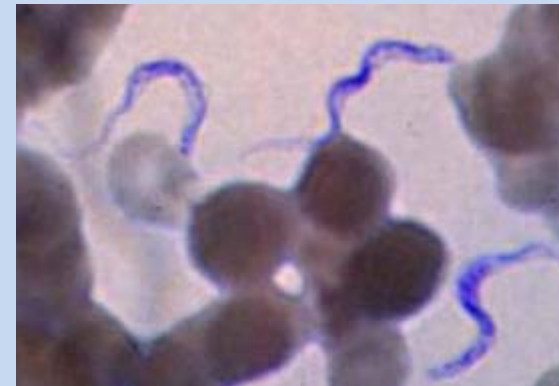
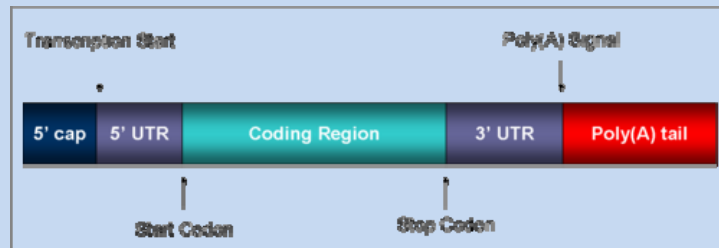
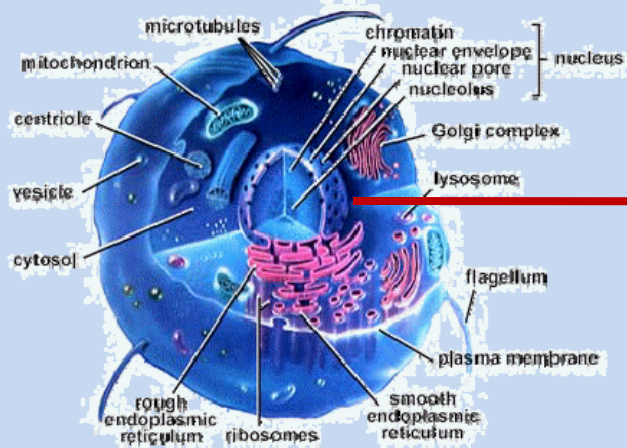


# UTR PREDICTION PROGRAMS FOR TRYPANOSOMA BRUCEI

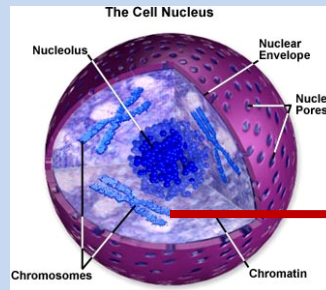


Maria Moutafis  
Masters Thesis Computational Biology  
December 2007

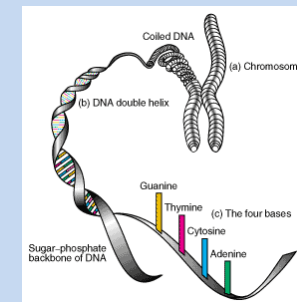
# Biology Basics



Cell

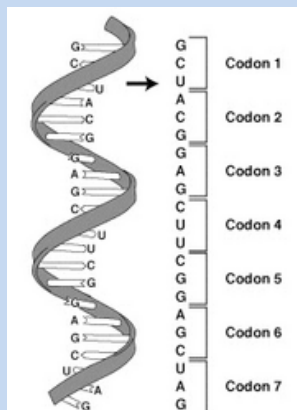


Nucleus



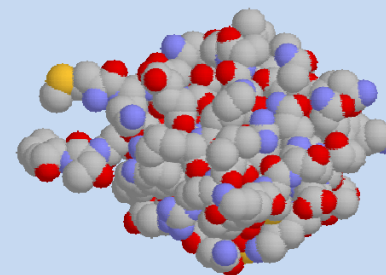
DNA

transcription



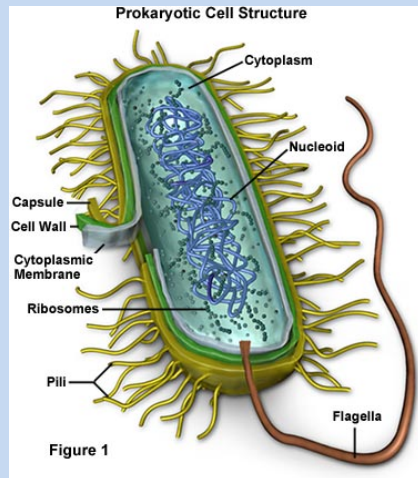
RNA

translation



Protein

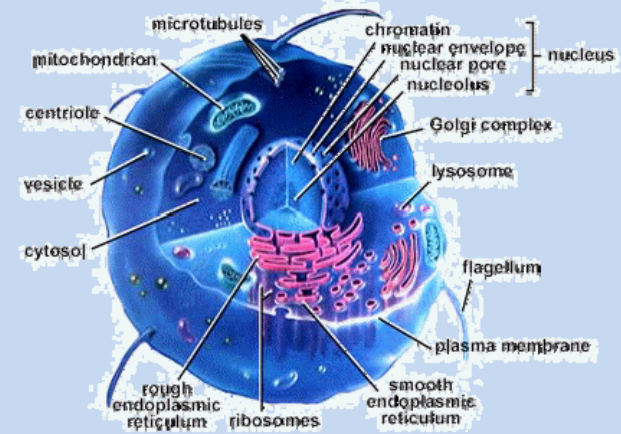
# Cell Type



[micro.magnet.fsu.edu/cells/bacteriacell.html](http://micro.magnet.fsu.edu/cells/bacteriacell.html)

## Prokaryotic cell

Bacteria  
Virus

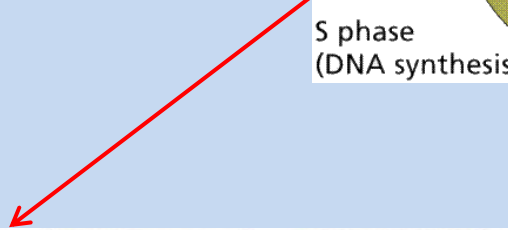
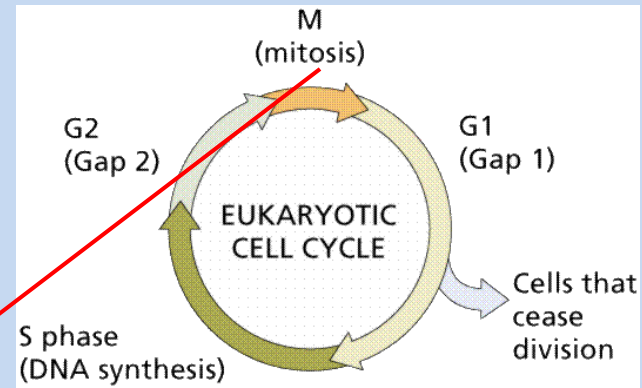
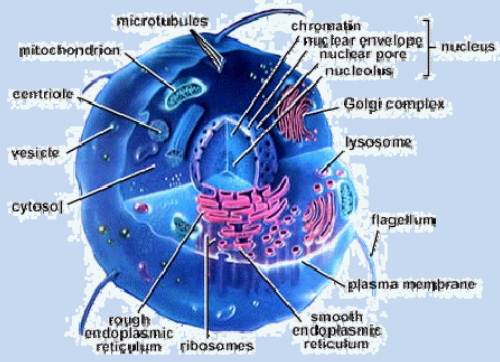


<http://www.modares.ac.ir/elearning/mnaderi/Genetic%20Engineering%20course%20II/images/wpe16.gif>

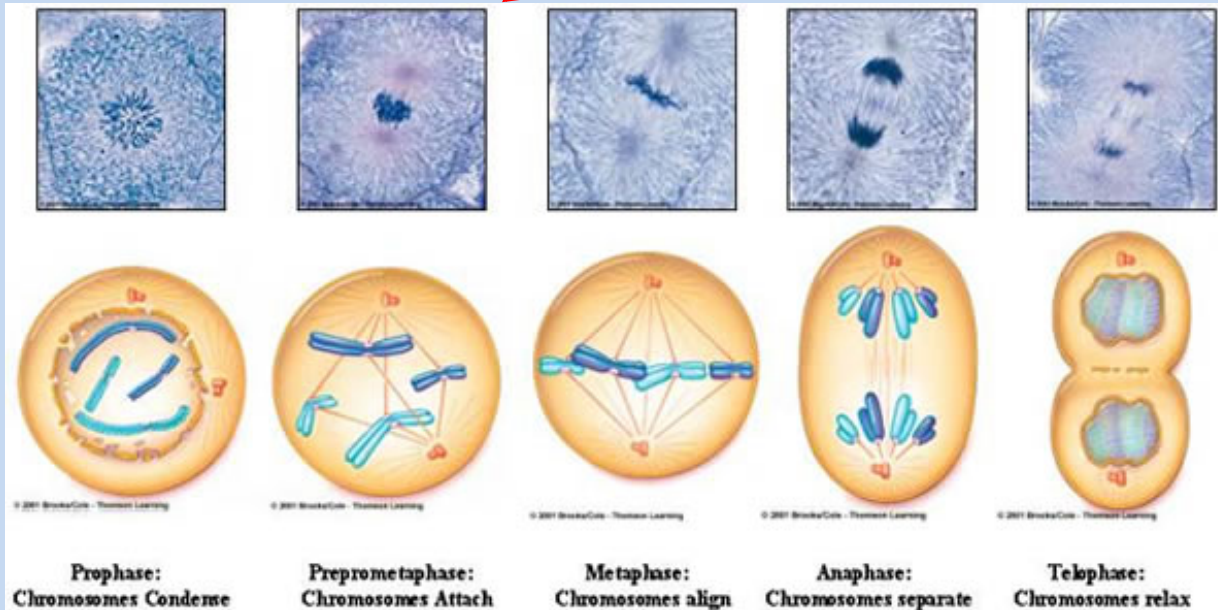
## Eukaryotic cell

Animals  
Fungi  
Plants  
Protists

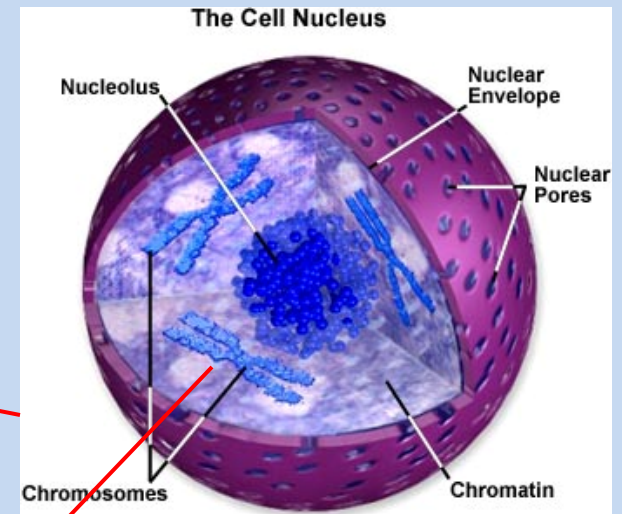
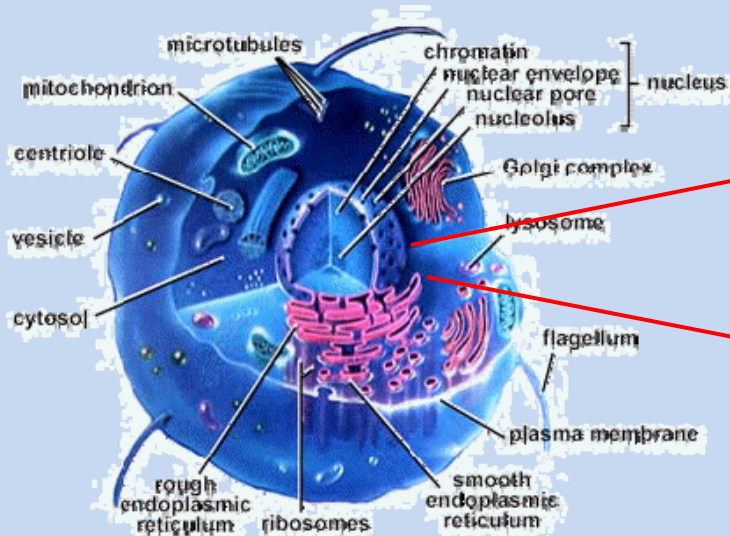
# Eukaryotic Cell



## Mitosis



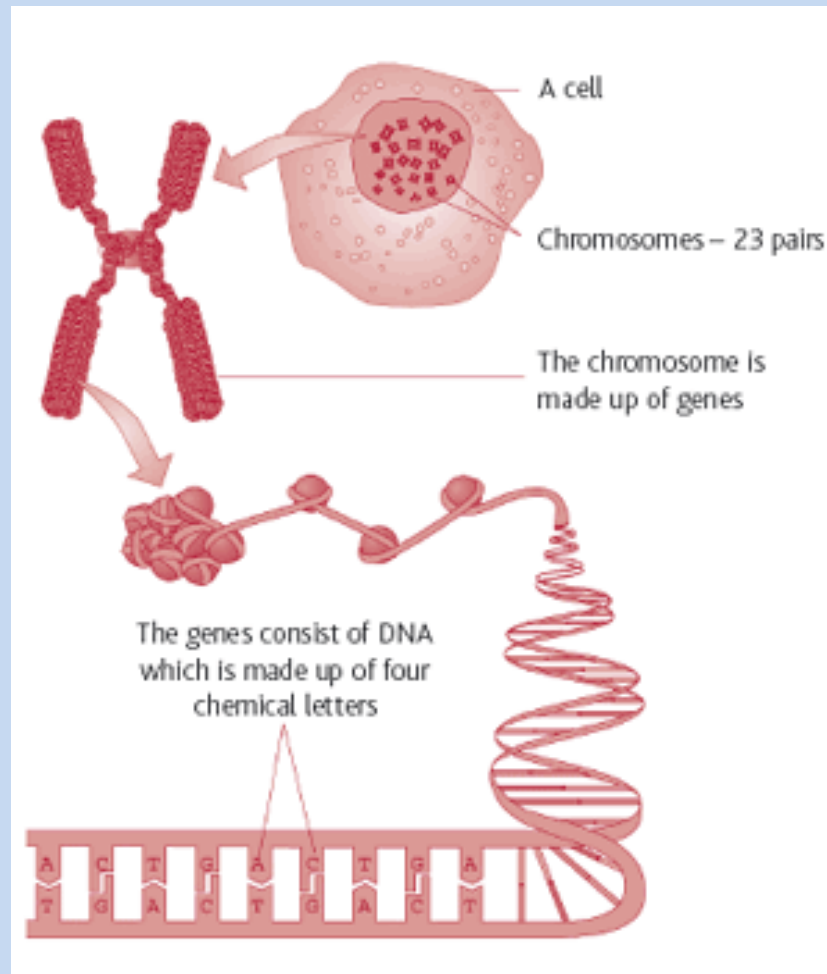
# The Nucleus



<http://micro.magnet.fsu.edu/cells/plants/images/cellnucleus.jpg>

Chromosomes are made of gene which contain the cell's genetic material

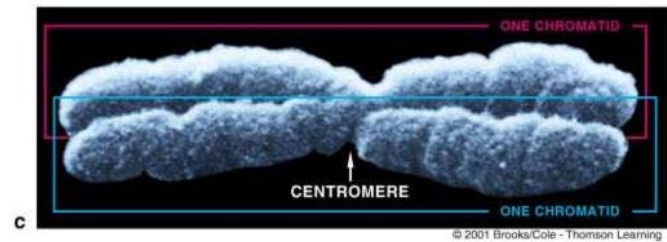
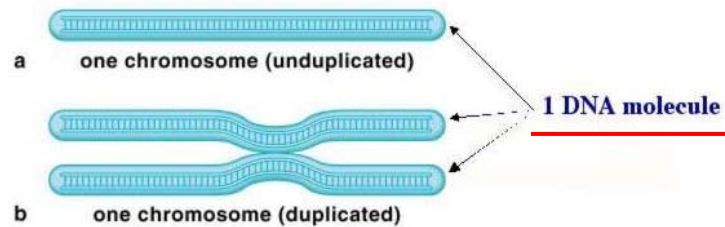
# Genes



<http://www.oncolink.com/library/images/id818-1.gif>

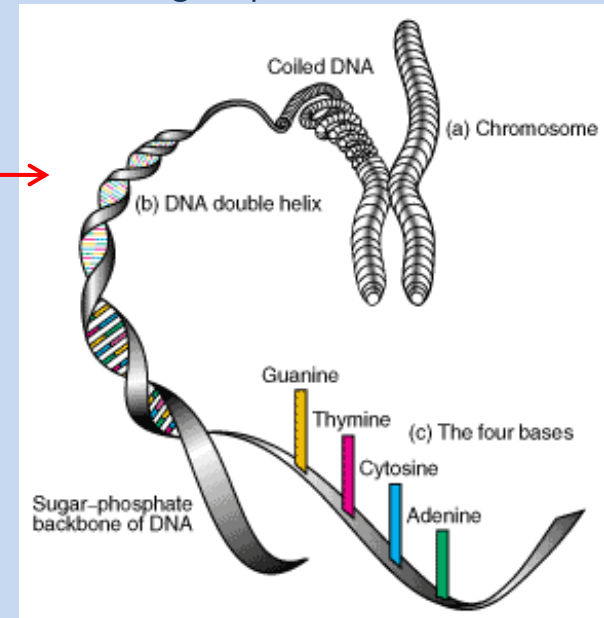
# DNA

## Chromosomes are made of DNA molecules



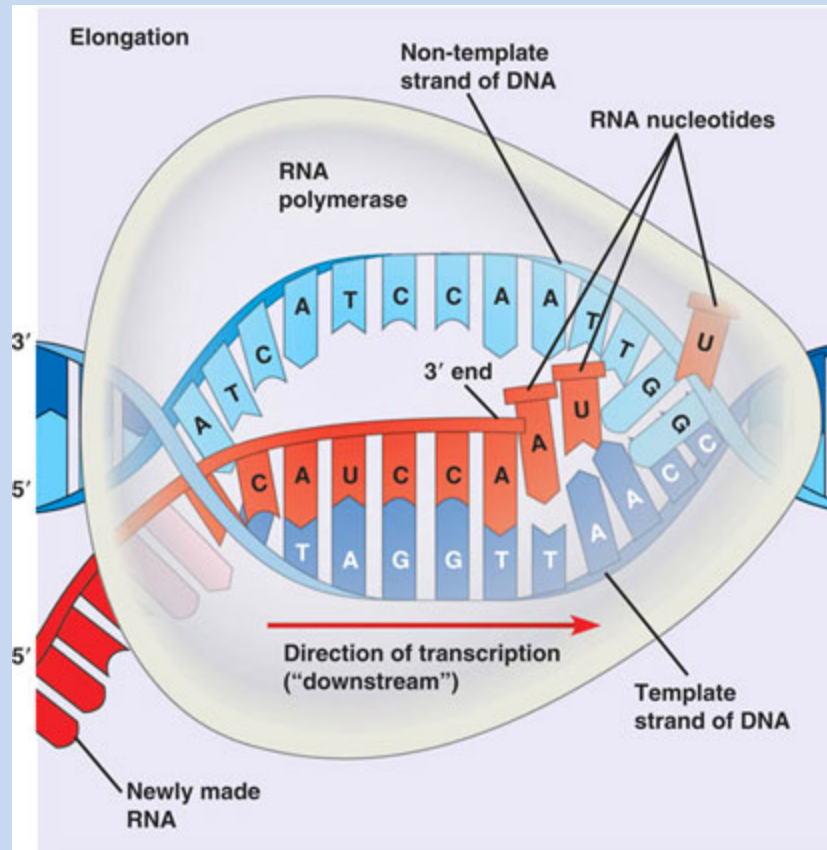
<http://www.ualr.edu/botany/chromosome.jpg>

Must unwind to undergo replication



<http://www.llnl.gov/str/June03/gifs/Stubbs1.gif>

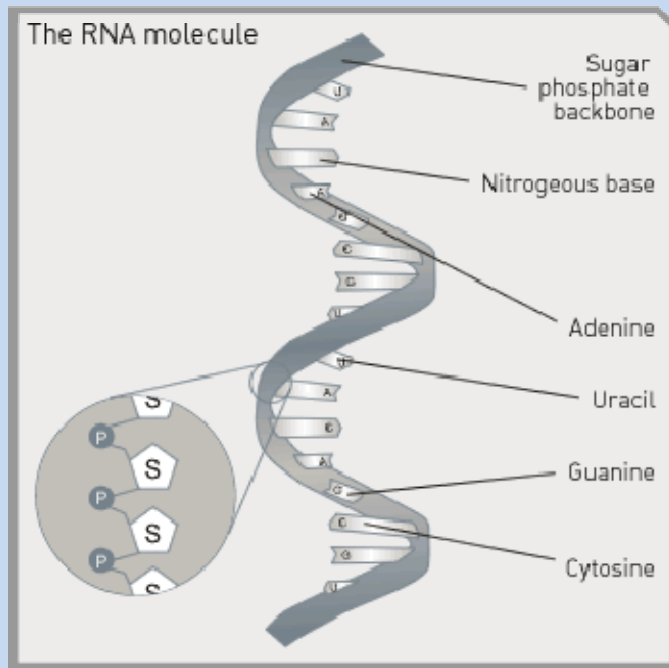
# Transcription



<http://fig.cox.miami.edu/~cmallery/150/gene/c7.17.7b.transcription.jpg>



# RNA



3 TYPES OF

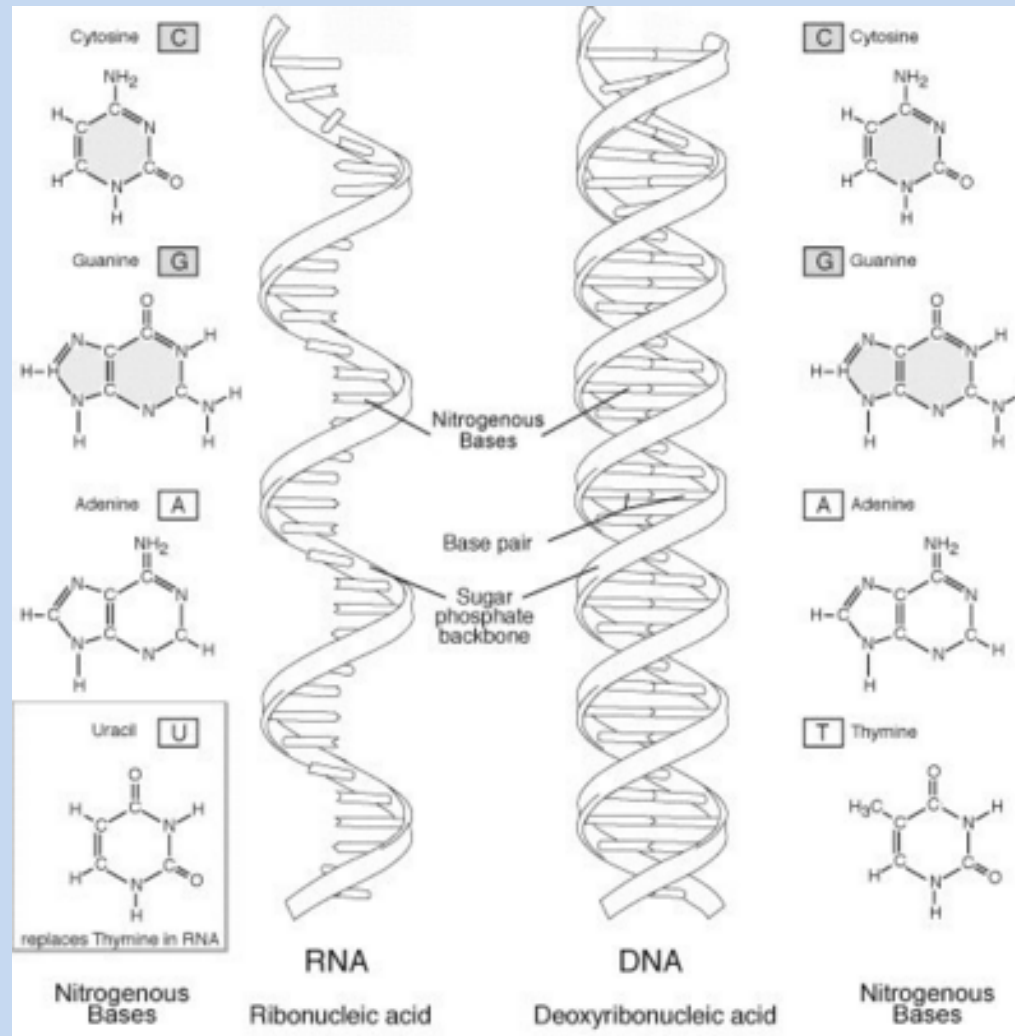
mRNA – messenger RNA

tRNA – transfer RNA

rRNA – ribosomal RNA

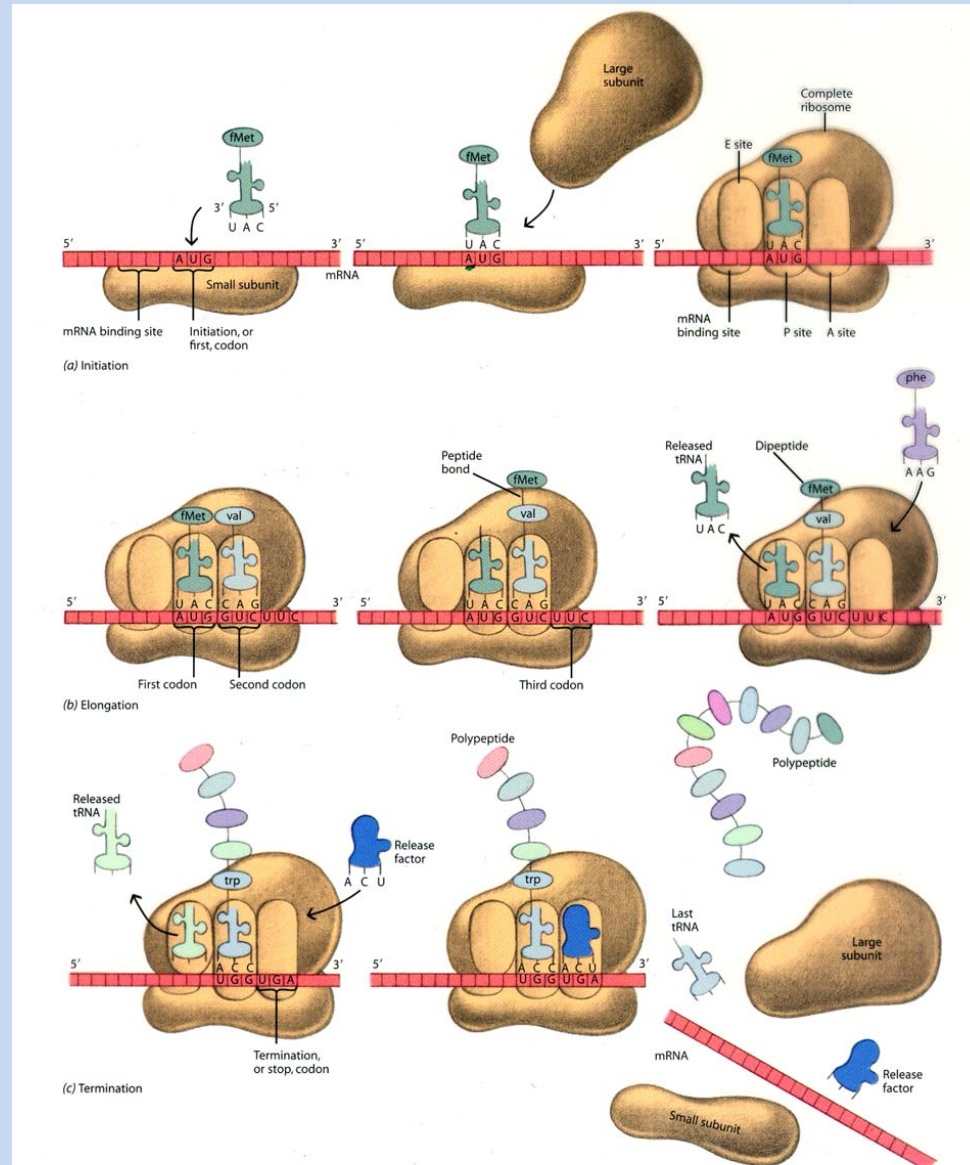
All of which assist in the process of translation

# RNA vs. DNA – structural differences



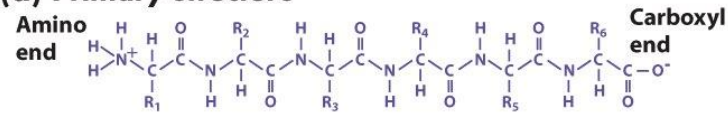
<http://www.search.com/reference/DNA>

# Translation

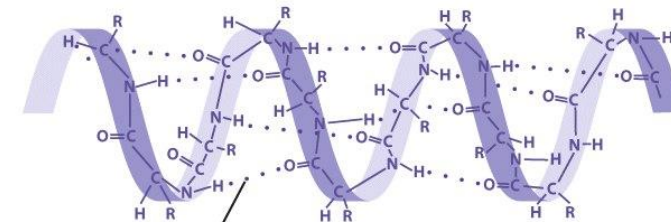


# Proteins

## (a) Primary structure

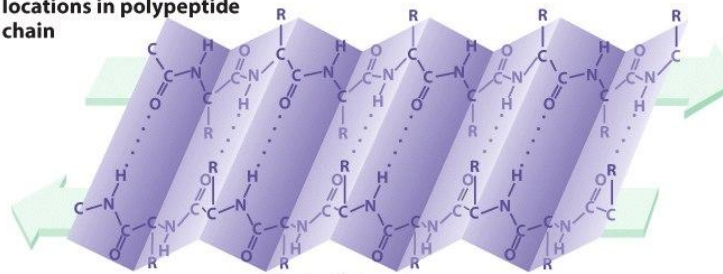


## (b) Secondary structure



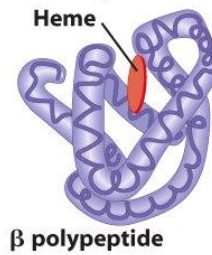
Hydrogen bonds between amino acids at different locations in polypeptide chain

$\alpha$  helix



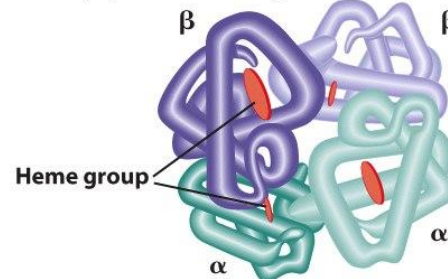
Pleated sheet

## (c) Tertiary structure



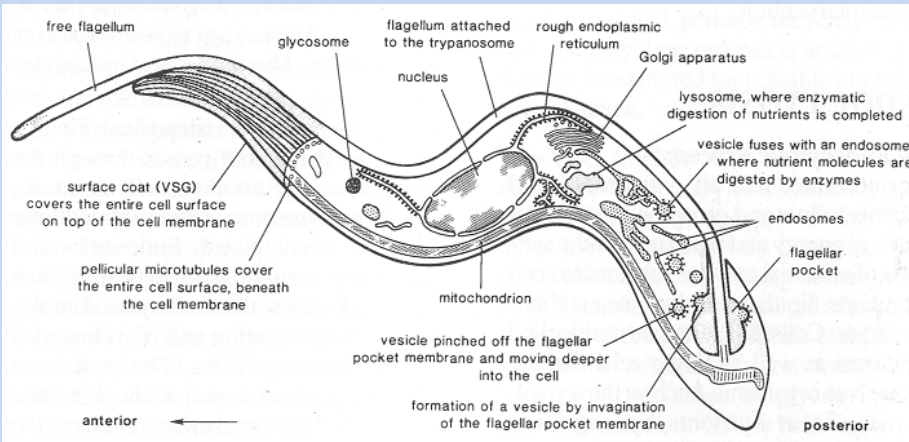
$\beta$  polypeptide

## (d) Quaternary structure



Heme group

# Trypanosoma brucei



<http://www.ilri.org/InfoServ/Webpub/Fulldocs/llradRe1989v7n1/FIG%201%20P1.gif>



<http://www.nature.com/ng/journal/v32/n3/images/ng1102-335-11.jpg>

- **Protists**
- **Cause African Trypanosomiasis also known as African Sleeping Sickness**
- **There are 2 subspecies of T. brucei; T.b.brucei, T.b.gambiense, T.b.rhodesiense**
- **Has 2 hosts: an insect vector (tsetse fly) and a mammalian host**
- **Found in Africa**

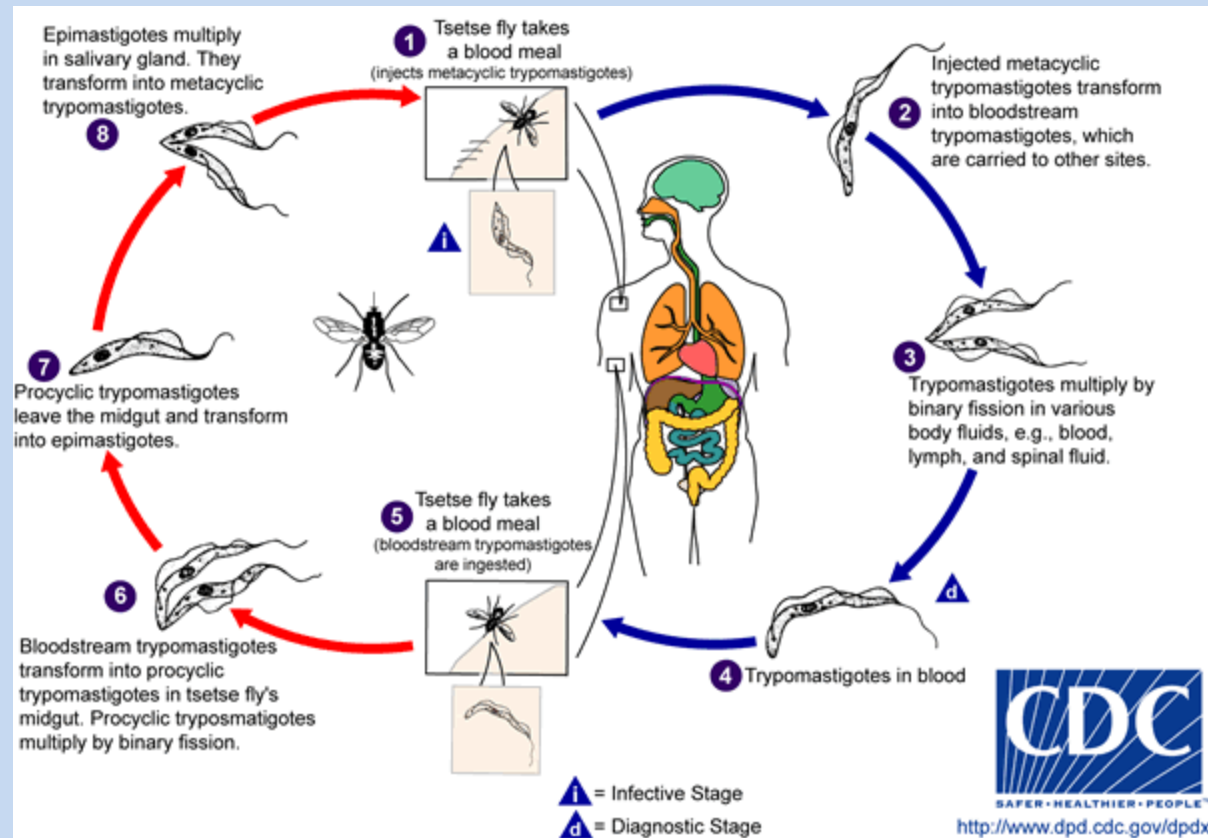


[http://www.sciam.com/media/inline/000E7E77-EEBB-1E36-82FC809EC5880000\\_1.jpg](http://www.sciam.com/media/inline/000E7E77-EEBB-1E36-82FC809EC5880000_1.jpg)

# T. Brucei – Life Cycle

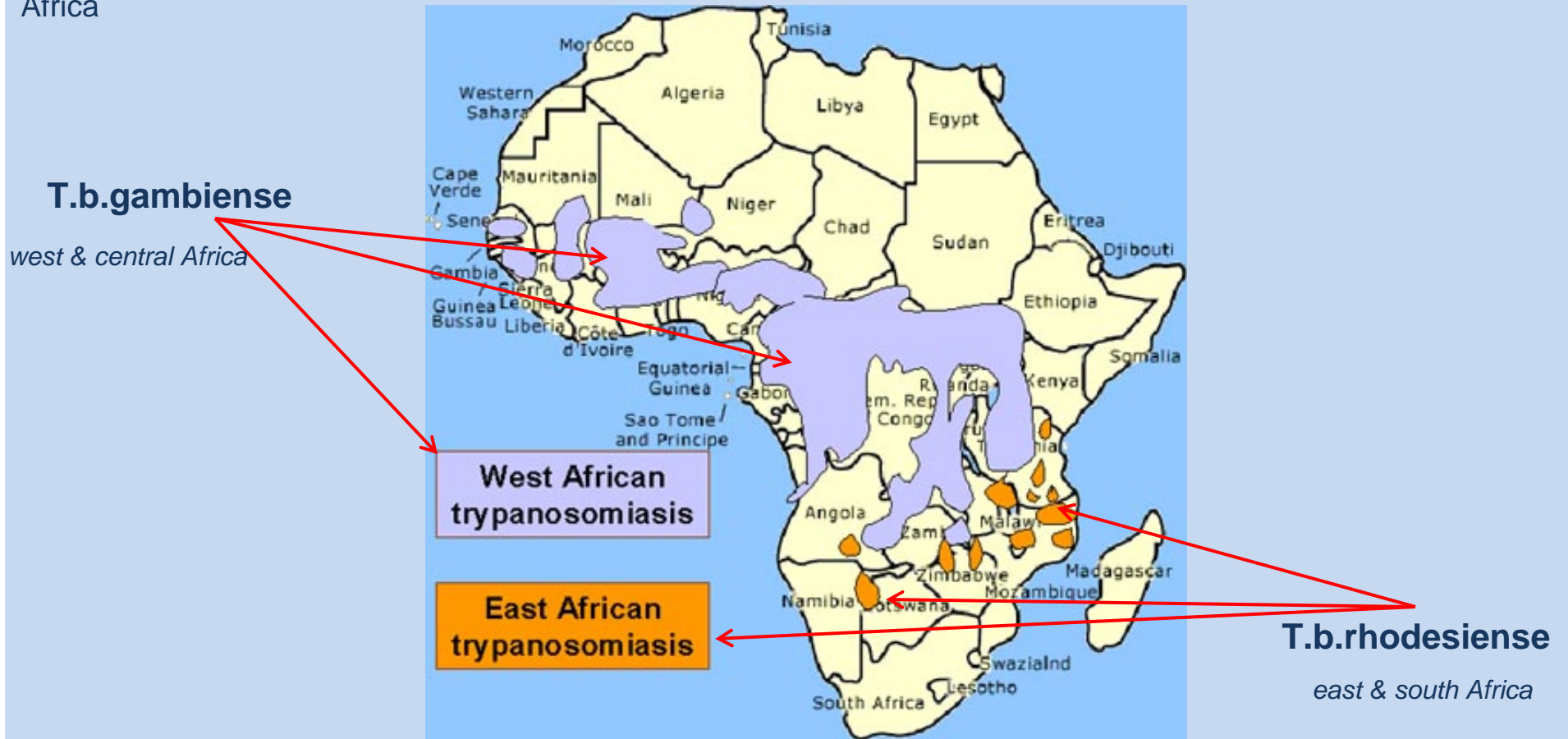
## Tsetse Fly Stages

## Human Stages



# Areas Affected by T. brucei

36 countries in sub-Saharan Africa



<http://pathmicro.med.sc.edu/parasitology/tryp-map.jpg>

Facts:  
<http://www.who.int/tdr/diseases/tryp/diseaseinfo.htm>

# African Trypanosomiasis

**General symptoms (appear 1-4 weeks after fly bite),**

**They include:**

- Swollen, red, painful nodule at site of fly bite
- Swollen lymph nodes
- Severe headaches
- Fever
- Sweating
- Irritability
- Insomnia at night
- Confusion
- Seizures
- Extreme fatigue
- Uncontrollable urge to sleep – in late stage, this leads to coma





# Disease Prevention & Control

**Prevention & Control** are extremely difficult.  
There is no vaccine for the disease.

WHO suggests:

- Wearing protective clothing, such as long-sleeved shirts and pants.
- Using insect repellent.
- Using bed netting when sleeping.
- Avoid bushes (they are the breeding and resting site for the fly).

Note: some of these suggestions are difficult to follow in such remote areas of the Africa

## **Treatment**

Medications used to treat this disorder include:

- Suramin (Antrypol) – IV form, adverse side-effects
- Melarsoprol – used for late stage, serious side-effects, may be fatal
- Pentamidine – not used in late stage, many strains are resistant to it
- Eflornithine

## **Exams and Tests**

A physical examination may show signs of meningoencephalitis (inflammation of the brain and meninges).

Common Tests include the following:

- Blood smear
- Lymph node aspiration
- CSF test
- CATT test

# Data Collection

## Part 1

Welcome to the GeneDB website Version 2.1

Database Entry Point

Searches

Sequence Searches

Datasets

Search for gene by ID/description in

Include description in search

Add wildcards to search term

omniBLAST (Multi-organism BLAST)

single organism BLAST

Fungi

Protozoa

Go To Choose...

Go To T. brucei

Go To Choose...

Go To D. discoideum

Go To E. histolytica

Go To L. tenella

Go To L. braziliensis

Go To L. infantum

Go To P. major

Go To P. berghei

Go To P. chabaudi

Go To P. falciparum

Go To P. knowlesi

Go To T. annulata

Go To T. brucei

Go To T. brucei strain 427

Go To T. congolense

Go To T. cruzi

Go To T. b. gambiense

Go To T. vivax

<http://www.genedb.org/>

## Part 2

What's New

New! A Eukaryotic Genome Annotation and Analysis Training Course is being offered at TIGR. More information and the registration form are available here.

The TIGR Trypanosoma brucei Genome Project

Listed below is a collection of information on *Trypanosoma brucei* genome research at TIGR. Please follow the links to get more detailed information for each project.

Funding for this project is being provided by the National Institute of Allergy and Infectious Diseases (NIAID), R01AI43062

Introduction and Sequencing methodology

A brief summary of the *Trypanosoma brucei* genome project and a detailed description of our sequencing strategy. Sequences from the ends of BAC clones provide highly specific markers and are used in a map-as-you-go strategy for genome sequencing.

End Sequencing and Gene Discovery

A database of ~47,000 end sequences generated at TIGR from BAC, F1 and small-insert random sheared libraries can be searched by clone name or sequence.

BAC Sequencing at TIGR

The progress of our contiguous sequencing project can be viewed here.

*Trypanosoma brucei* Genome Annotation Database (TGAD)

This database contains every contiguous sequence derived from the TIGR sequencing project. Annotated sequences are displayed based on their map position and are searchable by BAC clone name as well as by gene name.

*Trypanosoma brucei* Gene Index (TbGI)

All publicly available *T. brucei* ESTs as well as transcripts from Genbank have been assembled into contigs and built into the *Trypanosoma brucei* Gene Index (TbGI).

Data Release Policy

<http://www.tigr.org/tldb/tbgi/>

<http://www.tigr.org/tldb/e2k1/tba1/>

## Part 3

Computational Biology and Functional Genomics Laboratory

The Gene Index Project

Home Resourcer Gene Indices Genomic Maps EGO

Links

The Gene Indices

TGI Software

What's new?

Definitions

TGI Publications

TGI FAQ

TGI Disclaimer

Contact Us

The Gene Index Project Overview

The promise of genome projects is to provide a complete catalogue of genes in a wide range of organisms. While genome projects are successful in providing reference genome sequences, the problem of finding genes and their products remains an ongoing challenge.

The sequencing of Expressed Sequence Transcripts (ESTs), fragments of genes that have been copied from DNA to RNA, provides the most comprehensive evidence for the existence of genes and their structure.

The goal of The Gene Index Project is to use the available EST and gene sequences, along with the reference genomes wherever available, to provide an inventory of likely genes and their variants and to annotate these with information regarding the functional roles played by these genes and their products.

In addition, we are attempting to use these catalogues to find links between genes and pathways in different species and to provide lists of features within completed genomes that can aid in the understanding of how gene expression is regulated.

<http://compbio.dfci.harvard.edu/tgi/>

<http://compbio.dfci.harvard.edu/tgi/>

## Part 4

Computational Biology and Functional Genomics Laboratory

The Gene Index Project

Home Resourcer Gene Indices Genomic Maps EGO

The DFCI Trypanosoma brucei Gene Index (TbGI)

TbGI Library Sequence Download:

Please select cat# and choose the type of sequences to download

Library Description

Sequence Type:  EST  TC

Download Sequences

Reset

0191
0471
1038
1143
1487

[http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/cat\\_download.pl?db=tbest](http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/cat_download.pl?db=tbest)

# BLAST

## Part 1

NCBI **nucleotide-nucleotide BLAST**  
 Nucleotide Protein Translations Retrieve results for an RLC

>AA003459  
 CGAAATACTGTAAAACGTTCTCGACGGAGTGCACCGCCAGATAAACAAATAATGATGTG  
 GAACAAGGGAGAGAATTTTTTTTAAAAGAGAAAAAAGAGTAACTTCACTCCCTCCAG  
 ACTTATAGGTGGAACCTCTACGTCTTATGCGGTTCCCGCGGTANATTGTGTTGCTA  
 CCATCCGCGTGCCTCTTCTCGAGGCAACATCATCGTTGTCCTCCGACAAACGCAATTGGA

Search

Set subsequence From: To:

Choose database  
 Human genomic plus transcript  
 Mouse genomic plus transcript  
 Others (nr etc.):  
 nr

Now: **BLAST!** or [Reset query](#) [Reset all](#)

Options for advanced blasting

## Part 2

NCBI **formatting BLAST**  
 Nucleotide Protein Translations Retrieve results for an RLC

Your request has been successfully submitted and put into the Blast Queue.  
 Query = AA003459(404 letters)

The request ID is 1171214707-20372-195174200953.BLASTQ3

**Format!** or [Reset all](#)

The results are estimated to be ready in 4 seconds but may be done sooner.

Please press "FORMAT!" when you wish to check your results. You may change the formatting options for your result via the form below and press "FORMAT!" again. You may also request results of a different search by entering any other valid request ID to see other recent jobs.

Format

Show  Graphical Overview  Linkout  Sequence Retrieval  NCBI-gi Alignment in HTML [Format](#)

CDS feature

New View

## Part 3

Distance tree of results **NEW**

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer

Sequences producing significant alignments:  
 (Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
<a href="#">XM_824634.1</a>	Trypanosoma brucei TREU927 hypothetical protein (Tb11.01.8430)	498	498	69%	9e-138	98%	<a href="#">G</a>
<a href="#">AM408590.1</a>	Mycobacterium bovis BCG Pasteur 1173P2, complete genome	42.1	42.1	5%	1.6	100%	
<a href="#">XM_417442.2</a>	PREDICTED: Gallus gallus hypothetical LOC419269 (LOC419269),	42.1	42.1	5%	1.6	100%	
<a href="#">AC120359.8</a>	Mus musculus chromosome 5, clone RP24-285E7, complete sequence	42.1	42.1	5%	1.6	100%	<a href="#">G</a>
<a href="#">AF000516.2</a>	Mycobacterium tuberculosis CDC1551, complete genome	42.1	42.1	5%	1.6	100%	
<a href="#">AC121774.3</a>	Mus musculus BAC clone RP23-350E1 from chromosome 5, complete	42.1	42.1	5%	1.6	100%	
<a href="#">U00024.1</a>	Mycobacterium tuberculosis cosmid tbc2	42.1	42.1	5%	1.6	100%	
<a href="#">BX942581.1</a>	Mycobacterium tuberculosis H37rv complete genome; segment 10/13	42.1	42.1	5%	1.6	100%	
<a href="#">BX248344.1</a>	Mycobacterium bovis subsp. bovis AF2122/97 complete genome;	42.1	42.1	5%	1.6	100%	
<a href="#">AD000009.1</a>	Mycobacterium tuberculosis sequence from clone y2	42.1	42.1	5%	1.6	100%	
<a href="#">AM429234.1</a>	Vitis vinifera, whole genome shotgun sequence, contig	40.1	40.1	5%	6.2	95%	

## Part 4

> [\[ref\] XM\\_824634.1](#) [G](#) Trypanosoma brucei TREU927 hypothetical protein (Tb11.01.8430)  
 partial mRNA  
 length=2892

Score = 498 bits (251), Expect = 9e-138  
 Identities = 277/282 (98%), Gaps = 3/282 (1%)  
 Strand=Plus/Minus

Query 121 TTATAGGTGGAACCTCTACGTCTTATGCGGTTCCCGGGGTAN-ATTTGTGCTACC 179  
 Sbjct 282 TTATAGGTGGAACCTCTACGTCTTATGCGGTTCCCGGGGTAGATTTGTGCTACC 2833

Query 180 ATCCGGCGTGCCTCTTCTCGAGGCAACATCATCGTTGTCCCGGACAAACGCATTGGAATC 239  
 Sbjct 2832 ATCCGGCGTGCCTCTTCTCGAGGCAACATCATCGTTGTCCCGGACAAACGCATTGGAATC 2773

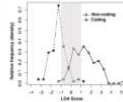
Query 240 TTGTCCTCCACAGCAGAGGAAAGCGGCGGGCGTGAATGTATCAGCATCATCTGTAT 299  
 Sbjct 2772 TTGTCCTCCACAGCAGAGGAAAGCGGCGGGCGTGAATGTATCAGCATCATCTGTAT 2713

Query 300 GCAGTCCTTTACCCCTTTGGTTTTGTGGAGGATATAAGTGTGCGGCTTCAACG-CCC 358  
 Sbjct 2712 GCAGTCCTTTACCCCTTTGGTTTTGTGGAGGATATAAGTGTGCGGCTTCAACGCCCC 2653

Query 359 ACGTGTCTTCGCGCTTGATGACGGG-CATCGCTTCAAAGCATT 399  
 Sbjct 2652 ACGTGTCTTCGCGCTTGATGACGGGCACTCTTCAAAGCATT 2611

# ORBIT

## ORBIT



ORBIT stands for **OR F B**inary **I**dentification **T**ool. ORBIT is an open reading frame (ORF) verification method based on statistical analysis of the nucleotide profiles of known coding and non-coding regions in *Trypanosoma brucei*, the causative agent of sleeping sickness. The method is described in the paper:

Gopal, S., Cross, G.A.M. and Gaasterland, T. (2003). "An organism-specific method to rank predicted coding regions in *Trypanosoma brucei*," *Nucleic Acids Research* **31**:5877-5885.

The result of ORBIT's analysis is a **binary** decision as to the likelihood that a given ORF is coding. ORBIT also provides an associated likelihood score which reflects the confidence in the predicted status of a given sequence.

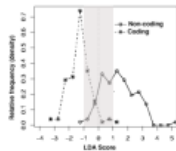
This web interface allows users to access ORBIT but has some restrictions on use. A total of **500,000 basepairs of total sequence data may be analyzed at any one time**. This can be pasted in as one contiguous sequence or as a multi-sequence FASTA formatted file. At this time, ORBIT does NOT predict open reading frames from raw genomic sequence data. To use this interface, you must have previously identified candidate ORFs by some other method. ORBIT will only provide a prediction of the likelihood of coding given the proposed open reading frame's sequence. It may also be used to ascertain whether a sequence is likely to be non-coding.

To evaluate a single sequence for its likelihood of coding, enter the sequence (raw sequence, no headers please):

```
TTATAGGTGGAACCTCTACGTCCTTATGCGGTTCCCGGCGTANATT
TGTGCTACCA
TCCGCCGTGCCTTCTCGAGGCAACATCATCGTTGTCGCCCGACAACG
```

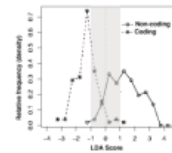
## Two Possible Outcomes

### ORBIT



Id	Predicted to be	Probability Coding	Probability Non-coding	LDA Score
your_seq	Non-coding	0.451	0.549	0.062

### ORBIT



Id	Predicted to be	Probability Coding	Probability Non-coding	LDA Score
your_seq	Coding	0.989	0.011	-1.419

# UTR Generation

	A	B	C	D	E	F	G	H
1								
2		<b>Origin Sequences</b>	<b>Blast Identifier</b>	<b>CDS</b>	<b>Subject</b>	<b>Orbit</b>		
3	1	>AA003459	<a href="#">ref M_824534.1</a>	1.2892	2892..2611	Non-Coding		
4		CGAAATACTGTAAAACGTTCCCTGACGGAGTGCACCCGCCAGATAAACAAATAATGATGTGG						
5		AACAAGGGAGAGAATTTTTTTTAAAAGAGAAAAAAGAGTAACCTTCACTCCTCCAGAC						
6		TTATAGGTGGAACCTCTACGTCCTTATGCGGTTCCCGGCGTANATTTGTTGCTACCA						
7		TCCGCGTGCCTTCTTCGAGGCAACATCATCGTTGTCCCCGACAACGCATTGGAATCT						
8		TCGTCCCTCACAGCAGAGGAACGGCGCGGCGTCAATGTGCATCAGCATCATGTGCATG						
9		CAGTCCTTACCCTCTTTGGTTTGTGGAGGATATAAGTGTGTCGGCTCAACGCCAC						
10		GTGTCTCCGCCCTTGATGACGGGCATCCTTCAAGCATTGTTGT						
11								
12	2	>AA003460	<a href="#">gil71744193ref M_798517.1</a>	1.1131	1..306	Coding		
13		CTAGTGAGGTGCAGGTGCCTCATGTCCGTTTTTCCCCCTGGCCGATTCGCAAAATGTG						
14		GTCTGCGCTCTAATGTGGGTGTGCCTTCGGGCCCGGATGTCACCTTGTGCGCCTGTGCT						
15		CCGTGCTTCTGGGTGCCACTGTGCGGTTCAATTCGTTGTGCGGGAGCCATTTAACTCA						
16		AGACCTCACATCTGCGCTGCGGTTCCGGCGATGTTGAGCCACAGGAGCTGTGTGCCA						
17		TCAGGGAGGATGCGTTGCGGGAGTATTTAAGGATGAAATGGCTGATGAGTTGTCGGATG						
18		TAGAGCGTATGGAGGCCGACGCGCGT ANCTCGTCGAGCAAGCACTATTTTTTTT						
19								
20	3	>AA003461	<a href="#">gil72387428ref M_839046.1</a>	1.2028	1982..2028	Non-Coding		
21		GGGAGGCGCGCAACTTAAGCATTATATTTGATATGCGTGAGTAA TTTACTTACCTGA						
22		AAAGGGTGTACACCGTTACACTTCGACATTTCCCTCCTGCTCCTTCTCTCTTCTCTTC						
23		CCTCCCCGCCCTTCCCTTCCCTTCCCTTCCAGATTGATTTCCCTTCACTATTTCTATTT						
24		TAAGAGAGGCGTAACAGTTTTTGTCTTTTCTCCTGTTTTTTCAGTGATGAGTACTTCGAA						
25		TTCGTACAGTAAACTGCTACTTTGTTGCGTCTTGTTCAGAGGAGGAGTAGGAGGGA						
26		TGTAAAAGAGAAAGAAAGTGGGTCCGGTGTACACTGGAGAGAGAGAGAGAAAAAGGGGAG						
27		GGTACAAAAGGAAAATTATAGAAAAGAAAAGGGGTGGACAACAGCAGCATTAGGGGG						
28								
29		>AA003462	<a href="#">gil74026213ref M_824580.1</a>	1.570	570..319	Non-Coding		
30		AACAACCCACATGATGTTAGGAGAGATTAGTCATTACCCTGATCCGCGCAACATCCCTT						
31		GTTCATTCAACATCCCATAGCTTGAAGTACCTTAATTACAGTCTCGGAGGCAAAATCGT						
32		CAGTTCACCAAACTCATCAACCCAATAAATCTGTGGAAGTGTCTTGTCTCTACTA						
33		GCATAAGCGTCGGCAGCATCAGAACATTAATCTTTCCGCAAGGAGGATTTCTCTCCA						
34		CATCAACATAGCAGAATCGCGTTTCGAAATGTTGAGG AAGAAAAAAGAAAGAGCAAA						
35		CAAAATGACTCGCGCTGGATTTAAGGGTAAGGTGCTCGGCAAGGAAAAGAACTTGCCT						
36		GCTTGAGGCCCGAAGAGGGCGCTGAAGTCG						
37								
38		>AA003463						

# True UTRs

NCBI Nucleotide

Search CoreNucleotide for Trypanosoma brucei

Limits mRNA

- Use All Fields pull-down menu to specify a field.
- Boolean operators AND, OR, NOT must be in upper case.
- If search fields tags are used enclose in square brackets, e.g., rubella [ti].
- More help on using limits is available [here](#).

Limited to:

Organism

exclude STSs  exclude TPA  exclude working draft  exclude patents  exclude all of the above

mRNA

Gene Location

Segmented Sequences

Only from

Modification Date

Modification Date From To

Use the format YYYY/MM/DD; month and day are optional.

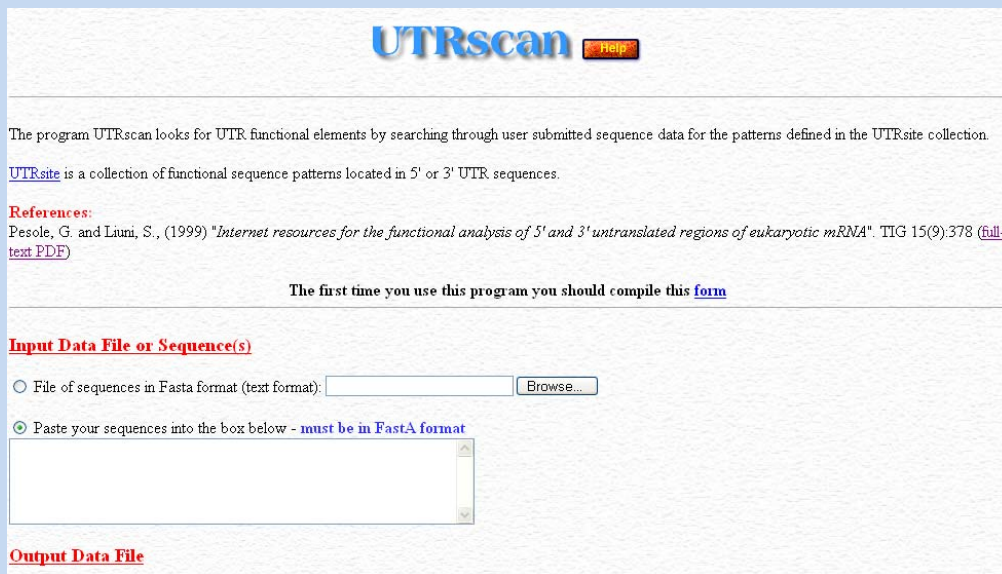
```

source          1..1362
                /organism="Trypanosoma brucei"
                /mol_type="mRNA"
                /isolate="AnTat 1.1"
                /db_xref="taxon:5691"
                /chromosome="9"
gene            1..1362
                /gene="BARP"
5' UTR          1..89
                /gene="BARP"
misc_RNA        1..33
                /gene="BARP"
                /note="trans-spliced leader sequence; mini-exon"
CDS             90..875
                /gene="BARP"
                /note="putative GPI-anchored protein; formerly bloodstream
                alanine-rich protein"
                /codon_start=1
                /product="alanine-rich protein"
                /protein_id="ABB49055.1"
                /db_xref="GI:78711810"
                /translation="MSITFHSWLWLLLTVLCTAGIRGDRVVYDCPEKGVDTSRDGIQAL
                CRAAEQFRGLSQTVTSAVETSATASSKAFEAKVQAEAEVELAESKGLNVTKAKEAAVR
                ATLAAEAAAATAASNVEINAANIAAVPWSQPSSDAGLQKLALCENIDKSLRQLASECSK
                RAENVTAQSLSEALEGLRKLRYNDVYVKEILEREDVEFHKEFMWLQHLREAVHARKQ
                AEDAAAAEANEIAGTNTGPGVSSVASPEGSVLLLMAGLFLSSLL"
3' UTR          876..1362
                /gene="BARP"
polyA_site      1356
                /gene="BARP"
ORIGIN
1  cgctattatt  agaacagttt  ctgtactata  ttgaatccac  tacaagacag  caggcacaag
61  cttcgtatcc  atccaaatta  acaacaatta  tgagcatcac  ttttoatagt  ttatggctac
121  ttctgcagct  gttgtgcact  gcaggtatcc  gtgggtgatg  agtctggtag  gattgtccag

```

# UTRscan

## Data Entry



The screenshot shows the UTRscan web interface. At the top, there is a logo for UTRscan with a 'Help' button. Below the logo, there is a paragraph explaining that the program looks for UTR functional elements by searching through user submitted sequence data for patterns defined in the UTRsite collection. A link to the UTRsite collection is provided. A reference is cited: Pesole, G. and Liuni, S., (1999) "Internet resources for the functional analysis of 5' and 3' untranslated regions of eukaryotic mRNA". TIG 15(9):378 (full-text PDF). A note states: "The first time you use this program you should compile this form". Under the heading "Input Data File or Sequence(s)", there are two options: "File of sequences in Fasta format (text format):" with a text input field and a "Browse..." button, and "Paste your sequences into the box below - must be in FastA format" with a large text area. At the bottom, there is a link for "Output Data File".

## Results

```
-----> UTRscan Results <-----
Processed sequences: 1
-----
Pattern = HISTONE3
-----
Pattern not found
-----
Pattern = IRE
-----
Pattern not found
-----
Pattern = SECIS
-----
Pattern not found
-----
Pattern = APP
-----
Pattern not found
-----
Pattern = CPE
-----
Pattern not found
-----
Pattern = TGE
-----
Pattern not found
-----
Pattern = NANOS
-----
Pattern not found
-----
Pattern = 15-LOX-DICE
-----
Found 13 matches in 1 sequences
EM_OM:M27214 :[2066,2084] :GCCACCCCTCT CCCCC AAG
EM_OM:M27214 :[2085,2103] :CCCTGCCCTCT TCCCC AAG
EM_OM:M27214 :[2104,2122] :CCCTGCCCTCT TTCCCC AAG
EM_OM:M27214 :[2123,2142] :CCCCATCCTCT TTCCCC AAG
EM_OM:M27214 :[2143,2161] :CCCCGCCCTCT TCCCC AAG
EM_OM:M27214 :[2162,2180] :CCCCGCCCTCT TCCCC AAG
EM_OM:M27214 :[2181,2199] :CCCCACCCCTCT TCCCC AAG
```

# PatSearch

## Data Entry

**BIG**  
Bioinformatics and Genomic Group - C.N.R. **PatSearch** [Help](#)

The PatSearch program is a pattern matching tool, that can find a well defined pattern against a given sequence(s) or database (primary or specialized) divisions.

The first time you use this program you should complete the form  
[Registration Form](#)

Results by e-mail to:  **Execute**

**Input**

Sequence Type: DNA  Protein

Enter sequence(s) below in FASTA/EMBL/Genbank/... format

or load it from disk:

or choose Database to search:

**Query Pattern**

Enter a query pattern:

## Submitted

**BIG**  
Bioinformatics and Genomic Group - C.N.R. **PatSearch** [Help](#)

**Results**

Ok, you are enabled to execute the program.  
Ok, your job is running

Thank you for using PatSearch. You will receive the output file at following address:

maria.moutafis@gmail.com

Your job code is : 1178427070

To see your job rank click [here](#)

## Results Received

Google Gmail Calendar Documents Photos Groups Reader mv.services a maria.moutafis@gmail.com | [Settings](#) | [Help](#) | [Sign out](#)

**Compose Mail** **Sanger PhD Programme** - www.sanger.ac.uk - Fully funded 4 year studentships in genomics. Deadline 10th Dec 2006 Sponsored Link

**Inbox (1)**    More actions... [Refresh](#) 1 - 50 of 75 [Older](#)

Select: All, None, Read, Unread, Starred, Unstarred

<input type="checkbox"/>	<input checked="" type="checkbox"/> bigstaff	Results of PatSearch program (code "1178427070") - This is the result from PatSearch program. The execution parameters are: INPUT: "[utr-Region ...	1:38 am
<input type="checkbox"/>	<input checked="" type="checkbox"/> bigstaff	Results of PatSearch program (code "1178420083") - This is the result from PatSearch program. The execution parameters are: INPUT: "[utr-div ...	11:02 pm
<input type="checkbox"/>	<input checked="" type="checkbox"/> bigstaff	Results of PatSearch program (code "1178418027") - This is the result from PatSearch program. The execution parameters are: INPUT: ...User File ...	10:20 pm



# Which To Choose?

- Which is more user-friendly?
- Which gives more accurate results?
- Which is faster?
- Which is more reliable?

I have decided, now it's your turn to choose!

