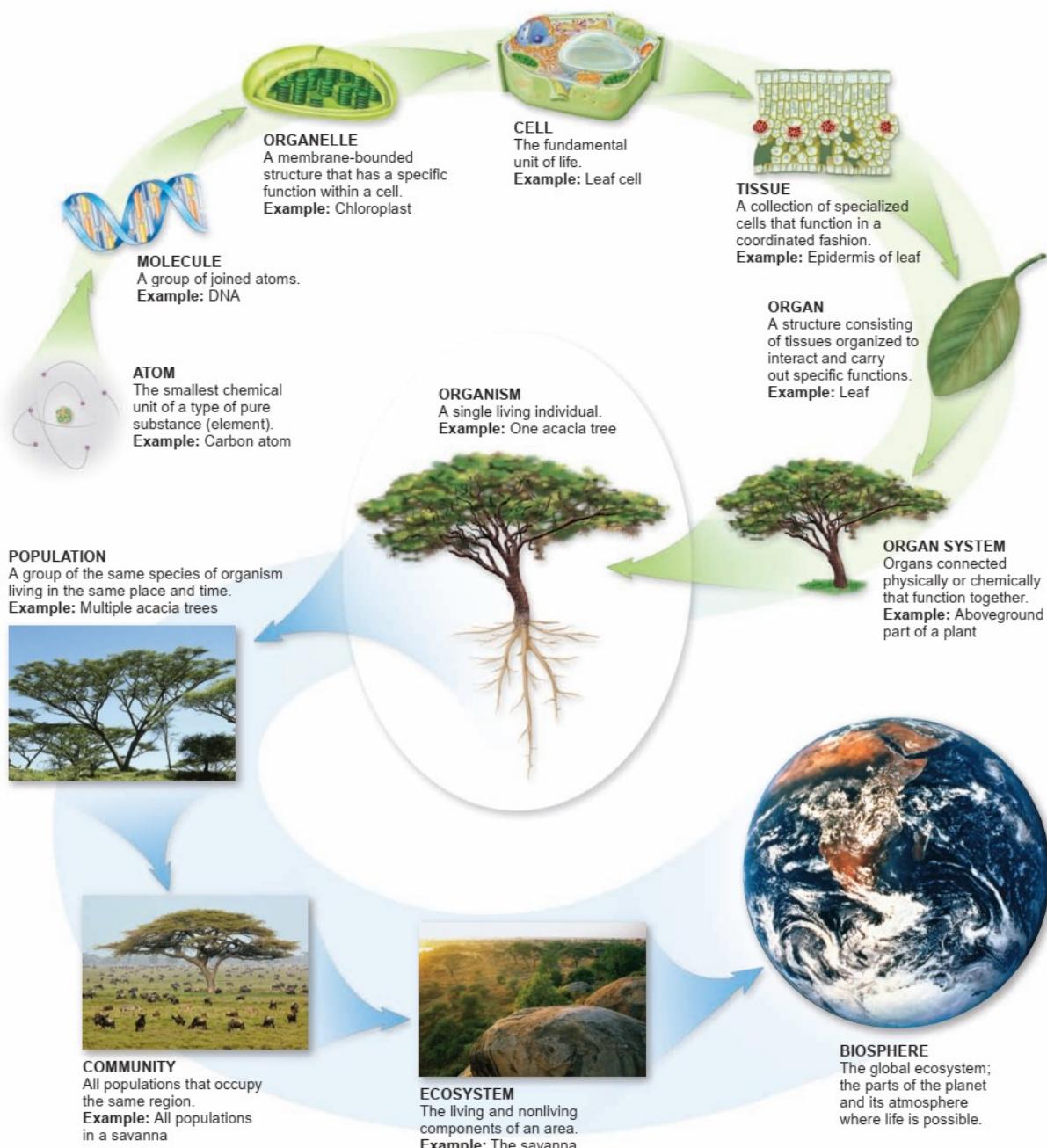


Figure 1.2 Levels of Biological Organization. Atoms arranged into molecules make up the parts of a cell. Multiple cells are organized into tissues, which make up organs and, in turn, organ systems. An individual organism may consist of one or many cells. A population consists of individuals of the same species, and communities are multiple populations sharing the same space. Communities interact with the nonliving environment to form ecosystems, and the biosphere consists of all places on Earth where life occurs.

refers to all parts of the planet that can support life. The biological organization is apparent in all life. Humans, eels, and evergreens, although outwardly very different, are all organized into specialized cells, tissues, organs, and organ systems. Single-

celled bacteria, although less complex than animals or plants, still contain DNA, proteins, and other molecules that interact in highly organized ways. An organism, however, is more than a collection of successively smaller parts. When those components interact, they create new, complex functions called emergent properties. These characteristics arise from physical and chemical interactions among a system's components, much like flour, sugar, butter, and chocolate can become brownies—something not evident from the parts themselves.

Emergent properties explain why the structural organization is closely tied to function. Disrupt a structure, and its function ceases. Shaking a fertilized hen's egg, for instance, disturbs critical interactions and stops the embryo from developing. Likewise, if a function is interrupted, the corresponding structure eventually breaks down, much as unused muscles begin to waste away. Biological function and form are interdependent.

1.2 types of biological organisms

axonomic category. Figure 1.3 depicts the three domains: Bacteria, Archaea, and Eukarya. Species in domains Bacteria and Archaea are superficially similar to one another; all are single-celled prokaryotes, meaning that their DNA is free in the cell and not confined to an organelle called a nucleus. Major differences in DNA sequences separate these two domains from each other. Domain Eukarya, on the other hand, Biologists have been studying life for centuries, documenting the existence of everything from bacteria to blue whales. An enduring problem has been how to organize the ever-growing list of known organisms into meaningful categories. Taxonomy is the biological science of naming and classifying organisms. The basic unit of classification is the species, which designates a distinctive “type” of organisms. Closely related species, in turn, are grouped into the same genus. Together, the genus and species denote the unique scientific name of each type of organism. A human, for example, is *Homo sapiens* (note that scientific names are always italicized). By assigning each type of organism a unique scientific name, taxonomists help other biologists communicate with one another. But taxonomy involves more than simply naming species.

Taxonomists also strive to classify organisms according to what we know about evolutionary relationships; that is, how recently one type of organism shared an ancestor with another type of organism. The more recently they diverged from a shared ancestor, the more closely related we presume the two types of organisms to be. Researchers infer these relationships by comparing anatomical, behavioral, cellular, genetic, and biochemical characteristics. It is enough to know that genetic evidence suggests that all species fall into one of three domains, the broadest (most inclusive)

contains all species of eukaryotes, which are unicellular or multicellular organisms whose cells contain a nucleus.

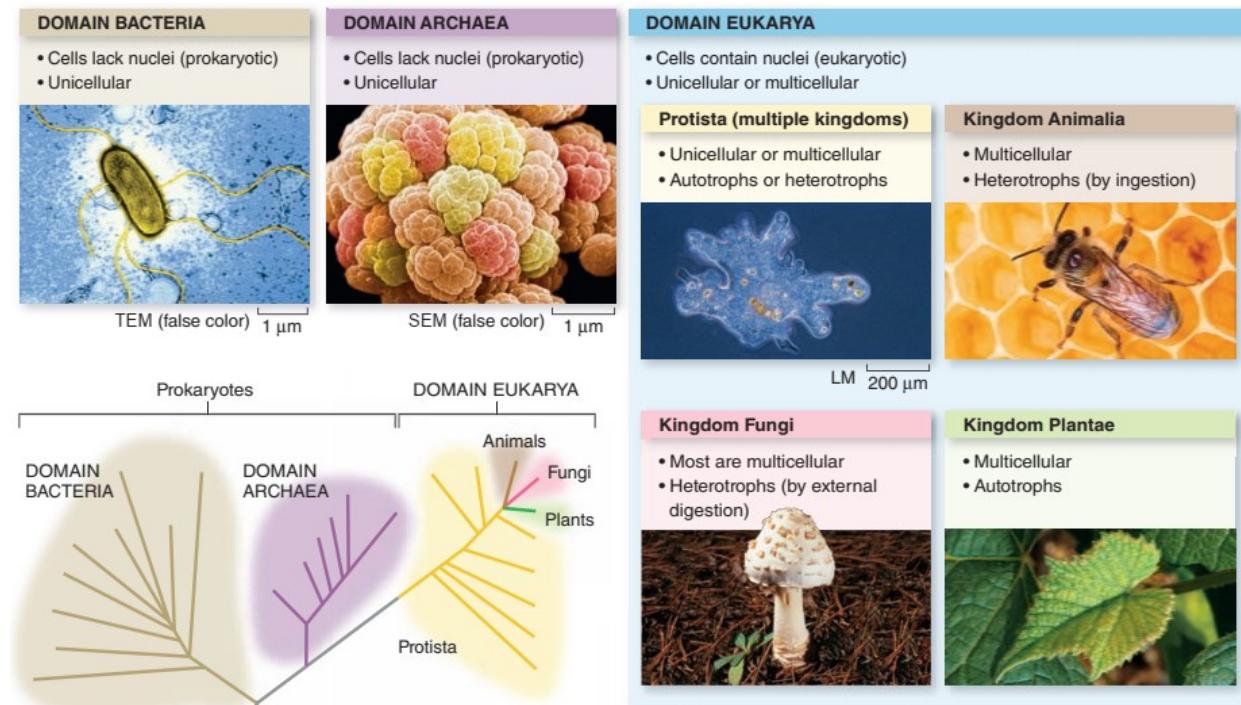


Figure 1.3 Life's Diversity. The three domains of life (Bacteria, Archaea, and Eukarya) arose from a hypothetical common ancestor

The species in each domain are further subdivided into kingdoms; figure 1.3 shows the kingdoms within the domain Eukarya. Three of these kingdoms—Animalia, Fungi, and Plantae—are familiar to most people. Within each one, organisms share the same general strategy for acquiring energy. For example, plants are autotrophs. Fungi and animals are consumers, although they differ in the details of how they obtain food. But the fourth group of eukaryotes, the Protista, contains a huge collection of unrelated species. Protista is a convenient but artificial “none of the above” category for the many species of eukaryotes that are not plants, fungi, or animals.

1.3 cell

A human, a hyacinth, a mushroom, and a bacterium appear to have little in common other than being alive. However, on a microscopic level, these organisms share many similarities. For example, all organisms consist of microscopic structures called cells, the smallest unit of life that can function independently. Within cells, highly coordinated biochemical activities carry out the basic functions of life.

The study of cells began in 1660 when English physicist Robert Hooke melted strands of spun glass to create lenses. He focused on bee stingers, fish scales, fly legs, feathers, and any type of insect he could hold still. When he looked at cork, which is the bark from a type of oak tree, it appeared to be divided into little boxes, left by cells that

were once alive. Hooke called these units “cells” because they looked like the cubicles (Latin, *cellae*) where monks studied and prayed. Although Hooke did not realize the significance of his observation, he was the first person to see the outlines of cells. His discovery initiated a new field of science, now called cell biology.

In 1673, Antony van Leeuwenhoek of Holland improved lenses further (figure 1.4a). He used only a single lens, but it produced a clearer and more highly magnified image than most two-lens microscopes then available. One of his first objects of the study was tartar scraped from his own teeth, and his words best describe what he saw there:

“To my great surprise, I found that it contained many very small animalcules, the motions of which were very pleasing to behold. The motion of these little creatures, one among another, may be likened to that of a great number of gnats or flies disporting in the air.”

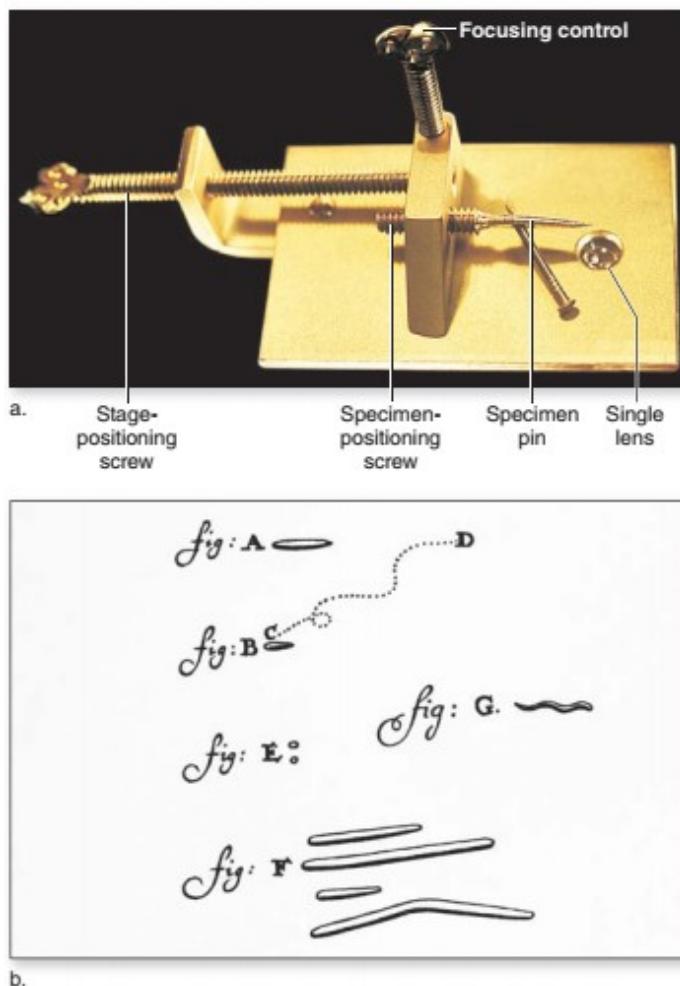


Figure 1.4 Early Microscope. (a) Antony van Leeuwenhoek made many simple microscopes like this one. The object he was studying would have been at the tip of the specimen pin. (b) Leeuwenhoek's sketches were the first record of microorganisms.

Leeuwenhoek opened a vast new world to the human eye and mind (figure 1.4b). He viewed bacteria and protists that people hadn't known existed. He also described, with remarkable accuracy, microscopic parts of larger organisms, including human red blood cells and sperm. However, he failed to see the single-celled "animalcules" reproduce.

Microscopes and other tools clearly reveal that although cells can appear very different, they all have some of the same features. All cells, from the simplest to the most complex, have the following structures and molecules in common that allow them to reproduce, grow, respond to stimuli, and obtain energy:

- DNA, the cell's genetic information;
- RNA, which participates in the production of proteins
- Ribosomes, structures that manufacture proteins;
- Proteins that carry out all of the cell's work, from orchestrating reproduction to processing energy to regulating what enters and leaves the cell;
- Cytoplasm, the fluid that occupies much of the volume of the cell
- a lipid-rich cell membrane (also called the plasma membrane) that forms a boundary between the cell and its environment

One other feature common to nearly all cells is small size, typically less than 0.1 millimeters in diameter. Why so tiny? The answer is that nutrients, water, oxygen, carbon dioxide, and waste products enter or leave a cell through its surface. Each cell must have an abundant surface area to accommodate these exchanges. As an object grows, however, its volume increases much faster than its surface area. Figure 1.5a illustrates this principle for a series of cubes, but the same applies to cells: small cell size maximizes the ratio of surface area to volume

Size of cube		
1 cm	2 cm	3 cm
Surface area = height x width x number of sides		
$1 \text{ cm} \times 1 \text{ cm} \times 6$ $= 6 \text{ cm}^2$	$2 \text{ cm} \times 2 \text{ cm} \times 6$ $= 24 \text{ cm}^2$	$3 \text{ cm} \times 3 \text{ cm} \times 6$ $= 54 \text{ cm}^2$
Volume = height x width x length		
$1 \text{ cm} \times 1 \text{ cm} \times 1 \text{ cm}$ $= 1 \text{ cm}^3$	$2 \text{ cm} \times 2 \text{ cm} \times 2 \text{ cm}$ $= 8 \text{ cm}^3$	$3 \text{ cm} \times 3 \text{ cm} \times 3 \text{ cm}$ $= 27 \text{ cm}^3$
Ratio of surface area to volume		
$6/1 = 6.0$	$24/8 = 3.0$	$54/27 = 2.0$



Figure 1.5 The Relationship Between Surface Area and Volume. (a) This simple example shows that smaller objects have more surface area relative to their volume than larger objects with the same overall size. (b) A micrograph of a amoeba-like protist, showing its irregular shape and internal organelles. Scale bars indicate 4 μm and 50 μm.

shape. (b) The membrane of this amoeba is highly folded, producing a large surface area relative to the cell's volume

Cells avoid surface area limitations in several ways. Nerve cells may be long (up to a meter or so), but they are also extremely thin, so the ratio of surface area to volume remains high. The flattened shape of a red blood cell maximizes its ability to carry oxygen, and the many microscopic extensions of an amoeba's membrane provide a large surface area for absorbing oxygen and capturing food (figure 1.5 b). A transportation system that quickly circulates materials throughout the cell also helps. The concept of the surface area is everywhere in biology. A pine tree's pollen grains have extensions that maximize flotation on air currents; root hairs have tremendous surface area for absorbing water; the broad, flat leaves of plants maximize exposure to light; a fish's feathery gills absorb oxygen from water; a jackrabbit's enormous ears help the animal lose excess body heat in the desert air—the list goes on and on. Conversely, low surface areas minimize the exchange of materials or heat with the environment. A hibernating animal, for example, conserves warmth by tucking its limbs close to its body.

1.4 Types of Cells

Until recently, biologists recognized just two types of cells, prokaryotic and eukaryotic. Prokaryotes, the simplest and most ancient forms of life, are organisms whose cells lack a nucleus (pro = before; karyon = kernel, referring to the nucleus). About 2.7 billion years ago, prokaryotes gave rise to eukaryotes, whose cells contain a nucleus and other membranous organelles (eu = true). In 1977, however, microbiologist Carl Woese studied key molecules in many cell types and detected differences that suggested that some prokaryotes represented a completely different form of life. Biologists subsequently divided life into three domains: Bacteria, Archaea, and Eukarya (figure 1.6).

	Cell Type	Nucleus	Membrane-bound Organelles	Membrane Chemistry	Cell Wall Chemistry	Typical Size
Domain Bacteria	Prokaryotic	Absent	Absent	Fatty acids	Peptidoglycan (if present)	1-10 µm
Domain Archaea	Prokaryotic	Absent	Absent	Nonfatty acid lipids	Pseudopeptidoglycan or protein	1-10 µm
Domain Eukarya	Eukaryotic	Present	Present	Fatty acids	Usually cellulose or chitin (if present)	1-100 µm

Figure 1.6 The Three Domains of Life. Biologists distinguish domains Bacteria, Archaea, and Eukarya based on unique features of cell structure and biochemistry. The small evolutionary tree shows that archaea are the closest relatives of the eukaryotes.

Domain Bacteria

Bacteria are the most abundant and diverse organisms on Earth. Some species, such as *Streptococcus* and *Escherichia coli*, can cause illnesses, but others living on your skin and inside your intestinal tract are essential for good health. Bacteria are also very valuable in research, food and beverage processing, and pharmaceutical production. In ecosystems, bacteria play critical roles as decomposers and producers. Bacterial cells are structurally simple (figure 1.7). The nucleoid is the area where the cell's circular DNA molecule congregates. Unlike a eukaryotic cell's nucleus, the bacterial nucleoid is not bounded by a membrane. Located near the DNA in the cytoplasm are the enzymes, RNA molecules, and ribosomes needed to produce the cell's proteins.

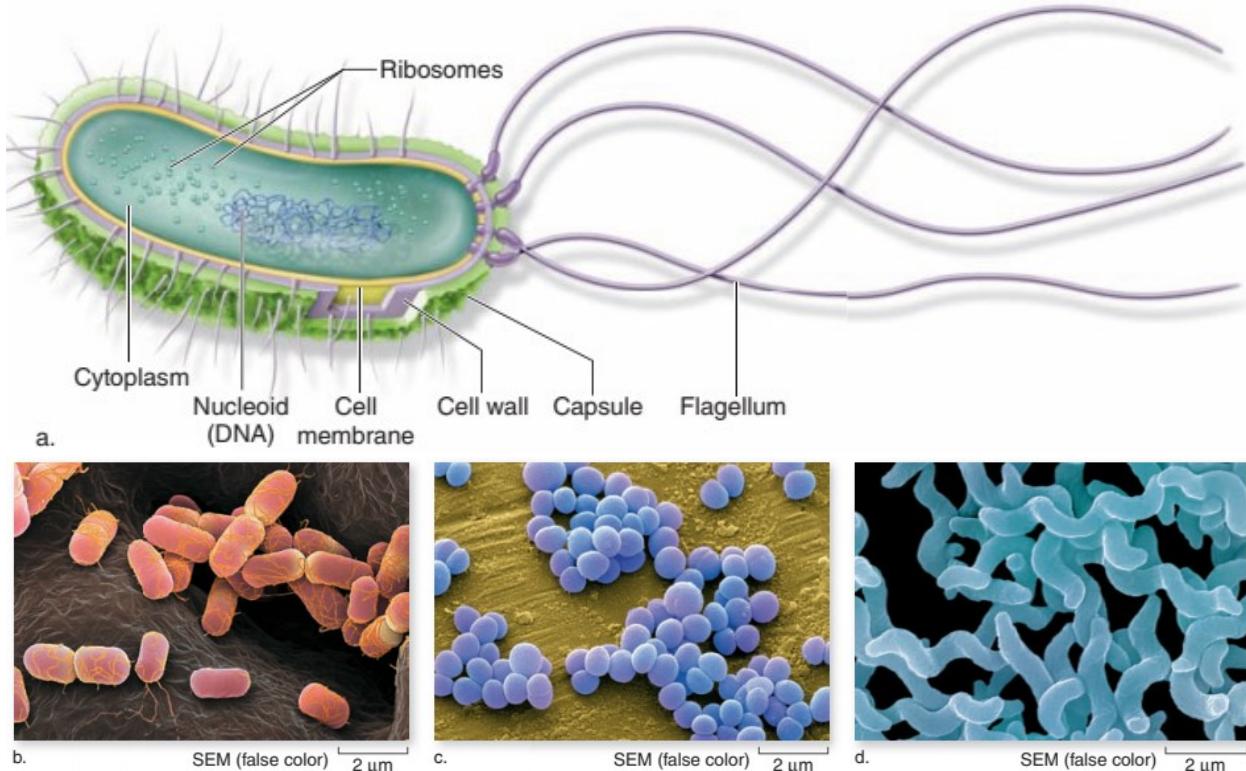


Figure 1.7 Anatomy of a Bacterium. (A) Bacterial cells lack internal compartments. (b) Rodshaped cells of *E. coli* inhabit human intestines. (c) Spherical *Staphylococcus aureus* cells cause infections that range from mild to deadly. (d) The corkscrew-shaped *Campylobacter jejuni* lives in the digestive tract of many animals.

A rigid cell wall surrounds the cell membrane of most bacteria, protecting the cell and preventing it from bursting if it absorbs too much water. This cell wall also gives the cell its shape: usually rod-shaped, round, or spiral. Many antibiotic drugs, including penicillin, halt bacterial infection by interfering with the microorganism's ability to construct its protective cell wall. In some bacteria, polysaccharides on the cell wall form a capsule that adds protection or enables the cell to attach to surfaces. Many bacteria can swim in fluids. Flagella (singular: flagellum) are tail-like appendages that enable these cells to move. One or more flagella are anchored in the cell wall and underlying

cell membrane. Bacterial flagella rotate like a propeller, moving the cell forward or backward.

Domain Archaea

Archaeal cells resemble bacterial cells in many ways. Like bacteria, they are smaller than most eukaryotic cells, and they lack a membrane-bounded nucleus and other organelles. Most have cell walls, and flagella are also common. Because of these similarities, Woese first named his newly recognized group Archaebacteria. The name later changed to Archaea when genetic sequences revealed that the resemblance to bacteria was only superficial. Archaea have their own domain because they build their cells out of biochemicals that are different from those in either bacteria or eukaryotes. Their phospholipids, cell walls, and flagella are all chemically unique. Their ribosomes, however, are more similar to those of eukaryotes than to those of bacteria. Archaea may therefore be the closest relatives of eukaryotes.

The first members of Archaea to be described were methanogens, microbes that use carbon dioxide and hydrogen from the environment to produce methane. Archaea subsequently became famous as “extremophiles” because scientists discovered many of them in habitats that are extremely hot, acidic, or salty. This characterization is somewhat misleading, however, because bacteria also occupy the same environments. Moreover, researchers have now discovered archaea in a variety of moderate habitats, including soil, swamps, rice paddies, oceans, and even the human mouth.

Domain Eukarya

An astonishing diversity of other organisms, including humans, belong to domain Eukarya. Our fellow animals are eukaryotes, as are yeasts, mushrooms, and other fungi. Plants are also eukaryotes, and so are one-celled protists such as *Amoeba* and *Paramecium*. Despite their great differences in external appearance, all eukaryotic organisms share many features on a cellular level. Figures 1.8 and 1.9 depict generalized animal and plant cells.

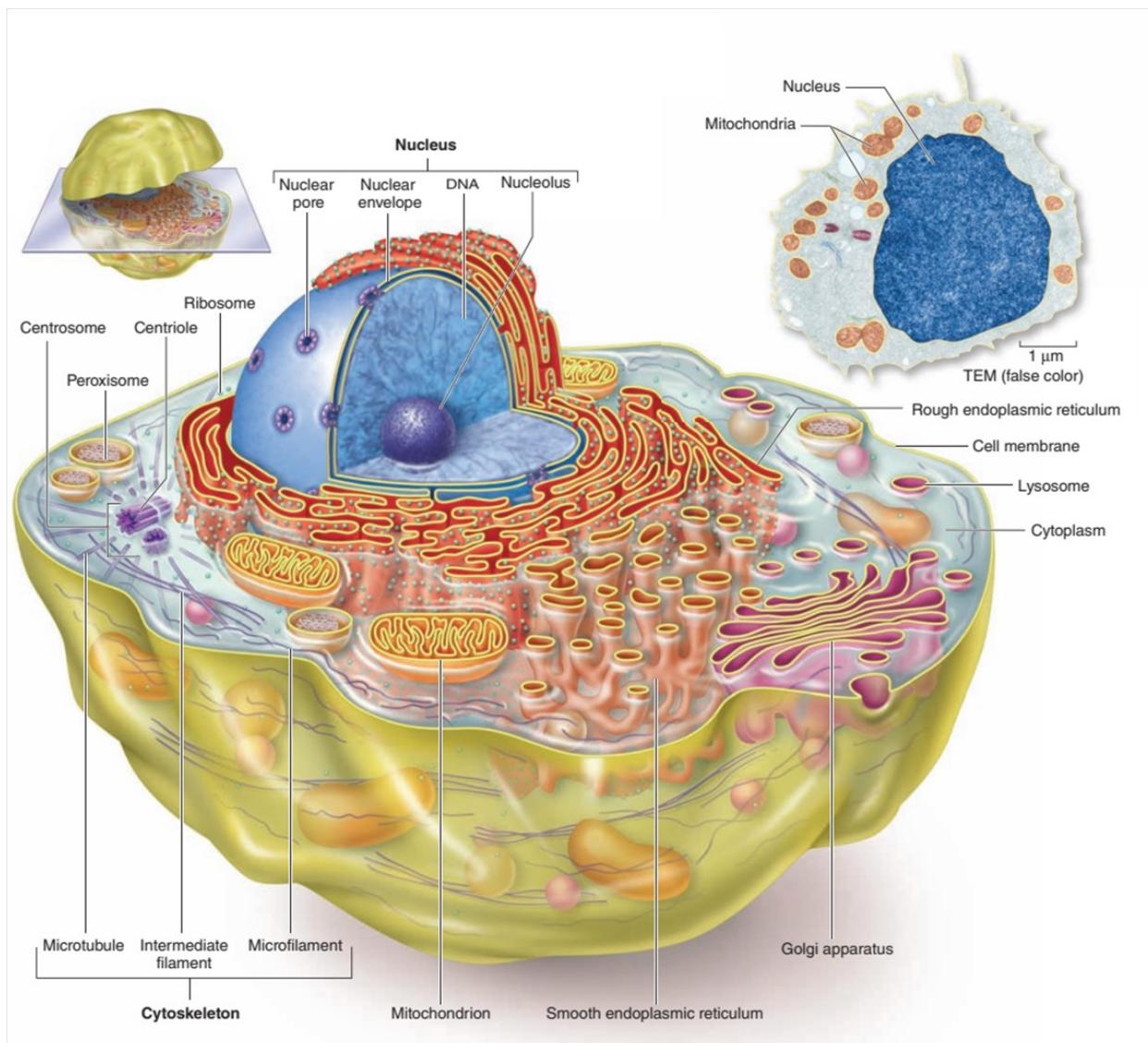


Figure 1.8 An Animal Cell. The large, generalized view shows the relative sizes and locations of a typical animal cell's components. The electron micrograph at the right shows a human white blood cell with a prominent nucleus and many mitochondria.

Although both of the illustrated cells have many structures in common, there are some differences. Most notably, plant cells have chloroplasts and a cell wall, which animal cells lack. One obvious feature that sets eukaryotic cells apart is their large size, typically 10 to 100 times greater than prokaryotic cells. The other main difference is that the cytoplasm of a eukaryotic cell is divided into organelles ("little organs"), compartments that carry out specialized functions. An elaborate system of internal membranes creates these compartments. In general, organelles keep related biochemicals and structures close enough to make them function efficiently. At the same time, they keep potentially harmful substances away from other cell contents. Compartmentalization also saves energy because the cell maintains high

concentrations of each biochemical only in certain organelles, not throughout the entire cell.

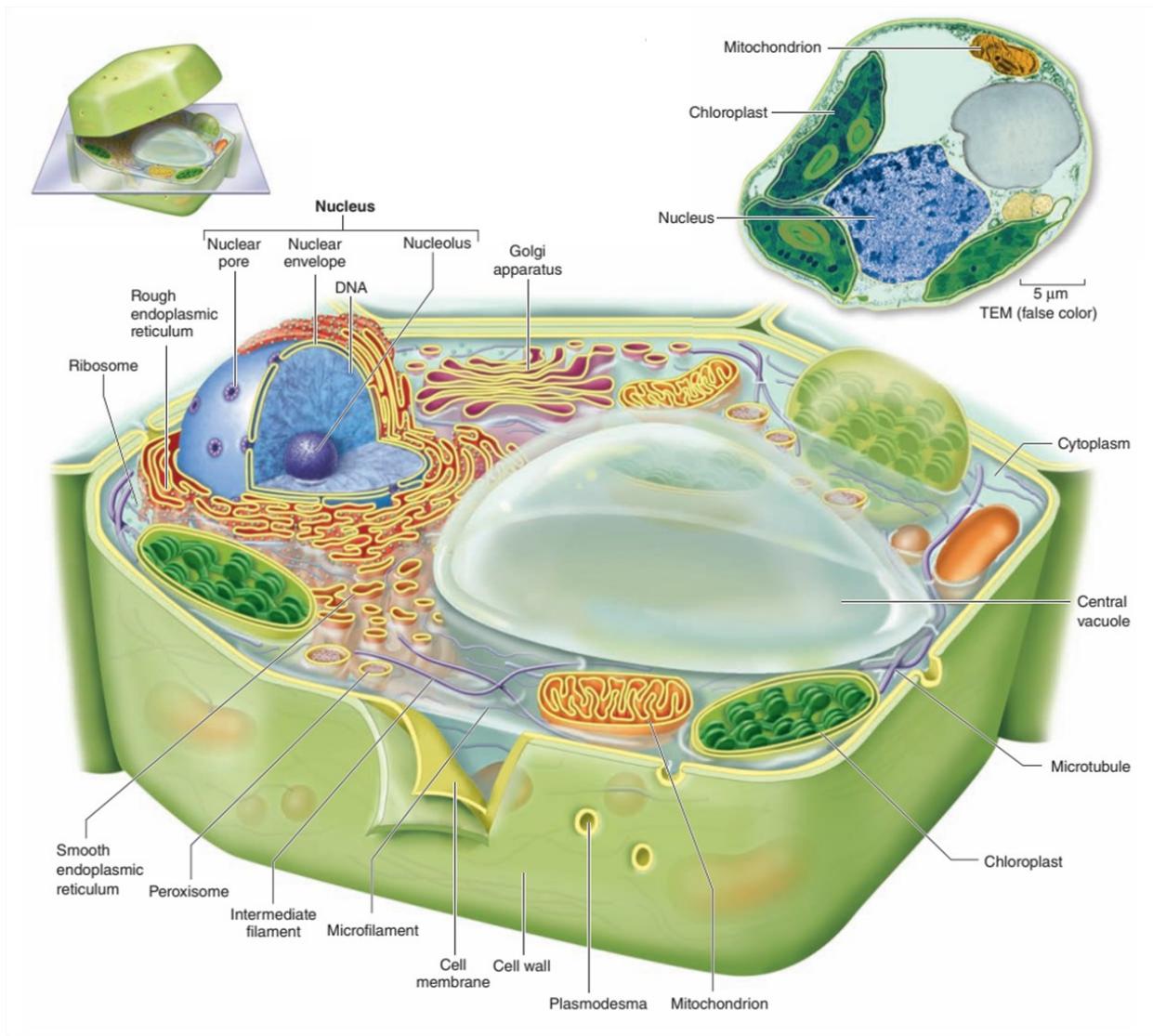


Figure 1.9 A Plant Cell. The large, generalized view illustrates key features of a typical plant cell. The electron micrograph at right shows a leaf cell; note the prominent nucleus, vacuole, chloroplasts, and cell wall.

2. Molecular biology

2.1 genes and genome

2.1.1 DNA and RNA

DNA is like a book, an encyclopedia that carries instructions for building a new individual. You already know the alphabet used to write the book: the four letters A, T, G, and C, for the four nucleotide bases adenine, thymine, guanine, and cytosine. A strand of DNA is a chain of those four kinds of nucleotides. The information it carries consists of which nucleotide follows the next along the strand—the DNA sequence.

Part of that information occurs in subsets called genes. A cell uses the sequence of a gene to build an RNA or protein product. Converting a gene's DNA sequence into its product starts with transcription: a process in which enzymes use the DNA sequence of a gene as a template to assemble a strand of RNA. Transcription makes an RNA copy of a gene. In other words, it transcribes the base sequence of a gene into a similar form: the base sequence of an RNA.

RNA is similar to a single strand of DNA in that both are chains of four kinds of nucleotides. However, the nucleotides that make up RNA differ slightly from those that make up DNA. The sugar in an RNA nucleotide is ribose, and that of a DNA nucleotide is a deoxyribose. Three of the bases (adenine, cytosine, and guanine) are the same in DNA and RNA nucleotides. The fourth base in RNA is uracil, not thymine.



Each strand of DNA is a polymer of nucleotides that have been linked into a chain. Even though a chain can be hundreds of millions of nucleotides long, only four kinds of nucleotides compose DNA. A DNA nucleotide has a five-carbon sugar, three phosphate groups, and one of four nitrogen-containing bases.

In 1950, Erwin Chargaff, one of many researchers who had been trying to solve the structure of DNA, made two discoveries. First, the amounts of thymine and adenine in all DNA are the same, as are the amounts of cytosine and guanine. We call this discovery Chargaff's first rule:

$$A = T \text{ and } G = C$$

Chargaff's second discovery, or rule, is that the proportion of adenine and guanine differs among DNA of different species. Watson and Crick proposed that DNA's structure consists of two chains (or strands) of nucleotides, running in opposite directions and coiled into a double helix. Bonds between the sugar of one nucleotide and the phosphate of the next form the backbone of each chain. Hydrogen bonds between the

internally positioned bases hold the two strands together. Only two kinds of base pairings form: A to T, and G to C, which explains the first of Chargaff's rules.

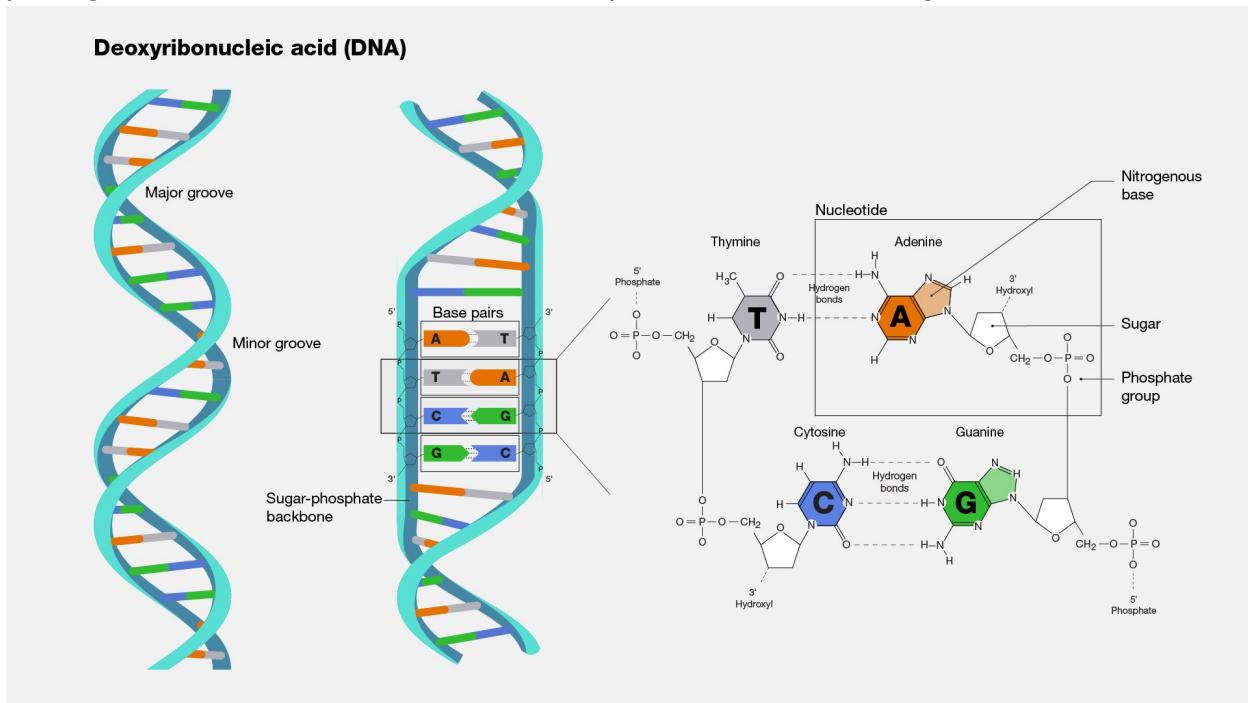
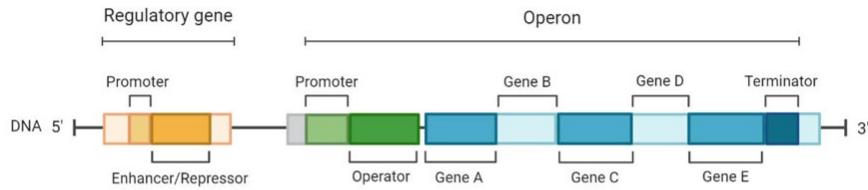


Fig. Double helix model of DNA

2.1.2 Gene

Gene structure is the organization of specialized sequence elements within a gene. Genes contain most of the information necessary for living cells to survive and reproduce. In most organisms, genes are made of DNA, where the particular DNA sequence determines the function of the gene. A gene is transcribed (copied) from DNA into RNA, which can either be non-coding (ncRNA) with a direct function, or an intermediate messenger (mRNA) that is then translated into protein. Each of these steps is controlled by specific sequence elements, or regions, within the gene. Every gene, therefore, requires multiple sequence elements to be functional.

Prokaryotic Gene Structure



Eukaryotic Gene Structure

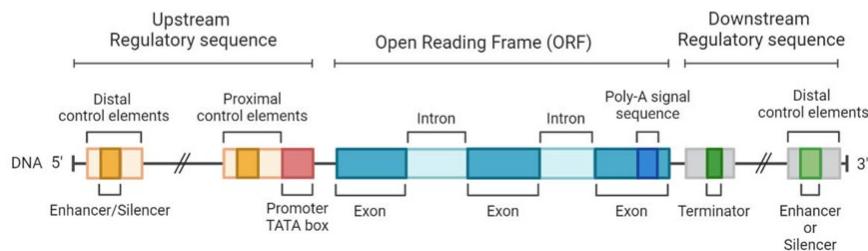


Fig. Structure of gene

2.1.3 Genome

A genome is the complete set of genetic instructions for an organism. It is a long molecule of DNA (deoxyribonucleic acid) that contains all the information needed to build and maintain an organism. The genome is the blueprint for an organism's traits, such as eye color, height, and susceptibility to certain diseases. The DNA molecule is made up of four chemical building blocks, called nucleotides, that are arranged in a specific sequence. This sequence determines the genetic code that is responsible for the characteristics of the organism. The genome is divided into smaller segments called genes, which are the functional units of DNA that code for specific proteins.

Different organisms have different genome sizes, ranging from small genomes of bacteria to complex genomes of humans. The human genome consists of approximately 3 billion nucleotide base pairs, which are organized into 23 pairs of chromosomes. Genomes can be studied using a variety of techniques, including DNA sequencing and genetic engineering. By understanding the genome of an organism, scientists can better understand its biology and develop new treatments for genetic diseases.

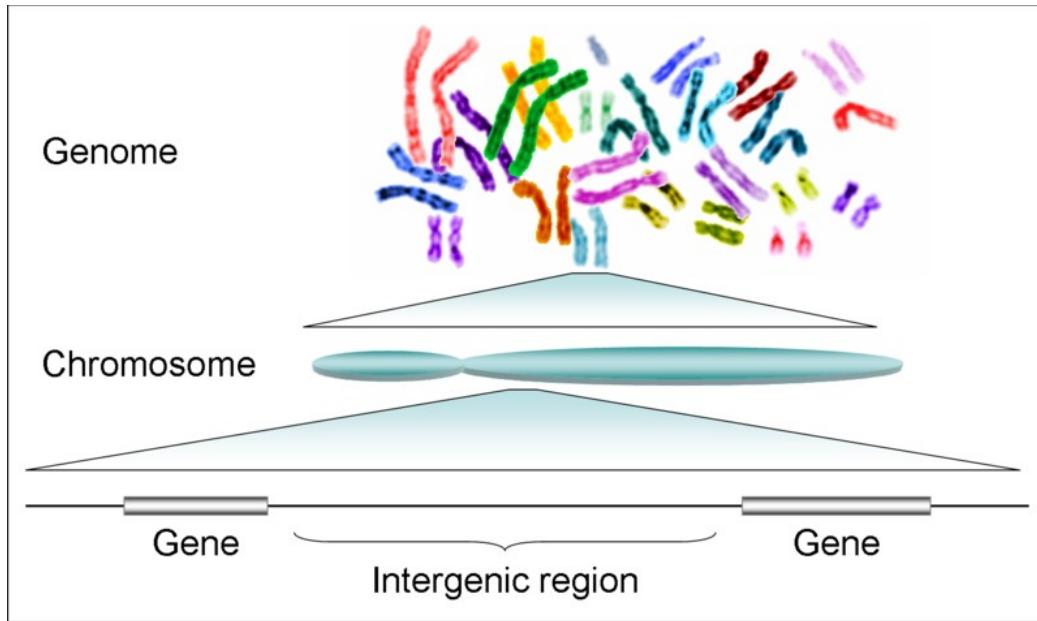


Fig. Human genome

2.2 Central Dogma

Despite the seemingly small differences in structure, DNA and RNA have very different functions. DNA's important but only role is to store a cell's heritable information. By contrast, cells make several kinds of RNAs, each of which has a different function. Although all RNAs are encoded by DNA, messenger RNA (mRNA) is the only kind of RNA that carries a protein-building message. That message is encoded within the sequence of the mRNA itself by sets of three bases, "genetic words" that follow one another along the length of the mRNA.

Like the words of a sentence, a series of genetic words can form a meaningful parcel of information—in this case, the sequence of amino acids of a protein. By the process of translation, the protein-building information in an mRNA is translated into a different language: a sequence of amino acids. The result is a polypeptide that twists and folds into a protein. The processes of transcription and translation are part of gene expression, a multistep process by which information encoded in a gene becomes converted to an RNA or protein product. During gene expression, genetic information flows from DNA to RNA to protein:



Fig. Central dogma of life

2.3 Transcription, translation, and genetic code

Gene expression is the process by which a gene product (an RNA or a polypeptide) is made. Two steps, called transcription and translation, are required to make a polypeptide from the instructions in a DNA gene. In the transcription step, an enzyme called RNA polymerase makes a copy of one of the DNA strands; this copy is not DNA, but its close cousin RNA. In the translation step, this RNA (messenger RNA, or mRNA) carries the genetic instructions to the cell's protein factories, called ribosomes. The ribosomes "read" the genetic code in the mRNA and put together a protein according to its instructions.

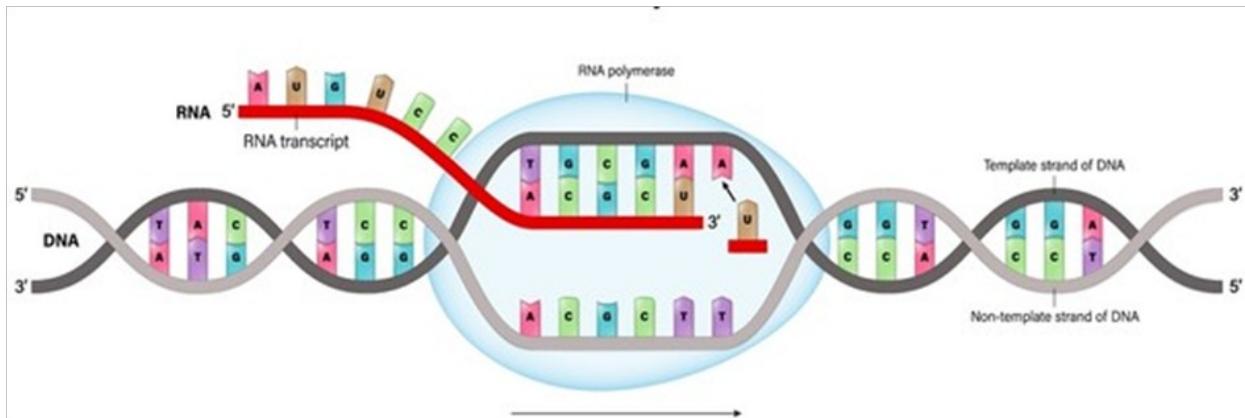


Fig. Transcription

Actually, the ribosomes already contain molecules of RNA, called ribosomal RNA (rRNA). Francis Crick originally thought that this RNA residing in the ribosomes carried the message from the gene. According to this theory, each ribosome would be capable of making only one kind of protein—the one encoded in its rRNA. The ribosomes are nonspecific translation machines that can make an unlimited number of different proteins, according to the instructions in the mRNAs that visit the ribosomes.

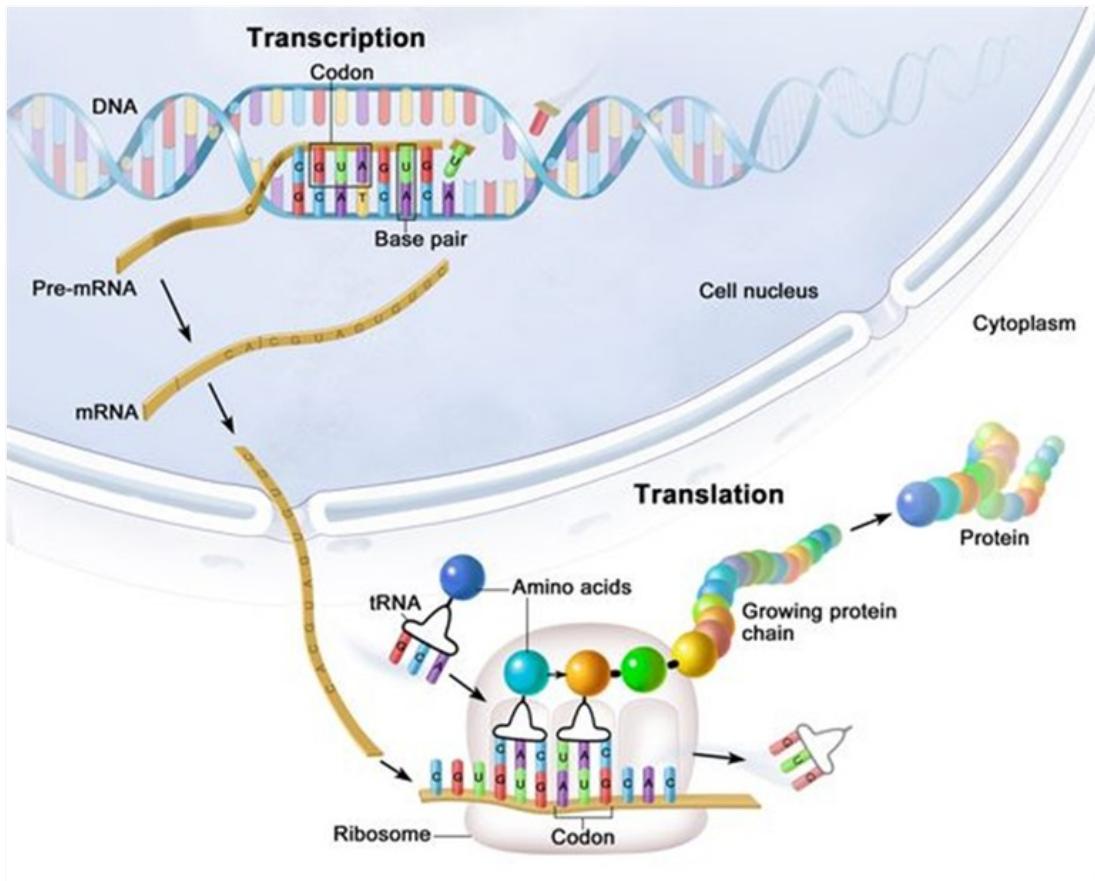


Fig. Translation

Marshall Nirenberg and Gobind Khorana, working independently with different approaches, cracked the genetic code in the early 1960s. They found that 3 bases constitute a code word, called a codon, that stands for one amino acid. Out of the 64 possible 3-base codons, 61 specify amino acids; the other three are stop signals. The ribosomes scan a messenger RNA 3 bases at a time and bring in the corresponding amino acids to link to the growing protein chain. When they reach a stop signal, they release the completed protein.

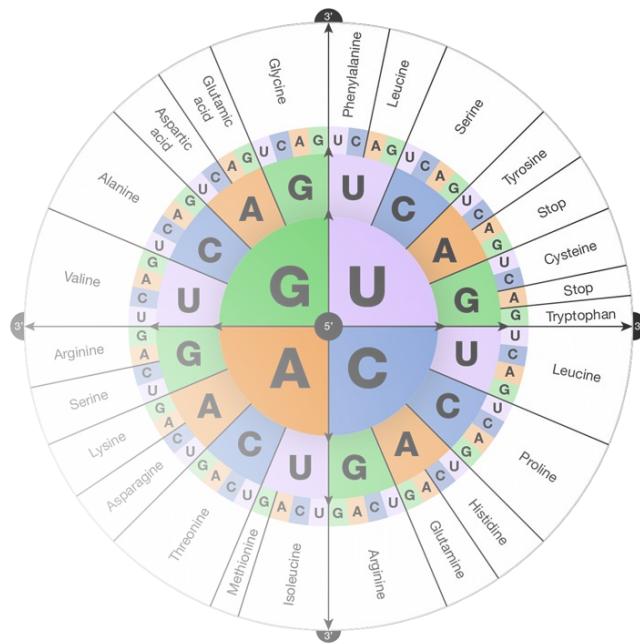


Fig. Genetic Code

3. Introduction to bioinformatics

Bioinformatics is the collection, classification, storage, and analysis of biochemical and biological information using computers especially as applied to molecular genetics and genomics[or] Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data, in particular when the data sets are large and complex. As a multidisciplinary field of science, bioinformatics combines biology, chemistry, physics, computer science, information engineering, mathematics, and statistics to analyze and interpret biological data.

Bioinformatics has been used for in silico analyses of biological queries using mathematical and statistical techniques. Bioinformatics includes biological studies that use computer programming as part of their methodology, as well as specific analysis "pipelines" that are repeatedly used, particularly in the field of genomics. Common uses of bioinformatics include the identification of candidate genes and single nucleotide polymorphisms (SNPs). Often, such identification is made with the aim of better understanding the genetic basis of disease, unique adaptations, desirable properties (esp. In agricultural species), or differences between populations. In a less formal way, bioinformatics also tries to understand the organizational principles within nucleic acid and protein sequences, called proteomics.

Image and signal processing allow the extraction of useful results from large amounts of raw data. In the field of genetics, it aids in sequencing and annotating genomes and their observed mutations. It plays a role in the text mining of biological literature and developing biological and gene ontologies to organize and query biological data. It also plays a role in the analysis of gene and protein expression and regulation. Bioinformatics tools aid in comparing, analyzing, and interpreting genetic and genomic data and more generally in the understanding of evolutionary aspects of molecular biology. At a more integrative level, it helps analyze and catalog the biological pathways and networks that are an essential part of systems biology. In structural biology, it aids in the simulation and modeling of DNA, RNA, proteins as well as biomolecular.

3.1 Primary and Secondary databases

In bioinformatics, and indeed in other data-intensive research fields, databases are often categorized as primary or secondary. Primary databases are populated with experimentally derived data such as nucleotide sequence, protein sequence, or macromolecular structure. Experimental results are submitted directly into the database by researchers, and the data are essentially archival in nature. Once given a database accession number, the data in primary databases are never changed: they form part of the scientific record.

By contrast, secondary databases comprise data derived from the results of analyzing primary data. They are often referred to as curated databases but this is a bit of a misnomer because primary databases are also curated to ensure that the data in them is consistent and accurate. Secondary databases often draw upon information from numerous sources, including other databases (primary and secondary), controlled vocabularies (see later section) and the scientific

literature. They are highly curated, often using a complex combination of computational algorithms and manual analysis and interpretation to derive new knowledge from the public record of science.

Secondary databases have become the molecular biologist's reference library over the past decade or so, providing a wealth of (often daunting) information on just about any gene or gene product that has been investigated by the research community. The potential for mining this information to make new discoveries is vast.

Many data resources have both primary and secondary characteristics. For example, UniProt accepts primary sequences derived from peptide sequencing experiments. However, UniProt also infers peptide sequences from genomic information, and it provides a wealth of additional information, some derived from automated annotation (TrEMBL), and even more from careful manual analysis (SwissProt).

3.2 NCBI

NCBI means or stands for National Center for Biotechnology Information. As a national resource for molecular biology information, NCBI's mission is to develop new information technologies to aid in the understanding of fundamental molecular and genetic processes that control health and disease.

The National Center for Biotechnology Information (NCBI) provides a large suite of online resources or biological information and data, including the GenBank nucleic acid sequence database and the PubMed database of citations and abstracts for published life science journals. The National Center for Biotechnology Information (NCBI) provides a large suite of online resources for biological information and data, including the GenBank nucleic acid sequence database and the PubMed database of citations and abstracts published in life science journals. The NCBI houses a series of databases relevant to biotechnology and biomedicine and is an important resource for bioinformatics tools and services. Major databases include GenBank for DNA sequences and PubMed, a bibliographic database for biomedical literature. Other databases include the NCBI Epigenomics database. All these databases are available online through the Entrez search engine. NCBI was directed by David Lipman, one of the original authors of the BLAST sequence alignment program and a widely respected figure in bioinformatics.

Entrez is an integrated database retrieval system that provides access to a diverse set of 37 databases that together contain 690 million records. Primary databases contain raw data as archival repositories such as the NCBI Sequence Read Archive (SRA), whereas secondary or derivative databases contain curated information as added value. NCBI's Genome resources

include information on large-scale genomics projects, genome sequences, and assemblies, and mapped annotations, such as variations, markers, and data from epigenomics studies. NCBI's Gene resources include collections of curated nucleotide sequences used as references, sequence clusters to predict and study homologs, and various databases and tools for the study of gene expression. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

3.3 UniProt

UniProtKB/Swiss-Prot is a manually annotated, non-redundant protein sequence database. It combines information extracted from scientific literature and biocurator-evaluated computational analysis. The aim of UniProtKB/Swiss-Prot is to provide all known relevant information about a particular protein. The UniProt Knowledgebase (UniProtKB) is an expertly curated database, a central access point for integrated protein information with cross-references to multiple sources. The UniProt Archive (UniParc) is a comprehensive sequence repository, that reflects the history of all protein sequences. More than 95% of the protein sequences provided by UniProtKB come from the translations of coding sequences (CDS) submitted to the ENA/GenBank/DDBJ nucleotide sequence resources of the International Nucleotide

3.4 RCSB

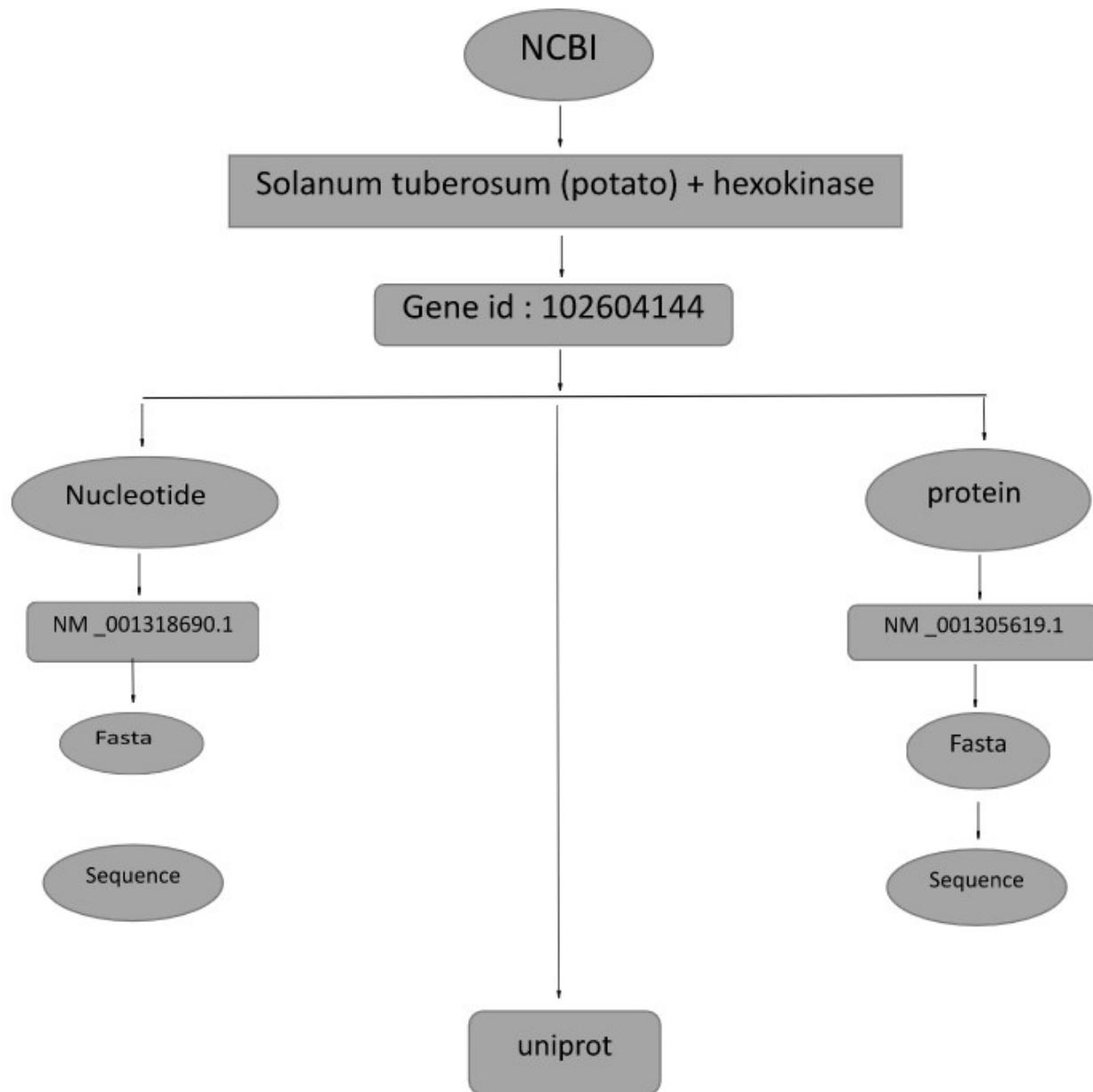
RCSB stands for Research Collaboratory for Structural Bioinformatics. The Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) supports scientific research and education worldwide by providing access to annotated information about three-dimensional (3D) structures of macromolecules (e.g., nucleic acids, proteins), and associated small molecules (e.g., drugs). The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond. RCSB PDB (Research Collaboratory for Structural Bioinformatics PDB) operates the US data center for the global PDB archive and makes PDB data available at no charge to all data consumers without limitations on usage (Policies).

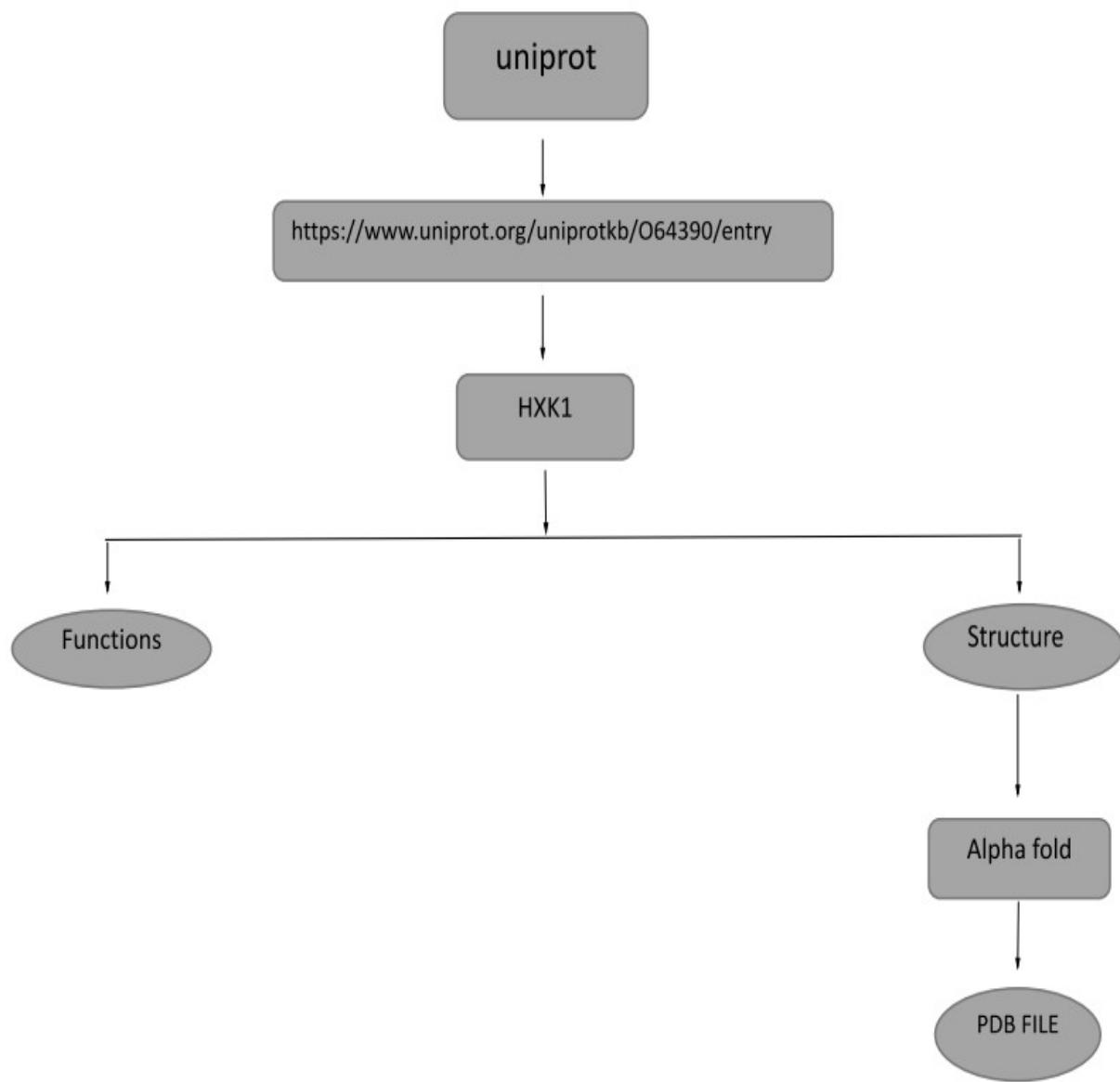
3.5 RasMol

RasMol is a molecular graphics program intended for the visualization of proteins, nucleic acids, and small molecules. The program is aimed at displaying, teaching, and generating publication-quality images. RasMol runs on a wide range of architectures and operating systems including Microsoft Windows, Apple Macintosh, UNIX, and VMS systems.

The program reads in a molecule coordinate file and interactively displays the molecule on the screen in a variety of colour schemes and molecule representations. Currently available representations include depth-cued wireframes, 'Dreiding' sticks, space-filling (CPK) spheres, ball and stick, solid and strand biomolecular ribbons, atom labels, and dot surfaces.

4 .Methodology





5.1 HXK1

Hexokinase-1

Hexokinase-1 (HK1) is an enzyme that helps to break down glucose in the body. It is encoded by the HK1 gene on chromosome 10 and is one of four highly homologous hexokinase isoforms in mammalian cells . HK1 catalyzes the rate-limiting and first obligatory step of glucose metabolism, which is the ATP-dependent phosphorylation of glucose to produce glucose-6-phosphate (G6P), the first step in most glucose metabolism pathways

In addition to HK1, there are other hexokinases that play important roles in metabolism. For example, glucokinase is found in the liver and requires a much greater concentration of glucose before it reacts. There is also a novel hexokinase called Hexokinase domain-containing protein-1 (HKDC1) that has been discovered recently. It functions similarly to known hexokinases like HK1 and helps with whole-body glucose use

HK1 is a ubiquitous form of hexokinase that localizes to the outer mitochondrial membrane (OMM). It is one of two mitochondrial isoforms of hexokinase and a member of the sugar kinase family. HK1 has a high affinity for glucose, meaning only small amounts of glucose are necessary for enzymatic activity.

Functions

Fructose and glucose phosphorylating enzyme. May be involved in the phosphorylation of glucose during the export from plastids to cytosol. Seems neither to be involved in cell sugar sensing nor in carbohydrate metabolism in tuber.



Fig .*Solanum tuberosum* (potato)

5.3 Gene id file

[Log in](#)

LOC102604144 hexokinase-1 [Solanum tuberosum (potato)]

[Download Datasets](#)

Gene ID: 102604144, updated on 26-May-2022

Summary

Gene symbol	LOC102604144
Gene description	hexokinase-1
Gene type	protein coding
RefSeq status	PROVISIONAL
Organism	Solanum tuberosum
Lineage	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae; Pentapetalae; asterids; lamiids; Solanales; Solanaceae; Solanoideae; Solaneae; Solanum
Also known as	hxx; HXX1; StHK1; hexokinase
NEW	Try the new Gene table Try the new Transcript table

Genomic context

Location: chromosome: Un

[See LOC102604144 in Genome Data Viewer](#)

Exon count: 9

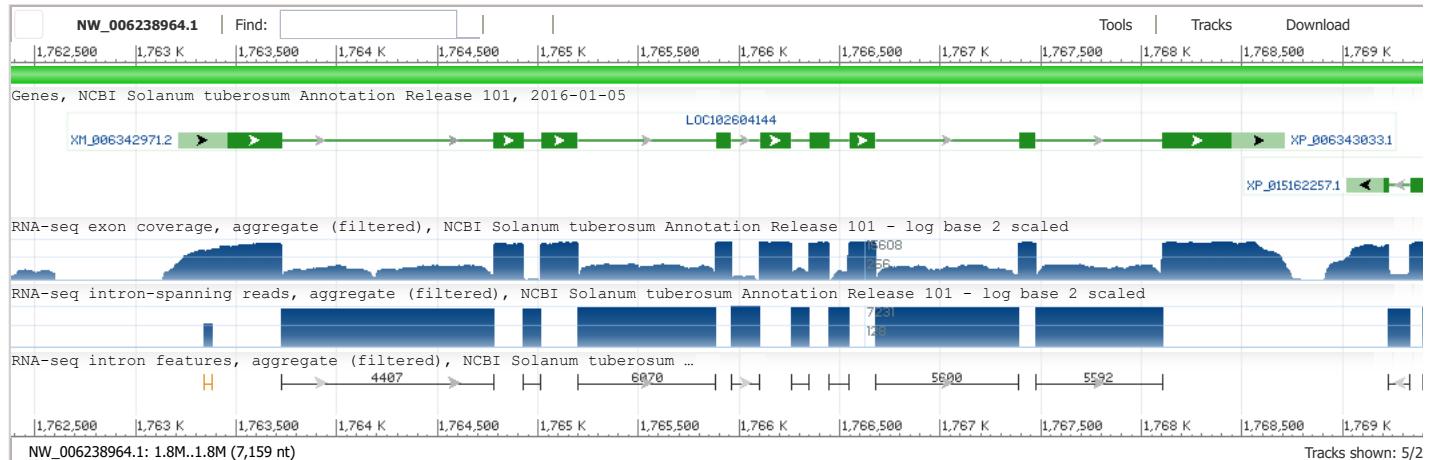
Annotation release	Status	Assembly	Chr	Location
101	current	SolTub_3.0 (GCF_000226075.1)	Unplaced Scaffold	NW_006238964.1 (1763210..1768716)



Genomic regions, transcripts, and products

[Go to reference sequence details](#)

Genomic Sequence: NW_006238964.1 Unplaced Scaffold Reference SolTub_3.0 Primary Assembly

[Go to nucleotide:](#) [Graphics](#) [FASTA](#) [GenBank](#)

Bibliography

[GeneRIFs: Gene References Into Functions](#)[What's a GeneRIF?](#)[Submit:](#) [New GeneRIF](#) [Correction](#)

General gene information

[Homology](#)[Orthologs](#)

General protein information

Preferred Names
hexokinase-1

XP_006343033.1

EC 2.7.1.1

NCBI Reference Sequences (RefSeq)NEW Try the new [Transcript table](#)**RefSeqs maintained independently of Annotated Genomes**These reference sequences exist independently of genome builds. [Explain](#)**mRNA and Protein(s)**

1. [NM_001318690.1](#) → [NP_001305619.1](#) hexokinase-1

Status: PROVISIONAL

Source sequence(s)	X94302
UniProtKB/Swiss-Prot	O64390
Conserved Domains (3) summary	
	PLN02405 Location:1 → 497
	pfam00349 Location:40 → 240
	pfam03727 Location:246 → 488
PLN02405; hexokinase	
Hexokinase_1; Hexokinase	
Hexokinase_2; Hexokinase	

RefSeqs of Annotated Genomes: Solanum tuberosum Annotation Release 101 [details...](#)The following sections contain reference sequences that belong to a specific genome build. [Explain](#)**Reference SolTub_3.0 Primary Assembly****Genomic**

1. [NW_006238964.1](#) Reference SolTub_3.0 Primary Assembly

Range	1763210..1768716
Download	GenBank , FASTA , Sequence Viewer (Graphics)

mRNA and Protein(s)

1. [XM_006342971.2](#) → [XP_006343033.1](#) hexokinase-1

Conserved Domains (1) summary	
	PLN02405 Location:1 → 496

Related sequences

Nucleotide	Accession and Version	Protein
Heading		
genomic	AEWC01007901.1 (12047..17553)	None
genomic	HB877598.1	CBF63127.1
genomic	HC935007.1	CBU88003.1
genomic	JA213451.1	CCA64912.1
genomic	JA214275.1	CCA65274.1
genomic	JA320204.1	CCA78534.1
genomic	JA321028.1	CCA78896.1
genomic	JB233357.1	CDH48157.1
genomic	JB768852.1	CDI44493.1
mRNA	X94302.1	CAA63966.1

Protein Accession	Links	
	GenPept Link	UniProtKB Link
O64390.1	GenPept	UniProtKB/Swiss-Prot:O64390

Additional links[Gene LinkOut](#)

5.4a nucleotide id file

[Log in](#)Nucleotide [GenBank](#)

Solanum tuberosum hexokinase-1 (LOC102604144), mRNA

NCBI Reference Sequence: NM_001318690.1

[FASTA](#) [Graphics](#)

LOCUS NM_001318690 1619 bp mRNA linear PLN 12-MAR-2022
 DEFINITION Solanum tuberosum hexokinase-1 (LOC102604144), mRNA.
 ACCESSION NM_001318690 XM_006342971
 VERSION NM_001318690.1
 KEYWORDS RefSeq.
 SOURCE Solanum tuberosum (potato)
 ORGANISM [Solanum tuberosum](#)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae;
 Pentapetalae; asterids; lamiids; Solanales; Solanaceae;
 Solanoideae; Solaneae; Solanum.
 COMMENT PROVISIONAL [REFSEQ](#): This record has not yet been subject to final
 NCBI review. The reference sequence was derived from [X94302.1](#).

On Jan 8, 2016 this sequence version replaced [XM_006342971.2](#).

```
##Evidence-Data-START##
Transcript exon combination :: X94302.1 [ECO:0000332]
RNAseq introns :: single sample supports all introns
                  SAMEA2149986, SAMEA2151923
                  [ECO:0000348]
```

```
##Evidence-Data-END##
```

FEATURES	Location/Qualifiers
source	1..1619 /organism="Solanum tuberosum" /mol_type="mRNA" /db_xref="taxon: 4113 "
gene	1..1619 /gene="LOC102604144" /gene_synonym="hexokinase; hxk; HXK1; StHK1" /note="hexokinase-1" /db_xref="GeneID: 102604144 "
misc_feature	1..3 /gene="LOC102604144" /gene_synonym="hexokinase; hxk; HXK1; StHK1" /note="upstream in-frame stop codon"
CDS	19..1515 /gene="LOC102604144" /gene_synonym="hexokinase; hxk; HXK1; StHK1" /EC_number=" 2.7.1.1 " /codon_start=1 /product="hexokinase-1" /protein_id=" NP_001305619.1 " /db_xref="GeneID: 102604144 " /translation="MKKVTVGAAVGAAACAVAALIVNHRMRKSSKWGRAMAILREF EEKCKTQDAKLKQVADAMTVEMHAGLASEGGQSSRCLSPMSIISQLVMKLGVFYALDL GGTNFRVLRVQLGGKGIIHQEFQAEASIPPSLMVGTSDALFDYIAELAKFVAEEEE KFHQPGKQRELGFHLLIPSNAFDNNSGTIMRWTKGSIDDAVGQDVVGELETKAMKEK VLDMRVSLVNNDTVGTLAGGKYTQKDVAVALGTGTNAAYVERQAIKPWHGPVPKS GEMVINMEWGNFRSSHPLTEYDHALDNESLNPAEQIFEKMTSGMYLGEILRRVLTRV AEEVLAFLAMRSLSQSLKDSFVLRTPDMSAMHHDTSPDLKVGEKLKDILEISNTSLKT RKLVLSLCNIVATRGARLDAAGVLGLKKGMDTPKQGGSERTVIAMDGGLYEHYTEY RMCLENSLKDLLGEELATSIVFVHSNDGSGIGAALLRASHSMYLEDQA"
misc_feature	28..90 /gene="LOC102604144" /gene_synonym="hexokinase; hxk; HXK1; StHK1" /note="propagated from UniProtKB/Swiss-Prot (064390.1); transmembrane region"
misc_feature	283..702 /gene="LOC102604144" /gene_synonym="hexokinase; hxk; HXK1; StHK1" /note="propagated from UniProtKB/Swiss-Prot (064390.1); Region: Hexokinase small subdomain."

```
/evidence=ECO:0000255|PROSITE-ProRule:PRU01084"
529..609
/gene="LOC102604144"
/gene_synonym="hexokinase; hкс; HXK1; StHK1"
/note="propagated from UniProtKB/Swiss-Prot (064390.1);
Region: Glucose-binding. /evidence=ECO:0000255"
703..1449
/gene="LOC102604144"
/gene_synonym="hexokinase; hкс; HXK1; StHK1"
/note="propagated from UniProtKB/Swiss-Prot (064390.1);
Region: Hexokinase large subdomain.
/evidence=ECO:0000255|PROSITE-ProRule:PRU01084"
```

ORIGIN

1 taaattccgg caaaaaaaaaat gaagaaaagtgc acgggtggag ccgcccgtgt tggtcagct
61 gcgggtgtgc cagtggcgc gttaatagtg aaccacccga tgccggaaatc tagtaaatgg
121 ggctcgctca tggctattct tcgtgaattt gaggaaaaagt gtaagactca agatgc当地
181 cttaaagcaag ttgcgtatgc tatgactgtt gagatgcacg ccggacttgtc ttctgaaggc
241 ggccagaatc caagatgtt atccatcatg tcgataatct cccaaatctt gtatggactgt
301 ggctgtttt atgcatttggaa tcttggatgaa acaaatttc gagttatgc agtgc当地
361 gttggaaaatc atgggtgttat tattatcaa gaatttgc aggcatcaat tcctcaagg
421 ttatgttgc ggactttcaga tgcaatcccc gattatattt ccgtcgatct tgctaaat
481 gttgtcgcc aagaggaaaa atttcatcaa cctctggta agcagagaga actaggat
541 caccttcata ttcccgatgaa tgccatctt aataacttg ggacttattt gcgggtggaca
601 aaaggcttct caatttgcatgaa tgccgtggc caagatgtt ttggagaact cacaaggct
661 atgaaaagaaa aggtgttcga tatgcgggtc tcagcttgc tgaatgtatc ttgtggaaaca
721 ttatgttgc gtaatatac gcaaaaggat gtactgttgc ctgttatctt aggtacagg
781 acgaatcgat cttatgttgc gccgggtgcag gcaatcttca agtgcgtatgg tcctgtggca
841 aaatctgttca aatgtttat caatatggaa tggggtaattt tttaggtc catctcgcc
901 tttagacagat atgtatcgc ttgttgcac gagatgttac atccctgcgaa agatatt
961 gagaagatcat cttctggcat gtactgttgc gaaatttac cgagatctt cactagg
1021 gctgaagaatg tcctggcgat tttggcgatg aggtccctcc aaagcttca ggattcatt
1081 gtgttaaggg cacctgatcat gtctgtatgc catcatgaca catcccttgc tctggaaatgt
1141 gtgggtggaaa agctgttgc gaaatgttgc tttttatcaat cctccgttgc gacaagg
1201 ttatgttgc gcttgcac tatctgttgc acacgtgggg caagacttgc cgctgcagg
1261 gtatggggc tttttttttt gatggggaaat gatggggaaat gatacgccatc agccggatgg
1321 acggttatgc ccatgttgc cgggtgtat gaggacttgc cagaatatgc aatgttgc
1381 gagaacttgc tgaaggactt gctcgagag gaatttgc gcaatgttgc ttttgc
1441 tcattatgtat gttctggcat ttgttgc tcttcgttgc cttcttcattt aatgttgc
1501 gaatgttgc ctttggcgat ttgttgc gttatgttgc ttttgc ttttgc
1561 gactctatcat tttgtcgatc acggatgttgc gttatgttgc ttttgc ttttgc

11

5.4b nucleotide id fasta file

[Log in](#)Nucleotide [FASTA](#)

Solanum tuberosum hexokinase-1 (LOC102604144), mRNA

NCBI Reference Sequence: NM_001318690.1

[GenBank](#) [Graphics](#)

```
>NM_001318690.1 Solanum tuberosum hexokinase-1 (LOC102604144), mRNA
TAAATCCGCAAAAAAATGAAGAAAGTGACGGTGGGAGCGCCGTGGTTGGTGCAGCTCGGGTGTG
CAGTGGCGCGTAAATGAAACCGGGATCGGAAATCTAGTAATGGGTCGTCTATGGCTATTCT
TCGTGAATTGAGGAAAAGTAAAGACTCAAGATGCAAGCTTAAGCAAGTTGCTGTGCTATGACTGTT
GAGATGCACGCCGACTGCTCTGAAGGCGGCAAGAGCTCAAGATGCTTATCACCTATGCGATAATCT
CCCAACTGGTGTGAAGCTGGCGTCTTTATGCATTGGATCTTGGAAACAAATTTCAGGTATTGCG
AGTGAATTGGGTGAGGAAAGATGGGTATTATTCAAGAATTGCTGAGGCATCAATTCTCCAAGT
TTGATGGTGGGACTTCAGATGCACTTTTGATTATATTGGCTGAGCTTGCTAAATTGTTGCTGCGG
AAGAGGAAAATTTCATCAACCTCTGGTAAGCAGAGAGAACACTAGGTTTACCTTCATTCCCAGTAA
TGCAGACTTCAATACTCTGGACTATTATGCGGTGGACAAAGGCTTCTCAATTGATGATGCGGTTGGC
CAAGATGTTGGAGAAACTCACAAAAGCTATGAAAGAAAAGGTGCTCGATATGCGGGTCTCAGCTCTGG
TGAATGATACTGTTGAACATTAGCTGGTAAATATACGCAAAGGATGTAAGCTGTTGCTGTTATCTT
AGGTACAGGGACGAATGCGCTTATGGAGCGGGTGCAGGCAATTCCAAAGTGGCATGGTCTGCCCCA
AAATCTGGTAAATGGTTATCAATATGGAATGGGTAAATTAGGTCACTGGCTGAGAAGTCTGGCATTGGCAT
GTACTGGGAGAAATTTCAGCAGAGTTCTCACTAGGGTGGCTGAAGAAGTCTGGCATTGGCATG
AGGCCCTCAAAGCCTCAAGGATTCAATTGTGTTAAGGACACCTGATATGCTGCTATGCATCATGACA
CATCCCCTGATGAAAGTGGTTGGAAAAGCTGAAGGATATTAGGATATCTAATACCTCTTGAA
GACAAGGAAAATTAGTGGTGAACGCTGCAATATCGTTGCAACACGGGGCAAGACTTGAACGCTGCAAGGG
GTATTGGGATCTGGTAAAGATGGGAAGAGATACGCTAAGCAGGGTGGTTCAGAAAGGACGGTTATAG
CCATGGATGGCGGGTTATGAGCACTACAGAATATAGAATGTTAGAGAAACTTTGAAGGACTT
GCTCGGAGAGGAATTGCAACGAGCATGTTTGCACTCAATGATGGTCTGGCATTGGCTGCT
CTTCTCGTGCCTCTCAATTCAATGACCTGAAAGATCAAGCTTAGGGCAGTGTGATCAAATCCATAACTT
GCTAGAGGGGACTACCTTAGACTCTATATTGCTCTCAAGCGTTTGGTATTCTCCTTATTG
TACCATCAA
```

5.5a protein id file

[Log in](#)Protein [GenPept](#)

hexokinase-1 [Solanum tuberosum]

NCBI Reference Sequence: NP_001305619.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)[Go to:](#)

LOCUS NP_001305619 498 aa linear PLN 12-MAR-2022
 DEFINITION hexokinase-1 [Solanum tuberosum].
 ACCESSION NP_001305619 XP_006343033
 VERSION NP_001305619.1
 DBSOURCE REFSEQ: accession [NM_001318690.1](#)
 KEYWORDS RefSeq.
 SOURCE Solanum tuberosum (potato)
 ORGANISM [Solanum tuberosum](#)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae;
 Pentapetalae; asterids; lamiids; Solanales; Solanaceae;
 Solanoideae; Solaneae; Solanum.
 COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final
 NCBI review. The reference sequence was derived from [X94302.1](#).

On Jan 8, 2016 this sequence version replaced [XP_006343033.1](#).

```
##Evidence-Data-START##
Transcript exon combination :: X94302.1 [ECO:0000332]
  RNAseq introns      :: single sample supports all introns
                        SAMEA2149986, SAMEA2151923
                        [ECO:0000348]
```

##Evidence-Data-END##

FEATURES Location/Qualifiers
 source 1..498
 /organism="Solanum tuberosum"
 /db_xref="taxon:[4113](#)"
 Protein 1..498
 /product="hexokinase-1"
 /EC_number="[2.7.1.1](#)"
 /calculated_mol_wt=54000
 Region 1..497
 /region_name="PLN02914"
 /note="hexokinase; cl42853"
 /db_xref="CDD:[456198](#)"
 Site 4..24
 /site_type="transmembrane region"
 /note="propagated from UniProtKB/Swiss-Prot (064390.1)"
 Region 89..228
 /region_name="Hexokinase small subdomain."
 /evidence=ECO:0000255|PROSITE-ProRule:PRU01084"
 /note="propagated from UniProtKB/Swiss-Prot (064390.1)"
 Region 171..197
 /region_name="Glucose-binding. /evidence=ECO:0000255"
 /note="propagated from UniProtKB/Swiss-Prot (064390.1)"
 Region 229..477
 /region_name="Hexokinase large subdomain."
 /evidence=ECO:0000255|PROSITE-ProRule:PRU01084"
 /note="propagated from UniProtKB/Swiss-Prot (064390.1)"
 CDS 1..498
 /gene="LOC102604144"
 /gene_synonym="hexokinase; hxk; HK1; StHK1"
 /coded_by="NM_001318690.1:19..1515"
 /db_xref="GeneID:[102604144](#)"

ORIGIN

```
1 mkkvtvgaav vgaaavcava alivnhrmrk sskwgramai lrefeekckt qdaklkqvad
 61 amtvemhagl aseggqssrc lspmisiql vmlkgvfyal dlggtnfrvl rvqlggkdg
121 iihqefaaeas ippslmvgts dalfdyiaae lakfvaeeee kfhpqppgkqr elgfhlips
181 nadffnnsgti mrwtkgfsid davgqdvvge ltkamkekvl dmrvsalnd tvgtlaggky
241 tqkdavavi lgtgtnaayv ervqaipkwh gpvpksgemv inmewgnfrs shplteydh
301 alndeslnpa eqifekmtsg mylgeilrrv ltrvaeevla flamrslqsl kdsfvrlrtpd
```

361 msamhhdtsp dlkvvgeklk dileisntsl ktrklvslc nivatrgarl daagvlgil
421 kmgrdtpkqg gsertviamd gglyehytey rmclenslkd llgeelatsi vfvhsndgsg
481 igaallrash smyledqa

//

5.5b protein id fasta file

[Log in](#)

Protein

[FASTA](#)

hexokinase-1 [Solanum tuberosum]

NCBI Reference Sequence: NP_001305619.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>NP_001305619.1 hexokinase-1 [Solanum tuberosum]
MKKVTVGAAVGAAAVCAVAALIVNHRMRKSSKWGRAMAILREFEEKCKTQDAKLKQVADAMTVEMHAGL
ASEGGQSSRCLSPMSIISQLVMKLGVFYALDLGGTNFRVLRVQLGGKDGGIIHQEFAEASIPPSLMVGTS
DALFDYIAAEALAKFVAEEEEKFHQPQPGKQRELGFHLLIPSNADEFNNSGTIMRWTKGSIDDAVGQDVVG
LTKAMKEVKLDMRVSALVNDTVGTLAGGKYTQKDVAVALGTGTAAYVERVQAIPKWHGPVPKGEMV
INMEWGNFRSSHLPLTEYDHALDNESLNPAEQIFEKMTSGMYLGEILRRVLTRVAEEVLAFLAMRSLSL
KDSFVLRTPDMSAMHHDTSPDLKVVGEEKLKDIILEISNTSLKTRKLVLSCNIVATRGARLDAAGVGLGILK
KMGRDTPKQGGSERVTIAMDGGLYEHYTEYRMCLENSLKDLLGEELATSIVFVHSNDGSGIGAALLRASH
SMYLEDQA
```

5.6 uniprot id file

Function

O64390 · HXK1_SOLTU

Names &

Taxonomy

Proteinⁱ

Hexokinase-1

Amino acids

498

Subcellular

Location

Statusⁱ

UniProtKB reviewed (Swiss-Prot)

Protein existenceⁱ

Evidence at protein level

Phenotypes &

Variants

Organismⁱ

Solanum tuberosum (Potato)

Annotation scoreⁱ

5/5

PTM/Processing

Geneⁱ

HXK1 (HXK)

Expression

Interaction

[Entry](#) [Feature viewer](#) [Publications](#) [External links](#) [History](#)

Structure

[Download](#) [Add a publication](#) [Entry feedback](#)

Family & Domains

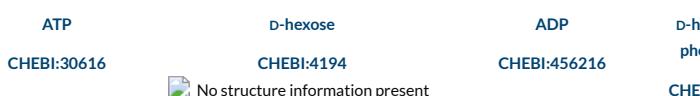
Sequence

Similar Proteins

Feedback

Functionⁱ

Fructose and glucose phosphorylating enzyme. May be involved in the phosphorylation of glucose during the export from plastids to cytosol. Seems neither to be involved in cell sugar sensing nor in carbohydrate metabolism in tuber. [1 Publication](#)

Catalytic ActivityATP + D-hexose = ADP + D-hexose 6-phosphate + H⁺This reaction proceeds in the forward direction. [1 Publication](#)EC:2.7.1.1 (UniProtKB | ENZYME [ENZYME](#) | Rhea [Rhea](#))Source: Rhea 22740 [Rhea](#)[^ Hide Rhea reaction](#)[zoom](#)[zoom](#)[zoom](#)

z

ATP + D-fructose = ADP + D-fructose 6-phosphate + H⁺ [1 Publication](#)This reaction proceeds in the forward direction. [1 Publication](#)EC:2.7.1.1 (UniProtKB | ENZYME [ENZYME](#) | Rhea [Rhea](#))Source: Rhea 16125 [Rhea](#)[^ View Rhea reaction](#)ATP + D-glucose = ADP + D-glucose 6-phosphate + H⁺ [1 Publication](#)This reaction proceeds in the forward direction. [1 Publication](#)EC:2.7.1.1 (UniProtKB | ENZYME [ENZYME](#) | Rhea [Rhea](#))Source: Rhea 17825 [Rhea](#)[^ View Rhea reaction](#)**Kinetics**

KM	SUBSTRATE	pH	TEMPERATURE[C]	NOTES	EVIDENCE
0.033 mM	glucose				1 Publication
1.47 mM	fructose				1 Publication
0.029 mM	mannose				1 Publication
Vmax	pH	TEMPERATURE[C]	NOTES		EVIDENCE
349 nmol/min/mg			with glucose as substrate		1 Publication
523 nmol/min/mg			with fructose as substrate		1 Publication
221 nmol/min/mg			with mannose as substrate		1 Publication

Measured in yeast lacking glucose and hexose kinase activity.

PathwayCarbohydrate metabolism; hexose metabolism. [1 Publication](#)Carbohydrate degradation; glycolysis; D-glyceraldehyde 3-phosphate and glyceraldehyde phosphate from D-glucose: step 1/4. [1 Publication](#)

Features

Showing features for domainⁱ, regionⁱ, binding siteⁱ.

TYPE	ID POSITION(S)	DESCRIPTION	
-- Select --			
► Domain	35-488	Hexokinase  	BLAST
► Region	89-228	Hexokinase small subdomain 	BLAST
► Binding site	101-106	ATP (UniProtKB ChEBI ) 	Sequence Analysis BLAST
► Region	171-197	Glucose-binding 	Sequence Analysis BLAST
► Region	229-477	Hexokinase large subdomain 	BLAST

[Expand table](#)

GO Annotationsⁱ

Slimming set:

plant

Feedback 

Cell color indicative of number of GO terms

ASPECT	TERM	Source
Cellular Component	chloroplast outer membrane 	Source:UniProtKB-SubCell
Cellular Component	cytosol  	Source:GO_Central 1 Publication
Cellular Component	integral component of membrane 	Source:UniProtKB-KW
Cellular Component	mitochondrion  	Source:GO_Central 1 Publication
Molecular Function	ATP binding 	Source:UniProtKB-KW
Molecular Function	fructokinase activity  	Source:GO_Central 1 Publication
Molecular Function	glucokinase activity  	Source:GO_Central 1 Publication

[Expand table](#)

[Complete GO annotation on QuickGO](#) 

Keywordsⁱ

Molecular function

#Kinase

#Transferase

Biological process

#Glycolysis

Ligand

#ATP-binding

#Nucleotide-binding

Enzyme and pathway databases

BRENDA

2.7.1.1  5757

SABIO-RK

O64390 

UniPathway

UPA00109 UER00180

UPA00242

ENZYME

[Search...](#) 

Names & Taxonomyⁱ

Protein namesⁱ

Recommended name

Hexokinase-1

EC number

EC:2.7.1.1  ([UniProtKB](#) | [ENZYME](#)  | [Rhea](#) )

Alternative names
StHK1

Gene namesⁱ

Name
HXK1

Synonyms
HXK

Organism names

Organismⁱ
[Solanum tuberosum \(Potato\)](#)

Taxonomic identifierⁱ
[4113 NCBI](#) ↗

Taxonomic lineageⁱ

Eukaryota > Viridiplantae > Streptophyta > Embryophyta > Tracheophyta > Spermatophyta > Magnoliopsida > eudicotyledons > Gunneridae > Pentapetalae > asterids > Iamiids > Solanales > Solanaceae > Solanoideae > Solaneae > Solanum

Feedback
↗

Accessions

Primary accession
O64390

Proteomesⁱ

Identifier
[UP000011115](#)

Componentⁱ
Unassembled WGS sequence

Subcellular Locationⁱ

[UniProt Annotation](#) [GO Annotation](#)

Plastid, chloroplast outer membrane ; **Single-pass membrane protein**

Features

Showing features for transmembraneⁱ.

TYPE	ID	POSITION(S)	DESCRIPTION
-- Select --			
▶ Transmembrane	4-24	Helical	BLAST

[Expand table](#)

Keywordsⁱ

Cellular component
[#Chloroplast](#)
[#Membrane](#)
[#Plastid](#)
[#Plastid outer membrane](#)

Phenotypes & Variantsⁱ

[Variants](#)

PTM/Processingⁱ

Features

Showing features for chainⁱ.

TYPE	ID	POSITION(S)	DESCRIPTION
-- Select --			
▶ Chain	PRO_0000197615	1-498	Hexokinase-1 BLAST

[Expand table](#)

Expressionⁱ

Tissue Specificityⁱ

Expressed in young and mature leaves, stems, roots, stolons, and developing and mature tubers.

1 Publication

Gene expression databases

ExpressionAtlas

O64390

Interactionⁱ

Protein-protein interaction databases

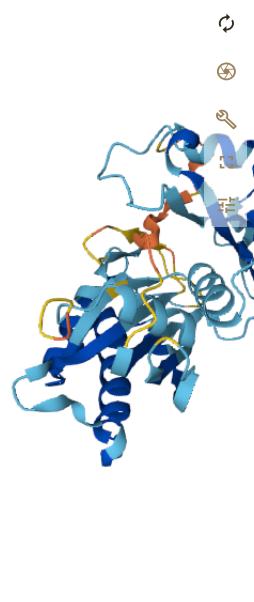
STRING

4113.PGSC0003DMT400006474

Structureⁱ

Model Confidence:Very high
(pLDDT > 90)Confident (90
> pLDDT > 70)Low (70 >
pLDDT > 50)Very low
(pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.



SOURCE

IDENTIFIER

METHOD

RESOLUTION

CHAIN

POSITIONS

LINKS

-- Select --

-- Select --

AlphaFold

AF-O64390-F1 Predicted

1-498

AlphaFold

3D structure databases

AlphaFoldDB

O64390

ModBase

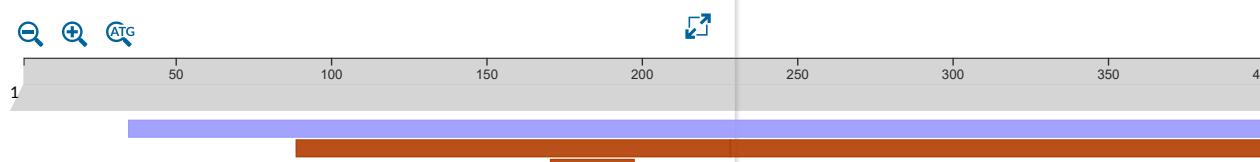
Search...

SMR

O64390

Family & Domainsⁱ

Features

Showing features for domainⁱ, regionⁱ.

TYPE

ID POSITION(S) DESCRIPTION

-- Select --

► Domain	35-488	Hexokinase	BLAST
► Region	89-228	Hexokinase small subdomain	BLAST
► Region	171-197	Glucose-binding	BLAST
► Region	229-477	Hexokinase large subdomain	BLAST

Sequence Similaritiesⁱ

Belongs to the hexokinase family.

[Expand table](#)

Keywordsⁱ

Domain

#Transmembrane

#Transmembrane helix

Phylogenomic databases

InParanoid

O64390

eggNOG

KOG1369 Eukaryota

Family and domain databases

InterPro

[View protein in InterPro](#)

IPR043129 ATPase_NBD

IPR001312 Hexokinase

IPR022673 Hexokinase_C

IPR022672 Hexokinase_N

PANTHER

PTHR19443 PTHR19443 1 hit

PROSITE

[View protein in PROSITE](#)

PS51748 HEXOKINASE_2 1 hit

Pfam

[View protein in Pfam](#)

PF00349 Hexokinase_1 1 hit

PF03727 Hexokinase_2 1 hit

SUPFAM

SSF53067 SSF53067 2 hits

MobiDB

[Search...](#)

ProtoNet

[Search...](#) [Feedback](#)

Sequenceⁱ

Length

498

Mass (Da)

54,130

Last updated

1998-08-01 v1

Checksumⁱ

D7F598CBB12762E4

```
MKKVTGAAVGAAVCAVAALIVNHRMRKSSKWGRAMAILREFEEKCKTQDAKLKQVADAMTVEM
HAGLASEGGQSSRCLSPMSIISQLVMKLGVFYALDLGGTNFRVLRVQLGGKDGGIIHQFEAEASIPPSLMV
GTSDALFDYIAEELAKFVAEEEFHQPPGKQRELGFHLIPSNADEFNNSTIMRWTKGFSIDDAVGQD
VVGELETAKMEKVLDMRVSLAVNDTVGLLAGGKYTQKDVA/AVILGTGTNAAY/VERVQAIPKWHGPV
PKSGEMVNMEWGNFRSSHPLTEYDHALDNESNPAEQIFEKMTSGMYLGEILRRVLTRVAEEVLAFLA
MRSLQLSKDSFVLRTPDMSAMHHDTSPDLKVVGKEKLKDILEISNTSLKTRKLVLSLCNIVATRGARLDA
GVLGILKKMGRDTPKQGGERTVIAMDGGLYEHYTEYRCMLENSLKDLLGEELATSIVFVHSNDGSGIG
AALLRASHSMYLEDQA
```

Keywordsⁱ

Technical term

#Reference proteome

Sequence databases

PIR

T07384 T07384

RefSeq

NP_001305619.1 NM_001318690.1

SEQUENCE	PROTEIN	MOLECULE TYPE	STATUS
X94302 (EMBL GenBank DDBJ)	CAA63966.1 (EMBL GenBank DDBJ)	mRNA	

[Expand table](#)

Genome annotation databases

GeneID

102604144

KEGG

sot:102604144

Similar Proteinsⁱ



Disclaimer

Any medical or genetic information present in this entry is provided for research, educational and informational purposes only. It is not in any way intended to be used as a substitute for professional

medical advice, diagnosis, treatment or care. Our staff consists of biologists and biochemists that are not trained to give medical advice.



© 2002 – 2022 UniProt consortium

[License & Disclaimer](#) | [Privacy Notice](#)

[Release 2022_04](#) | [Statistics](#)

Core data	Supporting data	Information
Proteins (UniProtKB)	Literature citations	Cite UniProt⁹⁹
Species (Proteomes)	Taxonomy	About & Help
Protein clusters (UniRef)	Keywords	UniProtKB manual
Sequence archive (UniParc)	Subcellular locations	Technical corner
	Cross-referenced databases	Expert biocuration
	Diseases	Statistics

[Get in touch](#)



UniProt is an ELIXIR core data resource

Main funding by: [National Institutes of Health](#) State Secretariat for Education, Research and Innovation SERI

Feedback

5.7a Alpha fold

AlphaFold Protein Structure Database

Search for protein, gene, UniProt accession or organism

BETA



Examples:

Free fatty acid receptor 2

(./search/uniprotDescription/Free%20fatty%20acid%20receptor%202)

At1g58602 (./search/text/At1g58602) Q5VSL9 (./search/text/Q5VSL9)

E. coli (./search/text/Escherichia%20coli) Help: AlphaFold DB search help (./faq#faq-2)

Hexokinase-1

AlphaFold structure prediction

Download

PDB file (https://alphafold.ebi.ac.uk/files/AF-O64390-F1-model_v4.pdb)

mmCIF file (https://alphafold.ebi.ac.uk/files/AF-O64390-F1-model_v4.cif)

Predicted aligned error (https://alphafold.ebi.ac.uk/files/AF-O64390-F1-predicted_aligned_error_v4.json)

Note: We have recently updated the PAE JSON format, please refer to our FAQ (/faq#faq-7) for a description of the updated format.

NEW Feedback on structure

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) (<https://www.ebi.ac.uk/data-protection/privacy-policy/embl-ebi-public-website>) and [Terms of Use](#) (<https://www.ebi.ac.uk/about/terms-of-use>).

[I agree, dismiss this banner](#)

Looks great

(https://docs.google.com/forms/d/e/1FAIpQLSel21_fnkopZuGGDiPF5T-fe-q4glwA17eW1WsMx2QLi0PX3Q/viewform?usp=pp_url&entry.62243469=064390)

Could be improved

(https://docs.google.com/forms/d/e/1FAIpQLScI4yiIyHQzpxND04Y47Kw-hyjsjCerb-w8jdCXFki2OW9Kxw/viewform?usp=pp_url&entry.62243469=064390)

Information ^

Protein	Hexokinase-1
Gene	HXK1
Source organism	Solanum tuberosum (Potato) go to search ↗
UniProt	064390 go to UniProt ↗
Experimental structures	None available in the PDB
Biological function	Fructose and glucose phosphorylating enzyme. May be involved in the phosphorylation of glucose during the export from plastids to cytosol. Seems neither to be involved in cell sugar sensing nor in carbohydrate metabolism in tuber. go to UniProt ↗

3D viewer ⓘ

Sequence of	AF-O64390-F1	Chain	1: Hexokinase-1	A	②	
1	11	21	31	41	51	61
MKKVTVGAAVGAAAVCAVAALIVNHRMRKSSKWGRAMAILREFEEKCKTQDAKLKVADAMTVEMH						
71	81	91	101	111	121	131
AGLASEGGQSSRCLSPMSIISQLVMKLGVFYALDLGGTNFRVRLRVQLGGKDGGIIHQEFAEASIPPS						
141	151	161	171	181	191	201
LMVGTSDALFDYIAAEELAKFVAAEEEKFHQPPGKQRELGFHLLIPSNADEFNNSGTIMRWTKGFSIDD						

Model

Confidence:

Very high (pLDDT > 90)

Confident (90 > pLDDT > 70)

Low (70 > pLDDT > 50)

Very low (pLDDT < 50)

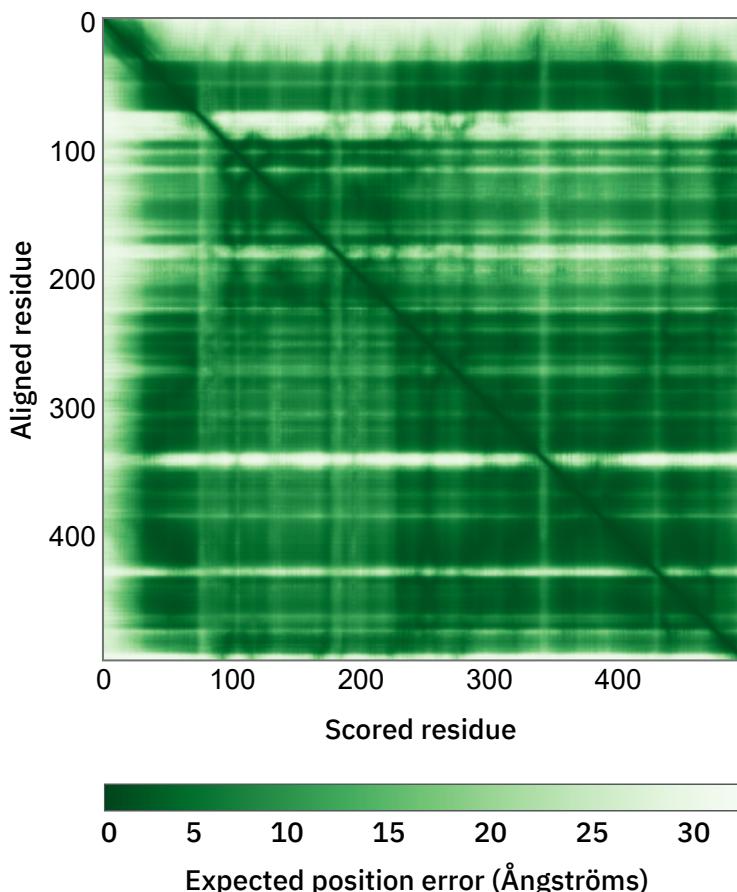
This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our Privacy Notice (<https://www.ebi.ac.uk/data-protection/privacy-policy/embl-ebi-public-website>) and Terms of Use (<https://www.ebi.ac.uk/about/terms-of-use>).

[I agree, dismiss this banner](#)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Views

Predicted aligned error



Predicted aligned error

The colour at position (x, y) indicates AlphaFold's expected position error at residue x, when the predicted and true structures are aligned on residue y.

This is useful for assessing inter-domain accuracy - see the tutorial below.

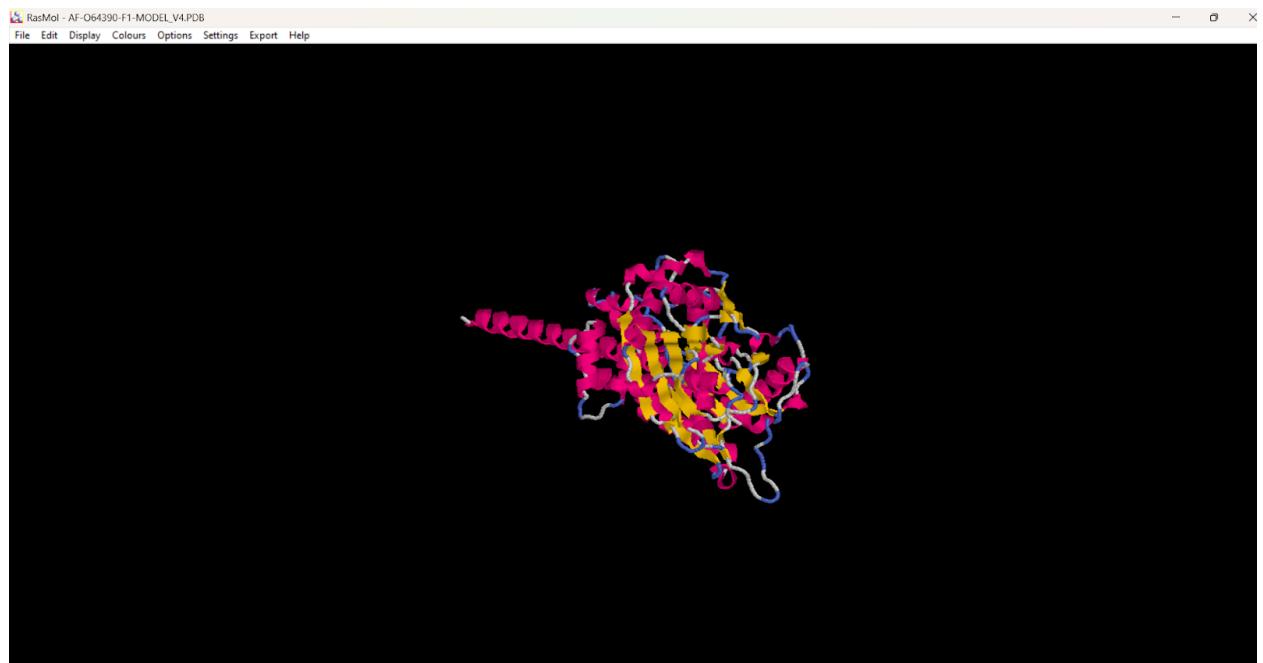
Predicted aligned error tutorial



This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) (<https://www.ebi.ac.uk/data-protection/privacy-policy/embl-ebi-public-website>) and [Terms of Use](#) (<https://www.ebi.ac.uk/about/terms-of-use>).

[I agree, dismiss this banner](#)

5.7b Rasmol



6. Conclusions

HXK1 (Hexokinase-1) is involved in the glycolysis and it has a length of 1619 base pairs. The length of protein is 498 amino acids . The structural data is not available in RCSB and is predicted by Alphafold