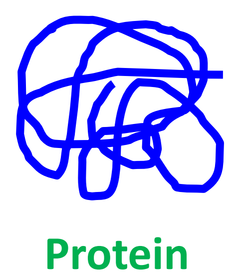
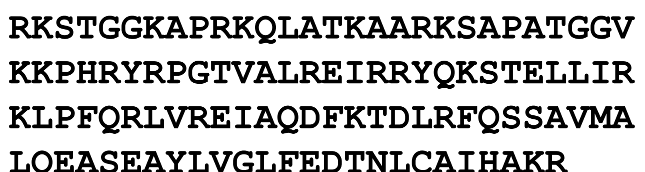
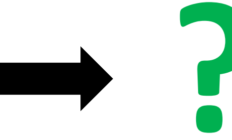
Predikce základních vlastností proteinů (a peptidů)

Aplikovaná bioinformatika, jaro 2023


Protein

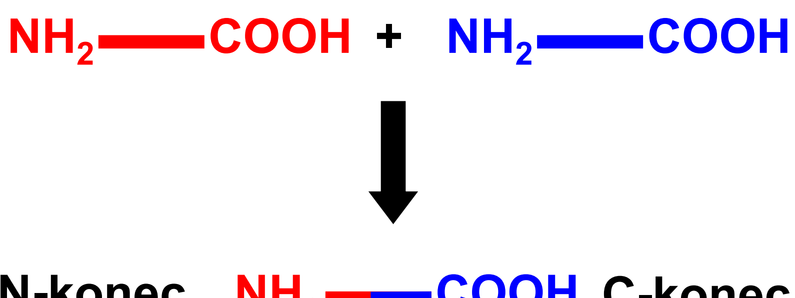




Lenka Malinovská

Proteiny

- Protein, polypeptid, bílkovina.
- Lineární polymer aminokyselin spojených peptidovými vazbami.
- Funkce: katalytická, regulační, transportní, zprostředkování pohybu, obranná, strukturální, zásobní.



N MTWCKTMIDQGRSWPHCYYGMAA
DTYYKKLTPGHTQVGITILMGAC
GCCCGTGCRNMSDETGCWWCGTA
HSPGCTDEQLRCGLVCGT **C**

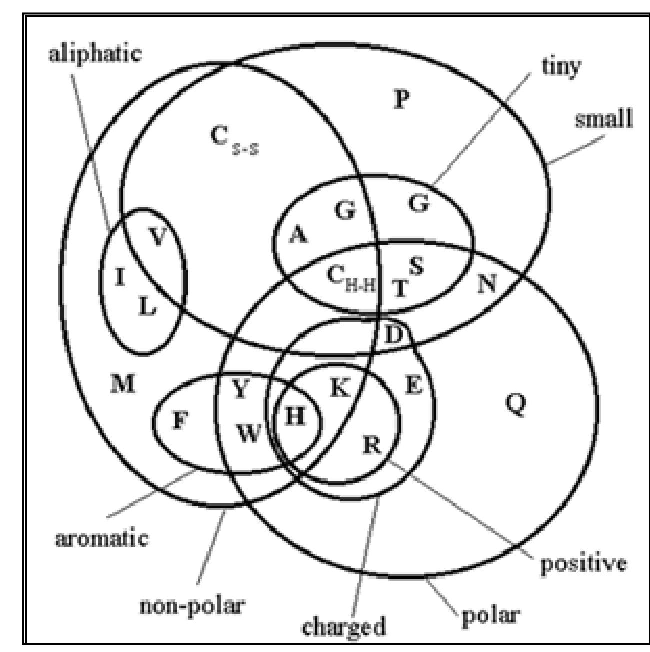
N → C, využíván jednopísmenný kód

$^{+}H_3N$–CH(R_1)–COO^- + $^{+}H_3N$–CH(R_2)–COO^- ⇌ $^{+}H_3N$–CH(R_1)–CO–NH–CH(R_2)–COO^- + H_2O

Peptide bond

Proteinogení aminokyseliny

- **Stavební jednotky proteinů: α-L-aminokyseliny.**
- **20 standardních proteinogenních aminokyselin.**
- **Alifatické (Gly, Ala, Val, Leu, Ile).**
- **Sirné (Cys, Met).**
- **S OH skupinou (Ser, Thr).**
- **Kyselé a z nich odvozené (Glu, Gln, Asp, Asn).**
- **Bazické (Lys, Arg).**



Aminokyseliny s podobnými vlastnostmi mohou plnit v proteinu stejné funkce – bývají vzájemně zastupitelné

D-aminokyseliny

An overview on D-amino acids

Giuseppe Genchi[1]

Abstract More than half a century ago researchers thought that D-amino acids had a minor function compared to L-enantiomers in biological processes. Many evidences have shown that D-amino acids are present in high concentration in microorganisms, plants, mammals and humans and fulfil specific biological functions. In the brain of mammals, D-serine (D-Ser) acts as a co-agonist of the *N*-methyl-D-aspartate (NMDA)-type glutamate receptors, responsible for learning, memory and behaviour. D-Ser metabolism is relevant for disorders associated with an altered function of the NMDA receptor, such as schizophrenia, ischemia, epilepsy and neurodegenerative disorders. On the other hand, D-aspartate (D-Asp) is one of the major regulators of adult neurogenesis and plays an important role in the development of endocrine function. D-Asp is present in the neuroendocrine and endocrine tissues and testes, and regulates the synthesis and secretion of hormones and spermatogenesis. Also food proteins contain D-amino acids that are naturally originated or processing-induced under conditions such as high temperatures, acid and alkali treatments and fermentation processes. The presence of D-amino acids in dairy products denotes thermal and alkaline treatments and microbial contamination. Two enzymes are involved in the metabolism of D-amino acids: amino acid racemase in the synthesis and D-amino acid oxidase in the degradation.

- Bakterie – antibiotika, složka buněčných stěn (peptidoglykany)
- Eukaryota - biologicky aktivní peptidy

Table 1 D-Aminoacid in eukaryotic peptides

Drug	D-Aminoacid	Source	Activity
ω-Agatoxin	D-Ser	Venom of funnel-web spider (*Agelenopsis aperta*)	Blocks sodium channels
Bombinins	D-Allo-Ile	Skin secretion of frogs (*Bombinatoridae*)	Antimicrobial and hemolytic activity
Contriphans	D-Trp	Venom of core snails (*Conus radiatus*)	Causes tremor and mucous secretions when injected into fish
Deltorphins	D-Ala	Skin secretions of three frogs (*Phyllomedusa bicolor*)	Binds to δ-type opiate receptors, acting as a hallucinogen
Dermorphins	D-Met	Skin secretions of three frogs (*Phyllomedusa sauvagi*)	Binds to μ-type opiate receptor and acts as an analgesic, more powerful than morphine
Achatin I	D-Phe	Ganglia and atrium of African snail (*Achatina fulica*)	Excitatory neurotransmitter controlling muscle contraction
Fucilin	D-Asn	Ganglia of African snail (*Achatina fulica*)	Excitatory neurotransmitter controlling penis contractions

- Savci – D-Ser, D-Asp
- D-aminokyseliny v potravě

A delicacy of traditional Chinese cuisine is represented by pidan. Duck or chicken eggs are immersed at room temperature for at least 30 days in an alkaline solution prepared with 4.2% NaOH and 5% NaCl, which leads to extensive racemization of all ovalbumin L-amino acids in D-amino acids with the concurrent formation of lysinoalanine (Chang et al. 1999).

D-aminokyseliny

An overview on D-amino acids

Giuseppe Genchi[1]

Abstract More than half a century ago researchers thought that D-amino acids had a minor function compared to L-enantiomers in biological processes. Many evidences have shown that D-amino acids are present in high concentration in microorganisms, plants, mammals and humans and fulfil specific biological functions. In the brain of mammals, D-serine (D-Ser) acts as a co-agonist of the *N*-methyl-D-aspartate (NMDA)-type glutamate receptors, responsible for learning, memory and behaviour. D-Ser metabolism is relevant for disorders associated with an altered function of the NMDA receptor, such as schizophrenia, ischemia, epilepsy and neurodegenerative disorders. On the other hand, D-aspartate (D-Asp) is one of the major regulators of adult neurogenesis and plays an important role in the development of endocrine function. D-Asp is present in the neuroendocrine and endocrine tissues and testes, and regulates the synthesis and secretion of hormones and spermatogenesis. Also food proteins contain D-amino acids that are naturally originated or processing-induced under conditions such as high temperatures, acid and alkali treatments and fermentation processes. The presence of D-amino acids in dairy products denotes thermal and alkaline treatments and microbial contamination. Two enzymes are involved in the metabolism of D-amino acids: amino acid racemase in the synthesis and D-amino acid oxidase in the degradation.

- D-aminokyseliny v potravě

D-Asn
D-Asp

D-Glu

D-Ala
D-Ser

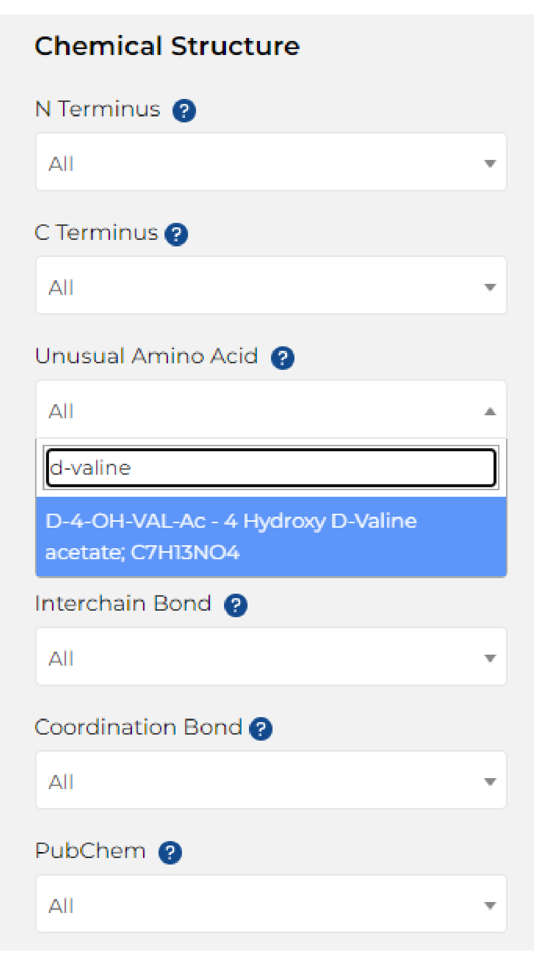
„Indeed, one of the biggest mysteries concerning the origin of life is why L-polypeptides were selected as proteins, while D-polypeptides were eliminated from primitive Earth.“

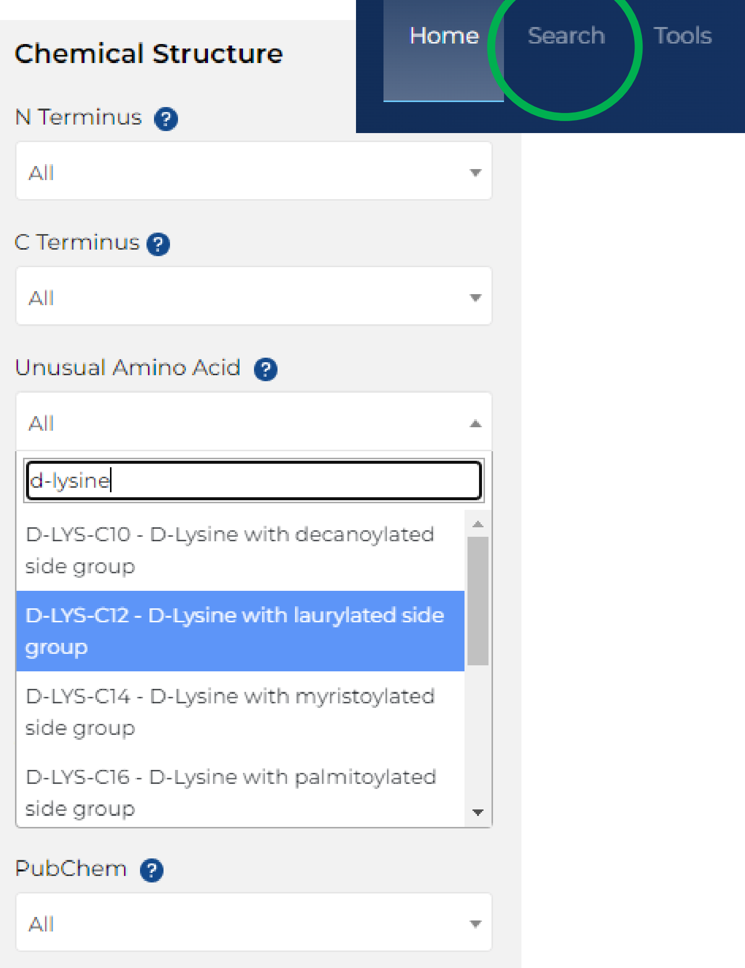
D-aminokyseliny

- Bakterie – antibiotika, složka buněčných stěn (peptidoglykany)

Database of Antimicrobial Activity and Structure of Peptides

https://dbaasp.org/home





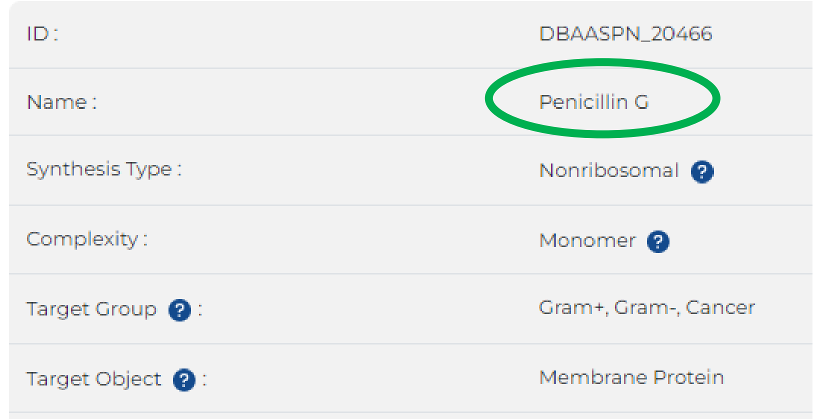
Úkol 1: Najděte v databázi „Penicillin“. Jakou D-aminokyselinu obsahuje? Jak jsou D-aminokyseliny vyznačené v sekvenci?

DBAASP Peptide Card

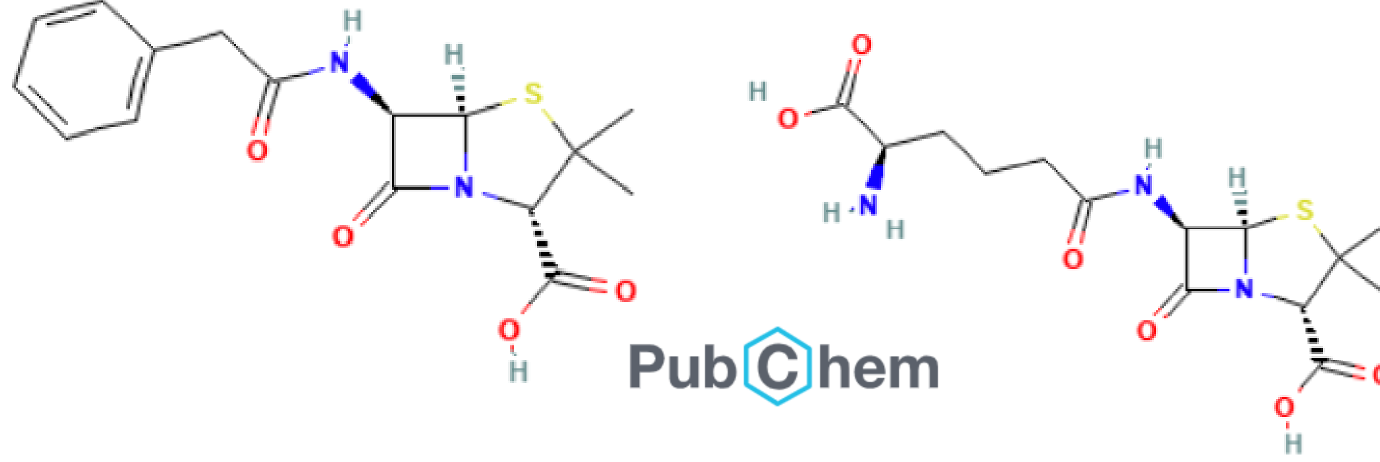
Database of Antimicrobial Activity and Structure of Peptides

https://dbaasp.org/home

General Information

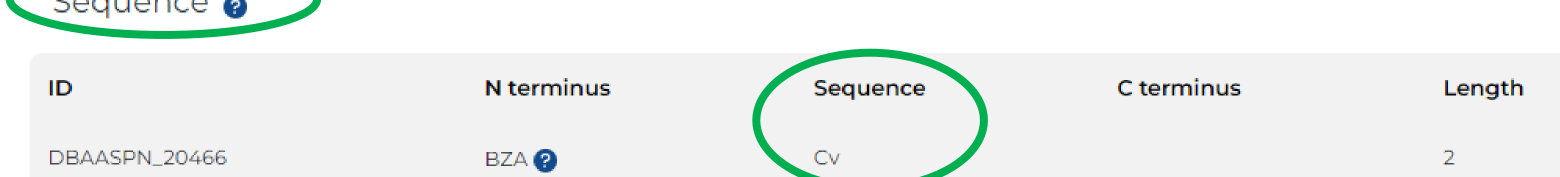


ID :	DBAASPN_20466
Name :	Penicillin G
Synthesis Type :	Nonribosomal
Complexity :	Monomer
Target Group :	Gram+, Gram-, Cancer
Target Object :	Membrane Protein



Chemical Structure

Sequence



ID	N terminus	Sequence	C terminus	Length
DBAASPN_20466	BZA	Cv		2

General Information

ID :	DBAASPN_20468
Name :	Penicillin N, Cephalosporin N, Adicillin
Synthesis Type :	Nonribosomal
Complexity :	Monomer
Target Group :	Gram+, Gram-, Fungus
Target Object :	Membrane Protein

D-peptidy

https://www.lifetein.com/Peptide-Synthesis-D-Amino-Acid.html

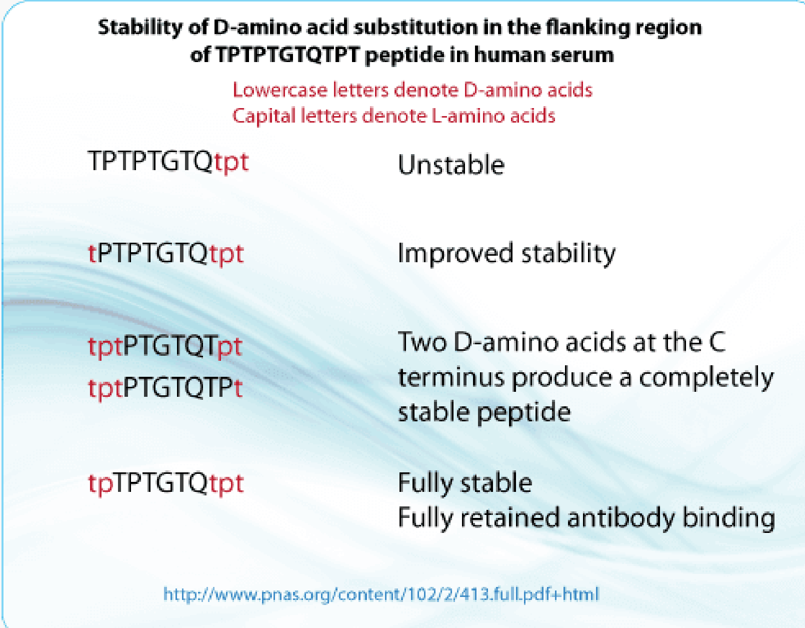
D Amino Acid and Retro-inverso Peptides

Get Quote

Proteins and most naturally occurring peptides are composed of amino acids in the L-configuration. However, D-amino acids have been detected in a variety of peptides synthesized in animal cells. Examples include opiate and antimicrobial peptides from frog skin, neuropeptides from snails, hormones from crustaceans, and venom from spiders. These D-amino acids form when L-amino acids undergo posttranslational alterations.

Case study 1: D-amino acid peptides are resistant to proteases

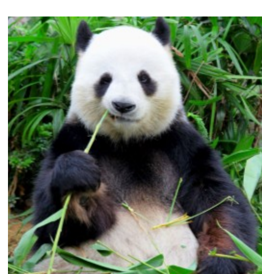
Many proteins are easily degraded or tagged for selective destruction in cells. Peptides that are at least partially made of D-amino acids have shown strong resistance to proteolytic degradation.



Improved Protease Stability of the Antimicrobial Peptide Pin2 Substituted with D-Amino Acids

G. Carmona · A. Rodriguez · D. Juarez · G. Corzo · E. Villegas

Pin2 = Pandanin 2



Pandinus imperator

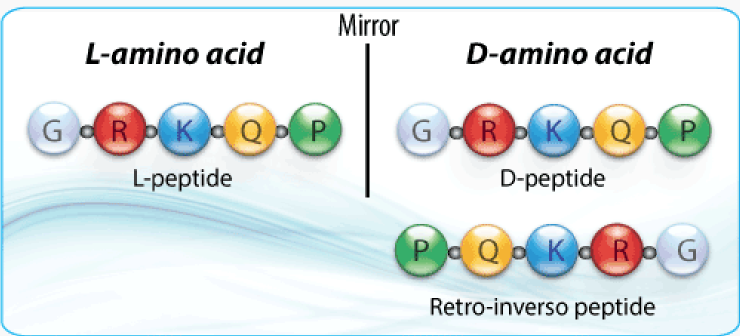
D-peptidy

https://www.lifetein.com/Peptide-Synthesis-D-Amino-Acid.html

D Amino Acid and Retro-inverso Peptides

Get Quote

Proteins and most naturally occurring peptides are composed of amino acids in the L-configuration. However, D-amino acids have been detected in a variety of peptides synthesized in animal cells. Examples include opiate and antimicrobial peptides from frog skin, neuropeptides from snails, hormones from crustaceans, and venom from spiders. These D-amino acids form when L-amino acids undergo posttranslational alterations.



- **L peptide**: GRKQP or Gly-Arg-Lys-Gln-Pro; (Example 2: RKDVY, Arg–Lys–Asp–Val–Tyr, TP5, Thymopentin)
- **D peptide**: (dG)(dR)(dK)(dQ)(dP) or D (Gly-Arg-Lys-Gln-Pro), (Example 2: dR-dK-dD-dV-dY, D(Arg–Lys–Asp–Val–Tyr), D-TP5, Thymopentin)
- **Retro-inverso**: (dP)(dQ)(dK)(dR)(dG) or D (Pro-Gln-Lys-Arg-Gly); (Example 2: d(YVDKR), D(Tyr-Val-Asp-Lys-Arg), Retro-inverso Thymopentin)

Retro-inverso peptides

Retro-inverso peptides are composed of D-amino acids that are assembled in the reverse order of their parental L-sequences. Retro-inverso peptides are obtained by replacing the normal L-amino acid residues with the corresponding D-amino acids and reversing the direction of the peptide backbone. Therefore, the original spatial orientation and the chirality of the side chains is unchanged. This results in a non-complementary side chain topochemistry between the analog and the parental L-peptide.

Advantages of retro-inverso peptides:

- Retention of protein bioactivity
- Long-lasting proteolytic stability in vivo
- Antigenic mimicry of natural L-peptides

Review

Recent Applications of Retro-Inverso Peptides

Nunzianna Doti [1], Mario Mardirossian [2], Annamaria Sandomenico [1], Menotti Ruvo [1,*] and Andrea Caporale [3,*]

Article

De Novo Design and In Vitro Testing of Antimicrobial Peptides against Gram-Negative Bacteria

Boris Vishnepolsky [1,*], George Zaalishvili [2], Margarita Karapetian [2], Tornike Nasrashvili [2], Nato Kuljanishvili [2], Andrei Gabrielian [3], Alex Rosenthal [3], Darrell E. Hurt [3], Michael Tartakovsky [3], Maya Grigolava [1] and Malak Pirtskhalava [1,*]

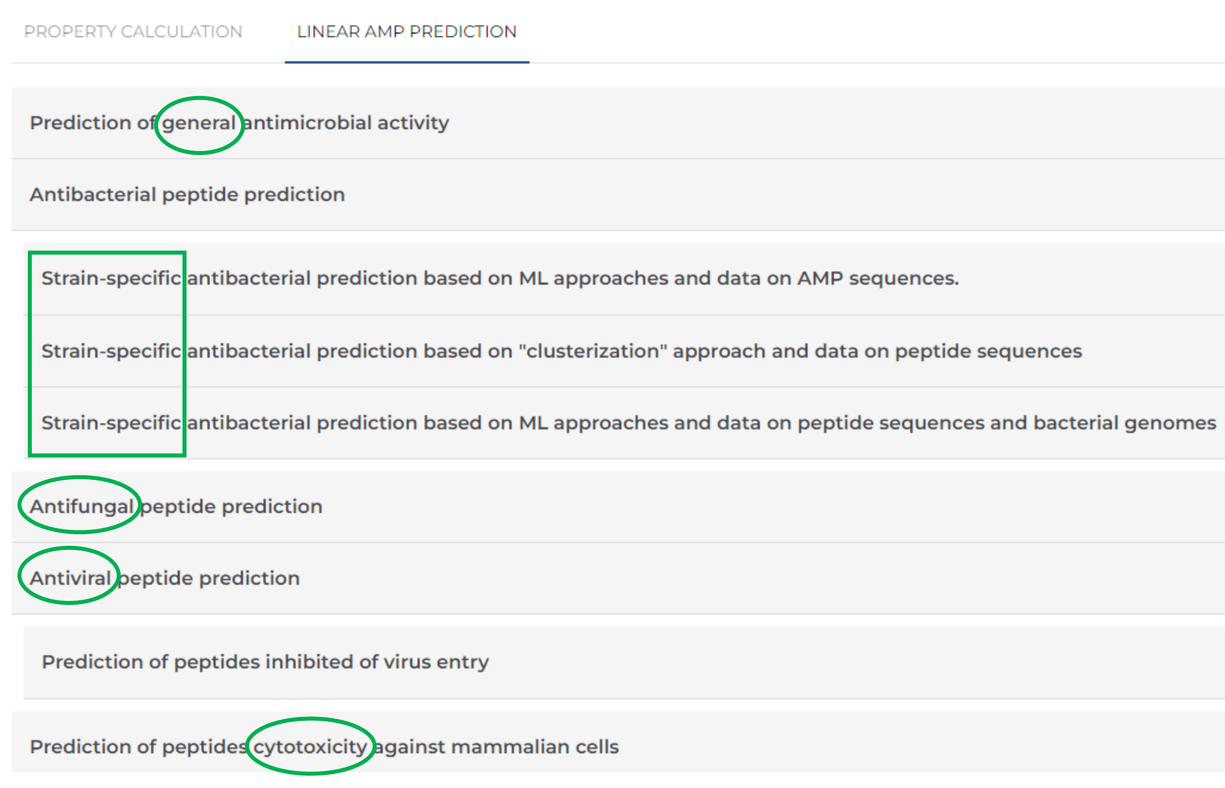
De novo design peptidů

Article

De Novo Design and In Vitro Testing of Antimicrobial Peptides against Gram-Negative Bacteria

Boris Vishnepolsky [1,*], George Zaalishvili [2], Margarita Karapetian [2], Tornike Nasrashvili [2], Nato Kuljanishvili [2], Andrei Gabrielian [3], Alex Rosenthal [3], Darrell E. Hurt [3], Michael Tartakovsky [3], Maya Grigolava [1] and Malak Pirtskhalava [1,*]

- Problémy AMP: degradace proteasami, toxické pro savčí buňky, produkce je drahá.
- Design vhodných AMP: experimentální a výpočetní metody.
- Mnoho různých nástrojů pro predikci antimikrobiální aktivity, toxicity a stability AMP.



https://dbaasp.org/tools?page=linear-amp-prediction

Úkol 2: Predikujte antimikrobiální aktivitu následujících peptidů vůči Staphylococcus aureus a Pseudomonas aeruginosa.

FLPLIGRVLSGIL
ALWKTLLKKVLKA

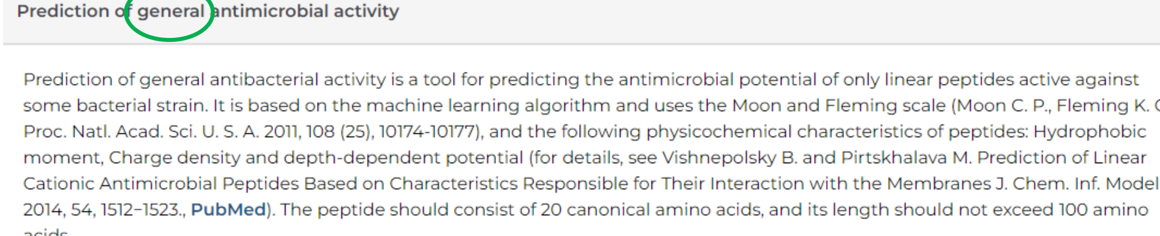
De novo design peptidů

Article

De Novo Design and In Vitro Testing of Antimicrobial Peptides against Gram-Negative Bacteria

Boris Vishnepolsky [1,*], George Zaalishvili [2], Margarita Karapetian [2], Tornike Nasrashvili [2], Nato Kuljanishvili [2], Andrei Gabrielian [3], Alex Rosenthal [3], Darrell E. Hurt [3], Michael Tartakovsky [3], Maya Grigolava [1] and Malak Pirtskhalava [1,*]

- Problémy AMP: degradace proteasami, toxické pro savčí buňky, produkce je drahá.
- Design vhodných AMP: experimentální a výpočetní metody.
- Mnoho různých nástrojů pro predikci antimikrobiální aktivity, toxicity a stability AMP.



Paste sequence(s) in FASTA format (the peptide sequence can contain the '+' sign to the end in case of C-terminal amidation):

```
>1
FLPLIGRVLSGIL

>2
ALWKTLLKKVLKA
```

Submit

Seq. ID	Class
1	AMP
2	AMP

2 out of 2 sequences are AMP(s)

Úkol 2: Predikujte antimikrobiální aktivitu následujících peptidů vůči Staphylococcus aureus a Pseudomonas aeruginosa.

FLPLIGRVLSGIL
ALWKTLLKKVLKA

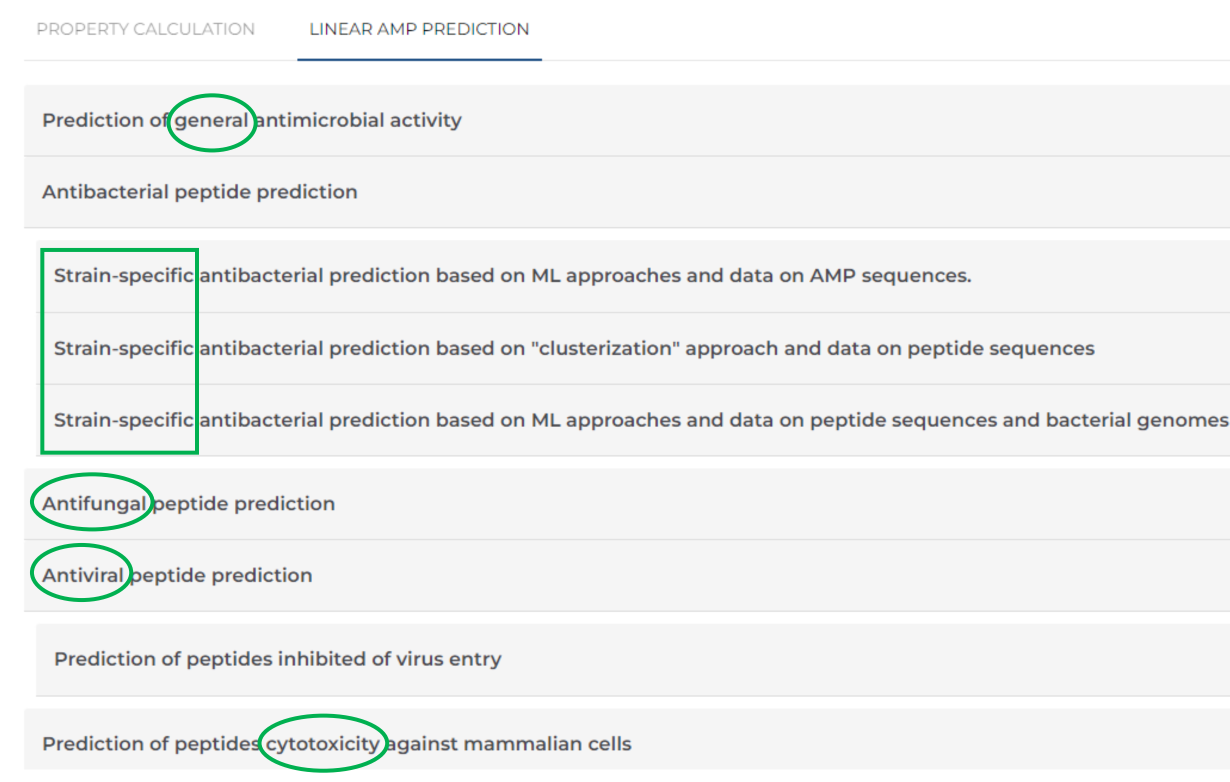
De novo design peptidů

Article

De Novo Design and In Vitro Testing of Antimicrobial Peptides against Gram-Negative Bacteria

Boris Vishnepolsky [1,*], George Zaalishvili [2], Margarita Karapetian [2], Tornike Nasrashvili [2], Nato Kuljanishvili [2], Andrei Gabrielian [3], Alex Rosenthal [3], Darrell E. Hurt [3], Michael Tartakovsky [3], Maya Grigolava [1] and Malak Pirtskhalava [1,*]

- Problémy AMP: degradace proteasami, toxické pro savčí buňky, produkce je drahá.
- Design vhodných AMP: experimentální a výpočetní metody.
- Mnoho různých nástrojů pro predikci antimikrobiální aktivity, toxicity a stability AMP.

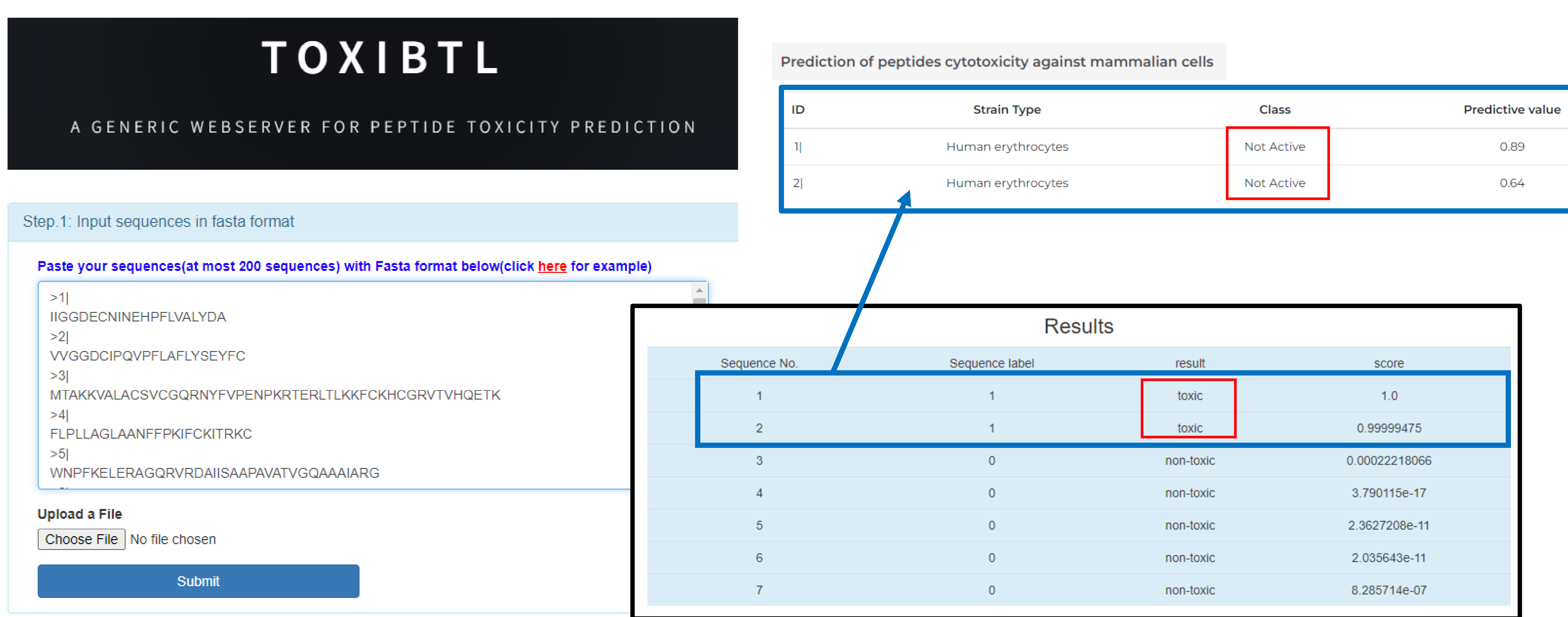


Úkol 3: Predikujte případné cytotoxické účinky peptidů vůči savčím buňkám.

https://dbaasp.org/tools?page=linear-amp-prediction

FLPLIGRVLSGIL
ALWKTLLKKVLKA

De novo design peptidů



ID	Strain Type	Class	Predictive value
1|	Human erythrocytes	Not Active	0.89
2|	Human erythrocytes	Not Active	0.64

Sequence No.	Sequence label	result	score
1	1	toxic	1.0
2	1	toxic	0.99999475
3	0	non-toxic	0.00022218066
4	0	non-toxic	3.790115e-17
5	0	non-toxic	2.3627208e-11
6	0	non-toxic	2.035643e-11
7	0	non-toxic	8.285714e-07

https://server.wei-group.net/ToxIBTL/Server.html

De novo design peptidů

Article

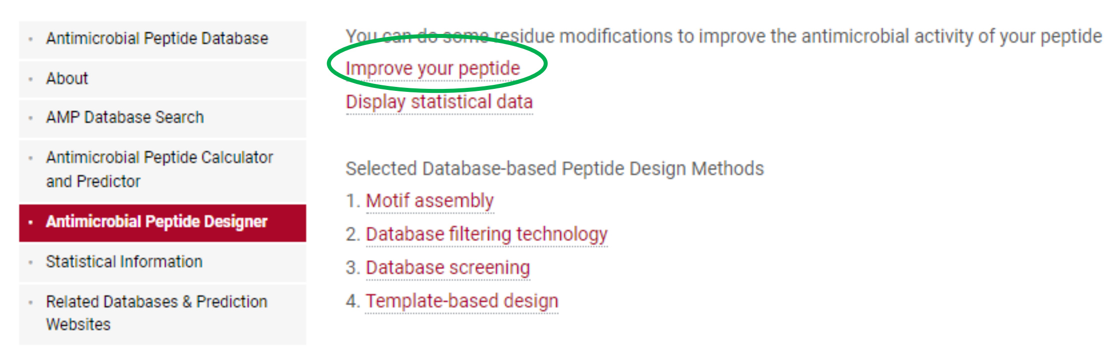
De Novo Design and In Vitro Testing of Antimicrobial Peptides against Gram-Negative Bacteria

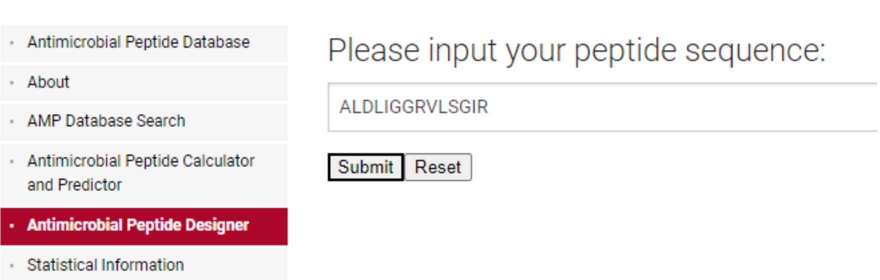
Boris Vishnepolsky [1,*], George Zaalishvili [2], Margarita Karapetian [2], Tornike Nasrashvili [2], Nato Kuljanishvili [2], Andrei Gabrielian [3], Alex Rosenthal [3], Darrell E. Hurt [3], Michael Tartakovsky [3], Maya Grigolava [1] and Malak Pirtskhalava [1,*]

- Problémy AMP: degradace proteasami, toxické pro savčí buňky, produkce je drahá.
- Design vhodných AMP: experimentální a výpočetní metody.
- Mnoho různých nástrojů pro predikci antimikrobiální aktivity, toxicity a stability AMP.

Úkol 4: Pokuste se „vylepšit" antimikrobiální aktivitu následujících peptidů.

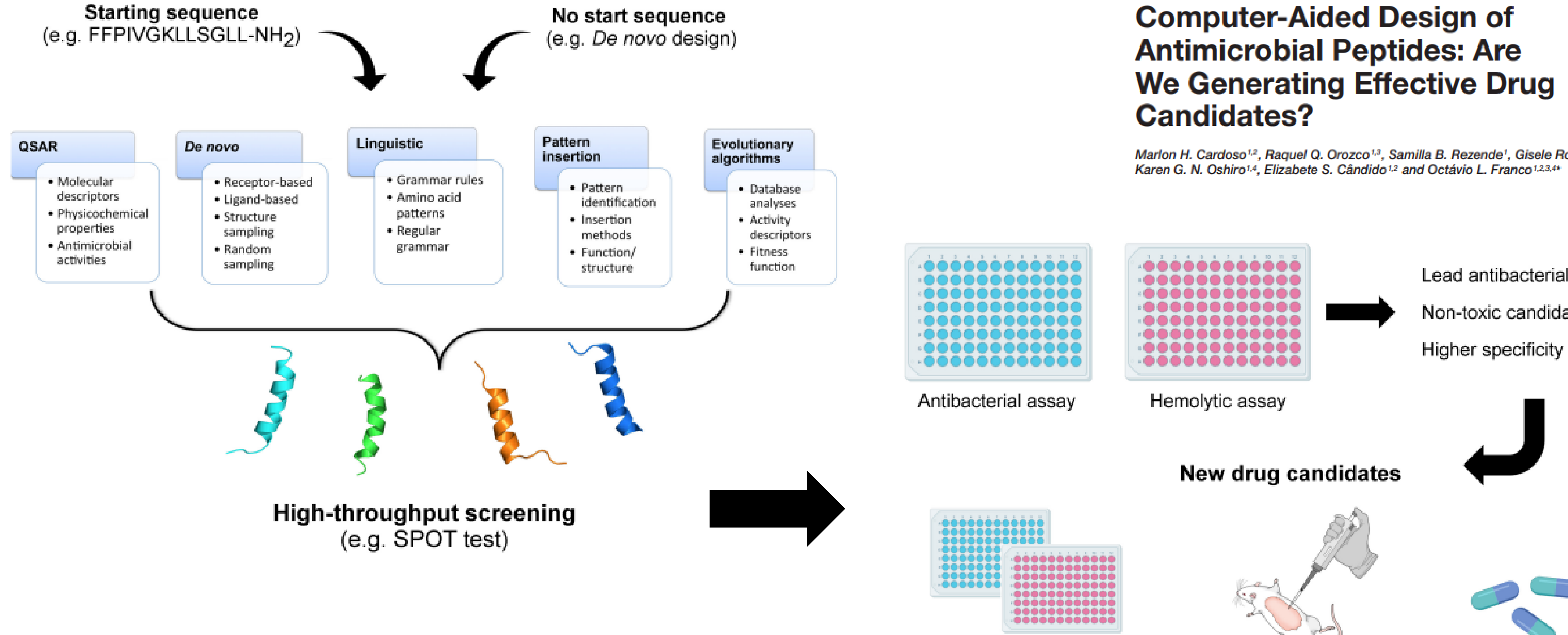
ALLLIGGLVLSGIL
FLDLIGGGVLSGIL





https://aps.unmc.edu/design

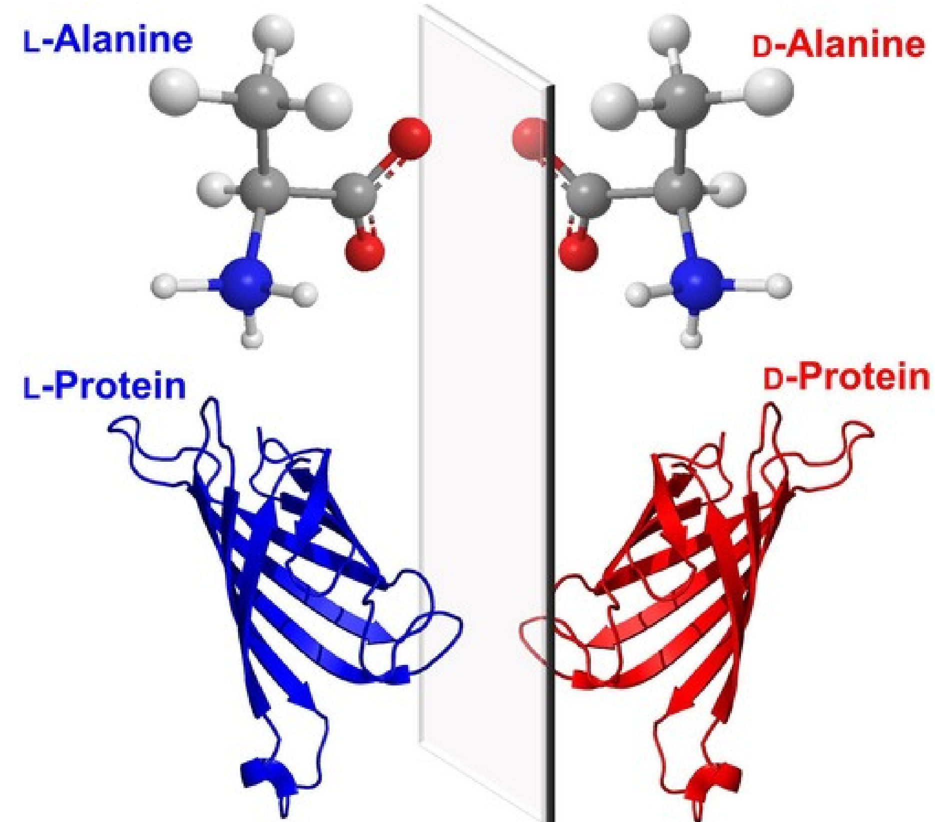
De novo design peptidů



Computer-Aided Design of Antimicrobial Peptides: Are We Generating Effective Drug Candidates?

Marlon H. Cardoso[1,2], Raquel Q. Orozco[1,3], Samilla B. Rezende[1], Gisele Rodrigues[2], Karen G. N. Oshiro[1,4], Elizabete S. Cândido[1,2] and Octávio L. Franco[1,2,3,4*]

D-proteiny



D-Peptide and D-Protein Technology: Recent Advances, Challenges, and Opportunities**

Alexander J. Lander,[a] Yi Jin,*[b] and Louis Y. P. Luk*[a]

„Identification of drug candidates through **mirror-image phage display** and related screening technologies“

4 Amplification & enrichment
3 Elute weak binders
1 Peptide library
Phage display
2 Binding to target
L-TARGET
D-TARGET
D-TARGET + L-binder
L-TARGET + D-binder

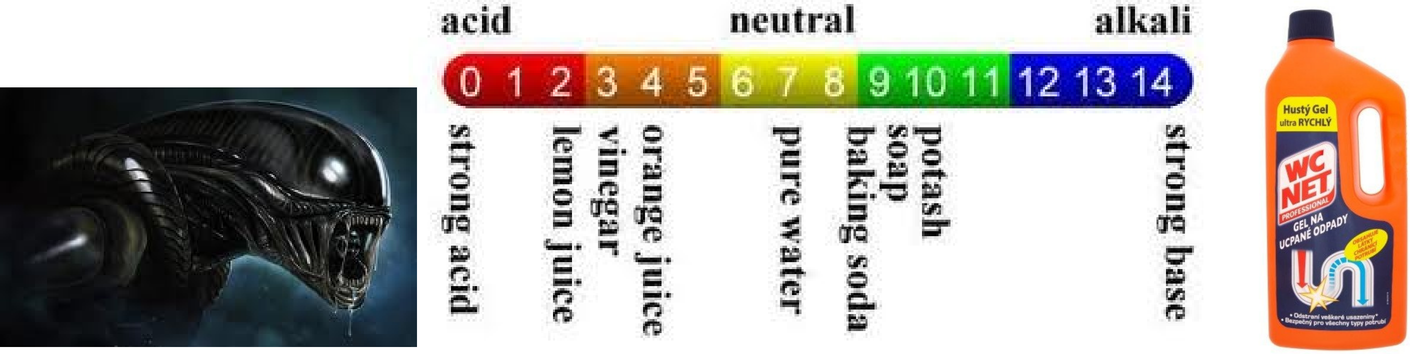
Phage display: vystavení peptidové sekvence na povrchu bakteriofágů, selekce peptidů na základě vazebných schopností

Jak se chovají proteiny (v laboratoři)?

- Limitované množství proteinu (cena, dostupnost).
- *In vitro* mohou rychle ztrácet aktivitu (nutná správná sekundární, terciární a někdy i kvarterní struktura).
- Mohou být nestabilní (některé velmi nestabilní) mimo své optimální prostředí v buňce (organismu).
- Při teplotě 95 °C dochází k úplné denaturaci téměř všech proteinů během několika minut. K výrazné destabilizaci a denaturaci může ale docházet již za laboratorní teploty (25 °C).
- Proteiny jsou štěpeny proteasami a peptidasami. Optimum těchto enzymů je 37 °C, za nižší teploty se jejich aktivita snižuje (ale jsou aktivní i při 4 °C). Proteasy se do vzorku dostanou neopatrnou manipulací, nedostatečnou purifikací a jsou také produkovány mikroorganismy.
- Kontaminace vzorků (bakterie, plísně).

Jak se chovají proteiny (v laboratoři)?

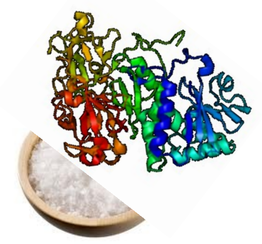
- Proteiny jsou aktivní (a stabilní) v určitém rozmezí pH.
  A to může být pro některé proteiny velmi úzké… Fyziologické pH pro většinu proteinů je cca 7,2-7,4. Silně kyselé nebo zásadité prostředí proteiny denaturuje.



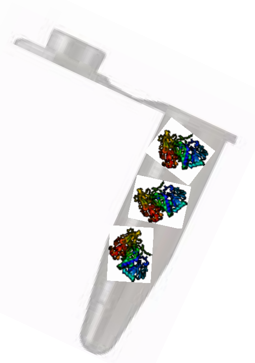
Pufr, tlumivý roztok, ústojný roztok, ústoj:
látka (směs látek) schopná udržovat stabilní pH po přídavku silné kyseliny nebo zásady do systému.
Příklad: slabá kyselina/její sůl, HA/A^-.
Pufry nesmí interagovat s proteiny nebo interferovat s jejich funkcí!

Jak se chovají proteiny (v laboratoři)?

- Proteiny vyžadují pro svou aktivitu (a stabilitu) určitou koncentraci „solí“. Vysoká i nízká koncentrace solí může způsobovat agregaci a precipitaci. Proteiny většinou nejsou stabilní v čisté vodě.



- Při práci s nízkými koncentracemi proteinů (< 1 mg/ml) se může výrazně projevit ztráta způsobená vazbou na stěny použité nádobky (zkumavky).
  Řešení: přídavek „inertního“ proteinu (BSA), použití vhodného plastu a skla.



Jak se chovají proteiny (v laboratoři)?

eppendorf

APPLICATION NOTE No. 382 I October 2016

Comparative Analysis of Protein Recovery Rates in Eppendorf LoBind® and Other "Low Binding" Tubes

Rafal Grzeskowiak[1], Sandrine Hamels[2], Eric Gancarek[2]
[1]Eppendorf AG, Hamburg, Germany; [2]Eppendorf Application Technologies SA, Namur, Belgium

Abstract

Protein preparation and storage poses a critical step in a wide range of laboratory applications. Unspecific adsorption of protein molecules or peptides to polymer surface of lab consumables has been shown to be a substantial factor contributing to sample loss during storage/handling and to influence experimental results. Binding of protein samples was investigated here by using a sensitive fluorescence assay, and recovery rates were compared between tubes of different manufacturers referred to as "low binding". The majority of tubes of different manufacturers tested showed very poor recovery rates (4 % - 12 %) after 24 h storage time and do not protect sufficiently against unspecific loss of protein samples. Eppendorf LoBind Tubes provided highest recovery rates of proteins (95 %) and thus ensure utmost protection of protein samples.

Certified Low Adsorption (LA) Vials and Kits for LC/MS

Supelco's new Low Adsorption (LA) vials are manufactured using a process that decreases the number of hydroxyl groups on the vial's glass surface, significantly reducing surface activity while improving analytical quantitation and minimizing pH shifts in the sample. This same process also removes unwanted surface metals such as sodium and boron that can contaminate samples and interfere with trace analysis. Unlike other methods used to decrease vial surface activity, the elimination of surface activity in LA vials is integral to the manufacturing process and is not a chemical surface treatment.

These vials and closures are designed, engineered, and manufactured for optimum performance in HPLC, GCMS, and LC/MS applications. The vials that are used become a critical part of the analysis of the sample. All of the parts of the vial system (vial, cap, and septum) can come into contact with the sample and are critical to the outcome of the analysis. The right choice for your vials is vital to ensure proper fit with the instrument.

We recommend using LA vials with MSQ polypropylene cap/PTFE silicone septa. This cap and septa combination was purposely designed for use with mass spectrometry. It shows little to no background contamination when compared to other cap and septa products in the marketplace.

Vial, 2 mL

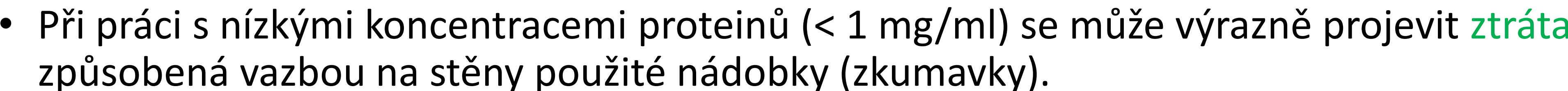
MRQ30 CD Vial, 0.2 µL

Center Drain (CD) Vial, 1.5 mL

Mikrozkumavky,
zkumavky,
špičky:
doba plastová
TUBES DE CENTRIFUGATION 50ml
Stérilisés par rayons gamma
TUBOS DE CENTRIFUGA 50ml
Esterilizado por radiación gamma
ZENTRIFUGE RÖHRCHEN 50ml
Gamma-Sterilisiert
TUBI PER CENTRIFUGA 50ml

Jak se chovají proteiny (v laboratoři)?

- Proteiny vyžadují pro svou aktivitu (a stabilitu) určitou koncentraci „solí“. Vysoká i nízká koncentrace solí může způsobovat agregaci a precipitaci. Proteiny většinou nejsou stabilní v čisté vodě.



- Při práci s nízkými koncentracemi proteinů (< 1 mg/ml) se může výrazně projevit ztráta způsobená vazbou na stěny použité nádobky (zkumavky).

- Proteiny mohou být rovněž poškozeny mechanicky při příliš energickém míchaní nebo třepání!

Vortex

Jak pracovat s proteiny?

- Práce za vhodné teploty
- Minimalizace kontaminace vzorku (ochranné pomůcky)
- Inhibitory proteas
- Přídavek antibakteriálních látek
- Kontrolované prostředí – pufry o vhodném pH
- Optimalizovaná koncentrace „solí“
- Přídavek inertních proteinů, speciální materiály se sníženou vazbou proteinů
- Opatrná manipulace!

Jak vybrat vhodné podmínky pro práci s proteinem?

- Informace o homologních proteinech (z literatury, od kolegů)
- Zkušenosti a/nebo metoda pokus-omyl
- Testování a výběr nejlepších podmínek – ideální je použití high-throughput metody, malá množství proteinu

Tabulka IV: 48 podmínek „buffer screenu“. Tabulka udává složení roztoků o různém pH (2,0–12,0) s různou koncentrací soli (0,5-5 M) a s přídavkem různých aditiv.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Voda	250 mM maleát, pH 2,0	500 mM glycin, pH 3,0	500 mM formiát, pH 4,0	500 mM citrát, pH 5,0	500 mM kakodylát, pH 6,0	500 mM Hepes, pH 7,0	500 mM bicin, pH 8,0	500 mM CHES, pH 9,0	65 mM borát, pH 10,0	500 mM CAPS, pH 11,0	500 mM fosfát, pH 12,0
	Voda	Vliv pH (2,0-12,0; nárůst = 1)										
B	500 mM acetát, pH 4,0	500 mM acetát, pH 4,5	500 mM acetát, pH 5,0	500 mM MES, pH 5,5	500 mM MES, pH 6,0	500 mM MES, pH 6,5	500 mM Na fosfát, pH 7,0	500 mM K fosfát, pH 7,5	500 mM Tris, pH 8,0	500 mM Tris, pH 8,5	500 mM glycin, pH 9,0	500 mM glycin, pH 9,5
	Vliv pH (4,0-9,5; nárůst = 0,5)											
C	500 mM MES, 500 mM NaCl, pH 6,0	500 mM MES, 1 M NaCl, pH 6,0	500 mM MES, 2,5 M NaCl, pH 6,0	500 mM MES, 5 M NaCl, pH 6,0	500 mM Na fosfát, 500 mM NaCl, pH 7,0	500 mM Na fosfát, 1 M NaCl, pH 7,0	500 mM Na fosfát, 2,5 M NaCl, pH 7,0	500 mM Na fosfát, 5 M NaCl, pH 7,0	500 mM Tris, 500 mM NaCl, pH 8,0	500 mM Tris, 1 M NaCl, pH 8,0	500 mM Tris, 2,5 M NaCl, pH 8,0	500 mM Tris, 5 M NaCl, pH 8,0
	Vliv soli - pH 6,0				Vliv soli - pH 7,0				Vliv soli - pH 8,0			
D	100 mM Tris, 750 mM NaCl, 500 uM CaCl2, pH 7,5	50 mM Hepes, 750 mM NaCl, 0,25% Tween 20, pH 7,5	60 mM fosfát, 685 mM NaCl, 13,5 mM KCl, pH 7,5	60 mM fosfát, 0,25% Tween 20, pH 7,5	1 M imidazol, pH 7,5	0,25% Tween 20	25% glycerol	25 mM bME	25% DMSO	25 mM trehalosa	100 mM arginin, 100 mM glutamin	25 mM EDTA
	Standardní pufry				Aditiva							

Vliv pH na vazebné vlastnosti lektinů, bakalářská práce, Barbora Mikšátková

Jak vybrat vhodné podmínky pro práci s proteinem?

- Informace o homologních proteinech (z literatury, od kolegů)
- Zkušenosti a/nebo metoda pokus-omyl
- Testování a výběr nejlepších podmínek – ideální je použití high-throughput metody, malá množství proteinu

Tabulka VIII: Teploty denaturace RSL (1. várka) ve všech podmínkách „buffer screenu" získané metodou nanoDSF. Podmínka s nejvíce stabilizujícím efektem a s nejvyšší teplotou denaturace (D7) má tmavě zelenou barvu. Barvy ostatních podmínek se odvíjí od srovnání s podmínkou D7 podle škály ve spodní části tabulky. Na příklad podmínka B4 má bledě modrou barvu, protože se od podmínky D7 liší o 7 °C.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Voda 94,3	25 mM maleát, pH 2,0 72,9	50 mM glycin, pH 3,0 84,2	50 mM formiát, pH 4,0 92,6	50 mM citrát, pH 5,0 90,7	50 mM kakodylát, pH 6,0 87,8	50 mM Hepes, pH 7,0 88,4	50 mM bicin, pH 8,0 84,1	50 mM CHES, pH 9,0 83,8	6,5 mM borát, pH 10,0 82,3	50 mM CAPS, pH 11,0 80,3	50 mM fosfát, pH 12,0 57,1
B	50 mM acetát, pH 4,0 92,7	50 mM acetát, pH 4,5 93,4	50 mM acetát, pH 5,0 92,8	50 mM MES, pH 5,5 90,2	50 mM MES, pH 6,0 92,4	50 mM MES, pH 6,5 89,5	50 mM Na fosfát, pH 7,0 83,6	50 mM K fosfát, pH 7,5 83,8	50 mM Tris, pH 8,0 86,4	50 mM Tris, pH 8,5 85,2	50 mM glycin, pH 9,0 83,9	50 mM glycin, pH 9,5 83,9
C	50 mM MES, 50 mM NaCl, pH 6,0 91,8	50 mM MES, 100 mM NaCl, pH 6,0 90,9	50 mM MES, 250 mM NaCl, pH 6,0 90,2	50 mM MES, 500 mM NaCl, pH 6,0 89,5	50 mM Na fosfát, 50 mM NaCl, pH 7,0 83,4	50 mM Na fosfát, 100 mM NaCl, pH 7,0 83,3	50 mM Na fosfát, 250 mM NaCl, pH 7,0 83,0	50 mM Na fosfát, 500 mM NaCl, pH 7,0 84,7	50 mM Tris, 50 mM NaCl, pH 8,0 85,0	50 mM Tris, 100 mM NaCl, pH 8,0 84,7	50 mM Tris, 250 mM NaCl, pH 8,0 84,1	50 mM Tris, 500 mM NaCl, pH 8,0 83,5
D	10 mM Tris, 75 mM NaCl, 50 uM CaCl2, pH 7,5 89,2	5 mM Hepes, 75 mM NaCl, 0,025% Tween 20, pH 7,5 86,0	6 mM fosfát, 68,5 mM NaCl, 1,35 mM KCl, pH 7,5 82,7	6 mM fosfát, 0,025% Tween 20, pH 7,5 84,7	100 mM imidazol, pH 7,5 86,5	0,025% Tween 20 95,4	2,5% glycerol 97,2	2,5 mM bME 94,4	2,5% DMSO 92,7	2,5 mM trehalosa 89,9	10 mM **arginin**, 10 mM **glutamin** 83,8	2,5 mM EDTA 84,3

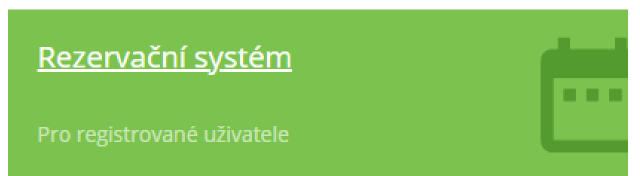
nejlepší stabilita	rozdíl do 2 °C	rozdíl 2 - 4 °C	rozdíl 4 - 6 °C	rozdíl 6 - 8 °C	rozdíl 8 - 10 °C	rozdíl 10 - 15 °C	rozdíl 15+ °C

Vliv pH na vazebné vlastnosti lektinů, bakalářská práce, Barbora Mikšátková

Jak vybrat vhodné podmínky pro práci s proteinem?

- **Informace o homologních proteinech (z literatury, od kolegů)**
- **Zkušenosti a/nebo metoda pokus-omyl**
- **Testování a výběr nejlepších podmínek – ideální je použití high-throughput metody, malá množství proteinu**

Sdílená laboratoř Interakce a krystalizace biomolekul



O nás:

Sdílená laboratoř CF BIC poskytuje služby umožňující strukturní a biofyzikální charakterizaci biomolekul a studium (bio)molekulárních interakcí. Přístrojové vybavení laboratoře umožňuje studium podmínek pro krystalizaci biomolekul a jejich komplexů, základní charakterizaci biofyzikálních vlastností molekul (analytická ultracentrifugace, dynamický rozptyl světla, CD spektroskopie, diferenciální skenovací kalorimetrie, diferenciální skenovací fluorimetrie) a studium termodynamiky a/nebo kinetiky vzájemných interakcí (isotermická titrační kalorimetrie, rezonance povrchového plasmonu, CD spektroskopie, analytická ultracentrifugace, microscale termoforéza).

Diferenciální skenovací fluorimetrie: Prometheus NT.48

Prometheus slouží pro charakterizaci teplotní nebo chemické denaturace proteinů a jejich agregace v jediném běhu v rozsahu až (15 - 110 °C). Během jednoho měření je možno analyzovat až 48 vzorků.

USER MODE

Eva Fujdiarová

specialistka na DSF, MST

http://bic.ceitec.cz/cs/

Fluorescence

Je fyzikální jev při kterém dochází vyzařování „světla“ látkou, která předtím pohltila elektromagnetické záření

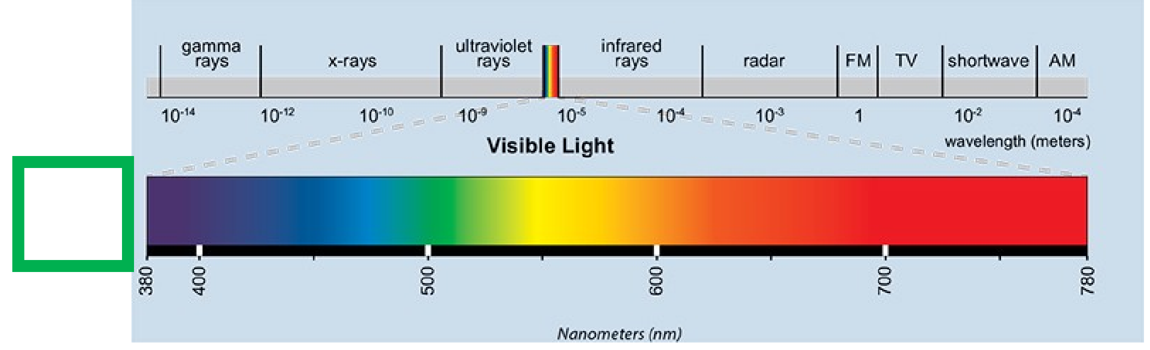
excitační záření

fluorofor

emitované záření

Opticky zjasňující prostředky

Vnitřní fluorescence proteinů (aromatických aminokyselin) v UV oblasti (λ = 300-360 nm)



Tryptofan (Trp, W)

Tyrosin (Tyr, Y)

Fenylalanin (Phe, F)

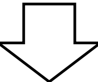
nanoDSF

Vnitřní fluorescence proteinů (aromatických aminokyselin) v UV oblasti (λ = 300-360 nm)

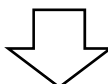
Mění se v závislosti na okolí aromatických aminokyselin

W, Y, F jsou hydrofobní a typicky se nacházejí uvnitř proteinů

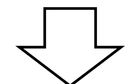
Se zvyšující se teplotou dochází k denaturaci proteinu



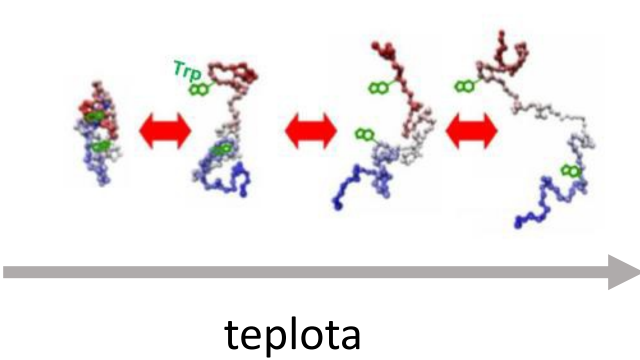
rozvolnění struktury

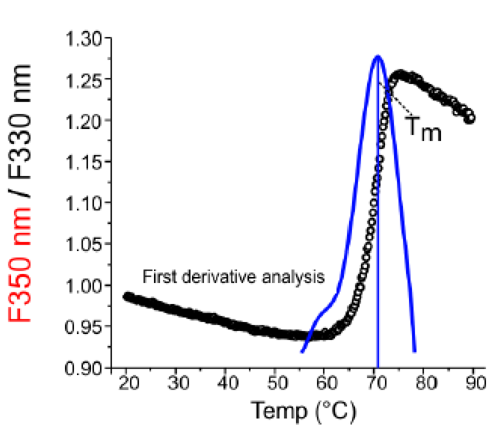


expozici W, Y, F na povrch



změny ve fluorescenci

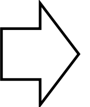




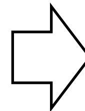
www.nanotempertech.com

nanoDSF v praxi

Protein v testovaném pufru (testovaných pufrech)



Nasaju do kapiláry

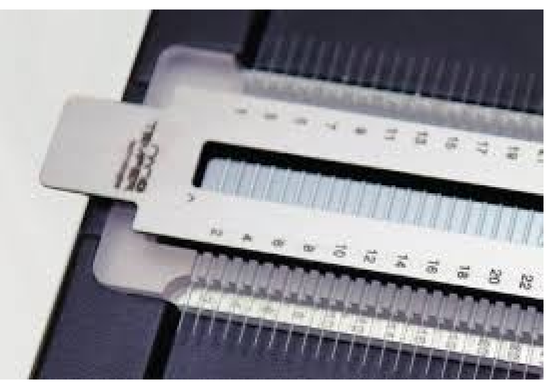


48 kapilár na 1 měření

www.nanotempertech.com

nanoDSF v praxi

Prometheus



www.nanotempertech.com

Uchovávání proteinů

	Storage Condition			
Characteristic	Solution at 4°C	Solution in 25-50% glycerol or ethylene glycol at -20°C	Frozen at -20° to -80°C or in liquid nitrogen	Lyophilized (usually also frozen)
Typical shelf life	1 month	1 year	Years	Years
Requires sterile conditions or addition of antibacterial agent	Yes	Usually	No	No
Number of times a sample may be removed for use	Many	Many	Once; repeated freeze-thaw cycles generally degrade proteins	Once; it is impractical to lyophilize a sample multiple times

- **Lyofilizace – mrazová sublimace.**
- **Odstranění vody ze zmraženého vzorku za sníženého tlaku.**
- **Nedostatek vody zabraňuje růstu mikroorganismů a inhibuje enzymy (proteasy).**
- **Nepoškozuje vzorek v takovém rozsahu jako jiné způsoby dehydratace (vysoká teplota, vysoušedla).**
- **Jednoduchá rehydratace.**

www.piercenet.com

- Lyofilizace – mrazová sublimace.
- Odstranění vody ze zmraženého vzorku za sníženého tlaku.
- Nedostatek vody zabraňuje růstu mikroorganismů a inhibuje enzymy (proteasy).
- Nepoškozuje vzorek v takovém rozsahu jako jiné způsoby dehydratace (vysoká teplota, vysoušedla).
- Jednoduchá rehydratace.

- Lyofilizace – mrazová sublimace.
- Odstranění vody ze zmraženého vzorku za sníženého tlaku.
- Nedostatek vody zabraňuje růstu mikroorganismů a inhibuje enzymy (proteasy).
- Nepoškozuje vzorek v takovém rozsahu jako jiné způsoby dehydratace (vysoká teplota, vysoušedla).
- Jednoduchá rehydratace.

Uchovávání proteinů

Proteiny mohou být lyofilizací nebo zamražením nevratně poškozeny!

Kryoprotektanty zabraňují tvorbě krystalků ledu a poškození proteinu.

	Storage Condition			
Characteristic	Solution at 4°C	Solution in 25-50% glycerol or ethylene glycol at -20°C	Frozen at -20° to -80°C or in liquid nitrogen	Lyophilized (usually also frozen)
Typical shelf life	1 month	1 year	Years	Years
Requires sterile conditions or addition of antibacterial agent	Yes	Usually	No	No
Number of times a sample may be removed for use	Many	Many	Once; repeated freeze-thaw cycles generally degrade proteins	Once; it is impractical to lyophilize a sample multiple times

Sterilní zkumavky, sterilizace filtrací. Inhibitory proteas.

Nutné připravit několik alikvotů (částí zásobního roztoku) proteinu.

www.piercenet.com

Práce s proteiny

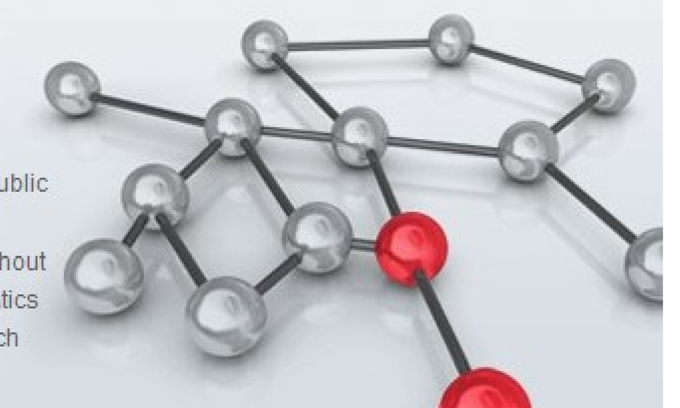
Expert Protein Analysis System

http://www.expasy.org

ExPASy is the **SIB Bioinformatics Resource Portal** which provides access to scientific databases and software tools (i.e., *resources*) in different areas of life sciences including proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc. (see Categories in the left menu). On this portal you find resources from many different SIB groups as well as external institutions.

https://www.sib.swiss/

The SIB Swiss Institute of Bioinformatics is an academic, non-profit foundation recognised of public utility and established in 1998. SIB coordinates research and education in bioinformatics throughout Switzerland and provides high quality bioinformatics services to the national and international research community.



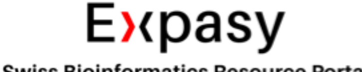
Expert Protein Analysis System

http://www.expasy.org

About Expasy

Expasy is the bioinformatics resource portal of the SIB Swiss Institute of Bioinformatics (more about its history).

It is an extensible and integrative portal which provides access to over 160 databases and software tools, developed by SIB Groups and supporting a range of life science and clinical research domains, from genomics, proteomics and structural biology, to evolution and phylogeny, systems biology and medical chemistry.



Swiss Bioinformatics Resource Portal

e.g. BLAST, UniProt, MSH6, Albumin...

- Genes & Genomes
  - Genomics
  - Metagenomics
  - Transcriptomics
- Proteins & Proteomes
- Evolution & Phylogeny
  - Evolution biology
  - Population genetics
- Structural Biology
  - Drug design
  - Medicinal chemistry
  - Structural analysis
- Systems Biology
  - Glycomics
  - Lipidomics
  - Metabolomics
- Text mining & Machine learning

SIB Resources

UniProtKB/Swiss-Prot
Protein knowledgebase

neXtProt
Human protein knowledgebase

SwissRegulon Portal
Tools and data for regulatory genomics

EPD
Eukaryotic Promoter Database

SwissDrugDesign
Widening access to computer-aided drug design

V-pipe
Viral genomics pipeline

SwissOrthology
One-stop shop for orthologs

SwissLipids
Knowledge resource for lipids

STRING
Protein-protein interaction networks and enrichment analysis

SWISS-MODEL
Protein structure homology-modelling

Bgee
Gene expression expertise

Other Resources of SIB Groups

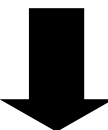
Swiss Mass Abacus
Intuitive calculator of peptide and glycopeptide masses

SwissADME
Pharmacokinetics properties and druglikeness

miROrtho
Catalogue of animal microRNA genes

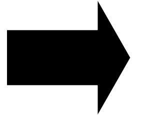
STRING-covid
New COVID-19 oriented version of the STRING database

CCTTTATTATCCGCTTCCATTGTTTCCGCTCCTGTTGTTACTTCCGAAACTTATGTTGATATTCCTGGTTTATATTTAGA
TGTTGCTAAAGCTGGTATTCGCGATGGTAAATTACAAGTTATTTTAAATGTTCCTACTCCTTATGCTACTGGTAATAATT
TTCCTGGTATTTATTTTGCTATTGCTACTAATCAAGGTGTTGTTGCTGATGGTTGTTTTACTTATTCCTCCAAAGTTCCT
GAATCCACTGGTCGCATGCCTTTTACTTTAGTTGCTACTATTGATGTTGGTTCCGGTGTTACTTTTGTTAAAGGTCAATG
GAAATCCGTTCGCGGTTCCGCTATGCATATTGATTCCTATGCTTCCTTATCCGCTATTTGGGGTACTGCTGCTCCTTCCT
CCCAAGGTTCCGGTAATCAAGGTGCTGAAACTGGTGGTACTGGTGCTGGTAATATTGGTGGTGGTGGTGAACGCGATGGT
ACTTTTAATTTACCTCCTCATATTAAATTTGGTGTTACTGCTTTAACTCATGCTGCTAATGATCAAACTATTGATATTTA
TATTGATGATGATCCTAAACCTGCTGCTACTTTTAAAGGTGCTGGTGCTCAAGATCAAAATTTAGGTACTAAAGTTTTAG
ATTCCGGTAATGGTCGCGTTCGCGTTATTGTTATGGCTAATGGTCGCCCTTCCCGCTTAGGTTCCCGCCAAGTTGATATT
TTTAAAAAATCCTATTTTGGTATTATTGGTTCCGAAGATGGTGCTGATGATGATTATAATGATGGTATTGTTTTTTTAAA



**PLLSASIVSAPVVTSETYVDIPGLYLDVAKAGIRDGKLQVILNVPTPYATGNNFPGIYFA
IATNQGVVADGCFTYSSKVPESTGRMPFTLVATIDVGSGVTFVKGQWKSVRGSAMHIDSY
ASLSAIWGTAAPSSQGSGNQGAETGGTGAGNIGGGERDGTFNLPPHIKFGVTALTHAAN
DQTIDIYIDDDPKPAATFKGAGAQDQNLGTKVLDSGNGRVRVIVMANGRPSRLGSRQVDI
FKKSYFGIIGSEDGADDDYNDGIVFL**

Nukleotidová a proteinová
sekvence hypotetických genů/proteinů



Predikce vlastností

In silico functional prediction of hypothetical proteins from the core genome of *Corynebacterium pseudotuberculosis* biovar *ovis*

Carlos Leonardo Araújo[1,*], Iago Blanco[1,*], Luciana Souza[1], Sandeep Tiwari[2], Lino César Pereira[1], Preetam Ghosh[3], Vasco Azevedo[2], Artur Silva[1] and Adriana Folador[1]

[1] Laboratory of Genomics and Bioinformatics, Center of Genomics and Systems Biology, Institute of Biological Sciences, Federal University of Pará, Belém, Pará, Brazil
[2] Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
[3] Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA
[*] These authors contributed equally to this work.

Příklad 1

ABSTRACT

Corynebacterium pseudotuberculosis is a pathogen of veterinary relevance diseases, being divided into two biovars: *equi* and *ovis*; causing ulcerative lymphangitis and caseous lymphadenitis, respectively. The isolation and sequencing of *C. pseudotuberculosis* biovar *ovis* strains in the Northern and Northeastern regions of Brazil exhibited the emergence of this pathogen, which causes economic losses to small ruminant producers, and condemnation of carcasses and skins of animals. Through the pan-genomic approach, it is possible to determine and analyze genes that are shared by all strains of a species—the core genome. However, many of these genes do not have any predicted function, being characterized as hypothetical proteins (HP). In this study, we considered 32 *C. pseudotuberculosis* biovar *ovis* genomes for the pan-genomic analysis, where were identified 172 HP present in a core genome composed by 1255 genes. We are able to functionally annotate 80 sequences previously characterized as HP through the identification of structural features as conserved domains and families. Furthermore, we analyzed the physicochemical properties, subcellular localization and molecular function. Additionally, through RNA-seq data, we investigated the differential gene expression of the annotated HP. Genes inserted in pathogenicity islands had their virulence potential evaluated. Also, we have analyzed the existence of functional associations for their products based on protein–protein interaction networks, and perform the structural prediction of three targets. Due to the integration of different strategies, this study can underlie deeper in vitro researches in the characterization of these HP and the search for new solutions for combat this pathogen.

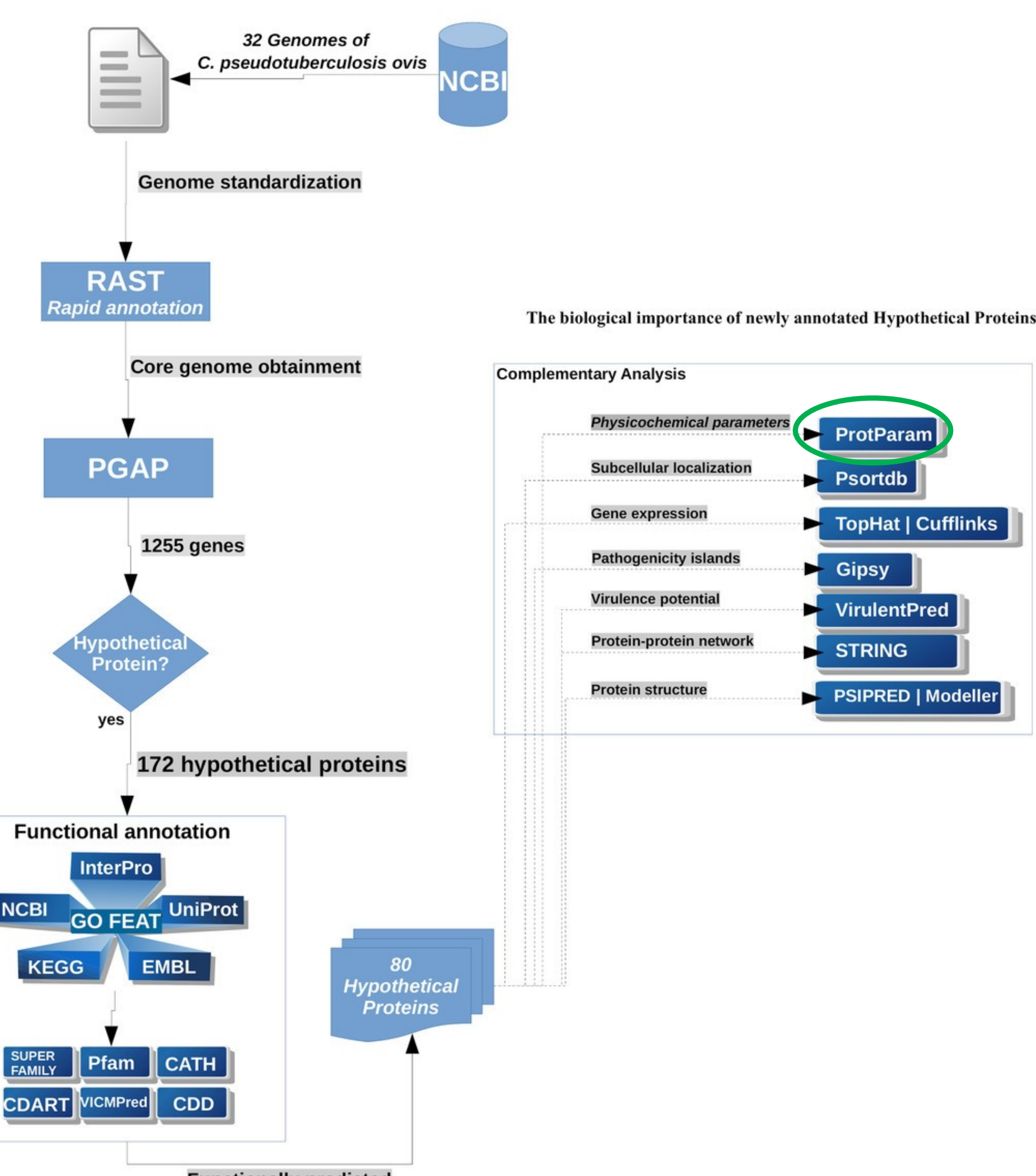


Table S2: Theoretical prediction of Physicochemical Parameters, Subcellular Localization and Protein Function for 172 HP

UNIPROT ID	Nº of AA	MW	IP	EC	II	-	AI	GRAVY	Subcellular Localization	Protein Function
A0A1L6CXU2	144	15424.71	5.43	8480	28.09	Stable	94.93	-0.013	Cytoplasm	Cellular Process
D9QD65	114	11830.05	4.95	15470	15.91	Stable	145.35	1 175	Cyt. Membrane	Cellular Process
D9QD66	259	28688.72	10.85	87805	38.93	Stable	132.47	0.874	Cyt. Membrane	Cellular Process
A0A1L3KSD4	189	19817.07	4.27	22460	26.68	Stable	78.47	-0.279	Cyt. Membrane	Cellular Process
D9QD72	223	24778.46	8.51	13410	50.79	Unstable	65.74	-0.835	Cyt. Membrane	Metabolism
D9QD92	317	34853.13	4.28	36900	54.75	Unstable	75.65	-0.51	Cytoplasm	Cellular Process
D9QD94	176	19506.58	8.74	30940	46.27	Unstable	94.83	0.212	Cyt. Membrane	Cellular Process
D9QDI6	84	9267.63	8.66	7115	42.76	Unstable	75.48	-0.308	Cytoplasm	Cellular Process
D9QDJ9	556	61675.71	4.76	96870	35.15	Stable	80.45	-0.393	Cytoplasm	Virulence
A0A0E3N3X9	106	11231.76	4.71	4470	12.58	Stable	94.62	0.182	Cytoplasm	Cellular Process
D9QDM2	180	19919.07	5.69	19605	36.73	Stable	100.17	0.014	Cyt. Membrane	Cellular Process
D9QDM3	300	32616.31	7.62	86775	36.00	Stable	107.90	0.666	Cyt. Membrane	Cellular Process
D9QDM9	74	8634.95	4.58	2980	60.19	Unstable	81.76	-0.446	Cytoplasm	Cellular Process
D9QDN0	153	16678.22	6.28	13980	22.29	Stable	113.40	0.116	Cyt. Membrane	Metabolism
D9QDN1	176	19175.82	4.88	31970	41.84	Unstable	89.32	-0.18	Cytoplasm	Metabolism
D9QDN2	295	32513.52	9.00	42315	48.82	Unstable	100.95	-0.107	Cytoplasm	Metabolism
D9QDS0	189	20508.19	4.21	23950	34.34	Stable	98.47	0.074	Cytoplasm	Cellular Process
D9QDT9	518	55588.43	4.93	61310	34.02	Stable	78.90	-0.323	Cell Wall	Metabolism
D9QDU3	281	29782.02	8.92	57410	38.51	Stable	119.18	0.746	Cyt. Membrane	Metabolism
D9QDV0	253	29432.13	4.87	44015	39.78	Stable	78.93	-0.383	Cytoplasm	Information
D9QDV3	207	22233.93	4.19	48595	43.96	Unstable	92.90	0.149	Cytoplasm	Cellular Process
D9QDW7	477	48187.87	5.81	44920	17.05	Stable	102.70	0.237	Cyt. Membrane	Information
D9QDY6	166	17814.82	8.80	13980	19.86	Stable	103.49	0.388	Cyt. Membrane	Cellular Process
D9QDZ4	64	6543.50	4.86	2980	30.64	Stable	99.37	0.516	Extracellular	Cellular Process
D9QE21	96	10681.79	9.45	1547	13.01	Stable	137.29	0.958	Cyt. Membrane	Metabolism
D9QE28	167	17712.91	4.36	15470	26.72	Stable	92.87	0.006	Cytoplasm	Metabolism
D9QE51	193	21446.27	9.87	33585	42.29	Unstable	111.71	0.338	Cyt. Membrane	Metabolism
D9QE89	317	33259.96	6.44	42065	36.21	Stable	71.07	-0.252	Cell Wall	Metabolism
D9QE91	344	37037.72	5.40	50990	37.97	Stable	80.00	-0.179	Cyt. Membrane	Metabolism
D9QE95	313	34589.40	5.63	39420	26.41	Stable	99.33	-0.093	Cyt. Membrane	Cellular Process
D9QEA2	241	26061.98	5.64	41160	41.84	Unstable	98.88	0.211	Cyt. Membrane	Metabolism
A0A1L6CSV7	218	24583.59	6.23	39085	34.78	Stable	103.72	-0.028	Cytoplasm	Metabolism
D9QEB2	181	20396.52	5.13	28990	55.88	Unstable	122.87	0.262	Cyt. Membrane	Metabolism
D9QEB3	183	20040.61	10.53	26470	36.22	Stable	113.61	0.233	Cyt. Membrane	Cellular Process
D9QEB4	153	16571.14	11.03	12490	55.95	Unstable	93.01	-0.087	Cytoplasm	Cellular Process
D9QEB5	68	7170.63	9.62	1490	30.69	Stable	146.32	0.946	Cyt. Membrane	Metabolism
D9QEB6	112	11812.43	4.74	7450	33.37	Stable	112.32	0.185	Cytoplasm	Cellular Process

Predikce základních fyzikálně-chemických vlastností proteinů

Základní predikované vlastnosti proteinů

- Počet aminokyselin
- Pozitivně (Arg + Lys)/záporně (Asp + Glu) nabitá rezidua
- Molekulová hmotnost
- Izoelektrický bod
- Extinkční koeficient
- „Aliphatic index“
- „Instability index“
- GRAVY (Grand Average of Hydropathy) index
- Poločas života

- Agregace proteinů
- Termostabilita – teplotní stabilita
- Posttranslační modifikace

- Peptidy: hydrofobicita, hydrofobní moment, „angle subtended by charged residues“

ProtParam References

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.;
Protein Identification and Analysis Tools on the ExPASy Server;
(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005).
pp. 571-607
Full text - Copyright Humana Press.

https://web.expasy.org/protparam/

Příklad: ProtParam

- Predikce/výpočet základních fyzikálně-chemických parametrů proteinu.
- Vychází pouze z **aminokyselinové** sekvence proteinu.
- ProtParam nebere v úvahu možné **posttranslační modifikace (PTM)** a **oligomerizaci** proteinů.
- Pro predikci PTM a oligomerizace existují specializované nástroje.
- Problematika PTM není stále dořešená, především u prokaryot.

Note: It is not possible to specify post-translational modification for your protein, nor will ProtParam know whether your mature protein forms dimers or multimers. If you do know that your protein forms a dimer, you may just duplicate your sequence (i.e. append a second copy of the sequence to the first), as all computations performed by ProtParam are based on either compositional data, or on the N-terminal amino acid.

ProtParam References

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.;
Protein Identification and Analysis Tools on the ExPASy Server;
(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005).
pp. 571-607
Full text - Copyright Humana Press.

https://web.expasy.org/protparam/

Příklad: ProtParam

ProtParam tool

ProtParam (References / Documentation) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Disclaimer).

Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) or a sequence identifier (ID) (for example **KPC1_DROME**):

Or you can paste your own amino acid sequence (in one-letter code) in the box below:

```
KITCQGDKYSLPEVAAGQQTFKLMQDVRRLQIPSTVVEPQ
DTLFSDCSPTGNCPKCRIPIRCPYKTERCDVKFHYFYIMP
MTQIALDGGYCLNHNQDIICARHKFENKFANKSRDMNHMM
MKLLIYQPEVAVHHDEELASGLQPSVARRMVTTSMVHLRD
HKYGIMNDNTNFEMNCFFYGGHEGTFAHYECGPYFSAQIS
```

RESET | Compute parameters

Úkol 5: Určete základní fyzikálně-chemické parametry tohoto proteinu

Number of amino acids: 145

Molecular weight: 15531.97

Theoretical pI: 8.65

Amino acid composition: CSV format

Amino acid	Count	%
Ala (A)	8	5.5%
Arg (R)	12	8.3%
Asn (N)	4	2.8%
Asp (D)	7	4.8%
Cys (C)	12	8.3%
Gln (Q)	5	3.4%
Glu (E)	4	2.8%
Gly (G)	8	5.5%
His (H)	1	0.7%
Ile (I)	5	3.4%
Leu (L)	12	8.3%
Lys (K)	4	2.8%
Met (M)	1	0.7%
Phe (F)	2	1.4%
Pro (P)	22	15.2%
Ser (S)	13	9.0%
Thr (T)	10	6.9%
Trp (W)	0	0.0%
Tyr (Y)	3	2.1%
Val (V)	12	8.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 11
Total number of positively charged residues (Arg + Lys): 16

Atomic composition:

Atom	Symbol	Count
Carbon	C	668
Hydrogen	H	1090
Nitrogen	N	196
Oxygen	O	203
Sulfur	S	13

Formula: $C_{668}H_{1090}N_{196}O_{203}S_{13}$
Total number of atoms: 2170

Extinction coefficients:

This protein does not contain any Trp residues. Experience shows that this could result in more than 10% error in the computed extinction coefficient.

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 5220
Abs 0.1% (=1 g/l) 0.336, assuming all pairs of Cys residues form cystines

Ext. coefficient 4470
Abs 0.1% (=1 g/l) 0.288, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is S (Ser).

The estimated half-life is: 1.9 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 64.41
This classifies the protein as unstable.

Aliphatic index: 75.24

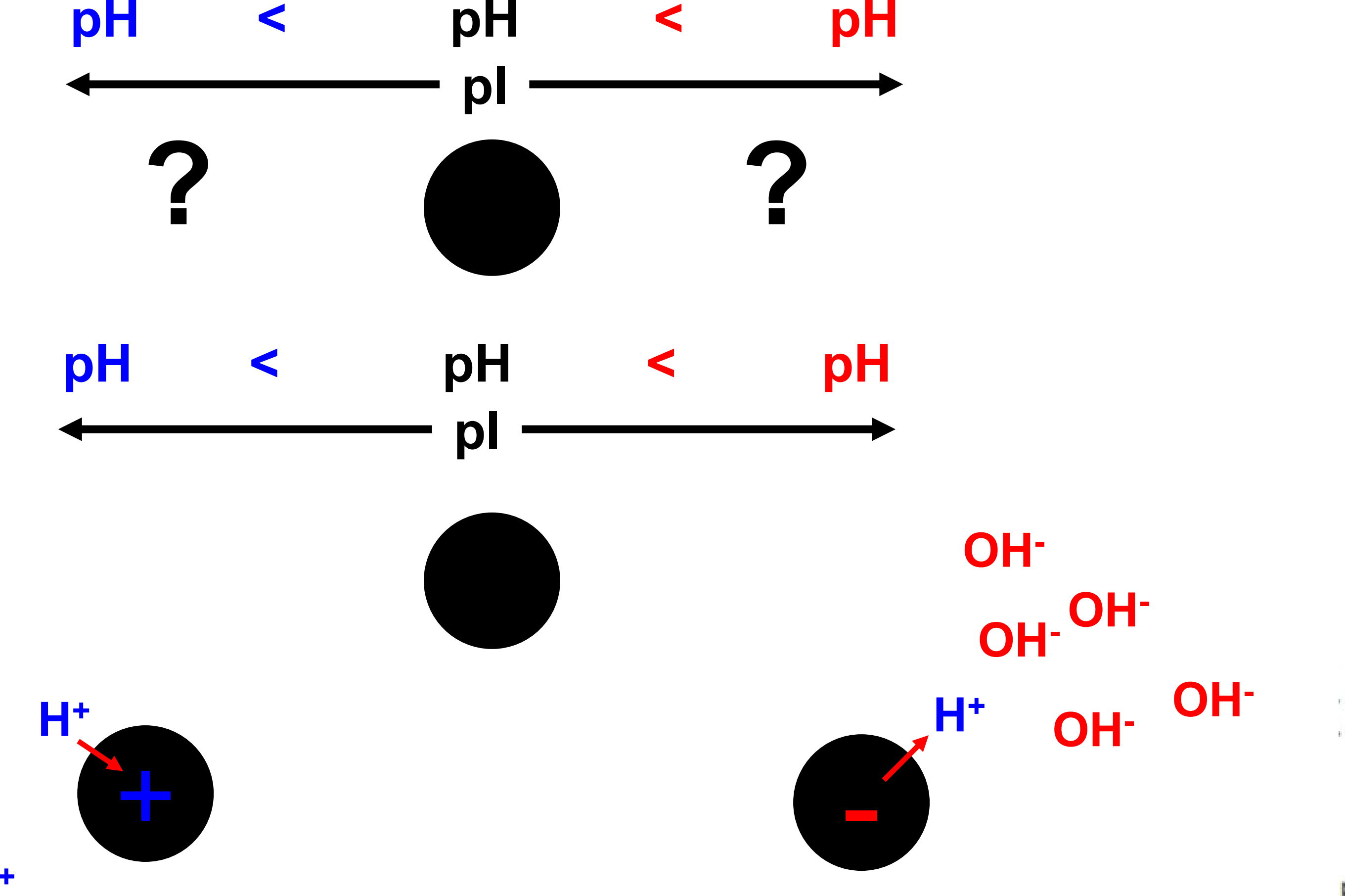
Grand average of hydropathicity (GRAVY): -0.221

Izoelektrický bod - pI

- **Izoelektrický bod = pH, při kterém má protein nulový sumární náboj, rozpustnost proteinů je při pH = pI nejmenší!**

4.6 Isoelectric Point Precipitation

The isoelectric point (pI) is the pH of a solution at which the net charge of a protein becomes zero. At solution pH that is above the pI, the surface of the protein is predominantly negatively charged, and therefore like-charged molecules will exhibit repulsive forces. Likewise, at a solution pH that is below the pI, the surface of the protein is predominantly positively charged, and repulsion between proteins occurs. However, at the pI, the negative and positive charges are balanced, reducing repulsive electrostatic forces, and the attraction forces predominate, causing aggregation and precipitation. The pI of most proteins is in the pH range of 4 to 7. Mineral acids, such as hydrochloric and sulfuric acids, are used as precipitants. The greatest disadvantage of isoelectric point precipitation is the irreversible denaturation caused by the mineral acids. For this reason isoelectric point precipitation is most often used to precipitate contaminant proteins rather than the target protein [4].



pH=2

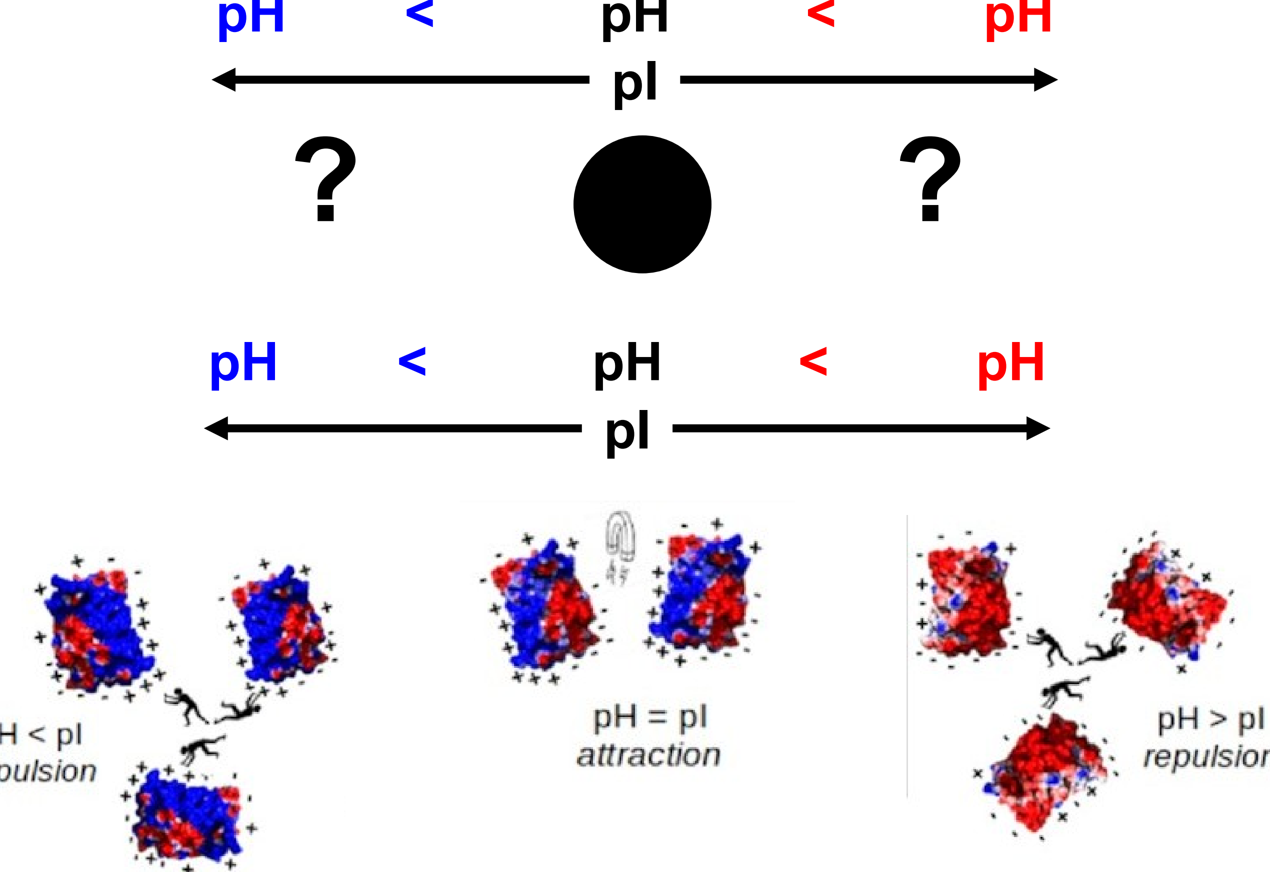
protein protonated positively charged

pH=10

protein deprotonated negatively charged

Izoelektrický bod - pI

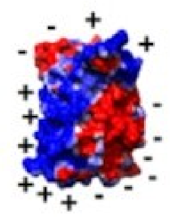
- **Izoelektrický bod = pH, při kterém má protein nulový sumární náboj, rozpustnost proteinů je při pH = pI nejmenší!**



4.6 Isoelectric Point Precipitation

The isoelectric point (pI) is the pH of a solution at which the net charge of a protein becomes zero. At solution pH that is above the pI, the surface of the protein is predominantly negatively charged, and therefore like-charged molecules will exhibit repulsive forces. Likewise, at a solution pH that is below the pI, the surface of the protein is predominantly positively charged, and repulsion between proteins occurs. However, at the pI, the negative and positive charges are balanced, reducing repulsive electrostatic forces, and the attraction forces predominate, causing aggregation and precipitation. The pI of most proteins is in the pH range of 4 to 7. Mineral acids, such as hydrochloric and sulfuric acids, are used as precipitants. The greatest disadvantage of isoelectric point precipitation is the irreversible denaturation caused by the mineral acids. For this reason isoelectric point precipitation is most often used to precipitate contaminant proteins rather than the target protein [4].

Izoelektrický bod - pI



- **Izoelektrický bod = pH, při kterém má protein nulový sumární náboj**

- pI proteinu – výpočet na základě znalostí acidobazických vlastností aminokyselin - pK (D,E,Y,C,H,K,R)
- Problémem jsou posttranslační modifikace!!!
- Použité hodnoty pK jednotlivých aminokyselin – různí autoři, různé podmínky, různé hodnoty…
- Vliv prostředí – sousední aminokyseliny, interakce (struktura proteinu)

Comments

1. Protein pI is calculated using pK values of amino acids described in Bjellqvist et al., which were defined by examining polypeptide migration between pH 4.5 to 7.3 in an immobilised pH gradient gel environment with 9.2M and 9.8M urea at 15°C or 25°C. Prediction of protein pI for highly basic proteins is yet to be studied and it is possible that current Compute pI/Mw predictions may not be adequate for this purpose.

Bengt Bjellqvist
Graham J. Hughes
Christian Pasquali
Nicole Paquet
Florence Ravier
Jean-Charles Sanchez
Séverine Frutiger
Denis Hochstrasser

Departments of Medicine and Biochemistry, Medical Center of the University of Geneva

The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences

The focusing positions in narrow range immobilized pH gradients of 29 polypeptides of known amino acid sequence were determined under denaturing conditions. The isoelectric points of the proteins calculated from their amino acid sequences matched with good accuracy the experimentally determined p*I* values. We show the advantages of being able to predict the position of a protein of known structure within a two-dimensional gel.

https://web.expasy.org/protparam/

Table 5. Comparison of experimental and calculated p*I*-values

Peptide	Experimental p*I*	Calculated p*I*
Apolipoprotein C-III-1	4.51	4.55
Apolipoprotein C-II	4.58	4.58
Apolipoprotein A-II	4.71	4.83
Haptoglobin α-1f	5.76	5.70
Haptoglobin α-1s	5.40	5.27
Apolipoprotein A-IV	5.13	5.20
α-Microglobulin	5.11	5.00
Haptoglobin α-2	5.69	5.69
Transthyretin	5.52	5.39
Serum albumin	5.87	5.81
4-Sialotransferrin	6.49	6.51
Apolipoprotein E	5.48	5.54
Retinol binding protein	5.28	5.30
β-Actin	5.26	5.26
Proapolipoprotein A-I	5.49	5.51
Complement factor C-3	4.80	4.80
C-Reactive protein	5.17	5.13
Serum amyloid protein A	6.17	6.08
β-Lactoglobulin A, bovine	4.82	4.77
Carbonic anhydrase I	6.63	6.69
Somatotropin	5.36	5.31
Thioredoxin	4.89	4.82
Elongation factor IB	4.53	4.50
Translationally controlled tumor protein	4.86	4.85
Heat shock protein 60	5.27	5.25
Cathepsin 81	5.22	5.25
ATP-synthase, coupling factor	5.31	5.46
Acyl-CoA dehydrogenase	6.35	6.34
β-2 Microglobulin	6.35	6.29

Izoelektrický bod - pI

Protein 1:

PTEFLYTSKIAAISWAATGGRQQRVYFQDLNGKIREAQRGGDNPWTGGSSQNVIGEAKLFSPLAAVTWKSAQGIQIRVYCVNKDNILSEFVYDGSK
WITGQLGSVGVKVGSNSKLAALQWGGSESAPPNIRVYYQKSNGSGSSIHEYVWSGKWTAGASFGSTVPGTGIGATAIGPGRLRIYYQATDNKIREH
CWDSNSWYVGGFSASASAGVSIAAISWGSTPNIRVYWQKGREELYEAAYGGSWNTPGQIKDASRPTPSLPDTFIAANSSGNIDISVFFQASGVSLQ
QWQWISGKGWSIGAVVPTGTPAGW

Protein 2:

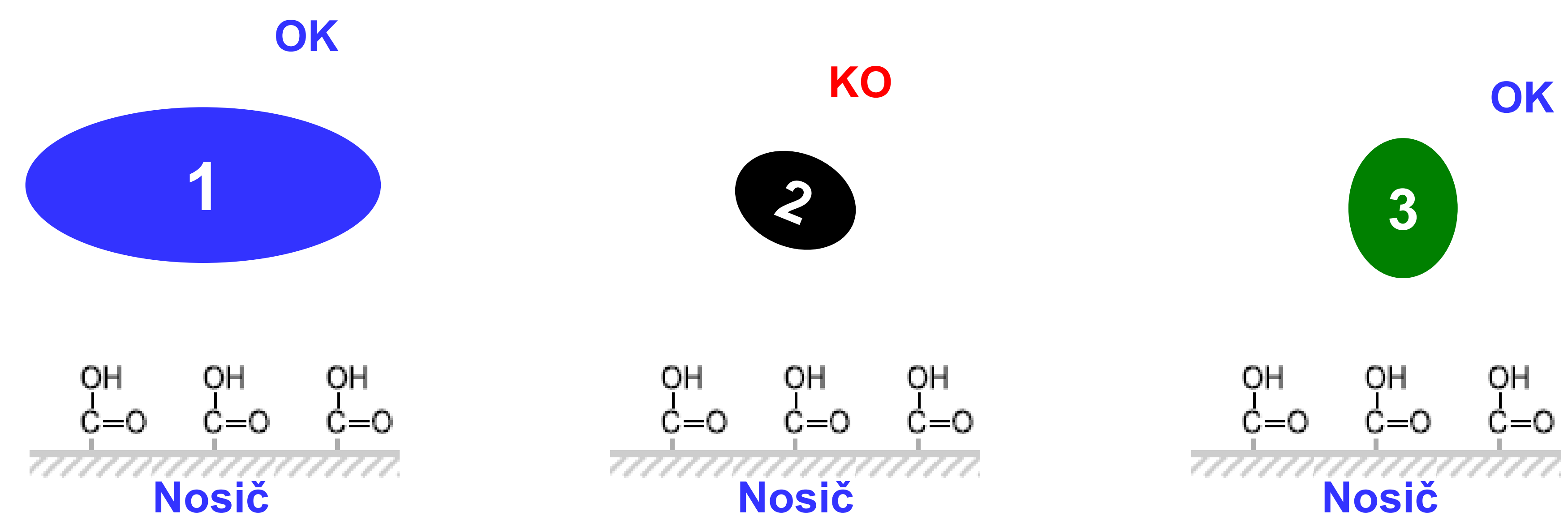
ATQGVFTLPANTFGVTAEFANESSGTQTVNVLVNNETAATFSGQSTNNAVIGTQVENSGSSGKVQVQVSVNGRPSDLVSAQVILTNELNFALVGSE
DDGTDNDYNDAVVVINWPLG

Protein 3:

SSVQTAATSWGTVPSIRVYTANNGKITERCWDGKGWYTGAFNEPGDNVSVTSWLVGSAIHIRVYASTGTTTTEWCWDGNGWTKGAYTATN

Úkol 6: *Student potřeboval pro následné experimenty imobilizovat 3 proteiny na matrici (karboxymethylovaný dextran). Nechtělo se mu ptát se na radu kolegů a tak proteiny rozpustil v doporučovaném komerčním pufru (10 mM octan sodný, pH 5,0) a provedl imobilizace. U proteinů 1 a 3 byla úspěšná, u proteinu 2 naprosto selhala. „Proč?“, ptá se zoufalý student.*

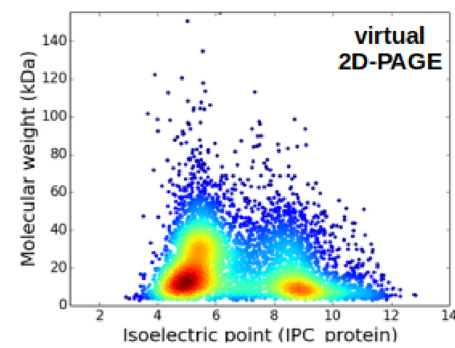
Izoelektrický bod - pI

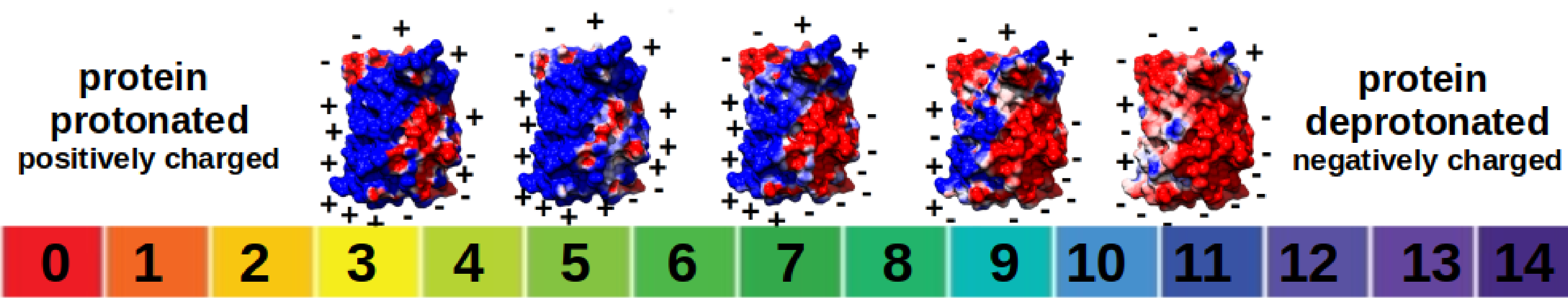


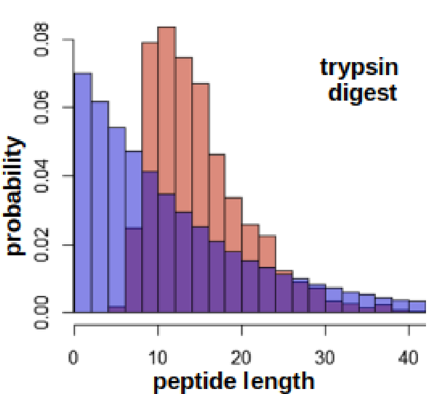
Úkol 6: Student potřeboval pro následné experimenty imobilizovat 3 proteiny na matrici (karboxymethylovaný dextran). Nechtělo se mu ptát se na radu kolegů a tak proteiny rozpustil v doporučovaném komerčním pufru (10 mM octan sodný, pH 5,0) a provedl imobilizace. U proteinů 1 a 3 byla úspěšná, u proteinu 2 naprosto selhala. „Proč?“, ptá se zoufalý student.

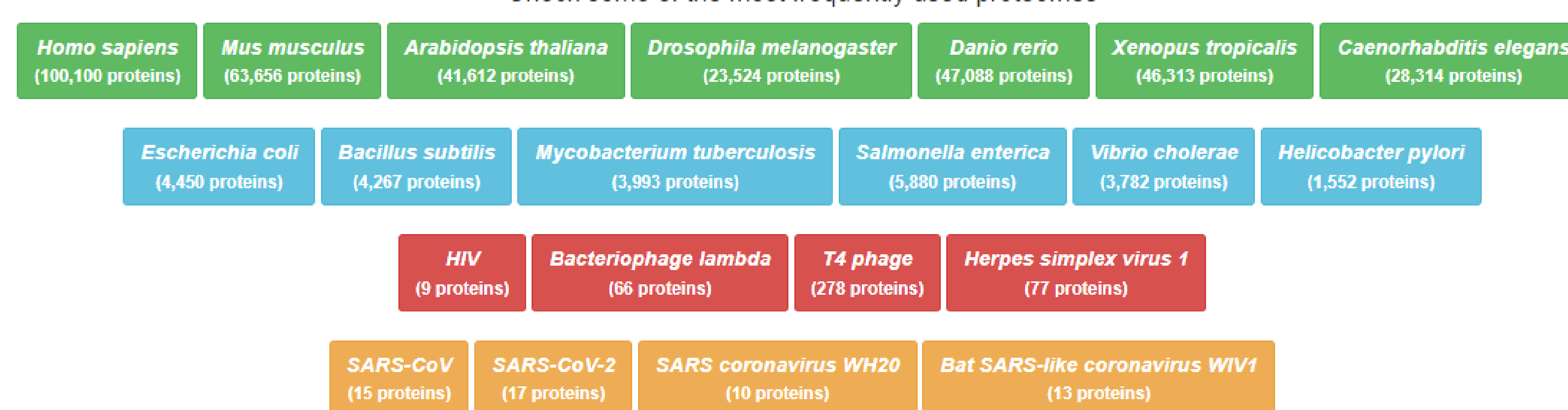
http://isoelectricpointdb2.mimuw.edu.pl/

Izoelektrický bod – pI



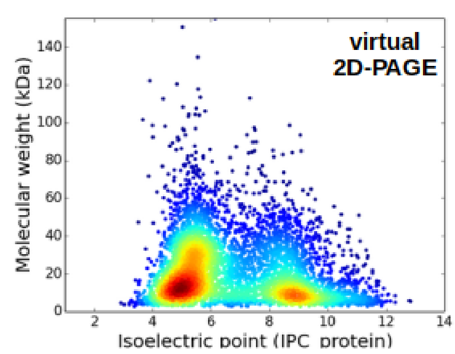


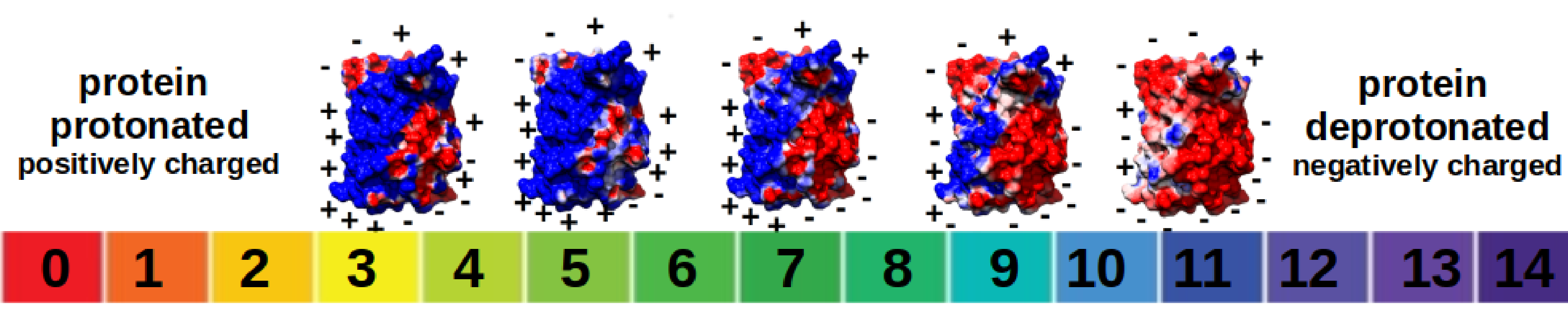


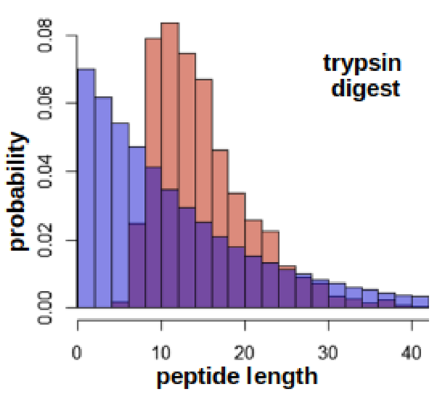


http://isoelectricpointdb2.mimuw.edu.pl/

Izoelektrický bod – pI







The goals of the database include making statistical comparisons of the various prediction methods (21 algorithms implemented) as well as facilitating the biological investigation of protein isoelectric point space. The isoelectric point, the pH at which a particular molecule carries no net electrical charge, is an important parameter for many analytical biochemistry and proteomics techniques, especially for 2D gel electrophoresis (2D-PAGE), capillary isoelectric focusing (cIEF), liquid chromatography–mass spectrometry (LC-MS) and X-ray protein crystallography

Using isoelectric point to determine the pH for initial protein crystallization trials

Jobie Kirkwood[1], David Hargreaves[2], Simon O'Keefe[3] and Julie Wilson[1,4,*]

Using data obtained from AstraZeneca and from the Structural Genomics Consortium (SGC), Oxford, we show that most proteins, both acidic and basic, do crystallize within one unit of their isoelectric point. This in turn allows for custom crystallization screens to be developed in instances where protein availability is scarce and allows deeper exploration of chemical parameter space as the pH is fixed.

Úkol 7: Jsou v databázi také proteiny z koronavirů?

Extinkční koeficient

- Extinkční koeficienty závisejí na okolí chromoforu!
- ProtParam nebere v úvahu sekundární a terciární strukturu.
- Přesné extinkční koeficienty je nutné získat experimentálně.

Extinction coefficients

The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. It is useful to have an estimation of this coefficient for following a protein which a spectrophotometer when purifying it.

Extinction coefficients

The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. It is useful to have an estimation of this coefficient for following a protein which a spectrophotometer when purifying it.

It has been shown [1c] that it is possible to estimate the molar extinction coefficient of a protein from knowledge of its amino acid composition. From the molar extinction coefficient of tyrosine, tryptophan and cystine (cysteine does not absorb appreciably at wavelengths >260 nm, while cystine does) at a given wavelength, the extinction coefficient of the native protein in water can be computed using the following equation:

```
E(Prot) = Numb(Tyr)*Ext(Tyr) + Numb(Trp)*Ext(Trp) + Numb(Cystine)*Ext(Cystine)
```

280 nm

Extinkční koeficient

- Predikovány **dvě** hodnoty (disulfidické můstky vs. bez disulfidických můstků).
- Poměrně spolehlivé pro proteiny obsahující **tryptofan** (absorbuje nejvíce, absorbance méně závislá na prostředí).
- **Předpoklad**: proteiny neobsahují jiný chromofor, který by absorboval při 280 nm.
- Proteiny bez **Trp**? Proteiny bez **Trp**, **Tyr** a disulfidových můstků?

Extinction coefficients

The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. It is useful to have an estimation of this coefficient for following a protein which a spectrophotometer when purifying it.

It has been shown [1c] that it is possible to estimate the molar extinction coefficient of a protein from knowledge of its amino acid composition. From the molar extinction coefficient of tyrosine, tryptophan and cystine (cysteine does not absorb appreciably at wavelengths >260 nm, while cystine does) at a given wavelength, the extinction coefficient of the native protein in water can be computed using the following equation:

```
E(Prot) = Numb(Tyr)*Ext(Tyr) + Numb(Trp)*Ext(Trp) + Numb(Cystine)*Ext(Cystine)
```

Protein Science (1995), *4*:2411–2423. Cambridge University Press. Printed in the USA.
Copyright © 1995 The Protein Society

How to measure and predict the molar absorption coefficient of a protein

C. NICK PACE,[1,2,3] FELIX VAJDOS,[2] LANETTE FEE,[2] GERALD GRIMSLEY,[1] AND THERONICA GRAY[1]
[1] Department of Medical Biochemistry and Genetics, [2] Department of Biochemistry and Biophysics, and [3] Center for Macromolecular Design, Texas A&M University, College Station, Texas 77843-1114

(RECEIVED July 12, 1995; ACCEPTED September 8, 1995)

Extinkční koeficient

Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm

Nicholas J. Anthis and G. Marius Clore*

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, 20892-0520

Received 1 February 2013; Revised 11 March 2013; Accepted 14 March 2013
DOI: 10.1002/pro.2253
Published online 23 March 2013 proteinscience.org

Abstract: Quantitative studies in molecular and structural biology generally require accurate and precise determination of protein concentrations, preferably via a method that is both quick and straightforward to perform. The measurement of ultraviolet absorbance at 280 nm has proven especially useful, since the molar absorptivity (extinction coefficient) at 280 nm can be predicted directly from a protein sequence. This method, however, is only applicable to proteins that contain tryptophan or tyrosine residues. Absorbance at 205 nm, among other wavelengths, has been used as an alternative, although generally using absorptivity values that have to be uniquely calibrated for each protein, or otherwise only roughly estimated. Here, we propose and validate a method for predicting the molar absorptivity of a protein or peptide at 205 nm directly from its amino acid sequence, allowing one to accurately determine the concentrations of proteins that do not contain tyrosine or tryptophan residues. This method is simple to implement, requires no calibration, and should be suitable for a wide range of proteins and peptides.

Keywords: protein; absorbance; UV; concentration; molecular biology; absorptivity; extinction coefficient

Protein Calculator

Reference (please quote when using this program):

Anthis, N.J. & Clore, G.M. (2013) Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm, *Protein Science, 22*, 851-858.
Pubmed PDF

Enter your amino acid sequence (in single letters) in the box below. It can be upper- or lower-case and in any format (the script will ignore any spaces, numbers, symbols, etc.), but any other text will be interpreted as amino acids. Input is limited to 5,000 characters in total.

LENKALENKALENKALENKALENKALENKALENKALENKA

submit reset

https://spin.niddk.nih.gov/clore/Software/A205.html

Extinkční koeficient

This script calculates molar absorptivities (extinction coefficients) at 205 nm and 280 nm from an amino acid sequence. It also calculates the molecular weight for various universal isotopic labeling schemes.

Reference:
Anthis N.J., Clore G.M. (2013) Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm, Protein Science 22, 851-8, doi:10.1002/pro.2253

Amino acid sequence:
LENKALENKALENKALENKALENKALENKALENKALENKA

Number of residues:
40

Molar absorptivity (extinction coefficient) at 280 nm =
0 M-1 cm-1

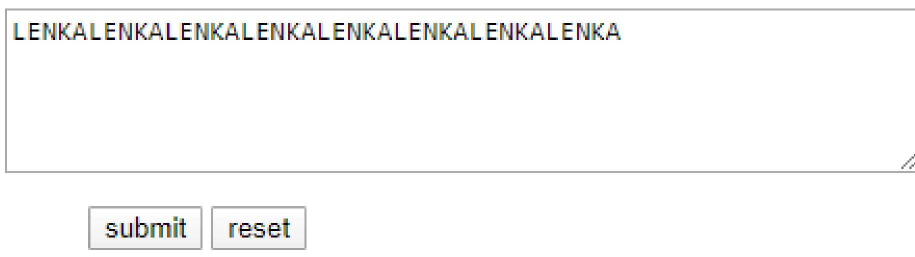
Molar absorptivity (extinction coefficient) at 205 nm =
111620 M-1 cm-1

Protein Calculator

Reference (please quote when using this program):

Anthis, N.J. & Clore, G.M. (2013) Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm, *Protein Science, 22*, 851-858.
Pubmed *PDF*

Enter your amino acid sequence (in single letters) in the box below. It can be upper- or lower-case and in any format (the script will ignore any spaces, numbers, symbols, etc.), but any other text will be interpreted as amino acids. Input is limited to 5,000 characters in total.



https://spin.niddk.nih.gov/clore/Software/A205.html

Extinkční koeficient

```
This script calculates molar absorptivities (extinction coefficients) at
205
nm and 280 nm from an amino acid sequence. It also calculates the
molecular
weight for various universal isotopic labeling schemes.

Reference:
Anthis N.J., Clore G.M. (2013) Sequence-specific determination of
protein
and peptide concentrations by absorbance at 205 nm, Protein Science 22,
851-8, doi:10.1002/pro.2253

Amino acid sequence:
LENKALENKALENKALENKALENKALENKALENKALENKA

Number of residues:
40
```

```
Molar absorptivity (extinction coefficient) at 280 nm =
0 M-1 cm-1

Molar absorptivity (extinction coefficient) at 205 nm =
111620 M-1 cm-1
```

https://spin.niddk.nih.gov/clore/Software/A205.html

Peptid 1:
AQQGVFTLP

Peptid 2:
NESEMELEMAKNAMELEPEPR

Peptid 3:
ELGKTNMSDFGHACPTRQELDFASIPQTVCHSDFGHLK

Peptid 4:
AWQGWVFTLP

Peptid 5:
ELGKTNMSDFGHAWCPTRQELDFASIPQTVCHSDFGHLK

Úkol 8: ***Predikujte extinkční koeficienty peptidů při 280 a 205 nm.***

Extinkční koeficient

http://bestsel.elte.hu/extcoeff.php

Prediction of Molar Extinction Coefficients of Proteins and Peptides Using UV Absorption of the Constituent Amino Acids at 214 nm To Enable Quantitative Reverse Phase High-Performance Liquid Chromatography–Mass Spectrometry Analysis

BAS J. H. KUIPERS AND HARRY GRUPPEN*

Department of Agrotechnology and Food Sciences, Laboratory of Food Chemistry, Wageningen University, P.O. Box 8129, 6700 EV Wageningen, The Netherlands

The molar extinction coefficients of 20 amino acids and the peptide bond were measured at 214 nm in the presence of acetonitrile and formic acid to enable quantitative comparison of peptides eluting from reversed-phase high-performance liquid chromatography, once identified with mass spectrometry (RP-HPLC-MS). The peptide bond has a molar extinction coefficient of 923 M^{-1} cm^{-1}. Tryptophan has a molar extinction coefficient that is ~30 times higher than that of the peptide bond, whereas the molar extinction coefficients of phenylalanine, tyrosine, and histidine are ~six times higher than that of the peptide bond. Proline, as an individual amino acid, has a negligible molar extinction coefficient. However, when present in the peptide chain (except at the N terminus), it absorbs ~three times more than a peptide bond. Methionine has a similar molar extinction coefficient as the peptide bond, while all other amino acids have much lower molar extinction coefficients. The predictability of the molar extinction coefficients of proteins and peptides, calculated by the amino acid composition and the number of peptide bonds present, was validated using several proteins and peptides. Most of the measured and calculated molar extinction coefficients were in good agreement, which shows that it is possible to compare peptides analyzed by RP-HPLC-MS in a quantitative way. This method enables a quantitative analysis of all peptides present in hydrolysates once identified with RP-HPLC-MS.

absorbs light (*7*). This absorbance can be used to estimate the relative amounts of peptides present in a hydrolysate. However, quantification based on the absorbance at 214 nm is not a common method, since in this wavelength region the absorption of the peptide bond is also conformation-dependent (*8*). Furthermore, besides the peptide bond, several amino acid residues are reported to contribute significantly to the absorption (*9*, *10*). Moreover, the absorbance of peptides is also pH-dependent due to the absorption of the carboxylic acid group, which depends on whether the carboxyl group is protonated or not. This results in a variation of the absorbance around the pK_a (~pH 3) of the free carboxylic acid in the peptides (*11*). This effect plays the most dominant role in dipeptides and decreases with increasing peptide length (*12*, *13*).

Because of the complexity of absorption at 214 nm, researchers are, in general, skeptic toward quantification of proteins at this wavelength. However, for peptides devoid of tryptophan or tyrosine, there is no good alternative (*14*). As explained above, the complexity of the absorbance is much higher as

Extinkční koeficient

Calculation of Extinction Coefficients at 205 and 214 nm

Sequence:

Number of S-S bonds:

0

Please, type the characters above or use your password.

Submit Reset

Protein concentration can be determined by the absorbance at 205 or 214 nm. It is especially useful when absorbance at 280 nm cannot be used in the lack of Trp and Tyr residues. CD samples can be directly measured at these wavelengths due to the high extinction coefficients. If the spectro-polarimeter is capable of converting the HT values to absorbances, then the concentrations can be determined right from the CD measurements after subtracting the baseline absorptions. Extinction coefficients at 205 and 214 nm can be calculated from the sequence.

References:

205 nm: Prot. Sci. 2013, 22, 851-858.
214 nm: J. Agric. Food Chem. 2007, 55, 5445-5451.

Peptid 1:
AQQGVFTLP

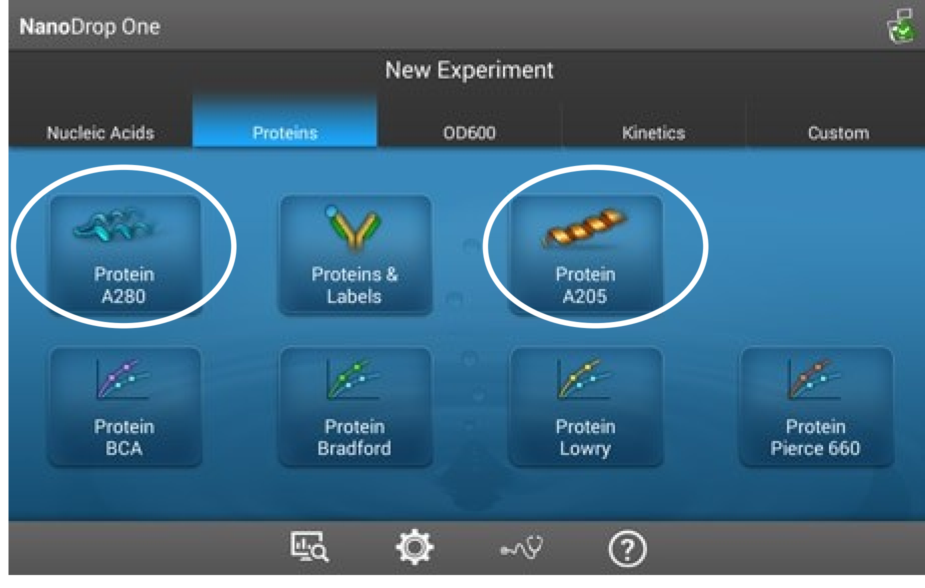
Peptid 2:
NESEMELEMAKNAMELEPEPR

Peptid 3:
ELGKTNMSDFGHACPTRQELDFASIPQTVCHSDFGHLK

http://bestsel.elte.hu/extcoeff.php

Úkol 9: Predikujte extinkční koeficienty peptidů při 214 a 205 nm.

Extinkční koeficient



Popis:
- pipetování vzorku přímo na měřící optiku
- ani vzorky s velmi vysokou koncentrací není třeba ředit
- zdroj světla: xenonová výbojka
- rychlé měření do 8 sekund
- přednastavené metody (např. A260, A280, značená DNA, proteiny)
- uživatelsky přívětivý SW
- verze s kyvetovým držákem (One c) je navíc vybavena inkubací
- nemusí být propojen s počítačem – obsahuje tablet pro ovládání přístroje a měření
- varianty: bez kyvety, s kyvetou.

Parametry:
- rozsah nastavení vlnových délek: 190 až 850 nm
- rozsah detekce: 2 až 27 500 ng/ul (ds DNA), 0,06 až 820 mg/ml (BSA)
- měřící rozsah: 0 až 550 Abs
- objem vzorku: 1,0 ul

A205	• varies from peptide to peptide. Approximate range 0.003 - 10.74 mg/mL	• Measures peptide backbone absorbance at 205 nm. For peptides that lack or have few Trp and Tyr residues • Scopes method option can be used to measure proteins that have significant amounts of Trp and Tyr. • More sensitive than A280 since A205, molar absorptivity is high.	• High salt protein buffers such as PBS and TE absorb in the low UV. Use a low salt buffer like the Brij® buffer diluted to 0.01%.

/Analysis Service /Protein Characterization /Protein Identification /Protein Sequence Analysis / Extinction Coefficient Determination Service

Extinction Coefficient Determination Service

TALK WITH AN EXPERT

The protein concentration is an important parameter affecting the efficacy of pharmaceutical products. Therefore, it is essential and necessary to determine the concentration of a protein solution during drug development. There are several methods for protein concentration determination, such as include UV absorption, the Braford assay, the bicinchoninic acid (BCA) assay, and so forth. To determine the protein concentration UV absorption, absorption in the UV region of the spectrum will be measure, and the concentration will be calculated according to the Beer-Lambert law, where extinction coefficient and path length are constant. Extinction coefficient, a measure of how strongly a substance absorbs light at a specific wavelength, is the intrinsic property of a protein depending on its composition and structure. Hence, to precisely determine protein concentration, it is fundamental to accurately determine extinction coefficient. Creative Proteomics offers reliable service to help you determine the extinction coefficient accurately, in accordance with the ICH Q6B Guidelines.

https://www.creative-proteomics.com/pronalyse/extinction-coefficient-determination.html

https://jordilabs.com/lab-testing/problems-we-solve/protein-analysis/protein-extinction-coefficient/

Protein concentration is a crucial parameter affecting the efficacy of a therapeutic. Accurate quantification provides key information supporting formulation studies and the determination of enzyme kinetics. Quantitation is critical for a wide variety of downstream applications in the preparation of commercial pharmaceuticals. Several approaches which are commonly applied to determine the protein concentration include UV absorption, the Bradford assay and the bicinchoninic acid (BCA) assay. Among these, UV absorption is widely applied due to its relative high accuracy, low cost and short analysis time.

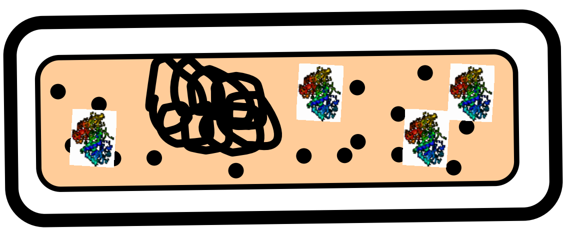
For UV absorption, the protein concentration is determined based on the measurement of the UV absorbance at a characteristic wavelength using the known protein extinction coefficient (Beer's Law). The strong UV-light absorption of the protein is due to the presence of aromatic side chains in the amino acids, such as tyrosine, tryptophan and cysteine. The absorbance measurement is usually conducted at 280 nm to reduce interferences from other compounds in the protein solutions. According to Beer's Law, the molar extinction coefficient ($M^{-1}cm^{-1}$) is a constant for a given substance dissolved in a given solution and measured at a given wavelength. For proteins, percent solution extinction coefficients $(g/100mL)^{-1}cm^{-1}$ are used for convenience instead of molar extinction coefficients. The percent solution extinction coefficient of a protein is a key parameter which can be determined by calibration with a commercial standard, estimated based on the chemical properties of the protein or it can be obtained from a database such as that provided by NIST.

Protein Analysis with Jordi Labs

Jordi Labs has extensive experience conducting UV-vis absorption and protein analysis. We offer high-quality services for the determination of the protein extinction coefficient and protein solution concentration. Our highly skilled chemists provide outstanding customer service and have the background needed to assist with simple or complex projects.

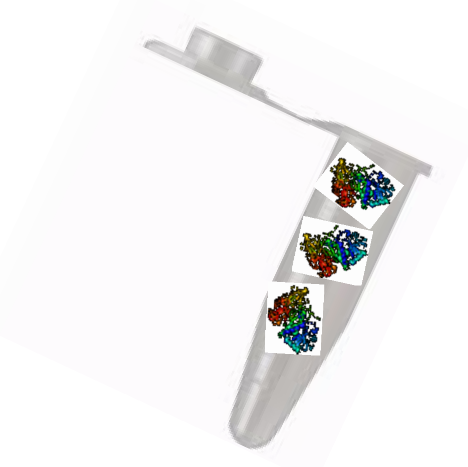
Jak stabilní je můj protein?

- Stabilita ***in vivo***
- Stabilita proteinu v buňce



- Degradace proteinů v buňce je **aktivní proces** („udělej svoji práci a zmiz, ať nezavazíš“).
- **„In-vivo half-life“**

- Stabilita ***in vitro***
- Stabilita proteinu ve zkumavce

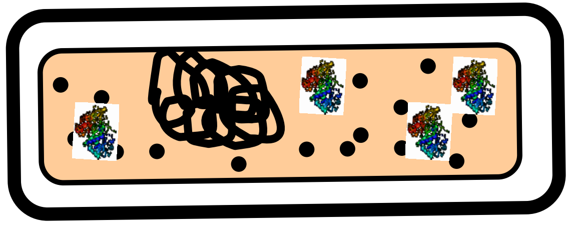


- **„Instability index“**

Jak stabilní je můj protein?

- Stabilita *in vivo*

- Stabilita proteinu v buňce



In vivo half-life
The half-life is a prediction of the time it takes for half of the amount of protein in a cell to disappear after its synthesis in the cell. ProtParam relies on the "N-end rule", which relates the half-life of a protein to the identity of its N-terminal residue; the prediction is given for 3 model organisms (human, yeast and E.coli). The N-end rule (for a review see [5],[6]) originated from the observations that the identity of the N-terminal residue of a protein plays an important role in determining its stability in vivo ([2],[3],[4]). The rule was established from experiments that explored the metabolic fate of artificial beta-galactosidase proteins with different N-terminal amino acids engineered by site-directed mutagenesis. The beta-gal proteins thus designed have strikingly different half-lives in vivo, from more than 100 hours to less than 2 minutes, depending on the nature of the amino acid at the amino terminus and on the experimental model (yeast in vivo; mammalian reticulocytes in vitro, Escherichia coli in vivo). In addition, it has been shown that in eukaryotes, the association of a destabilizing N-terminal residue and of an internal lysine targets the protein to ubiquitin-mediated proteolytic degradation [6]. Note that the program gives an **estimation** of the protein half-life and is not applicable for N-terminally modified proteins.

The N-end rule pathway of protein degradation

Alexander Varshavsky*

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

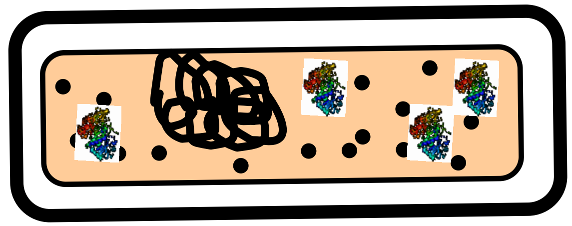
The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct versions of the N-end rule operate in all organisms examined, from mammals to fungi and bacteria. In eukaryotes, the N-end rule pathway is a part of the ubiquitin system. Ubiquitin is a 76-residue protein whose covalent conjugation to other proteins plays a role in many biological processes, including cell growth and differentiation. I discuss the current understanding of the N-end rule pathway.

- N-koncové pravidlo
- Odhad – osekání sekvenční informace na **jednu** aminokyselinu.
- Problém odštěpování **iniciačního methioninu**: která aminokyselina je ve skutečnosti první?
- Nástroje pro predikci odštěpení iniciačního methioninu.

Jak stabilní je můj protein?

- Stabilita *in vivo*

- Stabilita proteinu v buňce



Úkol 10: Predikujte in vivo half-life následujících proteinů:

Protein 1:
MERDGTFNLPPHIKFGVTALTHAANDQTIDIYIDDDPKPAATFKGAGAQDQNL
GTKVLDSGNGRVRVIVMANGRPSRLGSRQVDIFKKSYFGIIGSEDGADDDYND
GIVFLNWPLG

Protein 2:
ERDGTFNLPPHIKFGVTALTHAANDQTIDIYIDDDPKPAATFKGAGAQDQNLG
TKVLDSGNGRVRVIVMANGRPSRLGSRQVDIFKKSYFGIIGSEDGADDDYNDG
IVFLNWPLG

The N-end rule pathway of protein degradation

Alexander Varshavsky*

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct versions of the N-end rule operate in all organisms examined, from mammals to fungi and bacteria. In eukaryotes, the N-end rule pathway is a part of the ubiquitin system. Ubiquitin is a 76-residue protein whose covalent conjugation to other proteins plays a role in many biological processes, including cell growth and differentiation. I discuss the current understanding of the N-end rule pathway.

Úkol 11: Odštěpuje se u proteinu 1 iniciační methionin?

P1'	Cleavage	P2' Pro	Radius	Area
Glycine	Yes	No	0	0
Alanine	Yes	No	0.77	67
Serine	Yes	No	1.08	80
Cysteine	Yes	No	1.22	104
Theonine	Variable	No	1.24	102
Proline	Yes	No	1.25	105
Valine	Variable	No	1.29	117
Aspartic acid	No		1.43	106
Asparagine	No		1.45	113
Leucine	No		1.54	137
Isoleucine	No		1.56	140
Glutamine	No		1.75	144
Glutamic acid	No		1.77	138
Histidine	No		1.78	151
Methionine	No		1.80	160
Phenylanine	No		1.90	175
Lysine	No		2.08	167
Tyrosine	No		2.13	187
Trptophan	No		2.21	217
Arginine	No		2.38	196

Jak stabilní je můj protein?

Úkol 10: Predikujte in vivo half-life následujících proteinů:

Protein 1:
MERDGTFNLPPHIKFGVTALTHAANDQTIDIYIDDDPKPAATFKGAGAQDQNL
GTKVLDSGNGRVRVIVMANGRPSRLGSRQVDIFKKSYFGIIGSEDGADDDYND
GIVFLNWPLG

Protein 2:
ERDGTFNLPPHIKFGVTALTHAANDQTIDIYIDDDPKPAATFKGAGAQDQNLG
TKVLDSGNGRVRVIVMANGRPSRLGSRQVDIFKKSYFGIIGSEDGADDDYNDG
IVFLNWPLG

N-Terminal Methionine Processing

Paul Wingfield
Protein Expression Laboratory, NIAMD/NIH, Bldg. 6B, Room 1B130, 9000 Rockville Pike, Bethesda, MD 20892, Tel: 301-594-1313, pelpw@helix.nih.gov

Abstract

Protein synthesis is initiated by methionine in eukaryotes and formylmethionine in prokaryotes. N-terminal methionine can be co-translationally cleaved by the enzyme methionine aminopeptidase (MAP). When recombinant proteins are expressed in bacterial and mammalian expression systems there is a simple universal rule which predict whether the initiating methionine will be processed by MAP and is based on the size of the residue adjacent (penultimate) to the N-methionine. In general, if the side-chains of the penultimate residues have a radius of gyration of 1.29 Å or less methionine is cleaved. This rule was originally suggested by Sherman et al., (1985) based on their pioneering studies with yeast.

Úkol 11: Odštěpuje se u proteinu 1 iniciační methionin?

http://terminus.unige.ch/

https://bioweb.i2bc.paris-saclay.fr/terminator3/

„N-terminomics“

- N-terminální posttranslační modifikace ovlivňují stabilitu proteinů.
- Modifikace (např. acetylace), odštěpení iniciačního methioninu, odštěpení signálního peptidu.
- Analýza pomocí **MS** – **bioinformatické** nástroje pro analýzu dat.
- Bioinformatické nástroje pro **predikci** modifikací.
- **Databáze** věnované „terminálním“ modifikacím.

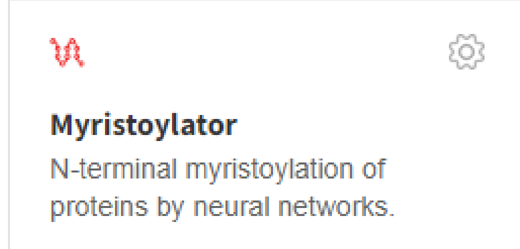
*„**N-terminomics** is a rapidly evolving branch of proteomics that encompasses the study of protein N-terminal sequence. A proteome-wide collection of such sequences has been widely used to understand the proteolytic cascades and in annotating the genome. Over the last two decades, various N-terminomic strategies have been developed for achieving high sensitivity, greater depth of coverage, and high-throughputness.“*

Review

N-terminomics – its past and recent advancements

Prashant Kaushal [a,b], Cheolju Lee [a,b,c,*]

[a] *Center for Theragnosis, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea*
[b] *Division of Bio-Medical Science & Technology, KIST School, Korea University of Science and Technology, Seoul 02792, Republic of Korea*
[c] *KHU-KIST Department of Converging Science and Technology, Kyung Hee University, 26 Kyunghee-daero, Dongdaemun-gu, Seoul 02447, Republic of Korea*



https://www.expasy.org/resources/myristoylator

https://topfind.clip.msl.ubc.ca/home/index

The public knowledgebase for protein termini and protease processing

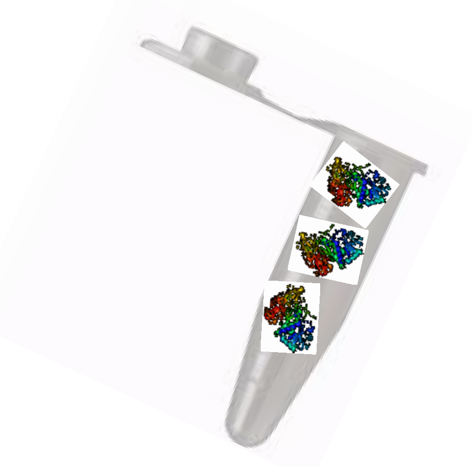
Jak stabilní je můj protein?

Instability index (II)
The instability index provides an estimate of the stability of your protein in a test tube. Statistical analysis of 12 unstable and 32 stable proteins has revealed [7] that there are certain dipeptides, the occurence of which is significantly different in the unstable proteins compared with those in the stable ones. The authors of this method have assigned a weight value of instability to each of the 400 different dipeptides (DIWV).

First amino acid of dipeptide	Second amino acid of dipeptide																			
	W	C	M	H	Y	F	Q	N	I	R	D	P	T	K	E	V	S	G	A	L
W	1.0	1.0	24.68	24.68	1.0	1.0	1.0	13.34	1.0	1.0	1.0	1.0	−14.03	1.0	1.0	−7.49	1.0	−9.37	−14.03	13.34
C	24.68	1.0	33.6	33.6	1.0	1.0	−6.54	1.0	1.0	1.0	20.26	20.26	33.6	1.0	1.0	−6.54	1.0	1.0	1.0	20.26
M	1.0	1.0	−1.88	58.28	24.68	1.0	−6.54	1.0	1.0	−6.54	1.0	44.94	−1.88	1.0	1.0	1.0	44.94	1.0	13.34	1.0
H	−1.88	1.0	1.0	1.0	44.94	−9.37	1.0	24.68	44.94	1.0	1.0	−1.88	−6.54	24.68	1.0	1.0	1.0	−9.37	1.0	1.0
Y	−9.37	1.0	44.94	13.34	13.34	1.0	1.0	1.0	1.0	−15.91	24.68	13.34	−7.49	1.0	−6.54	1.0	1.0	−7.49	24.68	1.0
F	1.0	1.0	1.0	1.0	33.6	1.0	1.0	1.0	1.0	1.0	13.34	20.26	1.0	−14.03	1.0	1.0	1.0	1.0	1.0	1.0
Q	1.0	−6.54	1.0	1.0	−6.54	−6.54	20.26	1.0	1.0	1.0	20.26	20.26	1.0	1.0	20.26	−6.54	44.94	1.0	1.0	1.0
N	−9.37	−1.88	1.0	1.0	1.0	−14.03	−6.54	1.0	44.94	1.0	1.0	−1.88	−7.49	24.68	1.0	1.0	1.0	−14.03	1.0	1.0
I	1.0	1.0	1.0	13.34	1.0	1.0	1.0	1.0	1.0	1.0	1.0	−1.88	1.0	−7.49	44.94	−7.49	1.0	1.0	1.0	20.26
R	58.28	1.0	1.0	20.26	−6.54	1.0	20.26	13.34	1.0	58.28	1.0	20.26	1.0	1.0	1.0	1.0	44.94	−7.49	1.0	1.0
D	1.0	1.0	1.0	1.0	1.0	−6.54	1.0	1.0	1.0	−6.54	1.0	1.0	−14.03	−7.49	1.0	1.0	20.26	1.0	1.0	1.0
P	−1.88	−6.54	−6.54	1.0	1.0	20.26	20.26	1.0	1.0	−6.54	−6.54	20.26	1.0	1.0	18.38	20.26	20.26	1.0	20.26	1.0
T	−14.03	1.0	1.0	1.0	1.0	13.34	−6.54	−14.03	1.0	1.0	1.0	1.0	1.0	1.0	20.26	1.0	1.0	−7.49	1.0	1.0
K	1.0	1.0	33.6	1.0	1.0	1.0	24.68	1.0	−7.49	33.6	1.0	−6.54	1.0	1.0	1.0	−7.49	1.0	−7.49	1.0	−7.49
E	−14.03	44.94	1.0	−6.54	1.0	1.0	20.26	1.0	20.26	1.0	20.26	20.26	1.0	1.0	33.6	1.0	20.26	1.0	1.0	1.0
V	1.0	1.0	1.0	1.0	−6.54	1.0	1.0	1.0	1.0	1.0	−14.03	20.26	−7.49	−1.88	1.0	1.0	1.0	−7.49	1.0	1.0
S	1.0	33.6	1.0	1.0	1.0	1.0	20.26	1.0	1.0	20.26	1.0	44.94	1.0	1.0	20.26	1.0	20.26	1.0	1.0	1.0
G	13.34	1.0	1.0	1.0	−7.49	1.0	1.0	−7.49	−7.49	1.0	1.0	1.0	−7.49	−7.49	−6.54	1.0	1.0	13.34	−7.49	1.0
A	1.0	44.94	1.0	−7.49	1.0	1.0	1.0	1.0	1.0	1.0	−7.49	20.26	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
L	24.68	1.0	1.0	1.0	1.0	1.0	33.6	1.0	1.0	20.26	1.0	20.26	1.0	−7.49	1.0	1.0	1.0	1.0	1.0	1.0

- Stabilita *in vitro*

- Stabilita proteinu ve zkumavce



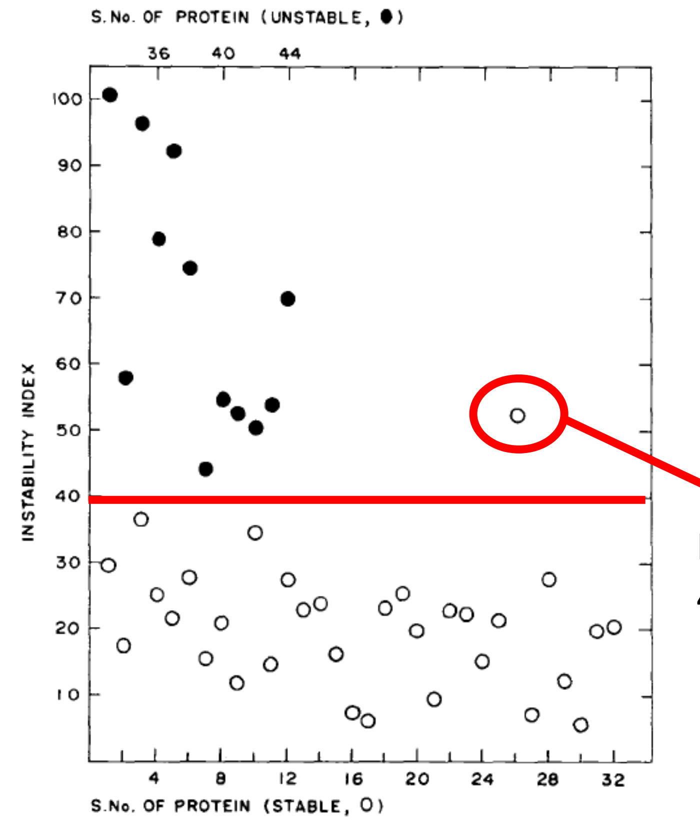
- **„Instability index“**

Jak stabilní je můj protein?

Instability index (II)
The instability index provides an estimate of the stability of your protein in a test tube. Statistical analysis of 12 unstable and 32 stable proteins has revealed [7] that there are certain dipeptides, the occurence of which is significantly different in the unstable proteins compared with those in the stable ones. The authors of this method have assigned a weight value of instability to each of the 400 different dipeptides (DIWV).

- Stabilita *in vitro*

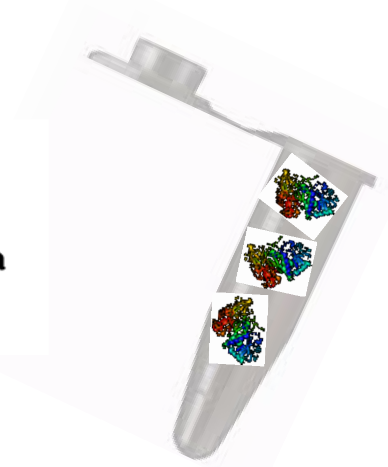
- Stabilita proteinu ve zkumavce



Protein Engineering vol.4 no.2 pp.155–161, 1990

Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence

RNasa A, stabilní protein,
4 disulfidové můstky



- „Instability index“

Jak stabilní je můj protein?

Instability index (II)
The instability index provides an estimate of the stability of your protein in a test tube. Statistical analysis of 12 unstable and 32 stable proteins has revealed [7] that there are certain dipeptides, the occurence of which is significantly different in the unstable proteins compared with those in the stable ones. The authors of this method have assigned a weight value of instability to each of the 400 different dipeptides (DIWV).

Protein Pept Lett. 2019;26(5):339-347. doi: 10.2174/0929866526666190228144219.

Applicability of Instability Index for In vitro Protein Stability Prediction.

Gamage DG[1], Gunaratne A[2], Periyannan GR[1], Russell TG[1].

Abstract

BACKGROUND: The dipeptide composition-based Instability Index (II) is one of the protein primary structure-dependent methods available for in vivo protein stability predictions. As per this method, proteins with II value below 40 are stable proteins. Intracellular protein stability principles guided the original development of the II method. However, the use of the II method for in vitro protein stability predictions raises questions about the validity of applying the II method under experimental conditions that are different from the in vivo setting.

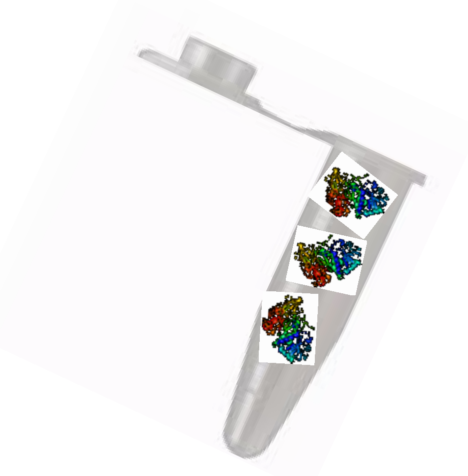
OBJECTIVE: The aim of this study is to experimentally test the validity of the use of II as an in vitro protein stability predictor.

METHODS: A representative protein CCM (CCM - Caulobacter crescentus metalloprotein) that rapidly degrades under in vitro conditions was used to probe the dipeptide sequence-dependent degradation properties of CCM by generating CCM mutants to represent stable and unstable II values. A comparative degradation analysis was carried out under in vitro conditions using wildtype CCM, CCM mutants and two other candidate proteins: metallo-β-lactamase L1 and α -S1- casein representing stable, borderline stable/unstable, and unstable proteins as per the II predictions. The effect of temperature and a protein stabilizing agent on CCM degradation was also tested.

RESULTS: Data support the dipeptide composition-dependent protein stability/instability in wt-CCM and mutants as predicted by the II method under in vitro conditions. However, the II failed to accurately represent the stability of other tested proteins. Data indicate the influence of protein environmental factors on the autoproteolysis of proteins.

CONCLUSION: Broader application of the II method for the prediction of protein stability under in vitro conditions is questionable as the stability of the protein may be dependent not only on the intrinsic nature of the protein but also on the conditions of the protein milieu.

- Stabilita *in vitro*
- Stabilita proteinu ve zkumavce



- „Instability index“

Agregace proteinů

- Agregace proteinů = shlukování proteinů
- Agregace proteinů *in vitro* vs. *in vivo*

- Agregace *in vitro* může být způsobena poškozením proteinů – narušení prostorového uspořádání (struktury), odhalení hydrofobních částí proteinu. Nevhodné pracovní podmínky, mechanické poškození.

- Agregace *in vivo* – hromadění špatně sbaleného proteinu (mutace)
- Špatně sbalené proteiny – neurodegenerativní choroby

- Vysoká koncentrace proteinu v bakteriální buňce - problémy při produkci rekombinantních proteinů (inkluzní tělíska)

Agregace proteinů

- Agregace *in vivo* – hromadění špatně sbaleného proteinu (mutace)
- Špatně sbalené proteiny – neurodegenerativní choroby

NEURODEGENERATIVNÍ CHOROBY ZPŮSOBENÉ UKLÁDÁNÍM NESPRÁVNĚ „SLOŽENÝCH" PROTEINŮ

(viz rovněž Vesmír 78, 328, 1999/6)

- Alzheimerova choroba je nejčastější ze všech demencí (tvoří zhruba 50–60 % všech případů). Podrobněji se jí zabývá rámeček T. Hájka na protější straně.
- Pickova choroba je velmi vzácná demence, rovněž podobná Alzheimerově nemoci. Atrofie mozkové kůry postihuje hlavně čelní a spánkový lalok, což způsobuje velmi nápadný klinický obraz. V počátečním stadiu choroby se objevují především hrubé změny osobnosti a postižení jedinci se často dopouštějí drobných sexuálních deliktů. Případné trestní stíhání je samozřejmě zastaveno pro nepříčetnost. Atrofie kortexu pokračuje poměrně rychle a nemoc končí smrtí.
- Parkinsonova choroba byla poprvé popsána r. 1817. Je častější u mužů a v pozdějším věku, ale nevyhýbá se ani mladým lidem. Klinicky se projevuje svalovým třesem (tremorem), který je nejvíce patrný v klidu a při pohybu mizí, dále ztuhlostí, pomalostí pohybů, šoupavou chůzí. Deprese se vyskytuje u 30 % pacientů a u 10–30 % se rozvine demence podobná Alzheimerově nemoci. Nemoc je způsobena úbytkem nervových buněk produkujících neurotransmiter dopamin v *substantia nigra*. Příznaky se objevují při destrukci více než 70 % dopaminergních neuronů v této oblasti mozku.
- Kuru je vzácná infekční prionová choroba vyskytující se na Papui-Nové Guinei u kmene Fore, který holdoval rituálnímu kanibalizmu. Priony se přenášejí především pojídáním mozků a jater zemřelých příslušníků kmene. Klinický i patologický obraz kuru je shodný s Creutzfeldtovou-Jakobovou chorobou.

V mozcích pacientů trpících Alzheimerovou chorobou se vyskytují dva různé proteinové agregáty: smotky chybně složeného cytoskeletárního proteinu *tau* a plaky z proteinových agregátů tvořených fragmenty membránového proteinu: β-amyloidového prekurzoru (β-APP). Přítomnost dvou druhů proteinových agregátů vedla ve vědecké obci k prudkým sporům o to, který z těchto proteinů je skutečnou příčinou onemocnění a který jen vedlejším příznakem. Zastánci jednotlivých teorií (posměšně nazývaní *tauisté* a *baptisté*) jdou ve svém nadšení pro oblíbenou teorii někdy tak daleko, že se v přednáškách na velkých vědeckých konferencích o protivném proteinu vůbec nezmiňují. Podle posledních prací (Science 293, 1487, 2001) se ale zdá, že oba mechanizmy jsou vzájemně provázány. Vzhledem k tomu, že autorem tohoto článku je přesvědčený *baptista*, neuslyšíte o proteinu *tau* v dalším textu už ani slovo.

- Huntingtonova chorea je autozomálně dominantní dědičné onemocnění. Defektní gen leží na 4. chromozomu. Postihuje osoby ve věku 20 až 50 let. Pacienti přežívají průměrně 15 let. Klinicky se projevuje mimovolními pohyby připomínajícími tanec (odtud *chorea*), psychotickými příznaky (halucinacemi, bludy) a postupující demencí. Takové klinické příznaky jsou důsledkem destrukce malých neuronů ve dvou částech mozku – v ocasatém jádře (*nucleus caudatus*) a skořápce (*putamen*).

Jan Konvalinka, Jana Peichlová | 1. 1. 2002 | Vesmír 81, 11, 2002/1

Agregace proteinů

- Agregace *in vivo* – hromadění špatně sbaleného proteinu (mutace)
- Špatně sbalené proteiny – neurodegenerativní choroby

Alzheimer's disease

JOSE A. SORIA LOPEZ[1,2], HECTOR M. GONZÁLEZ[1,2], AND GABRIEL C. LÉGER[1,2*]
[1]*Department of Neurosciences, University of California San Diego, La Jolla, CA, United States*
[2]*Shiley-Marcos Alzheimer's Disease Research Center, University of California San Diego, La Jolla, CA, United States*

Abstract

Alzheimer's disease (AD) dementia refers to a particular onset and course of cognitive and functional decline associated with age together with a particular neuropathology. It was first described by Alois Alzheimer in 1906 about a patient whom he first encountered in 1901. Modern clinical diagnostic criteria have been developed, and criteria have also been proposed to recognize preclinical (or presymptomatic) stages of the disease with the use of biomarkers. The primary neuropathology was described by Alzheimer, and in the mid-1980s subsequently evolved into a more specific neuropathologic definition that recognizes the comorbid neuropathologies that frequently contribute to clinical dementia. Alzheimer's disease is now the most common form of neurodegenerative dementia in the United States with a disproportionate disease burden in minority populations. Deficits in the ability to encode and store new memories characterizes the initial stages of the disease. Subsequent progressive changes in cognition and behavior accompany the later stages. Changes in amyloid precursor protein (APP) cleavage and production of the APP fragment beta-amyloid (Aβ) along with hyperphosphorylated tau protein aggregation coalesce to cause reduction in synaptic strength, synaptic loss, and neurodegeneration. Metabolic, vascular, and inflammatory changes, as well as comorbid pathologies are key components of the disease process. Symptomatic treatment offers a modest, clinically measurable effect in cognition, but disease-modifying therapies are desperately needed.

Handbook of Clinical Neurology, Vol. 167 (3rd series)
Geriatric Neurology
S.T. DeKosky and S. Asthana, Editors
https://doi.org/10.1016/B978-0-12-804766-8.00013-3
Copyright © 2019 Elsevier B.V. All rights reserved

V mozcích pacientů trpících Alzheimerovou chorobou se vyskytují dva různé proteinové agregáty: smotky chybně složeného cytoskeletárního proteinu *tau* a plaky z proteinových agregátů tvořených fragmenty membránového proteinu: β-amyloidového prekurzoru (β-APP). Přítomnost dvou druhů proteinových agregátů vedla ve vědecké obci k prudkým sporům o to, který z těchto proteinů je skutečnou příčinou onemocnění a který jen vedlejším příznakem. Zastánci jednotlivých teorií (posměšně nazývaní *tauisté* a *baptisté*) jdou ve svém nadšení pro oblíbenou teorii někdy tak daleko, že se v přednáškách na velkých vědeckých konferencích o protivném proteinu vůbec nezmiňují. Podle posledních prací (Science 293, 1487, 2001) se ale zdá, že oba mechanizmy jsou vzájemně provázány. Vzhledem k tomu, že autorem tohoto článku je přesvědčený *baptista*, neuslyšíte o proteinu *tau* v dalším textu už ani slovo.

- Huntingtonova chorea je autozomálně dominantní dědičné onemocnění. Defektní gen leží na 4. chromozomu. Postihuje osoby ve věku 20 až 50 let. Pacienti přežívají průměrně 15 let. Klinicky se projevuje mimovolními pohyby připomínajícími tanec (odtud *chorea*), psychotickými příznaky (halucinacemi, bludy) a postupující demencí. Takové klinické příznaky jsou důsledkem destrukce malých neuronů ve dvou částech mozku – v ocasatém jádře (*nucleus caudatus*) a skořápce (*putamen*).

Jan Konvalinka, Jana Peichlová | 1. 1. 2002 | Vesmír 81, 11, 2002/1

Agregace proteinů

- „Aggregation propensity prediction“
- Choroby způsobené mutovaným proteinem
- Predikce ze sekvence/ predikce ze struktury

In the cell, protein folding into stable globular conformations is in competition with aggregation into non-functional and usually toxic structures, since the biophysical properties that promote folding also tend to favor intermolecular contacts, leading to the formation of β-sheet-enriched insoluble assemblies. The formation of protein deposits is linked to at least 20 different human disorders, ranging from dementia to diabetes. Furthermore, protein deposition inside cells represents a major obstacle for the biotechnological production of polypeptides. Importantly, the aggregation behavior of polypeptides appears to be strongly influenced by the intrinsic properties encoded in their sequences and specifically by the presence of selective short regions with high aggregation propensity. This allows computational methods to be used to analyze the aggregation properties of proteins without the previous requirement for structural information. Applications range from the identification of individual amyloidogenic regions in disease-linked polypeptides to the analysis of the aggregation properties of complete proteomes. Herein, we review these theoretical approaches and illustrate how they have become important and useful tools in understanding the molecular mechanisms underlying protein aggregation.

Table 1. Online available computational methods to predict the aggregation properties of protein sequences

Method	Web	Reference
AGGRESCAN	http://bioinf.uab.es/aggrescan/	[19]
Betascan	http://betascan.csail.mit.edu/	[23]
FoldAmyloid	http://antares.protres.ru/fold-amyloid/	[29]
PASTA	http://biocomp.bio.unipd.it/pasta/	[32]
Waltz	http://waltz.switchlab.org/	[39]
ZipperDB	http://services.mbi.ucla.edu/zipperdb/	[45]
Zyggregator	http://www-vendruscolo.ch.cam.ac.uk/zyggregator.php	[49]
AmylPred	http://biophysics.biol.uoa.gr/AMYLPRED/input.html	[61]

Prediction of the aggregation propensity of proteins from the primary sequence: Aggregation properties of proteomes

Virginia Castillo, Ricardo Graña-Montes*, Raimon Sabate* and Salvador Ventura*

Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular and Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

Agregace proteinů

- „Aggregation propensity prediction“
- Choroby způsobené mutovaným proteinem
- Predikce ze sekvence/ predikce ze struktury

In the cell, protein folding into stable globular conformations is in competition with aggregation into non-functional and usually toxic structures, since the biophysical properties that promote folding also tend to favor intermolecular contacts, leading to the formation of β-sheet-enriched insoluble assemblies. The formation of protein deposits is linked to at least 20 different human disorders, ranging from dementia to diabetes. Furthermore, protein deposition inside cells represents a major obstacle for the biotechnological production of polypeptides. Importantly, the aggregation behavior of polypeptides appears to be strongly influenced by the intrinsic properties encoded in their sequences and specifically by the presence of selective short regions with high aggregation propensity. This allows computational methods to be used to analyze the aggregation properties of proteins without the previous requirement for structural information. Applications range from the identification of individual amyloidogenic regions in disease-linked polypeptides to the analysis of the aggregation properties of complete proteomes. Herein, we review these theoretical approaches and illustrate how they have become important and useful tools in understanding the molecular mechanisms underlying protein aggregation.

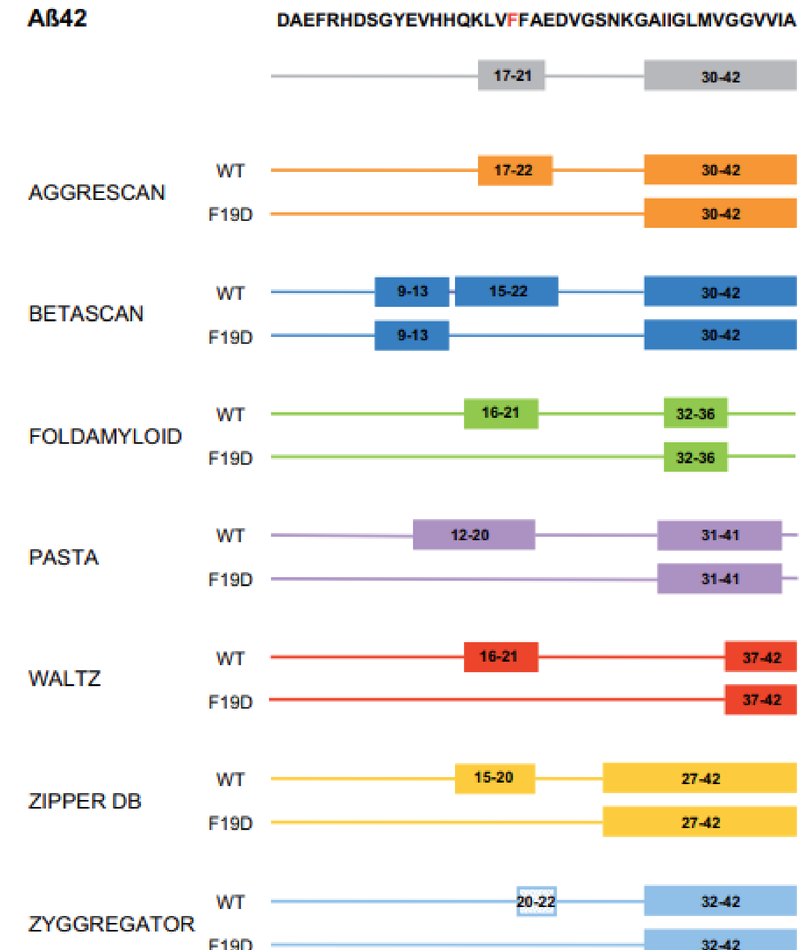
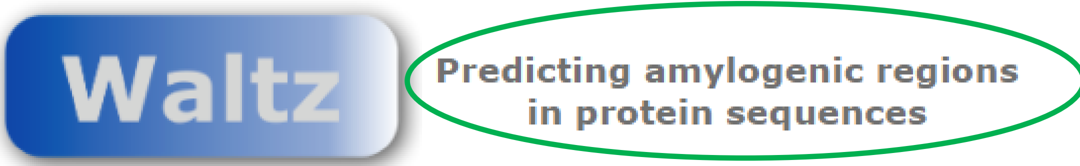


Figure 1. Comparison of the experimental and predicted aggregation-prone regions in two variants of the Alzheimer's disease related peptide Aβ42. Phe19 is shown in red in the sequence of the Aβ42 peptide. Changing this residue to Asp strongly reduces the aggregation propensity of the mutated variant (F19D), relative to the wild-type (WT) sequence [17]. Grey rectangles correspond to experimentally verified amyloidogenic sequences in the WT peptide. The numbers indicate the residues belonging to these regions. The different solid colored rectangles indicate aggregation-prone regions in the natural and mutated Aβ42 sequences, as predicted by the different algorithms discussed herein. The dashed rectangle corresponds to sequences with high aggregation propensity, but with values below the default threshold of the corresponding method.

Agregace proteinů



Sequence submission

Paste a single sequence or several sequences in FASTA format into the field below (max 10000 letters)

Threshold: Best Overall Performance / High Specificity (less false positives) / High Sensitivity / Custom : 0 [0-100]

pH: 7.0

Output Format: short text output

Submit sequences | Reset

Best Overall Performance / High Specificity (less false positives) / High Sensitivity / Custom : 0 [0-100]

7.0

detailed with graphics

short text output
detailed text output
detailed with graphics

Reset

https://waltz.switchlab.org/index.cgi

Protein 1:
SDVDIEAQDAGQTLVQVISIPSGETWVAIQLPSQYRYFDFVFENVSPTSSGSVLVA
QMAPQSGGVYGSNYSGSGWGNDLGGGGFYGYSEAKWMCLWPANRSGPSSKTGLYGT
CKLMNLNQSSAVPSVTSNLFAPTAYKNEPGYANVGGCCQKIRGLASSIQFAFALAG
GNVPQNTDTFNGGTIKVYGWN

Protein 2:
LVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVVSPGDAAQLGHNDSRLFTGLSPG
DQLHLRETALALRAEVSVLFIRFALKDAGIVAPIELEVRDAATAVPDADDLLHPSC
RPLKDHYWRSDVLAAGATTCTADFAVCDRDGTVSGYFRWETSIEIAGSQPDTKQPG
FKPSS

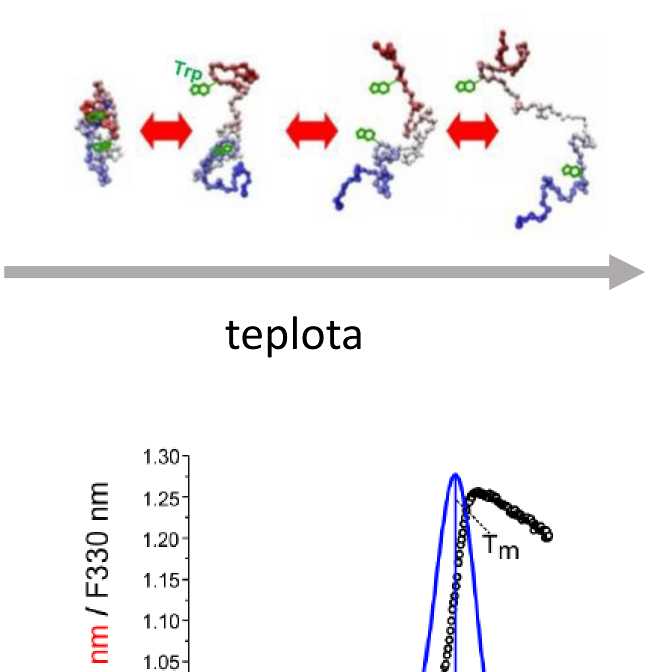
Protein 3:
PLLSASIVSAPVVTSETYVDIPGLYLDVAKAGIRDGKLQVILNVPTPYATGNNFPG
IYFAIATNQGVVADGCFTYSSKVPESTGRMPFTLVATIDVGSGVTFVKGQWKSVRG
SAMHIDSYASLSAIWGTAAPSSQGSGNQGAETGGTGAGNIGGGG

Protein 4:
ADSQTSSNRAGEFSIPPNTDFRAIFFANAAEQQHIKLFIGDSQEPAAYHKLTTRDG
PREATLNSGNGKIRFEVSVNGKPSATDARLAPINGKKSDGSPFTVNFGIVVSEDGH
DSDYNDGIVVLQWPIG

Úkol 12: Otestujte studované proteiny a predikujte, zda obsahují úseky náchylné k agregaci.

Jak pracovat s proteiny?

- **Práce za vhodné teploty**

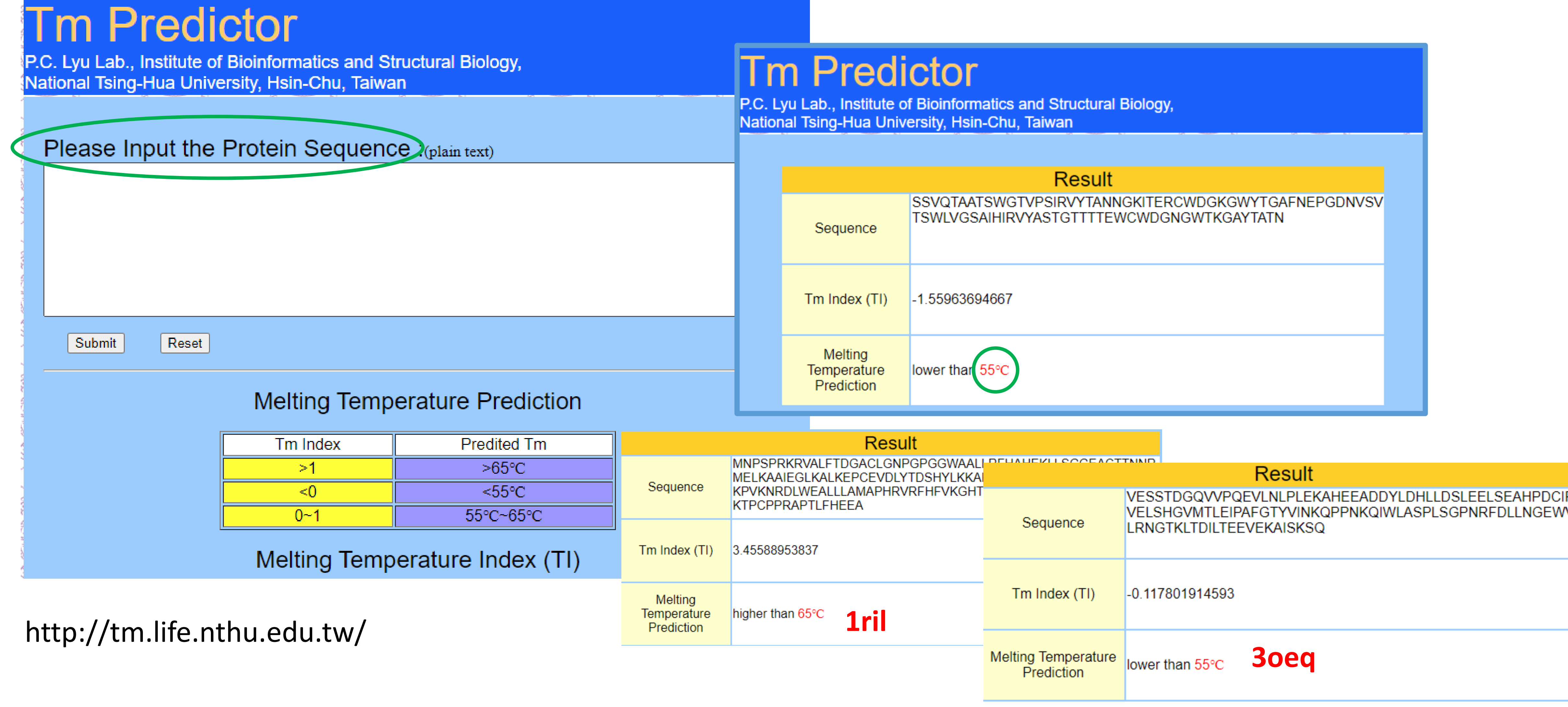


Tabulka VIII: Teploty denaturace RSL (1 várka) ve všech podmínkách „buffer screenu" získané metodou nanoDSF. Podmínka s nejvíce stabilizujícím efektem a s nejvyšší teplotou denaturace (D7) má tmavě zelenou barvu. Barvy ostatních podmínek se odvíjí od srovnání s podmínkou D7 podle škály ve spodní části tabulky. Na příklad podmínka B4 má bledě modrou barvu, protože se od podmínky D7 liší o 7 °C.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Voda 94,3	25 mM maleát, pH 2,0 72,9	50 mM glycin, pH 3,0 84,2	50 mM formiát, pH 4,0 92,6	50 mM citrát, pH 5,0 90,7	50 mM kakodylát, pH 6,0 87,8	50 mM Hepes, pH 7,0 88,4	50 mM bicin, pH 8,0 84,1	50 mM CHES, pH 9,0 83,8	6,5 mM borát, pH 10,0 82,3	50 mM CAPS, pH 11,0 80,3	50 mM fosfát, pH 12,0 57,1
B	50 mM acetát, pH 4,0 92,7	50 mM acetát, pH 4,5 93,4	50 mM acetát, pH 5,0 92,8	50 mM MES, pH 5,5 90,2	50 mM MES, pH 6,0 92,4	50 mM MES, pH 6,5 89,5	50 mM Na fosfát, pH 7,0 83,6	50 mM K fosfát, pH 7,5 83,8	50 mM Tris, pH 8,0 86,4	50 mM Tris, pH 8,5 85,2	50 mM glycin, pH 9,0 83,9	50 mM glycin, pH 9,5 83,9
C	50 mM MES, 50 mM NaCl, pH 6,0 91,8	50 mM MES, 100 mM NaCl, pH 6,0 90,9	50 mM MES, 250 mM NaCl, pH 6,0 90,2	50 mM MES, 500 mM NaCl, pH 6,0 89,5	50 mM Na fosfát, 50 mM NaCl, pH 7,0 83,4	50 mM Na fosfát, 100 mM NaCl, pH 7,0 83,3	50 mM Na fosfát, 250 mM NaCl, pH 7,0 83,0	50 mM Na fosfát, 500 mM NaCl, pH 7,0 84,7	50 mM Tris, 50 mM NaCl, pH 8,0 85,0	50 mM Tris, 100 mM NaCl, pH 8,0 84,7	50 mM Tris, 250 mM NaCl, pH 8,0 84,1	50 mM Tris, 500 mM NaCl, pH 8,0 83,5
D	10 mM Tris, 75 mM NaCl, 50 uM CaCl2, pH 7,5 89,2	5 mM Hepes, 75 mM NaCl, 0,025% Tween 20, pH 7,5 86,0	6 mM fosfát, 68,5 mM NaCl, 1,35 mM KCl, pH 7,5 82,7	6 mM fosfát, 0,025% Tween 20, pH 7,5 84,7	100 mM imidazol, pH 7,5 86,5	0,025% Tween 20 95,4	2,5% glycerol 97,2	2,5 mM bME 94,4	2,5% DMSO 92,7	2,5 mM trehalosa 89,9	10 mM **arginin**, 10 mM **glutamin** 83,8	2,5 mM EDTA 84,3
	nejlepší stabilita	**rozdíl do 2 °C**	**rozdíl 2 - 4 °C**	**rozdíl 4 - 6 °C**	**rozdíl 6 - 8 °C**	**rozdíl 8 - 10 °C**	**rozdíl 10 - 15 °C**	**rozdíl 15+ °C**				

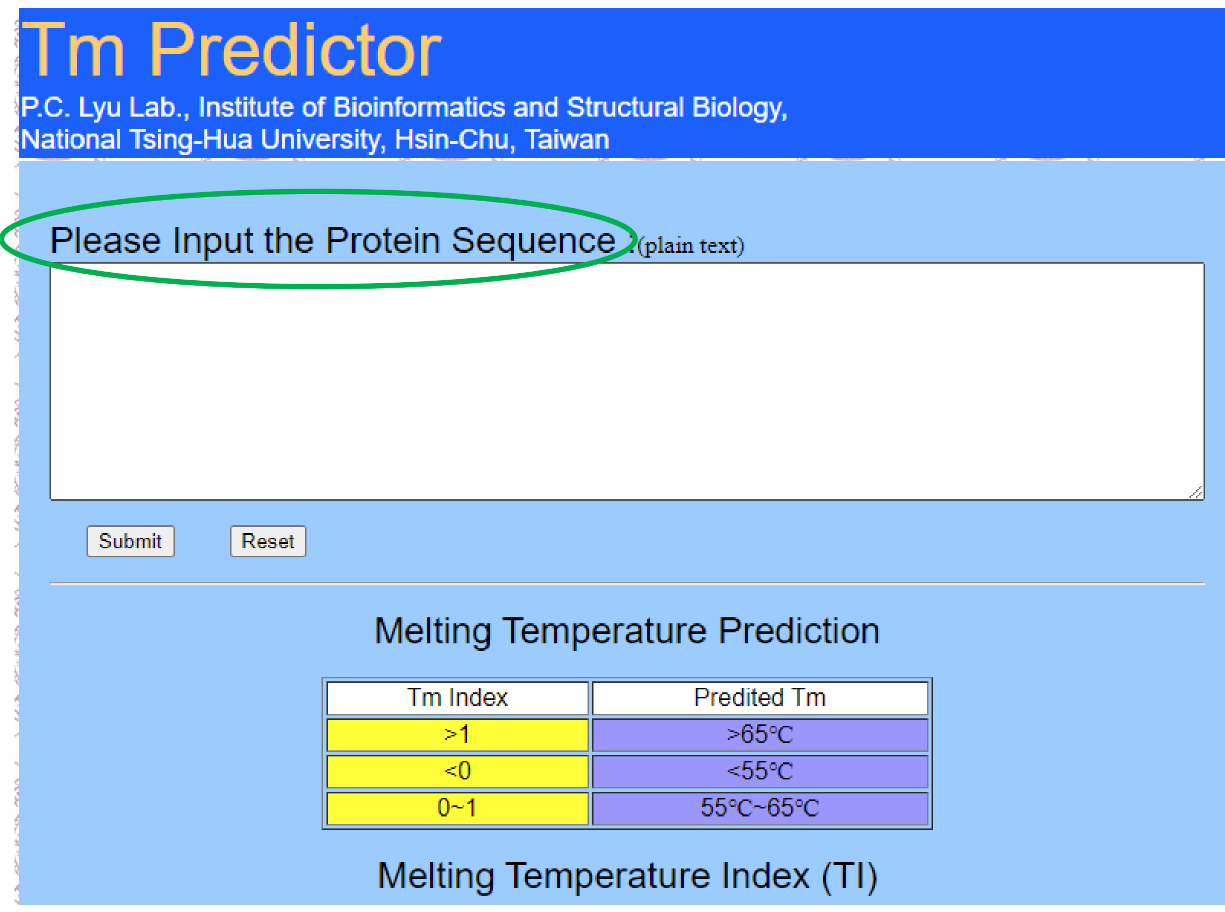
- Nejlepší podmínky vybíráme a stabilitu proteinu často posuzujeme s využitím **Tm** (melting temperature, teplota tání, 50% proteinu je denaturováno).

Predikce Tm



http://tm.life.nthu.edu.tw/

Predikce Tm



http://tm.life.nthu.edu.tw/

Protein 1:
ATQGVFTLPANTRFGVTAFANSSGTQTVNVLVNNETAATFSGQSTNNAVIGT
QVLNSGSSGKVQVQVSVNGRPSDLVSAQVILTNELNFALVGSEDGTDNDYND
AVVVINWPLG

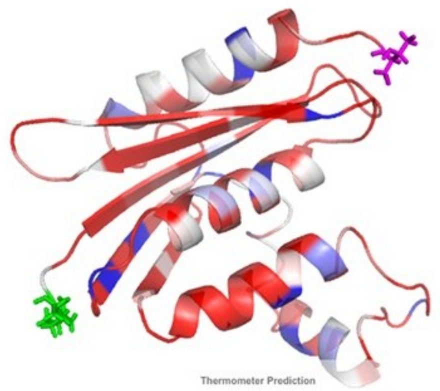
Protein 2:
ADSQTSSNRAGEFSIPPNTDFRAIFFANAAEQQHIKLFIGDSQEPAAYHKLT
TRDGPREATLNSGNGKIRFEVSVNGKPSATDARLAPINGKKSDGSPFTVNFG
IVVSEDGHDSDYNDGIVVLQWPIG

Protein 3:
LVIVDAVTLLSAYPEASRDPAAPTVIDGRHLYVVSPGDAAQLGHNDSRLFTG
LSPGDQLHLRETALALRAEVSVLFIRFALKDAGIVAPIELEVRDAATAVPDA
DDLLHPSCRPLKDHYWRSDVLAAGATTCTADFAVCDRDGTVSGYFRWETSIE
IAGSQPDTKQPGFKPSSDRNGNFSLPPNTAFKAIFYANAADRQDLKLFIDDA
PEPAATFVGNSEDGVRLFTLNSKGGKIRIEASANGRQSATDARLAPLSAGDT
VWLGWLGAEDGADADYNDGIVILQWPIT

Úkol 13: Predikujte Tm uvedených proteinů.

(Protein 1: termostabilní; Protein 2: „normální“; Protein 3: moc nevydrží)

Termostabilita



Portal Home

Thermostability predictor

[Thermostability predictor home - Documentation - Tutorial - Licensing Notice - Group

http://service.tartaglialab.com/update_submission/551650/

„The Thermometer web server is fast and for a ***50-residue-long protein****, results are available to the user in about* ***4 minutes****, while for a protein of average size (****250 resides****) the waiting time is of* ***nearly 20 minutes****. Moreover, the webserver is user-friendly and can be run without any a priori knowledge on theoretical or computational biology. We believe Thermometer can contribute to better understanding thermal stability and we hope that it could be useful in a number of practical applications.“*

Submission reference: 551650

Submission label (optional)

Email address (optional, used for notification)

PDB identifier (e.g. 3K8Y.pdb / 3K8Y / 3k8y)

1uzv

There is another version available that allows submission via file upload.

Submission status: Form not complete. Validate submission

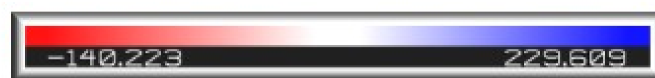
Send submission Submit

Sample data

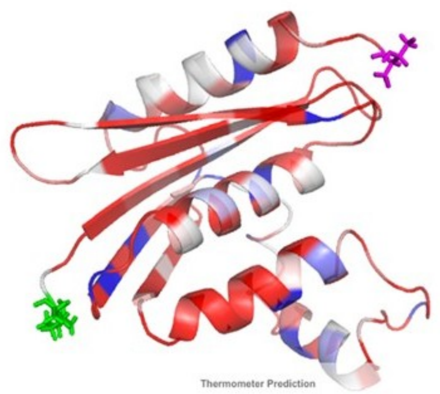
Just trying it out? Pre-populate the form with: (**Please, check that you do not have active proxies!**)

1ril - Thermostable example - Ribonuclease H from thermus thermophilus HB8

3oeq - Mesostable example - Trimeric frataxin from the yeast Saccharomyces cerevisiae



The weighted average of the 10 closest proteins is: **58.31**



Termostabilita

Portal Home

Thermostability predictor

[Thermostability predictor home - Documentation - Tutorial - Licensing Notice - Group page @ IIT]

http://service.tartaglialab.com/update_submission/551650/28c833ad26

„The Thermometer web server is fast and for a ***50-residue-long protein****, results are available to the user in about* ***4 minutes****, while for a protein of average size (****250 resides****) the waiting time is of* ***nearly 20 minutes****. Moreover, the webserver is user-friendly and can be run without any a priori knowledge on theoretical or computational biology. We believe Thermometer can contribute to better understanding thermal stability and we hope that it could be useful in a number of practical applications.“*

Submission reference: 551650

Submission label (optional)

Email address (optional, used for notification)

PDB identifier (e.g. 3K8Y.pdb / 3K8Y / 3k8y)

1uzv

There is another version available that allows submission via file upload.

Submission status: Form not complete. Validate submission

Send submission Submit

Sample data

Just trying it out? Pre-populate the form with: (**Please, check that you do not have active proxies!**)

1ril - Thermostable example - Ribonuclease H from thermus thermophilus HB8

3oeq - Mesostable example - Trimeric frataxin from the yeast Saccharomyces cerevisiae

3oeq

Úkol 13: Vyzkoušejte si predikci termostability pro tyto dva proteiny.

„Aliphatic index“ Termostabilita

Aliphatic index

Aliphatic index
The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins.

- Termostabilní proteiny – využití v průmyslu
- Termostabilní proteiny – proteiny z extremofilních organismů, proteinové inženýrství
- Využití bioinformatiky – predikce stabilizujících mutací
- Srovnání s proteiny z termofilních organismů
- **Sekvenční/strukturní úroveň**

Studies revealed the differences between amino acid and di-peptide composition in thermophilic and mesophilic proteins. For example, the frequency of Lys, Arg, Glu and Pro was higher in thermophilic than mesophilic protein [8,10]. These studies also show that the occurrences of EE, KK, RR, PP, KI, VV, VE, KE, and VK were higher in thermophilic proteins while QQ, AA, EQ, LL, NN, QT had lower occurrences [6]. In addition, the frequency of charged, hydrophobic and aromatic amino acids in thermophilic protein is higher than mesophilic ones [3]. Moreover, the correlation between protein amino acid composition and its biological function has been proven [1]. So, the protein sequence analysis provides valuable information to predict protein thermostability; particularly whenever the structural information of proteins is not available.

An Overview of the Protein Thermostability Prediction: Databases and Tools

Review: Engineering of thermostable enzymes for industrial applications

Termostabilita

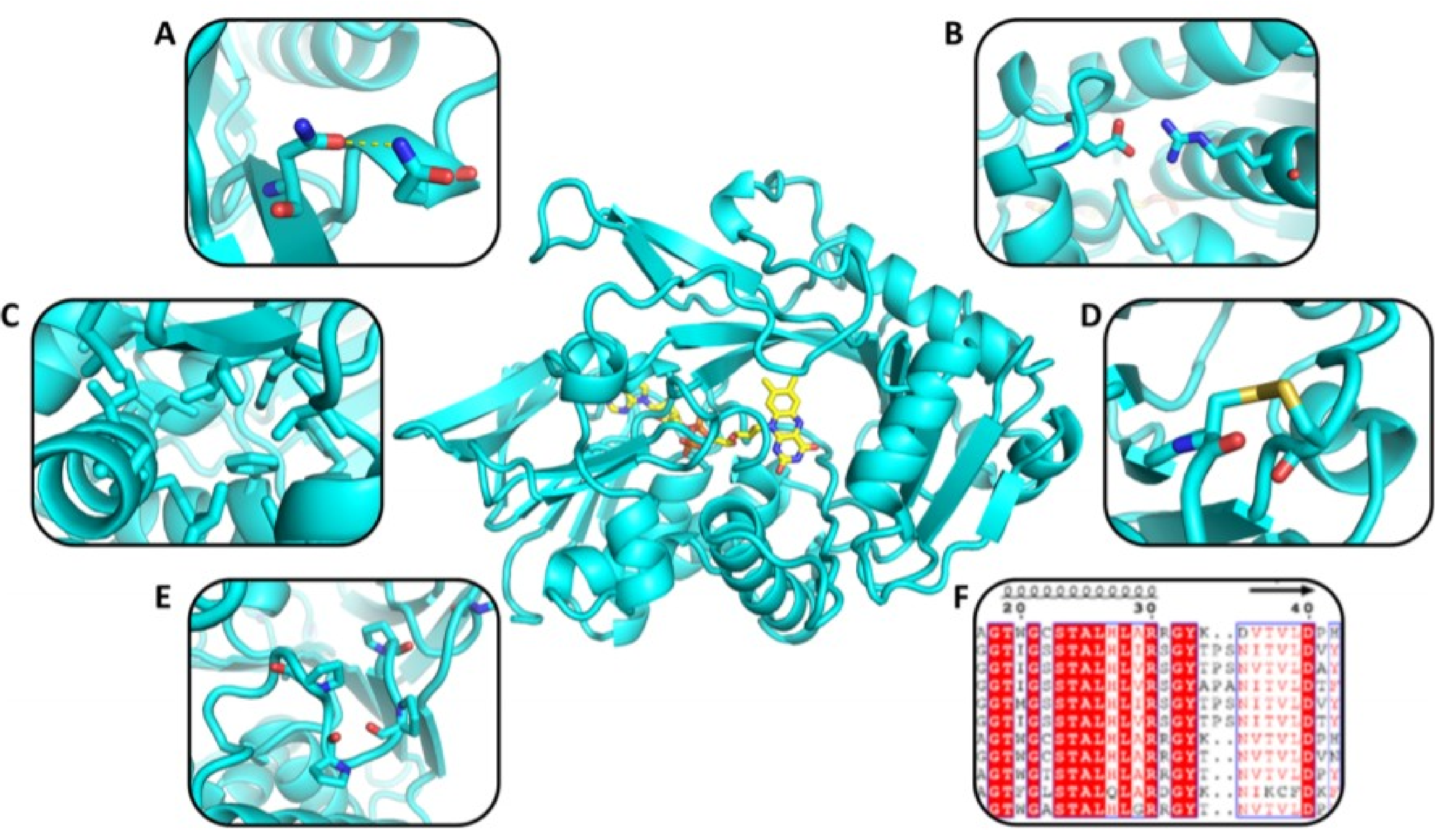
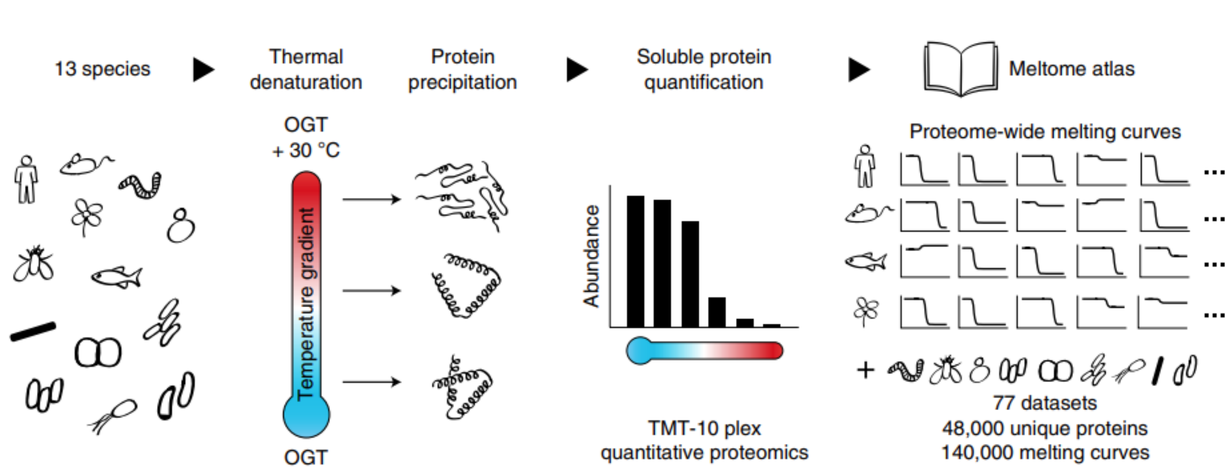


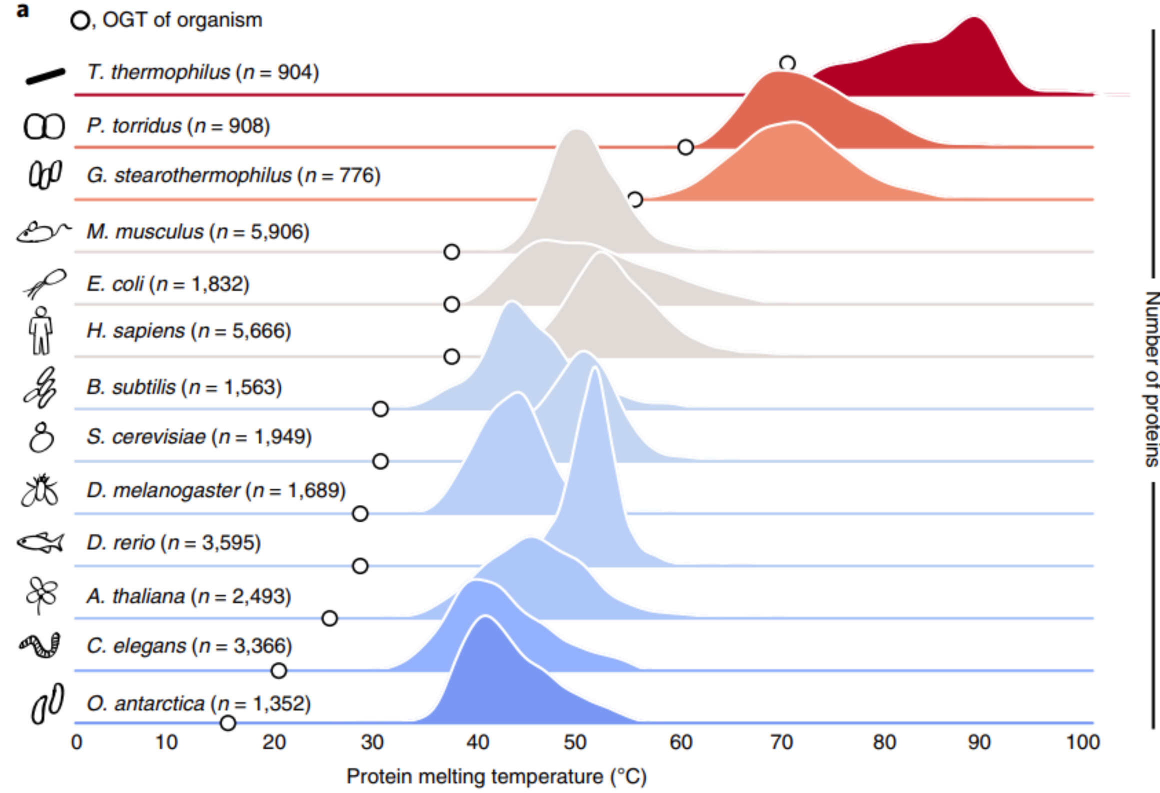
FIG. 2. Stabilizing strategies used in rational enzyme design. The most common strategies involve the introduction of surface hydrogen bonds (a) and salt bridges (b), the stabilization of the hydrophobic core (c), the introduction of disulfide bridges (d), and the stabilization of mobile loops using prolines (e). Phylogenetic analysis (f) can be used alone or in combination with previous strategies to guide the enzyme rational design process.

Meltome atlas—thermal proteome stability across the tree of life

„Meltome“



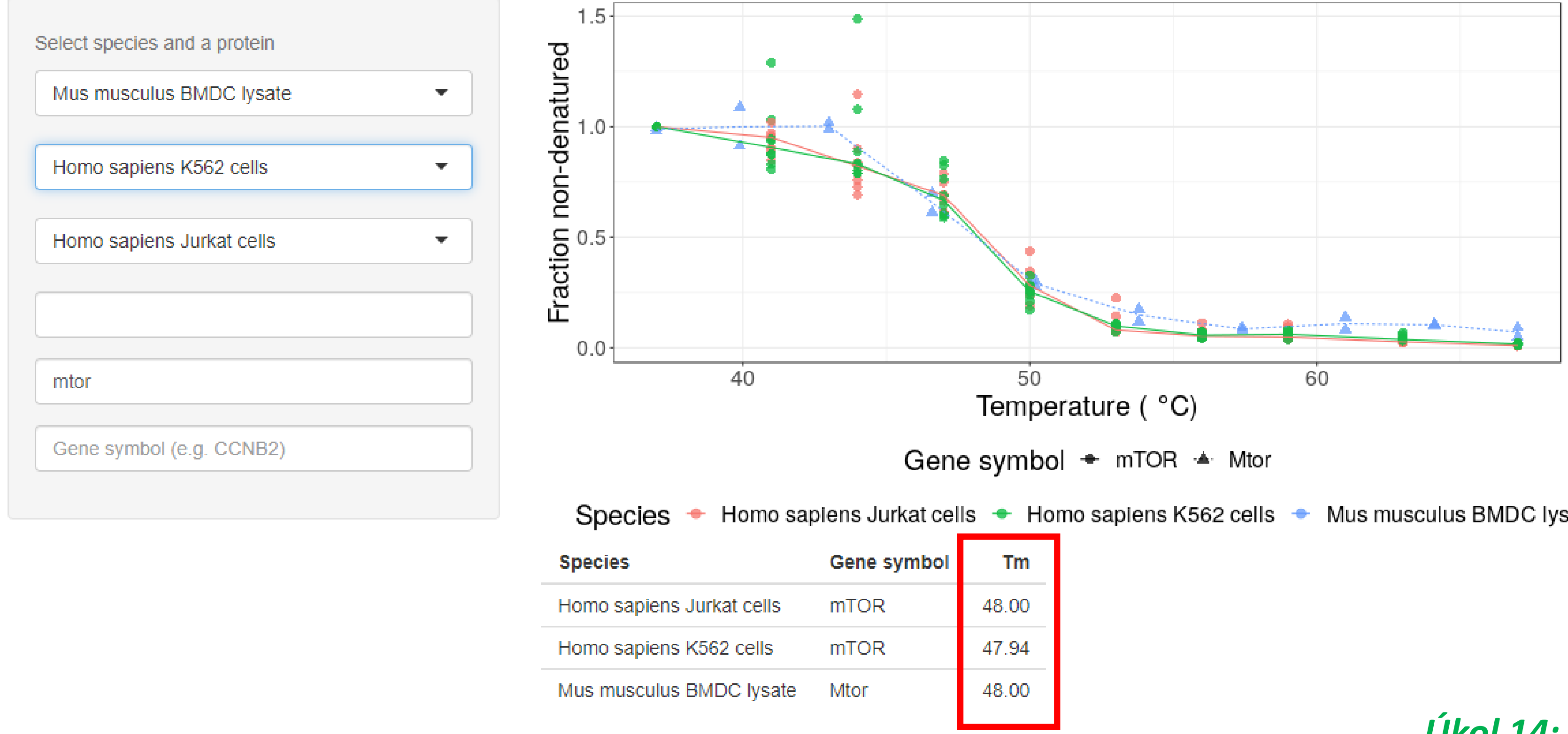
https://meltomeatlas.proteomics.wzw.tum.de/master_meltomeatlasapp/



We have used a mass spectrometry-based proteomic approach to compile an atlas of the thermal stability of 48,000 proteins across 13 species ranging from archaea to humans and covering melting temperatures of 30–90 °C. Protein sequence, composition and size affect thermal stability in prokaryotes and eukaryotic proteins show a nonlinear relationship between the degree of disordered protein structure and thermal stability. The data indicate that evolutionary conservation of protein complexes is reflected by similar thermal stability of their proteins, and we show examples in which genomic alterations can affect thermal stability. Proteins of the respiratory chain were found to be very stable in many organisms, and human mitochondria showed close to normal respiration at 46 °C. We also noted cell-type-specific effects that can affect protein stability or the efficacy of drugs. This meltome atlas broadly defines the proteome amenable to thermal profiling in biology and drug discovery and can be explored online at http://meltomeatlas.proteomics.wzw.tum.de:5003/ and http://www.proteomicsdb.org.

Meltome atlas—thermal proteome stability across the tree of life

„Meltome"



Species	Gene symbol	Tm
Homo sapiens Jurkat cells	mTOR	48.00
Homo sapiens K562 cells	mTOR	47.94
Mus musculus BMDC lysate	Mtor	48.00

https://meltomeatlas.proteomics.wzw.tum.de/master_meltomeatlasapp/

Úkol 14: Porovnejte Tm proteinu GAPDH u člověka a myši.

GRAVY

Grand average of hydropathy

GRAVY (Grand Average of Hydropathy)
The GRAVY value for a peptide or protein is calculated as the sum of hydropathy values [9] of all the amino acids, divided by the number of residues in the sequence.

Amino acid scale values:

Ala: 1.800	Gly: -0.400	Pro: -1.600
Arg: -4.500	His: -3.200	Ser: -0.800
Asn: -3.500	Ile: 4.500	Thr: -0.700
Asp: -3.500	Leu: 3.800	Trp: -0.900
Cys: 2.500	Lys: -3.900	Tyr: -1.300
Gln: -3.500	Met: 1.900	Val: 4.200
Glu: -3.500	Phe: 2.800	

J. Mol. Biol. (1982) **157**, 105–132

A Simple Method for Displaying the Hydropathic Character of a Protein

JACK KYTE AND RUSSELL F. DOOLITTLE

Hydrofobní/hydrofilní proteiny?

Membránové proteiny?

GRAVY

Hydrofobní/ hydrofilní proteiny?

Amino Acid	Kyte-Doolittle	Hopp-Woods
Alanine	1.8	-0.5
Arginine	-4.5	3.0
Asparagine	-3.5	0.2
Aspartic acid	-3.5	3.0
Cysteine	2.5	-1.0
Glutamine	-3.5	0.2
Glutamic acid	-3.5	3.0
Glycine	-0.4	0.0
Histidine	-3.2	-0.5
Isoleucine	4.5	-1.8
Leucine	3.8	-1.8
Lysine	-3.9	3.0
Methionine	1.9	-1.3
Phenylalanine	2.8	-2.5
Proline	-1.6	0.0
Serine	-0.8	0.3
Threonine	-0.7	-0.4
Tryptophan	-0.9	-3.4
Tyrosine	-1.3	-2.3
Valine	4.2	-1.5

Hydrofobní/**hydrofilní** proteiny?

Protein 1:

DPIALTAAVGADLLGDGRPETLWLGIGTLLMLIGTFYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVGSEMLDIYYARYADWLFTTPLLLLDLALLAKVDRVSIGTLVGV
DALMIVTGLVGALSHTPLARYTWWLFSTICMIVVLYFLATSLRAAAKERGPEVASTFNTLTALVLVLWTAYPILWIIGTEGAGVVGLGIETLLFMVLDVTAKVGFGFILLRSRAILGDTEAPEPS
AGAEASAAD Membránový protein

Protein 2:

KLAVYSTKQYDKKYLQQVNESFGFELEFFDFLLTEKTAKTANGCEAVCIFVNDDGSRPVLEELKKHGVKYIALRCAGFNNVDLDAAKELGLKVVRVPAYDPEAVAEHAIGMMMTLNRRIHRAYQR
TRDANFSLEGLTGFTMYGKTAGVIGTGKIGVAMLHILKGFGMRLLAFDPYPSAAALELGVEYVDLPTLFSESDVISLHCPLTPENYHLLNEAAFDQMKNGVMIVNTSRGALIDSQAAIEALKNQK
IGSLGMDVYENERDLFFEDKSNDVIQDDVFRRLSACHNVLFTGHQAFLTAEALTSISQTTLQNLSNLEKGETCPNELV Cytoplasmatický protein

Protein 1

Aliphatic index: 126.95

Grand average of hydropathicity (GRAVY): 0.812

Amino acid composition:

Amino acid		Count	%
Ala	(A)	31	11.9%
Arg	(R)	9	3.5%
Asn	(N)	1	0.4%
Asp	(D)	12	4.6%
Cys	(C)	1	0.4%
Gln	(Q)	1	0.4%
Glu	(E)	12	4.6%
Gly	(G)	26	10.0%
His	(H)	1	0.4%
Ile	(I)	19	7.3%
Leu	(L)	41	15.8%
Lys	(K)	5	1.9%
Met	(M)	7	2.7%
Phe	(F)	11	4.2%
Pro	(P)	9	3.5%
Ser	(S)	12	4.6%
Thr	(T)	23	8.8%
Trp	(W)	7	2.7%
Tyr	(Y)	10	3.8%
Val	(V)	22	8.5%
Pyl	(O)	0	0.0%
Sec	(U)	0	0.0%

Amino acid composition:

Amino acid		Count	%
Ala	(A)	30	9.1%
Arg	(R)	13	4.0%
Asn	(N)	18	5.5%
Asp	(D)	19	5.8%
Cys	(C)	6	1.8%
Gln	(Q)	11	3.3%
Glu	(E)	24	7.3%
Gly	(G)	23	7.0%
His	(H)	8	2.4%
Ile	(I)	14	4.3%
Leu	(L)	38	11.6%
Lys	(K)	19	5.8%
Met	(M)	10	3.0%
Phe	(F)	18	5.5%
Pro	(P)	9	2.7%
Ser	(S)	16	4.9%
Thr	(T)	19	5.8%
Trp	(W)	0	0.0%
Tyr	(Y)	11	3.3%
Val	(V)	23	7.0%
Pyl	(O)	0	0.0%
Sec	(U)	0	0.0%

Protein 2

Aliphatic index: 91.31

Grand average of hydropathicity (GRAVY): -0.103

Úkol 15: Ověřte, že se skutečně jedná o membránový a cytoplasmatický protein.

Hydrofobicita

50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification

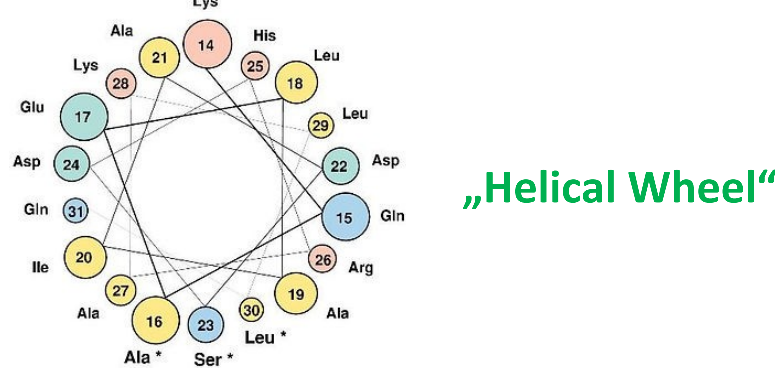
Stefan Simm[1], Jens Einloft[2], Oliver Mirus[1] and Enrico Schleiff[3*]

- Hydrofobicita aminokyselin – minimálně **98** (2016) stupnic hydrofobicity, velmi rozdílné (různé způsoby určení hodnot).
- Hydrofobicita využívána pro **predikci vlastností** proteinů: membránové proteiny/části proteinů, sekundární struktura, folding proteinů, rozpustnost.
- Různé parametry založené na hydrofobicitě aminokyselin (např. GRAVY).
- **Hydrofobní moment** – vzájemné prostorové rozložení hydrofilních a hydrofobních postranních řetězců jednotlivých aminokyselin v sekvenci, je závislý na sekundární struktuře daného **peptidu**.
- „Hydrofobní parametry“ – využívány pro popis a predikci vlastností **peptidů** (částí proteinu).
- Parametry související s interakcemi s membránami (antimikrobiální peptidy).

50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification

Stefan Simm[1], Jens Einloft[2], Oliver Mirus[1] and Enrico Schleiff[3*]

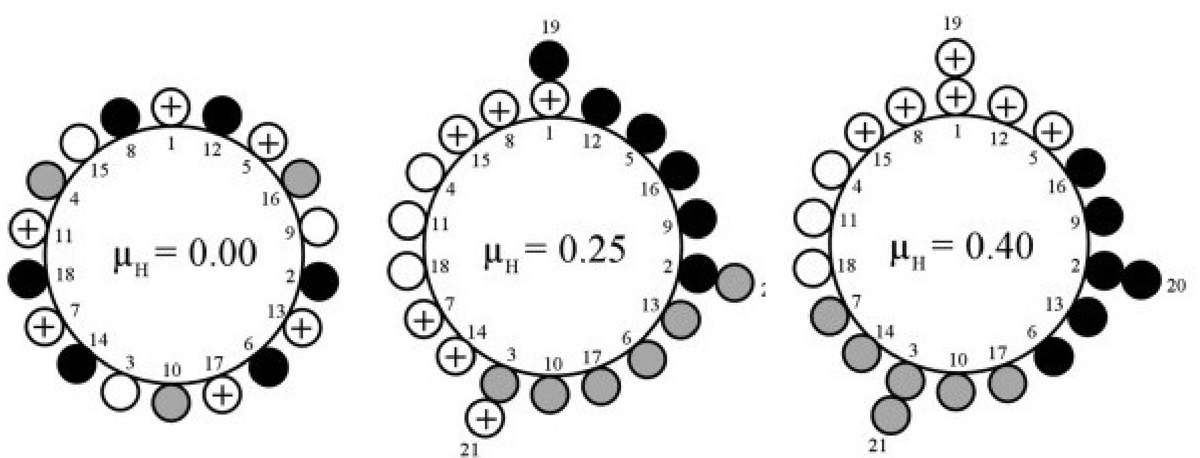
Hydrofobicita



„Helical Wheel“

- Hydrofobicita aminokyselin – minimálně **98** (2016) stupnic hydrofobicity, velmi rozdílné (různé způsoby určení hodnot).
- Hydrofobicita využívána pro **predikci vlastností** proteinů: membránové proteiny/části proteinů, sekundární struktura, folding proteinů, rozpustnost.
- Různé parametry založené na hydrofobicitě aminokyselin (např. GRAVY).

- **Hydrofobní moment** – vzájemné prostorové rozložení hydrofilních a hydrofobních postranních řetězců jednotlivých aminokyselin v sekvenci, je závislý na sekundární struktuře daného **peptidu**.



https://doi.org/10.1074/jbc.M102865200

50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification

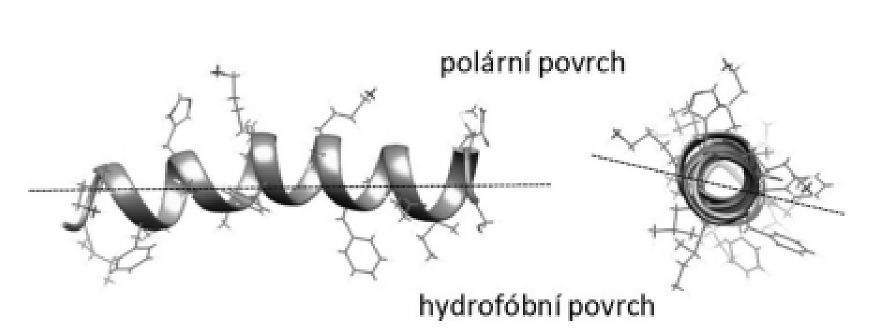
Stefan Simm[1], Jens Einloft[2], Oliver Mirus[1] and Enrico Schleiff[3*]

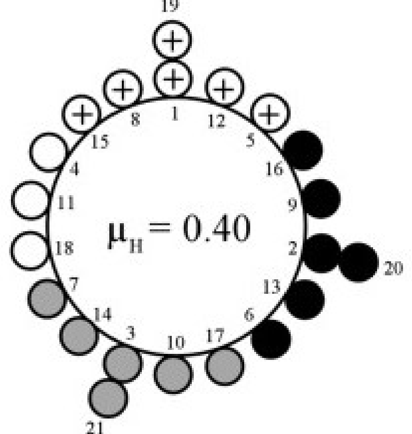
Hydrofobicita

„Helical Wheel“

- Hydrofobicita aminokyselin – minimálně **98** (2016) stupnic hydrofobicity, velmi rozdílné (různé způsoby určení hodnot).
- Hydrofobicita využívána pro **predikci vlastností** proteinů: membránové proteiny/části proteinů, sekundární struktura, folding proteinů, rozpustnost.
- Různé parametry založené na hydrofobicitě aminokyselin (např. GRAVY).

- **Hydrofobní moment** – vzájemné prostorové rozložení hydrofilních a hydrofobních postranních řetězců jednotlivých aminokyselin v sekvenci, je závislý na sekundární struktuře daného **peptidu**.



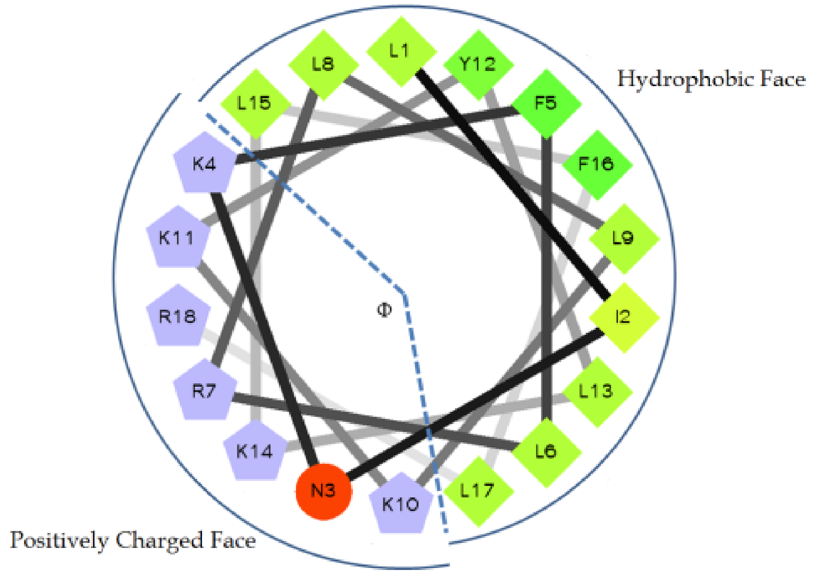


ANTIMIKROBIÁLNÍ PEPTIDY: NADĚJNÁ ZBRAŇ V BOJI PROTI ANTIBIOTICKÉ REZISTENCI

Ondřej Nešuta, Václav Čeřovský
Ústav organické chemie a biochemie AV ČR, v. v. i., skupina Antimikrobiální peptidy; ondrej.nesuta@uochb.cas.cz

DOI: 10.5772/38023

Hydrofobicita



FEBS Letters 403 (1997) 208–212 FEBS 18200

Hydrophobicity, hydrophobic moment and angle subtended by charged residues modulate antibacterial and haemolytic activity of amphipathic helical peptides

Margitta Dathe[a,*], Torsten Wieprecht[a], Heike Nikolenko[a], Liselotte Handel[a], W. Lee Maloy[b], Dorothy L. MacDonald[b], Michael Beyermann[a], Michael Bienert[a]

[a]*Institute of Molecular Pharmacology, Alfred-Kowalke-Strasse 4, 10315 Berlin, Germany*
[b]*Magainin Pharmaceuticals, Inc., 5110 Campus Drive, Plymouth Meeting, PA 19462, USA*

- **„Angle subtended by charged residues“**

„It was observed that analogues with large angles show greater antibacterial action, for example, MIC value against E. coli decreased to eight times when the angle is increased from 120° (Figure 2A) to 180° (Figure 2B). However, these analogues showed larger hemolytic activity than magainin 2.“

https://doi.org/10.3329/dujps.v20i3.59806

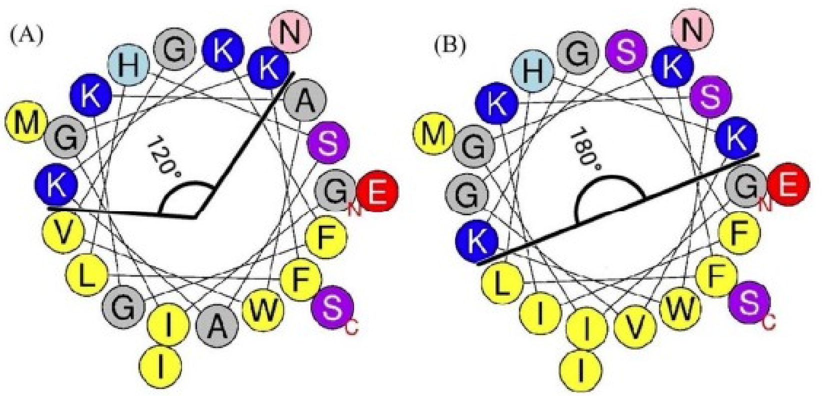
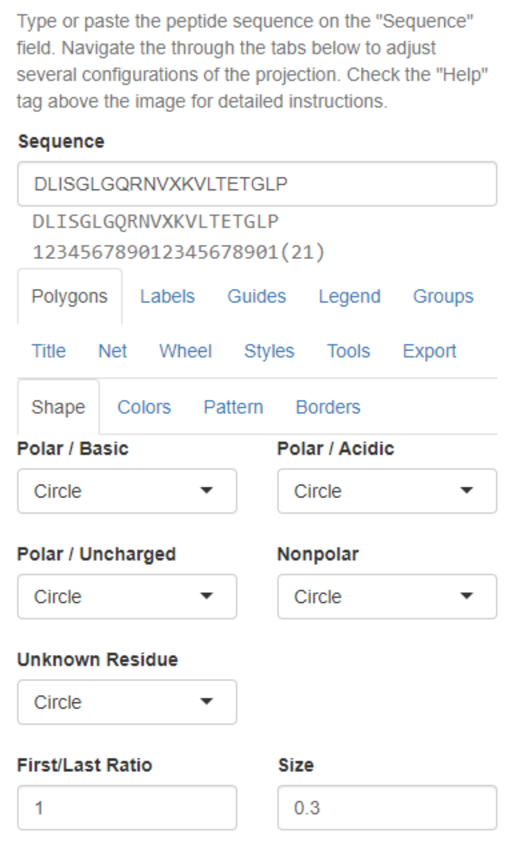
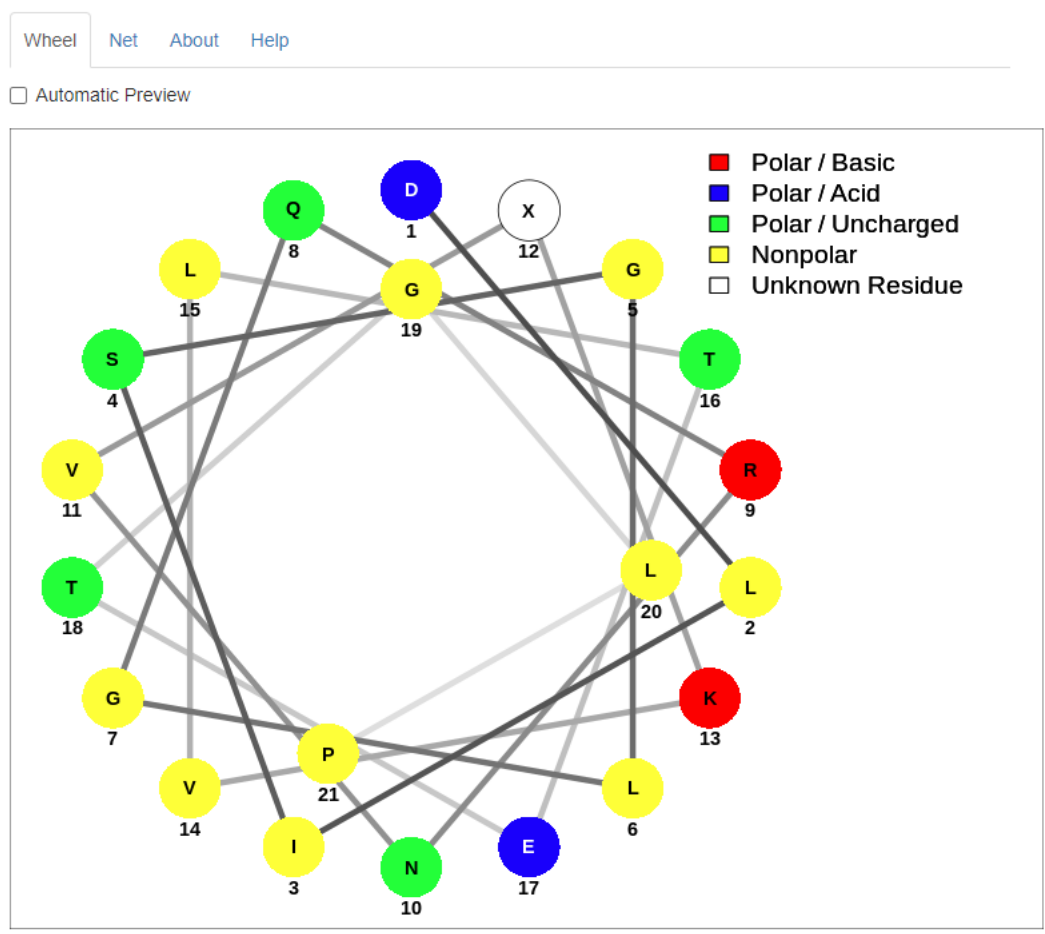


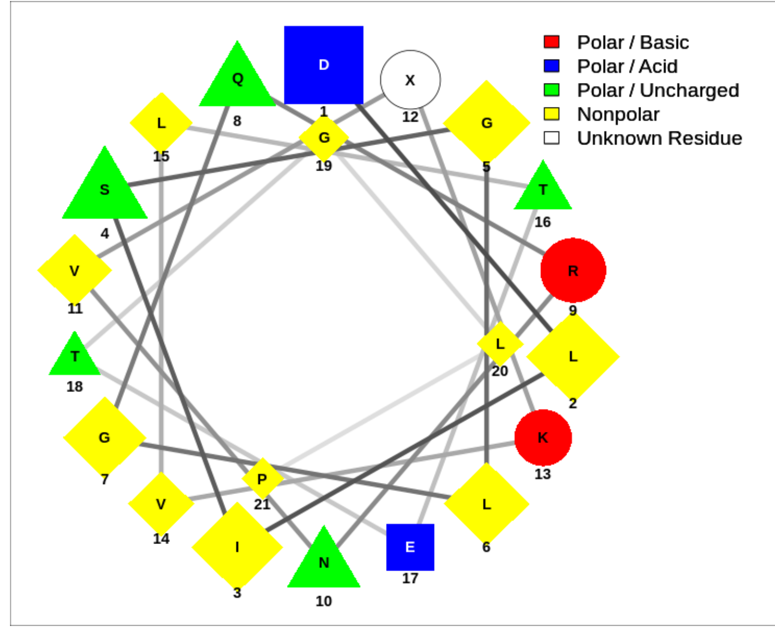
Figure 2. Angle comprised by the positively charged residues (four lysine represented by blue color and angle measured in degree) in the helical wheel projection of two magainin 2 analogues.

„Helical Wheel“

NetWheels: Peptides Helical Wheel and Net projections maker





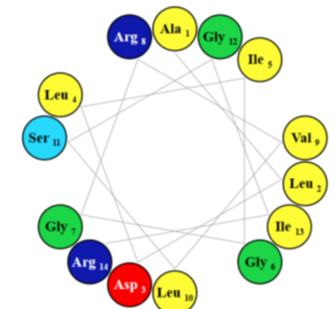


Úkol 16: Pohrajte si s nástrojem NetWheels. Vytvořte diagram pro peptidy:

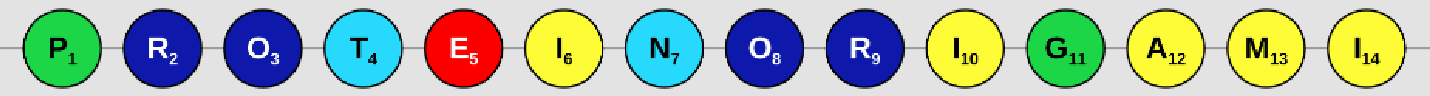
FLPLIGRVLSGIL
ALWKTLLKKVLKA

http://lbqp.unb.br/NetWheels/

„Helical Wheel“



https://www.ibg.kit.edu/protein_origami/?page=index



Home

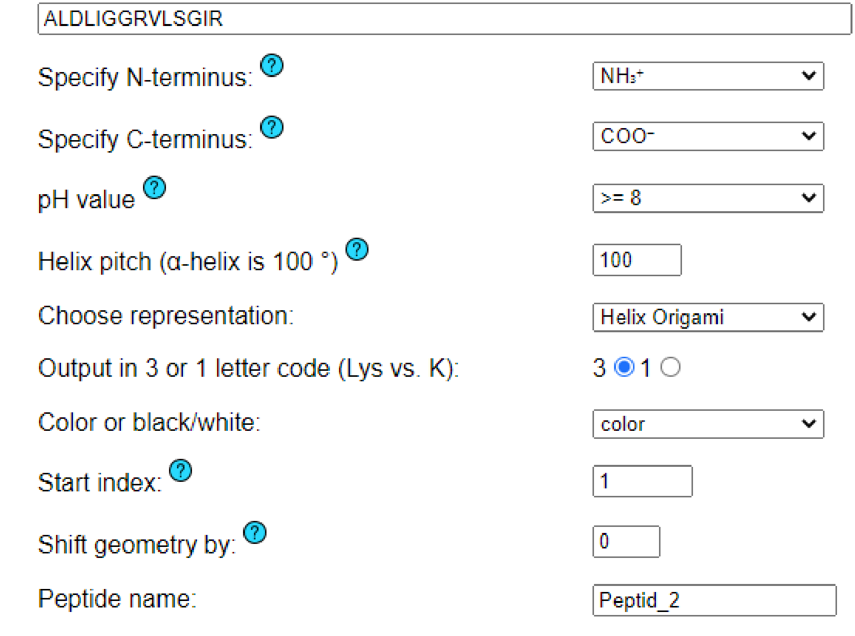
Submit Sequence

Reference

Manual

Contact

Legals



Create Protein ORIGAMI

Submit the sequence of the peptide as one-letter code (capital letters for L-form amino acids (default), small letters for D-form (use "Caps lock" or "Shift" key)).
If you have non-standard amino acids in your sequence, you can use the letters B, J, U, X or Z, and assign 3-letter codes and colors to them.

ALDLIGGRVLSGIR

Specify N-terminus: NH_3^+

Specify C-terminus: COO^-

pH value: >= 8

Helix pitch (α-helix is 100 °): 100

Choose representation: Helix Origami

Output in 3 or 1 letter code (Lys vs. K): 3 1

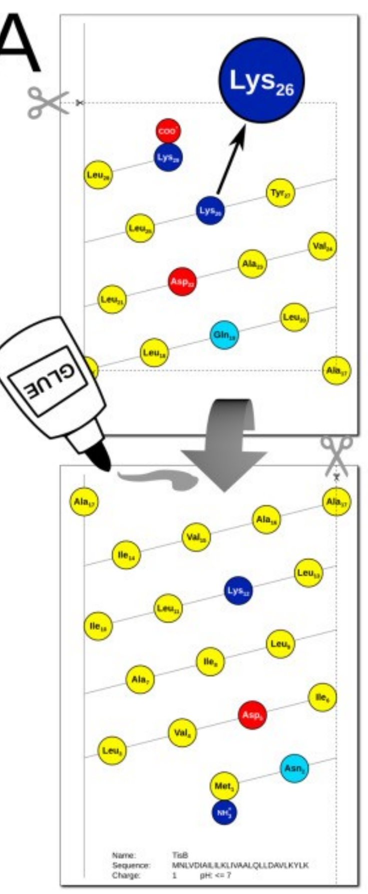
Color or black/white: color

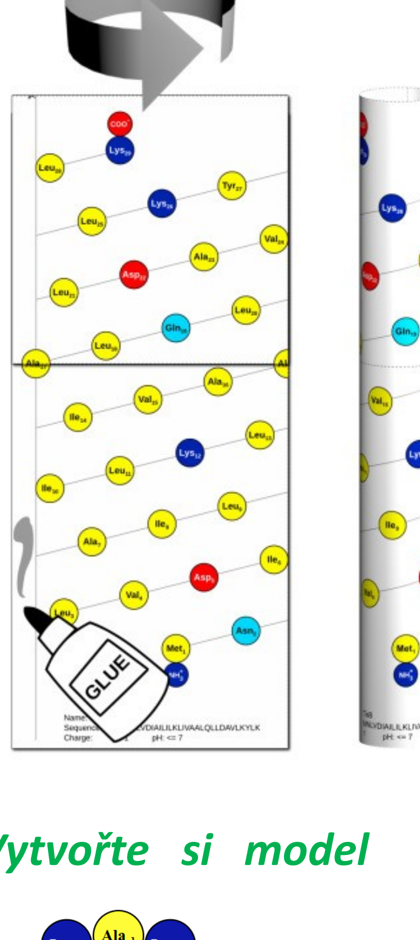
Start index: 1

Shift geometry by: 0

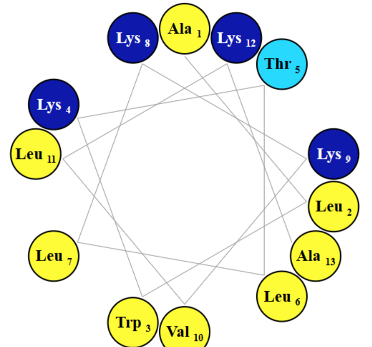
Peptide name: Peptid_2

Submit!





Volitelný úkol: Vytvořte si model peptidu…



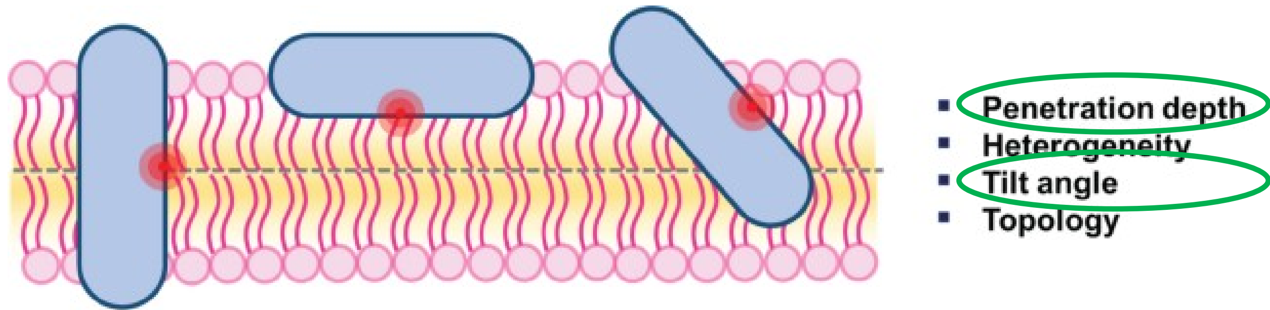
50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification

Stefan Simm[1], Jens Einloft[2], Oliver Mirus[1] and Enrico Schleiff[3*]

Hydrofobicita

- Hydrofobicita aminokyselin – minimálně **98** (2016) stupnic hydrofobicity, velmi rozdílné (různé způsoby určení hodnot).
- Hydrofobicita využívána pro **predikci vlastností** proteinů: membránové proteiny/části proteinů, sekundární struktura, folding proteinů, rozpustnost.
- Různé parametry založené na hydrofobicitě aminokyselin (např. GRAVY).

- Parametry související s interakcemi s membránami („penetration depth, tilt angle“).



https://doi.org/10.1007/s00232-022-00224-2

Proteom a proteomika

- Proteom – soubor všech forem proteinů existujících v buňce (organismu, biologickém vzorku) v určitém čase a za určitých podmínek.

  Množství kovalentních forem proteinů přesahuje množství proteinů predikovaných z DNA (genom).

  Proteomy jsou složitější než genomy.

  1 genom – mnoho proteomů.

- 1 gen může být exprimován ve více než 20 různých variantách proteinu. Například α1-antitrypsin se může vyskytovat ve 22 různých formách.

- 25 000 genů – 0,5 - 1 milion proteinů.

Proteom a proteomika

- Proteom – soubor všech forem proteinů existujících v buňce (organismu, biologickém vzorku) v určitém čase a za určitých podmínek.

  Množství kovalentních forem proteinů přesahuje množství proteinů predikovaných z DNA (genom).

  Proteomy jsou složitější než genomy.

  1 genom – mnoho proteomů.

- Navýšení kódovací kapacity genomu: alternativní sestřih, posttranslační modifikace (PTM).

Posttranslační modifikace

- Posttranslační modifikace – kovalentní modifikace proteinů po transkripci DNA a translaci RNA.

  Posttranslační modifikace – probíhají i u prokaryot

  Posttranslační modifikace jsou prováděny enzymy. Enzymy rozeznávají specifické signály – aminokyselinové sekvence v proteinech. Identifikace těchto sekvenčních motivů umožňuje predikci PTM.

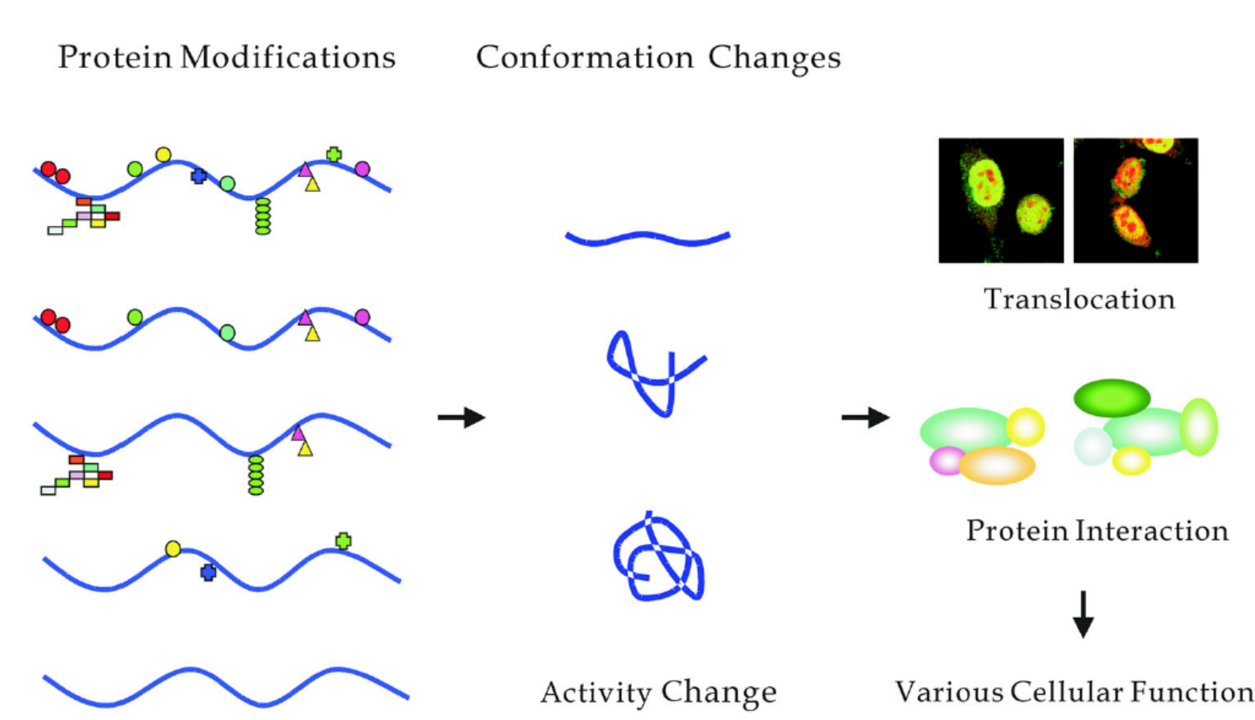
  Člověk: 500 proteinkinas, 150 proteinfosfatas, 500 proteas.

  5 % genomu vyšších eukaryot – zapojení do PTM.

- Klasifikace posttranslačních modifikací – typ modifikované aminokyseliny, podle modifikujícího enzymu, reverzibilita modifikací.

Posttranslační modifikace - význam

- Ovlivňují 3D a 4D strukturu proteinů, aktivitu a funkci (rozpustnost, stabilita, interakce, vypnuto/zapnuto).
- Mohou ovlivňovat lokalizaci proteinu v buňce (prenylace a jiné – připojení hydrofobní skupiny umožňuje lokalizaci do membrány).
- Tvorba disulfidických můstků může být nezbytná pro správné sbalení proteinů.
- Význam pro imunitní systém – glykosylace.



Post-translational Modifications and Their Biological Functions: Proteomic Analysis and Systematic Approaches

Jawon Seo and Kong-Joo Lee*

Posttranslační modifikace
Predikce

- Posttranslační modifikace jsou prováděny enzymy. Enzymy rozeznávají specifické signály – aminokyselinové sekvence v proteinech. Identifikace těchto sekvenčních motivů umožňuje predikci PTM.

- Problémy predikce:

- Může být těžké vytvořit „průměrný“ sekvenční motiv vhodný pro predikci.
- Proteiny jsou modifikovány různými enzymy s různou specifitou.
- Vliv okolních aminokyselin – ovlivnění náboje, hydrofility části proteinu v kontaktu s enzymem.
- Vliv 3D/4D struktury.

Posttranslační modifikace
Experimentální identifikace

- Problémy při identifikaci posttranslačních modifikací:

  modifikovaná je jen frakce proteinu (nízká koncentrace, nutné citlivé metody), kovalentní vazba PTM je labilní – nemusí vydržet zpracování vzorku a analýzu.

- Identifikace a analýza PTM: kombinace LC a MS analýzy.

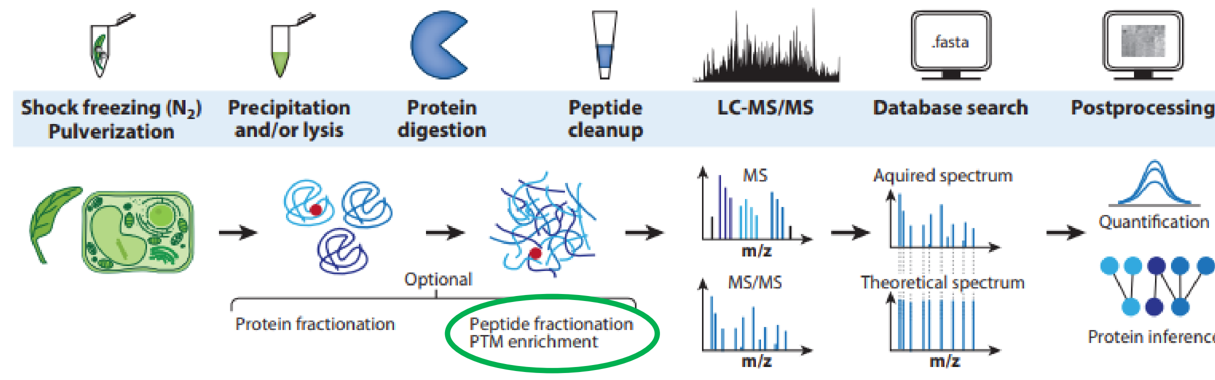


Figure 2

Generic workflow for proteomic sample preparation, data acquisition, and analysis. Proteins are extracted from whole-plant samples or enriched subfractions using optimized homogenization and protein extraction methods. Proteins are digested into peptides and measured on an LC-MS/MS system. Peptides are identified from MS/MS spectra by database matching and quantified based on the peak areas of intact peptide (survey MS spectra) or fragment ion signals (tandem MS spectra) provided by the mass spectrometer along the chromatographic time scale. Protein identity and quantity are then inferred based on peptide sequence assignments. Abbreviations: LC-MS/MS, liquid chromatography–coupled tandem mass spectrometry; MS, mass spectrometry; MS/MS, tandem mass spectrometry; m/z, mass-to-charge ratio; PTM, posttranslational modification.

Annual Review of Plant Biology

Plant Proteome Dynamics

Julia Mergner[1,2] and Bernhard Kuster[2,3]

[1]Bavarian Center for Biomolecular Mass Spectrometry at Klinikum rechts der Isar (BayBioMS@MRI), Technical University of Munich, Munich, Germany; email: Julia.mergner@tum.de

[2]Chair of Proteomics and Bioanalytics, Technical University of Munich, Freising, Germany; email: kuster@tum.de

[3]Bavarian Center for Biomolecular Mass Spectrometry (BayBioMS), Technical University of Munich, Freising, Germany

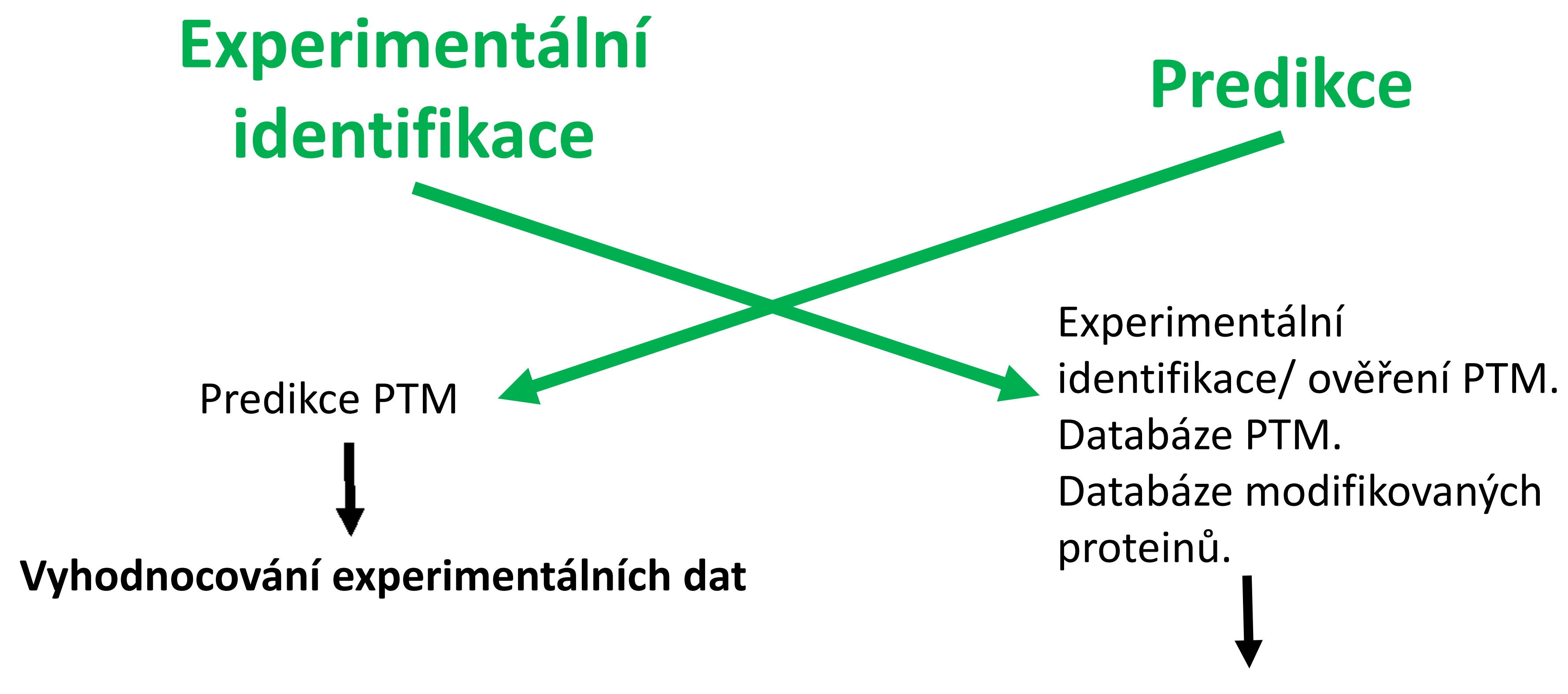
Annu. Rev. Plant Biol. 2022. 73:14.1–14.26

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

https://doi.org/10.1146/annurev-arplant-102620-031308

Copyright © 2022 by Annual Reviews.
All rights reserved

Posttranslační modifikace



Vyhodnocování experimentálních dat

Predikce PTM na základě homologie.
Predikce na základě sekvenčních motivů.

Post-translational modifications of proteins

DictyOGlyc	*O-(alpha)-GlcNAc glycosylation sites (trained on Dictyostelium discoideum proteins)*
NetAcet	*N-terminal acetylation in eukaryotic proteins*
NetCGlyc	*C-mannosylation sites in mammalian proteins*
NetCorona	*Coronavirus 3C-like proteinase cleavage sites in proteins*
NetGPI	*GPI Anchor predictions*
NetNGlyc	*N-linked glycosylation sites in human proteins*
NetOGlyc	*O-GalNAc (mucin type) glycosylation sites in mammalian proteins*
NetPhorest	*Linear motif atlas for phosphorylation-dependent signaling*
NetPhos	*Generic phosphorylation sites in eukaryotic proteins*
NetPhosBac	*Generic phosphorylation sites in bacterial proteins*
NetPhosYeast	*Serine and threonine phosphorylation sites in yeast proteins*
NetPhospan	*Prediction of phosphorylation using convolutional neural networks (CNNs).*
NetworKIN	*In vivo kinase-substrate relationships*
ProP	*Arginine and lysine propeptide cleavage sites in eukaryotic protein sequences*

DTU - Technical University of Denmark

https://services.healthtech.dtu.dk/

Swiss Bioinformatics Resource Portal

Sulfinator

Predict tyrosine sulfation sites in protein sequences

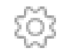
The Sulfinator is a software tool able to predict tyrosine sulfation sites in protein sequences. It employs four different Hidden Markov Models that were built to recognise sulfated tyrosine residues located N-terminally within sequence windows of more than 25 amino acids and C-terminally, as well as sulfated tyrosines clustered within 25 amino acid windows, respectively. All four HMMs contain the distilled information from one multiple sequence alignment.

PeptideCutter

Potential cleavage sites in a protein

GlycoMod

Possible oligosaccharide structures on proteins from masses



FindMod

Potential PTMs and single amino acid substitutions

Myristoylator

N-terminal myristoylation of proteins by neural networks.

Expert Protein Analysis System
http://www.expasy.org

Posttranslační modifikace - štěpení

PeptideCutter

Potential cleavage sites in a protein

https://web.expasy.org/peptide_cutter/

PeptideCutter

PeptideCutter [references / documentation] predicts potential cleavage sites cleaved by proteases or chemicals in a given protein sequence. PeptideCutter returns the query sequence with the possible cleavage sites mapped on it and /or a table of cleavage site positions.

Enter a UniProtKB (Swiss-Prot or TrEMBL) protein identifier, ID (e.g. ALBU_HUMAN), or accession number, AC (e.g. P04406), **or** an amino acid sequence (e.g. 'SERVELAT'):

```
PLLSASIVSAPVVTSETYVDIPGLYLDVAKAGIRDGKLQVILNVPTPYATG
NNFPGIYFAIATNQGVVADGCFTYSSKVPESTGRMPFTLVATIDVGSGVTF
VKGQWKSVRGSAMHIDSYASLSAIWGTAAPSSQGSGNQGAETGGTGAGNIG
GGG
```

Perform the cleavage of the protein. Reset the fields.

Protein 1

```
MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRY
PGQGSPGGNRYPPQGGGGWGQPHGGGWGQPHGGGWGQP
HGGGWGQPHGGGWGQGGGTHSQWNKPSKPKTNMKHMAG
AAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRE
NMHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVT
TTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQR
GSSMVLFSSPPV ILLISFLIFLIVG
```

PRIO_HUMAN

Úkol 17: Predikujte potenciální místa štěpení v uvedeném proteinu.

Použitá a doporučená literatura

Published online 28 November 2015 Nucleic Acids Research, 2016, Vol. 44, Database issue D27–D37
doi: 10.1093/nar/gkv1310

The SIB Swiss Institute of Bioinformatics' resources: focus on curated databases

Published online 31 May 2012 Nucleic Acids Research, 2012, Vol. 40, Web Server issue W597–W603
doi:10.1093/nar/gks400

ExPASy: SIB bioinformatics resource portal

Panu Artimo[1], Manohar Jonnalagedda[1,2], Konstantin Arnold[3], Delphine Baratin[4], Gabor Csardi[5], Edouard de Castro[4], Séverine Duvaud[4], Volker Flegel[1], Arnaud Fortier[1], Elisabeth Gasteiger[4], Aurélien Grosdidier[2], Céline Hernandez[1], Vassilios Ioannidis[1], Dmitry Kuznetsov[1], Robin Liechti[1], Sébastien Moretti[1,6], Khaled Mostaguir[4], Nicole Redaschi[4], Grégoire Rossier[1], Ioannis Xenarios[1,4,7] and Heinz Stockinger[1,*]

Improved Annotations of 23 Differentially Expressed Hypothetical Proteins in Methicillin Resistant *S. aureus*

Jessica Marklevitz[1] and Laura K. Harris[1,2*]

Protein Identification and Analysis Tools on the ExPASy Server

Elisabeth Gasteiger, Christine Hoogland, Alexandre Gattiker, Séverine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch

The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences

Bengt Bjellqvist
Graham J. Hughes
Christian Pasquali
Nicole Paquet
Florence Ravier
Jean-Charles Sanchez
Séverine Frutiger
Denis Hochstrasser

Departments of Medicine and Biochemistry, Medical Center of the University of Geneva

The focusing positions in narrow range immobilized pH gradients of 29 polypeptides of known amino acid sequence were determined under denaturing conditions. The isoelectric points of the proteins calculated from their amino acid sequences matched with good accuracy the experimentally determined p*I* values. We show the advantages of being able to predict the position of a protein of known structure within a two-dimensional gel.

Protein post-translational modifications: *In silico* prediction tools and molecular modeling

Martina Audagnotto *, Matteo Dal Peraro *
Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

How to measure and predict the molar absorption coefficient of a protein

C. NICK PACE,[1,2,3] FELIX VAJDOS,[2] LANETTE FEE,[2] GERALD GRIMSLEY,[1] AND THERONICA GRAY[1]
[1] Department of Medical Biochemistry and Genetics, [2] Department of Biochemistry and Biophysics, and [3] Center for Macromolecular Design, Texas A&M University, College Station, Texas 77843-1114

(Received July 12, 1995; Accepted September 8, 1995)

The N-end rule pathway of protein degradation

Alexander Varshavsky*
Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Protein Engineering vol.4 no.2 pp.155–161, 1990

Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence

A Simple Method for Displaying the Hydropathic Character of a Protein

JACK KYTE AND RUSSELL F. DOOLITTLE

FEBS Letters 403 (1997) 208–212 FEBS 18200

Hydrophobicity, hydrophobic moment and angle subtended by charged residues modulate antibacterial and haemolytic activity of amphipathic helical peptides

Margitta Dathe[a,*], Torsten Wieprecht[a], Heike Nikolenko[a], Liselotte Handel[a], W. Lee Maloy[b], Dorothy L. MacDonald[b], Michael Beyermann[a], Michael Bienert[a]
[a] *Institute of Molecular Pharmacology, Alfred-Kowalke-Strasse 4, 10315 Berlin, Germany*
[b] *Magainin Pharmaceuticals, Inc., 5110 Campus Drive, Plymouth Meeting, PA 19462, USA*

ANTIMIKROBIÁLNÍ PEPTIDY: NADĚJNÁ ZBRAŇ V BOJI PROTI ANTIBIOTICKÉ REZISTENCI

Ondřej Nešuta, Václav Čeřovský
Ústav organické chemie a biochemie AV ČR, v. v. i., skupina Antimikrobiální peptidy; ondrej.nesuta@uochb.cas.cz

Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications

Christopher T. Walsh, Sylvie Garneau-Tsodikova, and Gregory J. Gatto, Jr.*

Post-translational Modifications and Their Biological Functions: Proteomic Analysis and Systematic Approaches

Jawon Seo and Kong-Joo Lee*

Prediction of the aggregation propensity of proteins from the primary sequence: Aggregation properties of proteomes

Virginia Castillo, Ricardo Graña-Montes*, Raimon Sabate* and Salvador Ventura*
Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular and Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification

Stefan Simm[1], Jens Einloft[2], Oliver Mirus[1] and Enrico Schleiff[3*]