

UNIVERZITA SV. CYRILA A METODA V TRNAVE

Fakulta prírodných vied



Štefan Janeček

PROTEÍNOVÝ DIZAJN

Trnava 2014

Autor: Ing. Štefan Janeček, DrSc.

Publikácia bola schválená edičnou radou Univerzity sv. Cyrila a Metoda v Trnave a vedením Fakulty prírodných vied Univerzity sv. Cyrila a Metoda v Trnave ako učebné texty pre študentov vysokých škôl.

© Univerzita sv. Cyrila a Metoda v Trnave

© Ing. Štefan Janeček, DrSc.

Všetky práva vyhradené.

Vydala: Univerzita sv. Cyrila a Metoda v Trnave, 2014.

Vydanie: prvé.

Web: http://fpv.ucm.sk/images/ucebne_texty/Proteinovy_dizajn.pdf

ISBN 978-80-8105-594-2

Predslov

Predmet „Proteínový dizajn“ je venovaný proteínom v najširšom zmysle slova.

Základným cieľom výučby je porozumieť týmto unikátnym stavebným jednotkám živej hmoty, t.j. zákonitostiam výstavby ich molekúl.

Ďalším cieľom je pochopiť zásahy do štruktúry proteínov na molekulárnej úrovni v snahe cielene meniť ich vlastnosti.

K tomu je potrebné zvládnuť poznatky o aminokyselinách, peptidovej väzbe, vlastnostiach proteínov, detaily zo štruktúry proteínov (primárna, sekundárna, supersekundárna, terciárna, kvartérna), ich predikcie, stability a evolúcie, ako aj základy vzťahov medzi živými organizmami (*Eucarya*, *Bacteria* a *Archaea*).

Obsah

1. Proteíny, ich základná charakteristika, aminokyseliny	3
2. Vznik proteínov, peptidová väzba, peptidy	14
3. Primárna a sekundárna štruktúra proteínov	18
4. Terciárna a kvartérna štruktúra proteínov	24
5. Supersekundárna štruktúra proteínov, motívy a domény	40
6. Katalogizácia štruktúr proteínov	44
7. Predikcie štruktúry proteínov	52
8. Molekulová evolúcia proteínov	64
9. Stabilita a stabilizácia proteínov	72
10. Faktory stability proteínov	77
11. Extrémofilné mikroorganizmy a extrémostabilné proteíny	83
12. Proteínové inžinierstvo a dizajn	87
13. Praktické úlohy a cvičenia	93

1. Proteíny, ich základná charakteristika, aminokyseliny

Proteíny sú základnými stavebnými kameňmi živej hmoty.

Sú to biopolyméry zložené z aminokyselín.

Vznikajú transláciou po transkripcii genetického kódu.

Triplet nukleotidov kóduje jednu aminokyselinu.

Základná charakteristika:

- všetky proteíny sú vytvorené z 20 aminokyselín;
- aminokyselina obsahuje -NH_2 a -COOH skupinu;
- v živých organizmoch sú len α -aminokyseliny;
- všetky aminokyseliny (okrem Gly) sú opticky aktívne;
- v živých organizmoch sú len L-formy aminokyselín.

Ako fungujú proteíny:

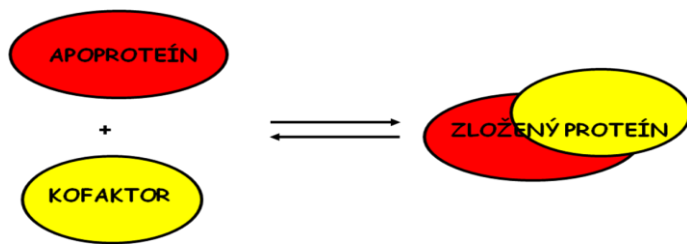
- imunitný systém – špecifické protilátky (rozpoznávajú cudzie molekuly);
- receptory – v bunkovej membráne (informujú bunku);
- enzýmy – zvýšenie reakčnej rýchlosti $\sim 10^6\times$ (biokatalyzátory);
- tvorba a prenos nervového vzruchu – neurotransmitery a zachytávače;
- transport a uchovávanie – kanály a póry v membránach (hemoglobín);
- koordinácia pohybu – svaly, spermie (bičík);
- mechanická podpora – pokožka a kosti (kolagén);
- iná obrana – zrážanie krvi, toxíny, repelenty...

Zložené proteíny:

Proteín (prostetická skupina, t.j. kofaktor):

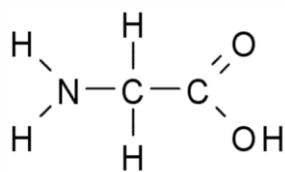
- metaloproteín (ión kovu);
- lipoproteín (lipid);
- glykoproteín (sacharid);

- fosfoproteín (fosfát);
- hemoproteín (hém, Fe).

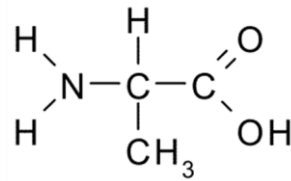


Kofaktory sú dôležité pre biologickú aktivitu proteínov.

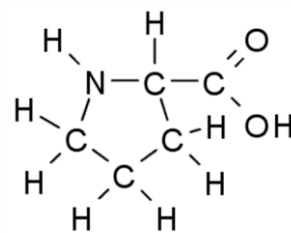
Aminokyseliny:



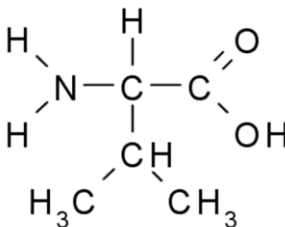
Glycín (glycine) Gly G



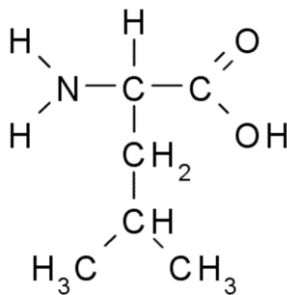
Alanín (alanine) Ala A



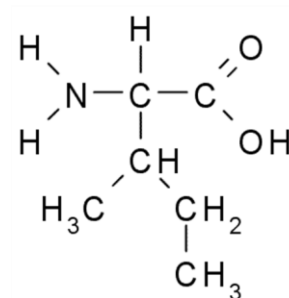
Prolín (proline) Pro P



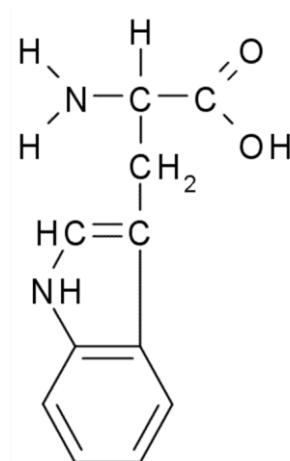
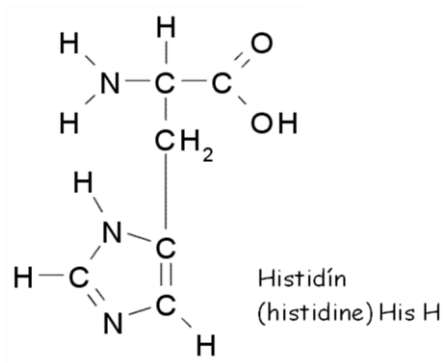
Valín (valine) Val V



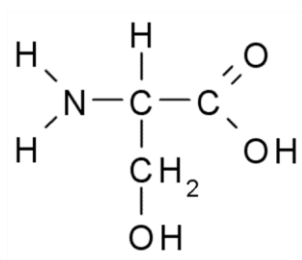
Leucín (leucine) Leu L



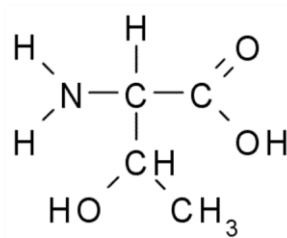
Izoleucín (isoleucine) Ile I



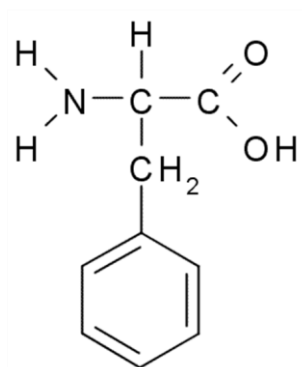
Tryptofán (tryptophan) Trp W



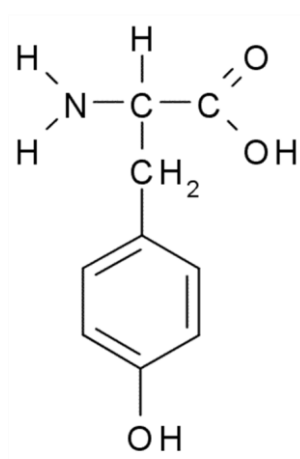
Serín (serine) Ser S



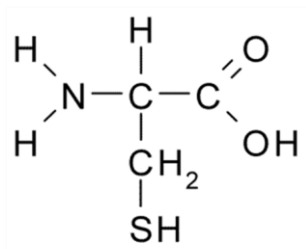
Treonín (threonine) Thr T



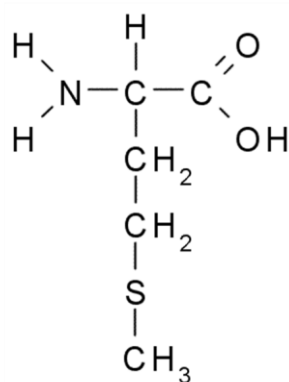
Fenylalanín (phenylalanine) Phe F



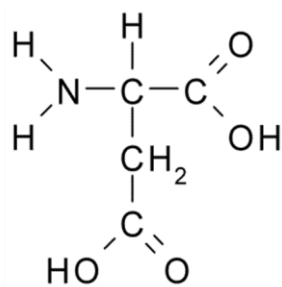
Tyrozín (tyrosine) Tyr Y



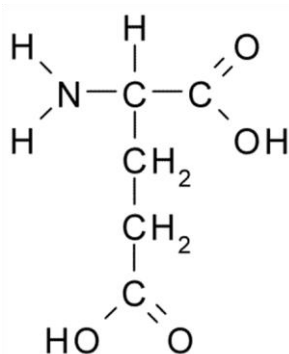
Cysteín (cysteine) Cys C



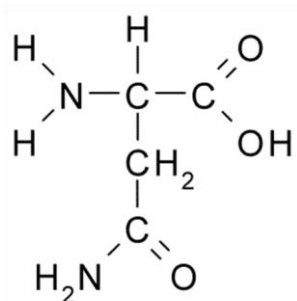
Metionín (methionine) Met M



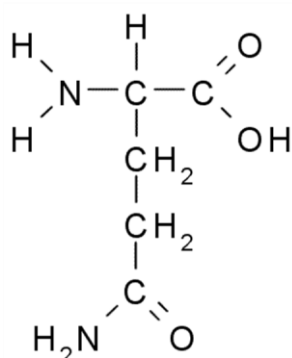
Kyselina asparágová
(aspartate) Asp D



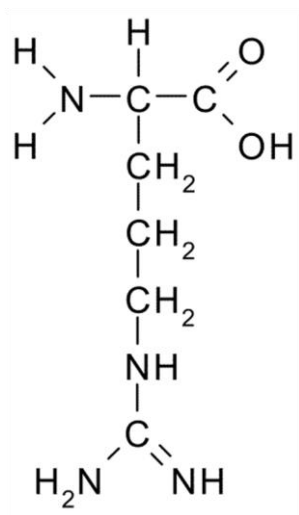
Kyselina glutámová
(glutamate) Glu E



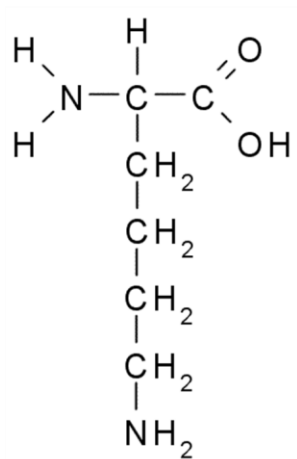
Asparagín (asparagine) Asn N



Glutamín (glutamine) Gln Q



Arginín (arginine) Arg R



Lyzín (lysine) Lys K

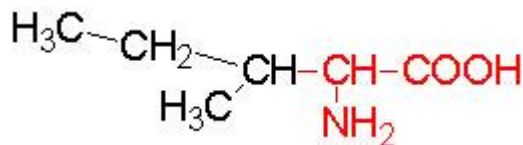
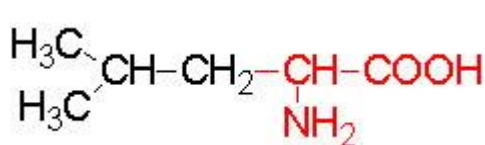
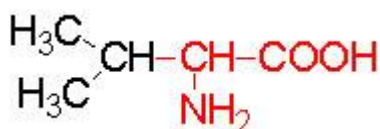
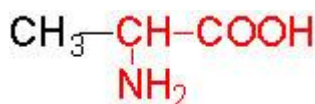
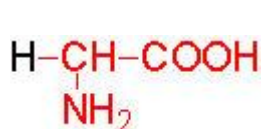


Dr. Margaret Oakley Dayhoff (1925-1983)

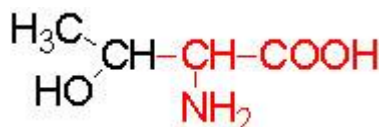
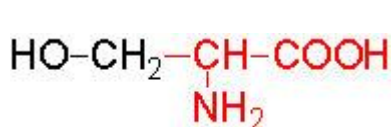
- autorka jednopísmenového kódu aminokyselin

(Upravené podľa: http://en.wikipedia.org/wiki/Margaret_Oakley_Dayhoff)

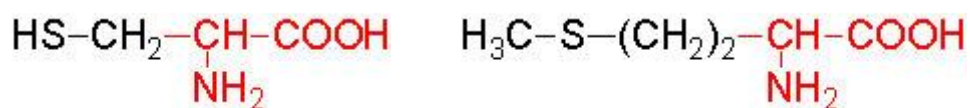
Aminokyseliny s alifatickým bočným reťazcom: Gly, Ala, Val, Leu, Ile



Nearomatické aminokyseliny s hydroxylovou skupinou: Ser, Thr



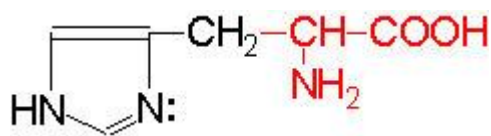
Aminokyseliny obsahujúce bočný reťazec so sírou: Cys, Met



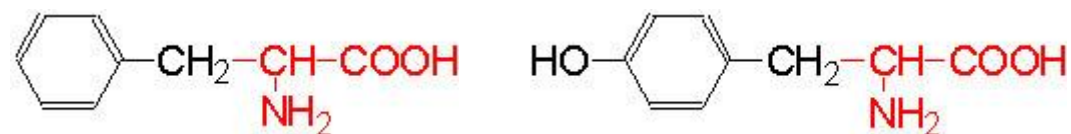
Kyslé aminokyseliny a ich amidy: Asp, Asn, Glu, Gln

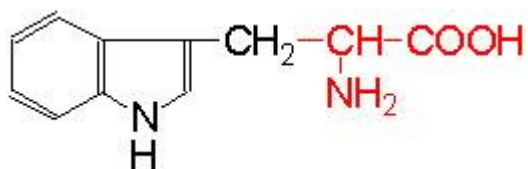


Bázické aminokyseliny: Arg, Lys, His

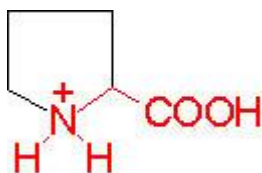


Aminokyseliny s aromatickým kruhom: Phe, Tyr, Trp





Iminokyselina: Pro

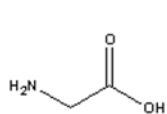


Rozdelenie aminokyselín podľa fyzikálno-chemických vlastností:

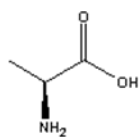
nepolárne hydrofóbne					
<chem>CC(N)C(=O)[O-]</chem> Alanine A	<chem>CC(C)C(N)C(=O)[O-]</chem> Valine V	<chem>CC(C)CC(N)C(=O)[O-]</chem> Leucine L	<chem>CC(C)CC(N)C(=O)[O-]</chem> Isoleucine I	<chem>C1CCNC1C(=O)[O-]</chem> Proline P	
<chem>CSCC(N)C(=O)[O-]</chem> Methionine M	<chem>c1ccc(cc1)CC(N)C(=O)[O-]</chem> Phenylalanine F	<chem>c1ccc2c(c1)c(c[nH]2)CC(N)C(=O)[O-]</chem> Tryptophan W	<chem>NCC(N)C(=O)[O-]</chem> Glycine G	<chem>OC(CN)C(N)C(=O)[O-]</chem> Serine S	polárne hydrofilné
<chem>CC(O)C(N)C(=O)[O-]</chem> Threonine T	<chem>SCC(N)C(=O)[O-]</chem> Cysteine C	<chem>NC(=O)CC(N)C(=O)[O-]</chem> Asparagine N	<chem>NC(=O)CC(N)C(=O)[O-]</chem> Glutamine Q	<chem>c1ccc(cc1)C(CN)C(N)C(=O)[O-]</chem> Tyrosine Y	
<chem>[O-]C(=O)CC(N)C(=O)[O-]</chem> Aspartic Acid D	<chem>[O-]C(=O)CCC(N)C(=O)[O-]</chem> Glutamic Acid E	<chem>NC(CCC(N)C(=O)[O-])C(N)C(=O)[O-]</chem> Lysine K	<chem>NC(=[NH2+])CCC(N)C(=O)[O-]</chem> Arginine R	<chem>c1c[nH]c2c(c1)c(c[nH]2)CC(N)C(=O)[O-]</chem> Histidine H	zásadité
kyslé					

(Upravené podľa: http://mol-biol4masters.masters.grkraj.org/html/Protein_Synthesis1-Prokaryotes_Introduction.htm)

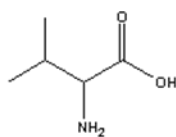
Sumarizácia štruktúr 20 aminokyselín



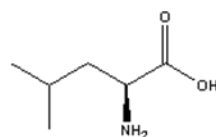
Gly



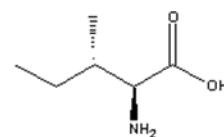
Ala



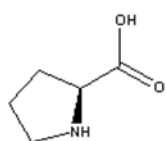
Val



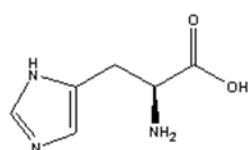
Leu



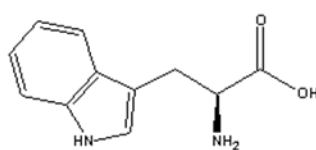
Ile



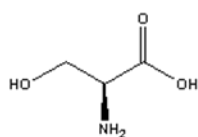
Pro



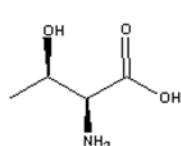
His



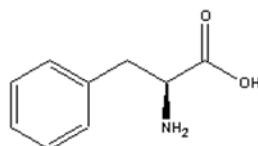
Trp



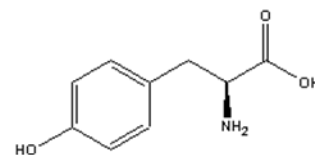
Ser



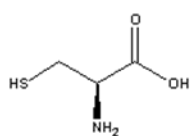
Thr



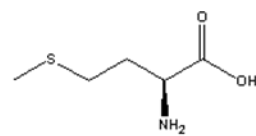
Phe



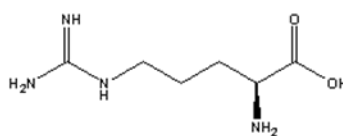
Tyr



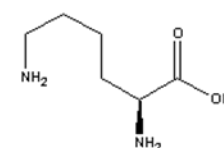
Cys



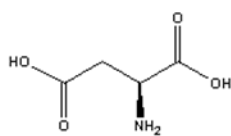
Met



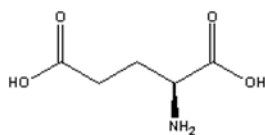
Arg



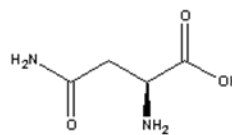
Lys



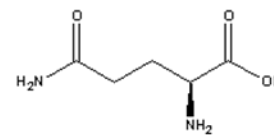
Asp



Glu



Asn

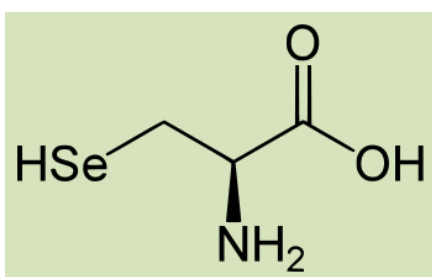


Gln

Extra aminokyseliny

21. aminokyselina: selenocysteín (Sec, U):

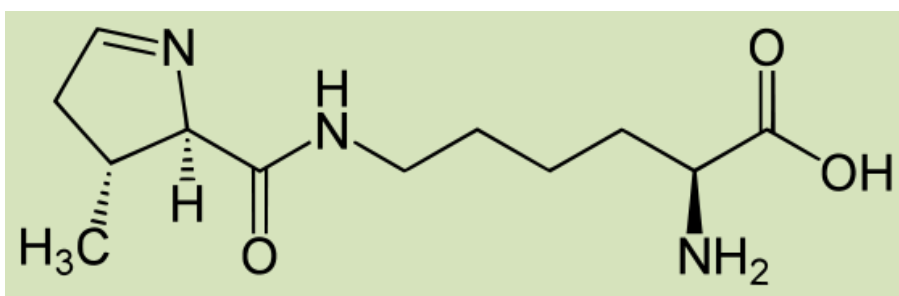
Selenocysteín sa podobá na cysteín, ale namiesto atómu síry má atóm selénu. Selenocysteín je kódovaný kodónom UGA (stop kodón; opál); zaradenie selenocysteínu je spôsobené sekvenciou v bezprostrednom okolí kodónu UGA. Vyskytuje sa napr. v glutationperoxidázach.



(Upravené podľa: <http://en.wikipedia.org/wiki/Selenocysteine>)

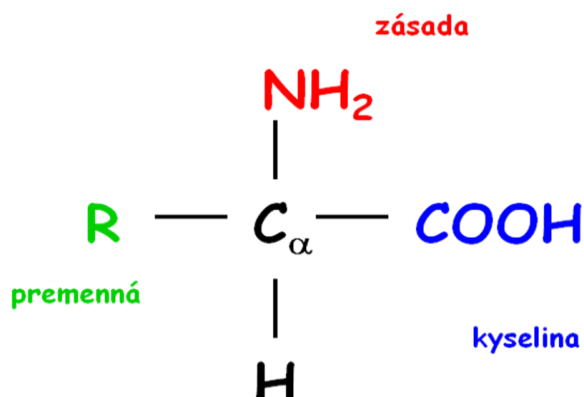
22. aminokyselina: pyrolyzín (Pyl, O):

Pyrolyzín sa podobá na lyzín, ale má pridaný pyrrolínový kruh viazaný na koniec postranného reťazca lyzínu. Pyrolyzín je kódovaný kodónom UAG (stop kodón; jantár); prirodzene sa vyskytuje v metanogénnych archaebaktériách ako súčasť ich metán-produkujúceho metabolizmu.



(Upravené podľa: <http://en.wikipedia.org/wiki/Pyrrolysine>)

Všeobecný vzorec α -aminokyseliny



R – postranný (bočný) reťazec:

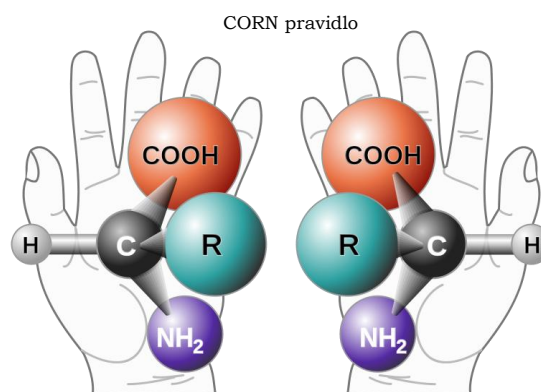
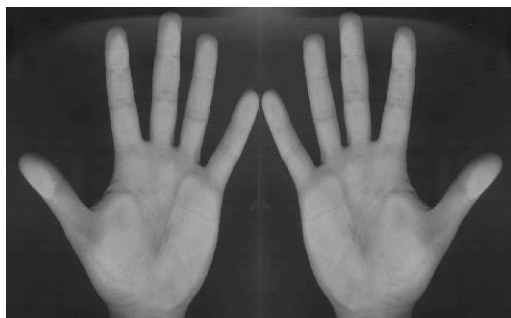
- ním sa odlišujú jednotlivé aminokyseliny;
- určuje, či je daná aminokyselina pri fyziologickom pH nabitá (protonovaná) alebo nenabitá (deprotonovaná);
- α -uhlík (C_α) je chirálny (okrem glycínu).

Všetky živé formy majú zabudované L-formy aminokyselín.

Proteín s D-formou aminokyselín by v podstate nemohol existovať, pretože by bočné reťazce boli usporiadané zlým smerom. Je to dané aj translačnou mašinériou proteosyntézy.

L-forma je zrkadlovým obrazom D-formy (rovnaký molekulový vzorec, ale podobná štruktúra).

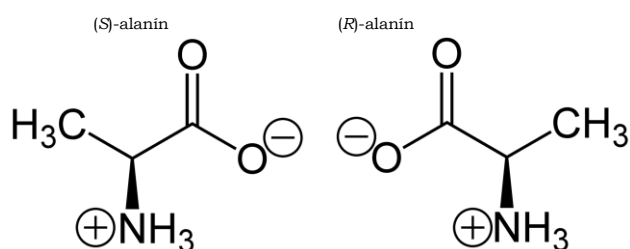
Nemožno ich preložiť jednu cez druhú.



(Upravené podľa: http://en.wikipedia.org/wiki/Chirality_%28chemistry%29)

CORN pravidlo:

- L-forma: COOH, R, NH₂ a H – v smere otáčania hodinových ručičiek
- D-forma: COOH, R, NH₂ a H – proti smeru otáčania hodinových ručičiek



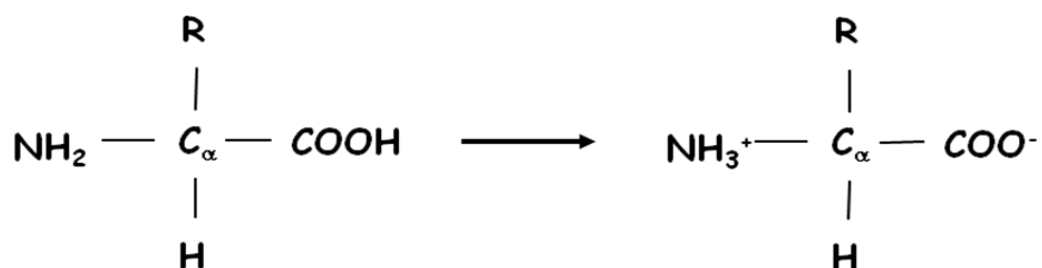
(Upravené podľa: http://en.wikipedia.org/wiki/Chirality_%28chemistry%29)

(S)-Alanín a (R)-alanín v zwitteriónovej forme (neutrálne pH).

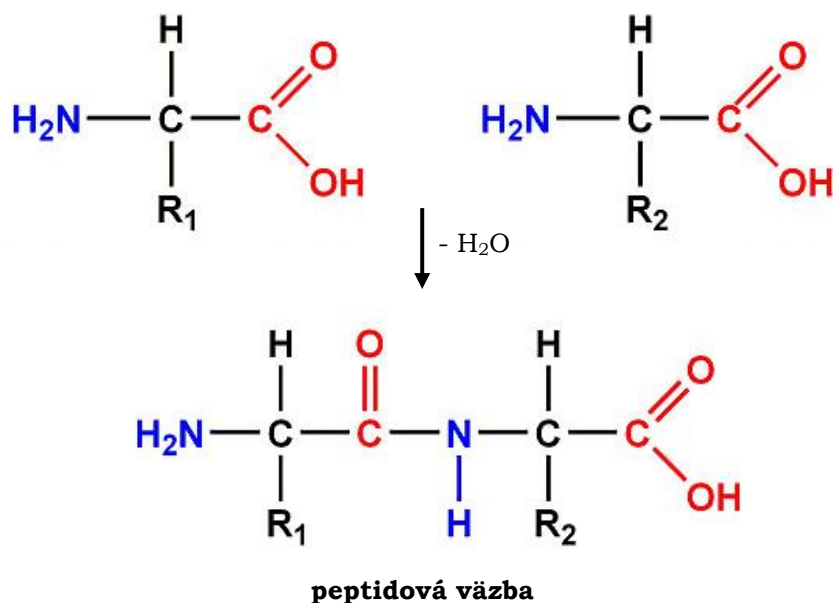
Zwitterión (z nemčiny "Zwitter" - hybrid) je zlúčenina s kyslou, aj zásaditou skupinou v tej istej molekule.

Pri neutrálnom (fyziologickom) pH zwitterión obsahuje naraz negatívne nabité anióny (COO⁻) a pozitívne nabité katióny (NH₃⁺).

Navonok však je to častica neutrálna (hoci nabitá).



2. Vznik proteínov, peptidová väzba, peptidy



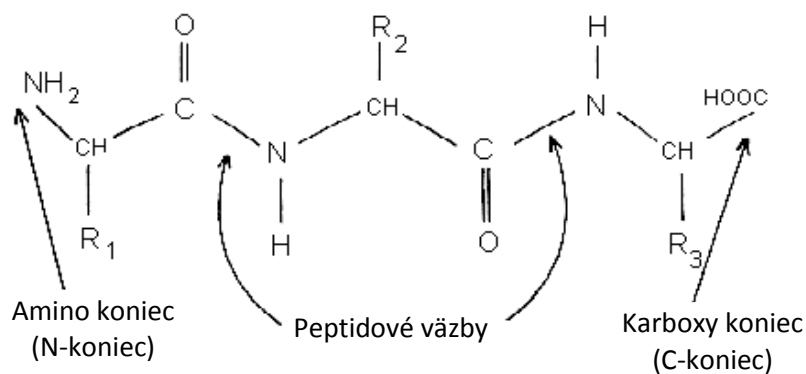
(Upravené podľa: <http://www.chemicalconnection.org.uk/chemistry/topics/view.php?topic=5&headingno=3>)

Peptidová väzba $-C(O)-N(H)-$ je chemická väzba tvorená medzi dvoma molekulami aminokyselín.

Karboxylová skupina jednej (1.) aminokyseliny reaguje s amino skupinou druhej (2.) aminokyseliny, pričom sa uvoľní molekula vody (H_2O).

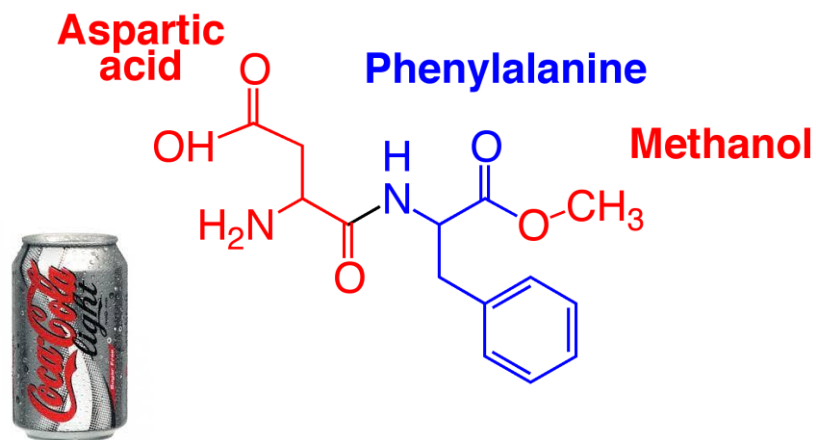
Je to dehydratácia.

- napr. tripeptid



(Upravené podľa: http://www.cryst.bbk.ac.uk/PPS95/course/3_geometry/peptide2.html)

Aspartám (umelé sladidlo) – L-aspartyl-L-fenylalanín metyl ester



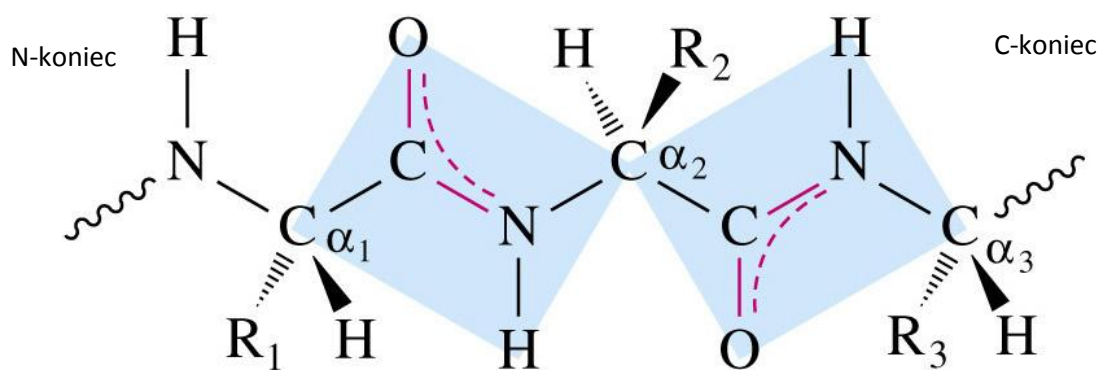
(Upravené podľa: http://www.sweetenerbook.com/aspartame_2.html)

Oligopeptidy (~ 20-50 aminokyselinových zvyškov).

Viac ako 50 zvyškov – proteín.

N-koniec (začiatok proteínu) a C-koniec (koniec proteínu).

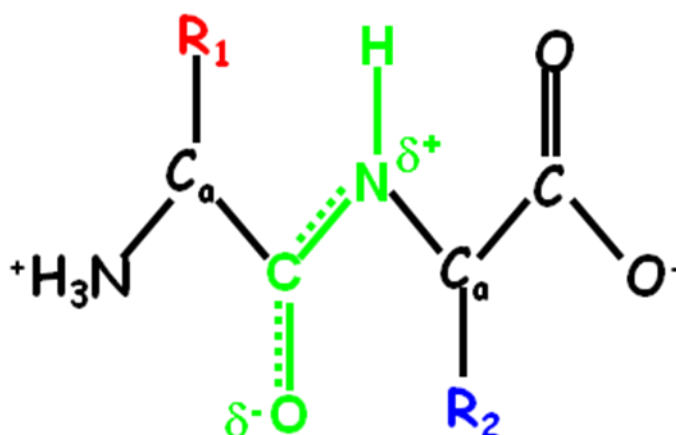
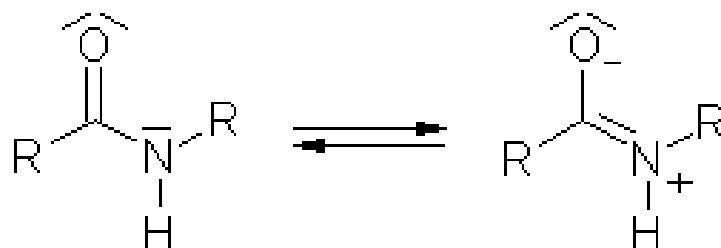
Peptidové roviny (jednotky)



(Upravené podľa: Moran a kol. *Principles of Biochemistry*, Pearson Education, Inc., Boston, 2011)

Väzba medzi karbonylovým uhlíkom a dusíkom má čiastočne charakter dvojitej väzby, rotácia je obmedzená (znemožnená).

Zároveň je peptidová jednotka planárna, pričom rotácia je dovolená na C _{α} .



Dĺžka C-N väzby je kratšia než dĺžka väzby $\text{C}_\alpha\text{-N}$.

Dĺžka C=O väzby je v skutočnosti dlhšia než normálna karbonylová väzba.

Dôsledok mezomérneho účinku atómu kyslíka na voľný elektrónový pár na atóme dusíka; vzniknuté náboje sú parciálne.

Atómy peptidovej väzby N, C, O a atómy k nim pripojené ležia v jednej rovine.

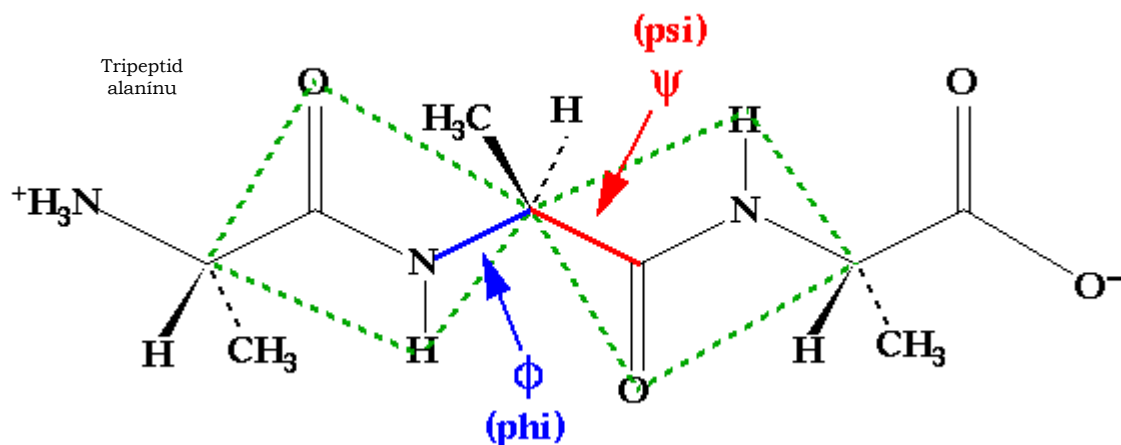
Peptidová väzba je planárna.

Rotovať môžu iba väzby $\text{C}_\alpha\text{-N}$ a $\text{C}_\alpha\text{-C}(\text{O})$ (rotácie na C_α uhlíkoch).

Dihedrálne (torzné) uhly φ a ψ

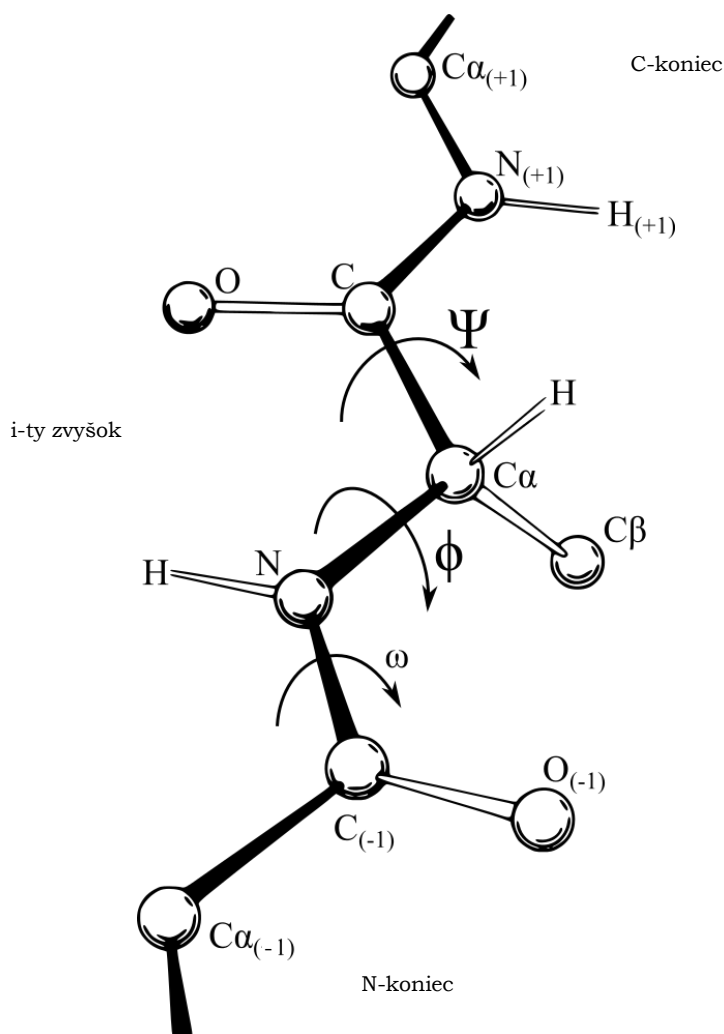
Dihedrálne uhly φ a ψ (voľné otáčanie skupín na C_α).

Jedna dvojica torzných uhlov sa vždy priraduje jednému C_α uhlíku.



(Upravené podľa: <http://www.iop.vast.ac.vn/theor/conferences/smp/1st/kaminuma/UCSFComputerGraphicsLab/AAA.html>)

Rotácie sú definované torznými uhlami ϕ (C_{α} -N väzba) a ψ (C_{α} -C(O) väzba). Hodnoty torzných uhlov sú ovplyvňované susednými aminokyselinovými zvyškami.



(Upravené podľa: http://en.wikipedia.org/wiki/Ramachandran_plot)

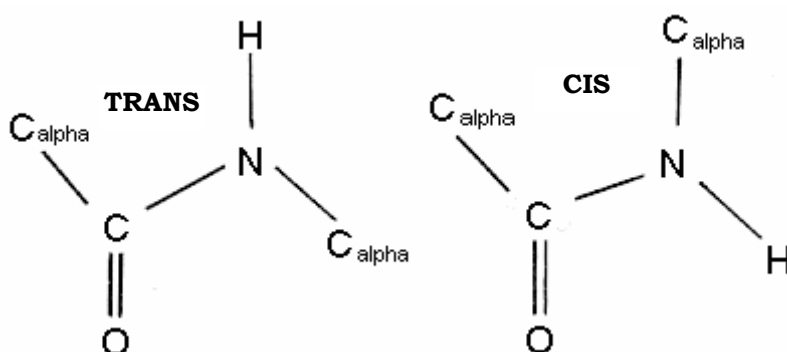
3. Primárna a sekundárna štruktúra proteínov

Definície

- primárna – poradie monomérových jednotiek, t.j. aminokyselinová sekvencia;
- sekundárna – geometrické usporiadania úsekov peptidových reťazcov (α -špirála, skladaný list – β -vlákna, ohyby, slučky);
- terciárna – priestorové usporiadanie polypeptidového reťazca (domény, motívy...);
- kvartérna – priestorové usporiadanie podjednotiek (diméry, triméry, ... oligoméry).

Vyššie typy štruktúr (sekundárna a terciárna) sú determinované primárnou štruktúrou.

Konfigurácie peptidovej väzby *cis* a *trans*



(Upravené podľa: http://ictwiki.iitk.ernet.in/wiki/index.php/Peptides_and_proteins)

Energeticky výhodnejšia je konfigurácia *trans* (~8 kJ/mol).

Konfigurácia *cis* – ak N-atóm poskytuje prolín.

Primárna štruktúra

α -Amyláza z ľudských slín

```
MKLFWLLFTIGFCWAQYSSNTQQGRTSIVHLFEWRWVDIALECERYLAPKGFGGVQVSP 60
NENVAIHNPFRPWERYQPVSYKLCTRSGNEDEFNMVTRCANNVGVRIYVDAVINHMCN 120
AVSAGTSSTCGSYFNPGRDFPAVPYSGWDFNDGKCKTGSGDIENYNDATQVRDCRLSGL 180
LDLALGKDYVRSKIAEYMNHLIDIGVAGFRIDASKHMWPGDIKAILDKLHNLNSNWFPEG 240
SKPFIYQEVIDLGGEPIKSSDYFGNGRVTEFKYGAKLGTVIRKWNGEKMSYLNWGEWG 300
FMPSDRALVFVDNHDNQRGHGAGGASILTFWDARLYKMAVGFMALHPYGFTRVMSSYRWP 360
RYFENGKDVNDWVGPPNDNGVTKEVTINPDTTGNDWVCEHRWRQIRNMVFNFRNVVDGQP 420
FTNWDNGSNQVAFGRNGRFIVFNDDWTFSLTLQTGLPAGTYCDVISGDKINGNCTGI 480
KIYVSDDGKAHFSISNSAEDPFIAIHAESKL 511
```

```
MKLFWLLFTIGFCWA sp
QYSSNTQQGRTSIVHLFEWRWVDIALECERYLAPKGFGGVQVSPNENVAIHNPFRPWWE 60
RYQPVSYKLCTRSGNEDEFNMVTRCANNVGVRIYVDAVINHMCNNAVSAGTSSTCGSYFN 120
PGSRDFPAVPYSGWDFNDGKCKTGSGDIENYNDATQVRDCRLSGLLDLALGKDYVRSKIA 180
EYMNHLIDIGVAGFRIDASKHMWPGDIKAILDKLHNLNSNWFPEGSKPFIYQEVIDLGGE 240
PIKSSDYFGNGRVTEFKYGAKLGTVIRKWNGEKMSYLNWGEWGFMPSDRALVFVDNHD 300
NQRGHGAGGASILTFWDARLYKMAVGFMALHPYGFTRVMSSYRWPYRYFENGKDVNDWVG 360
PNDNGVTKEVTINPDTTGNDWVCEHRWRQIRNMVFNFRNVVDGQPFNTNWDNGSNQVAFG 420
RGNRGFIVFNDDWTFSLTLQTGLPAGTYCDVISGDKINGNCTGIKIYVSDDGKAHFSIS 480
NSAEDPFIAIHAESKL 496
```

UniProt databáza: P04745 (<http://www.uniprot.org/>)

- accession number (prístupové číslo)

The screenshot displays the UniProt entry for P04745 - AMY1_HUMAN. The header includes the UniProt logo, a search bar, and navigation links. The main content area shows the protein name, gene, organism, and status. A sidebar on the left allows filtering by various categories. The 'Function' section describes the catalytic activity of the protein. The 'Sites' section contains a table of binding and active sites.

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Metal binding ⁱ	115 – 115	1	Calcium			
Metal binding ⁱ	173 – 173	1	Calcium; via carbonyl oxygen			
Metal binding ⁱ	182 – 182	1	Calcium			
Binding site ⁱ	210 – 210	1	Chloride			
Active site ⁱ	212 – 212	1	Nucleophile			
Metal binding ⁱ	216 – 216	1	Calcium; via carbonyl oxygen			
Active site ⁱ	248 – 248	1	Proton donor			

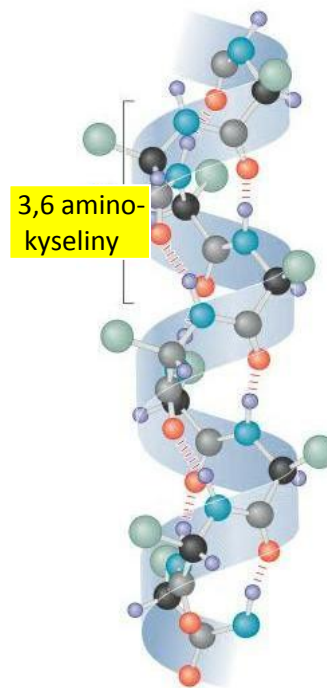
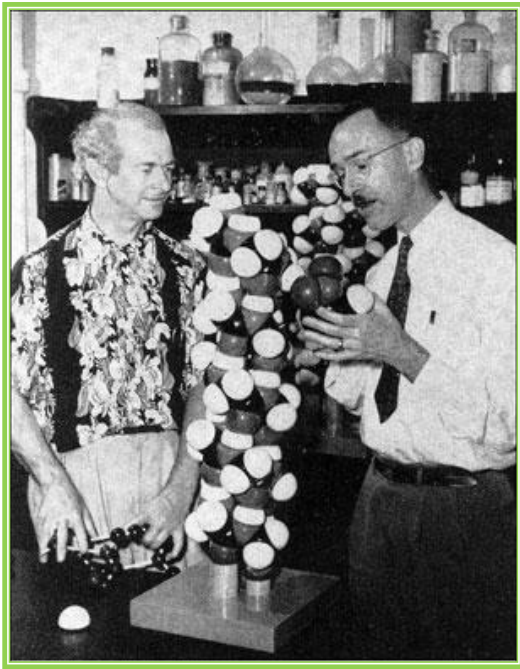
Sekundárna štruktúra

- α -helix
- β -skladaný list (β -vlákna)
- ohyby a slučky

α -Helixy spolu sú spolu s β -listami (vláknami) základné – pravidelné – elementy sekundárnej štruktúry proteínov.

Helix (špirála)

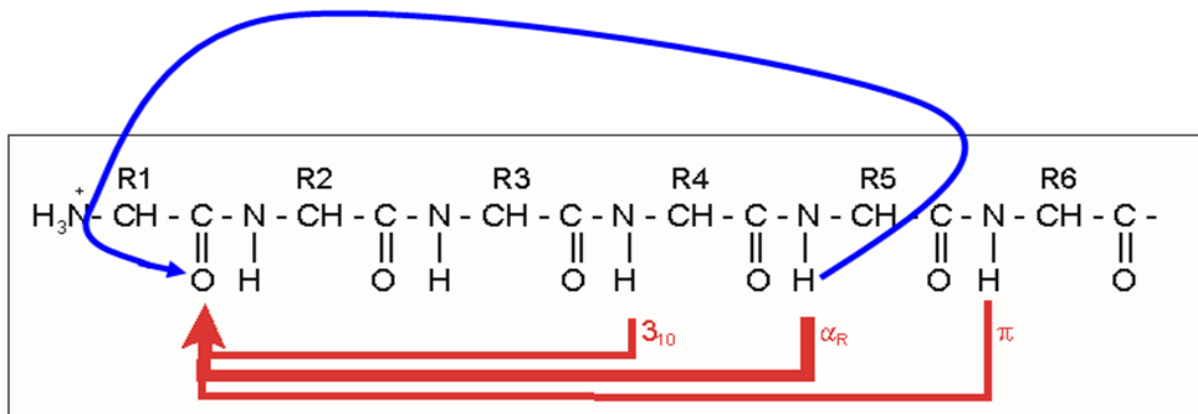
- aminokyseliny sú v α -helixe usporiadané v pravotočivej závitnicovej štruktúre (~0,5 nm šírka);
- každá aminokyselina sa zatača asi v 100° ohybe (posun ~0,15 nm pozdĺž osi špirály);
- helix je tesne zbalený, takmer bez voľného miesta;
- na jednu otáčku špirály pripadá v α -helixe 3,6 aminokyselinového zvyšku;
- všetky aminokyselinové zvyšky sú orientované von zo špirály;
- N-H skupina aminokyseliny n môže tvoriť vodíkovú väzbu s C=O skupinou aminokyseliny $n+4$;
- krátke polypeptidy zvyčajne nemôžu adoptovať α -helix (strata entropie je príliš vysoká);
- niektoré aminokyseliny – prolín a glycín – narúšajú špirálovú štruktúru (tzv. helix breakers);
- helix má pozitívny náboj na N-konci a negatívny náboj na C-konci (helixový dipól; destabilizačný vplyv); kompenzácia je obsadením N-konca negatívne nabitým zvyškom (napr. Glu) – stabilizácia dipólu. Menej bežné (a menej efektívne) je obsadenie C-konca pozitívne nabitým zvyškom (napr. Lys);
- štruktúra bola prvý-krát navrhnutá v roku 1951 (Pauling a Corey).



Linus Pauling a Robert Corey
s ich modelom α -helixu (1951).

(Upravené podľa: Eisenberg *Proc. Natl. Acad. Sci. USA* 2003, 100: 11207 a
http://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins)

α -helix = 3,6₁₃ helix (α_R) – číslo 13 udáva počet atómov medzi atómami O a H



Myoglobín

Prvý proteín, ktorého štruktúra bola vyriešená Rtg kryštalografiou (1958),
Max Perutz a John Kendrew; Nobelova cena za chémiu; terciárna štruktúra
myoglobínu obsahuje len α -špirály (a štatistické kľbko); žiadne β -listy.

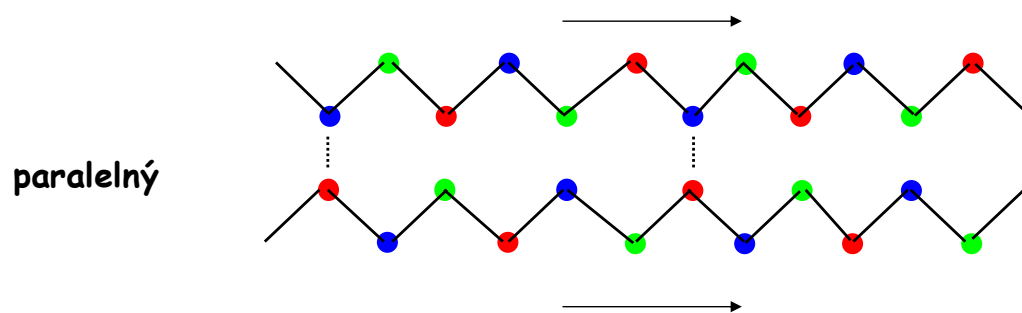
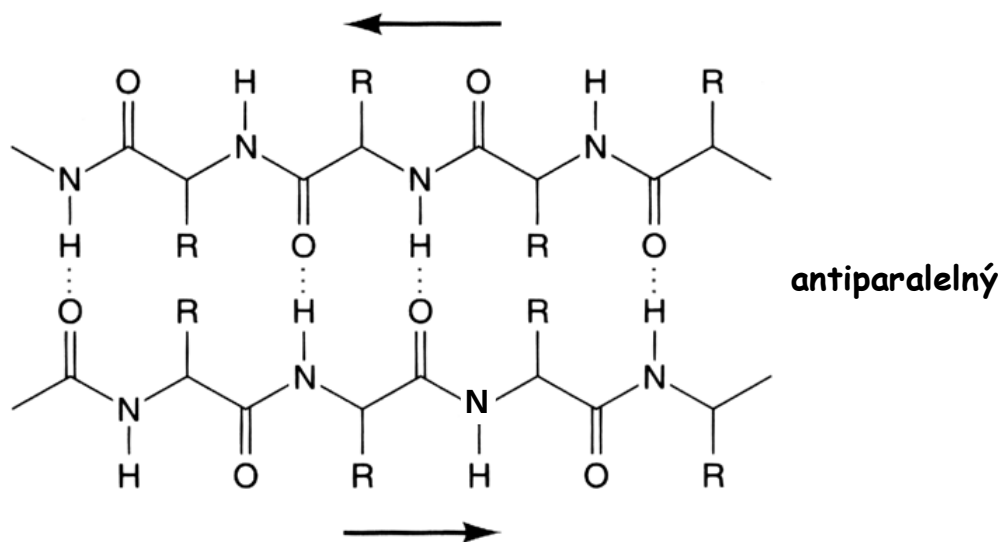
Skladaný list (β -vlákna)

- štruktúra β -listu bola tiež prvý-krát navrhnutá v roku 1951 (Pauling a Corey);
- pozostáva z dvoch a viac aminokyselinových úsekov (β -vlákien) v tom istom proteíne, ktoré sú navzájom rovnobežné a susedia spolu;
- orientácia môže byť tým istým smerom (paralelné vlákna) alebo môže alternovať (antiparalelné vlákna);
- vodíkové väzby sa tvoria medzi dvomi vláknami toho istého β -listu;
- postranné reťazce aminokyselín sú takmer úplne rozvinuté v celom β -vlákne;
- N-H skupiny kostry jedného vlákna tvoria H-väzby s C=O skupinami v kostre susediaceho vlákna (efekt sa kumuluje a prispieva k stabilite β -listu a jeho štruktúrnej rigidite);
- C $_{\alpha}$ atómy susedných β -vlákien sú od seba vzdialené 0,35 nm;
- bočné reťazce aminokyselinových zvyškov v β -listovej štruktúre môžu byť usporiadané aj tak, že na jednej strane listu sú hydrofóbne postranné reťazce, kým na protiľahlej strane listu sú polárne alebo nabité (t.j. hydrofilné) reťazce;
- niektoré sekvencie β -listu sa najmä na prolínových zvyškoch pozdĺž svojej kostry otáčajú v tzv. vlásočnicovom ohybe (*hairpin turn*);
- β -listy v skutočnosti nie sú až také otvorené štruktúry, ako naznačuje ich názov – β -vlákna možno považovať za veľmi voľne zvinuté pravotočivé špirály.

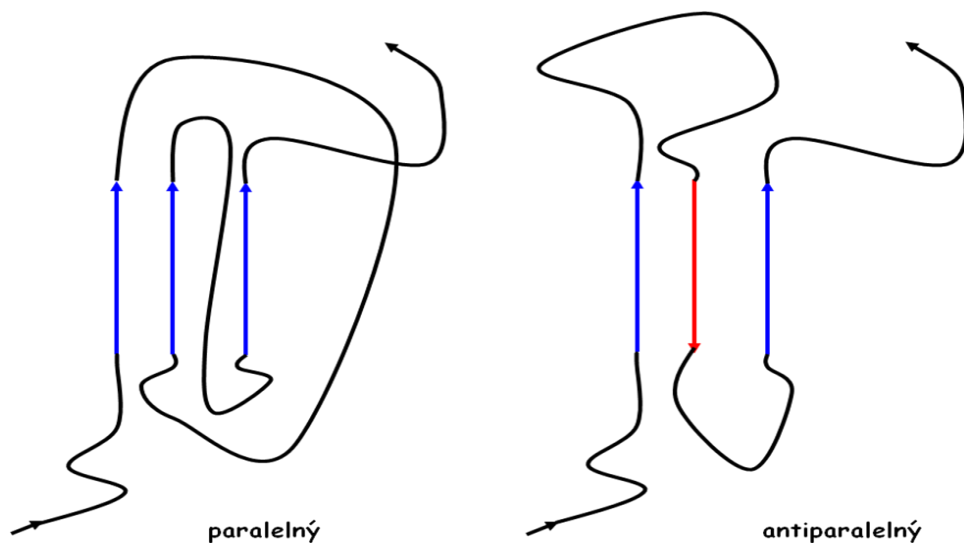
Vodíkové väzby

Väčší počet H-väzieb sa tvorí pri antiparalelnom β -liste.

Intermolekulárne H-väzby sú medzi rôznymi polypeptidovými reťazcami, kým intramolekulárne sú medzi rôznymi časťami toho istého polypeptidového reťazca.



Trojvláknový β -list



4. Terciárna a kvartérna štruktúra proteínov

Terciárna štruktúra

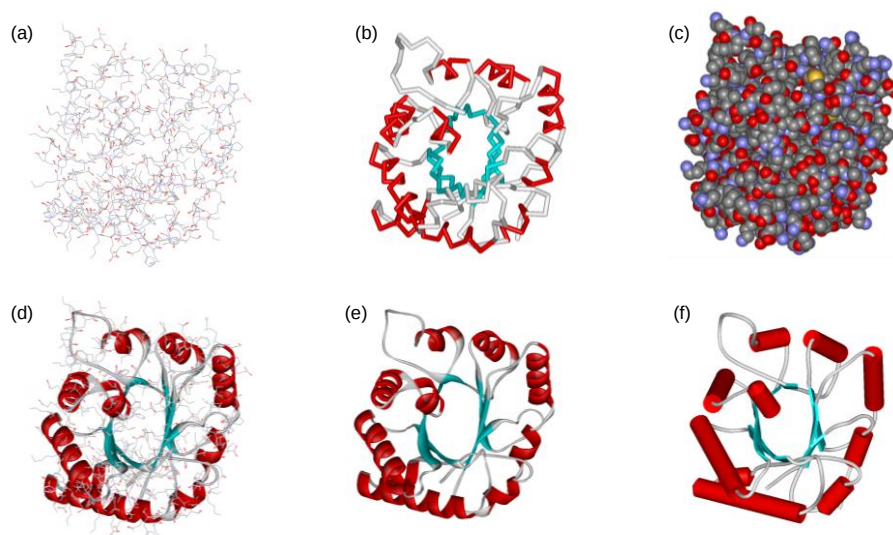
Je to priestorová štruktúra proteínu, jeho molekuly.

Vyriešiť, resp. poznať terciárnu štruktúru proteínu znamená získať pozíciu – t.j. koordináty v priestore (x, y, z) – pre každý atóm, ktorý tvorí molekulu proteínu. Najdôležitejšie sú atómy tvoriace kostru polypeptidového reťazca, t.j. atómy N, C_α a C(O), ale rovnako je potrebné určiť koordináty pre všetky ostatné atómy (okrem atómov vodíka – kvôli malej veľkosti).

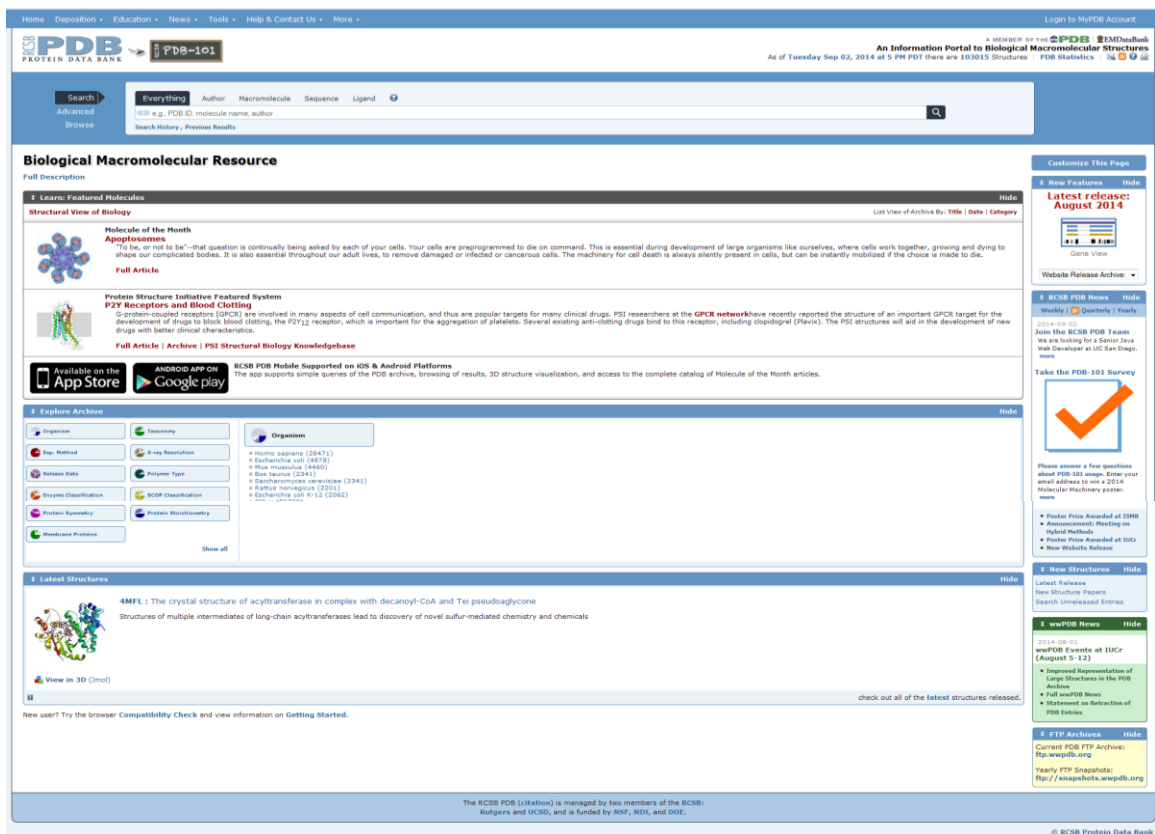
Priestorová štruktúra proteínu sa určuje (rieši):

- metódami röntgenovej kryštalografie (proteín v tuhom stave);
- metódami NMR spektroskopie (v roztoku; menšie proteíny).

Pre metódy röntgenovej kryštalografie je potrebné získať proteín vysokej čistoty a následne kryštál, ktorý vhodne difraktuje. Presnosť stanovenia štruktúry (rozlíšenie) sa udáva v „Å“ (10⁻¹⁰ m). Čím je hodnota v „Å“ nižšia, tým je štruktúra určená presnejšie (t.j. s väčším rozlíšením).

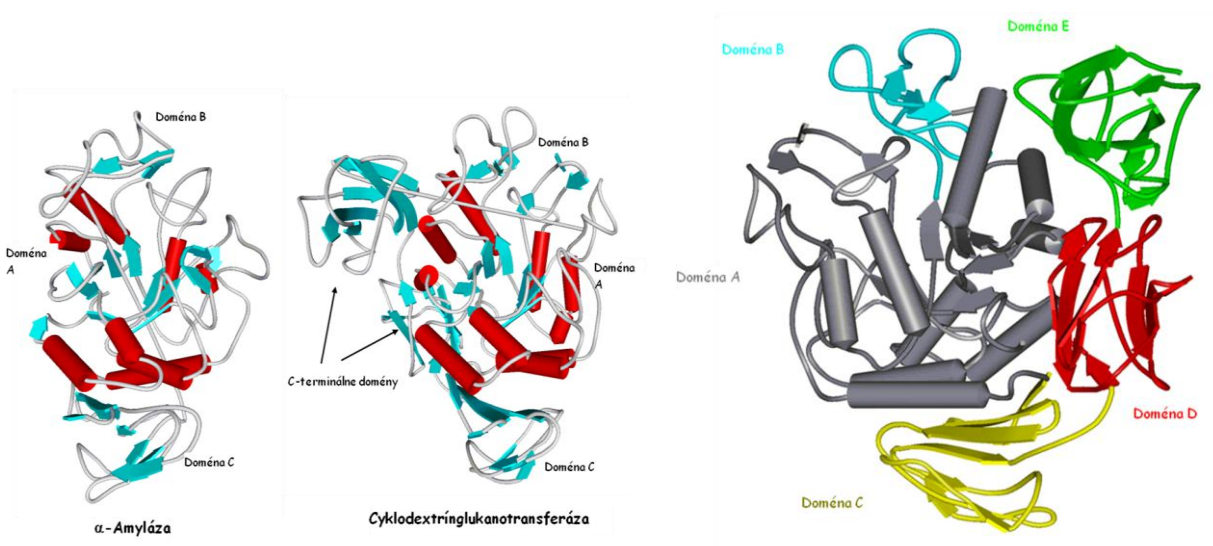


Štruktúra triózafozfátizomerázy (PDB kód: 1TIM). Zobrazenia: (a) atómový model, (b) C_α kostra (c) „CPK“ model, (d) model stuhy so zapnutými atómami, (e) model stuhy s vypnutými atómami, (f) schéma - α-helixy a β-vlákná.



Databáza PDB (Protein Data Bank) – <http://www.rcsb.org/pdb/>

Vyriešené štruktúry jednotlivých proteínov sú v PDB uložené pod tzv. PDB (prístupovými) kódmi.



Trojdoménnová α -amyláza (PDB kód: 2TAA) a 5-doménnová cyklodextrínoglukanotransferáza (PDB kód: 1CDG); vpravo sú farebne odlišené jednotlivé domény A-E.

[Home](#) [Deposition](#) [Education](#) [News](#) [Tools](#) [Help & Contact Us](#) [More](#)

PDB PROTEIN DATA BANK

PDB-101

An Information Portal to Biological Macromolecular Structures
As of Tuesday Sep 09, 2014 at 5 PM EDT there are 103199 Structures PDB Statistics

Login to PDB Account

Search: **Everything** Author Macromolecule Sequence Ligand ⓘ
 e.g. PDB ID, molecule name, author
[Search History](#) [Previous Results](#)

[Advanced](#) [Browse](#)

Summary [3D View](#) [Sequence](#) [Annotations](#) [Gap Prediction](#) [2D Domains](#) [Connectors](#) [Aids & Downloads](#) [Related Links](#)

STRUCTURE AND POSSIBLE CATALYTIC RESIDUES OF TAKA-AMYLASE A

DOI:10.2310/jpdb.taa/pdb

ENTRY 2TAA SUPERSEDES 1TAA

Primary Citation

Structure and possible catalytic residues of Taka-amylase A.
Hatawara, Y., Kawasumi, H., Harada, W., Kakuho, H.
Jour Biol Chem (1984) 259(1):50-52

PubMed 6559312 ⓘ
[Search Related Articles in PubMed](#) ⓘ

Published Abstract:
 A tentative molecular model of Taka-amylase A consisting of 476 amino acid residue was built with the aid of amino acid sequence data. Some typical structural features of the molecule are described. A model fitting of an enzyme chain in the catalytic site of the enzyme showed a possible productive binding mode between substrate and enzyme. On the basis of the difference Fourier study, glutamic acid (Glu230) and aspartic acid (Asp297), which are located at the bottom of the cleft, were concluded to be the catalytic residues, serving as the general acid and base, respectively.

Keywords:
 Binding Sites, Catalysis, Hydrogen-Ion Concentration, Models, Molecular, Peptide Fragments, Structure-Activity Relationship, Substrate Specificity, Alpha-Amylases

Click on abstract words and keywords to add them to the search box.
[\[Search Published Abstracts\]](#)

[Hide Abstract]

1 Molecular Description

Classification: [Get information on Classification](#)

3DResource Weight: 1.97422 AN 13

Name	Type	Length
Protein	protein	476

Chemical Component

EC: 3.2.1.1 ⓘ
Enzyme Class: Amylogalactosidase ⓘ
Gene Name: amy1 amy1 Taa-G1 AG09023000944 amy1 Taa-G2 AG090120000196
UniprotRef: ⓘ
[Protein Feature View](#) | [Search PDB](#) | PDB183 ⓘ

Protein Processing

Processing: signal → 3D,SS-alpha-system

UP sites: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 6

```

HEADER          HYDROLASE (O-GLYCOSYL)                  18-OCT-82   2TAA
TITLE           STRUCTURE AND POSSIBLE CATALYTIC RESIDUES OF TAKA-AMYLASE A
COMPND          MOL_ID: 1;
COMPND          2 MOLECULE: TAKA-AMYLASE A;
COMPND          3 CHAIN: A, B, C;
COMPND          4 EC: 3.2.1.1;
COMPND          5 ENGINEERED: YES
SOURCE          MOL_ID: 1;
SOURCE          2 ORGANISM_SCIENTIFIC: ASPERGILLUS ORYZAE;
SOURCE          3 ORGANISM_TAXID: 5062
KEYWDS          HYDROLASE (O-GLYCOSYL)
EXPDTA          X-RAY DIFFRACTION
AUTHOR          M.KUSUNOKI,Y.MATSUURA,N.TANAKA,M.KAKUDO
REVDAT          8   13-JUL-11  2TAA   1   VERSN
REVDAT          7   24-FEB-09  2TAA   1   VERSN
REVDAT          6   15-OCT-89  2TAA   3   MTRIX
REVDAT          5   22-OCT-84  2TAA   1   SEQRES SHEET
REVDAT          4   17-JUL-84  2TAA   2   REMARK SSBOND CONECT
REVDAT          3   31-JAN-84  2TAA   1   REMARK
REVDAT          2   30-SEP-83  2TAA   1   REVDAT
REVDAT          1   21-OCT-82  2TAA   0
SPRSDE          21-OCT-82  2TAA   1TAA
JRNL            AUTH   Y.MATSUURA,M.KUSUNOKI,W.HARADA,M.KAKUDO
JRNL            TITL   STRUCTURE AND POSSIBLE CATALYTIC RESIDUES OF TAKA-AMYLASE A
JRNL            REF    J.BIOCHEM.(TOKYO)                V. 95    697 1984
JRNL            REFN                               ISSN 0021-924X
JRNL            PMID   6609921
REMARK          1
REMARK          1 REFERENCE 1
REMARK          1 AUTH   Y.MATSUURA,M.KUSUNOKI,W.HARADA,N.TANAKA,Y.IGA,N.YASUOKA,
REMARK          1 AUTH 2 H.TODA,K.NARITA,M.KAKUDO

```

REMARK 1 TITL MOLECULAR STRUCTURE OF TAKA-AMYLASE A. I. BACKBONE CHAIN
 REMARK 1 TITL 2 FOLDING AT 3 ANGSTROMS RESOLUTION
 REMARK 1 REF J.BIOCHEM.(TOKYO) V. 87 1555 1980
 REMARK 1 REFN ISSN 0021-924X
 REMARK 1 REFERENCE 2
 REMARK 1 AUTH Y.MATSUURA,M.KUSUNOKI,W.DATE,S.HARADA,S.BANDO,N.TANAKA,
 REMARK 1 AUTH 2 M.KAKUDO
 REMARK 1 TITL LOW RESOLUTION CRYSTAL STRUCTURES OF TAKA-AMYLASE A AND ITS
 REMARK 1 TITL 2 COMPLEXES WITH INHIBITORS
 REMARK 1 REF J.BIOCHEM.(TOKYO) V. 86 1773 1979
 REMARK 1 REFN ISSN 0021-924X
 REMARK 2
 REMARK 2 RESOLUTION. 3.00 ANGSTROMS.
 REMARK 3
 REMARK 3 REFINEMENT.
 REMARK 3 PROGRAM : NULL
 REMARK 3 AUTHORS : NULL
 REMARK 3
 REMARK 3 DATA USED IN REFINEMENT.
 REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 3.00
 REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS) : NULL
 REMARK 3 DATA CUTOFF (SIGMA(F)) : NULL
 REMARK 3 DATA CUTOFF HIGH (ABS(F)) : NULL
 REMARK 3 DATA CUTOFF LOW (ABS(F)) : NULL
 REMARK 3 COMPLETENESS (WORKING+TEST) (%) : NULL
 REMARK 3 NUMBER OF REFLECTIONS : NULL
 REMARK 3
 REMARK 3 FIT TO DATA USED IN REFINEMENT.
 REMARK 3 CROSS-VALIDATION METHOD : NULL
 REMARK 3 FREE R VALUE TEST SET SELECTION : NULL
 REMARK 3 R VALUE (WORKING SET) : NULL
 REMARK 3 FREE R VALUE : NULL
 REMARK 3 FREE R VALUE TEST SET SIZE (%) : NULL
 REMARK 3 FREE R VALUE TEST SET COUNT : NULL
 REMARK 3 ESTIMATED ERROR OF FREE R VALUE : NULL
 REMARK 3
 REMARK 3 FIT IN THE HIGHEST RESOLUTION BIN.
 REMARK 3 TOTAL NUMBER OF BINS USED : NULL
 REMARK 3 BIN RESOLUTION RANGE HIGH (A) : NULL
 REMARK 3 BIN RESOLUTION RANGE LOW (A) : NULL
 REMARK 3 BIN COMPLETENESS (WORKING+TEST) (%) : NULL
 REMARK 3 REFLECTIONS IN BIN (WORKING SET) : NULL
 REMARK 3 BIN R VALUE (WORKING SET) : NULL
 REMARK 3 BIN FREE R VALUE : NULL
 REMARK 3 BIN FREE R VALUE : NULL
 REMARK 3 BIN FREE R VALUE TEST SET SIZE (%) : NULL
 REMARK 3 BIN FREE R VALUE TEST SET COUNT : NULL
 REMARK 3 ESTIMATED ERROR OF BIN FREE R VALUE : NULL
 REMARK 3
 REMARK 3 NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
 REMARK 3 PROTEIN ATOMS : 11070
 REMARK 3 NUCLEIC ACID ATOMS : 0
 REMARK 3 HETEROGEN ATOMS : 3
 REMARK 3 SOLVENT ATOMS : 0
 REMARK 3
 REMARK 3 B VALUES.
 REMARK 3 FROM WILSON PLOT (A**2) : NULL
 REMARK 3 MEAN B VALUE (OVERALL, A**2) : NULL
 REMARK 3 OVERALL ANISOTROPIC B VALUE.
 REMARK 3 B11 (A**2) : NULL
 REMARK 3 B22 (A**2) : NULL
 REMARK 3 B33 (A**2) : NULL
 REMARK 3 B12 (A**2) : NULL
 REMARK 3 B13 (A**2) : NULL
 REMARK 3 B23 (A**2) : NULL
 REMARK 3
 REMARK 3 ESTIMATED COORDINATE ERROR.
 REMARK 3 ESD FROM LUZZATI PLOT (A) : NULL
 REMARK 3 ESD FROM SIGMAA (A) : NULL
 REMARK 3 LOW RESOLUTION CUTOFF (A) : NULL
 REMARK 3
 REMARK 3 CROSS-VALIDATED ESTIMATED COORDINATE ERROR.
 REMARK 3 ESD FROM C-V LUZZATI PLOT (A) : NULL

```

REMARK 3 ESD FROM C-V SIGMAA (A) : NULL
REMARK 3
REMARK 3 RMS DEVIATIONS FROM IDEAL VALUES.
REMARK 3 BOND LENGTHS (A) : NULL
REMARK 3 BOND ANGLES (DEGREES) : NULL
REMARK 3 DIHEDRAL ANGLES (DEGREES) : NULL
REMARK 3 IMPROPER ANGLES (DEGREES) : NULL
REMARK 3
REMARK 3 ISOTROPIC THERMAL MODEL : NULL
REMARK 3
REMARK 3 ISOTROPIC THERMAL FACTOR RESTRAINTS. RMS SIGMA
REMARK 3 MAIN-CHAIN BOND (A**2) : NULL ; NULL
REMARK 3 MAIN-CHAIN ANGLE (A**2) : NULL ; NULL
REMARK 3 SIDE-CHAIN BOND (A**2) : NULL ; NULL
REMARK 3 SIDE-CHAIN ANGLE (A**2) : NULL ; NULL
REMARK 3
REMARK 3 NCS MODEL : NULL
REMARK 3
REMARK 3 NCS RESTRAINTS. RMS SIGMA/WEIGHT
REMARK 3 GROUP 1 POSITIONAL (A) : NULL ; NULL
REMARK 3 GROUP 1 B-FACTOR (A**2) : NULL ; NULL
REMARK 3
REMARK 3 PARAMETER FILE 1 : NULL
REMARK 3 TOPOLOGY FILE 1 : NULL
REMARK 3
REMARK 3 OTHER REFINEMENT REMARKS: NULL
REMARK 4
REMARK 4 2TAA COMPLIES WITH FORMAT V. 3.30, 13-JUL-11
REMARK 100
REMARK 100 THIS ENTRY HAS BEEN PROCESSED BY BNL.
REMARK 200
REMARK 200 EXPERIMENTAL DETAILS
REMARK 200 EXPERIMENT TYPE : X-RAY DIFFRACTION
REMARK 200 DATE OF DATA COLLECTION : NULL
REMARK 200 TEMPERATURE (KELVIN) : NULL
REMARK 200 PH : NULL
REMARK 200 NUMBER OF CRYSTALS USED : NULL
REMARK 200
REMARK 200 SYNCHROTRON (Y/N) : NULL
REMARK 200 RADIATION SOURCE : NULL
REMARK 200 BEAMLINE : NULL
REMARK 200 X-RAY GENERATOR MODEL : NULL
REMARK 200 MONOCHROMATIC OR LAUE (M/L) : NULL
REMARK 200 WAVELENGTH OR RANGE (A) : NULL
REMARK 200 MONOCHROMATOR : NULL
REMARK 200 OPTICS : NULL
REMARK 200
REMARK 200 DETECTOR TYPE : NULL
REMARK 200 DETECTOR MANUFACTURER : NULL
REMARK 200 INTENSITY-INTEGRATION SOFTWARE : NULL
REMARK 200 DATA SCALING SOFTWARE : NULL
REMARK 200
REMARK 200 NUMBER OF UNIQUE REFLECTIONS : NULL
REMARK 200 RESOLUTION RANGE HIGH (A) : NULL
REMARK 200 RESOLUTION RANGE LOW (A) : NULL
REMARK 200 REJECTION CRITERIA (SIGMA(I)) : NULL
REMARK 200
REMARK 200 OVERALL.
REMARK 200 COMPLETENESS FOR RANGE (%) : NULL
REMARK 200 DATA REDUNDANCY : NULL
REMARK 200 R MERGE (I) : NULL
REMARK 200 R SYM (I) : NULL
REMARK 200 <I/SIGMA(I)> FOR THE DATA SET : NULL
REMARK 200
REMARK 200 IN THE HIGHEST RESOLUTION SHELL.
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE HIGH (A) : NULL
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE LOW (A) : NULL
REMARK 200 COMPLETENESS FOR SHELL (%) : NULL
REMARK 200 DATA REDUNDANCY IN SHELL : NULL
REMARK 200 R MERGE FOR SHELL (I) : NULL
REMARK 200 R SYM FOR SHELL (I) : NULL
REMARK 200 <I/SIGMA(I)> FOR SHELL : NULL

```

REMARK 200
REMARK 200 DIFFRACTION PROTOCOL: NULL
REMARK 200 METHOD USED TO DETERMINE THE STRUCTURE: NULL
REMARK 200 SOFTWARE USED: NULL
REMARK 200 STARTING MODEL: NULL
REMARK 200
REMARK 200 REMARK: NULL
REMARK 280
REMARK 280 CRYSTAL
REMARK 280 SOLVENT CONTENT, VS (%): 65.65
REMARK 280 MATTHEWS COEFFICIENT, VM (ANGSTROMS**3/DA): 3.58
REMARK 280
REMARK 280 CRYSTALLIZATION CONDITIONS: NULL
REMARK 285
REMARK 285 THE ENTRY COORDINATES
REMARK 285 ARE NOT PRESENTED IN THE STANDARD CRYSTAL FRAME.
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY
REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: P 1 21 1
REMARK 290
REMARK 290 SYMOP SYMMETRY
REMARK 290 NNNMMM OPERATOR
REMARK 290 1555 X,Y,Z
REMARK 290 2555 -X,Y+1/2,-Z
REMARK 290
REMARK 290 WHERE NNN -> OPERATOR NUMBER
REMARK 290 MMM -> TRANSLATION VECTOR
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS
REMARK 290 THE FOLLOWING TRANSFORMATIONS OPERATE ON THE ATOM/HETATM
REMARK 290 RECORDS IN THIS ENTRY TO PRODUCE CRYSTALLOGRAPHICALLY
REMARK 290 RELATED MOLECULES.
REMARK 290 SMTRY1 1 1.000000 0.000000 0.000000 0.000000
REMARK 290 SMTRY2 1 0.000000 1.000000 0.000000 0.000000
REMARK 290 SMTRY3 1 0.000000 0.000000 1.000000 0.000000
REMARK 290 SMTRY1 2 -1.000000 0.000000 0.000000 0.000000
REMARK 290 SMTRY2 2 0.000000 1.000000 0.000000 66.66667
REMARK 290 SMTRY3 2 0.000000 0.000000 -1.000000 0.000000
REMARK 290
REMARK 290 REMARK: NULL
REMARK 300
REMARK 300 BIOMOLECULE: 1, 2, 3
REMARK 300 SEE REMARK 350 FOR THE AUTHOR PROVIDED AND/OR PROGRAM
REMARK 300 GENERATED ASSEMBLY INFORMATION FOR THE STRUCTURE IN
REMARK 300 THIS ENTRY. THE REMARK MAY ALSO PROVIDE INFORMATION ON
REMARK 300 BURIED SURFACE AREA.
REMARK 300 REMARK: THE CRYSTALLOGRAPHIC ASYMMETRIC UNIT FOR THIS ENTRY
REMARK 300 CONTAINS THREE AMYLASE MOLECULES. THEY CAN BE GENERATED
REMARK 300 FROM CHAIN A OF THIS ENTRY BY APPLYING THE
REMARK 300 NON-CRYSTALLOGRAPHIC THREE-FOLD SCREW AXIS GIVEN BY THE
REMARK 300 MTRIX RECORDS BELOW.
REMARK 350
REMARK 350 COORDINATES FOR A COMPLETE MULTIMER REPRESENTING THE KNOWN
REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF THE
REMARK 350 MOLECULE CAN BE GENERATED BY APPLYING BIOMT TRANSFORMATIONS
REMARK 350 GIVEN BELOW. BOTH NON-CRYSTALLOGRAPHIC AND
REMARK 350 CRYSTALLOGRAPHIC OPERATIONS ARE GIVEN.
REMARK 350
REMARK 350 BIOMOLECULE: 1
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: MONOMERIC
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A
REMARK 350 BIOMT1 1 1.000000 0.000000 0.000000 0.000000
REMARK 350 BIOMT2 1 0.000000 1.000000 0.000000 0.000000
REMARK 350 BIOMT3 1 0.000000 0.000000 1.000000 0.000000
REMARK 350
REMARK 350 BIOMOLECULE: 2
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: MONOMERIC
REMARK 350 APPLY THE FOLLOWING TO CHAINS: B
REMARK 350 BIOMT1 1 1.000000 0.000000 0.000000 0.000000
REMARK 350 BIOMT2 1 0.000000 1.000000 0.000000 0.000000
REMARK 350 BIOMT3 1 0.000000 0.000000 1.000000 0.000000
REMARK 350

REMARK 350 BIOMOLECULE: 3
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: MONOMERIC
REMARK 350 APPLY THE FOLLOWING TO CHAINS: C
REMARK 350 BIOMT1 1 1.000000 0.000000 0.000000 0.000000
REMARK 350 BIOMT2 1 0.000000 1.000000 0.000000 0.000000
REMARK 350 BIOMT3 1 0.000000 0.000000 1.000000 0.000000
REMARK 620
REMARK 620 METAL COORDINATION
REMARK 620 (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER;
REMARK 620 SSEQ=SEQUENCE NUMBER; I=INSERTION CODE):
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 CA A 479 CA
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 ASP A 175 OD2
REMARK 620 2 ASN A 121 OD1 92.1
REMARK 620 3 ASP A 175 OD1 43.0 81.5
REMARK 620 N 1 2
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 CA B 479 CA
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 ASP A 476 O
REMARK 620 2 ASN B 121 OD1 56.1
REMARK 620 3 SER A 477 N 1.2 56.5
REMARK 620 4 SER A 477 O 2.8 58.7 3.1
REMARK 620 5 ASP B 175 OD1 101.8 81.5 100.8 103.6
REMARK 620 6 ASP B 175 OD2 73.7 92.1 72.4 74.0 43.0
REMARK 620 7 SER A 477 OG 2.2 56.5 3.3 2.5 104.0 75.6
REMARK 620 N 1 2 3 4 5 6
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 CA C 479 CA
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 ASN C 121 OD1
REMARK 620 2 ASP C 175 OD1 81.5
REMARK 620 3 ASP C 175 OD2 92.1 43.0
REMARK 620 N 1 2
REMARK 700
REMARK 700 SHEET
REMARK 700 THE NINE-STRANDED SHEET BS1 DESCRIBED BELOW IS ACTUALLY
REMARK 700 AN EIGHT-STRANDED BETA BARREL. THIS IS DENOTED BY THE
REMARK 700 FIRST STRAND RECURRING AS THE LAST STRAND.
REMARK 800
REMARK 800 SITE
REMARK 800 SITE_IDENTIFIER: AC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE CA A 479
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC2
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE CA B 479
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC3
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE CA C 479
DBREF 2TAA A 1 478 UNP P10529 AMYA_ASPOR 22 499
DBREF 2TAA B 1 478 UNP P10529 AMYA_ASPOR 22 499
DBREF 2TAA C 1 478 UNP P10529 AMYA_ASPOR 22 499
SEQRES 1 A 478 ALA THR PRO ALA ASP TRP ARG SER GLN SER ILE TYR PHE
SEQRES 2 A 478 LEU LEU THR ASP ARG PHE ALA ARG THR ASP GLY SER THR
SEQRES 3 A 478 THR ALA THR CYS ASN THR ALA ASP GLN LYS TYR CYS GLY
SEQRES 4 A 478 GLY THR TRP GLN GLY ILE ILE ASP LYS LEU ASP TYR ILE
SEQRES 5 A 478 GLN GLY MET GLY PHE THR ALA ILE TRP ILE THR PRO VAL
SEQRES 6 A 478 THR ALA GLN LEU PRO GLN ASP CYS ALA TYR GLY ASP ALA
SEQRES 7 A 478 TYR THR GLY TYR TRP GLN THR ASP ILE TYR SER LEU ASN
SEQRES 8 A 478 GLU ASN TYR GLY THR ALA ASP ASP LEU LYS ALA LEU SER
SEQRES 9 A 478 SER ALA LEU HIS GLU ARG GLY MET TYR LEU MET VAL ASP
SEQRES 10 A 478 VAL VAL ALA ASN HIS MET GLY TYR ASP GLY ALA GLY SER
SEQRES 11 A 478 SER VAL ASP TYR SER VAL PHE LYS PRO PHE SER SER GLN
SEQRES 12 A 478 ASP TYR PHE HIS PRO PHE CYS PHE ILE GLN ASN TYR GLU
SEQRES 13 A 478 ASP GLN THR GLN VAL GLU ASP CYS TRP LEU GLY ASP ASN

SEQRES	14	A	478	THR	VAL	SER	LEU	PRO	ASP	LEU	ASP	THR	THR	LYS	ASP	VAL
SEQRES	15	A	478	VAL	LYS	ASN	GLU	TRP	TYR	ASP	TRP	VAL	GLY	SER	LEU	VAL
SEQRES	16	A	478	SER	ASN	TYR	SER	ILE	ASP	GLY	LEU	ARG	ILE	ASP	THR	VAL
SEQRES	17	A	478	LYS	HIS	VAL	GLN	LYS	ASP	PHE	TRP	PRO	GLY	TYR	ASN	LYS
SEQRES	18	A	478	ALA	ALA	GLY	VAL	TYR	CYS	ILE	GLY	GLU	VAL	LEU	ASP	GLY
SEQRES	19	A	478	ASP	PRO	ALA	TYR	THR	CYS	PRO	TYR	GLN	ASN	VAL	MET	ASP
SEQRES	20	A	478	GLY	VAL	LEU	ASN	TYR	PRO	ILE	TYR	TYR	PRO	LEU	LEU	ASN
SEQRES	21	A	478	ALA	PHE	LYS	SER	THR	SER	GLY	SER	MET	ASP	ASP	LEU	TYR
SEQRES	22	A	478	ASN	MET	ILE	ASN	THR	VAL	LYS	SER	ASP	CYS	PRO	ASP	SER
SEQRES	23	A	478	THR	LEU	LEU	GLY	THR	PHE	VAL	GLU	ASN	HIS	ASP	ASN	PRO
SEQRES	24	A	478	ARG	PHE	ALA	SER	TYR	THR	ASN	ASP	ILE	ALA	LEU	ALA	LYS
SEQRES	25	A	478	ASN	VAL	ALA	ALA	PHE	ILE	ILE	LEU	ASN	ASP	GLY	LEU	PRO
SEQRES	26	A	478	ILE	ILE	TYR	ALA	GLY	GLN	GLU	GLN	HIS	TYR	ALA	GLY	GLY
SEQRES	27	A	478	ASN	ASP	PRO	ALA	ASN	ARG	GLU	ALA	THR	TRP	LEU	SER	GLY
SEQRES	28	A	478	TYR	PRO	THR	ASP	SER	GLU	LEU	TYR	LYS	LEU	ILE	ALA	SER
SEQRES	29	A	478	ALA	ASN	ALA	ILE	ARG	ASN	TYR	ALA	ILE	SER	LYS	ASP	THR
SEQRES	30	A	478	GLY	PHE	VAL	THR	TYR	LYS	ASN	PRO	TYR	ILE	LYS	ASP	ASP
SEQRES	31	A	478	THR	THR	ILE	ALA	MET	ARG	LYS	GLY	THR	ASP	GLY	SER	GLN
SEQRES	32	A	478	ILE	VAL	THR	ILE	LEU	SER	ASN	LYS	GLY	ALA	SER	GLY	ASP
SEQRES	33	A	478	SER	TYR	THR	LEU	SER	LEU	SER	GLY	ALA	SER	TYR	THR	ALA
SEQRES	34	A	478	GLY	GLN	GLN	LEU	THR	GLU	VAL	ILE	GLY	CYS	THR	THR	VAL
SEQRES	35	A	478	THR	VAL	GLY	SER	ASP	GLY	ASN	VAL	PRO	VAL	PRO	MET	ALA
SEQRES	36	A	478	GLY	GLY	LEU	PRO	ARG	VAL	LEU	TYR	PRO	THR	GLU	LYS	LEU
SEQRES	37	A	478	ALA	GLY	SER	LYS	ILE	CYS	SER	ASP	SER	SER			
SEQRES	1	B	478	ALA	THR	PRO	ALA	ASP	TRP	ARG	SER	GLN	SER	ILE	TYR	PHE
SEQRES	2	B	478	LEU	LEU	THR	ASP	ARG	PHE	ALA	ARG	THR	ASP	GLY	SER	THR
SEQRES	3	B	478	THR	ALA	THR	CYS	ASN	THR	ALA	ASP	GLN	LYS	TYR	CYS	GLY
SEQRES	4	B	478	GLY	THR	TRP	GLN	GLY	ILE	ILE	ASP	LYS	LEU	ASP	TYR	ILE
SEQRES	5	B	478	GLN	GLY	MET	GLY	PHE	THR	ALA	ILE	TRP	ILE	THR	PRO	VAL
SEQRES	6	B	478	THR	ALA	GLN	LEU	PRO	GLN	ASP	CYS	ALA	TYR	GLY	ASP	ALA
SEQRES	7	B	478	TYR	THR	GLY	TYR	TRP	GLN	THR	ASP	ILE	TYR	SER	LEU	ASN
SEQRES	8	B	478	GLU	ASN	TYR	GLY	THR	ALA	ASP	ASP	LEU	LYS	ALA	LEU	SER
SEQRES	9	B	478	SER	ALA	LEU	HIS	GLU	ARG	GLY	MET	TYR	LEU	MET	VAL	ASP
SEQRES	10	B	478	VAL	VAL	ALA	ASN	HIS	MET	GLY	TYR	ASP	GLY	ALA	GLY	SER
SEQRES	11	B	478	SER	VAL	ASP	TYR	SER	VAL	PHE	LYS	PRO	PHE	SER	SER	GLN
SEQRES	12	B	478	ASP	TYR	PHE	HIS	PRO	PHE	CYS	PHE	ILE	GLN	ASN	TYR	GLU
SEQRES	13	B	478	ASP	GLN	THR	GLN	VAL	GLU	ASP	CYS	TRP	LEU	GLY	ASP	ASN
SEQRES	14	B	478	THR	VAL	SER	LEU	PRO	ASP	LEU	ASP	THR	THR	LYS	ASP	VAL
SEQRES	15	B	478	VAL	LYS	ASN	GLU	TRP	TYR	ASP	TRP	VAL	GLY	SER	LEU	VAL
SEQRES	16	B	478	SER	ASN	TYR	SER	ILE	ASP	GLY	LEU	ARG	ILE	ASP	THR	VAL
SEQRES	17	B	478	LYS	HIS	VAL	GLN	LYS	ASP	PHE	TRP	PRO	GLY	TYR	ASN	LYS
SEQRES	18	B	478	ALA	ALA	GLY	VAL	TYR	CYS	ILE	GLY	GLU	VAL	LEU	ASP	GLY
SEQRES	19	B	478	ASP	PRO	ALA	TYR	THR	CYS	PRO	TYR	GLN	ASN	VAL	MET	ASP
SEQRES	20	B	478	GLY	VAL	LEU	ASN	TYR	PRO	ILE	TYR	TYR	PRO	LEU	LEU	ASN
SEQRES	21	B	478	ALA	PHE	LYS	SER	THR	SER	GLY	SER	MET	ASP	ASP	LEU	TYR
SEQRES	22	B	478	ASN	MET	ILE	ASN	THR	VAL	LYS	SER	ASP	CYS	PRO	ASP	SER
SEQRES	23	B	478	THR	LEU	LEU	GLY	THR	PHE	VAL	GLU	ASN	HIS	ASP	ASN	PRO
SEQRES	24	B	478	ARG	PHE	ALA	SER	TYR	THR	ASN	ASP	ILE	ALA	LEU	ALA	LYS
SEQRES	25	B	478	ASN	VAL	ALA	ALA	PHE	ILE	ILE	LEU	ASN	ASP	GLY	LEU	PRO
SEQRES	26	B	478	ILE	ILE	TYR	ALA	GLY	GLN	GLU	GLN	HIS	TYR	ALA	GLY	GLY
SEQRES	27	B	478	ASN	ASP	PRO	ALA	ASN	ARG	GLU	ALA	THR	TRP	LEU	SER	GLY
SEQRES	28	B	478	TYR	PRO	THR	ASP	SER	GLU	LEU	TYR	LYS	LEU	ILE	ALA	SER
SEQRES	29	B	478	ALA	ASN	ALA	ILE	ARG	ASN	TYR	ALA	ILE	SER	LYS	ASP	THR
SEQRES	30	B	478	GLY	PHE	VAL	THR	TYR	LYS	ASN	PRO	TYR	ILE	LYS	ASP	ASP
SEQRES	31	B	478	THR	THR	ILE	ALA	MET	ARG	LYS	GLY	THR	ASP	GLY	SER	GLN
SEQRES	32	B	478	ILE	VAL	THR	ILE	LEU	SER	ASN	LYS	GLY	ALA	SER	GLY	ASP
SEQRES	33	B	478	SER	TYR	THR	LEU	SER	LEU	SER	GLY	ALA	SER	TYR	THR	ALA
SEQRES	34	B	478	GLY	GLN	GLN	LEU	THR	GLU	VAL	ILE	GLY	CYS	THR	THR	VAL
SEQRES	35	B	478	THR	VAL	GLY	SER	ASP	GLY	ASN	VAL	PRO	VAL	PRO	MET	ALA
SEQRES	36	B	478	GLY	GLY	LEU	PRO	ARG	VAL	LEU	TYR	PRO	THR	GLU	LYS	LEU
SEQRES	37	B	478	ALA	GLY	SER	LYS	ILE	CYS	SER	ASP	SER	SER			
SEQRES	1	C	478	ALA	THR	PRO	ALA	ASP	TRP	ARG	SER	GLN	SER	ILE	TYR	PHE
SEQRES	2	C	478	LEU	LEU	THR	ASP	ARG	PHE	ALA	ARG	THR	ASP	GLY	SER	THR
SEQRES	3	C	478	THR	ALA	THR	CYS	ASN	THR	ALA	ASP	GLN	LYS	TYR	CYS	GLY
SEQRES	4	C	478	GLY	THR	TRP	GLN	GLY	ILE	ILE	ASP	LYS	LEU	ASP	TYR	ILE
SEQRES	5	C	478	GLN	GLY	MET	GLY	PHE	THR	ALA	ILE	TRP	ILE	THR	PRO	VAL
SEQRES	6	C	478	THR	ALA	GLN	LEU	PRO	GLN	ASP	CYS	ALA	TYR	GLY	ASP	ALA
SEQRES	7	C	478	TYR	THR	GLY	TYR	TRP	GLN	THR	ASP	ILE	TYR	SER	LEU	ASN
SEQRES	8	C	478	GLU	ASN	TYR	GLY	THR	ALA	ASP	ASP	LEU	LYS	ALA	LEU	SER
SEQRES	9	C	478	SER	ALA	LEU	HIS	GLU	ARG	GLY	MET	TYR	LEU	MET	VAL	ASP
SEQRES	10	C	478	VAL	VAL	ALA	ASN	HIS	MET	GLY	TYR	ASP	GLY	ALA	GLY	SER
SEQRES	11	C	478	SER	VAL	ASP	TYR	SER	VAL	PHE	LYS	PRO	PHE	SER	SER	GLN

```

SEQRES 12 C 478 ASP TYR PHE HIS PRO PHE CYS PHE ILE GLN ASN TYR GLU
SEQRES 13 C 478 ASP GLN THR GLN VAL GLU ASP CYS TRP LEU GLY ASP ASN
SEQRES 14 C 478 THR VAL SER LEU PRO ASP LEU ASP THR THR LYS ASP VAL
SEQRES 15 C 478 VAL LYS ASN GLU TRP TYR ASP TRP VAL GLY SER LEU VAL
SEQRES 16 C 478 SER ASN TYR SER ILE ASP GLY LEU ARG ILE ASP THR VAL
SEQRES 17 C 478 LYS HIS VAL GLN LYS ASP PHE TRP PRO GLY TYR ASN LYS
SEQRES 18 C 478 ALA ALA GLY VAL TYR CYS ILE GLY GLU VAL LEU ASP GLY
SEQRES 19 C 478 ASP PRO ALA TYR THR CYS PRO TYR GLN ASN VAL MET ASP
SEQRES 20 C 478 GLY VAL LEU ASN TYR PRO ILE TYR TYR PRO LEU LEU ASN
SEQRES 21 C 478 ALA PHE LYS SER THR SER GLY SER MET ASP ASP LEU TYR
SEQRES 22 C 478 ASN MET ILE ASN THR VAL LYS SER ASP CYS PRO ASP SER
SEQRES 23 C 478 THR LEU LEU GLY THR PHE VAL GLU ASN HIS ASP ASN PRO
SEQRES 24 C 478 ARG PHE ALA SER TYR THR ASN ASP ILE ALA LEU ALA LYS
SEQRES 25 C 478 ASN VAL ALA ALA PHE ILE ILE LEU ASN ASP GLY LEU PRO
SEQRES 26 C 478 ILE ILE TYR ALA GLY GLN GLU GLN HIS TYR ALA GLY GLY
SEQRES 27 C 478 ASN ASP PRO ALA ASN ARG GLU ALA THR TRP LEU SER GLY
SEQRES 28 C 478 TYR PRO THR ASP SER GLU LEU TYR LYS LEU ILE ALA SER
SEQRES 29 C 478 ALA ASN ALA ILE ARG ASN TYR ALA ILE SER LYS ASP THR
SEQRES 30 C 478 GLY PHE VAL THR TYR LYS ASN PRO TYR ILE LYS ASP ASP
SEQRES 31 C 478 THR THR ILE ALA MET ARG LYS GLY THR ASP GLY SER GLN
SEQRES 32 C 478 ILE VAL THR ILE LEU SER ASN LYS GLY ALA SER GLY ASP
SEQRES 33 C 478 SER TYR THR LEU SER LEU SER GLY ALA SER TYR THR ALA
SEQRES 34 C 478 GLY GLN GLN LEU THR GLU VAL ILE GLY CYS THR THR VAL
SEQRES 35 C 478 THR VAL GLY SER ASP GLY ASN VAL PRO VAL PRO MET ALA
SEQRES 36 C 478 GLY GLY LEU PRO ARG VAL LEU TYR PRO THR GLU LYS LEU
SEQRES 37 C 478 ALA GLY SER LYS ILE CYS SER ASP SER SER
HET CA A 479 1
HET CA B 479 1
HET CA C 479 1
HETNAM CA CALCIUM ION
FORMUL 4 CA 3 (CA 2+)
HELIX 1 H1 GLY A 44 GLN A 53 1HELIX IS BENT AT 49 TO 50 10
HELIX 2 H2 ASP A 99 ARG A 110 1 12
HELIX 3 H3 VAL A 182 VAL A 195 1 14
HELIX 4 H4 PHE A 215 ALA A 222 1 8
HELIX 5 H5 ALA A 237 TYR A 242 1 6
HELIX 6 H6 PRO A 253 ALA A 261 1 9
HELIX 7 H7 ASP A 270 LYS A 280 1 11
HELIX 8 H8 ILE A 308 LEU A 320 1 13
HELIX 9 H9 GLU A 357 SER A 374 1 18
HELIX 10 H11 GLY B 44 GLN B 53 1HELIX IS BENT AT 49 TO 50 10
HELIX 11 H12 ASP B 99 ARG B 110 1 12
HELIX 12 H13 VAL B 182 VAL B 195 1 14
HELIX 13 H14 PHE B 215 ALA B 222 1 8
HELIX 14 H15 ALA B 237 TYR B 242 1 6
HELIX 15 H16 PRO B 253 ALA B 261 1 9
HELIX 16 H17 ASP B 270 LYS B 280 1 11
HELIX 17 H18 ILE B 308 LEU B 320 1 13
HELIX 18 H19 GLU B 357 SER B 374 1 18
HELIX 19 H21 GLY C 44 GLN C 53 1HELIX IS BENT AT 49 TO 50 10
HELIX 20 H22 ASP C 99 ARG C 110 1 12
HELIX 21 H23 VAL C 182 VAL C 195 1 14
HELIX 22 H24 PHE C 215 ALA C 222 1 8
HELIX 23 H25 ALA C 237 TYR C 242 1 6
HELIX 24 H26 PRO C 253 ALA C 261 1 9
HELIX 25 H27 ASP C 270 LYS C 280 1 11
HELIX 26 H28 ILE C 308 LEU C 320 1 13
HELIX 27 H29 GLU C 357 SER C 374 1 18
SHEET 1 BS1 9 SER A 10 LEU A 15 0
SHEET 2 BS1 9 THR A 58 PRO A 64 1
SHEET 3 BS1 9 MET A 112 VAL A 118 1
SHEET 4 BS1 9 GLY A 202 ASP A 206 1
SHEET 5 BS1 9 TYR A 226 VAL A 231 1
SHEET 6 BS1 9 GLY A 248 TYR A 252 1
SHEET 7 BS1 9 LEU A 289 VAL A 293 1
SHEET 8 BS1 9 GLY A 323 ALA A 329 1
SHEET 9 BS1 9 SER A 10 LEU A 15 1
SHEET 1 BS2 3 HIS A 122 GLY A 127 0
SHEET 2 BS2 3 THR A 170 ASP A 175 -1
SHEET 3 BS2 3 TRP A 165 ASN A 169 -1
SHEET 1 BS3 8 ASN A 384 ASP A 390 0
SHEET 2 BS3 8 THR A 392 GLY A 398 -1

```

SHEET	3	BS3	8	ILE	A	404	ASN	A	410	-1			
SHEET	4	BS3	8	SER	A	417	GLY	A	424	-1			
SHEET	5	BS3	8	GLY	A	430	VAL	A	436	-1			
SHEET	6	BS3	8	THR	A	440	GLY	A	445	-1			
SHEET	7	BS3	8	VAL	A	450	GLY	A	456	-1			
SHEET	8	BS3	8	LEU	A	458	THR	A	465	-1			
SHEET	1	BS4	9	SER	B	10	LEU	B	15	0			
SHEET	2	BS4	9	THR	B	58	PRO	B	64	1			
SHEET	3	BS4	9	MET	B	112	VAL	B	118	1			
SHEET	4	BS4	9	GLY	B	202	ASP	B	206	1			
SHEET	5	BS4	9	TYR	B	226	VAL	B	231	1			
SHEET	6	BS4	9	GLY	B	248	TYR	B	252	1			
SHEET	7	BS4	9	LEU	B	289	VAL	B	293	1			
SHEET	8	BS4	9	GLY	B	323	ALA	B	329	1			
SHEET	9	BS4	9	SER	B	10	LEU	B	15	1			
SHEET	1	BS5	3	HIS	B	122	GLY	B	127	0			
SHEET	2	BS5	3	THR	B	170	ASP	B	175	-1			
SHEET	3	BS5	3	TRP	B	165	ASN	B	169	-1			
SHEET	1	BS6	8	ASN	B	384	ASP	B	390	0			
SHEET	2	BS6	8	THR	B	392	GLY	B	398	-1			
SHEET	3	BS6	8	ILE	B	404	ASN	B	410	-1			
SHEET	4	BS6	8	SER	B	417	GLY	B	424	-1			
SHEET	5	BS6	8	GLY	B	430	VAL	B	436	-1			
SHEET	6	BS6	8	THR	B	440	GLY	B	445	-1			
SHEET	7	BS6	8	VAL	B	450	GLY	B	456	-1			
SHEET	8	BS6	8	LEU	B	458	THR	B	465	-1			
SHEET	1	BS7	9	SER	C	10	LEU	C	15	0			
SHEET	2	BS7	9	THR	C	58	PRO	C	64	1			
SHEET	3	BS7	9	MET	C	112	VAL	C	118	1			
SHEET	4	BS7	9	GLY	C	202	ASP	C	206	1			
SHEET	5	BS7	9	TYR	C	226	VAL	C	231	1			
SHEET	6	BS7	9	GLY	C	248	TYR	C	252	1			
SHEET	7	BS7	9	LEU	C	289	VAL	C	293	1			
SHEET	8	BS7	9	GLY	C	323	ALA	C	329	1			
SHEET	9	BS7	9	SER	C	10	LEU	C	15	1			
SHEET	1	BS8	3	HIS	C	122	GLY	C	127	0			
SHEET	2	BS8	3	THR	C	170	ASP	C	175	-1			
SHEET	3	BS8	3	TRP	C	165	ASN	C	169	-1			
SHEET	1	BS9	8	ASN	C	384	ASP	C	390	0			
SHEET	2	BS9	8	THR	C	392	GLY	C	398	-1			
SHEET	3	BS9	8	ILE	C	404	ASN	C	410	-1			
SHEET	4	BS9	8	SER	C	417	GLY	C	424	-1			
SHEET	5	BS9	8	GLY	C	430	VAL	C	436	-1			
SHEET	6	BS9	8	THR	C	440	GLY	C	445	-1			
SHEET	7	BS9	8	VAL	C	450	GLY	C	456	-1			
SHEET	8	BS9	8	LEU	C	458	THR	C	465	-1			
SSBOND	1	CYS	A	30			CYS	A	38		1555	1555	1.79
SSBOND	2	CYS	A	150			CYS	A	164		1555	1555	1.88
SSBOND	3	CYS	A	240			CYS	A	283		1555	1555	2.06
SSBOND	4	CYS	A	439			CYS	A	474		1555	1555	2.23
SSBOND	5	CYS	B	30			CYS	B	38		1555	1555	1.79
SSBOND	6	CYS	B	150			CYS	B	164		1555	1555	1.88
SSBOND	7	CYS	B	240			CYS	B	283		1555	1555	2.06
SSBOND	8	CYS	B	439			CYS	B	474		1555	1555	2.23
SSBOND	9	CYS	C	30			CYS	C	38		1555	1555	1.79
SSBOND	10	CYS	C	150			CYS	C	164		1555	1555	1.88
SSBOND	11	CYS	C	240			CYS	C	283		1555	1555	2.06
SSBOND	12	CYS	C	439			CYS	C	474		1555	1555	2.23
LINK		CA		CA	A	479			OD2 ASP A 175		1555	1555	3.03
LINK		CA		CA	A	479			OD1 ASN A 121		1555	1555	2.11
LINK		CA		CA	A	479			OD1 ASP A 175		1555	1555	3.07
LINK		CA		CA	B	479			O ASP A 476		1555	1556	61.82
LINK		CA		CA	B	479			OD1 ASN B 121		1555	1555	2.11
LINK		CA		CA	B	479			N SER A 477		1555	1556	59.95
LINK		CA		CA	B	479			O SER A 477		1555	1556	59.93
LINK		CA		CA	B	479			OD1 ASP B 175		1555	1555	3.07
LINK		CA		CA	B	479			OD2 ASP B 175		1555	1555	3.03
LINK		CA		CA	B	479			OG SER A 477		1555	1556	58.46
LINK		CA		CA	C	479			OD1 ASN C 121		1555	1555	2.11
LINK		CA		CA	C	479			OD1 ASP C 175		1555	1555	3.07
LINK		CA		CA	C	479			OD2 ASP C 175		1555	1555	3.03
SITE	1	AC1	3	ASN	A	121	GLU	A	162	ASP A 175			

SITE	1	AC2	3	ASN	B	121	GLU	B	162	ASP	B	175						
SITE	1	AC3	3	ASN	C	121	GLU	C	162	ASP	C	175						
CRYST1	91.900	133.300				94.300	90.00	102.70	90.00	P	1	21	1				6	
ORIGX1		1.000000		0.000000		0.000000		0.000000		0.000000								
ORIGX2		0.000000		1.000000		0.000000		0.000000		0.000000								
ORIGX3		0.000000		0.000000		1.000000		0.000000		0.000000								
SCALE1		0.007110		0.000000		0.008600		0.000000		0.31300								
SCALE2		0.000000		0.007500		0.000000		0.000000		0.13700								
SCALE3		-0.006650		0.000000		0.008600		0.000000		0.41400								
MTRIX1	1	-0.537662		0.842761		0.003491		-13.11357		1								
MTRIX2	1	-0.843195		-0.537874		-0.015704		15.05393		1								
MTRIX3	1	-0.011223		-0.011613		0.999697		37.58095		1								
MTRIX1	2	0.494189		0.868623		0.024926		-18.69613		1								
MTRIX2	2	0.869382		-0.494682		0.002191		40.90763		1								
MTRIX3	2	0.014088		0.020356		-0.999516		4.56432		1								
ATOM	1	N	ALA	A	1	-8.065	17.989	-10.906	1.00	0.00							N	
ATOM	2	CA	ALA	A	1	-7.235	16.799	-11.163	1.00	0.00							C	
ATOM	3	C	ALA	A	1	-8.117	15.784	-10.473	1.00	0.00							C	
ATOM	4	O	ALA	A	1	-8.999	16.398	-9.862	1.00	0.00							O	
ATOM	5	CB	ALA	A	1	-6.004	16.975	-10.304	1.00	0.00							C	
ATOM	6	N	THR	A	2	-8.015	14.404	-10.731	1.00	0.00							N	
ATOM	7	CA	THR	A	2	-8.896	13.246	-10.275	1.00	0.00							C	
ATOM	8	C	THR	A	2	-9.123	12.010	-11.254	1.00	0.00							C	
ATOM	9	O	THR	A	2	-9.377	12.191	-12.428	1.00	0.00							O	
ATOM	10	CB	THR	A	2	-10.220	13.886	-10.054	1.00	0.00							C	
ATOM	11	OG1	THR	A	2	-10.752	14.492	-11.225	1.00	0.00							O	
ATOM	12	CG2	THR	A	2	-10.711	14.297	-8.666	1.00	0.00							C	
ATOM	13	N	PRO	A	3	-9.514	10.811	-10.733	1.00	0.00							N	
ATOM	14	CA	PRO	A	3	-9.710	9.413	-11.371	1.00	0.00							C	
ATOM	15	C	PRO	A	3	-9.923	9.115	-12.890	1.00	0.00							C	
ATOM	16	O	PRO	A	3	-9.607	8.022	-13.351	1.00	0.00							O	
ATOM	17	CB	PRO	A	3	-10.985	8.930	-10.675	1.00	0.00							C	
ATOM	18	CG	PRO	A	3	-11.690	10.193	-10.224	1.00	0.00							C	
ATOM	19	CD	PRO	A	3	-10.510	10.925	-9.647	1.00	0.00							C	
ATOM	20	N	ALA	A	4	-10.732	9.869	-13.615	1.00	0.00							N	
ATOM	21	CA	ALA	A	4	-11.084	9.554	-15.029	1.00	0.00							C	
ATOM	22	C	ALA	A	4	-10.120	10.164	-16.059	1.00	0.00							C	
ATOM	23	O	ALA	A	4	-10.000	9.792	-17.230	1.00	0.00							O	
ATOM	24	CB	ALA	A	4	-12.470	10.143	-15.170	1.00	0.00							C	
ATOM	25	N	ASP	A	5	-9.434	11.136	-15.575	1.00	0.00							N	
ATOM	26	CA	ASP	A	5	-8.405	11.693	-16.441	1.00	0.00							C	
ATOM	27	C	ASP	A	5	-7.125	10.824	-16.324	1.00	0.00							C	
ATOM	28	O	ASP	A	5	-6.799	10.080	-17.224	1.00	0.00							O	
ATOM	29	CB	ASP	A	5	-8.065	13.135	-16.005	1.00	0.00							C	
ATOM	30	CG	ASP	A	5	-9.163	14.202	-15.732	1.00	0.00							C	
ATOM	31	OD1	ASP	A	5	-10.034	14.483	-16.589	1.00	0.00							O	
ATOM	32	OD2	ASP	A	5	-9.000	14.932	-14.744	1.00	0.00							O	
ATOM	33	N	TRP	A	6	-6.514	10.824	-15.184	1.00	0.00							N	
ATOM	34	CA	TRP	A	6	-5.387	9.985	-14.770	1.00	0.00							C	
ATOM	35	C	TRP	A	6	-5.008	8.680	-15.543	1.00	0.00							C	
ATOM	36	O	TRP	A	6	-4.044	8.664	-16.250	1.00	0.00							O	
ATOM	37	CB	TRP	A	6	-5.563	9.719	-13.334	1.00	0.00							C	
ATOM	38	CG	TRP	A	6	-4.994	10.840	-12.496	1.00	0.00							C	
ATOM	39	CD1	TRP	A	6	-5.610	11.803	-11.679	1.00	0.00							C	
ATOM	40	CD2	TRP	A	6	-3.649	11.181	-12.475	1.00	0.00							C	
ATOM	41	NE1	TRP	A	6	-4.659	12.657	-11.141	1.00	0.00							N	
ATOM	42	CE2	TRP	A	6	-3.517	12.280	-11.589	1.00	0.00							C	
ATOM	43	CE3	TRP	A	6	-2.549	10.625	-13.112	1.00	0.00							C	
ATOM	44	CZ2	TRP	A	6	-2.282	12.801	-11.292	1.00	0.00							C	
ATOM	45	CZ3	TRP	A	6	-1.297	11.203	-12.844	1.00	0.00							C	
ATOM	46	CH2	TRP	A	6	-1.175	12.255	-11.940	1.00	0.00							C	
ATOM	47	N	ARG	A	7	-5.891	7.718	-15.470	1.00	0.00							N	
ATOM	48	CA	ARG	A	7	-6.089	6.461	-16.239	1.00	0.00							C	
ATOM	49	C	ARG	A	7	-5.225	6.355	-17.529	1.00	0.00							C	
ATOM	50	O	ARG	A	7	-4.551	5.388	-17.745	1.00	0.00							O	
ATOM	51	CB	ARG	A	7	-7.616	6.542	-16.606	1.00	0.00							C	
ATOM	52	CG	ARG	A	7	-8.668	5.409	-16.216	1.00	0.00							C	
ATOM	53	CD	ARG	A	7	-10.181	5.879	-16.077	1.00	0.00							C	
ATOM	54	NE	ARG	A	7	-11.291	4.895	-16.498	1.00	0.00							N	
ATOM	55	CZ	ARG	A	7	-12.622	4.978	-16.258	1.00	0.00							C	
ATOM	56	NH1	ARG	A	7	-13.455	3.980	-16.429	1.00	0.00							N	
ATOM	57	NH2	ARG	A	7	-13.143	6.195	-16.040	1.00	0.00							N	

ATOM	58	N	SER	A	8	-5.225	7.424	-18.329	1.00	0.00	N
ATOM	59	CA	SER	A	8	-4.566	7.763	-19.675	1.00	0.00	C
ATOM	60	C	SER	A	8	-3.072	8.225	-19.778	1.00	0.00	C
ATOM	61	O	SER	A	8	-2.441	8.022	-20.800	1.00	0.00	O
ATOM	62	CB	SER	A	8	-5.457	8.753	-20.414	1.00	0.00	C
ATOM	63	OG	SER	A	8	-6.747	8.883	-19.839	1.00	0.00	O
ATOM	64	N	GLN	A	9	-2.553	8.911	-18.763	1.00	0.00	N
ATOM	65	CA	GLN	A	9	-1.207	9.537	-18.464	1.00	0.00	C
ATOM	66	C	GLN	A	9	-0.031	8.553	-18.263	1.00	0.00	C
ATOM	67	O	GLN	A	9	-0.144	7.650	-17.474	1.00	0.00	O
ATOM	68	CB	GLN	A	9	-1.409	10.315	-17.146	1.00	0.00	C
ATOM	69	CG	GLN	A	9	-0.764	11.668	-17.010	1.00	0.00	C
ATOM	70	CD	GLN	A	9	-1.828	12.759	-16.804	1.00	0.00	C
ATOM	71	OE1	GLN	A	9	-1.676	13.832	-17.277	1.00	0.00	O
ATOM	72	NE2	GLN	A	9	-2.786	12.419	-16.083	1.00	0.00	N
ATOM	73	N	SER	A	10	0.993	8.688	-19.038	1.00	0.00	N
ATOM	74	CA	SER	A	10	2.170	7.791	-18.881	1.00	0.00	C
ATOM	75	C	SER	A	10	2.984	8.202	-17.615	1.00	0.00	C
ATOM	76	O	SER	A	10	2.982	9.368	-17.294	1.00	0.00	O
ATOM	77	CB	SER	A	10	2.844	8.034	-20.254	1.00	0.00	C
ATOM	78	OG	SER	A	10	3.385	6.881	-21.006	1.00	0.00	O

- pokračovanie

ATOM	3664	N	SER	A	475	6.073	8.445	-40.258	1.00	0.00	N
ATOM	3665	CA	SER	A	475	6.937	8.896	-41.427	1.00	0.00	C
ATOM	3666	C	SER	A	475	6.284	10.033	-42.265	1.00	0.00	C
ATOM	3667	O	SER	A	475	6.259	9.939	-43.468	1.00	0.00	O
ATOM	3668	CB	SER	A	475	7.321	7.635	-42.229	1.00	0.00	C
ATOM	3669	OG	SER	A	475	8.116	7.756	-43.413	1.00	0.00	O
ATOM	3670	N	ASP	A	476	5.776	11.090	-41.626	1.00	0.00	N
ATOM	3671	CA	ASP	A	476	5.133	12.284	-42.240	1.00	0.00	C
ATOM	3672	C	ASP	A	476	6.174	13.363	-42.554	1.00	0.00	C
ATOM	3673	O	ASP	A	476	6.007	14.035	-43.563	1.00	0.00	O
ATOM	3674	CB	ASP	A	476	4.007	12.817	-41.342	1.00	0.00	C
ATOM	3675	CG	ASP	A	476	3.954	14.385	-41.297	1.00	0.00	C
ATOM	3676	OD1	ASP	A	476	3.952	15.004	-40.235	1.00	0.00	O
ATOM	3677	OD2	ASP	A	476	3.942	15.017	-42.330	1.00	0.00	O
ATOM	3678	N	SER	A	477	7.036	13.599	-41.577	1.00	0.00	N
ATOM	3679	CA	SER	A	477	8.055	14.683	-41.595	1.00	0.00	C
ATOM	3680	C	SER	A	477	7.610	16.037	-41.023	1.00	0.00	C
ATOM	3681	O	SER	A	477	7.178	16.874	-41.791	1.00	0.00	O
ATOM	3682	CB	SER	A	477	8.633	14.815	-42.986	1.00	0.00	C
ATOM	3683	OG	SER	A	477	9.626	15.837	-43.066	1.00	0.00	O
ATOM	3684	N	SER	A	478	7.619	16.181	-39.701	1.00	0.00	N
ATOM	3685	CA	SER	A	478	7.120	17.394	-38.962	1.00	0.00	C
ATOM	3686	C	SER	A	478	7.396	17.343	-37.426	1.00	0.00	C
ATOM	3687	O	SER	A	478	8.188	18.098	-36.921	1.00	0.00	O
ATOM	3688	CB	SER	A	478	5.662	17.627	-39.000	1.00	0.00	C
ATOM	3689	OG	SER	A	478	5.356	18.313	-37.828	1.00	0.00	O
ATOM	3690	OXT	SER	A	478	6.700	16.633	-36.724	1.00	0.00	O
TER	3691		SER	A	478						

- pokračovanie

HETATM	11074	CA	CA	A	479	-0.075	-0.525	10.425	1.00	0.00	CA
--------	-------	----	----	---	-----	--------	--------	--------	------	------	----

- pokračovanie

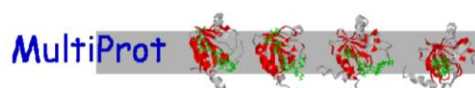
CONECT 235 299
 CONECT 299 235
 CONECT 94411074
 CONECT 1176 1296
 CONECT 1296 1176
 CONECT 138011074
 CONECT 138111074
 CONECT 1898 2239
 CONECT 2239 1898
 CONECT 3418 3663
 CONECT 3663 3418

- pokračovanie

MASTER	276	1	3	27	60	0	3	1211073	3	36	111
END											

Porovnávanie terciárnych štruktúr (structure alignment, overlay):

- vzájomné preloženie terciárnych štruktúr proteínov;
- superpozícia atómových koordinát s udaním počtu korešpondujúcich C_α atómov a hodnoty rmsd (Å; root-mean square deviation).



Napr. program MultiProt: <http://bioinfo3d.cs.tau.ac.il/MultiProt/>

[\[Back to Home Page\]](#)

PDB IDs separated by a space: (e.g. 1adj:A 1hc7:A 1v95)
Upload zip file of pdb structures: Nie je zvolený súbor.
Accuracy (in Angstroms):
Sequence Order: Yes * No ☐

* for NMR structures only the first model is considered *

Contact: max.shatsky@gmail.com

MultiProt is a fully automated highly efficient technique which detects the multiple structural alignments of protein structures. It finds the common geometrical cores between the input molecules. Most of the existing methods require that all the input molecules be aligned, while MultiProt does not require that all the input molecules participate in the alignment. Actually, it efficiently detects high scoring partial multiple alignments for all possible number of molecules from the input.

The final structural alignment can either preserve the sequence order (like sequence alignment), or be sequence order independent.

Reference:

Shatsky, M. and Nussinov, R. and Wolfson, H.J., *A method for simultaneous alignment of multiple protein structures*. *Proteins: Structure, Function, and Bioinformatics*. 2004 Jul 1;56(1):143-56.

Kvartérna štruktúra

Je to priestorové usporiadanie podjednotiek proteínu (jeho jednotlivých polypeptidových reťazcov).

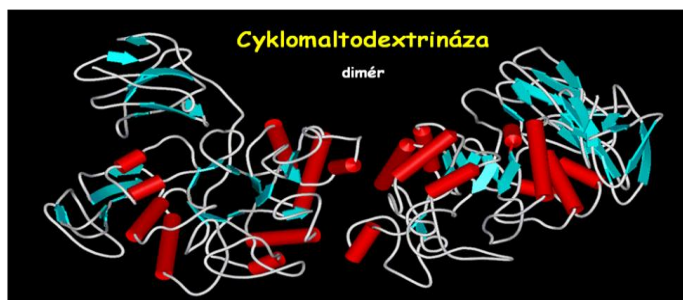
Z monoméru sa vytvárajú oligoméry – diméry, triméry, ...

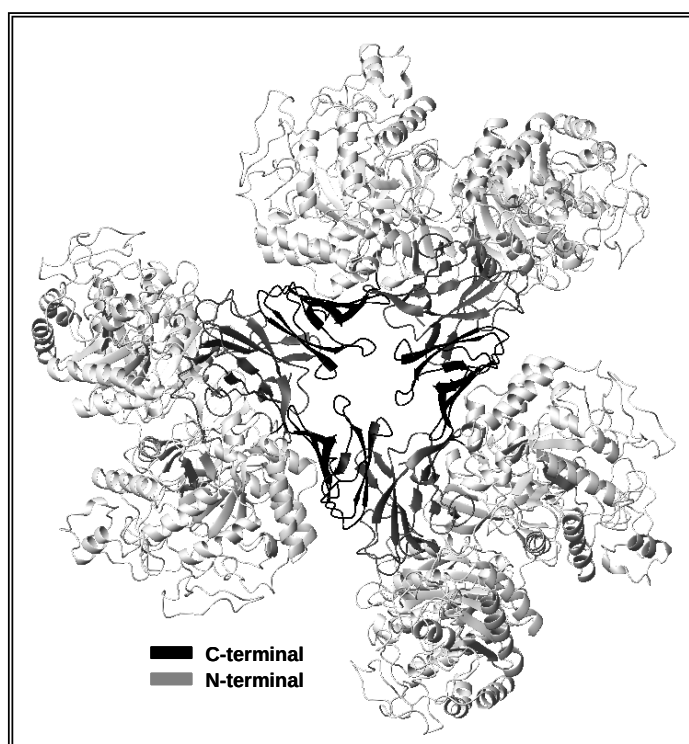
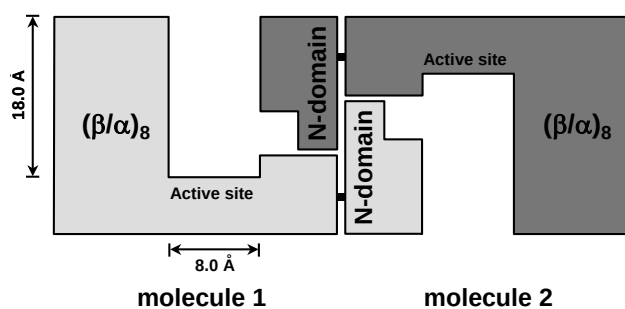
Tvorba kvartérnej štruktúry je často podmienená zvýšenou výkonnosťou proteínu (napr. pri enzýmoch – viac spolupracujúcich aktívnych centier).

Nie každý proteín musí mať kvartérnu štruktúru.

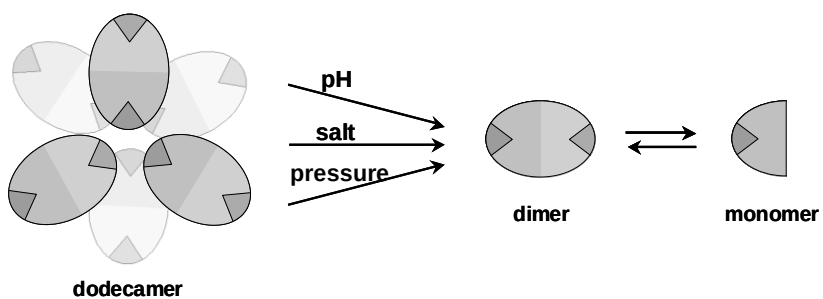
Príklad:

- cyklomaltodextrináza z *Bacillus* sp. I-5 (PDB kód: 1EA9) vytvára kvartérnu štruktúru – v závislosti podmienok – monomér – dimér – hexamér – dodekamér – pH <6 dimér, pH >6,5 dodekamér (zložený zo 6 homodimérov).





Dodekamér cyklomaltodextrinázy (6 homodimérov).



Disociácia-asociácia oligomerizácie cyklomaltodextrinázy.

(Upravené podľa: Lee a kol. *Biologia* 2005, 60 (Suppl. 16): 73)

Ramachandranov diagram

Spôsob vizualizácie dihedrálnych uhlov ϕ a ψ aminokyselinových zvyškov proteínu.

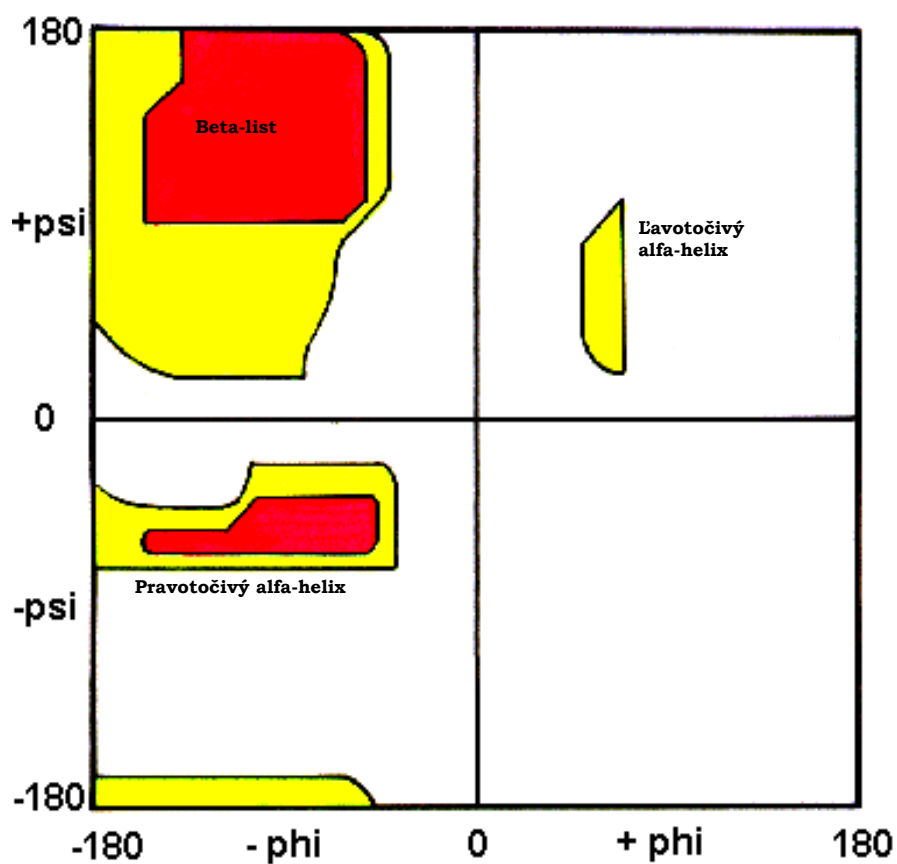
Udáva možné konformácie uhlov ϕ a ψ v reálnom proteíne.

V proteíne väzby hlavného reťazca $N-C_\alpha$ a $C_\alpha-C(O)$ môžu rotovať.

Tieto rotácie sú reprezentované dihedrálnymi (torznými) uhlami ϕ a ψ .

Určité kombinácie uhlov ϕ a ψ nie sú dovolené.

Ramachandranov diagram.



(Upravené podľa: http://www.cryst.bbk.ac.uk/PPS95/course/3_geometry/rama.html)

Biele oblasti v diagrame korešpondujú konformáciám, pri ktorých sa atómy v proteíne dostanú bližšie, než je suma ich van der Waals polomerov.

Tieto oblasti sú stéricky zakázané pre všetky aminokyselinové zvyšky okrem glycínu (nemá postranný reťazec).

Červené oblasti korešpondujú konformáciám, pre ktoré neexistujú stérické prekážky, t.j. sú dovolené – α -helikálna a β -listová konformácia.

Žlté oblasti naznačujú čiastočne dovolené kombinácie uhlov φ a ψ – prípady, kedy na výpočet boli použité skrátené van der Waals polomery, t.j. atómy sa môžu dostať navzájom bližšie spolu. Toto vysvetľuje oblasť ľavotočivého α -helixu.

L-aminokyseliny zvyčajne (okrem glycínu) nemôžu tvoriť rozvinuté (pretiahnuté) formy ľavotočivej špirály (nevýhodná konformácia); občas tiež platí pre kyselinu asparágovú alebo asparagín – schopné tvoriť H-väzby s hlavným reťazcom (stabilizujú inak nevýhodnú konformáciu).

Nedovolené oblasti existujú v dôsledku stérických prekážok medzi skupinami na C_β uhlíkoch a atómami hlavného reťazca.

Glycín, ktorý nemá postranný reťazec, môže adoptovať kombinácie uhlov φ a ψ vo všetkých kvadrantoch Ramachandranovho diagramu (napr. ohyby).

Fibrilárne proteíny

- na ich opis stačí oblasť α_R alebo β -región

Globulárne proteíny

- všetky dovolené oblasti, pretože reťazec mení smer

5. Supersekundárna štruktúra proteínov, motívy a domény

Zjednodušená charakterizácia štruktúry proteínov

- primárna – sekvencia;
- sekundárna – geometria;
- terciárna – priestor;
- kvartérna – oligoméry.

Vyššie typy štruktúr (2D, 3D) sú determinované primárnou štruktúrou, ktorá je (vo všeobecnosti) kódovaná DNA.

Pridané úrovne štruktúry proteínov

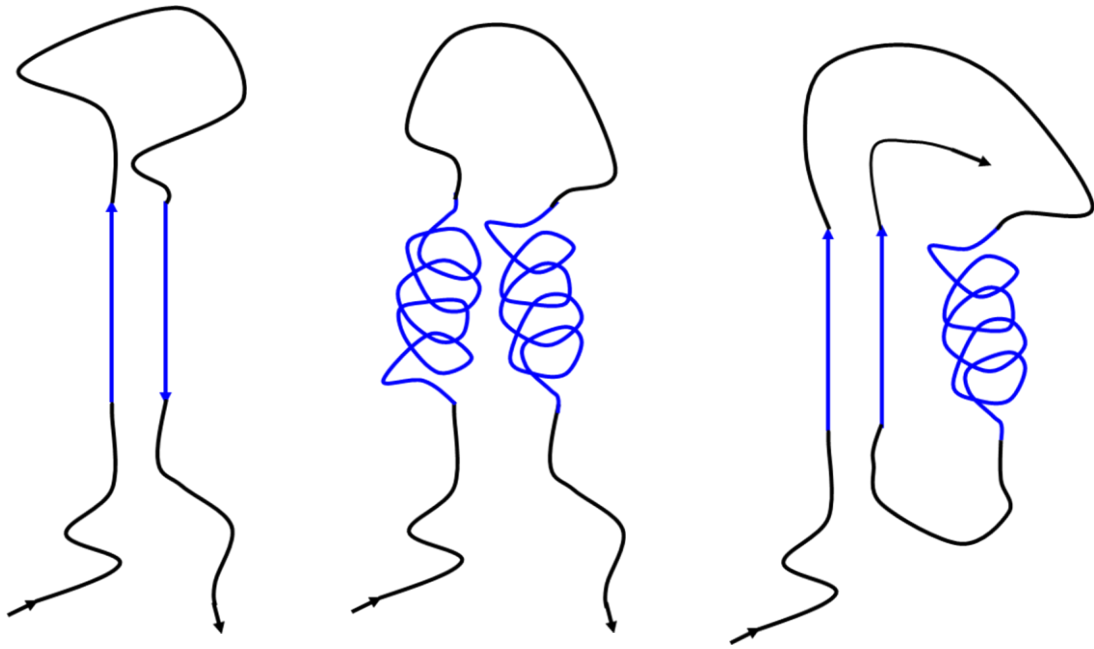
- supersekundárna štruktúra – určitý vzor interakcií medzi helixami a vláknami, ktoré sú navzájom blízko v sekvencii; môžu mať vlastnú biologickú funkciu (napr. väzbu DNA), prípadne sú časťou väčších štruktúrnych zoskupení;
- doména (modul) – väčšie zoskupenie s vlastným vzorom zvinutia so zdanlivo nezávislou stabilitou; v hierarchii je medzi supersekundárnou a terciárnou štruktúrou;
- modulárny proteín – multidoménný proteín.

Slučky a ohyby

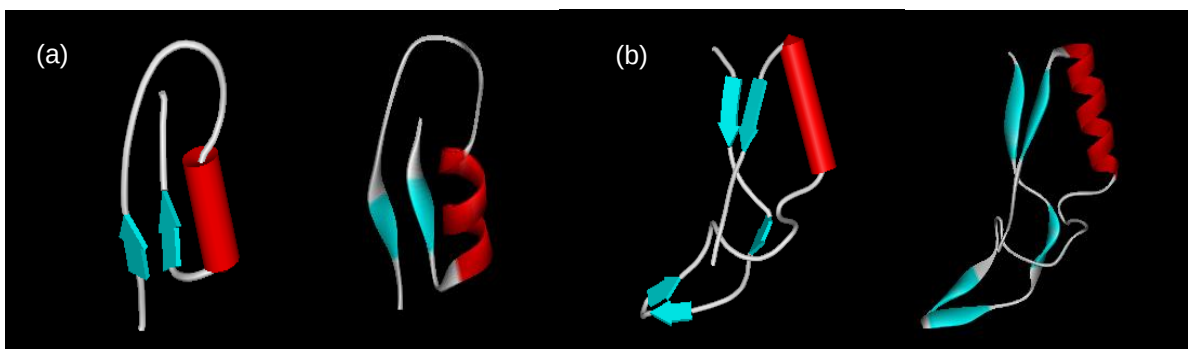
Slučky a ohyby (loops, turns) sú nepravidelné elementy sekundárnej štruktúry proteínov. Sú lokalizované najmä na povrchu molekúl globulárnych proteínov.

Motívy

- β -vlásočnica (β -hairpin);
- α -vlásočnica (α -hairpin);
- β - α - β -jednotka (β - α - β -unit).



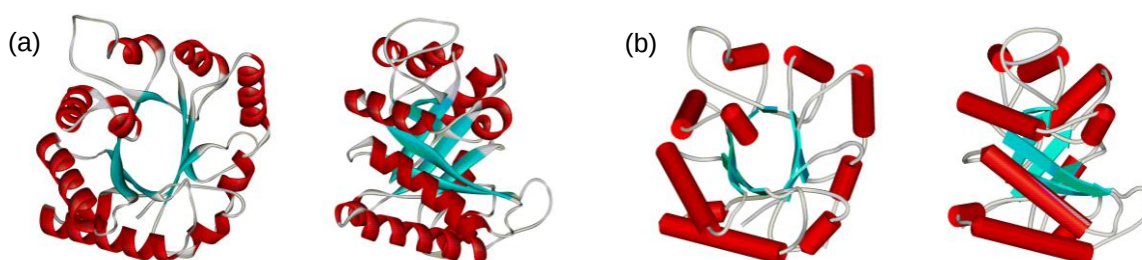
Jednoduché motívy (a) sa kombinujú a tvoria zložené motívy (b), z ktorých potom môže byť vytvorený modul (doména), ktorá môže byť celým proteínom alebo jeho súčasťou.





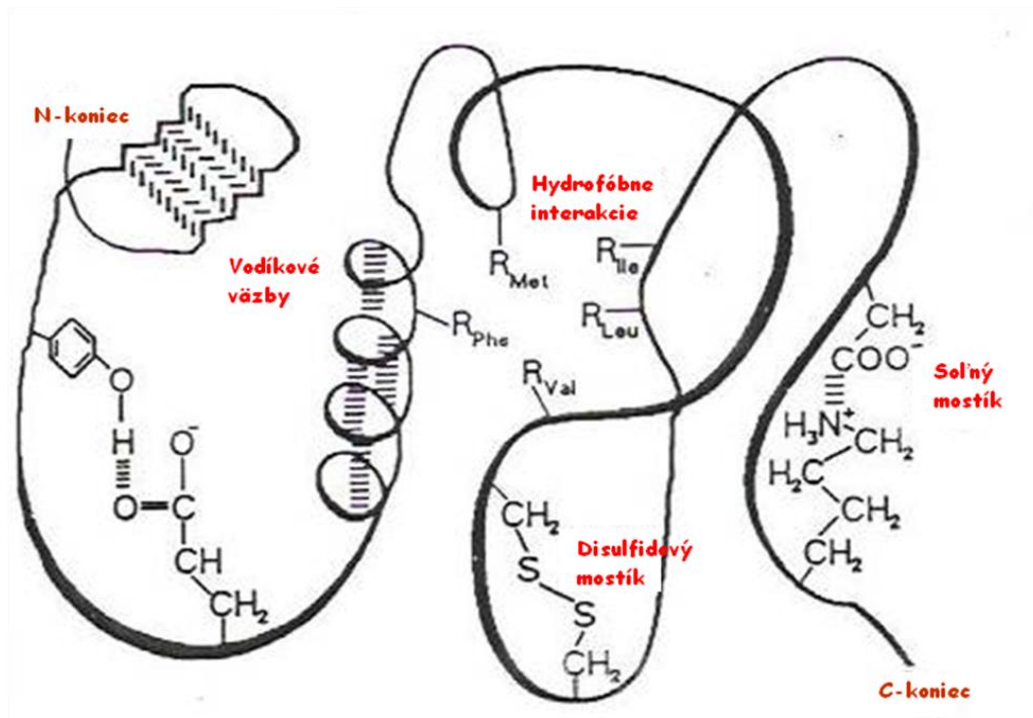
V molekule cyklodextrínglukanotransferázy z *Bacillus circulans* (PDB kód: 1CDG) sú motívy (a) a (b) z predchádzajúceho obrázku zvýraznené zelenou, resp. žltou farbou.

Nie všetky proteíny majú modulárnu štruktúru, t.j. sú zložené z viacerých domén. Menšie proteíny (aj vo veľkosti 200-300 aminokyselinových zvyškov) môžu byť tvorené len jednou doménou, napr. TIM-barel – $(\beta/\alpha)_8$ -barel.



TIM-barel – katalytická doména tvoriaca celú proteínovú molekulu enzýmu TriózafosfátlzoMeráza; z názvu enzýmu pochádza aj názov domény (1. raz bola táto častá proteínová štruktúra určená práve v tomto enzýme). Štruktúra ako stuha (a) a so zvýraznením α -špirál a β -vlákien (b), pričom pohľad je zhora (vľavo) a zboku (vpravo).

Väzby a interakcie v proteíne



(Upravené podľa: Bohinski *Modern Concepts in Biochemistry*, Allyn and Bacon, Boston, 1983)

Ideálny proteín

- hydrofóbne jadro (interiér) a hydrofilný povrch (exteriér)

Reálny proteín

- hydrofóbne jadro (interiér), v ktorom ale môžu byť hydrofilné molekuly vody (napr. kvôli nutnosti transportu substrátu do aktívneho miesta enzýmu);
- hydrofilný povrch (exteriér), na ktorom ale môžu byť rôzne hydrofóbne skupiny (napr. aminokyselinové zvyšky s hydrofóbnymi reťazcami zabezpečujúce väzbu dôležitých látok na povrchu proteínu).

Odklon od ideálneho stavu v realite je podmienený biologickou funkčnosťou.

6. Katalogizácia štruktúr proteínov

Štruktúra proteínu vs. štruktúra textu

AMINOKYSELINY

SEKUNDÁRNA ŠTRUKTÚRA

SUPERSEKUNDÁRNA ŠTRUKTÚRA

DOMÉNY

TERCIÁRNA ŠTRUKTÚRA

KVARTÉRNA ŠTRUKTÚRA

PÍSMENÁ

SLOVÁ

FRÁZY

VETY A ODSTAVCE

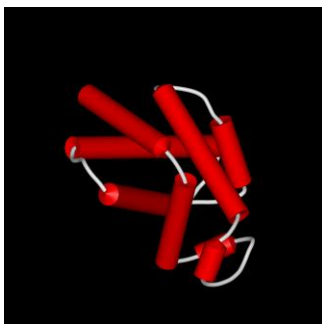
KAPITOLY

CELÁ KNIHA

Klasifikácia proteínov podľa obsahu sekundárnej štruktúry

- α -helix – sekundárna štruktúra iba alebo takmer iba z α -helixov;
- β -list – sekundárna štruktúra iba alebo takmer iba z β -listov;
- $(\alpha+\beta)$ – α -helixy a β -listy separované v odlišných častiach molekuly proteínu (absencia β - α - β motívov);
- α/β – štruktúra je vytvorená alternovaním α -helixov a β -listov (prítomnosť β - α - β motívov);
 - α/β lineárna – čiara cez stred β -vlákien listu je ~lineárna;
 - α/β barely – čiara cez stred β -vlákien listu je ~kruhová (elipsa)
- málo alebo žiadna sekundárna štruktúra.

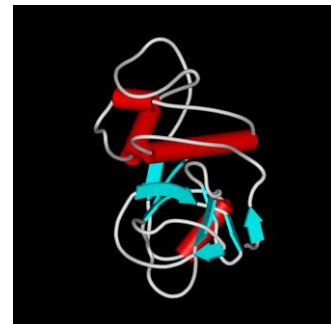
myoglobín (hlavne α)



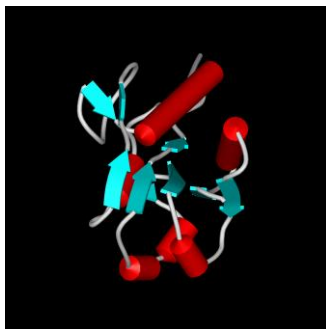
chymotrypsín (hlavne β)



papain ($\alpha+\beta$)

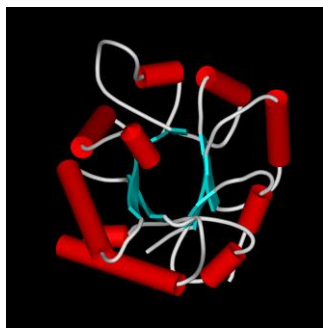


flavodoxín (α/β)



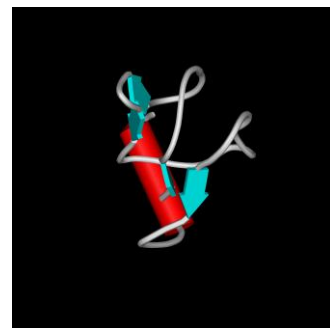
lineárna

triózafosfátizomeráza (α/β)



barel (elipsa)

ferredoxín (málo 2D)



Katalógy štruktúr proteínov

Existencia množstva proteínových štruktúr priniesla potrebu ich katalogizácie.

www-systém

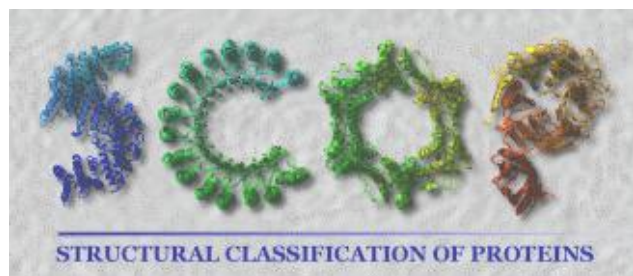
- vhodnosť, dostupnosť, umožňuje prehľadávanie a vyberanie dát, ich uchovávanie, prezentáciu (zobrazovanie), vzájomné prepojenie (vrátane bibliografických databáz).

CATH – Class, Architecture, Topology, Homologous superfamily

SCOP – Structural Classification Of Proteins



<http://www.cathdb.info/>



<http://scop.mrc-lmb.cam.ac.uk/scop/>

CATH

- sekvenčná superrodina (podobné sekvencie; > ako 35%; domény);
- homologická superrodina (spoločný predok);
- topológia (fold; podobné 3D štruktúry);
- architektúra (rôzne spojenie α a β);
- trieda (α , β , zmes α - β).

CATH Home Search Browse Download About Support

CATH

Classification of Protein Structures

CATH provides information on protein structures that are related by evolution. Protein structures (from the PDB) are chopped into structural domains then grouped into evolutionary superfamilies.

[Get Started »](#) [Search »](#) [Download »](#) [Take the Tour »](#)

Gene3D

Structures Assigned to Genomes

Gene3D takes CATH domains and uses Hidden Markov Models (HMMs) to assign these structures to the millions of protein sequences without structural information.

What's New?

The CATH website has recently undergone a big overhaul. We really hope you find the new pages more useful, easier to use and quicker to load. Please [get in touch](#) and let us know what you think.

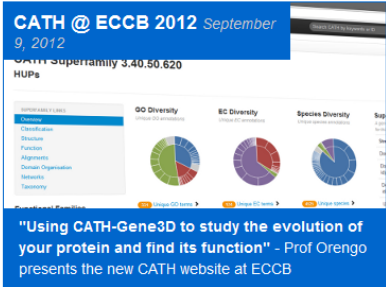
Searching CATH

- Search by ID / keyword
- Search by FASTA sequence
- Search by PDB structure

Example pages

- PDB "1dan"
- Domain "1cukA01"
- Relatives of "1cukA01"
- Superfamily "HUPs"
- Functional Family
- FunFam Alignment
- Search for "enolase"
- Superfamily Comparison

Latest News



"Using CATH-Gene3D to study the evolution of your protein and find its function" - Prof Orengo presents the new CATH website at ECCB

Release Statistics

CATH v3.5 based on PDB dated September 20, 2011	
173,536	CATH Domains
2,626	CATH Superfamilies
51,334	PDBs

Gene3D v11 released March 18, 2012	
1,639	Cellular Genomes
1,016	Viral Genomes
14,963,305	Protein Sequences
16,297,076	CATH Domain Predictions

CATH News
[Support](#)
[Jobs](#)

Get Started
[Documentation Home](#)
[Tutorials](#)

Download
[WebServices](#)
[Software](#)

About
[The Orengo Group](#)
[Web accessibility](#)

CATH je personálne spravovaná hierarchická klasifikácia doménových štruktúr proteínov, ktorá usporadúva proteíny na štyroch hlavných úrovniach:

Class (C), Architecture (A), Topology (T) and Homologous superfamily (H).

Trieda (C), architektúra (A), topológia (T) a homologická superrodina (H).

V súčasnosti sa pridáva ešte sekvenčná rodina – sequence family (S, O, L, I, D) – viac ako 35% identita: S – 35%; O – 60%; L – 95%; I – 100%.

Popis jednotlivých úrovní v CATH.

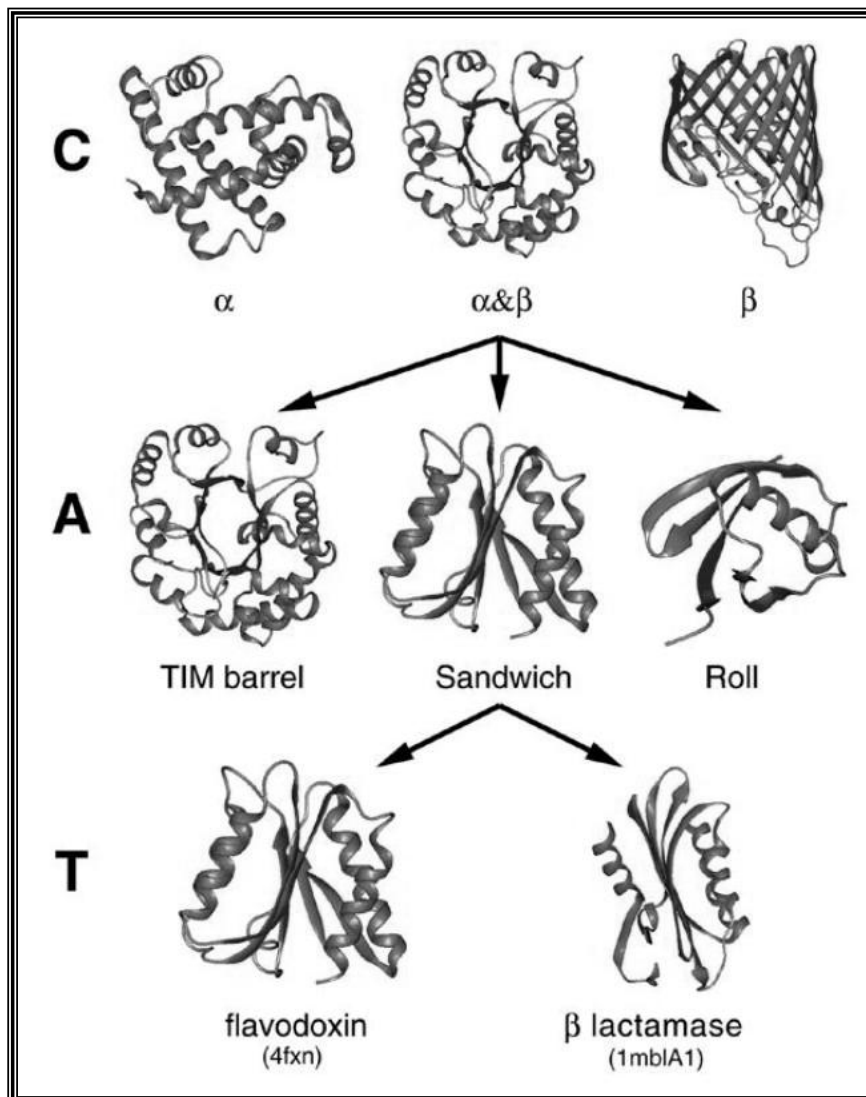
Úroveň	Písmeno	Názov	Kritériá
1	C	Class	Secondary structure content
2	A	Architecture	General spatial arrangement of secondary structures
3	T	Topology	Spatial arrangement and connectivity of secondary structures (fold)
4	H	Homologous Superfamily	Manual curation of evidence of evolutionary relationship
5	S	Sequence Family (S35)	$\geq 35\%$ sequence similarity
6	O	Orthologous Family (S60)	$\geq 60\%$ sequence similarity
7	L	"Like" domain (S95)	$\geq 95\%$ sequence similarity
8	I	Identical domain (S100)	100% sequence similarity
9	D	Domain counter	Unique domains

- trieda, odvodená z obsahu sekundárnej štruktúry, je pripísaná pre viac než 90% proteínových štruktúr automaticky;
- architektúra, ktorá opisuje hrubú orientáciu elementov sekundárnej štruktúry nezávisle od spojenia, je v súčasnosti prisudzovaná manuálne;
- topológia zoskupuje štruktúry podľa ich topologického spojenia a počtu prvkov sekundárnej štruktúry (fold);
- homologické superrodiny obsahujú proteíny s veľmi podobnými štruktúrami a funkciami (spoločný pôvod);
- prisúdenie štruktúr k úrovniam topológia a homologická superrodina sú robené pomocou porovnania sekvencií a štruktúr;
- celá klasifikácia je robená až na úroveň jednotlivých domén, t.j. končí až na úrovni sekvenčných superrodín; každá doména má svoje unikátne CATHSOLID číslo.

CATH štatistika

- verzia 4.0 na základe PDB (Protein Data Bank) z 26. marca 2013;
- 2 738 superrodín;
- 277 688 domén;
- 69 058 anotovaných PDB štruktúr.

Ilustrácia klasifikačnej schémy (postupnosti) v CATH



(Upravené podľa: Orengo a kol. *Structure* 1997, 5: 1093)

SCOP

- organizuje proteínové štruktúry v hierarchii podľa ich evolučnej príbuznosti a štruktúrnej podobnosti;
- doména (z PDB);
- rodina (> ako 30% sekvenčná identita);
- superrodina (< ako 30% sekvenčná identita);
- fold (podobná sekundárna štruktúra);
- trieda (α , β , $\alpha+\beta$, α/β).

Popis jednotlivých úrovní v SCOP.

Trieda	Počet foldov	Počet superrodín	Počet rodín
All alpha proteins	284	507	871
All beta proteins	174	354	742
Alpha and beta proteins (α/β)	147	244	803
Alpha and beta proteins ($\alpha+\beta$)	376	552	1,055
Multi-domain proteins	66	66	89
Membrane and cell surface proteins	58	110	123
Small proteins	90	129	219
Total	1,195	1,962	3,902

Ilustrácia klasifikačnej schémy (postupnosti) v SCOP

Structural Classification of Proteins



Hierarchia v katalógu SCOP

Root: scop

Classes:

1. [All alpha proteins](#) [46456] (284)
2. [All beta proteins](#) [48724] (174)
3. [Alpha and beta proteins \(a/b\)](#) [51349] (147)
Mainly parallel beta sheets (beta-alpha-beta units)
4. [Alpha and beta proteins \(a+b\)](#) [53931] (376)
Mainly antiparallel beta sheets (segregated alpha and beta regions)
5. [Multi-domain proteins \(alpha and beta\)](#) [56572] (66)
Folds consisting of two or more domains belonging to different classes
6. [Membrane and cell surface proteins and peptides](#) [56835] (58)
Does not include proteins in the immune system
7. [Small proteins](#) [56992] (90)
Usually dominated by metal ligand, heme, and/or disulfide bridges
8. [Coiled coil proteins](#) [57942] (7)
Not a true class
9. [Low resolution protein structures](#) [58117] (26)
Not a true class
10. [Peptides](#) [58231] (121)
Peptides and fragments. Not a true class
11. [Designed proteins](#) [58788] (44)
Experimental structures of proteins with essentially non-natural sequences. Not a true class

Enter [search](#) key:

(Upravené podľa: <http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.html>)

SCOP štatistika

- verzia 1.75 z júna 2009;
- 110 800 domén;
- s vylúčením nukleových kyselín a teoretických modelov;
- 38 221 anotovaných PDB štruktúr.

CATH vs. SCOP

- SCOP a CATH nie sú úplne kompatibilné;
- jednotlivé kategórie nie sú definované rovnako;
- nie vždy sa totiž myslí to isté tvrdením, že „niečo je podobné“ alebo „niečo je rozdielne“;
- SCOP: 4 triedy, resp. CATH: 3 triedy.

Sumarizácia

- všetky proteíny majú spoločnú chemickú črtu: polypeptidový reťazec (kostru);
- každý proteín má ale rozdielnu sekvenciu bočných reťazcov;
- v dôsledku flexibility kostry môže mať reťazec neobmedzené množstvo konformácií;
- v prirodzenom proteíne ale interakcie medzi bočnými reťazcami vyselektujú jednu trojrozmernú štruktúru;
- skutočná terciárna štruktúra sa od ostatných možných odlišuje pri daných fyziologických podmienkach a teplote omnoho väčšou termodynamickou stabilitou;
- každý proteín má teda jedinečnú konformáciu;
- presná terciárna štruktúra je nevyhnutná pre plnenie biologickej funkcie proteínu.

- jednorozmerná genetická informácia;
- jednorozmerná aminokyselinová sekvencia;
- trojrozmerná štruktúra proteínu.

Natívny proteín pozostáva z 1 alebo viac domén (kompaktné zoskupenie aminokyselinových zvyškov s vlastnou funkciou).

Akýkoľvek úsek v proteíne:

- nízka voľná energia;
- husté zbalenie zvyškov dovnútra;
- uspokojenie potenciálu polárnych skupín (H-H väzby);
- ukrytie hydrofóbneho povrchu.

Kostra opisuje krivku traverzujúcu priestor zaberaný molekulou proteínu. Tvar tejto krivky zahŕňa štandardné elementy sekundárnej štruktúry.

7. Predikcie štruktúry proteínov

Vznik nových proteínov

- prirodzený vývoj proteínov
- súčasnosť
- génové manipulácie
- proteínové inžinierstvo
- proteínový dizajn

Sekvenovanie kompletných genómov

Haemophilus influenzae - 1,83 Mb - 1743 génov - júl 1995

Methanococcus jannaschii - 1,66 Mb - 1738 génov - august 1996

Saccharomyces cerevisiae - 12,07 Mb - 5885 génov - október 1996

Mycoplasma genitalium - 0,58 Mb - 470 génov - október 1995

Escherichia coli - 4,64 Mb - 4288 génov - september 1997

Bacillus subtilis - 4,21 Mb - 4100 génov - november 1997

Homo sapiens - ~3 000 Mb

- chromozóm 22 - 33,4 Mb - december 1999

- 99% - 3000 Mb - 23 tisíc génov - 1,5% kóduje proteíny

Kľúčová je terciárna štruktúra (3D)

Jej experimentálne určovanie je stále pomalé v porovnaní so stanovením primárnej štruktúry (aminokyselinovej sekvencie - 1D)

- údaje z konca augusta a začiatku septembra 2014:

Počet sekvencií (GenBank): ~ 174 miliónov



www.ncbi.nlm.nih.gov

(~166 giga báz)

Počet štruktúr (PDB): ~ 103 tisíc



www.rcsb.org

Existujú snahy o predikciu terciárnej štruktúry z poznania štruktúry primárnej. Toto je ale stále nie úplne vyriešený problém.

Pre počítačový program to znamená 2 možnosti:

- (i) hľadať medzi všetkými možnými konformáciami konformáciu s najnižšou energiou;
- (ii) hľadať určité vzory zvinutia pozdĺž polypeptidového reťazca.

Homologické proteíny

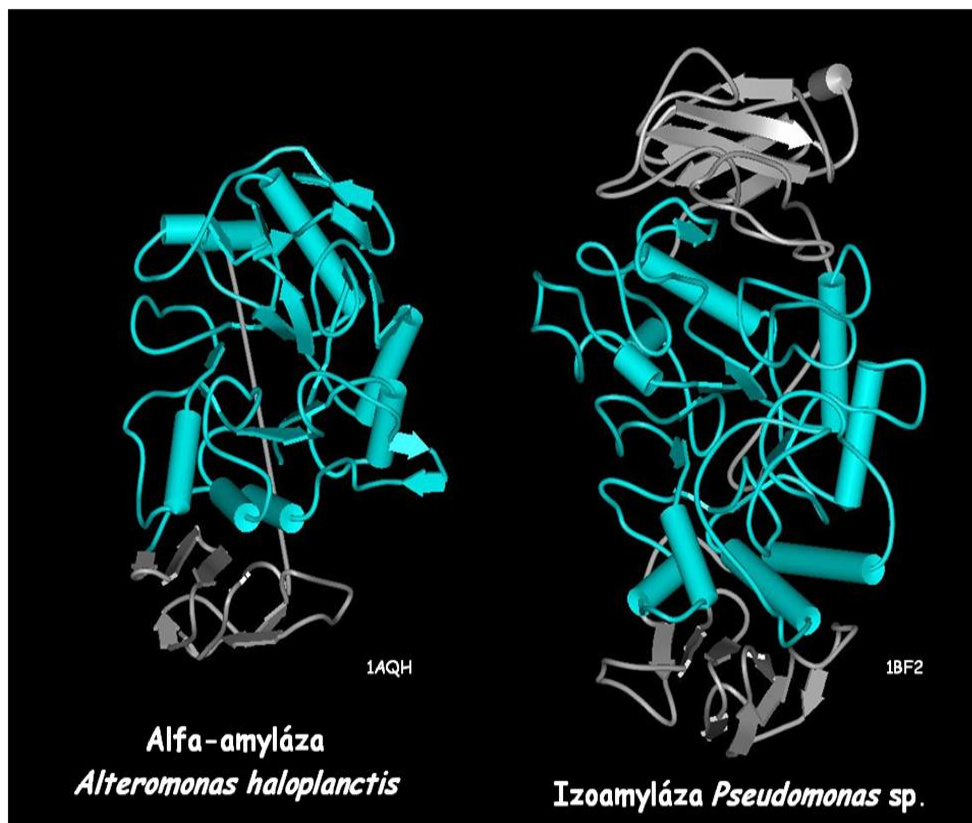
- proteíny s podobnými aminokyselinovými sekvenciami;
- proteíny s evolučnou príbuznosťou.

Máme proteín, ktorého sekvenciu (primárnu štruktúru) poznáme a chceme teoreticky určiť jeho štruktúru (3D).

V prípade predikcie terciárnej štruktúry „homologických proteínov“ už existuje proteín, ktorý je sekvenčne podobný s mojím proteínom, (t.j. je s ním homologický) a jeho terciárna štruktúra už bola vyriešená.

Homologické proteíny majú jadro (elementy sekundárnej štruktúry, vnútro) podobné. Aj vzdialene príbuzné proteíny majú hrubé rysy jadra (tzv. fold) podobné, ale majú nízku sekvenčnú podobnosť. Čím je vyššia sekvenčná podobnosť, tým viac príbuzné (podobné) sú aj trojrozmerné štruktúry. Toto je veľmi dôležité pri tvorbe modelov proteínov (pomocou známej terciárnej

štruktúry). Ak je identita sekvencií pod 20%, je veľmi ťažko predikovať (modelovať) štruktúru sekvencie pomocou známej štruktúry proteínu s podobnou sekvenciou.



Homologické proteíny - amylolytické enzýmy α -amyláza (PDB kód: 1AQH) a izoamyláza (PDB kód: 1BF2) - líšiac sa v najmä absenciou-prítomnosťou N-terminálnej domény (vpravo hore).

Slučky spájajú elementy sekundárnej štruktúry (helixy a vlákna), a zároveň varujú v dĺžke a sekvencii (ale aj štruktúre!); a to aj pre homologické proteíny.

Problém predikcie terciárnej štruktúry proteínu, ktorého sekvencia je homologická so sekvenciou proteínu so známou terciárnou štruktúrou, je vlastne otázka predikcie štruktúry pre slučky po tom, čo fold bol v hlavných rysoch určený (na základe homológie).

Slučky tiež nemajú náhodnú štruktúru. Štruktúra slučky závisí od:

- (i) počtu aminokyselín, ktoré ju tvoria;
- (ii) od typu sekundárnej štruktúry, ktoré spája ($\alpha \rightarrow \alpha$, $\alpha \rightarrow \beta$, $\beta \rightarrow \alpha$, $\beta \rightarrow \beta$);
- (iii) menej od aktuálnej sekvencie slučky.

Databáza slučiek. Čiže sa získa model jadra – fold (hrubý) – potom sú z databázy „dolepené“ slučky – získa sa hlavný reťazec – potom sa robia minimalizácie energie (výpočty) – získajú sa konformácie bočných reťazcov – a nasleduje spresnenie („doladenie“, refinement) modelu.

Neexistencia homologického proteínu

Čo vtedy, ak sa pre sekvenciu proteínu, ktorého štruktúru chceme teoreticky spoznať, nenájde žiadny homologický proteín, t.j. proteín s podobnou sekvenciou a už známou (vyriešenou) terciárnou štruktúrou?

- t.j., máme proteín, ktorého primárnu štruktúru (t.j. aminokyselinovú sekvenciu; 1D) poznáme a chceme teoreticky určiť jeho terciárnu štruktúru (3D), ale nemáme k dispozícii homologický proteín s už vyriešenou terciárnou štruktúrou

Je možné modelovať terciárnou štruktúru len zo samotnej sekvencie?

Dnes nie je možné získať presný model len zo znalosti aminokyselinovej sekvencie. Je to ale veľmi aktívna oblasť výskumu.

Je však možné získať aspoň predstavu o type proteínu (α , β , $\alpha+\beta$, resp. α/β).

Takéto prediktívne metódy spočívajú v predikcii sekundárnej (2D) štruktúry. To znamená: ktoré aminokyseliny sú v α -špirálach a ktoré v β -vláknach?

Teda základné obmedzenie pre predikciu terciárnej štruktúry je v nemožnosti 100% správne predpovedať sekundárnu štruktúru. Ak je však sekundárna štruktúra správne určená, terciárna štruktúra sa dá v dobrom priblížení predpovedať na základe zákonitostí ovládajúcich zbalenie elementov sekundárnej štruktúry, aspoň v zmysle, že sa určí malý súbor možných štruktúr.

Predikcia sekundárnej štruktúry je v centre predikcie terciárnej štruktúry.

Nanešťastie pre prediktívne metódy: oba typy štruktúry proteínov - aj sekundárna, aj terciárna - sú navzájom veľmi prepojené; terciárna štruktúra je vlastne sekundárna štruktúra v danej oblasti aranžovaná do priestoru.

To, ako sa daný úsek sekvencie proteínu zachová, t.j., či bude α -helix, β -vlákno alebo slučka, je ovplyvňované nielen sekvenciou, ale aj okolitým prostredím v priestorovej štruktúre. Napr. určitý sekvenčný región je v jednom proteíne α -helix, kým v druhom proteíne je to β -vlákno, prípadne slučka v inej štruktúre.

Táto medzizávislosť často komplikuje predikcie sekundárnej štruktúry proteínov, ale môže aj pomôcť tak, že sa predbežne priradí sekundárna štruktúra, ktorá pomôže v hrubom určení štruktúry terciárnej (t.j. foldu – napr. TIM-barel); čo následne obmedzí ďalšie možnosti pre sekundárnu štruktúru, ktorú možno takto spresniť.

Predikcie sekundárnej štruktúry proteínov

Rozdielne algoritmy:

- stereochemia
- štatistika
- neurónové siete

Stereochemický prístup

- založený na poznaní fyzikálno-chemických vlastností;
- dôraz na kompaktnosť štruktúry;
- predpoklad: tesne zbalené hydrofóbne jadro s polárnym povrchom.

Limova metóda – Lim V.I. (1974) Structural principles of the globular organization of protein chains. A stereochemical theory of globular protein secondary structure. *J. Mol. Biol.* **88**: 857-872.

Štatistický prístup

- založený na empirii;
- vychádza z analýzy známych terciárnych štruktúr;
- predpoklad: lokálna sekvencia určuje lokálnu štruktúru.

GOR metóda – Garnier J., Osguthorpe D.J. & Robson B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**: 97-120.

Predpoklad lokálnosti obmedzoval všetky metódy. Náhodná hodnota predikcie ~33%. Začiatkom 90. rokov: $65 \pm 2\%$ – maximum!

BIOINFORMATICS APPLICATIONS NOTE Vol. 21 no. 11 2005, pages 2787–2788 doi:10.1093/bioinformatics/bti408

Structural bioinformatics

GOR V server for protein secondary structure prediction

Taner Z. Sen,^{1,2} Robert L. Jernigan,^{1,2} Jean Garnier³ and Andrzej Kloczkowski^{1,*}

¹L.H. Baker Center for Bioinformatics and Biological Statistics, Iowa State University Ames, IA 50011, USA,

²Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011, USA

and ³INRA—Unite MIG, Bat. 233, Domaine de Vilvert, 78352 Jouy en Josas Cedex, France

Received on February 23, 2005; revised on March 21, 2005; accepted on March 22, 2005

Advance Access publication March 29, 2005

ABSTRACT

Summary: We have created the GOR V web server for protein secondary structure prediction. The GOR V algorithm combines information theory, Bayesian statistics and evolutionary information. In its fifth version, the GOR method reached (with the full jack-knife procedure) an accuracy of prediction Q_3 of 73.5%. Although GOR V has been among the most successful methods, its online unavailability has been a deterrent to its popularity. Here, we remedy this situation by creating the GOR V server.

Availability: The GOR V server is freely accessible to public users and private institutions at <http://gor.bb.iastate.edu/>

from crystallographically determined coordinates. Although DSSP defines eight different structural elements, these eight states are commonly translated into three secondary structure states: α -helix, β -sheet and coil. This translation is usually performed in the following manner: (1) α -helix in the three letter code corresponds to H (α -helix), G (3_{10} helix) and I (π -helix) from the DSSP 8-letter code, (2) sheet corresponds to B (bridge—single residue sheet), and E (extended β -strand) in DSSP nomenclature and finally, (3) coil in 3-letter code corresponds to the remaining three DSSP states: T (β -turn), S (bend) and C (coil). In the GOR V output, α -helix is represented by a letter H, β -sheet by E and coil by C.

GOR metóda na internete: <http://gor.bb.iastate.edu/> – 73,5%

IOWA STATE UNIVERSITY

PLANT SCIENCES INSTITUTE

Laurence H. Baker Center for Bioinformatics and Biological Statistics



!!! Please use our improved CDM Secondary Structure Prediction Server that also includes GOR V results !!!

GOR V PROTEIN SECONDARY STRUCTURE PREDICTION SERVER

The GOR (Garnier-Osguthorpe-Robson) method uses both information theory and Bayesian statistics for predicting the secondary structure of proteins. Over the years, the method has been improved by including larger databases and more detailed statistics, which account not only for amino acid composition, but also for amino acid pairs and triplets. The most crucial change in the algorithm was the inclusion of evolutionary information using PSI-BLAST (Altschul *et al.* Nucl. Acids Res. 25, 3389, 1997) to increase the information content for improved discrimination among secondary structures. In GOR V, the prediction accuracy Q₃ using full-jackknifing reached 73.5%.

Sequence name (optional):

Your e-mail address :

Paste a protein sequence below:
(Please use one-letter amino acid codes with no comment line--the submission is limited to 1000 residues)

Neurónové siete

- rok 1993 – Burkhard Rost >70% – metóda založená na neurónových sieťach;
- nový a kľúčový aspekt bol využitie evolučnej informácie vo forme zrovnania sekvencií, ktoré je použité namiesto jednotlivej sekvencie ako vstup pre predikciu;
- pre ~ 36% zvyškov je predikcia ~ 88%;
- sekvencia → zrovnanie → predikcia – výsledok je spoľahlivosť (0-9; reliability), že daný aminokyselinový zvyšok je α-špirála (H; helix), β-vláknko (E; extended structure) alebo slučka (L; loop).

PHD metóda – Profile network prediction Heidelberg – Rost B. & Sander C. (1993)
Prediction of protein secondary structure at better than 70% accuracy . *J. Mol. Biol.* **232**: 584-599.

The PredictProtein server

Burkhard Rost^{1,2,3,*}, Guy Yachdav^{1,3} and Jinfeng Liu^{1,2}

¹CUBIC and ²North East Structural Genomics Consortium (NESG), Department of Biochemistry and Molecular Biophysics, Columbia University, 650 West 168th Street BB217, New York, NY 10032, USA and ³Columbia University Center for Computational Biology and Bioinformatics (C2B2), Russ Berrie Pavilion, 1150 St Nicholas Avenue, New York, NY 10032, USA

Received February 13, 2004; Revised and Accepted March 15, 2004

ABSTRACT

PredictProtein (<http://www.predictprotein.org>) is an Internet service for sequence analysis and the prediction of protein structure and function. Users submit protein sequences or alignments; PredictProtein returns multiple sequence alignments, PROSITE sequence motifs, low-complexity regions (SEG), nuclear localization signals, regions lacking regular structure (NORS) and predictions of secondary structure, solvent accessibility, globular regions, transmembrane helices, coiled-coil regions, structural switch regions, disulfide-bonds, sub-cellular localization and functional annotations. Upon request fold recognition by prediction-based threading, CHOP domain assignments, predictions of transmembrane strands and inter-residue contacts are also available. For all services, users can submit their query either by electronic mail or interactively via the World Wide Web.

Attempt to simplify output by incorporating a hierarchy of thresholds

The attempt to 'pre-digest' as much information as possible to simplify the ease of interpreting the results is another unique pillar of PP. For example, by default PP returns only those proteins found in the database that are very likely to have a similar structure to the query protein (1). Particular predictions, such as those for membrane helices, coiled-coil regions, signal peptides and nuclear localization signals, are not returned if found to be below given probability thresholds. Over the years, we have added so many methods into the output of PP that our original 'easy-to-interpret' goal is challenged. We hope that a variety of improvements in the near future will reduce this problem.

Each request triggers the application of over 20 different methods

Currently, users receive a single output file with the following results (some of these are optional, Table 1). Database searches: similar sequences are reported and aligned by a

PHd metóda na internete: <http://www.predictprotein.org/>



Enter an amino-acid sequence (one letter code) or click a link in the bar on the right to see an example input.

PredictProtein

WHAT WE PROVIDE

Analysis Methods	PredictProtein integrates feature prediction for secondary structure, solvent accessibility, transmembrane helices, globular regions, coiled-coil regions, structural switch regions, B-values, disorder regions, intra-residue contacts, protein-protein and protein-DNA binding sites, sub-cellular localization, domain boundaries, beta-barrels, cysteine bonds, metal binding sites and disulphide bridges.
Visualization Features	PredictProtein features data visualizations using the Jbio interactive visualization library
Data Storage and Caching	PredictProtein caches the prediction for each new query sequence it processes for quick and easy retrieval. Currently the PredictProtein cache contains 8,425,519 annotated proteins.
Job Managment	PredictProtein users can store their jobs for later reference
Extensive Support	The service features: 1. On screen help tooltips 2. Complete user documentation 3. Online discussion board XML RSS



[Questions? Contact Us](#)

Licensing Options We currently provide two licensing options for your convenience:

1. [Commercial License](#) - SAFEGUARDING YOUR INTELLECTUAL PROPERTY
2. [Academic License](#) - SUPPORTING ACADEMIC STUDY AND RESEARCH

Predikcie terciárnej štruktúry proteínov

S ohľadom na existenciu homologických proteínov, resp. nemožnosť ich využitia – ak ich niet – predikčné metódy delíme na dve skupiny:

- (1) induktívne – využívajú databázy sekvencií a štruktúr (homologické proteíny);
- (2) deduktívne – *ab initio* prístupy (nevyžívajú databázy, neexistencia homologického proteínu).

Deduktívne metódy

- snaha predikovať štruktúru proteínu na základe všeobecných princípov fyziky, chémie a biológie, bez zrejmého využitia známych sekvencií a štruktúr;
- avšak vývoj aj týchto metód závisí na stave poznania známych sekvencií a štruktúr proteínov; iba sa táto znalosť priamo nevyužíva pri predikcii (iba pri metodike);
- rozdiel je v tom, že dosiahnuté poznanie sa premietne do všeobecných princípov, ktoré sa aplikujú bez „pozerania“ do databáz;
- deduktívne metódy sa zdokonaľujú veľmi pomaly.

Induktívne metódy

- tieto metódy sa rozvíjajú veľmi rýchlo a ich výkonnosť je taktiež veľká; sčasti závisí na náraste údajov v databázach;
- viac známych sekvencií a štruktúr znamená väčšiu pravdepodobnosť, že nový proteín (o predikciu štruktúry ktorého je záujem) bude podobný už známemu proteínu (so známou terciárnou štruktúrou).

Dva typy induktívnych metód:

- (i) homologické modelovanie (*homology modelling*);
- (ii) rozpoznanie štruktúry (*fold recognition*).

Homologické modelovanie

Predikcia cieľovej štruktúry na základe blízko príbuzného proteínu so

známou štruktúrou. Napr. Swiss-Model: <http://swissmodel.expasy.org/>

Schwede T., Kopp J., Guex N. & Peitsch M.C. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* 31: 3381-3385 .

Nucleic Acids Research, 2003, Vol. 31, No. 13 3381–3385
DOI: 10.1093/nar/gkg520

SWISS-MODEL: an automated protein homology-modeling server

Torsten Schwede^{1,2,*}, Jürgen Kopp^{1,2}, Nicolas Guex³ and Manuel C. Peitsch^{2,4}

¹Biozentrum der Universität Basel, Klingelbergstr. 50-70, CH 4056 Basel, Switzerland, ²Swiss Institute of Bioinformatics, Basel, Switzerland, ³GlaxoSmithKline, Research Triangle Park, NC 27709, USA and ⁴Novartis AG, 4002 Basel, Switzerland

Received February 15, 2003; Revised and Accepted March 20, 2003

ABSTRACT

SWISS-MODEL (<http://swissmodel.expasy.org>) is a server for automated comparative modeling of three-dimensional (3D) protein structures. It pioneered the field of automated modeling starting in 1993 and is the most widely-used free web-based automated modeling facility today. In 2002 the server computed 120 000 user requests for 3D protein models. SWISS-MODEL provides several levels of user interaction through its World Wide Web interface: in the 'first approach mode' only an amino acid sequence of a protein is submitted to build a 3D model. Template selection, alignment and model building are done

experimental protein structures are deposited in the Protein Data Bank (PDB) (1). Nevertheless, the number of structurally characterized proteins is low compared to the number of known protein sequences: taken together, the SWISS-PROT and TrEMBL databases hold about 850 000 sequence entries (2), which exceeds the number of known different structures by about two orders of magnitude. Thus, no structural information is available for the vast majority of protein sequences. Therefore, it is an obvious demand to bridge this 'structure knowledge gap' and computational methods for protein structure prediction have gained much interest in recent years.

Among all current theoretical approaches, comparative modeling is the only method that can reliably generate a 3D model of a protein (target) from its amino acid sequence (3,4). Successful model building requires at least one experimentally



Swiss Institute of Bioinformatics



BIOZENTRUM
Universität Basel
The Center for Molecular Life Sciences



SWISS-MODEL

Modelling

- myWorkspace
- Automated Mode
- Alignment Mode
- Project Mode

Tools

- Template Identification
- Domain Annotation
- Structure Assessment
- Template Library

Repository

- Search by Sequence
- Search by AC
- Search by full text

Documentation

- SWISS-MODEL Workspace
- SWISS-MODEL Repository
- Structures & Models
- Helpdesk

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists worldwide.

What's new?

- Find more news on [SWISS-MODEL Blog](#)
- ... faster news on [Twitter](#)
- Follow us on [Facebook](#)

SWISS-MODEL Team

Torsten Schwede: Project Leader
Florian Kiefer: SWISS-MODEL Repository
Lorenza Bordoli: Method Development and user support
Konstantin Arnold: SWISS-MODEL Workspace

References:
When you publish or report results using SWISS-MODEL, please cite the relevant publications:

- Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22, 195-201.
- Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*, 37, D387-D392.
- Peitsch, M. C. (1995) Protein modeling by E-mail *Bio/Technology* 13: 658-660.

Rozpoznanie štruktúry

Zhodnotenie kompatibilnosti študovanej aminokyselinovej sekvencie s knižnicou známych vzorov zvinutia proteínov.

Napr. PHYRE: <http://www.sbg.bio.ic.ac.uk/phyre2/>

Kelley L.A. & Sternberg M.J. (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols* 4: 363-371.

PROTOCOL

Protein structure prediction on the Web: a case study using the Phyre server

Lawrence A Kelley & Michael J E Sternberg

Structural Bioinformatics Group, Division of Molecular Biosciences, Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. Correspondence should be addressed to L.A.K. (l.a.kelley@imperial.ac.uk).

Published online 26 February 2009; doi:10.1038/nprot.2009.2

Determining the structure and function of a novel protein is a cornerstone of many aspects of modern biology. Over the past decades, a number of computational tools for structure prediction have been developed. It is critical that the biological community is aware of such tools and is able to interpret their results in an informed way. This protocol provides a guide to interpreting the output of structure prediction servers in general and one such tool in particular, the protein homology/analogy recognition engine (Phyre). New profile–profile matching algorithms have improved structure prediction considerably in recent years. Although the performance of Phyre is typical of many structure prediction systems using such algorithms, all these systems can reliably detect up to twice as many remote homologies as standard sequence–profile searching. Phyre is widely used by the biological community, with >150 submissions per day, and provides a simple interface to results. Phyre takes 30 min to predict the structure of a 250-residue protein.

NATURE PROTOCOLS | VOL.4 NO.3 | 2009 | 363

Phyre²

Protein Homology/analogy Recognition Engine V 2.0

[Subscribe to Phyre at Google Groups](#)
Email:
[Visit Phyre at Google Groups](#)



[What's New in Phyre2](#)

E-mail Address

Optional Job description

Amino Acid Sequence 

Modelling Mode 

Normal ☒ Intensive ☐

331214 submissions since Feb 14 2011

Phyre is for **academic use only**

Please cite: Protein structure prediction on the web: a case study using the Phyre server
Kelley LA and Sternberg MJE. *Nature Protocols* 4, 363 - 371 (2009) [[pdf](#)] [[Import into BibTeX](#)]

© Structural Bioinformatics Group, Imperial College, London
[Lawrence Kelley](#), [Benjamin Jefferys](#)



8. Molekulová evolúcia proteínov

Jedným zo základov molekulárnej biológie je fakt, že aminokyselinová sekvencia proteínu určuje jeho štruktúru, a to znamená aj jeho funkciu.

Jednorozmerná genetická informácia sa preloží do trojrozmerného sveta, ktorý obývame.

Zákonitosti výstavby molekúl proteínov:

- hierarchia;
- štruktúry (primárna, sekundárna, terciárna, ...);
- motívy, domény;
- vzory zvinutia štruktúry.

Rozlišujeme:

- protein folding - *proces* zaujatia natívnej štruktúry;
- protein folding motif – *motív* zvinutia natívnej štruktúry (~fold, napr. TIM-barel).

Klasifikácia proteínov – na základe ich sekundárnej a terciárnej štruktúry (all α , all β , α/β , $\alpha+\beta$).

Klasifikácia z hľadiska štruktúry – do jednotlivých rodín proteínov – odráža evolúciu.

Divergentná evolúcia – zo spoločného predka – podobné:

- aminokyselinové sekvencie;
- priestorové (terciárne) štruktúry;
- funkcie.

Konvergentná evolúcia – vo všeobecnosti nezávislá tvorba výhodnej štruktúry, resp. štruktúrnych črt, pri ktorej:

- priestorové štruktúry môžu byť podobné;
- aminokyselinové sekvencie a funkcie sú nepodobné.

Porovnanie sekvencií proteínov

Aký je vzájomný vzťah medzi dvomi a viac proteínmi?

Je potrebné určiť korešpondencie medzi aminokyselinovými sekvenciami, t.j. navzájom si odpovedajúce úseky, t.j. zrovnanie sekvencií (amino acid sequence alignment).

- párové zrovnanie (pair-wise alignment);
- násobné zrovnanie (multiple alignment).

Je nevyhnutné uvažovať:

- delécie;
- inzercie;
- substitúcie.

Násobné zrovnanie – multiple alignment:

- viac sekvencií;
- viac informatívny;
- odhaľuje vzory, sekvenčné motívy, konzervované regióny;
- naznačuje, čo je, resp. čo by mohlo byť významné z hľadiska štruktúry a funkcie porovnávaných proteínov.

Pre vzdialene príbuzné proteíny môže byť nemožné určiť správne zrovnanie len zo sekvencií. Je to dôsledok toho, že trojrozmerná štruktúra proteínu je viac konzervovaná ako aminokyselinová sekvencia. Preto sa porovnávajú štruktúry proteínov.

Vzniká tzv. background noise (šum pozadia), resp. hovoríme o náraste šumu pozadia (tzv. background noise increase). Šum pozadia pri zrovnávaní viacerých sekvencií menej príbuzných proteínov je vždy, to znamená, že nedokážeme spraviť ich jediné správne zrovnanie – zrovnáme konzervované oblasti a zvyšky sekvencií (môže to byť podstatná časť ich dĺžok) sú zrovnané len s určitou pravdepodobnosťou.

Zrovnanie založené na štruktúre (structure-based alignment)

- hľadajú sa korešpondencie medzi zvyškami, ktoré majú tú istú priestorovú dispozíciu;
- môže odhaliť evolučnú príbuznosť medzi tak veľmi vzdialene príbuznými proteínmi, že na úrovni sekvencií už nič nemožno detekovať.

Primárne udalosti v tvorbe biologickej diverzity:

- mutácie, inzercie a delécie nukleotidov v DNA;
- transpozície (väčšie preskupenia kusov genetického materiálu).

Štádiá a zmeny v evolúcii proteínových štruktúr:

- príbuzné proteíny si zachovávajú väčšinu elementov sekundárnej štruktúry (α -špirály a β -vlákna);
- jadro štruktúry – zhluk centrálnych α -helixov a/alebo β -listov – si udržiava svoju topológiu (vzor zvinutia):
 - pre blízko príbuzné proteíny jadro predstavuje takmer celú štruktúru
 - pre vzdialene príbuzné proteíny jadro môže byť tvorené menej ako polovicou zvyškov
- okrajové regióny mimo jadra môžu meniť svoj vzor zvinutia úplne;
- relatívna geometria elementov sekundárnej štruktúry, dokonca aj v jadre, sa mení;
- v dôsledku mutácií sa α -helixy a β -listy môžu posúvať alebo navzájom rotovať;

- pre evolúciu, pri ktorej sa konzervuje funkcia, štrukturálne zmeny sú obmedzované, aby sa chránila funkcia, napr. udržanie integrity aktívneho miesta;
- pre evolúciu, pri ktorej sa funkcia mení, tie obmedzenia sú nahradené inými obmedzeniami, ktoré sú vyžadované zmenenou funkciou, čo rezultuje vo väčšej štrukturálnej zmene.

Molekulárna úroveň

- vznik – široká jednota (jednotnosť) – DNA → proteín (všetko je rovnaké);
- vývoj – evolúcia – veľká diverzita (rozmanitosť) – DNA → proteín (každý je iný).

Možno dať do súvisu molekulové údaje s výsledkami tradičných makroskopických evolučných štúdií?

- resp. do akej miery molekulové údaje potvrdzujú, dopĺňajú alebo modifikujú naše pochopenie tradičných fylogenetických vzájomných vzťahov, získaných z porovnávacej anatómie, embryológie a fosílií?

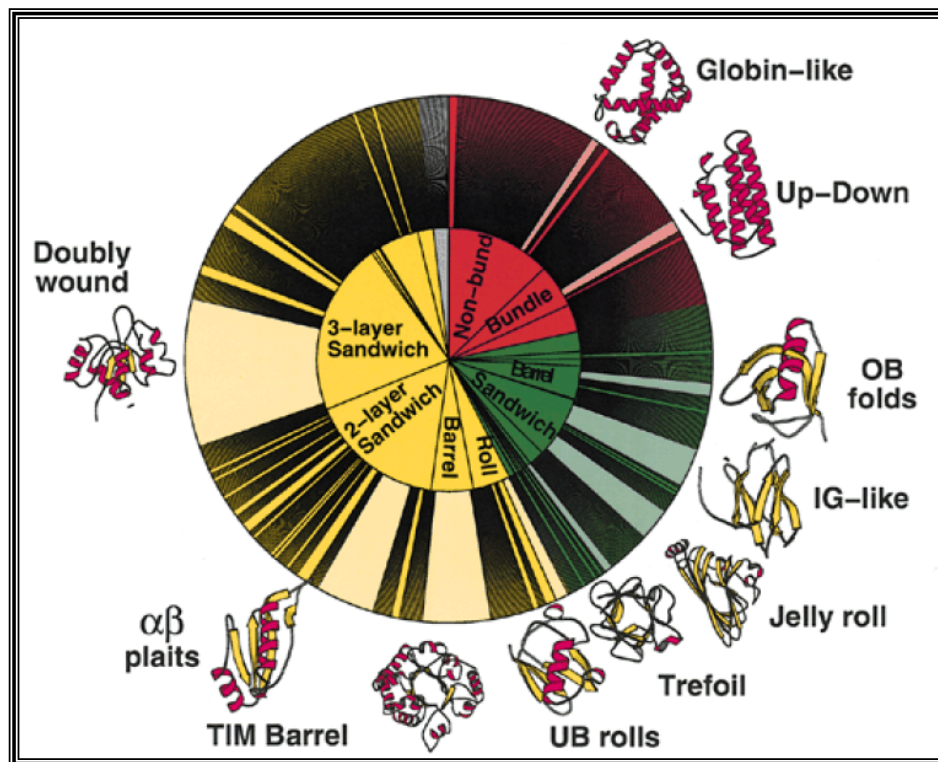
Divergencia sa deje paralelne na molekulárnej, aj makroskopickej úrovni

Homologické proteíny z príbuzných druhov majú podobné, ale nie identické aminokyselinové sekvencie.

Vo všeobecnosti:

- $>10^x$ génov ($x \geq 5$) → $\sim 10^4$ sekvenčných rodín → $\sim 10^3$ štrukturálnych rodín → $\sim 10^2$ architektúr → ~ 10 motívov (! tzv. superfoldy).

Proteínové motívy (superfoldy).



(Upravené podľa: Orenge a kol. *Structure* 1997, 5: 1093)

Divoký gén (tzv. wild-type gene) a mutácie:

- (i) alternatívny proteín s ekvivalentnou funkciou (neutrálna mutácia);
- (ii) proteín s tou istou funkciou, ale so zmeneným rýchlostným alebo špecifitným profilom;
- (iii) proteín so zmenenou funkciou;
- (iv) proteín nefunkčný, prípadne nezvinutý (letálna mutácia).

Príklad:

- divoký gén kóduje α -amylázu – štiepi α -1,4-glukozidové väzby v škrobe;
- mutáciami môže vzniknúť:
 - (i) izozým α -amylázy s nezmenenou funkciou;
 - (ii) príbuzná α -amyláza, ktorá štiepi tie isté väzby v tom istom substráte (škrob), ale inou rýchlosťou, prípadne je viac špecifická na rovnaké väzby napr. v kratších

maltooligosacharidoch – G3, G4, G5, G6 a G7 (t.j. zmena v špecifite);

- (iii) pululanáza neschopná štiepiť väzby α -1,4, ale schopná štiepiť α -1,6-glukozidové väzby v pululane;
- (iv) mutáciou v katalytickom mieste (napr. Asp→Asn) proteín nie je schopný katalyzovať danú biochemickú reakciu vôbec, alebo sú problémy s jeho expresiou, prípadne adoptuje nesprávnu konformáciu.

Divergentná evolúcia je evolúcia zo spoločného predka (common ancestor).

Homologické sekvencie:

- sekvencie, ktoré majú spoločného evolučného predka;
- inak, pokiaľ spoločný pôvod nie je zrejmý, sa podobné sekvencie označujú iba ako podobné.

Ortológy a paralógy

- ortológy – homologické sekvencie existujúce v rôznych druhoch ako dôsledok vývoja druhov;
- paralógy – homologické sekvencie existujúce v tom istom druhu ako dôsledok multiplikácií génov.

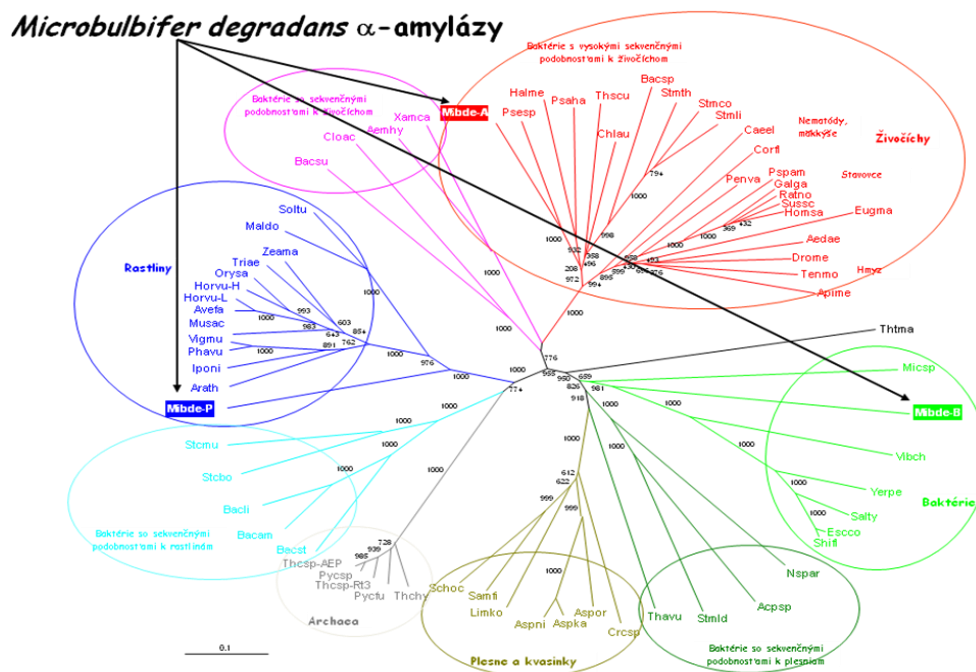
Horizontálny prenos génov

Horizontálny prenos génov (horizontal or lateral gene transfer) je proces, v ktorom organizmus prenáša genetický materiál (DNA) to inej bunky, ktorá nie je jeho potomstvom. Pri klasickom vertikálnom prenose organizmus získava genetický materiál od svojho predka, t.j. od rodiča alebo druhu, z ktorého pochádza (resp. sa vyvinul). V genetike sa uvažovalo najmä o vertikálnom prenose génov, ktorý je dominantný, avšak podľa najnovšieho poznania je zrejmé, že horizontálny prenos je najmä u mikroorganizmov významným fenoménom.

Príklady

Microbulbifer degradans – tri kópie α -amylázy:

- bakteriálny typ (*Escherichia coli*: SI=34.1%, SS=49.8%);
- rastlinný typ (jačmeň: SI=42.6%, SS=55.9%);
- živočíšny typ (človek: SI=45.1%, SS=56.4%).



(Upravené podľa: Da Lage a kol. *Cell. Mol. Life Sci.* 61: 97)

α -Amylaza	Cytochrome oxidase	Maltogénna α -amyláza	Maltotetraohydroláza	Maltotriohydroláza	Maltotriohydroláza syntáza
Heopululanáza	Maltogénna amyláza	Cytochrome oxidase	Amylosacharáza	Sacharóza fosforyláza	Maltotriohydroláza
Oligo-1,6-glukozidáza	Izomaltulóza syntáza	Izoamyláza	Glukan vetviaci enzým	4- α -Glukanotransferáza	Maltotriohydroláza

α -Amylázová rodina – galéria terciárnych štruktúr príbuzných enzýmov

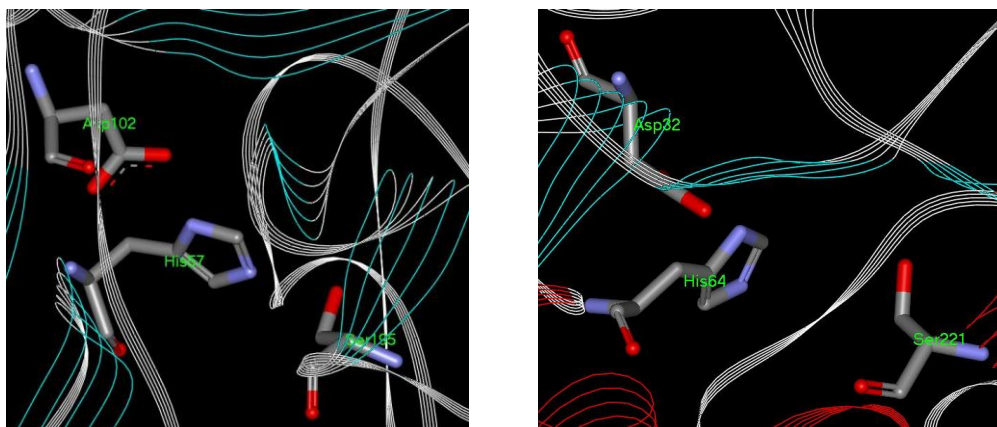
Konvergentná evolúcia je evolúcia k niečomu spoločnému

Konvergencia je adaptívna zmena, v ktorej menej príbuzné entity sa zdajú byť viac príbuzné, ako v skutočnosti sú.

Funkčná konvergencia – rovnaká molekulová funkcia sa objaví nezávisle viac ako 1× (napr. metaloproteázy, Asp-proteázy, Ser-proteázy).

Mechanistická konvergencia – rovnaký chemický mechanizmus – napr. serínové proteázy chymotrypsín (s katalytickou triádou His57, Asp102, Ser195) a subtilizín (Asp32, His64 a Ser 221).

Štruktúrna konvergencia – rovnaká štruktúra – vnútorná stabilita, jednoduchá tvorba, napr. TIM-barelové proteíny – TIM barelová (katalytická) doména sa vyskytuje v cca 10% proteínov (enzýmov), ktorých terciárna štruktúra už bola určená.



Príklad mechanistickej konvergenzie: prasačí trypsín (vľavo): His57, Asp102 a Ser195 (PDB kód: 1MCT), resp. subtilizín z *Bacillus subtilis* (vpravo): Asp32, His64 a Ser221 (PDB kód: 1SUP). Obe dva enzýmy sú serínové proteínázy, ale poradie v katalytickej triáde je rozdielne, t.j. nie je možné tieto štruktúry preložiť tak, aby si navzájom korešpondovali katalytické zvyšky.

9. Stabilita a stabilizácia proteínov

Rovnováha medzi stabilizujúcimi a destabilizujúcimi interakciami, t.j. rozdiel v Gibbsovej energii medzi zvinutým a nezvinutým proteínom je ~40 kJ/mol.

Extrémoustabilné proteíny – nič nové, t.j. žiadna iná aminokyselina, ani žiadna iná interakcia.

Väzby a interakcie v proteíne:

- vodíkové väzby;
- hydrofóbne interakcie;
- soľné mostíky;
- disulfidové mostíky.

Všetky tieto väzby a interakcie sa musia v proteíne stabilnom v extrémnom prostredí zosilniť, alebo sa musí ich počet zvýšiť.

Ako získať stabilné enzýmy?

- 1. izolácia z extrémofilných (mikro-)organizmov;
- 2. produkcia v geneticky manipulovaných neextrémofilných hostiteľských producentoch;
- 3. stabilizácia nestabilných enzýmov modifikáciou:
 - imobilizácia;
 - chemická modifikácia;
 - proteínové inžinierstvo;
- 4. použitie aditív.

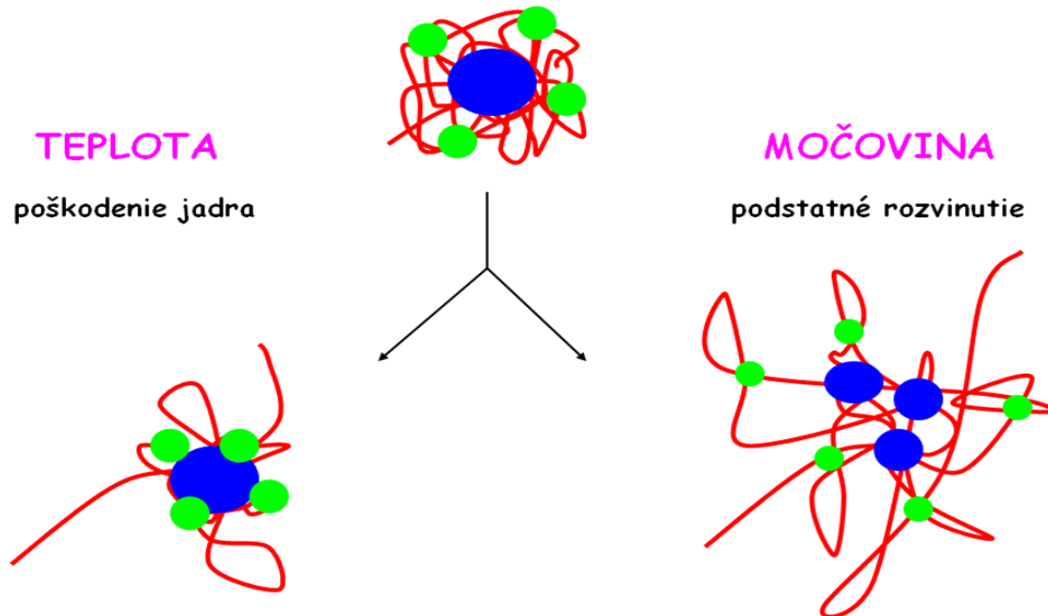
GMO – geneticky modifikované organizmy

Treba brať do úvahy riziko spôsobené možným uvoľnením GMO do životného prostredia.

GRAS – generally recognised as safe

Stabilizácia proteínov – je to najlepší spôsob ako získať stabilné proteíny (hoci nie vždy ten najjednoduchší a najefektívnejší).

Denaturácia proteínov



(Upravené podľa: Janecek *Process Biochem.* 1993, 28: 435)

Imobilizácia proteínov

- premena z mobilného stavu na nemobilný (imobilizovaný) stav;
- stabilizácia sa dosahuje v dôsledku:
 - fixácie (proti agregácii alebo autolýze);
 - zvýšenia rigidity (proti rozvinutiu);
 - ochrany pred inaktiváciou (pH, O₂, H₂O₂ ...).

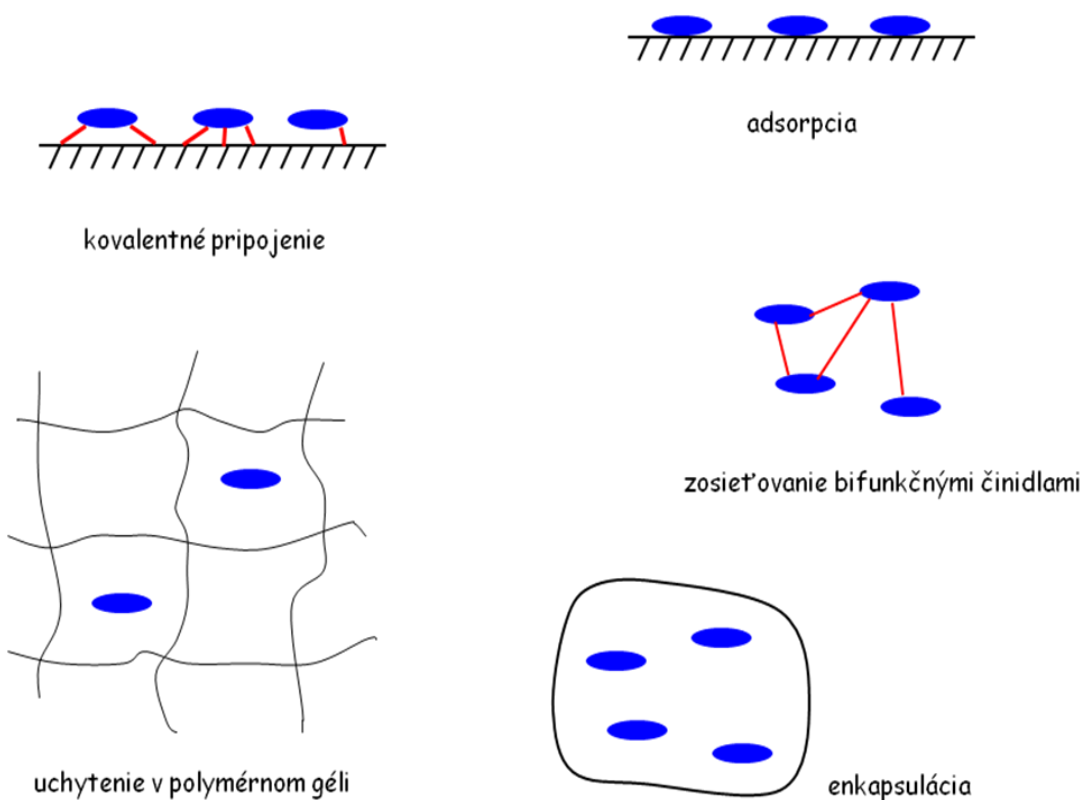
Výhody imobilizácie:

- kontinuálnosť procesu;
- vysoká stabilita;
- vysoká reakčná rýchlosť;
- vysoká koncentrácia enzýmu;
- jednoduchá separácia;
- násobné použitie.

Nevýhody:

- odpor proti prestupu látky;
- imobilizovanie;
- cena imobilizácie.

Ilustračné príklady imobilizácie proteínov.



(Upravené podľa: Klibanov *Science* 1983, 219: 722)

Chemická modifikácia proteínov

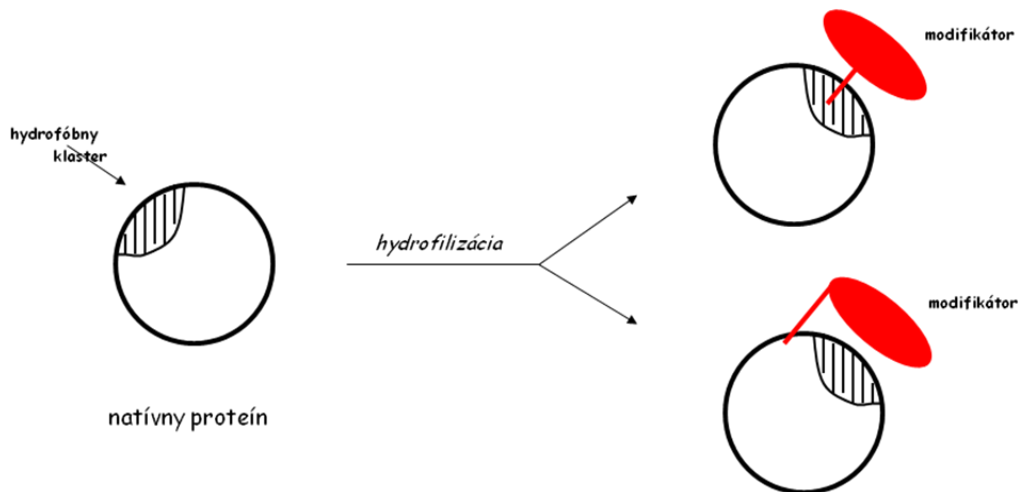
- ide o zachovanie pohyblivej (mobilnej) formy;
- vonkajšie (povrchové) funkčné skupiny ľahko reagujú s modifikujúcimi činidlami.

Možnosti:

- modifikácia s bifunkčnými činidlami (zosieťovanie);
- modifikácia s nepolárnymi činidlami (hydrofóbne interakcie);

- zavedenie nových polárnych alebo nabitých skupín (doplňujúce vodíkové a iónové väzby);
- hydrofilizácia povrchu (ochrana pred nevýhodným hydrofóbnym efektom – redukcia nevýhodného kontaktu s vodou).

Schéma stabilizácie chemickou modifikáciou hydrofilizácie povrchu proteínu.



(Upravené podľa: Mozhaev a kol. *Eur. J. Biochem.* 1988, 173: 147)

Proteínové inžinierstvo

- vo všeobecnosti možno napr. meniť:
 - kinetické vlastnosti;
 - termostabilita a T_{opt} ;
 - stabilita a aktivita v organických rozpúšťadlách;
 - substrátová a produktová špecificita;
 - požiadavky na kofaktory;
 - pH optimum;
 - odolnosť voči proteázam;
 - molekulová hmotnosť a podjednotková (kvartérna) štruktúra.

Ako to možno meniť:

- najmä metódy cielenej mutagenézy – napodobňovanie prirodzene stabilných proteínov alebo na základe predikcie stabilizujúcich efektov.

Príklady:

- zavedenie vnútorných alebo povrchových disulfidových mostíkov – konformačná stabilita – vhodná pozícia v štruktúre (aby S-S mostík mohol vzniknúť);
- zlepšenie vnútorného hydrofóbného zbalenia (Leu→Val – $1 \times -CH_2$ medzera destabilizuje $\sim 4,6$ kJ/mol);
- nárast vnútorných vodíkových väzieb – príspevok môže byť porovnateľný s hydrofóbnym efektom;
- nárast povrchových soľných mostíkov (izolovaný náboj vo vnútri proteínu môže destabilizovať >15 kJ/mol);
- aromatické interakcie – zavedenie aromatických zvyškov (tzv. „stacking“ interakcie).

10. Faktory stability proteínov

Rigidita

- rigidita – je nevyhnutná na ochranu štruktúry aktívnych centier a chráni ich pred rozvinutím;
- aminokyselinové substitúcie – nie sú systematické, ale z porovnania veľkého počtu enýmov je možné zovšeobecniť, že stabilizujúce substitúcie: (i) zdokonaľujú účinnosť zbalenia štruktúry (vyplnením dier/kavít v jadre a zvýšením jeho hydrofóbnosti) a (ii) zvyšujú jej celkovú rigiditu (stabilizácia α -špirál, optimalizácia elektrostatických interakcií a redukcia konformačného pnutia).

Hydrofóbne interakcie

- poskytujú energiu nevyhnutnú pre zvinutie proteínu vo vodných roztokoch;
- v priemere každá ukrytá $-\text{CH}_2$ skupina prispieva ku konformačnej stabilite príspevkom $\sim 5,4$ kJ/mol;
- tesne zbalené aromatické „jadro-jadro“ interakcie (aromatické interakcie – „stacking“) $\sim 4,2$ kJ/mol;
- napr. 3-izopropylmalátdehydrogenáza z *Thermus thermophilus* obsahuje dva hydrofóbne zvyšky (Leu246 a Val249) na dimérovom medzipovrchu, ktoré sú v mezozýme z *Escherichia coli* nahradené menej hydrofóbnymi zvyškami (Glu a Met). Výsledkom je ľahšia disociácia diméru z *E. coli* vplyvom močoviny. Zavedenie Leu246 a Val249 do *E. coli* enzýmu ho naozaj stabilizovalo voči disociácii a denaturácii.

Vodíkové väzby

- dlho známy príspevok vodíkových väzieb – podporený cieľenou mutagenézou – mnoho prípadov;
- jedno z najlepších porovnaní bolo umožnené porovnaním štruktúr glyceraldehydfosfátdehydrogenázy z psychrofila (morský rak), mierneho termofila (*Bacillus stearothermophilus*), termofila (*Thermus aquaticus*) a hypertermofila (*Thermotoga maritima*);
- najlepšia korelácia bola nájdená medzi termostabilitou a počtom vodíkových väzieb typu „nabitý-neutrálny“;
- tento typ vodíkových väzieb je v termozýmoch uprednostňovaný pred soľnými mostíkmi a H väzbami typu „neutrálny-neutrálny“.

Účinnosť zbalenia

- keďže jadrá proteínov sú zväčša hydrofóbne, zvýšená účinnosť zbalenia je často v korelácii so zvýšenou hydrofóbnosťou;
- potenciál zvýšiť termostabilitu proteínu vyplnením hydrofóbných dier (kavít) v zvinutej štruktúre (t.j. zvýšením hydrofóbnosti zbalenia štruktúry) bol demonštrovaný cieľenou mutagenézou vychádzajúcou z detailnej znalosti trojrozmernej štruktúry;
- napr. tvorba vnútorných kavít mutáciami (zámenami Leu a Phe za Ala) destabilizovala T4 fágový lyzozým s priemerným príspevkom ~8,4 kJ/mol, pričom pri dvojito mutante L99A/F153A bol pokles v stabilite ~34,8 kJ/mol;
- ukryté polárne aminokyselinové zvyšky však nemusia nevyhnutne destabilizovať proteín – A166S substitúcia v neutrálnej proteáze z *Bacillus stearothermophilus* viedla k tomu, že –OH skupina Ser nahradila ukrytú molekulu H₂O a zdokonalila lokálnu sieť vodíkových väzieb, čím bola zvýšená aj účinnosť zbalenia, aj stabilita.

Soľné mostíky

- veľké množstvo experimentálnych štúdií identifikovalo soľné mostíky ako stabilizujúce interakcie;
- často stabilizujú elementy sekundárnej štruktúry, čím zvyšujú rigiditu celkovej štruktúry;
- štatisticky bolo zistené, že termozýmy majú vyšší pomer Arg:Lys ako mezozýmy (Arg je lepší v udržiavaní soľných mostíkov pri vyšších teplotách – má vyššiu hodnotu pK_a ako Lys a dlhší postranný reťazec);
- napr. fosfoglycerátkináza z *Bacillus stearothermophilus* a z kvasinky sa líši o 15 soľných mostíkov;
- glyceraldehydfosfátdehydrogenáza z *Thermus aquaticus* má stabilizujúci soľný Arg mostík (Arg22 a Arg248), čo chýba v menej termostabilnom enzýme z *Bacillus stearothermophilus*, ktorý má na korešpondujúcich miestach dva Lys zvyšky s kratším reťazcom – neschopné analogickej stabilizujúcej interakcie.

Stabilizácia slučiek

- slučky a ohyby boli dlho považované za slabé miesta štruktúry proteínov medzi elementami jadra;
- veľmi pravdepodobne sa proces rozvinutia pri teplotnej denaturácii začína práve na slučkách;
- slučky termozýmov sú často kratšie než slučky v odpovedajúcich mezozýmoch;
- skrátenie slučiek zvyšuje kompaktnosť štruktúry a redukuje jej flexibilitu (priamy efekt);
- nepriamy efekt je, že slučky a ohyby sa podieľajú na stabilizácii interakcií medzi elementami sekundárnej štruktúry a jednotlivými doménami;
- prolín môže udržiavať slučky alebo ohyby spôsobom podobným zipsu;

- napr. citrátsyntázy z *Pyrococcus furiosus* a *Thermoplasma acidophilum* obsahujú skrátenu verziu slučky, ktorá je prítomná v „nestabilnej“ citrátsyntáze z prasťa.

Redukcia konformačného pnutia

- neglycínové zvyšky prítomné v proteínoch, pokiaľ nie sú stabilizované intramolekulovými nekovalentnými interakciami, svojou ľavotočivou špirálovou konformáciou často spôsobujú nestabilitu;
- lokálne konformačné pnutie vzniká v dôsledku blízkeho kontaktu β -uhlíka s karbonylovým kyslíkom;
- v týchto prípadoch často pomôže substitúcia za Gly, najmä ak sú zvyšky spôsobujúce pnutie v ohybových oblastiach spájajúcich elementy sekundárnej štruktúry;
- napr. DNA-väzbový proteín HU z *Bacillus subtilis* obsahuje Glu15, kým jeho termostabilný analóg z *Bacillus stearothermophilus* na korešpondujúcom mieste v ohybe medzi dvomi α -špirálami obsahuje Gly, t.j. mutácia Gly15Glu v enzýme z *B. stearothermophilus* otvára $\alpha 1$ -ohyb- $\alpha 2$ konformáciu, čím potláča jednu zo stabilizujúcich vodíkových väzieb (helix-ohyb-helix).

Stabilizácia α -špirál

- prístup je napr. substitúcia tých zvyškov v α -špirálach, ktoré majú nízku schopnosť vytvárať α -špirálové úseky (alebo zabudovávať sa do nich), za zvyšky, ktoré majú vysoké predpoklady pre α -špirály (Ala);
- nie je však zistený ich skutočný príspevok, ktorý je nízky (pod 4 kJ/mol);
- alanínový zvyšok dominuje v špirálach termozýmov;
- do úvahy sa berie aj tzv. helixový dipól – substitúcie na koncoch špirál (elektrostatické interakcie) – helix má pozitívny náboj na N-konci

a negatívny náboj na C-konci (helixový dipól; destabilizačný vplyv) – kompenzácia je obsadením N-konca negatívne nabitým zvyškom (napr. Glu) – stabilizácia dipólu;

- menej bežné (a menej efektívne) je obsadenie C-konca pozitívne nabitým zvyškom (napr. Lys).

Redukcia entropie rozvinutia

- v nezvinutom stave proteínu je to Gly (nemá β -uhlík), ktorý je zvyškom s najväčšou konformačnou entropiou;
- v procese zvinutia je preto oveľa viac energie treba na vymedzenie presnej konfigurácie pre Gly ako pre všetky ostatné zvyšky;
- Pro so svojim pyrrolidínovým kruhom môže adoptovať iba niekoľko konfigurácií, ktoré obmedzujú konfigurácie dovolené pre predchádzajúci zvyšok;
- substitúcie Gly najmä za Pro v akomkoľvek mieste by mala znížiť entropiu proteínu v nezvinutom stave;
- aby však došlo k stabilizácii, je potrebné starostlivo vybrať vhodné miesto, čo nie je jednoduché (3-D štruktúra);
- mutačné miesto by malo dovoliť zvyšku konformáciu, ktorá je podobná pre dovolené konformácie nového zvyšku, pričom nový zvyšok by nemal vytvárať objemové interferencie a nemal by rušiť stabilizujúce nekovalentné interakcie;
- napr. T4 lyzozým – účinok Pro zvyškov na stabilizáciu proteínu – Ala82Pro a Ile3Pro mutanty: (i) v prvom prípade (Ala82Pro) boli podmienky starostlivo preskúmané a mutácia stabilizovala proteín najmä znížením entropie rozvinutia, a (ii) ale v druhom prípade (Ile3Pro) substitúcia eliminovala hydrofóbnú interakciu tvorenú Ile s výsledkom, že celkovo mutácia bola destabilizujúca, pretože stupeň entalpickej destabilizácie bol väčší než entropický stabilizujúci príspevok zo zavedenia prolínu.

Rezistencia voči kovalentnej deštrukcii

- nevratná denaturácia molekuly proteínu vplyvom kovalentnej modifikácie vyvolanej teplotou je aktuálna iba pri vysokých (zvyšných) teplotách;
- deštrukcia a oxidácia Cys, deamidácia Asn a Gln, hydrolýza peptidových väzieb;
- kým mezozýmy sa inaktivujú rozvinutím (kolaps vedúci k inaktívnej forme proteínu), termozýmy sa inaktivujú kovalentnou modifikáciou;
- napr. α -amylázy z *Bacillus amyloliquefaciens* a *Bacillus stearothermophilus* sú inaktivované rozvinutím, za inaktiváciu rigidnejšej a stabilnejšej α -amylázy z *Bacillus licheniformis* je zodpovedná deamidácia;
- platí, že viac stabilné enzýmy sú viac odolné voči kovalentnej modifikácii;
- napr. disulfidové mostíky stabilizujúce proteíny pri teplotách ~60-80 °C sú často neprítomné v termozýmoch izolovaných z hypertermofilov;
- hydrolýza peptidovej väzby medzi Asn-Xaa (doprevádzaná uvoľnením amoniaku) je bežný spôsob inaktivácie termozýmov.

11. Extrémofilné mikroorganizmy a extrémstabilné proteíny

Podľa optimálnej rastovej teploty možno mikroorganizmy klasifikovať do štyroch skupín:

- (1) hypertermofily (nad 80 °C);
- (2) termofily (50-80 °C);
- (3) mezofily (20-50 °C);
- (4) psychofily (0-20 °C).

Extrémofilné mikroorganizmy

Kde je horný (dolný) limit pre život?

- teplota:
 - >100 °C (hypertermofily; horúce pramene)
 - ~ 15 °C (psychofily; zemské póly)
- pH:
 - ~1 (acidofily; kyslé sírne pramene)
 - ~9-11 (alkalifily; soľné jazerá)
- salinita:
 - ~20% (halofily; extra slané moria)
- tlak:
 - ~700 atm (barofily; priekopy v oceánoch)

Enzýmy z termofilných mikroorganizmov sú termozýmy.

Sú to enzýmy, ktoré sa vyvinuli:

- v termofiloch (rast pri 60-80 °C);
- v hypertermofiloch (rast nad 80 °C).

Termozýmy – $T_{\text{opt}} \sim 60-125$ °C – záujem:

- vedecký (stabilita);
- priemyselný (biotechnológie).

Je záujem aj o psychrofilny (nizko-teplotné extrémofily) a ich enzýmy (enzýmy aktívne pri nízkych teplotách; „chlad-aktívne enzýmy“):

- psychrotolerantné mikroorganizmy – rastú dobre aj pri teplotách blízko 0 °C, ale najlepšie pri teplote nad 20 °C;
- psychrofilné mikroorganizmy – najlepšie rastú pri 15 °C a menej, pričom nerastú nad 20 °C.

Homologické chlad-aktívne enzýmy, mezozýmy, aj termozýmy sú si podobné:

- v aminokyselinovej sekvencii;
- ich terciárne štruktúry sú navzájom preložiteľné;
- v katalytickom mechanizme.

Aké sú proteín stabilizačné mechanizmy?

Boli identifikované štúdiom modelov mezofilných enzýmov (mezozýmov) a boli potvrdené aj najnovšími štrukturálnymi porovnaniami mezozýmov a termozýmov (ale aj chlad-aktívnych enzýmov).

Zvýšená stabilita termozýmov musí byť výsledkom špecifických rozdielov v ich aminokyselinových sekvenciách.

Klonované termozýmy možno úspešne exprimovať v mezofilných organizmoch.

Rekombinantné termozýmy si zachovávajú svoje natívne vlastnosti.

Hoci vonkajšie faktory (soli, polyamíny, glykozylácia) pomáhajú stabilizovať niektoré termozýmy, väčšina termozýmov zostáva vnútorne termostabilná.

Zvýšená aktivita enzýmov aktívnych pri nízkych teplotách (chlad-aktívnych enzýmov) práve pri nízkych teplotách musí byť tiež výsledkom špecifických rozdielov v ich aminokyselinových sekvenciách.

Nízke teploty – nízka kinetická energia reagujúcich molekúl – toto je kompenzované *flexibilnými* štruktúrami.

Flexibilita sa dosahuje kombináciou štruktúrnych črt:

- redukcia hydrofóbneho jadra;
- znížené iónové a elektrostatické interakcie;
- zvýšený náboj zvyškov na povrchu, ktorý zlepšuje interakcie s rozpúšťadlami;
- dopĺňujúce povrchové slučky;
- Pro→Gly substitúcie v povrchových slučkách;
- menej medzidoménových a medzipodjednotkových interakcií;
- znížený pomer Arg:Lys.

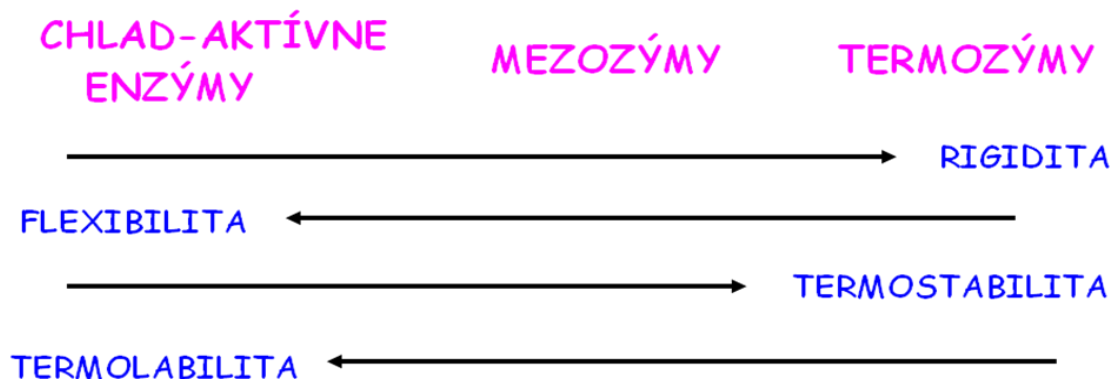
Toto má za následok, že aktívne miesto a príľahlé regióny zostávajú flexibilné.

Zvýšená konformačná flexibilita je ale doprevádzaná zvýšenou termolabilitou (t.j. zníženou termostabilitou).

Ako je to pri termozýmoch?

Termozýmy sú stabilizované rovnako ako meozýmy, t.j. stabilizačnými faktormi – rôznymi interakciami, ale aj štruktúrnymi faktormi.

Ich príspevky sú vo všeobecnosti malé, ale relevantné.



Neexistuje spoločný (všeobecný) determinant termostability!

Termodynamická stabilita je definovaná ako:

- ΔG_{stab} – voľná energia stabilizácie enzýmu;
- T_m – teplota, pri ktorej je proteín na 50% rozvinutý.

Kinetická stabilita závisí na energetickej bariére rozvinutia, t.j. aktivačnej energii rozvinutia (E_a) a často sa vyjadruje ako čas ($t_{1/2}$), pri ktorom je proces v polovici pri definovaných podmienkach.

Hypertermofilné proteíny:

- extracelulárne ($T_{\text{opt}} \sim 20\text{-}30\text{ }^{\circ}\text{C}$ vyššia ako teplota optimálna pre rast);
- intracelulárne (zvyčajne ich T_{opt} je porovnateľná s teplotou rastu).

12. Proteínové inžinierstvo a dizajn

Príklady

Zmena cyklodextrínglukanotransferázy na α -amylázu (zmena EC triedy)

PubMed: 11943149

FEBS 25836

FEBS Letters 514 (2002) 189–192

Mutations converting cyclodextrin glycosyltransferase from a transglycosylase into a starch hydrolase

Hans Leemhuis^a, Bauke W. Dijkstra^b, Lubbert Dijkhuizen^{a,*}

^aDepartment of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands

^bBIOSON Research Institute and Laboratory of Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received 2 January 2002; revised 23 January 2002; accepted 23 January 2002

First published online 12 February 2002

Edited by Hans Eklund

Abstract Cyclodextrin glycosyltransferase (CGTase) efficiently catalyzes transglycosylation of oligo-maltodextrins, although the enzyme also has a low hydrolytic activity. Its +2 substrate binding subsite, which contains the conserved Phe184 and Phe260 residues, has been shown to be important for this

CGTase is first of all a transglycosylase. The hydrolysis activity of CGTases is in general much lower than the disproportionation and cyclization activities. Only the CGTases from *Thermoanaerobacter* [6] and *Thermoanaerobacterium thermosulfurigenes* strain EM1 (*Tabium*) [7] have relatively high hy-

Zmena α -xylozidázy na α -glukozidázu (substrátová preferencia)

PuBmed: 16631751

FEBS Letters 580 (2006) 2707–2711

Structural elements to convert *Escherichia coli* α -xylosidase (YicI) into α -glucosidase

Masayuki Okuyama, Akira Kaneko, Haruhide Mori, Seiya Chiba, Atsuo Kimura*

Division of Applied Bioscience, Graduate School of Agriculture, Hokkaido University, Kita-9 Nishi-9, Kita-ku, Sapporo 060-8589, Japan

Received 27 January 2006; revised 7 April 2006; accepted 10 April 2006

Available online 19 April 2006

Edited by Gianni Cesareni

Abstract *Escherichia coli* YicI, a member of glycoside hydrolase family (GH) 31, is an α -xylosidase, although its amino-acid sequence displays approximately 30% identity with α -glucosidases. By comparing the amino-acid sequence of GH 31 enzymes and through structural comparison of the $(\beta/\alpha)_8$ barrels of GH 27 and GH 31 enzymes, the amino acids Phe277, Cys307, Phe308, Trp345, Lys414, and $\beta \rightarrow \alpha$ loop 1 of $(\beta/\alpha)_8$ barrel of YicI have been identified as elements that might be important for YicI substrate specificity. In attempt to convert YicI into an α -gluco-

lyze the non-reducing terminal α -glucoside unit of substrate to release α -glucose. There have been no previous studies of the key residues that determine the substrate specificity of GH 31 enzymes. The present study was performed to elucidate the structural elements involved in the control of substrate recognition of α -xylosidase and α -glucosidase.

The three-dimensional structure of a GH 31 enzyme has only been solved for YicI, the catalytic domain of which shows a modified $(\beta/\alpha)_8$ barrel fold [11]. Many carbohydrate hydrolases

Zmena substrátovej špecificity lipázy (dĺžka reťazca mastných kyselín)

PubMed: 12200542

Protein Engineering vol.15 no.7 pp.595–601, 2002

Blocking the tunnel: engineering of *Candida rugosa* lipase mutants with short chain length specificity

Jutta Schmitt¹, Stefania Brocca², Rolf D.Schmid¹ and Jürgen Pleiss^{1,3}

¹Institute of Technical Biology, University of Stuttgart, Allmandring 31, D-70569 Stuttgart, Germany and ²Department of Biotechnology and Biosciences, University of Milano-Bicocca, I-20126 Milano, Italy

³To whom correspondence should be addressed.
E-mail: juergen.pleiss@po.uni-stuttgart.de

The molecular basis of chain length specificity of *Candida rugosa* lipase 1 was investigated by molecular modeling and site-directed mutagenesis. The synthetic lip1 gene and the lipase mutants were expressed in *Pichia pastoris* and assayed for their chain length specificity in single substrate assays using triglycerides as well as in a competitive substrate assay using a randomized oil. Mutation of amino

1993). CRL has a broad chain length profile ranging from short chain esters to long, saturated and even polyunsaturated fatty acids (Osada *et al.*, 1990) with a distinct preference for fatty acids of chain lengths C4, C8, C10 and C12 (Janssen and Halling, 1994). Since CRL is an unspecific lipase towards a broad range of fatty acid chain lengths but has low activity towards long, polyunsaturated fatty acids, it has been used for the enrichment of these valuable food ingredients. CRL has been used for the enrichment of eicopentaenoic and docosa-hexaenoic acids from fish oil (McNeill *et al.*, 1996; Moore and McNeill, 1996) or of gamma-linolenic acid from borage oil (Shimada *et al.*, 1998; Schmitt-Rozieres *et al.*, 1999). To understand the structural basis of substrate specificity, the structure of complexes of CRL with substrate-analogous covalent inhibitors was studied by X-ray crystallography (Cveter

Termostabilizácia celulázy (zavedenie soľných mostíkov)

PubMed: 11522924

Protein Engineering vol.14 no.7 pp.501–504, 2001

Thermostabilization by replacement of specific residues with lysine in a *Bacillus* alkaline cellulase: building a structural model and implications of newly formed double intrahelical salt bridges

Tadahiro Ozawa¹, Yoshihiro Hakamada¹, Yuji Hatada¹, Tohru Kobayashi¹, Tsuyoshi Shirai² and Susumu Ito^{1,3}

¹Tochigi Research Laboratories, Kao Corporation, 2606 Akabane, Ichikai, Haga, Tochigi 321-3497 and ²Department of Biotechnology and Biomaterial Chemistry, Graduate School of Engineering, Nagoya University, Chikusa-ku, Nagoya 464-8603, Japan

³To whom correspondence should be addressed.
E-mail: itos@jamstec.go.jp

An alkaline, mesophilic *endo*-1,4- β -glucanase from alkaliphilic *Bacillus* sp. strain KSM-64 was significantly thermostabilized by replacement of both Asn179 and Asp194 with lysine by site-directed mutagenesis. Structural remodeling of the mutant enzyme newly generated by the double mutation suggested that Glu175 $\rightarrow$ Lys179 and Glu190 $\rightarrow$ Lys194 were the most plausible ion pairs, both of

increase in ionic interactions is often observed in thermophilic enzymes (Vogt *et al.*, 1997; Wallon *et al.*, 1997; Hashimoto *et al.*, 1999; Kumar *et al.*, 2000).

Recently, we found a thermostable, alkaline Egl (Egl-237) from the alkaliphilic *Bacillus* isolate KSM-S237 (Hakamada *et al.*, 1997) and sequenced the cloned gene of the enzyme (Hakamada *et al.*, 2000). The deduced amino acid sequence of Egl-237 showed high homology with those of Egl-K (Ozaki *et al.*, 1990) and Egl-64 (Sumitomo *et al.*, 1992). Based on the difference of a few amino acids, we constructed several chimeric genes from Egl-64 and Egl-237, and Lys179 and Lys194 residues in Egl-237, not conserved in Egl-64, were found to contribute to thermostability after site-directed mutagenesis. As a result, the double mutation Asn179 $\rightarrow$ Lys (NK)/Asp194 $\rightarrow$ Lys (DK) (NK/DK) significantly improved thermo-

Hyperthermostabilizácia α -amylázy (o viac ako 20 °C)

PubMed: 12736372

Protein Engineering vol.16 no.4 pp.287–293, 2003
DOI: 10.1093/proeng/gzg032

Hyperthermostabilization of *Bacillus licheniformis* α -amylase and modulation of its stability over a 50°C temperature range

Nathalie Declerck^{1,2}, Mischa Machius³, Philippe Joyet¹, Georg Wiegand⁴, Robert Huber⁴ and Claude Gaillardin¹

¹Génétique Moléculaire et Cellulaire, CNRS-URA1925, INRA-UMR216, F-78850 Thiverval-Grignon, France, ²University of Texas Southwestern Medical Center, Dallas, TX 75390, USA and ⁴Max-Planck-Institut für Biochemie, D-85152 Planegg-Martinsried, Germany

³To whom correspondence should be addressed. Present address: Centre de Biochimie Structurale, CNRS-5048, INSERM-554, 29 rue de Navacelles, F-34090 Montpellier, France.
E-mail: nathalie@cbs.cnrs.fr

Bacillus licheniformis α -amylase (BLA) is a highly thermostable starch-degrading enzyme that has been extensively studied in both academic and industrial laboratories. For

multiple amino acid substitutions (Declerck *et al.*, 1990, 1995, 2000; Joyet *et al.*, 1992). This mutational analysis, combined with structural studies (Declerck *et al.*, 1995, 1997; Machius *et al.*, 1995, 1998, 2003), allowed the probing of the contributions of various molecular features to BLA thermostability. In particular, a key role could be attributed to an unusual calcium/sodium binding site where amino acid substitutions drastically decreased the overall stability of the protein. In addition to thermosensitive mutations, we also identified several amino acid changes that render the enzyme even more thermostable than the wild-type.

In this paper, we report the cumulative effect of some of the most stabilizing and destabilizing mutations that we and others have identified in BLA. Comparison of the hyperther-

Zmena farby emitovaného svetla luciferázy (červená farba na zelenú)

PubMed: 18251715

BJ www.biochemj.org

Biochem. J. (2008) **412**, 27–33 (Printed in Great Britain) doi:10.1042/BJ20070733



27

Site-directed mutagenesis of firefly luciferase: implication of conserved residue(s) in bioluminescence emission spectra among firefly luciferases

Narges Kh. TAFRESHI*, Majid SADEGHIZADEH*¹, Rahman EMAMZADEH†, Bijan RANJBAR‡, Hossein NADERI-MANESH‡ and Saman HOSSEINKHANI†¹

*Department of Genetics, Faculty of Basic Sciences, Tarbiat Modares University, Tehran, Iran, 14115-175, †Department of Biochemistry, Faculty of Basic Sciences, Tarbiat Modares University, Tehran, Iran, 14115-175, and ‡Department of Biophysics, Faculty of Basic Sciences, Tarbiat Modares University, Tehran, Iran, 14115-175

The bioluminescence colours of firefly luciferases are determined by assay conditions and luciferase structure. Owing to red light having lower energy than green light and being less absorbed by biological tissues, red-emitting luciferases have been considered as useful reporters in imaging technology. A set of red-emitting mutants of *Lampyrus turkestanicus* (Iranian firefly) luciferase has been made by site-directed mutagenesis. Among different beetle luciferases, those from *Phrixothrix* (railroad worm) emit either green or red bioluminescence colours naturally. By substitution of three specific amino acids using

bioluminescence colour. Meanwhile, the luciferase reaction took place with relative retention of its basic kinetic properties such as K_m and relative activity. Structural comparison of the native and mutant luciferases using intrinsic fluorescence, far-UV CD spectra and homology modelling revealed a significant conformational change in mutant forms. A change in the colour of emitted light indicates the critical role of these conserved residues in bioluminescence colour determination among firefly luciferases. Relatively high specific activity and emission of red light might make these mutants suitable as reporters for the study of gene

Určenie úlohy aminokyselinového zvyšku v β -glukozidáze (W262)

PubMed: 16274659



Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Archives of Biochemistry and Biophysics 444 (2005) 66–75

ABB

www.elsevier.com/locate/yabbi

Trp-262 is a key residue for the hydrolytic and transglucosidic reactivity of the *Aspergillus niger* family 3 β -glucosidase: Substitution results in enzymes with mainly transglucosidic activity[☆]

Heather F. Seidle^a, Kyle McKenzie^a, Ira Marten^b, Oded Shoseyov^b, Reuben E. Huber^{a,*}

^a Division of Biochemistry, Faculty of Science, University of Calgary, 2500 University Dr. NW, Calgary, Alta., Canada T2N 1N4

^b The Institute of Plant Sciences and Genetics in Agriculture, The Faculty of Agriculture, The Hebrew University of Jerusalem, P.O. Box 12, Rehovot 76100, Israel

Received 23 August 2005, and in revised form 22 September 2005
Available online 24 October 2005

Určenie úlohy aminokyselinového zvyšku v 4- α -glukanotransferáze (W229)

PubMed: 17035108



Available online at www.sciencedirect.com

ScienceDirect

Biochimica et Biophysica Acta 1764 (2006) 1633–1638

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.elsevier.com/locate/bbapap

Contribution of W229 to the transglycosylation activity of 4- α -glukanotransferase from *Pyrococcus furiosus*

Shuang-Yan Tang^a, Sung-Jae Yang^a, Hyunju Cha^a, Eui-Jeon Woo^b,
Cheonseok Park^c, Kwan-Hwa Park^{a,*}

^a Center for Agricultural Biomaterials and Department of Food Science and Biotechnology, School of Agricultural Biotechnology, Seoul National University, Seoul 151-742, Korea

^b Systemic Proteomics Research Center, Korea Research Institute of Bioscience and Biotechnology, Taejeon 305-333, Korea

^c Department of Food Science and Biotechnology and Institute of Life Sciences and Resources, Kyunghee University, Yongin 449-701, Korea

Received 30 June 2006; received in revised form 22 August 2006; accepted 28 August 2006
Available online 1 September 2006

Zmena účinku polygalakturonázy (procesívne vs. neprocesívne pôsobenie)

PubMed: 11445590

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 2001 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 276, No. 36, Issue of September 7, pp. 33652–33656, 2001
Printed in U.S.A.

Changing a Single Amino Acid Residue Switches Processive and Non-processive Behavior of *Aspergillus niger* Endopolygalacturonase I and II*

Received for publication, June 21, 2001
Published, JBC Papers in Press, July 9, 2001, DOI 10.1074/jbc.M105770200

Sandrine Pagès[‡], Harry C. M. Kester, Jaap Visser[§], and Jacques A. E. Benen[¶]

From the Section Molecular Genetics of Industrial Microorganisms, Wageningen University, Dreyenlaan 2, 6703 HA Wageningen, The Netherlands

Processivity, also known as multiple attack on a single chain, is a feature commonly encountered only in enzymes in which the substrate binds in a tunnel. However, of the seven *Aspergillus niger* endopolygalacturonases, which have an open substrate binding cleft, four enzymes show processive behavior, whereas the other endopolygalacturonases are randomly acting enzymes. In a previous study (Benen, J.A.E., Kester, H.C.M., and Visser, J. (1999) *Eur. J. Biochem.* 259, 577–585) we proposed that the high affinity for the substrate of subsite

exo-enzymes that attack the chains from the extremities. In *Aspergillus niger*, *Sclerotinia sclerotiorum*, and *Botrytis cinerea* families of genes encoding endoPGs were discovered (3–8); but actually little is known about the precise role of the different endoPGs in these organisms. However, the characterization of the seven highly homologous endoPGs secreted by *A. niger* (PGI, PGII, PGA-E) suggested that individual enzymes harbor specific enzymic and physiological functions. One of the enzymic characteristics is the processive behavior observed for endoPGs A, C, and D (9, 11). Processivity first described by

Určenie príspevku molekúl vody k stabilite (molekuly H₂O vo vnútri)

PubMed: 12646687

Protein Engineering vol.16 no.1 pp.5–9, 2003
DOI: 10.1093/proeng/gzg001

Buried water molecules contribute to the conformational stability of a protein

Kazufumi Takano^{1,2}, Yuriko Yamagata³ and Katsuhide Yutani^{1,4}

¹Institute for Protein Research, Osaka University, Yamadaoka, Suita, Osaka 565-0871 and ²Graduate School of Pharmaceutical Sciences, Kumamoto University, Oe-honmachi, Kumamoto 862-0973, Japan

⁴Present address: Graduate School of Sciences, Kwansei Gakuin University, Gakuen, Sanda, Hyogo 669-1337, Japan

²To whom correspondence should be addressed. Present address: Graduate School of Engineering, Osaka University, Yamadaoka, Suita, Osaka 565-0871, Japan.
E-mail: ktakano@mls.eng.osaka-u.ac.jp

This study sought to attain a better understanding of the contribution of buried water molecules to protein stability. The 3SS human lysozyme lacks one disulfide bond between

the role of buried water molecules in protein stability (Wade *et al.*, 1991; Madan and Lee, 1994; Fischer and Verma, 1999). However, there have been only a few experimental studies because it is difficult to estimate experimentally. We previously succeeded in estimating the role of buried water molecules by examining Ile to Ala/Gly mutants of the human lysozyme (Takano *et al.*, 1997a). The results demonstrated that mutants with additional water molecules in the created cavity are destabilized less than mutants of the human lysozyme with empty cavities, indicating that the buried water molecules stabilize the protein structure (Takano *et al.*, 1997a). It is important to verify empirically whether this conclusion is common to globular proteins. In the present study, we examined one Ile to Ala mutant of the 3SS human lysozyme and compared the stability between this mutant with Ile and Ala

Zvýšenie termostability a katalytickej účinnosti pululanázy (bodový mutant)

PubMed: 23624477



Improving the Thermostability and Catalytic Efficiency of *Bacillus deramificans* Pullulanase by Site-Directed Mutagenesis

Xuguo Duan,^{a,b} Jian Chen,^{a,b} Jing Wu^{a,b}

State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi, China^a; School of Biotechnology and Key Laboratory of Industrial Biotechnology Ministry of Education, Jiangnan University, Wuxi, China^b

Pullulanase (EC 3.2.1.41) is a well-known starch-debranching enzyme. Its instability and low catalytic efficiency are the major factors preventing its widespread application. To address these issues, Asp437 and Asp503 of the pullulanase from *Bacillus deramificans* were selected in this study as targets for site-directed mutagenesis based on a structure-guided consensus approach. Four mutants (carrying the mutations D503F, D437H, D503Y, and D437H/D503Y) were generated and characterized in detail. The results showed that the D503F, D437H, and D503Y mutants had an optimum temperature of 55°C and a pH optimum of 4.5, similar to that of the wild-type enzyme. However, the half-lives of the mutants at 60°C were twice as long as that of the wild-type enzyme. In addition, the D437H/D503Y double mutant displayed a larger shift in thermostability, with an optimal temperature of 60°C and a half-life at 60°C of more than 4.3-fold that of the wild-type enzyme. Kinetic studies showed that the K_m values for the D503F, D437H, D503Y, and D437H/D503Y mutants decreased by 7.1%, 11.4%, 41.4%, and 45.7% and the K_{cat}/K_m values increased by 10%, 20%, 140%, and 100%, respectively, compared to those of the wild-type enzyme. Mechanisms that could account for these enhancements were explored. Moreover, in conjunction with the enzyme glucoamylase, the D503Y and D437H/D503Y

4072 aem.asm.org

Applied and Environmental Microbiology p. 4072–4077

July 2013 Volume 79 Number 13

Schopnosť viazať natívny škrob α -amylázou (mimo katalytickej domény)

PubMed: 19052787

J Ind Microbiol Biotechnol (2009) 36:341–346
DOI 10.1007/s10295-008-0502-y

ORIGINAL PAPER

A single residue mutation abolishes attachment of the CBM26 starch-binding domain from *Lactobacillus amylovorus* α -amylase

Romina Rodríguez-Sanoja · N. Oviedo · L. Escalante ·
B. Ruiz · S. Sánchez

Received: 14 August 2008 / Accepted: 6 November 2008 / Published online: 4 December 2008
© Society for Industrial Microbiology 2008

Abstract Starch is degraded by amylases that frequently have a modular structure composed of a catalytic domain

Introduction

13. Praktické úlohy a cvičenia

13.1. Zobrazovanie terciárnej štruktúry proteínov

Amylolytické enzýmy

Z proteínovej databanky PDB (<http://www.rcsb.org/pdb/>) získajte koordináty nasledujúcich enzýmov a na ich molekulách v programe WebLabViewerLite precvičujte zobrazovanie štruktúr proteínov:

- alfa-amyláza z *Aspergillus oryzae* - PDB kód: 7TAA (7TAA.pdb);
- cyklodextrínglukanotransferáza z *Bacillus circulans* - PDB kód: 1CDG (1CDG.pdb);
- neopululanáza z *Bacillus stearothermophilus* - PDB kód: 1J0I (1J0I.pdb).

Na ploche priečinok: PDB - tam ukladať všetky súbory; hotové obrázky uložte ako „*.gif“ a „*.msv“.

Alfa-amyláza

Celú štruktúru proteínu zobrazte ako „solid ribbon“ (pri súčasnom vypnutí atómov). Jednotlivé domény potom zvýraznite podľa farebného kódu uvedeného nižšie. Katalytické aminokyselinové zvyšky zobrazte ako „stick“ a farbte ako „element“. Molekulu akarbózy zobrazte ako „scaled ball and stick“ a farbte ako „element“. Atóm vápnika zobrazte ako „CPK“ a farbte na modro.

Jednotlivé domény:

doména A: 1-121 a 177-380 - tyrkysová

doména B: 122-176 - žltá

doména C: 381-476 - oranžová

Katalytické aminokyselinové zvyšky: Asp206, Glu230 a Asp297.

Molekula akarbózy: ABC479.

Atómy vápnika: Ca480.

Cyklodextrínglukanotransferáza

Celú štruktúru proteínu zobrazte ako „solid ribbon“ (pri súčasnom vypnutí atómov). Jednotlivé domény potom zvýraznite podľa farebného kódu uvedeného nižšie. Katalytické aminokyselinové zvyšky zobrazte ako „stick“ a farbte ako „element“. Molekuly maltózy zobrazte ako „scaled ball and stick“ a farbte ako „element“. Atómy vápnika zobrazte ako „CPK“ a farbte na modro.

Jednotlivé domény:

doména A: 1-138 a 203-406 - tyrkysová

doména B: 139-202 - žltá

doména C: 407-495 - oranžová

doména D: 496-582 - zelená

doména E (SBD): 583-686 - červená

Katalytické aminokyselinové zvyšky: Asp229, Glu257 a Asp328.

Molekuly maltózy: Mal688, Mal689 a Mal690.

Atómy vápnika: Ca691 a Ca692.

Neopululanáza

Celú štruktúru proteínu zobrazte ako „solid ribbon“ (pri súčasnom vypnutí atómov). Jednotlivé domény potom zvýraznite podľa farebného kódu uvedeného nižšie. Katalytické aminokyselinové zvyšky zobrazte ako „stick“ a farbte ako „element“. Molekulu panózy zobrazte ako „scaled ball and stick“ a farbte ako „element“. Z originálneho PDB súboru (obsahuje dimér) treba najskôr odstrániť: molekulu B neopululanázy aj s panózou (GLC604-GLC605-GLC606) a molekuly vody; 1JOI_A.pdb.

Jednotlivé domény:

doména A: 124-246 a 300-506 - tyrkysová

doména B: 247-299 - žltá

doména C: 507-588 - oranžová

doména N: 1-123 - sivá

Katalytické aminokyselinové zvyšky: Asp328, Glu357 a Asp424.

Molekula panózy: GLC601-GLC602-GLC603.

DNA-binding protein HU from Bacillus stearothermophilus

- (1) Utvorte si priečinok PDB. Z databázy PDB si stiahnite súbor DNA-viažuceho proteínu HU - 1HUU.pdb a uložte v priečinku PDB. Zo súboru 1HUU.pdb si pripravte súbor, v ktorom bude len podjednotka (molekula) A proteínu HU a uložte ako 1HUU_A.pdb.
- (2) V programe WebLabViewerLite zobrazte súbor 1HUU_A.pdb ako „Protein“ (pri vypnutí atómov) ako (i) „Solid Ribon“ a (ii) „Schematic“. Uložte ako súbory programu WebLabViewerLite: 1HUU_A_sr.msv a 1HUU_A_sch.msv.
- (3) V zobrazení „Solid Ribon“ zobrazte súbor 1HUU.pdb (t.j. iba ako proteín, t.j. bez atómov) a každú z troch molekúl proteínu zvýraznite inou farbou a súbor uložte ako 1HUU_sr_colors.msv.
- (4) Otvorte súbor 1HUU_A_sr.msv a pre aminokyselinové zvyšky Met69, Glu70, Ile71 a Pro72 zobrazte aj ich atómy s ich van der Waalsovými polomermi („CPK“) a súbor uložte ako 1HUU_A_sr_cpk.msv.
- (5) Zistite klasifikáciu HU proteínu v rámci CATH (1huuA00) a SCOP (d1huua_) - informácia sa nachádza v PDB pre HU proteín (1HUU) a klasifikáciu si pozrite v systémoch CATH a SCOP.

13.2. Evolúcia enzýmov metabolickej dráhy

Štyri enzýmy z glykolýzy

- triózafosfátizomeráza (TIM)
- glyceraldehyd-3-fosfátdehydrogenáza (GAPD)
- fosfoglycerátkináza (PGK)
- enoláza (ENOL)

Najskôr postupne a potom spolu všetky 4 enzýmy naraz.

Zadanie:

- (1) Vytvorte si priečink „Glycolysis“ a v rámci neho 4 priečinky: TIM, GAPD, PGK a ENOL.
- (2) Vstupné súbory si vytvorte pomocou údajov v Tabuľke a uložte do príslušných priečinkov TIM, GAPD, PGK a ENOL ako textové súbory („*.txt“; t.j. napr. TIM.txt) vhodné pre program CLUSTAL-Omega.
- (3) Na EBI serveri zrovnajte postupne sekvencie pre TIM, GAPD, PGK a ENOL v programe CLUSTAL-Omega - získajte vždy dva súbory („*_aln.txt“ a „*_fas.txt“; t.j. napr. TIM_aln.txt a TIM_fas.txt). Vypočítajte hodnoty konsenzuálnej dĺžky (CL), sekvenčnej podobnosti (SS) a sekvenčnej identity (SI).
- (4) Súbory „*_aln.txt“ otvorte vždy v programe MS-Word a v zrovnaných sekvenciách zvýraznite žltým podfarbením aminokyselinové zvyšky aktívneho miesta:
 - (i) TIM - Asn10, Lys12, His95 a Glu165 v TIM zo *Saccharomyces cerevisiae*;
 - (ii) Cys151 a His178 v GAPD z *Bacillus anthracis* (pre baktérie a eukaryoty) a Cys139 a His219 v GAPD zo *Sulfolobus solfataricus* (pre archaebaktérie);
 - (iii) Lys218, Asn335 a Asp373 v PGK zo *Saccharomyces cerevisiae* (pre baktérie a eukaryoty) a Glu342 pre všetky sekvencie;
 - (iv) Glu169, Asp247, Glu296, Asp321, Asp322 a Arg375 v ENOL zo *Saccharomyces cerevisiae*a súbory uložte vždy ako dokument „*_aln.doc“; t.j. napr. TIM_aln.doc.
- (5) Na EBI serveri v rámci programu CLUSTAL-W2 Phylogeny vypočítajte vždy (t.j. pre každý enzým) dva evolučné stromy - s uvažovaním medzier v sekvenciách („*_off.ph“) a s ich ignorovaním („*_on.ph“) - t.j. napr. TIM_off.ph a TIM_on.ph.
- (6) Na základe vstupných súborov pre TIM, GAPD, PGK a ENOL vytvorte spoločný vstupný súbor pre všetky štyri študované enzýmy glykolýzy („Glycolysis.txt“; sekvencie ukladať pre každý organizmus vždy v tom istom poradí - napr. 1. TIM, 2. GAPD, 3. PGK a 4. ENOL), uložte ho do priečinku „Glycolysis“ a na EBI serveri vykonajte zrovnanie sekvencií

programom CLUSTAL-Omega, t.j. získajte súbory „Glycolysis_aln“.txt a „Glycolysis_fas.txt“.

- (7) Na EBI serveri v rámci programu CLUSTAL-W2 Phylogeny vypočítajte pre zrované štyri enzýmy glykolýzy dva evolučné stromy - s uvažovaním medzier v sekvenciách („Glycolysis_off.ph“) a s ich ignorovaním („Glycolysis_on.ph“). Vypočítajte CL, SS a SI a tieto hodnoty porovnajte s údajmi pre jednotlivé enzýmy.
- (8) V rámci diskusie písomne porovnajte všetky získané evolučné stromy (zobrazené v programe TreeView).

Záver:

TIM - triózafozfátizomomeraza

- | | |
|-----------------------|------------------------|
| (a) najdlhší proteín: | (b) najkratší proteín: |
| (c) CL = | (d) SI = |
| | (e) SS = |

GAPD - glyceraldehyd-3-fosfátdehydrogenáza

- | | |
|-----------------------|------------------------|
| (a) najdlhší proteín: | (b) najkratší proteín: |
| (c) CL = | (d) SI = |
| | (e) SS = |

PGK - fosfoglycerátkináza

- | | |
|-----------------------|------------------------|
| (a) najdlhší proteín: | (b) najkratší proteín: |
| (c) CL = | (d) SI = |
| | (e) SS = |

ENOL - enoláza

- | | |
|-----------------------|------------------------|
| (a) najdlhší proteín: | (b) najkratší proteín: |
| (c) CL = | (d) SI = |
| | (e) SS = |

Glycolysis - glykolýza (4 enzýmy)

- | | |
|-------------------------|--------------------------|
| (a) najdlhšie proteíny: | (b) najkratšie proteíny: |
| (c) CL = | (d) SI = |
| | (e) SS = |

Tabuľka enzýmov z glykolýzy.

Č.	Zdroj	Skratka	TIM		GAPD		PGK		ENOL	
			UniProt	Dĺžka	UniProt	Dĺžka	UniProt	Dĺžka	UniProt	Dĺžka
<u>Bacteria</u>										
1	<i>Agrobacterium tumefaciens</i>	Agrtu	Q8UEY3	256	A9CFL0	336	Q8U9I9	400	Q8UFH1	424
2	<i>Anabaena sp. PCC 7120</i>	Anasp	Q8YP17	241	P80506	343	Q8YPR1	400	Q8YRB0	429
3	<i>Bacillus anthracis</i>	Bacan	Q81X76	251	Q81L07	342	Q81X75	394	Q81X78	431
4	<i>Bifidobacterium longum</i>	Biflo	Q8G6D5	267	Q8G4L6	352	Q8G6D6	401	Q8G5I9	432
5	<i>Borrelia burgdorferi</i>	Borbu	Q59182	253	P46795	335	Q59181	393	O51312	433
6	<i>Chlamydia pneumoniae</i>	Chlpn	Q9Z6J6	254	Q9Z7T0	335	Q9Z7M5	402	Q9Z7A6	428
7	<i>Deinococcus radiodurans</i>	Deira	Q9RUP5	244	Q9RUP1	330	Q9RUP2	411	Q9RR60	422
8	<i>Escherichia coli</i>	Escco	P0A858	255	P0A9B2	331	P0A799	387	P0A6Q1	432
9	<i>Haemophilus influenzae</i>	Haein	P43727	263	P44304	339	P43726	386	P43806	436
10	<i>Mycobacterium tuberculosis</i>	Myctu	P66940	261	P64178	339	P65700	412	P96377	429
11	<i>Mycoplasma genitalium</i>	Mypge	P47670	244	P47543	337	P47542	416	P47647	458
12	<i>Neisseria meningitidis</i>	Neime	Q9JW31	251	A1IP92	334	Q9JWS8	392	Q9JU46	428
13	<i>Salmonella typhimurium</i>	Salty	Q8ZKP7	255	P0A1P0	331	P65702	387	P64076	432
14	<i>Staphylococcus aureus</i>	Staa	Q5HHP3	253	P0A036	336	P68819	396	P64078	434
15	<i>Streptococcus mutans</i>	Stcmu	P72484	252	Q8DVV3	337	Q8DVV2	398	Q8DTS9	432
16	<i>Streptomyces coelicolor</i>	Strco	Q9Z520	258	Q93J08	332	Q9Z519	403	Q9F2Q3	426
17	<i>Synechocystis sp. PCC 6803</i>	Synsp	Q59994	242	P49433	339	P74421	401	P77972	432
18	<i>Thermotoga maritima</i>	Thtma	P36204	255 (400-654)	P17721	333	P36204	399 (1-399)	P42848	429
19	<i>Vibrio cholerae</i>	Vibch	Q9KNR1	257	Q9KQJ8	331	A5F9G2	387	Q9KPC5	433
20	<i>Yersinia pestis</i>	Yerpe	Q8ZJK9	255	Q7CIH3	334	Q8ZHH3	387	Q8ZBN2	431
Archaea										
21	<i>Aeropyrum pernix</i>	Aerpe	Q9YBR1	223	Q9YFS9	343	Q9YFS7	415	Q9Y927	432
22	<i>Methanosarcina mazei</i>	Metma	Q8PXE2	222	Q8PTD3	335	Q8PZK7	416	Q8PT81	428
23	<i>Pyrobaculum aerophilum</i>	Pybae	Q8ZX28	227	Q8ZWK7	344	Q8ZWK6	408	Q8ZYE7	419
24	<i>Pyrococcus furiosus</i>	Pycfu	P62002	228	P61879	334	P61883	410	Q8U477	430
25	<i>Sulfolobus solfataricus</i>	Sulso	Q97VM8	227	P39460	340	P50317	408	Q97ZJ3	419
Eucarya										
26	<i>Arabidopsis thaliana</i>	Arath	Q9SKP6	255 (61-315)	P25856	336 (61-396)	P50318	405 (74-478)	P25696	444
27	<i>Caenorhabditis elegans</i>	Caeel	Q10657	247	P17331	341	P91427	417	Q27527	434
28	<i>Drosophila melanogaster</i>	Drome	P29613	247	P07486	332	Q01604	415	P15007	433 (68-500)
29	<i>Homo sapiens</i>	Homsa	P60174	249	P04406	335	P00558	417	P06733	434
30	<i>Saccharomyces cerevisiae</i>	Sacce	P00942	248	P00359	332	P00560	416	P00924	437

13.3. Tréning s molekulárno-biologickými databázami

Na základe prístupových čísiel do databázy UniProt, doplňte chýbajúce údaje pre 10 enzýmov a proteínov z tabuľky:

- prístupové číslo sekvencie v databáze GenBank a jej prekladu v GenPept;
- EC číslo - enzýmové číslo z enzýmovej nomenklatúry (pokiaľ nie je uvedené, pomôcť si informáciou o danom enzýme priamo z tabuľky);
- záznam v databáze PubMed - relevantné literárne údaje;
- dĺžku aminokyselinovej sekvencie proteínového reťazca;
- pozíciu domény CBM20 - pokiaľ je to uvedené.

Tabuľka.

Č.	Zdroj	Enzým	Skratka	UniProt	GenBank	GenPept	EC	PubMed	Dĺžka	CBM20
1	<i>Anaerobranca gottschalkii</i>	CGTáza	Anago_CGT	Q5ZEQ7						
2	<i>Bacillus agaradhaerens</i>	CGTáza	Bacag_CGT	Q7X3T0						
3	<i>Aspergillus awamori</i>	Alfa-amyláza	Aspaw_AMY	Q76L96						
4	<i>Bacillus</i> sp. TS-23	Alfa-amyláza	Bacsp_AMY	Q59222						
5	<i>Streptomyces griseus</i>	Alfa-amyláza	Strgr_AMY	P30270						
6	<i>Bacillus cereus</i>	Beta-amyláza	Bacce_BMY	P36924						
7	<i>Aspergillus niger</i>	Glukoamyláza G1	Aspni_GMY	P69328						
8	<i>Athelia rolfsii</i>	Glukoamyláza	Athro_GMY	Q12596						
9	<i>Homo sapiens</i>	Laforín - izoformy 1 a 2	Homsa_LAF	Q6IS15						
10	<i>Homo sapiens</i>	Genetonín-1	Homsa_GEN	Q95210						

13.4. Bioinformatická analýza škrob-viažucich domén

Teoretická časť

Škrob-viažuca doména (starch-binding domain; SBD) typických mikrobiálnych amylolytických enzýmov (α -amylázy - rodina GH13, β -amylázy - GH14, glukoamylázy - GH15) je klasifikovaná ako tzv. modul z rodiny CBM20, t.j. carbohydrate-binding module family 20. Táto doména umožňuje enzýmu viazať a degradovať surový (tepelne neupravený) škrob. Modul sa môže nachádzať aj v rôznych blízko, ale aj vzdialene príbuzných rastlinných a živočíšnych enzýmoch a proteínoch, ako napr. fosfoglukán vodná dikináza-3, laforín, genetónin-1, apod. V amylolytických enzýmoch, ako sú pululanáza a glukán-odvetvujúce enzýmy, sa nachádza sekvenčne príbuzný SBD modul, klasifikovaný do rodiny CBM48. Tento typ SBD je tiež prítomný v rôznych rastlinných a živočíšnych enzýmoch a proteínoch, ako napr. proteín nadbytku škrobu-4 (SEX-4) alebo β -podjednotka AMP-aktivovanej proteínkinázy, apod.

Zadanie

- (1) Všetkých 45 sekvencií SBD (skratky podľa tabuľky - 30× CBM20 a 15× CBM48) zhromaždíte z databázy UniProt do vstupného súboru (SBD.txt v priečinku SBD) vhodného pre program CLUSTAL-Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).
- (2) Na EBI serveri zrovnajte sekvencie SBD v programe CLUSTAL-Omega; získajte dva súbory: SBD_aln.txt a SBD_fas.txt (a uložte ich do priečinku SBD).
- (3) V rámci programu CLUSTAL-W2 Phylogeny (http://www.ebi.ac.uk/Tools/phylogeny/clustalw2_phylogeny/) vypočítajte dva evolučné stromy pre SBD - s uvažovaním medzier v sekvenciách (SBD_off.ph) a s ich ignorovaním (SBD_on.ph).
- (4) Stromy zobrazte v programe TreeView a diskutujte; pre zrovnané SBD sekvencie zistite konsenzuálnu dĺžku (CL), sekvenčnú identitu (SI) a podobnosť (SS).
- (5) Podľa informácií o dvoch škrob-viažucich miestach v sekvenciách SBD z rodín CBM20 a CBM48 - CBM20 (č. 9: 20-GH15-Aspni-GMY): W567 (1), K602 (1) a W614 (1) a W587 (2); CBM48 (č. 38: 48-Ratno-AMPK1): W100 (1), K126 (1) a W133 (1) a F112 (2) - upravte zrovnanie (SBD_aln.doc zo súboru SBD_aln.txt) v programe MS-Word a uložte ako SBD_aln_2.doc. Farebne zvýraznite všetky 4 aminokyselinové zvyšky zapojené do väzby sacharidov (pokiaľ sú prítomné - miesto 1 - žltá, miesto 2 - zelená).
- (6) Z upraveného zrovnania SBD_aln_2.doc vyrobte manuálne súbor vo formáte FASTA/Pearson (SBD_fas_2.txt) a na jeho základe znovu vypočítajte oba evolučné stromy (SBD_off_2.ph a SBD_on_2.ph), zobrazte a porovnajte ich s pôvodnými stromami. Skontrolujte, či a ako sa zmenili hodnoty CL, SI a SS.

Tabuľka. Reprezentatívne škrob-viažuce domény z rodín CBM20 a CBM48.

Číslo	Enzým/proteín ^a	EC	Zdroj	Skratka	UniProt	Dĺžka	Pozícia
<i>CBM20:</i>							
1.	GH13 α-amyláza	3.2.1.1	<i>Streptomyces griseus</i>	20-GH13-Stmgr-AMY	P30270	566	470-566
2.	GH13 α-amyláza	3.2.1.1	<i>Aspergillus kawachii</i>	20-GH13-Aspka-AMY	O13296	640	538-640
3.	GH13 cyklodextrín glukano-transferáza	2.4.1.19	<i>Bacillus circulans</i> strain 251	20-GH13-Bacci-CGT	P43379	713	608-713
4.	GH13 cyklodextrín glukano-transferáza	2.4.1.19	<i>Thermococcus</i> sp. B1001	20-GH13-Thcsp-CGT	Q9UWN2	739	629-739
5.	GH13 cyklodextrín glukano-transferáza	2.4.1.19	<i>Nostoc</i> sp. PCC 9229	20-GH13-Nosp-CGT	Q8RMG0	642	539-642
6.	GH13 izocyklomaltooligosacharid glukano-transferáza	2.4.1.-	<i>Bacillus circulans</i>	20-GH13-Bacci-ICGT	A0P8W9	995	893-995
7.	GH14 β-amyláza	3.2.1.2	<i>Bacillus cereus</i>	20-GH14-Bacce-BMY	Q9Z4N9	546	444-546
8.	GH14 β-amyláza	3.2.1.2	<i>Thermoanaerobacterium thermosulfurogenes</i>	20-GH14-Thbth-BMY	P19584	551	448-551
9.	GH15 glukoamyláza	3.2.1.3	<i>Aspergillus niger</i>	20-GH15-Aspni-GMY	P69328	640	533-640
10.	GH31 6-α-glukozyltransferáza	2.4.1.-	<i>Arthrobacter globiformis</i>	20-GH31-Arbgl-6AGT	Q6BD65	965	864-965
11.	GH57 amylopululanáza (kópia 1)	3.2.1.1/41	<i>Kosmotoga olearia</i>	20-GH57-Kotol-APU-1	C5CEB0	1354	37-138
12.	GH57 amylopululanáza (kópia 2)	3.2.1.1/41	<i>Kosmotoga olearia</i>	20-GH57-Kotol-APU-2	C5CEB0	1354	160-259
13.	GH57 amylopululanáza (kópia 3)	3.2.1.1/41	<i>Kosmotoga olearia</i>	20-GH57-Kotol-APU-3	C5CEB0	1354	272-373
14.	GH77 4-α-glukano-transferáza DPE2 (kópia 1)	2.4.1.25	<i>Bacteroides thetaiotaomicron</i>	20-GH77-Bctth-4AGT-1	Q8A5U2	893	3-92
15.	GH77 4-α-glukano-transferáza DPE2 (kópia 2)	2.4.1.25	<i>Bacteroides thetaiotaomicron</i>	20-GH77-Bctth-4AGT-2	Q8A5U2	893	127-224
16.	GH77 4-α-glukano-transferáza DPE2 (kópia 1)	2.4.1.25	<i>Chlamydomonas reinhardtii</i>	20-GH77-Chlre-4AGT-1	A8JEI0	941	8-110
17.	GH77 4-α-glukano-transferáza DPE2 (kópia 2)	2.4.1.25	<i>Chlamydomonas reinhardtii</i>	20-GH77-Chlre-4AGT-2	A8JEI0	941	160-254
18.	GH77 4-α-glukano-transferáza DPE2 (kópia 1)	2.4.1.25	<i>Solanum tuberosum</i>	20-GH77-Soltu-4AGT-1	Q6R608	948	8-103
19.	GH77 4-α-glukano-transferáza DPE2 (kópia 2)	2.4.1.25	<i>Solanum tuberosum</i>	20-GH77-Soltu-4AGT-2	Q6R608	948	152-244
20.	Fosfoglukán, vodná dikináza 3	2.7.9.5	<i>Chlamydomonas reinhardtii</i>	20-Chlre-PGW3	A8J6C3	978	20-123
21.	Fosfoglukán, vodná dikináza 3	2.7.9.5	<i>Arabidopsis thaliana</i>	20-Arath-PGW3	Q6ZY51	1196	66-166
22.	Genetonin-1		<i>Tetraodon nigroviridis</i>	20-Teoni-GEN1	Q4TBR2	174	73-174
23.	Genetonin-1		<i>Homo sapiens</i>	20-Homsa-GEN1	O95210	358	258-358
24.	Laforín	3.1.3.48/3.1.3.16	<i>Nematostella vectensis</i>	20-Nemve-LAF	A7SVW9	324	1-118
25.	Laforín	3.1.3.48/3.1.3.16	<i>Tetraodon nigroviridis</i>	20-Teoni-LAF	Q4S6Z3	312	1-106
26.	Laforín	3.1.3.48/3.1.3.16	<i>Homo sapiens</i>	20-Homsa-LAF	Q6IS15	331	1-124
27.	Glycerofosfodiester fosfodiesteráza-5	3.1.-.-	<i>Branchiostoma floridae</i>	20-Brsfl-GPDP5	C3Y330	680	1-107
28.	Glycerofosfodiester fosfodiesteráza-5	3.1.-.-	<i>Danio rerio</i>	20-Danre-GPDP5	Q1LVM1	666	1-113
29.	Glycerofosfodiester fosfodiesteráza-5	3.1.-.-	<i>Homo sapiens</i>	20-Homsa-GPDP5	Q9NPB8	672	1-115
30.	CE1 sacharidová esteráza		<i>Thermococcus kodakarensis</i>	20-CE1-Thcko	Q5JF12	449	88-189

Tabuľka. Reprezentatívne škrob-viažuce domény z rodín CBM20 a CBM48 (pokračovanie).

Číslo	Enzým/proteín ^a	EC	Zdroj	Skratka	UniProt	Dĺžka	Pozícia
<u>CBM48:</u>							
31.	Proteín nadbytku škrobu-4 (SEX-4)	3.1.3.48	<i>Chlamydomonas reinhardtii</i>	48-Chlre-SEX4	A8J2N4	428	276-360
32.	Proteín nadbytku škrobu-4 (SEX-4)	3.1.3.48	<i>Arabidopsis thaliana</i>	48-Arath-SEX4	Q9FEB5	379	249-335
33.	Proteín nadbytku škrobu-4 (SEX-4)	3.1.3.48	<i>Solanum tuberosum</i>	48-Soltu-SEX4	Q2VCK3	370	239-325
34.	AMP-aktivovaná proteinkináza β-podjednotka		<i>Brachyspira hyodysenteriae</i>	48-Brshy-AMPK	C0QYJ7	249	51-138
35.	AMP-aktivovaná proteinkináza β-podjednotka		<i>Chlamydomonas reinhardtii</i>	48-Chlre-AMPK	A8I0Q1	271	66-151
36.	AMP-aktivovaná proteinkináza β-podjednotka		<i>Nematostella vectensis</i>	48-Nemve-AMPK	A7SRX9	274	61-146
37.	AMP-aktivovaná proteinkináza β1-podjednotka		<i>Danio rerio</i>	48-Danre-AMPK1	Q6DHM2	268	68-151
38.	AMP-aktivovaná proteinkináza β1-podjednotka		<i>Rattus norvegicus</i>	48-Ratno-AMPK1	P80386	270	71-153
39.	AMP-aktivovaná proteinkináza β2-podjednotka		<i>Danio rerio</i>	48-Danre-AMPK2	B3DLH3	269	68-151
40.	AMP-aktivovaná proteinkináza β2-podjednotka		<i>Rattus norvegicus</i>	48-Ratno-AMPK2	Q9QZH4	271	69-152
41.	GH13 izoamyláza	3.2.1.68	<i>Pseudomonas amyloclavata</i>	48-GH13-Pseam-IAM	P10342	776	27-170
42.	GH13 glykogén odvetvujúci enzým	2.4.1.25/3.2.1.33	<i>Sulfolobus solfataricus</i>	48-GH13-Sulso-GDE	P95868	718	19-110
43.	GH13 glykogén vetviaci enzým	2.4.1.18	<i>Escherichia coli</i>	48-GH13-Escco-GBE	P07762	728	117-223
44.	GH13 pululanáza	3.2.1.41	<i>Klebsiella aerogenes</i>	48-GH13-Kleae-PUL	P07811	1096	310-414
45.	GH13 maltooligozyltrehalóza trehalohydroláza	3.2.1.141	<i>Sulfolobus solfataricus</i>	48-GH13-Sulso-MOTH	Q55088	559	1-88

13.5. Porovnanie terciárnych štruktúr škrob-viažucich domén

Škrob-, resp. glykogén-viažuce domény z rodín CBM20 a CBM48

(1) Podľa dostupných publikovaných informácií o škrob-, resp. glykogén-viažucich doménach (SBD, resp. GBD) z rodín CBM20 (publikácia Sorimachi a kol 1997; PubMed ID: 9195884 pre CBM20) a CBM48 (publikácia Polekhina a kol 2005; PubMed ID: 16216577 pre CBM48) sa pokúste identifikovať aminokyselinové zvyšky zapojené do 1. a prípadne aj 2. škrob/glykogén-viažuceho miesta v sekvenciách SBD/GBD z rodín CBM20 a CBM48.

(2) Identifikované funkčne dôležité väzbové zvyšky si zapíšte:

CBM20 - väzbové miesto 1: W543, K578 a W590

CBM20 - väzbové miesto 2: W563

CBM48 - väzbové miesto 1: W100, K126 a W133

CBM48 - väzbové miesto 2: F112

(3) Z proteínovej databanky PDB (<http://www.rcsb.org/pdb/>) získajte koordináty SBD z rodiny CBM20 a GBD z rodiny CBM48 a pripravte ich zobrazenia podľa zadania. Obrázky uložte ako 1AC0.msv a 1AC0.gif, resp. 1Z0M.msv a 1Z0M.gif.

- škrob-viažuca doména glukoamylázy z *Aspergillus niger* (rodina CBM20)
 - PDB kód: 1AC0 (1AC0.pdb). Štruktúru celej domény CBM20 zobrazte ako „solid ribbon“ (pri súčasnom vypnutí atómov) a farbite šedo. Väzbové zvyšky zobrazte ako „stick“ a farbite modro. Molekuly β -cyklodextrínu (GLC1-BGC2-...-GLC7 a GLC617B-GLC623B) zobrazte ako „stick“ a farbite tyrkysovou. Pri zobrazovaní odstraňte (vypnite) atómy vodíka.
- glykogén-viažuca doména β -podjednotky AMP-aktivovanej proteínovej kinázy z potkana (rodina CBM48) - PDB kód: 1Z0M (1Z0M.pdb). Pracujte iba s molekulou A a jej príslušným ligandom (β -cyklodextrínom). Štruktúru celej domény CBM48 zobrazte ako „solid ribbon“ (pri súčasnom vypnutí atómov) a farbite oranžovo. Väzbové zvyšky zobrazte ako „stick“ a farbite červeno. Molekulu β -cyklodextrínu (BCD201) zobrazte ako „stick“ a farbite fialovo. Pri zobrazovaní odstraňte (vypnite) molekuly vody.

13.6. Sledovanie evolúcie jednotlivých domén modulárneho proteínu

Teoretická časť

Enzým α -amyláza (EC 3.2.1.1) patrí medzi hydrolázy. Katalyzuje hydrolytické štiepenie α -1,4-glukozidových väzieb (napr. v škrobe a príbuzných poly- a oligo-sacharidoch). α -Amyláza je multidoménný enzým (pozri schému), ktorý sa skladá v zásade z troch domén (viac-menej samostatných štruktúrno-funkčných častí): (i) doména A - katalytická - TIM-barel zložený z 8 vnútorných β -vlákien obklopených zvonku 8 α -helixami; (ii) doména B - vsunutá do katalytickej domény medzi vláknom β 3 a helixom α 3; a (iii) doména C - C-terminálna doména - antiparalelná β -sendvičová doména.

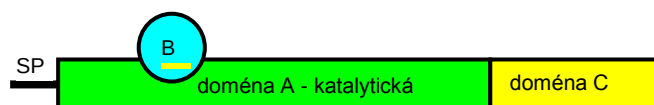


Schéma α -amylázy.



Zadanie

- (1) Vytvorte si priečinok „Amylase“. Všetkých 16 sekvencií domén A+B+C α -amyláz (podľa tabuľky) zhromaždíte z databázy UniProt do vstupného súboru „Amylase.txt“ (v priečinku Amylase), vhodného pre program CLUSTAL-Omega.
- (2) Vytvorte aj dva vstupné súbory α -amyláz pre sekvencie domén A+B: „Amylase_A_B.txt“ a pre sekvenciu domény C: „Amylase_C.txt“.
- (3) Na EBI serveri zrovnajte sekvencie domén „A+B+C“, „A+B“ a „C“ v programe CLUSTAL-Omega; pre každé zrovnanie získajte dva súbory: „*_aln.txt“ a „*_fas.txt“ (a uložte ich do priečinku Amylase).
- (4) Na EBI serveri v rámci programu CLUSTAL-W2 (Phylogeny) vypočítajte po dva evolučné stromy pre všetky tri súbory α -amyláz - s uvažovaním medzier v sekvenciách („*_off.ph“) a s ich ignorovaním („*_on.ph“).
- (5) Pre zrovnané sekvencie α -amyláz zistite konsenzuálnu dĺžku (CL), vypočítajte sekvenčnú identitu (SI) a podobnosť (SS). Vypočítané evolučné stromy zobrazte v programe TreeView a postupne diskutujte.
- (6) Údaje CL, SI a SS spolu s obrázkami všetkých evolučných stromov (s ich označením v legende), ako aj ich diskusiou uložte do súboru „Amylase_results.doc“.
- (7) Súbor zrovnaných domén A+B, t.j. „Amylase_A_B_aln.txt“ uložte ako dokument typu „*_aln.doc“ a podfarbite v ňom žltým zvýraznením tri regióny okolo katalytických zvyškov: Asp206, Glu230 a Asp297 (α -amyláza z *Aspergillus oryzae*; GLRIDTVKH, EVLD a FVENHD). Identifikujte katalytickú triedu pre α -amylázu z *Thermococcus hydrothermalis*.

Tabuľka. α -Amylázy a ich charakteristika.

Č.	Organizmus	Skratka	Taxonómia	Dĺžka	SP	A+B+C	A+B	C	UniProt
1	<i>Pyrococcus furiosus</i>	Pycfu	Archaea	460	26	27-460	27-388	389-460	O08452
2	<i>Pyrococcus woesei</i>	Pycwo	Archaea	460	26	27-460	27-388	389-460	Q7LYT7
3	<i>Thermococcus hydrothermalis</i>	Thchy	Archaea	457	23	24-457	24-385	386-457	O93647
4	<i>Thermococcus onnurineus</i>	Thcon	Archaea	458	24	25-458	25-386	387-458	B6YUG5
5	Jačmeň AMY1 (izozým s nízkym pI)	Horvu_1	Eucarya; rastlina	438	24	25-438	25-367	368-438	P00693
6	Jačmeň AMY2 (izozým s vysokým pI)	Horvu_2	Eucarya; rastlina	427	24	25-427	25-365	366-427	P04063
7	Fazuľa	Phavu	Eucarya; rastlina	420	23	24-420	24-360	361-420	Q9ZP43
8	Zemiak	Soltu	Eucarya; rastlina	407	18	19-407	19-348	349-407	Q41442
9	<i>Bacillus licheniformis</i>	Bacli	Bacteria; termofil	512	29	30-512	30-452	453-512	P06278
10	<i>Bacillus subtilis</i>	Bacsu	Bacteria; termofil	660	27	42-477	42-392	393-477	P00691
11	<i>Pseudoalteromonas haloplanktis</i>	Psaha	Bacteria; psychrofil	669	24	25-477	25-384	385-477	P29957
12	<i>Streptomyces limosus</i>	Strli	Bacteria; aktinomycéta	566	28	29-464	29-382	383-464	P09794
13	<i>Thermotoga maritima</i>	Thema	Bacteria; termofil	553	22	23-553	23-460	461-553	P96107
14	<i>Aspergillus oryzae</i>	Aspor	Eucarya; huba	499	21	22-499	22-406	407-499	P0C1B3
15	Ovocná muška	Drome	Eucarya; hmyz	494	18	19-494	19-406	407-494	P08144
16	Prasa (pankreas)	Sussc	Eucarya; cicavec	511	15	16-511	16-423	424-511	P00690

Dĺžka, počet aminokyselinových zvyškov kompletného proteínu; SP, dĺžka reálneho alebo hypotetického signálneho peptidu; A+B+C, dospelý α -amylázový proteín - katalytická doména A (TIM-barel), doména B (včlenená do domény A v oblasti medzi vláknom β 3 a helixom α 3) a doména C (C-terminálna doména nasledujúca katalytickú doménu); A+B, oblasť domén A a B; C, oblasť domény C; UniProt, prístupové číslo z UniProt databázy.

13.7. Predikcia sekundárnej a terciárnej štruktúry proteínov

Proteín: genetonín-1

- UniProt Acc. No.: O95210;
- cicavčí proteín z kostrového svalu s doteraz neznámou funkciou (väzba glykogénu);
- 358 aminokyselinových zvyškov;
- C-terminálna oblasť 260-358 je homologická so škrob-viažucou doménou mikrobiálnych amyláz z rodiny CBM20 (napr. cyklodextrín glukano-transferáza z *Bacillus circulans*: P43379; škrob-viažuca doména: 610-713 → 583-686 po odpočítaní signálneho peptidu 27 aa).

Zadanie

(1) Zo sekvencie genetonínu-1 vystrihnite C-terminálnu oblasť 260-358 a vykonajte predikciu sekundárnej štruktúry danej časti genetonínu metódou GOR (server: <http://gor.bb.iastate.edu/>). Získanú predpovedanú sekundárnu štruktúru (C, coil; H, helix a E, extended) zapíšte nad aminokyselinovú sekvenciu a uložte do súboru Gen_SBD_2D.txt (podľa bodu 2).

(2) Z predikcie sekundárnej štruktúry sekvencie C-terminálnej oblasti 260-358 genetonínu-1 metódou Profile Network Prediction Heidelberg (PROF, PHD; dostupnej na adrese: http://imb.savba.sk/~janecek/UCM/piataci/Gen_PHd_results.htm) vyberte výsledky predikcie PROF_sec (spolu s REL_sec), zapíšte pod aminokyselinovú sekvenciu a uložte do súboru Gen_SBD_2D.txt:

No.	1.....2.....3.....4.....5.....6.....7.....8.....9.....
GOR-V	CCCCCEEEEECCCCCCCCCEEECCCCCCCCCCCCCCCCCCCCCEEECCCCCHHHHHHHCCCCCEEECCCCCEEECCCCCEEECCCC
Sequence	GSQQVSVRFQVHYVTSTDVQFIAVTGDHECLGRWNTYIPLHYNKDGFWSHSIFLPADTVVEWKFVLVENGGVTRWEECSNRFLETGHEDKVVHAWWGIH
PROF_sec	EEEEEEEEEEEE EEEEEEE EEE EEEEEEE EEEEEEE EEEEE EEEEE EEEEEEE
Rel_sec	984256777776520146407888632332243243201220367426778882477436788888606842665305641565135632566311125

(3) Zo sekvencie genetonínu-1 vystrihnite C-terminálnu oblasť 260-358 a vykonajte predikciu terciárnej štruktúry danej časti genetonínu metódou homologického modelovania na serveri Swiss-Model - <http://swissmodel.expasy.org/> (tzv. „Automated Mode“). Ako homologický proteín („Specific template“) použite CGTázu z *Bacillus circulans*; PDB: 1CDG, reťazec „A“. Získaný model štruktúry C-koncovej domény genetonínu-1 uložte ako Gen_SBD_3D_SwissModel.txt (Gen_SBD_3D_SwissModel.pdb) a v programe WebLabViewerLite zobrazte vo viacerých zobrazovacích štýloch (samostatný tréning práce zobrazovania štruktúr proteínov).

- (4) Pre C-terminálnu oblasť 260-358 genetónínu-1 vykonajte predikciu terciárnej štruktúry aj metódou rozpoznania potenciálnej terciárnej štruktúry (tzv. „Fold Recognition“) na serveri Phyre-2 - <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>. Po prezretí výsledkov vyberte 2 najlepšie (najpravdepodobnejšie) modely pre C-terminálnu oblasť genetónínu-1. Modely uložte ako Gen_SBD_3D_Phyre2_1.txt a Gen_SBD_3D_Phyre2_2.txt (*_1.pdb a *_2.pdb) a v programe WebLabViewerLite zobrazte vo viacerých zobrazovacích štýloch. Modely vizuálne porovnajte navzájom, ako aj s výsledkami z modelovania na serveri Swiss-Model.
- (5) Z proteínovej databanky PDB (<http://www.rcsb.org/pdb/>) získajte koordináty cyklodextrínglukanotransferázy (CGTázy) z *Bacillus circulans* - PDB kód: 1CDG a pripravte si súbor 1cdg_SBD.txt (resp. 1cdg_SBD.pdb) štruktúry jej skutočnej škrob-viažucej domény (doména E: zvyšky Ser583-Pro686).
- (6) Všetky 4 proteínové domény (1× genetónínová SBD zo SwissModel; 2× genetónínová SBD z Phyre-2 a 1× reálna SBD z CGTázy) zobrazte pri vypnutí atómov ako „schematic“ a farbite podľa „secondary type“. Potom zobrazte aminokyselinové zvyšky vybraných funkčne dôležitých zvyškov - ako „stick“ a farbenie“ nasledovne: genetónín-1 - Trp34 a Trp75 (293 a 334) (zelená), Trp48 (307) (červená), Trp96 (355) (modrá) a Lys63 (321) (čierna); CGTáza - Trp616 a Trp662 (zelená), Trp 636 (červená), Trp684 (modrá) a Lys651 (čierna). Všetky 4 zobrazené štruktúry uložte ako „*.msv“ súbory a pokúste sa ich vizuálne porovnať.

13.8. Modelovanie a porovnávanie terciárnej štruktúry proteínov

Máte k dispozícii, resp. získajte

- (1) Koordináty modelu terciárnej štruktúry hypotetickej škrob-viažucej domény (SBD; CBM20) ľudského genetonínu-1 (zvyšky Gly260-His358), získané metódou homologického modelovania (Swiss-Model; „Automated Mode“; homologický proteín: „Specific template“ - CGTáza z *Bacillus circulans*; PDB kód: 1CDG, reťazec „A“); tento súbor označte: Gen_CBM20_Swiss_1CDG.pdb;
- (2) Koordináty modelu terciárnej štruktúry hypotetickej škrob-viažucej domény (SBD; CBM20) ľudského genetonínu-1 (zvyšky Gly260-His358), získané metódou homologického modelovania (Swiss-Model; „Automated Mode“; homologický proteín: „Specific template“ - škrob-viažuca doména CBM20 glukoamylázy z *Aspergillus niger* (PDB kód: 1AC0, reťazec „A“); tento súbor označte: Gen_CBM20_Swiss_1AC0.pdb;
- (3) Koordináty experimentálne určenej terciárnej štruktúry reálnej škrob-viažucej domény (SBD; CBM20) CGTázy z *Bacillus circulans* (doména E: zvyšky Ser583-Pro686): 1CDG_SBD.pdb; získané z databázy PDB (kód: 1CDG; vystrihnutím príslušných riadkov pre dané aminokyselinové zvyšky); tento súbor označte: 1CDG_CBM20.pdb;
- (4) Koordináty experimentálne určenej terciárnej štruktúry reálnej škrob-viažucej domény (SBD; CBM20) glukoamylázy z *Aspergillus niger* (PDB kód: 1AC0); súbor je pripravený s dvoma molekulami β -cyklodextrínu; tento súbor označte: 1AC0_CBM20.pdb;
- (5) Koordináty experimentálne určenej terciárnej štruktúry reálnej glykogén-viažucej domény (GBD; CBM48) z β -podjednotky AMP-aktivovanej proteín-kinázy z potkana (PDB kód: 1Z0M) - iba molekula A s príslušnou molekulou β -cyklodextrínu; súbor je pripravený s jednou molekulou β -cyklodextrínu; tento súbor označte: 1Z0M_CBM48.pdb;
- (6) Koordináty modelu terciárnej štruktúry hypotetickej škrob-viažucej domény (SBD; CBM20) ľudského genetonínu-1 (zvyšky Gly260-His358), získané metódou Phyre-2 (tzv. „fold recognition“ - rozpoznanie štruktúry) pomocou templátu škrob-viažucej domény glukoamylázy z *Aspergillus niger* (model č. 1; PDB kód: 1KUL); tento súbor označte: Gen_CBM20_Phyre2_1KUL.pdb.

Úlohy

- (1) Vytvorte si 15 priečinkov a v nich 15 ZIP súborov obsahujúcich všetky príslušné dvojice štruktúr škrob-viažucej domény podľa nasledovného členenia:

- 1: Gen_CBM20_Swiss_1CDG.pdb a Gen_CBM20_Swiss_1AC0.pdb;
- 2: Gen_CBM20_Swiss_1CDG.pdb a 1CDG_CBM20.pdb;
- 3: Gen_CBM20_Swiss_1CDG.pdb a 1AC0_CBM20.pdb;
- 4: Gen_CBM20_Swiss_1CDG.pdb a 1Z0M_CBM48.pdb;
- 5: Gen_CBM20_Swiss_1CDG.pdb a Gen_CBM20_Phyre2_1KUL.pdb;
- 6: Gen_CBM20_Swiss_1AC0.pdb a 1CDG_CBM20.pdb;
- 7: Gen_CBM20_Swiss_1AC0.pdb a 1AC0_CBM20.pdb
- 8: Gen_CBM20_Swiss_1AC0.pdb a 1Z0M_CBM48.pdb
- 9: Gen_CBM20_Swiss_1AC0.pdb a Gen_CBM20_Phyre2_1KUL.pdb
- 10: 1CDG_CBM20.pdb a 1AC0_CBM20.pdb
- 11: 1CDG_CBM20.pdb a 1Z0M_CBM48.pdb
- 12: 1CDG_CBM20.pdb a Gen_CBM20_Phyre2_1KUL.pdb
- 13: 1AC0_CBM20.pdb a 1Z0M_CBM48.pdb
- 14: 1AC0_CBM20.pdb a Gen_CBM20_Phyre2_1KUL.pdb
- 15: 1Z0M_CBM48.pdb a Gen_CBM20_Phyre2_1KUL.pdb

(2) Na serveri MultiProt: <http://bioinfo3d.cs.tau.ac.il/MultiProt/> podľa požiadaviek (upload a ZIP file of PDB structures) spravte postupne vzájomné preloženia všetkých 15 dvojíc štruktúr a poskytnuté údaje zapíšte do tabuľky (znázornená nižšie) - hodnotu RMSD a počet korešpondujúcich Ca atómov (v zátvorke):

(3) Všetkých 15 preložených štruktúr (PDB alignment) spolu s príslušným zrovnaním (Alignment size) a výsledkami uložte s názvom ZIP súboru (napr. „1_overlap.pdb“, „1_alignment.htm“ a „1_vysledky.htm“ - ako kompletnú web-stránku; atď.). Zrovnania vizuálne prezrite s ohľadom na funkčne dôležité zvyšky (udané nižšie) a preklady zobrazte v programe WebLab pri vypnutí atómov ako „solid ribbon“ a farbte vždy jednu štruktúru tyrkysovou a druhú oranžovou farbou. Potom zobrazte aminokyselinové zvyšky vybraných funkčne dôležitých zvyškov - ako „stick“: (i) genetonín-1 - Trp34 (293), Trp48 (307), Lys63 (321) a Trp75 (334); (ii) CGTáza - Trp616, Trp 636, Lys651 a Trp662; (iii) glukaoamyláza: - Trp543, Trp563, Lys578 a Trp590; a (iv) β -podjednotka AMPK: Trp100, Phe112, Lys126 a Trp133; a farbte vždy modrou v tyrkysovej štruktúre a fialovou v oranžovej štruktúre. Molekuly β -cyklodextrínov zobrazujte ako „stick“ a farbte zelenou v tyrkysovej štruktúre a žltou v oranžovej štruktúre. Všetky zobrazenia uložte ako príslušné „*.msv“ a „*.gif“ súbory.

Tabuľka s výsledkami po superpozícii terciárnych štruktúr.

	Gen_CBM20_Swiss_1CDG.pdb	Gen_CBM20_Swiss_1AC0.pdb	1CDG_CBM20.pdb	1AC0_CBM20.pdb	1Z0M_CBM48.pdb
Gen_CBM20_Swiss_1AC0.pdb	1,94 (74)	---	---	---	---
1CDG_CBM20.pdb	0,50 (93)	1,87 (70)	---	---	---
1AC0_CBM20.pdb	1,74 (71)	0,79 (85)	2,00 (79)	---	
1Z0M_CBM48.pdb	1,34 (66)	1,94 (55)	1,44 (68)	1,87 (58)	---
Gen_CBM20_Phyre_1KUL.pdb	1,55 (79)	1,86 (70)	1,53 (79)	1,69 (72)	1,65 (66)

13.9. *In silico* analýza proteínu pomocou primárnej štruktúry

Teoretická časť

Predchádzajúcim výskumom ste zistili, že hypotetická amylomaltáza z *Borrelia burgdorferi* (t.j. sekvencia, ktorá je známa iba zo sekvenovania genómu tejto baktérie; MalQ, BB0166; GenBank Acc. No. AAC66547.1) obsahuje zaujímavé (unikátne) sekvenčné črty, ktorými sa odlišuje od ostatných typických amylomaltáz. Podobné črty obsahuje aj sekvencia z *Borrelia garinii*, pričom sekvencia z *Borrelia turicatae* môže byť považovaná za akýsi „intermediát“ (medzičlánok) - má črty aj typických amylomaltáz, aj amylomaltáz z borélií. Sú to najmä (v sekvencii amylomaltázy z *B. burgdorferi* BB0166): Ser49 (namiesto Gly), Ala57 (namiesto Pro), Asn227 (namiesto Asp), Lys305 (namiesto Arg), Phe356-Gln357, resp. Phe-Glu (namiesto Leu-Gly) a Gly406 (namiesto His). Rozhodli ste sa preto izolovať proteín amylomaltázy z *B. burgdorferi* a v rámci tejto práce po presekvenovaní reálneho génu overiť, či unikátne sekvenčné črty sú naozaj prítomné, t.j. či nie sú dôsledkom nejakej sekvenačnej chyby pri sekvenovaní genómu.

Úlohy

- (1) Stiahnite si aminokyselinové sekvencie (tabuľka) z databázy GenBank („Protein“ database) a zrovnajte ich na EBI serveri pre program Clustal-Omega. Pripravte aj ich evolučný strom počítaný zo zrovnania programom Clustal-W2, v ktorom sa medzery neignorujú (súbory: AGT.txt, AGT_aln.txt, AGT_fas.txt, AGT_off.ph).

Tabuľka s amylomaltázami.

Č.	Zdroj	Skratka	GenBank	Proteín	Dĺžka
Borélie					
1.	<i>Borrelia burgdorferi</i>	Borbu	AAC66547.1	Hypotetický	506
2.	<i>Borrelia garinii</i>	Borga	AAU07022.1	Hypotetický	506
3.	<i>Borrelia turicatae</i>	Bortu	AAX17503.1	Hypotetický	493
Baktérie					
4.	<i>Aquifex aeolicus</i>	Aquae	CAE51853.1	Reálny	485
5.	<i>Clostridium butyricum</i>	Clobu	AAB84229.1	Reálny	487
6.	<i>Escherichia coli</i>	Escco	AAA24106.1	Reálny	694
7.	<i>Thermus aquaticus</i>	Theaq	BAA33728.1	Reálny	500
8.	<i>Thermus thermophilus</i>	Theth	CAE51852.1	Reálny	500
Archaeobaktérie					
9.	<i>Pyrobaculum aerophilum</i>	Pyrae	AAL63326.1	Reálny	468
Rastliny					
10.	<i>Chlamydomonas reinhardtii</i>	Chlre	AAG29840.1	Reálny	585
11.	<i>Solanum tuberosum</i>	Soltu	CAA48630.1	Reálny	576
12.	<i>Triticum aestivum</i>	Triae	AAZ23612.1	Reálny	574

- (2) V zrovnání (v programe MS-Word; súbor AGT_aln.doc) vyznačte unikátne sekvenčné črty v enzýmoch z borélií (*B. burgdorferi* a *B. garinii*) žltým zvýraznením a odpovedajúce aminokyselinové zvyšky v enzýmoch z ostatných zdrojov zeleným zvýraznením. Črty v amylomaltáze z *B. turicatae* farbte podľa toho, či sú typické pre jednu alebo druhú skupinu.
- (3) Po presekvenovaní génu z Vašej reálnej amylomaltázy z *B. burgdorferi* ste získali jej nukleotidovú sekvenciu dlhú 1527 nukleotidov (http://imb.savba.sk/~janecek/UCM/Piataci/BB0166_real.txt). Na proteomickom serveri (<http://www.expasy.ch/>; v sekcii „Proteomics“ - „Protein sequences and identification“ - „Translate“) získajte jej správny preklad, t.j. správnu aminokyselinovú sekvenciu, a túto porovnajte s aminokyselinovou sekvenciou hypotetickej amylomaltázy z *B. burgdorferi*. Skontrolujte zrovnanie, či sa unikátne sekvenčné črty potvrdili alebo nie (súbor: Borelie_aln.txt).
- (4) Poznáte hodnoty pI a Mw pre hypotetickú amylomaltázu z *B. burgdorferi*: 8.67 a 59,354 kDa. Na proteomickom serveri (<http://www.expasy.ch/>; v sekcii „Proteomics“ - „Protein characterization and function“ - „Compute pI/Mw“) získajte tieto hodnoty pre Vašu reálnu amylomaltázu a výsledky porovnajte pre hypotetický a reálny proteín. Výsledky zapíšte a diskutujte.

Použitá a odporúčaná literatúra

- Banner D.W., Bloomer A., Petsko G.A., Phillips D.C. & Wilson I.A. 1976. Atomic coordinates for triose phosphate isomerase from chicken muscle. *Biochem. Biophys. Res. Commun.* **72**: 146-155.
- Benson D.A., Clark K., Karsch-Mizrachi I., Lipman D.J., Ostell J. & Sayers E.W. 2014. GenBank. *Nucleic Acids Res.* **42 (Database issue)**: D32-D37.
- Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. & Bourne P.E. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28**: 235-242.
- Bohinski R.C. 1983. *Modern Concepts in Biochemistry*, 4th Edition. Allyn & Bacon, Boston, ISBN 0205079059.
- Chothia C. 1992. Proteins. One thousand families for the molecular biologist. *Nature* **357**: 543-544.
- Duan X., Chen J. & Wu J. 2013. Improving the thermostability and catalytic efficiency of *Bacillus deramificans* pullulanase by site-directed mutagenesis. *Appl. Environ. Microbiol.* **79**: 4072-4077.
- Eisenberg D. 2003. The discovery of the α -helix and β -sheet, the principal structural features of proteins. *Proc. Natl. Acad. Sci. USA* **100**: 11207-11210.
- Da Lage J.L., Feller G. & Janecek S. 2004. Horizontal gene transfer from Eukarya to bacteria and domain shuffling: the α -amylase model. *Cell. Mol. Life Sci.* **61**: 97-109.
- Declerck N., Machius M., Joyet P., Wiegand G., Huber R. & Gaillardin C. 2003. Hyperthermostabilization of *Bacillus licheniformis* α -amylase and modulation of its stability over a 50 degrees C temperature range. *Protein Eng.* **16**: 287-293.
- Farber G.K. & Petsko G.A. 1990. The evolution of α/β barrel enzymes. *Trends Biochem. Sci.* **15**: 228-234.
- Janecek S. 1993. Strategies for obtaining stable enzymes. *Process Biochemistry* **28**: 435-445.
- Janecek S. (2002) A motif of a microbial starch-binding domain found in human genethonin. *Bioinformatics* **18**: 1534-1537.
- Janecek S., Svensson B. & MacGregor E.A. 2003. Relation between domain evolution, specificity, and taxonomy of the α -amylase family members containing a C-terminal starch-binding domain. *Eur. J. Biochem.* **270**: 635-645.
- Janecek S., Svensson B. & MacGregor E.A. 2014. α -Amylase - an enzyme specificity found in various families of glycoside hydrolases. *Cell. Mol. Life Sci.* **71**: 1149-1170.
- Kelley L.A. & Sternberg M.J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**: 363-371.
- Klibanov A.M. 1983. Immobilized enzymes and cells as practical catalysts. *Science* **219**: 722-727.

- Lee H.S., Kim J.S., Shim K.H., Kim J.W., Park C.S. & Park K.H. 2005. Quaternary structure and enzymatic properties of cyclomaltodextrinase from alkalophilic *Bacillus* sp. I-5. *Biologia* **60** (Suppl. 16): 73-77.
- Leemhuis H., Dijkstra B.W. & Dijkhuizen L. 2002. Mutations converting cyclodextrin glycosyltransferase from a transglycosylase into a starch hydrolase. *FEBS Lett.* **514**: 189-192.
- Lesk A.M. 2001. *Introduction to Protein Architecture*. Oxford University Press, Oxford, ISBN 0198504748.
- Lesk A.M. 2004. *Introduction to Protein Science*. Oxford University Press, Oxford, ISBN 9780199265114.
- Matsuura Y., Kusunoki M., Harada W. & Kakudo M. 1984. Structure and possible catalytic residues of Taka-amylase A. *J. Biochem.* **95**: 697-702.
- Moran L.A., Horton R.A., Scrimgeour G. & Perry M. 2011. *Principles of Biochemistry*, 5th Edition. Pearson Education, Inc., Boston, ISBN 0321707338.
- Mozhaev V.V., Siksnis V.A., Melik-Nubarov N.S., Galkantaite N.Z., Denis G.J., Butkus E.P., Zaslavsky B.Y., Mestechkina N.M. & Martinek K. 1988. Protein stabilization via hydrophilization. Covalent modification of trypsin and alchyhotrypsin. *Eur. J. Biochem.* **173**: 147-154.
- Okuyama M., Kaneko A., Mori H., Chiba S. & Kimura A. 2006. Structural elements to convert *Escherichia coli* α -xylosidase (YicI) into α -glucosidase. *FEBS Lett.* **580**: 2707-2711.
- Orengo C.A., Michie A.D., Jones S., Jones D.T., Swindells M.B., Thornton J.M. 1997. CATH - a hierarchic classification of protein domain structures. *Structure* **5**: 1093-1108.
- Orengo C.A., Pearl F.M., Bray J.E., Todd A.E., Martin A.C., Lo Conte L. & Thornton J.M. 1999. The CATH database provides insights into protein structure/function relationships. *Nucleic Acids Res.* **27**: 275-279.
- Ozawa T., Hakamada Y., Hatada Y., Kobayashi T., Shirai T. & Ito S. 2001. Thermostabilization by replacement of specific residues with lysine in a *Bacillus* alkaline cellulase: building a structural model and implications of newly formed double intrahelical salt bridges. *Protein Eng.* **14**: 501-504.
- Pages S., Kester H.C., Visser J. & Benen J.A. 2001. Changing a single amino acid residue switches processive and non-processive behavior of *Aspergillus niger* endopolygalacturonase I and II. *J. Biol. Chem.* **276**: 33652-33656.
- Rodríguez-Sanoja R., Oviedo N., Escalante L., Ruiz B. & Sánchez S. 2009. A single residue mutation abolishes attachment of the CBM26 starch-binding domain from *Lactobacillus amylovorus* α -amylase. *J. Ind. Microbiol. Biotechnol.* **36**: 341-346.
- Rost B., Yachdav G. & Liu J. 2004. The PredictProtein server. *Nucleic Acids Res.* **32** (Web Server issue): W321-W326.
- Schmitt J., Brocca S., Schmid R.D. & Pleiss J. 2002. Blocking the tunnel: engineering of *Candida rugosa* lipase mutants with short chain length specificity. *Protein Eng.* **15**: 595-601.

- Schwede T., Kopp J., Guex N. & Peitsch M.C. 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **31**: 3381-3385.
- Seidle H.F., McKenzie K., Marten I., Shoseyov O. & Huber R.E. 2005. Trp-262 is a key residue for the hydrolytic and transglucosidic reactivity of the *Aspergillus niger* family 3 β -glucosidase: substitution results in enzymes with mainly transglucosidic activity. *Arch. Biochem. Biophys.* **444**: 66-75.
- Sen T.Z., Jernigan R.L., Garnier J. & Kloczkowski A. 2005. GOR V server for protein secondary structure prediction. *Bioinformatics* **21**: 2787-2788.
- Shatsky M., Nussinov R. & Wolfson H.J. 2004. A method for simultaneous alignment of multiple protein structures. *Proteins: Struct. Funct. Bioinform.* **56**: 143-156.
- Tafreshi N.K., Sadeghizadeh M., Emamzadeh R., Ranjbar B., Naderi-Manesh H. & Hosseinkhani S. 2008. Site-directed mutagenesis of firefly luciferase: implication of conserved residue(s) in bioluminescence emission spectra among firefly luciferases. *Biochem. J.* **412**: 27-33.
- Takano K., Yamagata Y. & Yutani K. 2003. Buried water molecules contribute to the conformational stability of a protein. *Protein Eng.* **16**: 5-9.
- Tang S.Y., Yang S.J., Cha H., Woo E.J., Park C. & Park K.H. 2006. Contribution of W229 to the transglycosylation activity of 4- α -glucanotransferase from *Pyrococcus furiosus*. *Biochim. Biophys. Acta* **1764**: 1633-1638.
- UniProt Consortium 2014. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **42 (Database issue)**: D191-D198.

PROTEÍNOVÝ DIZAJN

Autor: Ing. Štefan Janeček, DrSc.

Vydala: Univerzita sv. Cyrila a Metoda v Trnave, 2014

Vydanie: prvé.

Web: http://fpv.ucm.sk/images/ucebne_texty/Proteinovy_dizajn.pdf

Počet strán: 114

ISBN 978-80-8105-594-2