# Особливості патентного пошуку біологічних послідовностей в заявках на винаходи в галузі молекулярної біотехнології

Краузе Антоніна
Головний експерт
Відділу хіміко - біологічних технологій
31.03.2016

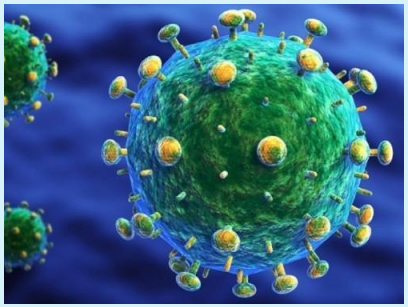
Біологічні послідовності — це первинна структура біологічних макромолекул.

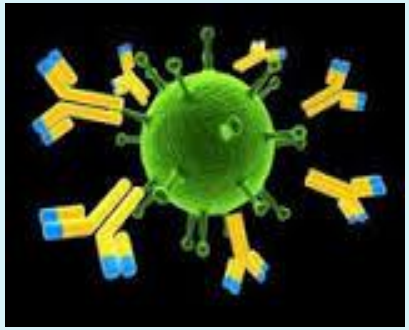
До біологічних макромолекул належать зокрема білки та ДНК / РНК.

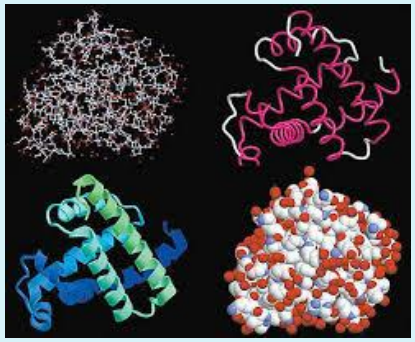
# Галузі застосування:

- Медицина
- Фармацевтика
- Харчова промисловість
- Сільське господарство
- Біоенергетика

# Об’єктами яких є:







## Вимоги щодо складання формули та опису винаходу:

Відповідно до п. 11.3.1 Правил складання у формулу винаходу, яка характеризує індивідуальну хімічну сполуку з визначеною структурою, включають її структурну формулу.

Відповідно до п. 11.3.2 Правил складання для об’єктів генетичної інженерії до формули винаходу включають їх визначені амінокислотні та нуклеотидні послідовності.

Відповідно до п. 11.2.3 Правил складання амінокислотні та нуклеотидні послідовності індивідуальної хімічної сполуки, що є об’єктом генетичної інженерії, які зазначають в формулі винаходу, мають бути представленими та розкритими в описі винаходу.

# Стандарт ВОІВ ST.25

Цілями зазначеного стандарту є:

- полегшити пошук даних про послідовності
- підвищити достовірність і якість представлення послідовностей для спрощення їх поширення серед заявників, широкого загалу та експертів,
- полегшити обмін даними про послідовності в електронній формі, а також введення відповідних даних в комп’ютеризовані бази даних
- дозволити заявнику складати перелік послідовностей в уніфікованій формі в заявці на винахід, який був би прийнятним як для міжнародних, так і для національних або регіональних процедур;

# **CLUSTAL OMEGA (Clus**ter **al**ignment **)**

http://www.ebi.ac.uk/Tools/msa/clustalo

Одна з самих широко використовуваних комп’ютерних програм для множинного вирівнювання (порівняння) нуклеотидних і амінокислотних послідовностей.

CLUSTAL забезпечує можливість:

- Парного вирівнювання послідовностей, що аналізуються
- Побудови їх філогенетичного дерева (guidetree) та
- Множинного вирівнювання послідовностей.

Параметри програми визначають її роботу на трьох етапах побудови вирівнювання.

# Приклад застосування CLUSTAL OMEGA

1. Поліпептид X, що містить SEQ ID NO: 1 та його варіанти, що мають 90 % ідентичності.
2. Поліпептид X за п.1, що містить SEQ ID NOs: 2 або 3.

# Clustal Omega

http://www.ebi.ac.uk/Tools/msa/clustalo

Input form | Web services | Help & Documentation | Share | Feedback

Tools > Multiple Sequence Alignment > Clustal Omega

## Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools.

STEP 1 - Enter your input sequences

Enter or paste a set of PROTEIN sequences in any supported format:

```
>IMGA|Medtr6g025160.1
MNDTGMINFAFVCLLNGDRTINASVLGCWKLYLECGAALSLSAKMRFAYRPFQDIF
 >IMGA|Medtr6g023700.1
MASLGSMNIVTLTFCVIILTTCNHQAHASSRVFLNKKNDKSPIQGLCASSVTIHGFK
```

Or, upload a file: Выберите файл Файл не выбран

STEP 2 - Set your parameters

OUTPUT FORMAT Clustal w/o numbers

*The default settings will fulfill the needs of most users and, for that reason, are not visible.*

More options... *(Click here, if you want to view or change the default settings.)*

STEP 3 - Submit your job

Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

Submit

# Результати порівняння №1

```
#
#
#  Percent Identity  Matrix - created by Clustal2.1
#
#

    1: IMGA|Medtr6g025160.1  100.00   99.73
    2: I|Medtr6g025160.1      99.73  100.00
```

# Результати порівняння №2

```
#
#
#  Percent Identity  Matrix - created by Clustal2.1
#
#

   1: IMGA|Medtr6g025160.1  100.00   13.33
   2: IMGA|Medtr6g023700.1   13.33  100.00
```

# Програми для пошуку послідовностей

ALIGN, AMAS, BLAST, BLAT, CLUSTAL, DiAlign, FASTA, HI, HMMER, MAP, MGA, OWEN, PipMarker, MultiPipMarker, REGISTRY та інші.

# REGISTRY

http://www.stn-international.com/stn_biosequence_searching_mfs.html

STN®

**Найбільш точними безкоштовними є програми:**

**BLAST**
http://blast.ncbi.nlm.nih.gov/Blast.cgi та

**FASTA**
http://www.ebi.ac.uk/Tools/msa/fasta

# BLAST

(**B**asic **L**ocal **A**lignment **S**earch **T**ool )

http://blast.ncbi.nlm.nih.gov/Blast.cgi

Основний засіб пошуку, заснований на локальних вирівнюваннях.

BLAST – це інструмент, який дозволяє швидко знайти подібні послідовності до шуканих послідовностей білків або нуклеїнових кислот і забезпечує можливість:

* Передбачення функції білка
* Передбачення просторової структури білка
* Передбачення доменної структури
* Ідентифікація гомологів (схожих білків) в інших організмах

## BLAST Assembled Genomes

Find Genomic BLAST pages:

Enter organism name or id--completions will be suggested GO

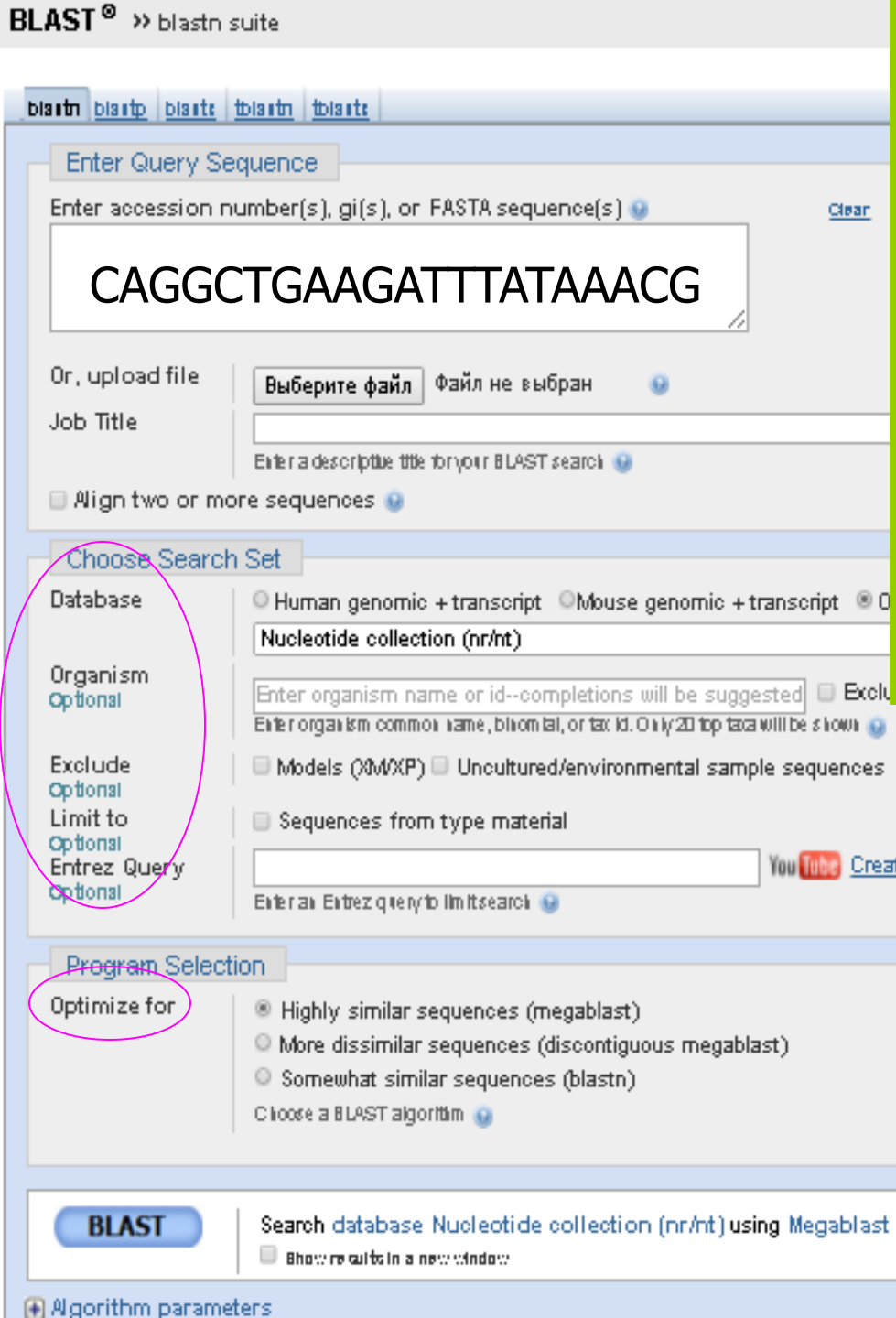
- Human
- Mouse
- Rat
- Cow
- Pig
- Dog
- Rabbit
- Chimp
- Guinea pig
- Fruit fly
- Honey bee
- Chicken
- Zebrafish
- Clawed frog
- *Arabidopsis*
- Rice
- Yeast
- Microbes

## Basic BLAST

Choose a BLAST program to run.

| | |
|---|---|
| nucleotide blast | пошук нуклеотидних послідовностей по нуклеотидним базам |
| protein blast | пошук амінокислотних послідовностей по амінокислотним базам |
| blastx | пошук трансльованих нуклеотидних послідовностей по амінокислотним базам |
| tblastn | пошук амінокислотних послідовностей по базах трансльованих нуклеотидів |
| tblastx | пошук трансльованих нуклеотидних послідовностей по базах трансльованих нуклеотидів |

Спосіб визначення ДНК культури бактерії *E. coli* у харчових продуктах методом ПЛР, який відрізняється тим, що для визначення ДНК культури бактерії *E. coli* застосовують пару олігонуклеотидних праймерів до гену recN DNA repair protein: прямий праймер recN F: 5´-CAGGCTGAAGATTTATAAACG-3´ та зворотній праймер recN R:GGGCCCTTTAATTTAAAGGGCC-3´ для ампліфікації 529 bp фрагмента ДНК культури бактерії *E. coli*.



**Sequences producing significant alignments:**

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

| Description | Max score | Total score | Query cover | E value | Ident | Accession |
|---|---|---|---|---|---|---|
| Drosophila busckii chromosome 3R sequence | 36.2 | 92.7 | 85% | 2.9 | 100% | CP012526.1 |
| Ovis canadensis canadensis isolate 43U chromosome 8 sequence | 36.2 | 352 | 95% | 2.9 | 100% | CP011893.1 |
| Parascaris equorum genome assembly P. equorum, scaffold PEQ_scaffold0000133 | 36.2 | 36.2 | 85% | 2.9 | 100% | LM462891.1 |
| Echinostoma caproni genome assembly E. caproni_Egypt, scaffold ECPE_scaffold0000462 | 36.2 | 36.2 | 85% | 2.9 | 100% | LL233557.1 |
| Treponema putidum strain OMZ 758, complete genome | 36.2 | 157 | 95% | 2.9 | 100% | CP009228.1 |
| Neotyphodium coenophialum dimethylallyltryptophan N-methyltransferase (easF), agroclavine dehydrogenase (easG), chanoclavine aldehyde oxid | 36.2 | 36.2 | 85% | 2.9 | 100% | KC989607.1 |
| PREDICTED: Coturnix japonica vascular endothelial growth factor A (VEGFA), transcript variant X4, misc_RNA | 34.2 | 34.2 | 80% | 12 | 100% | XR_001553983.1 |
| Bacterium L21-Spi-D4, complete genome | 34.2 | 34.2 | 80% | 12 | 100% | CP013118.1 |
| Ovis canadensis canadensis isolate 43U chromosome 25 sequence | 34.2 | 181 | 95% | 12 | 100% | CP011910.1 |
| Ovis canadensis canadensis isolate 43U chromosome 18 sequence | 34.2 | 149 | 80% | 12 | 100% | CP011903.1 |
| Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold27 | 34.2 | 179 | 95% | 12 | 100% | LK391419.1 |
| PREDICTED: Fundulus heteroclitus pleckstrin homology domain-containing family G member 4B-like (LOC105936417), transcript variant X1, mRN | 34.2 | 34.2 | 80% | 12 | 100% | XM_012877270.1 |
| TPA: Oryzias latipes strain Hd-rR, complete genome assembly, chromosome 11 | 34.2 | 92.7 | 80% | 12 | 100% | HG313989.1 |
| Cellulophaga baltica 18, complete genome | 34.2 | 62.4 | 80% | 12 | 100% | CP009976.1 |
| Spirometra erinaceieuropaei genome assembly S. erinaceieuropaei, scaffold SPER_scaffold0230958 | 34.2 | 34.2 | 80% | 12 | 100% | LN358867.1 |
| Spirometra erinaceieuropaei genome assembly S. erinaceieuropaei, scaffold SPER_scaffold0149618 | 34.2 | 34.2 | 80% | 12 | 100% | LN168366.1 |

# FASTA

http://www.ebi.ac.uk/Tools/msa/fasta

FASTA це програма для пошуку нуклеотидних або білкових послідовностей у вибраних базах даних.

Можливості FASTA:

- Знаходить регіони локальної або глобальної подібностей між білками або ДНК-послідовностями.
- Порівнює послідовність білка з іншою білковою послідовністю або з цілою білковою базою даних, або послідовність ДНК з іншою послідовністю ДНК або з ДНК-бібліотекою.

Як програма BLAST, так і програма FASTA може бути використаною для виведення функціональних і еволюційних взаємовідносин між послідовностями.

# FASTA

Protein | Nucleotide | Genomes | Proteomes | Whole Genome Shotgun | Web services

Tools > Sequence Similarity Searching > FASTA

## Protein Similarity Search

This tool provides sequence similarity searching against protein databases using the FAST
FASTY translate a DNA query. Optimal searches are available with SSEARCH (local), GGSEA

STEP 1 - Select your databases

PROTEIN DATABASES

1 Databank Selected

- [x] UniProt Knowledgebase
- [ ] UniProtKB/Swiss-Prot
- [ ] UniProtKB/Swiss-Prot isoforms
- [ ] UniProtKB/TrEMBL
- UniProtKB Taxonomic Subsets
- UniProt Clusters
- Patents
- Structure
- Other Protein Databases

STEP 1 - Select your databases

PROTEIN DATABASES

1 Databank Selected  X Clear Selection

- [x] UniProt Knowledgebase
- [ ] UniProtKB/Swiss-Prot
- [ ] UniProtKB/Swiss-Prot isoforms
- [ ] UniProtKB/TrEMBL
- **UniProtKB Taxonomic Subsets**
- **UniProt Clusters**
- **Patents**
- **Structure**
- **Other Protein Databases**

STEP 2 - Enter your input sequence

Enter or paste a PROTEIN sequence in any supported format:

```
>IMGA|Pedtr256965.1
MNDTGMINFAFVCLLNGDRTINASVLGCWKLYLECGAALSLSAKMRFAYRPFQDIF
```

or Upload a file: Выберите файл Файл не выбран

STEP 3 - Set your parameters

PROGRAM

FASTA

| Align. | DB:ID | Source | Length | Score (Bits) | Identities % | Positives % | E() |
|---|---|---|---|---|---|---|---|
| 50 | TR:Q6UVN9_LOXCY | Cytochrome b OS=Loxodonta cyclotis GN=CYTB PE=3 SV=1<br>*Cross-references and related information in:*<br>Small molecules · Nucleotide sequences · Samples & ontologies · Protein families · Literature · Protein sequences | 378 | 398.1 | 96.1 | 98.2 | 9.8E-108 |
| 49 | TR:Q6UVN4_LOXCY | Cytochrome b OS=Loxodonta cyclotis PE=3 SV=1<br>*Cross-references and related information in:*<br>Small molecules · Nucleotide sequences · Samples & ontologies · Protein families · Literature · Protein sequences | 378 | 398.1 | 96.1 | 98.2 | 9.8E-108 |
| 48 | TR:Q4VQE0_LOXAF | Cytochrome b OS=Loxodonta africana GN=cytb PE=3 SV=1<br>*Cross-references and related information in:*<br>Small molecules · Nucleotide sequences · Samples & ontologies · Protein families · Protein sequences | 378 | 398.1 | 96.1 | 98.2 | 9.8E-108 |
| 47 | TR:L7Q7V0_9EUTH | Cytochrome b OS=Loxodonta cyclotis x africana GN=CYTB PE=3 SV=1<br>*Cross-references and related information in:*<br>Small molecules · Nucleotide sequences · Samples & ontologies · Protein families · Literature · Protein sequences | 378 | 398.1 | 96.1 | 98.2 | 9.8E-108 |

Державне підприємство

**УКРАЇНСЬКИЙ ІНСТИТУТ ІНТЕЛЕКТУАЛЬНОЇ ВЛАСНОСТІ**

**(УКРПАТЕНТ)**

# Дякую за увагу