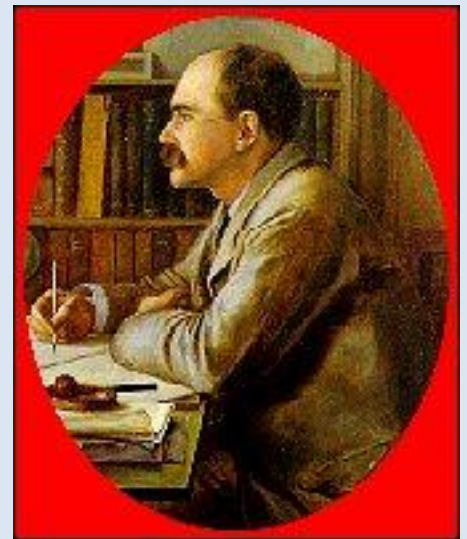


Функции белков и генов

*I keep six honest serving-men
(They taught me all I knew);
Their names are **What** , and **Why**, and **When**,
And **How**, and **Where**, and **Who**.*

Rudyard Kipling



План

- Что важно знать функции данного белка и где это найти?
- Как найти все белки с определенной функцией?
- Как бороться с синонимами?
- База данных GO: структура; использование
- Что такое “Protein evidence”?

Задача I: узнать функцию белка

Пример: что за белок Q71SG9_KLEPN

- Название (кто?)
 - Uniprot:
 - Cytosine-specific methyltransferase
 - ген kpn2k1M
- Откуда?
 - Klebsiella pneumoniae
 - плазмида

Q71SG9_KLEPN

- Что делает?

- Pubmed:

- статей нет

- ссылка на [Gene DB \(NCBI\)](#) (часть проекта RefSeq):

- type II DNA-methyltransferase

- Кто – ближайший изученный гомолог?

- BLAST

- MTS2_SHISO, Identity 100% по всей длине (!?)

- M.Ssoll, *Shigella sonnei*, плаزمида

Description and origin of the Protein

| | | |
|-----------------|--|--|
| Description | Recommended | Full=Modification methylase SsoII; Short=M.SsoII; EC= 2.1.1.37 ; |
| | Synonym | Full=Cytosine-specific methyltransferase SsoII; |
| Gene name(s) | ssoIIM | |
| Organism source | Shigella sonnei. | |
| Taxonomy | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella | |
| Organelle | Plasmid P4. | |

M.SsoII

Database cross-references

| | |
|--------------------|--|
| EMBL | M86545 ; AAA98279.1 ; -; Genomic_DNA. |
| PIR | JT0744 ; JT0744. |
| ProteinModelPortal | P34879; -. |
| SMR | P34879 ; 7-53. |
| REBASE | 3509; M.SsoII . |
| BRENDA | 2.1.1.37 ; 74546. |
| GO | GO:0003886 ; F:DNA (cytosine-5-)-methyltransferase activity; IEA:EC. GO:0043565 ; F:sequence-specific DNA binding; IEA:InterPro. GO:0009307 ; P:DNA restriction-modification system; IEA:UniProtKB-KW. GO:0045449 ; P:regulation of transcription; IEA:UniProtKB-KW. GO:0006350 ; P:transcription; IEA:UniProtKB-KW. |

| | |
|----------|---|
| InterPro | IPR018117 ; C5_DNA_meth_AS. IPR001525 ; C5_MeTfrase. IPR001387 ; HTH_3. IPR010982 ; Lambda_DNA-bd. |
| PANTHER | PTHR10629 ; C5_DNA_meth; 1. |
| Pfam | PF00145 ; DNA_methylase ; 1. PF01381 ; HTH_3 ; 1. |
| PRINTS | PR00105 ; C5METTRFRASE. |
| SMART | SM00530 ; HTH_XRE; 1. |
| SUPFAM | SSF47413 ; Lambda_like_DNA; 1. |
| TIGRFAMs | TIGR00675 ; dcm; 1. |
| PROSITE | PS00094 ; C5_MTASE_1; 1. PS00095 ; C5_MTASE_2; 1. PS50943 ; HTH_CROC1; 1. |

Protein Existence

3: Inferred from homology;

M.SsoII

- Что делает?

- Pubmed:

- есть статьи; нет обзоров

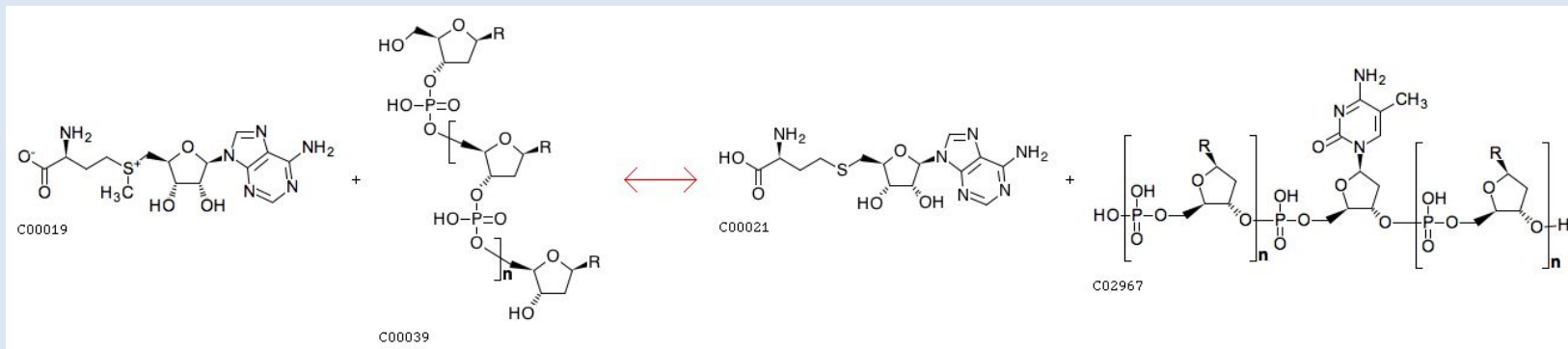
UniProt ссылается на разные БД

- GO:0003886 (F)

- Molecular Function:

Catalysis of the reaction: S-adenosyl-L-methionine + DNA containing cytosine = S-adenosyl-L-homocysteine + DNA containing 5-methylcytosine

- Переносит метильную группу с кофактора на цитозин ДНК
- приведен список из 20 синонимов



M.SsoII

Uniprot ссылается на разные БД (продолжение)

– GO:0043565: (F)

- Name: sequence-specific DNA binding
- Molecular Function: Interacting selectively and non-covalently with DNA of a specific nucleotide composition, e.g. GC-rich DNA binding, or with a specific sequence motif or type of DNA e.g. promotor binding or rDNA binding.

– EC 2.1.1.37

- Catalysis of the reaction: S-adenosyl-L-methionine + DNA containing cytosine = S-adenosyl-L-homocysteine + DNA containing 5-methylcytosine
- **COMMENT** * See the REBASE database for a complete list of these enzymes: <http://rebase.neb.com/rebase/>
- приведен список из 20 синонимов

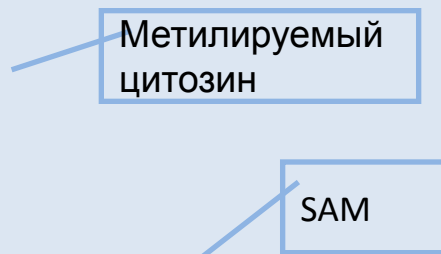
Мы описали
МОЛЕКУЛЯРНУЮ ФУНКЦИЮ
белка M.SsoII (и гомолога из
K.pn.)

M.SsoII

- Каков механизм работы?
 - кофактор SAM (S-adenosyl-L-methionine)
 - на какой атом ДНК переносится – углерод в 5-м положении цитозина
 - как фермент, кофактор и субстрат расположены в пространстве? 3D структуры нет; есть модель, построенная по гомологам доменов с известной структурой. Цитозин выворачивается, см. рис.
 - как устроен активный центр? – известны консервативные мотивы и остатки
 - как происходит узнавание последовательности ДНК? – известна узнаваемая последовательность – CCNGG - и домен, отвечающий за узнавание (TRD)
 -

Методы:

- Пришивка (cross-linking)
- Мутагенез
-
- моделирование структуры
- консервативность последовательностей
- множество всего



Описали
МЕХАНИЗМ
действия M.SsolI

M.SsoII: зачем нужно метилировать ДНК?

- Зачем?

- GO:0009307: (P)

- Name: DNA restriction-modification system
 - Biological Process: Any process that protect an organism from invading foreign DNA by methylation of self DNA at specific sequences and nucleolytic cleavage of unmethylated foreign DNA.
 - Система рестрикции-модификации: любой процесс, защищающий организм от вторжения чужеродной ДНК посредством метилирования специфических последовательностей и гидролиза неметилированной чужеродной ДНК
Чужеродная ДНК – ДНК бактериофагов, плазмид и др.

- Какие гены включает система РМ?

- Rebase:

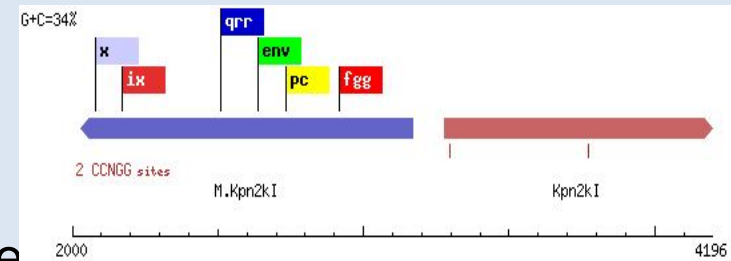
- M.SsoII:

- » M.SsoII – ДНК метилтрансфераза

- » SsoII – эндонуклеаза рестрикции

- SsoII - вносит двухцепочечный разрыв ДНК в неметилированных сайтах CCNCGG

M.SsoII метилирует цитозин только в сайтах той же последовательности CCNCGG; метилируя хозяйскую ДНК M.SsoII предотвращает фрагментацию собственного генома



Мы описали
БИОЛОГИЧЕСКИЙ ПРОЦЕСС
И
систему генов
его осуществляющих

M.Kpn2kl и M.SsoII


- В каком организме?
 - *Klebsiella pneumoniae*
 - плазида
 - *Shigella sonnei*, бактерия из семейства *Enterobacteriaceae*; патоген
 - плазида
 - Сходство 100%; как объяснить?

M.SsoII

• Какие еще функции известны?

– Pfam:

- два домена



| Source | Domain | Start^ | End |
|--------|-------------------------------|--------|-----|
| Pfam A | HTH_3 | 9 | 65 |
| Pfam A | DNA_methylase | 72 | 375 |

– GO:0045449: (P)

- regulation of transcription; IEA:UniProtKB-KW.
- регулятор транскрипции:
- связывается с сайтами NNNNNN;
- такие сайты расположены между генами M.SsoII и SsoII
- репрессор гена M.SsoII и активатор SsoII

– для чего?

- для “размножения” плазмиды: проникновения в нового хозяина

– механизм регуляции?

- имеет N-концевой HTH домен, гомологичный известным доменам транскрипционных факторов
- димеризуется при связывании с ДНК – еще одна функция: способность димеризации

M.SsoII

- Где локализуется?
 - в цитоплазме хозяйской бактериальной клетки
 - а точнее?
- Когда экспрессируется?
 - небольшая поддерживается постоянно; авторегуляция
 - при проникновении в нового хозяина увеличивается

ЛОКАЛИЗАЦИЯ и
ЭКСПРЕССИЯ

Итог: M.SsoII (и M.Kpn2kI)

- **Функции:**
 - метилирование цитозина в сайтах CCN₂GG
 - связывание с ДНК в сайтах определенной последовательности
 - гомодимеризация при связывании с ДНК
- **Механизм:**
 - узнавание последовательности – TRD домен
 - кофактор SAM
 - мотивы I - X
 - выворачивание основания
 - ход реакции
- **Биологический процесс:**
 - защита от чужеродной ДНК
 - регуляция транскрипции системы PM SsoII
- **Система** рестрикции-модификации II типа; состоит из двух генов на трансмиссивной плазмиде
- **Локализация:**
 - цитоплазма хозяйской клетки
- **Экспрессия**
 - негативная авторегуляция

Базы данных

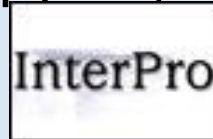
- Uniprot
- Pubmed
- GO – gene ontology
- EC - enzyme classification
- Pfam, Interpro, а также Prosite, Panther, Prints, TIGERFam, SMART, Supfam – семейства белков/доменов
- Brenda - The Comprehensive Enzyme Information System
- Rebase – специализированная БД

Где искать описание функции

- Краткое описание функций одного белка и ссылки на другие ресурсы см.



- Краткое описание функций семейств белков и доменов см. в



- Подробное описание функций генов и их продуктов см в энциклопедиях, таких как

BIOCYC



- Подробное описание отдельных классов функций и соответствующих белков см. в специализированных БД, таких как ENZYME, Rebase

BRENDA

TCDB

...

Как нам узнать функцию интересующего нас и б.м. известного науке белка или гена?

- +/+++ Читайте оригинальные статьи хороших авторов в хороших журналах!
Ищите в PubMed
- -/++ Читайте аннотации записи Uniprot
Используйте SRS
- -/+ Читайте аннотации записей Pfam и InterPro, содержащие описания семейств доменов, к которым принадлежат домены белка
- -/+ Ищите ваш белок в специализированных БД (БД и энциклопедии, в которых подробно описаны функции генов и их продуктов : KEGG, BIOSYS, ENZYME, TC-DB, REACTOME.....)
- Ищите, предположительно, гомологичные белки самостоятельно. Используйте BLAST, psiBLAST, профили и паттерны

Задача II: найти белки с той же функцией

- Та же молекулярная функция?
 - в аннотации записи Uniprot
 - нужные термины GO, помеченные **F** (molecular function)
 - нужные коды EC
- Тот же механизм?
 - функция + гомология
- Тот же биологический процесс?
 - нужные термины GO, помеченные **P** (biological process)
- Та же система генов?
 - БД систем: KEGG, SEED, ...
- Та же локализация?
 - нужные термины GO, помеченные **C** (cellular component)

Что бывает

- Верно ли, что у белки всегда есть “главная” функция?
 - нет; цитохром с: окислительное фосфорилирование
индукция апоптоза
- Верно ли, что гомологичные белки имеют ту же основную функцию?
 - часто, но не всегда!
- Верно ли, что белки с той же функцией гомологичны?
 - часто, но не всегда! (карбоангидраза)
- Верно ли, что белки с одинаковой функцией участвуют в тех же биологических процессах?
 - не обязательно! (ДНК-метилтрансферазы)
- Верно ли, что один и тот же биологический процесс реализуется сходными системами генов?
 - не обязательно! (лекарственная устойчивость)

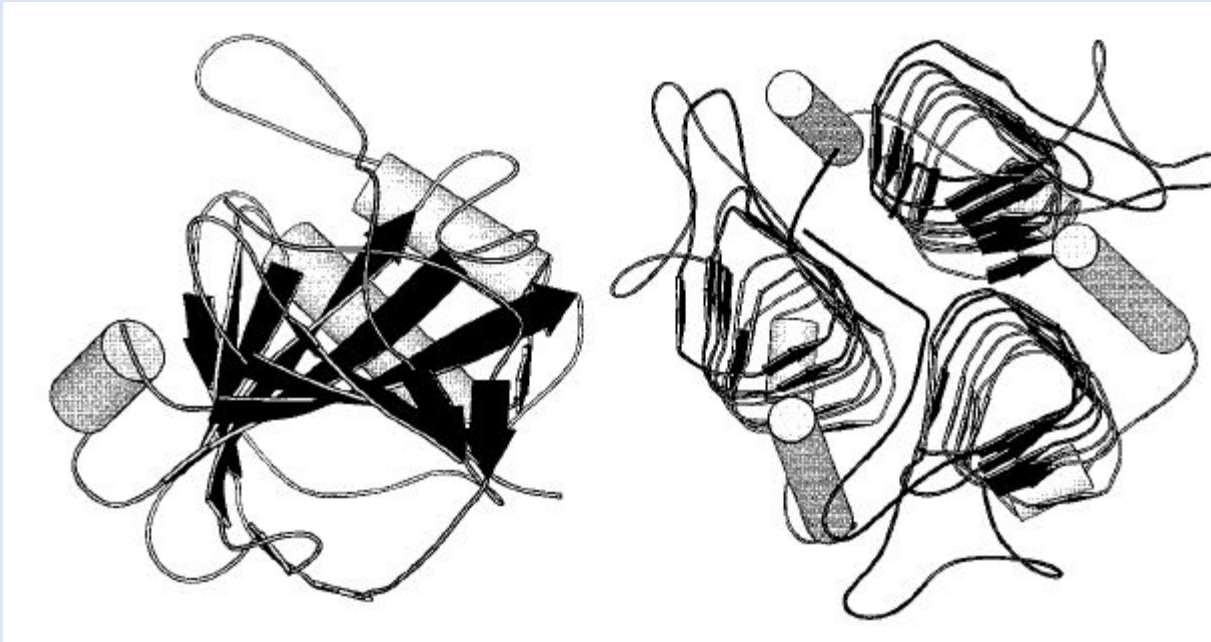


Figure 1. Specific example of convergent and divergent evolution. Top, an example of convergent evolution, showing structures of two carbonic anhydrases with the same enzymatic function (**EC number 4.2.1.1**), but with different folds. The Figure was drawn with Molscript (Kraulis, 1991) from **1THJ** (left-handed beta helix) and **1DMX** (**right beta sheet**).

Hedi Hegyi and Mark Gerstein
J. Mol. Biol. (1999) 288, 147±164

Пробуем выделить классы белков по функции

- Молекулярные машины – рибосома [&&&]
- Ферменты - РНК-зависимая РНК полимераза [рабочие на производстве]
- Регуляторные белки – регулируют биологические процессы, например, активность ферментов – TetR [бюрократия]
- Хранение и транспорт (ионов, маленьких молекул) – гемоглобин [складские рабочие]
- Транспорт через мембраны – TetA [таможенники]
- Секреторные, взаимодействие с другими клетками – инсулин [командировочные?]
- Структурные [атланты 😊]
- Сигнальные []
- Рецепторы []
- Мотор []

Проблема неоднозначности

терминологии

UniProtKB/Swiss-Prot entry P52303 [AP1B1_RAT] AP-1 complex subunit beta-1 - Mozilla Firefox

Файл Правка Вид Журнал Закладки Инструменты Справка

http://cn.expasy.org/uniprot/P52303

Начальная страница Последние заголовки FBB NCBI HomePage Mrs ExPASy China SRS

note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.

Entry information

| | |
|-----------------------------------|--------------------------------------|
| Entry name | AP1B1_RAT |
| Primary accession number | P52303 |
| Secondary accession numbers | None |
| Integrated into Swiss-Prot on | October 1, 1996 |
| Sequence was last modified on | October 1, 1996 (Sequence version 1) |
| Annotations were last modified on | February 5, 2008 (Entry version 62) |

Name and origin of the protein

| | |
|-------------------|--|
| Protein name | AP-1 complex subunit beta-1 |
| Synonyms | Adapter-related protein complex 1 subunit beta-1 Beta1-adaptin Beta-adaptin 1 Adaptor protein complex AP-1 subunit beta-1 Golgi adaptor HA1/AP1 adaptin beta subunit Clathrin assembly protein complex 1 beta large chain |
| Gene name | Name: Ap1b1 Synonyms: Adtb1 |
| From | Rattus norvegicus (Rat) [TaxID: 10116] |
| Taxonomy | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus. |
| Protein existence | 1: Evidence at protein level; |

Готово



Питер Брейгель старший
"Вавилонская башня" 1563
Музей истории искусств, Вена



Проект GO (Gene Ontology)

- Цель: Создание унифицированной терминологии для аннотации генов
- БД GO Включает три независимых словаря
 - Молекулярные функции (molecular Function)
(Как? С чем?) Например, *carbohydrate binding* или *ATPase activity*
 - Биологические процессы (biological Process)
(Зачем?) Например, *митоз* или *биосинтез пуринов*
 - Клеточные компоненты (cellular Component)
(Где?) Например, *ядро* или *холофермент РНК-полимераза II*
- В консорциум GO входит EBI (БД Uniprot, Interpro) и много других организаций
 - GOA – проект описания записей Uniprot терминами GO
 - 94% записей Swissprot и 65% Trembl имеют хотя бы один термин GO

Запись GO называется “термин GO”

- Термины имеют определение и перечень синонимов.
- Термины в пределах одной онтологии (словаря) связаны отношениями “is_a”, “is_part_of”
- Термины имеют стандартные идентификаторы:
GO:0000093 (пример)

tricarboxylic acid cycle

Accession: GO:0006099

Ontology: biological_process

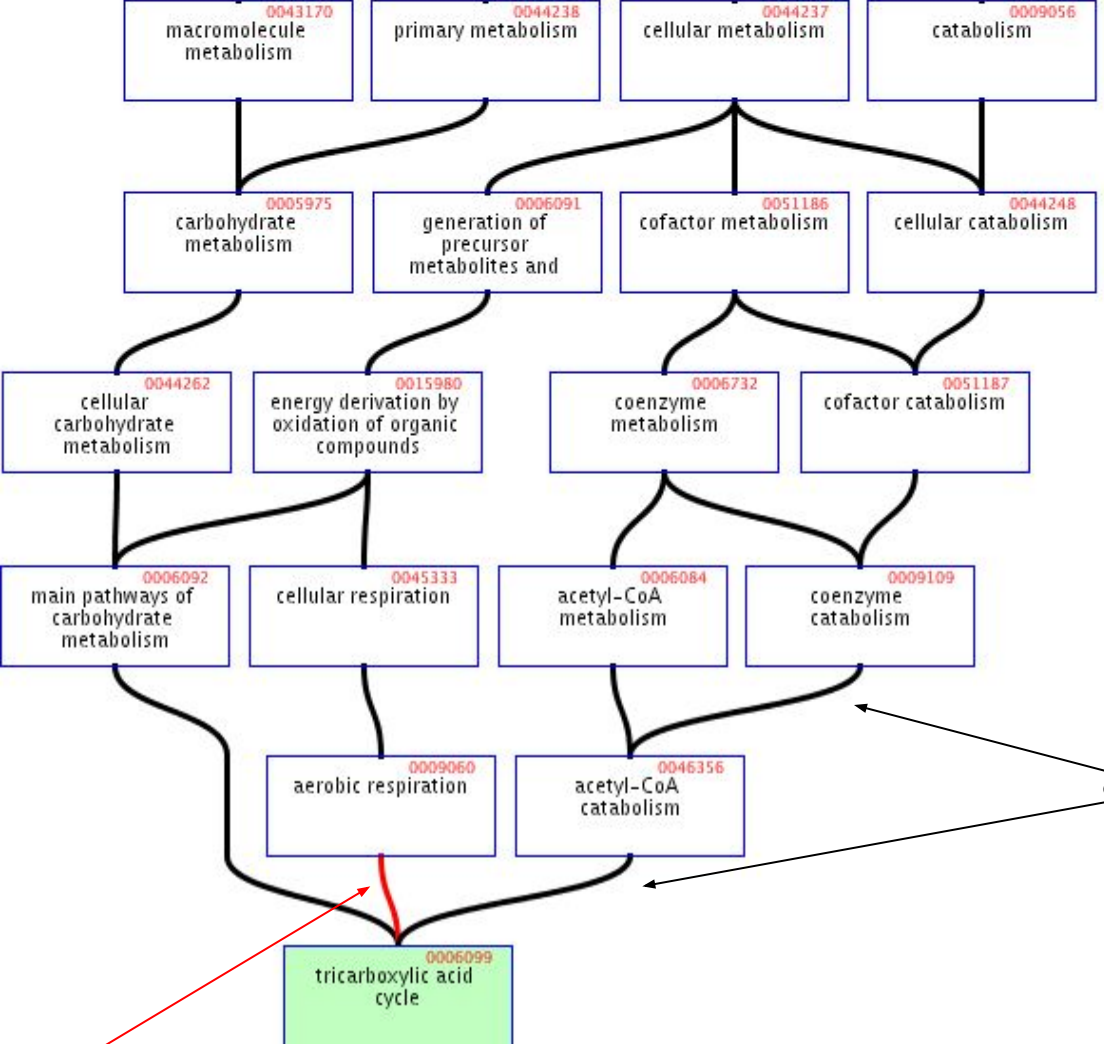
Synonyms:

exact: citric acid cycle

exact: Krebs cycle

exact: TCA cycle

Definition: A nearly universal metabolic pathway in which the acetyl group of acetyl coenzyme A is effectively oxidized to two CO₂ and four pairs of electrons are transferred to coenzymes. The acetyl group combines with oxaloacetate to form citrate, which undergoes successive transformations to isocitrate, 2-oxoglutarate, succinyl-CoA, succinate, fumarate, malate, and oxaloacetate again, thus completing the cycle. In eukaryotes the tricarboxylic acid is confined to the mitochondria. See also glyoxylate cycle.

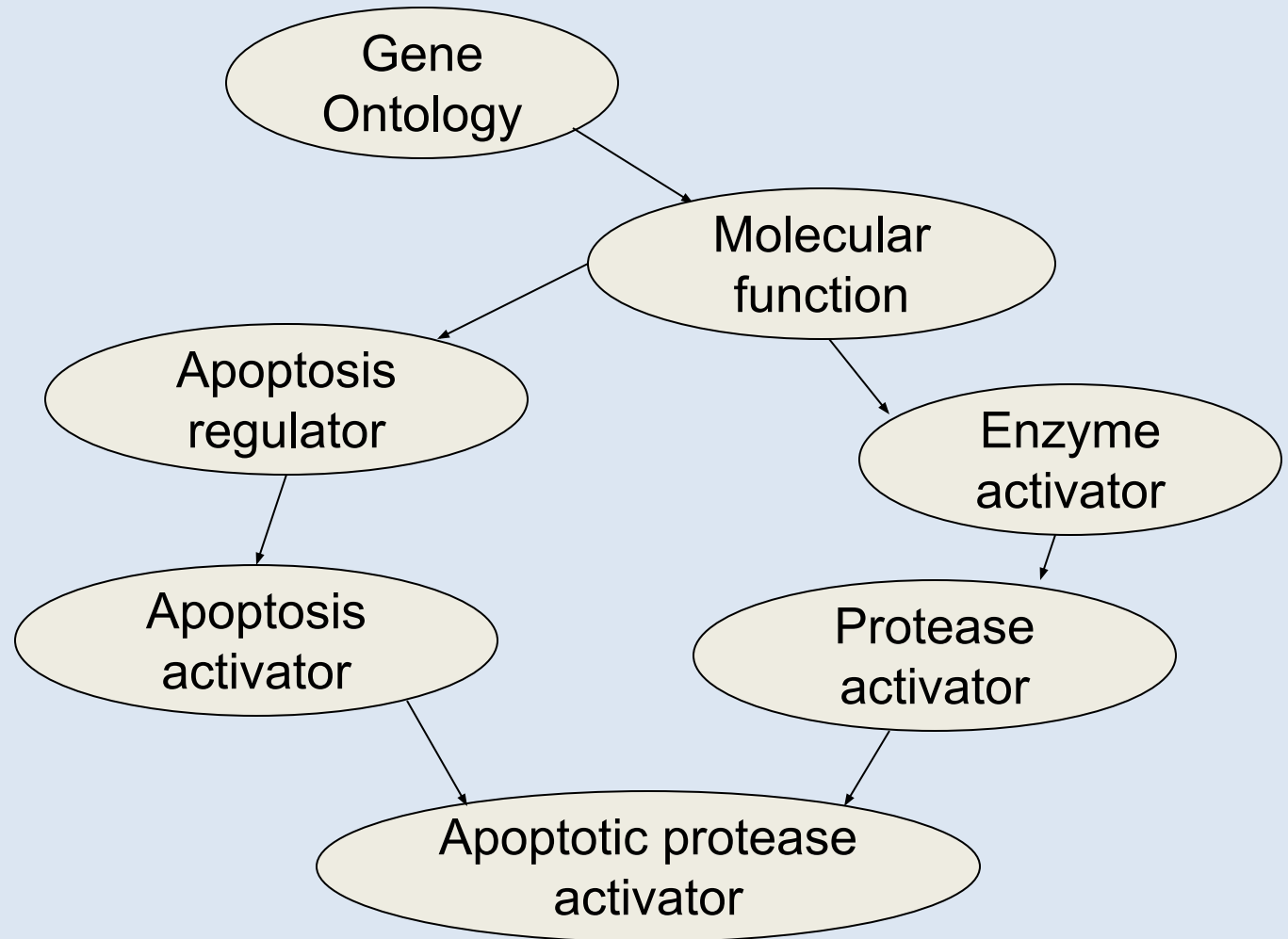


Directed acyclic graph DAG — ориентированный ациклический граф

• отношение "_is_a":
"A is B" означает, что
A — частный случай B;

- отношение "is_part_of":
"A is part of B" означает, что A — часть B, но
B не обязательно содержит A.

The Gene Ontology database



Аннотация GO записей Uniprot

- Два принципа:
 - каждая аннотация должна ссылаться на ИСТОЧНИК:
 - литературная ссылка
 - другая база данных
 - компьютерное предсказание
 -
 - аннотация должна указывать на достоверность применимости термина GO к данному белку в источнике (kind of evidence)

Аннотация GO для MTS2_SHISO (UniProt)

GO

GO:0003886; F:DNA (cytosine-5-)-methyltransferase activity; IEA:EC.

GO:0043565; F:sequence-specific DNA binding; IEA:InterPro.

GO:0009307; P:DNA restriction-modification system; IEA:UniProtKB-KW.

GO:0045449; P:regulation of transcription; IEA:UniProtKB-KW.

GO:0006350; P:transcription; IEA:UniProtKB-KW.

Evidence Codes

- **IDA** Inferred from **D**irect **A**ssay
- **TAS** **T**raceable **A**uthor **S**tatement
- **IMP** Inferred from Mutant Phenotype
- **IGI** Inferred from Genetic Interaction
- **IPI** Inferred from Physical Interaction
- **RCA** Inferred from Reviewed Computational Analysis
- **ISS** Inferred from Sequence Similarity
- **IEP** Inferred from Expression Pattern
- **NAS** Non-traceable Author Statement
- **IEA** Inferred from Electronic Annotation
- **IC** Inferred by Curator

- **ND** No biological Data available

Предостережение:

- GO не является номенклатурой генов или их продуктов (белков). Словари описывают биологические феномены (например, программируемую клеточную смерть), а не конкретные биологические объекты

Есть и другие онтологии, например,



[Home Page](#)

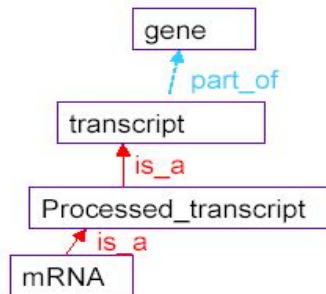
The Sequence Ontology Project

Summary

The Sequence Ontology is a set of terms used to describe features on a nucleotide or protein sequence. It encompasses both "raw" features, such as nucleotide similarity hits, and interpretations, such as genes. It is also intended to have a rich set of attributes, such as "dicistronic" gene and "pseudogene."

The is_a relationship

- The is_a relation is like inheritance.
- Children terms inherit the properties and relationships of the parent term.



mRNA inherits the attributes of *transcript*
Therefore *mRNA* sequence is part of the *gene* sequence

exon,
promoter, binding_site,
non_canonical_splice_site,
stop_codon. pseudogene

Резюме

- Функциональная аннотация геномов — задача биоинформатики
- Существуют энциклопедии, где можно узнать о функциях генов и их продуктов, например, **BioCyc**.
- Полное описание функции — это ответы на вопросы "где?", "зачем?", "как?", "с чем?".
- **GO** — перспективный подход к
 - разработке общего языка (решение проблема синонимов),
 - разработке формализованного описания функций, общего для всех организмов.

The End