

PRACOWNIA WYKORZYSTANIA ZASOBÓW INTERNETOWYCH

Dr Anna Modrak-Wójcik

Celem zajęć jest zapoznanie studentów z ogólnodostępnymi zasobami sieci Internet oraz zasobami bibliotek wirtualnych, związanymi z biofizyką molekularną, genetyką, farmakologią i medycyną. Uczestnicy w sposób praktyczny będą mogli wykorzystać Internet do poszukiwania i analizowania informacji naukowych.

I. BAZY BIBLIOGRAFICZNE – WSTĘP

1. Przydatne linki:

PubMed

<http://www.ncbi.nlm.nih.gov/pubmed>

Bookshelf

<http://www.ncbi.nlm.nih.gov/books/>

Wirtualna Biblioteka Nauki (ICM)

<http://vls.icm.edu.pl/>

Biblioteka Uniwersytecka w Warszawie (Biblioteka UW)

<http://www.buw.uw.edu.pl>

Google Scholar

<http://scholar.google.pl/>

2. Słowniczek

PubMed – baza danych obejmująca artykuły z dziedziny medycyny i nauk przyrodniczych założona w 1996 roku przez **National Center for Biotechnology Information (NCBI)**, będący częścią **National Library of Medicine**, wchodzącego w skład **National Institutes of Health**. Zapewnia bezpłatny dostęp do artykułów znajdujących się w bazie **MEDLINE** oraz niektórych artykułów z czasopism do niej nienależących. Publikuje głównie abstrakty artykułów oraz łączy do strony wydawcy czasopisma, w którym dany artykuł się ukazał (gdzie w niektórych przypadkach dostępna jest jego pełna wersja). PubMed korzysta z

systemu **Entrez** (<http://www.ncbi.nlm.nih.gov/sites>), służącego do uzyskiwania informacji również z innych projektów NCBI, takich jak PubChem czy Genbank. Artykuły bazy danych PubMed są indeksowane w systemie **Medical Subject Headings (MeSH)**. Każdemu artykulowi przyporządkowany jest unikalny **identyfikator PMID**.

Baza **MEDLINE** stanowi największy składnik PubMed. Obejmuje około 5400 czasopism biomedycznych z ponad 80 krajów, zawiera ponad 18 milionów pozycji datujących się od 1946 roku.

Wirtualna Biblioteka Nauki – jeden z pierwszych w Polsce systemów udostępniania naukowych baz danych przez Internet, zarządzany przez **Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego (ICM)** Uniwersytetu Warszawskiego, którego początki sięgają 1996 roku. Poszczególne zasoby są udostępniane jednostkom naukowym w ramach WBN na podstawie umów licencyjnych zawieranych pomiędzy ICM a wydawcami. W chwili obecnej (2011 r.) Biblioteka Wirtualna Nauki zawiera pełne teksty czasopism największych światowych wydawców (Elsevier, Springer, ACS, AIP/APS, Nature, Science), abstrakty, cytowania oraz zasoby faktograficzne, a także polskie zasoby pełnotekstowe. Z zasobów Wirtualnej Biblioteki Nauki korzysta obecnie ponad 170 jednostek naukowych w Polsce.

II. PUBMED

Zadanie 1

Wejść na stronę główną bazy PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>). Zarejestruj się i zaloguj do **My NCBI** (kliknij na **My NCBI** na górnym pasku po prawej stronie)

Zadanie 2

Dowiedziałas/ęś się, w jednym z wydań czasopisma BMJ można znaleźć artykuł, porównujący korzyści zdrowotne wynikające z picia martini wstrząśniętego (shaken) lub mieszanego (stirred). Znajdź ten artykuł. Posłuż się opcją **Single Citation Matcher**. (strona główna => **PubMed Tools**)

Zadanie 3

Znajdź pełny tytuł czasopisma BBRC. Kto jest jego wydawcą, jak często się ukazuje, w którym roku ukazało się pierwsze wydanie? Wskazówka: znajdź jakikolwiek artykuł w tym

czasopiśmie, wpisując BBRC w okienku wyszukiwarki; ze strony z wynikami wyszukiwania przejdź do strony informacyjnej dotyczącej tego artykułu, kliknij lewym klawiszem myszy na nazwę czasopisma i wybierz *NLM Catalog*.

Zadanie 4

Znajdź artykuły, których autorem jest S. R. Kimball. Jaka jest jego afiliacja?

Zadanie 5

Znajdź odnośniki do artykułów o Jamesie D Watsonie. Skorzystaj z tagu *Personal name as subject*. Wyeliminuj wszystkie pozycje, których autorem jest Watson. Wybierz następnie kilka pozycji, które są dla Ciebie interesujące i umieść je w schowku (*Sent to => Clipboard*). Możesz skorzystać z *Search Builder* w *Advance Search* lub wpisać frazy z odpowiednimi tagami ([...]) i kwantyfikatorami logicznymi (AND, OR, NOT). Pełną listę tagów i opis odpowiadających im obszarów znajdziesz w *PubMed Help* w zakładce *Search Field Descriptions and Tags*

Zadanie 6

Znajdź artykuły w BBRC, których autorem jest S. R. Kimball. Wykorzystaj historię wyszukiwania (*Search History* w *Advanced Search*). Używając filtru, wyszukaj artykuły przeglądowe (*Filter your Results => Review*) i utwórz zawierającą je nową kolekcję w *My NCBI* (użyj opcji *Sent to => Collections*). Nadaj jej odpowiednią nazwę.

Zadanie 7

Znajdź artykuły S. R. Kimballa, które ukazały się po 2007 roku (opcja *Limits*) i zawierające w tytule lub abstrakcie termin *mRNA*. Dodaj je do kolekcji utworzonej w zadaniu 6.

Zadanie 8

- A. W jednym wyszukiwaniu znajdź cytowania dotyczące białka GFP (Green Fluorescent Protein) i procesu tworzenia struktury trzeciorzędowej białek (protein folding). Obejrzyj w jaki sposób zostało przeprowadzone wyszukiwanie (*Search Details*).
- B. Przeprowadź analogiczne wyszukiwanie używając *MeSH Database* do konstrukcji zapytania (również przy ograniczeniu do *Major Topic headings only*).
- C. Spośród cytowań znalezionych w punkcie B wybierz te, które powstały w placówkach naukowych mających swą siedzibę w Nowym Yorku (ogranicz wyszukiwanie do

odpowiedniej dziedziny – box *Field w Advanced Search* lub zastosuj odpowiedni **tag**). Zapisz wynik wyszukiwania w *My NCBI (Save search)* oraz stwórz plik tekstowy zawierający spis cytowań (użyj funkcji *Sent to => File*).

- D. Wybierz trzy odnośniki z punktu B i wyślij je do siebie poprzez e-mail (*Sent to => E-mail*) w formacie, który zawiera abstrakty.

Zadanie 9

Wykorzystując PubMed, Wirtualną Bibliotekę Nauki oraz Bibliotekę UW, postaraj się znaleźć pełne teksty odnośników znalezionych w **zadaniu 8 C**.

Zadanie 10

Znajdź pełne teksty artykułów przeglądowych dotyczących białka GFP, które ukazały się w czasopiśmie *Science*.

III. BIOLOGICZNE BAZY DANYCH – WSTĘP

Biologiczne i biomedyczne bazy danych to „kopalnie” informacji z dziedziny nauk przyrodniczych, pochodzących z eksperymentów naukowych i projektów wielkoskalowych, artykułów naukowych i analiz komputerowych. Zawierają m. in. dane dotyczące sekwencji, funkcji i struktury genów oraz białek, efektów klinicznych mutacji, ścieżek metabolicznych jak również podobieństw między sekwencjami i strukturami.

Raz w roku czasopismo „*Nucleic Acids Research*” publikuje listę wszystkich dostępnych baz biologicznych, wraz z ich adresami internetowym i krótkim opisem bazy. W roku 2011 lista ta obejmowała 1330 pozycji (<http://www.oxfordjournals.org/nar/database/c/>).

Przydatne linki:

ExPASy (Expert Protein Analysis System) <http://www.expasy.ch/>

serwer proteomiczny Swiss Institute of Bioinformatics (SIB, Genewa, Szwajcaria), współpracujący z European Bioinformatics Institute (EBI) oraz z Protein Information Resource (PIR)

European Bioinformatics Institute (EBI, Hinxton, UK) <http://www.ebi.ac.uk/>

wchodzi w skład European Molecular Biology Laboratory (EMBL), jednostki wspieranej przez 16 krajów (Europa zachodnia i Izrael)

National Center for Biotechnology Information (NCBI, Bethesda, USA)
<http://www.ncbi.nlm.nih.gov/>

wchodzi w skład National Library of Medicine (NLM) przy National Institutes of Health (NIH). Gromadzi informacje dotyczące sekwencji nukleotydowych (w bazie **GenBank**), sekwencji białkowych (**Protein**), artykułów biomedycznych (**PubMed**), małych cząsteczek biologicznych (**PubChem**) a także inne informacje dotyczące biotechnologii. Wszystkie te bazy danych są dostępne w sieci przez wyszukiwarkę **Entrez**

International Nucleotide Sequence Database Collaboration <http://insdc.org/>

międzynarodowa baza wszystkich publicznie dostępnych sekwencji nukleotydowych oparta na współpracy między trzema bazami: **DDBJ** (DNA Data Bank of Japan, <http://www.ddbj.nig.ac.jp/index-e.html>), **ENA** (European Nucleotide Archive, <http://www.ebi.ac.uk/ena/>), **GenBank** (<http://www.ncbi.nlm.nih.gov/genbank/>)

Worldwide Protein Data Bank (wwPDB) <http://www.wwpdb.org/index.html>

ogólnoświatowa baza struktur molekularnych, zawierająca w lutym 2011 r. 71415 struktur, założona przez **RCSB PDB** (<http://www.rcsb.org/pdb/home/home.do>; USA), **PDBe** (<http://www.ebi.ac.uk/pdbe/>; Europa) i **PDBj** (<http://www.pdbj.org/>; Japonia)

IV. BAZY SEKWENCJI BIAŁKOWYCH – UNIPROT

1. Przeszukiwanie UniProt przy użyciu *Text Search*.

Zadanie 1

Na stronie domowej Uniprot (<http://www.uniprot.org/>) w okienku *Query* (u góry strony) wpisz: **Myosin light chain kinase**. Upewnij się, że w okienku *Search in* figuruje „**Protein Knowledgebase (UniProtK)**”. Wciśnij *Search*. Ile jest wszystkich wyników wyszukiwania? Ile znaleziono w bazie Swiss-Prot? Aby ograniczyć liczbę wyników przejdź do *Advanced Search* (u góry z prawej). W okienku *Field* wybierz „**Organism [OS]**”, okienku *Term* wpisz

„Human”, wybierz „Human[9606]”, kliknij *Add & Search*. Ile tym razem wyników otrzymałaś/eś?

2. Opis wyniku wyszukiwania – informacje ogólne

Aby obejrzeć opis dotyczący konkretnej pozycji, należy kliknąć na hiperlink w pierwszej kolumnie (*Accession*).

Zadanie 1

Kliknij na **Q15746**. Na pasku znajdującym się bezpośrednio nad tekstem opisu, kliknij *Entry info*. W którym roku białko znalazło się w bazie Swiss-Prot? Ile było różnych wersji sekwencji?

Zadanie 2

Przejdź do *Names and origin*. Jakie są inne nazwy białka? Jak nazywa się kodujący je gen?

Zadanie 3

Kliknij na “*EC=2.7.11.18*” (kod enzymu; *Ezyme Nomenclature Committee*). Przenosisz się do bazy **ENZYME**. Czy są podane inne nazwy dotyczące tego enzymu? Jaką reakcję katalizuje ta kinaza?

Zadanie 4

Przejdź do *Protein attributes*. Czy są przesłanki świadczące o obecności kinazy na poziomie białkowym?

3. Opis wyniku wyszukiwania – komentarze

Zadanie 1

Przejdź do *Ontologies*. Terminy, które tu znajdziesz dadzą Ci pogląd na temat czym jest poszukiwane białko. Jakie ligandy wiąże kinaza? W jakich procesach biologicznych bierze udział? Jaka jest jej molekularna funkcja?

Zadanie 2

Przejdź do *General annotation*. Ta część opisu jest sporządzana na podstawie danych literaturowych. Jaką funkcję pełni nasze białko? Jaką rolę odgrywają modyfikacje post-translacyjne? Jakie domeny posiada?

4. Opis wyniku wyszukiwania – sekwencja

Zadanie 1

Przejdź do *Alternative products*. Ile jest różnych izoform naszego białka? Czym się różnią? Czy wszystkie posiadają aktywność katalityczną? Jakiego fragmentu brakuje w Izofornie 6? Jaka jest jej alternatywna nazwa?

Zadanie 2

Przejdź do *Sequences annotation (Features)*. Jest to miejsce, gdzie możesz dowiedzieć się więcej o charakterystycznych cechach sekwencji (domenach, miejscu aktywnym, modyfikowanych aminokwasach, różnicach między izoformami). Jaka jest pozycja aminokwasów należących do domeny mającej aktywność kinazy białkowej? Wyjaśnij, dlaczego telokina nie posiada aktywności katalitycznej. Jaka jest różnica między izoformami 1 i 5? Jaka jest pozycja miejsca aktywnego (akceptor protonu)? Jaki to aminokwas? Jaka reszta odpowiada za wiązanie ATP? Ile modyfikowanych aminokwasów występuje w białku? Na czym polegają modyfikacje?

Zadanie 3

Przejdź do *Sequences*. Wybierz *ProtParam* (w okienku *Tools*) i wciśnij *Go*. Kliknij na „CHAIN 1-1914”. Ile wynosi masa cząsteczkowa naszego białka, punkt izoelektryczny, współczynnik ekstynkcyjny w 280 nm? Czy kinaza jest stabilnym białkiem? Jaki jest przewidywany połowkowy czas życia? Ile reszt tryptofanowych i tyrozynowych zawiera?

Zadanie 4

Porównaj sekwencję izoformy 1 i 2. Zaznacz pola po lewej stronie I1 i I2 i na zielonym pasku na dole strony, kliknij *Align*.

5. Opis wyniku wyszukiwania – struktura



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Zadanie 1

Przejdź ponownie do *Sequences annotation (Features)*. W części *Secondary Structure* kliknij *Details* Co możesz powiedzieć o strukturze drugorzędowej białka?

Zadanie 2

Przejdź do *Cross references* i znajdź podrozdział dotyczący struktur trójwymiarowych (*3D structure database*), Znajdziesz tu wszystkie eksperymentalnie wyznaczone struktury zdeponowane w PDB oraz przewidziane na podstawie homologii (ModBase), jeśli takie istnieją. Jakimi metodami uzyskano struktury? Czy znana jest struktura całego białka? Obejrzyj struktury w różnych bazach (PDB, PDBe, PDBj).

6. Przeszukiwanie bazy UniProt przy pomocy narzędzia BLAST

Zadanie 1

Kliknij na okienko *Blast* (u góry strony, na szarym pasku). Sekwencja ludzkiej kinazy MYLK powinna znajdować się w okienku wyszukiwania. W okienku *Database* wybierz *UniProtKB/Swiss-Prot*. Jakie białka są najbardziej podobne do ludzkiej kinazy (nie licząc izoform MYLK)? Jaki jest % podobieństwa?

Zadanie 2

Na stronie z wynikami wyszukiwania zaznacz okienka (z lewej strony) odpowiadające MYLK (Q15746) i dwóm najbardziej podobnym białkom i wciśnij *Align*. Zlokalizuj rejony o wysokim i niskim stopniu podobieństwa.

(„*” – identyczne reszty, „:” – podstawienie konserwatywne, „.” – semi-konserwatywne, „-” – przerwa)

Zadanie 3

Czy białka te mają podobne rejony hydrofobowe i polarne? Co możesz powiedzieć o położeniach reszt tyrozynowych i treoninowych? (zaznacz odpowiednie okienko w *Amino acid properties*)

Zadanie 4



Jakie informacje dotyczące cech sekwencji są dostępne dla naszych białek (*Sequence annotation*). Znajdź ich położenie w sekwencji.

V. BAZY SEKWENCJI NUKLEOTYDOWYCH – GENBANK

Zadanie 1

Jakie są różnice między genami i genomami pochodzącymi z organizmów prokariotycznych i eukariotycznych.?

Zadanie 2

Przejdź na stronę **GenBanku** (na stronie głównej NCBI <http://www.ncbi.nlm.nih.gov/> wybierz *Nucleotide* w okienku *Popular Resources*, alternatywnie w okienku wyszukiwania wybierz *Nucleotide*). W okienku wyszukiwania wpisz **X01714** (GenBank ID). Odpowiedz na pytania:

- Z jakiego organizmu pochodzi wyszukana sekwencja i jakiego jest typu (RNA, DNA)?
- Co koduje? Ile par zasad zawiera?
- Jakie jest położenie promotora? W którym miejscu rozpoczyna się transkrypcja?
- Ile sekwencji kodujących zawiera? Która koduje znane białko?

objaśnienia skrótów: **RBS** – Ribosome Binding Site, **CDS** – CoDing Segment

Zadanie 3

Wyszukaj informację na temat sekwencji **AF018430**.

- Z jakiego organizmu pochodzi?
- Z jakim genem jest związana? Czy zawiera pełen gen?
- Na którym chromosomie znajduje się gen zawierający wyszukaną sekwencję?
- Ile różnych sekwencji zawartych w GenBank potrzeba do rekonstrukcji genu i sekwencji mRNA kodującej białko?
- Ile różnych wersji mRNA powstaje w wyniku alternatywnego splicingu. Jakie są ich produkty białkowe?
- Jaki fragment wyszukanej sekwencji to ekson? Czy występuje on w obu wersjach CDS?

Oznaczenia: znak „<” na początku formuły dotyczącej opisu sekwencji danego obiektu (gene, mRNA) oznacza, że gen może zacząć się prze oznaczoną pozycją; znak „>” na końcu formuły oznacza, że gen może rozciągać się za oznaczoną pozycję.

Zadanie 4

Wyszukaj informację na temat sekwencji **AB001981**. Odpowiedz na pytania:

- A. Z jakiego organizmu pochodzi sekwencja?
- B. Ile genów zawiera sekwencja? Jakie białka kodują te geny?
- C. Ile eksonów zawiera każdy z genów? Jakim fragmentom sekwencji odpowiadają?
- D. Jaką trójką nukleotydową zaczyna się i kończy każdy z regionów CDS?
- E. Obejrzyj sekwencję w modzie graficznym.

Zadanie 5

Wyszukaj informacje dotyczące genu kodującego ludzką insulinę. Przy konstrukcji zapytania użyj odpowiednich tagów lub ogranicz wyszukiwanie do konkretnych obszarów. Pomocny może być opis obszarów wyszukiwania i odpowiadających im tagów (kliknij na zakładkę **HELP** na stronie GenBank i wyszukaj **Search Field**).

Odpowiedz na pytania:

- A. Z ilu sekwencji zdeponowanych w bazie GenBank składa się poszukiwany gen? Jakie są ich numery ID (ACCESSION)? Które nukleotydy tworzą gen?
- B. Jaki jest numer ID poszukiwanego genu?
- C. Na którym chromosomie leży gen inuliny?
- D. Co możesz powiedzieć o modyfikacjach potranslacyjnych pierwotnego produktu białkowego na podstawie opisu genu?

Sprawdź jakie informacje dotyczące budowy ludzkiej insuliny dostępne są w bazie Uniprot.

Zadanie 6

Znajdź pełnej długości sekwencję cDNA (mRNA) kodującą dehydrogenazę GAPDH (glyceraldehyde 3-phosphate dehydrogenase) z zarodźca sierpowego (*Plazmodium falciparum*; pierwotniak wywołujący malarię u ludzi).

Ile par zasad zawiera? Na którym chromosomie leży odpowiadający jej gen?

VI. BAZY MIEJSC FUNKCYJNYCH, MOTYWÓW, DOMEN I RODZIN BIAŁKOWYCH

1. PROSITE

(<http://expasy.org/prosite/>; user manual: <http://expasy.org/prosite/prosuser.html>)

Zadanie 1

Wejdź na stronę **EXPASY** (<http://expasy.org/>). Wybierz *ScanProsite* w *Tools&Software* lub *PROSITE* (w *Databases*) i wejdź w *ScanProsite*. Wpisz **P12259** w okienku po lewej stronie (Swiss-Prot accession numer), odhacz „*Exclude motifs with a high probability of occurrence*”, zaznacz „*Do not scan profiles*” i wciśnij **START THE SCAN**.

- Jakiego białka dotyczyyszukana sekwencja?
- Ile uzyskałaś/eś wszystkich wyników wyszukiwania? Ile wśród nich to motywy o wysokim prawdopodobieństwie występowania?
- Jakie charakterystyczne motywy zostały zidentyfikowane dla naszej sekwencji? Jeśli istnieją, obejrzyj ich strukturę 3D
- Jakie domeny zawiera i do jakiej rodziny białek należy?
- Jaka wygląda wzór sekwencyjny dla motywu *Coagulation factors 5/8 type C domain (FA58C) signature 1*?

Oznaczenia: x – dowolny aminokwas, [] – różnorodność, () – powtórzenia, {} – wykluczenie

- Przyjrzyj się motywowi dotyczącemu *N-myristoylation site*. Czy jest możliwe, by motyw ten znajdował się w naszej sekwencji?

Zadanie 2

Przeprowadź w PROSITE wyszukiwanie profili i wzorów (patterns) dla poniższej sekwencji (zaznacz „*Exclude motifs with a high probability of occurrence*”):

```
MERCGAVVSCVFAVILVVLQLGLTPIMAQQEQQIGIAGKEVILSCKAINNQKDGCTCTWKYK
YKEVSSSTIIISFSKAQVFVFKGKAPMTHRSELNSNSKCLKVSDLSLDDAGIYTCACYSPVVS I
SLHVFKLTISSNGHFLTNEDELELTLMQNSSHSQPHLSIKLFNINNDIVTTEILQEEAPQK
YILKLLKQLKAIDSGTWMCHVYNSNSPSINQNISFDVKVLGFEEKERLEIIYTTVGNTAILSW
RLNFRKIKWKEGFTGKLNWEPQGNIAHELLNFVSVTTHQELHKTKKSNHIWFEISEGKTD
GTMDVKIIPKVQLNHSGQYKQCQLEINGRRTESVVRALVVMQVTAIPAGPLSRGGKMTLLCQV
SGPLPSNAHLLWERVNGTQMEMKSKQHEAKVEVNVSAAPGLWNCHLVEDNKKISLNYTV
EEAHVWNSYAVIGIIIGASVLVIGLACMCIITGMRWQRRRKRARRMAQAKQYLLEKKTQCQ
CQRRMYK
```

- Jaką domenę/domeny zostały zidentyfikowane dla powyższej sekwencji? Czy występują w nich mostki disiarczkowe? Jeśli tak, to między którymi aminokwasami?




- B. Wejdź do opisu profilu. Co na ich podstawie możesz powiedzieć o poszukiwanym białku?
- C. Używając bazy UniProt zidentyfikuj powyższą sekwencję. Jakiego białka dotyczy? Z jakiego pochodzi organizmu?
- D. Używając bazy PROSITE odpowiedz, jak wiele ludzkich białek zawiera domenę zidentyfikowaną w punkcie A) powtórzoną co najmniej czterokrotnie? Przy wyszukiwaniu posłuż się charakterystycznym numerem profilu (ACCESSION NUMBER), zastosuj odpowiedni filtr (TAXONOMY – homo sapiens), a w części „*Output*” w „*Show only sequences with at least ___ hit(s)*” wpisz odpowiednią liczbę

2. INTERPRO (<http://www.ebi.ac.uk/interpro/>)

Zadanie 1

Wejdź na stronę InterPro (<http://www.ebi.ac.uk/interpro/>)

- A. Jakie bazy możesz przeszukiwać? Wskazówka: wejdź w zakładkę *InterPro Scan*
- B. Wpisz w okienku wyszukiwania **O15075** (UniProt accession numer). Ile różnych pozycji bazy InterPro (ale nie indywidualnych sygnatur) otrzymałaś/eś? Ile domen możesz wyróżnić w sekwencji?
- C. Kliknij na link **IPR000719**, aby zobaczyć opis dotyczący tej pozycji. Jak jest nazwa domeny? Ile i jakie sygnatury składają się na tę domenę? W ilu białkach występują te sygnatury? W jakiego typu organizmach występuje ta domena?
- D. Wróć na stronę z wynikami wyszukiwania i przejdź do części *Structural Features*. Dla jakiego fragmentu białka znana jest struktura 3D? Jakiej domenie odpowiada?
- E. Obejrzyj tę domenę używając *AstexViewer* (ikonka  przy strukturze domeny w **SCOP** lub **CATCH**). Jaka jest struktura drugorzędowa domeny? Ile widzisz ligandów i jakie? Z jakimi resydami tworzą wiązania wodorowe?

Zadanie 2

Używając *InterPro Scan* wyszukaj charakterystyczne motywy w poniżej sekwencji:

```
MDVFSFGQGN NTTASQEPFG TGGNVTISID VTFSYQVITS LLLGTLIFCA VLGNACVVAA
IALERSLQNV ANYLIGSLAV TDLMVSVLVL PMAALYQVLN KWTLGQVTC D LFIALDVLCC
TSSILHLCAI ALDRYWAITD PIDYVNRTP RRAALISLT WLIGFLISIP PMLGWRTPED
RSDPDACTIS KDHGYTIYST FGAFYIPLLL MLVLYGRIFR AARFRIRKTV RKVEKKGAGT
SLGTSSAPPP KKSLNGQPGS GDWRRCAENR AVGTPCTNGA VRQGDDEATL EVIEVHRVGN
SKEHLPLPSE SGSNSYAPAC LERKNERNAE AKRKMALARE RKTVKTLGII MGTFILCWLP
```



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



FFIVALVLPF CESSCHMPAL LGAIINWLGYSNSLLNPVIY AYFNKDFQNA FKKIIKCKFC
R

- A. Określ, które pozycje bazy InterPro dotyczą domen, rodzin, miejsc itp. Jakie są powiązania między nimi?
- B. W ilu sekwencjach występują? W jakich organizmach?
- C. Która z pozycji jest najbardziej specyficzna?

VII. BAZY STRUKTUR MOLEKULARNYCH – PROTEIN DATA BANK (PDB)

<http://www.pdb.org/pdb/home/home.do>

Zadanie 1

Wpisz w okienku wyszukiwania kod PDB **3CLN**.

- A. Jakiego białka z jakiego organizmu dotyczy wyszukana struktura?
- B. Jaką metodą i z jaką rozdzielczością została wyznaczona?
- C. Ile łańcuchów polipeptydowych zawiera? Czy zawiera ligandy?
- D. Obejrzyj jej file PDB. Kiedy struktura trafiła do bazy? Jaką funkcję pełni białko, którego strukturę oglądasz? Z jakich części składa się file PDB?
- E. Na stronie głównej PDB znajdź link do „przewodnika po PDB File” (*Tools => File Formats => PDB File Format => <http://www.wwpdb.org/docs.html>*).
- F. Czy z komentarzy wynika, że białko wiąże jony wapnia? *Wskazówka: przeczytaj **REMARK 800**.*
- G. Ile jonów wapnia i cząsteczek wody występuje w strukturze? *Wskazówka: patrz **HET, FORMUL, HETATM**. Sprawdź w opisie filu PDB co dokładnie znaczą te noty.*
- H. Które reszty aminokwasowe biorą udział w wiązaniu jonów wapnia? Czy potrafisz zidentyfikować atomy biorące udział w tych wiązaniach? Jakie są odległości między atomami białka a jonami wapnia? *Wskazówka: patrz **LINK i SITE**.*
- I. Dużą część filu PDB zajmuje szczegółowy opis współrzędnych atomowych (**ATOM**). Wyjaśnij znaczenie poszczególnych pozycji linii opisu.
- J. Dlaczego lista współrzędnych atomowych białka zawiera residua 5-147 (**ATOM**), a część sekwencyjna 1-148 (**SEQRES**)? *Wskazówka: przeczytaj **REMARK 465**.*

Zadanie 2

Wpisz w okienku wyszukiwania kod PDB **1CFC**.



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



- A. Jakiego białka z jakiego organizmu dotyczy wyszukana struktura?
- B. Jaką metodą została wyznaczona?
- C. Ile zawiera różnych konformerów? Obejrzyj te konformery. Który fragment łańcucha jest „sztywny”, a który „ruchliwy”?
- D. Obejrzyj file PDB (może być sam „header”)

Zadanie 3

Porównaj powyższe struktury kalmoduliny. *Wskazówka:* Użyj zakładki **Compare Structures** w **Tools**.

Zadanie 4

Znajdź wszystkie struktury ludzkiej trypsyny (wpisz w okienko wyszukiwania: **human trypsin**).

- A. Ile struktur otrzymałaś/eś? Czy wszystkie dotyczą trypsyny człowieka?
- B. Przeprowadź wyszukiwanie wpisując nazwę enzymu w cudzysłowie („**human trypsin**”). Ile struktur zostało znalezionych tym razem? Zidentyfikuj numer EC trypsyny
- C. Przeprowadź wyszukiwanie zaawansowane (**Advanced Search**) struktur ludzkiej trypsyny posługując się numerem EC enzymu i nazwą organizmu, z którego pochodzi (**homo sapiens**). Ile tym razem struktur otrzymałaś/eś? Wyłumacz dlaczego w poprzednim wyszukiwaniu było ich mniej.
- D. Wejdz do opisu struktury **1TRN**. Czy występują w niej ligandy? A zmodyfikowane aminokwasy? Obejrzyj jak wyglądają. Ile łańcuchów polipeptydowych zawiera struktura? Jakiej są długości?
- E. Obejrzyj wykres Ramachandrana dla tej struktury. Co możesz powiedzieć o jakości oglądanej struktury na jego podstawie? *Wskazówka:* wejdz w zakładkę **Geometry**.
- F. Przejdź do filu PDB. Ile zmodyfikowanych reszt występuje w strukturze? Podaj ich numery.
- G. Ile mostków disiarczkowych zaobserwowano w strukturze? Czy łączą ze sobą poszczególne łańcuchy? *Wskazówka:* patrz **SSBOND**

VIII. PYMOL – PROGRAM DO WIZUALIZACJI STRUKTUR MOLEKULARNYCH

Pymol manual: <http://pymol.sourceforge.net/html/>

Pymol Wiki: www.pymolwiki.org/

1. Uruchamianie programu PyMOL i „ściągnięcie” zbioru PDB

Zadanie 1

W bazie PDB znajdź strukturę **2biw** i zapamiętaj ją w formacie *.pdb (na dysku sieciowym). Ile łańcuchów polipeptydowych zawiera ta struktura?

Zadanie 2

Otwórz PyMOL dwukrotnie klikając na jego ikonkę.

Wpisz komendę *log_open log-file-name.pml*, by zapamiętać sesję w postaci skryptu. Komendy możesz wpisywać zarówno białym-szarym oknie (**external GUI** – Graphical User Interface), jak i w czarnym oknie (**PyMol Viewer**)

Zadanie 3

Załaduj file PDB zapamiętany w zadaniu 1, wybierając *File => Open* na pasku menu górnego okienka (**external GUI** – Graphical User Interface).

2. Manipulacje na cząsteczce (przesuwanie, obracanie, zmienianie reprezentacji, kolorowanie)

Zadanie 1

Wypróbuj zastosowanie poszczególnych przycisków myszy do obracania, przesuwania, powiększania i pomniejszania cząsteczki.

Zadanie 2

Przetestuj różne sposoby przedstawiania cząsteczki posługując się przyciskami **S/H** wewnętrznego interfejsu graficznego (**internal GUI**) oraz poprzez napisanie odpowiednich komend (*hide/show representation*).

Zadanie 3

Przedstaw cząsteczkę w reprezentacji *cartoon*. Pokoloruj poszczególne elementy struktury drugorzędowej na inne kolory: helisy na zielono, beta-kartki na czerwono, skręty na żółto. Użyj komendy *color*.

Zadanie 4

Wypróbuj różne sposoby rysowania α -helisy, β -kartek oraz pętli. Użyj przycisków *Settings => Cartoon*

Zadanie 5

Zmień kolor tła na biały posługując się przyciskami *Display => Background*, a następnie znów na czarny używając komendy *bg_color color-name*.

Zadanie 6

Pokoloruj każdy z łańcuchów na inny kolor: łańcuch A – na czerwono, B – na cyjan, C – na żółto, D – na zielono

3. Wyodrębnianie fragmentów cząsteczki i tworzenie nowych obiektów

Zadanie 1

W każdym z łańcuchów wyodrębnij fragment odpowiadający aminokwasom od 150 do 250. Nadaj mu nazwę *frag*. Skorzystaj z komendy *select*. Pokoloruj wybrany fragment na różowo i przedstaw w reprezentacji *sticks*.

Zadanie 2

Wybierz ten sam fragment, ale tylko w łańcuchu A. Nazwij go *frag_A*. Definiując „selekcję” skorzystaj z kwantyfikatorów logicznych lub z poniżej zdefiniowanych hierarchicznych konstrukcji (*macra*):

/object-name/semi-indentyfier/chain-indentyfier/resi-indentyfier/name-indentyfier lub
object-name/semi-indentyfier/chain-indentyfier/resi-indentyfier/name-indentyfier

Druga opcja (*macra*) jest bardzo przydatna przy skomplikowanych selekcjach, np.,:

select pept and segi lig and chain b and resi 142 and name ca = select /pept/lig/b/142/ca

Obecność lub brak „/” na początku determinuje sposób odczytania *macra* – ze slash-em czytane jest „od góry” hierarchii (pierwsze wyrażenie za „/” to nazwa obiektu), bez slash-a – „od dołu” hierarchii (więcej szczegółów znajdziesz w manualu).

Zadanie 3

Pokoloruj wybrany fragment łańcucha A na biało, przedstaw w postaci powierzchni i powiększ (*zoom selection-name*).

Zadanie 4

Sprawdź w bazie PDB jakie ligandy i w jakiej ilości występują w oglądanej strukturze. Znajdź w filu pdb odpowiadające im identyfikatory i numery.

Zadanie 4

Utwórz nowy obiekt o nazwie *ligand*, zawierający cząsteczki karotenolu (komenda *create*), przedstaw je w postaci sfer i pokoloruj na odpowiedni, kontrastujący kolor. Pokoloruj atomy żelaza na wybrany kolor i przedstaw je w postaci sfer. Zapamiętaj sesję.

4. Analiza miejsca aktywnego

Zadanie 1

Sprawdź w filu pdb, które reszty aminokwasowe i cząsteczki wody tworzą miejsce wiążące karotenol w łańcuchu B. Wyodrębnij je, nadając nazwę *active_B*.

Zadanie 2

Sporządź rysunek zawierający miejsce aktywne łańcucha B (tylko reszty miejsca aktywnego i ligand). Przedstaw je w reprezentacji *sticks*. Pokoloruj reszty „*by element*”, a ligand tak, by dobrze kontrastował z białkiem.

Zadanie 3

Przypisz etykiety z nazwami reszt miejsca aktywnego (przycisk **L** – label)

Zadanie 4

Korzystając z przycisków *Wizard* => *Measurement* zmierz odległości między wybranymi, nie połączonymi wiązaniami kowalencyjnymi, atomami miejsca aktywnego. Postaraj się znaleźć najbliższe kontakty. Jaka jest odległość między tlenem grupy OH Tyr 322 a węglem C30 karotenolu?

5. Sporządzenie rysunków dobrej jakości i zapamiętanie

Zadanie 1

Wykonaj rysunek przedstawiający cząsteczkę **2BIW**. Przedstaw białko w reprezentacji *cartoon*, każdy z łańcuchów w innym kolorze, a ligandy (3ON i Fe) jako sfery odpowiednio pokolorowane. Zmień tło na białe. Przed zapamiętaniem (*File* => *Save Image As* => *PNG*), popraw jakość rysunku używając komendy *ray* (np. *ray 2000,2000*).

Przed przystąpieniem do kolejnego zadania, możesz zapamiętać obecną orientację przestrzenną wciskając *Scene* => *Store* => *Fi* (przywołanie zapamiętanej orientacji *Scene* => *Recall* => *Fi*).

Zadanie 2

Wykonaj wizualizację miejsca aktywnego (może być dla łańcucha B). Przedstaw reszty miejsca aktywnego oraz cząsteczkę ligandu tak jak w zadaniu 1, a pozostałą część łańcucha białkowego w reprezentacji *cartoon*. Zmień kolor tła na biały. Pokoloruj odpowiednio poszczególne elementy. Komenda *zoom active_B* pozwoli Ci umieścić atomy miejsca aktywnego pośrodku okienka w odpowiednim powiększeniu (możesz skorzystać z przycisku *A* => *zoom* przy selekcji *active_B*). Możesz umieścić niektóre odległości międzyatomowe. Przed zapamiętaniem grafiki, popraw jej jakość. Zapamiętaj sesję.

IX. QTIPTOT – PROGRAM DO ANALIZY I WIZUALIZACJI DANYCH (linuksowy odpowiednik programu ORIGIN)

Zadanie 1

Otwórz program QtiPlot. Używając opcji *Import ASCII* (zakładka *File* => *Import*) wczytaj dane dotyczące pomiarów absorpcji światła laserowego przez szklane filtry (dane otrzymałaś/eś emailem, znajdziesz je także w **Tabeli 1** na końcu). Nazwij odpowiednie

kolumny tak jak w tabeli 1, a jednostki wpisz jako *Comment*. Przypisz kolumnom ich „przeznaczenie” (x , y , dx , dy). Zmian tych możesz dokonać klikając na nagłówek z nazwą kolumny lub *Table* => *Column Options*. Zapamiętaj projekt (*File* => *Save Project As*).

Zadanie 2

Zakładając, że na powierzchni filtra odbija się ok. 4% promieniowania, oblicz wartość absorbancji i jej niepewności dla każdego z filtrów. Pomocne będą zakładki *Table* => *Add Column* oraz *Table* => *Set Column Values*.

Zadanie 3

Wykonaj wykres $A(l)$ (absorbancja w funkcji grubości filtra) w postaci niepołączonych punktów (*Plot* => *Scatter*). Nanieś na wykres również niepewności absorbancji (*Graph* => *Add Error Bars*). Nazwij odpowiednio osie, dobierz skalę, znaczniki przedziałów na osiach skieruj do wnętrza, wybierz krój i kolor punktów, zatytułuj rysunek. Zapamiętaj projekt.

Zadanie 4

Do danych doświadczalnych dopasuj zależność liniową (*Analysis* => *Fit Linear*). Zmień kolor, krój i grubość linii. Umieść na rysunku w czytelnej formie parametry dopasowania prostej (*Graph* => *Add Text*). Usuń legendę. Zapamiętaj projekt.

Zadanie 5

Wyprowadź zależność łączącą $U/(0,96*U_0)$ i l . Dopasuj ją do danych doświadczalnych. Najpierw stwórz nową kolumnę, zawierającą wartości doświadczalne $U/(0,96*U_0)$. Wykonaj wykres $U/(0,96*U_0)$ w funkcji l . Następnie zdefiniuj funkcję, którą chcesz dopasować (*Analysis* => *Fit Wizard, Category* => *User defined, Save*), zaznacz „*Fit with selected user function*” i przejdź do „*Fitting Session*”. Wprowadź startowe wartości parametrów, sprawdź czy podałeś dobre dane do dopasowania i wciśnij „*Fit*”. Popraw i zapamiętaj rysunek.

Tabela 1



Zależność napięcia na fotodiodzie U po przejściu wiązki lasera przez szklany filtr w funkcji grubości filtra l . U_0 to napięcie zmierzone przed umieszczeniem danego filtra w wiązce lasera, u_{U0} i u_U to odpowiednio niepewności U_0 i U .

l [mm]	U_0 [V]	u_{U0} [V]	U [V]	u_U [V]
1,03	6,85	0,155	3,49	0,05095
2,5	6,88	0,155	1,49	0,02208
3,86	6,87	0,155	0,664	0,01016
4,82	6,85	0,155	0,402	0,00638
6,56	6,85	0,155	0,1482	0,00272
7,7	6,85	0,155	0,0737	0,00164
8,73	6,58	0,155	0,0386	0,00113
10,2	6,59	0,155	0,0165	8,15507E-4
11,38	6,59	0,155	0,0085	7,00037E-4
12,52	6,58	0,155	0,0046	6,43746E-4

Napięcie na fotodiodzie (odczytane za pomocą woltomierza) jest wprost proporcjonalne do natężenia światła padającego na fotodiodę $\Rightarrow U_0/U = I_0/I$, gdzie I_0 to natężenie światła padającego, a I – natężenie światła po przejściu przez filtr.

Prawo Lamberta-Beera opisuje pochłanianie promieniowania elektromagnetycznego przy przechodzeniu przez ośrodek absorbujący:

$$\log(I_0/I) = A = \alpha cl$$

gdzie $A = \log(I_0/I)$ to wielkość zwana absorbancją, α – współczynnik absorpcji, a c – koncentracja substancji absorbującej.