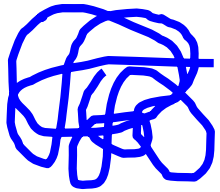
Predikce základních vlastností proteinů

C2131 Úvod do bioinformatiky, jaro 2024



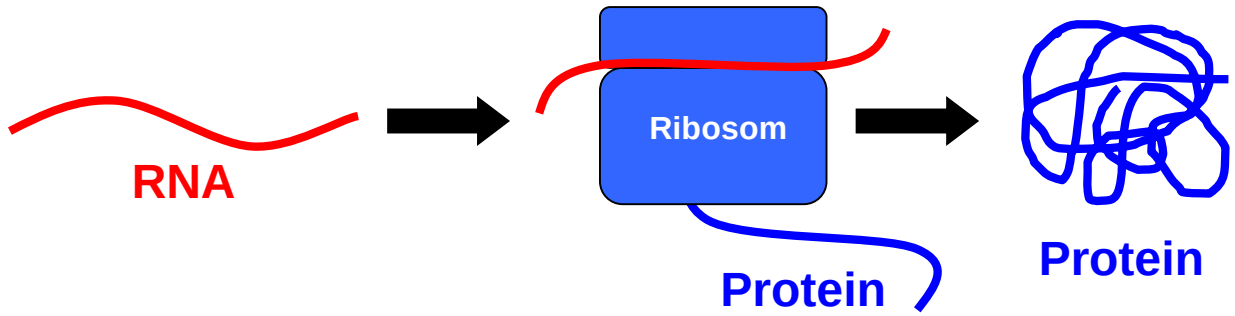
Protein

```
RKSTGGKAPRKQLATKAARKSAPATGGV
KKPHRYRPGTVALREIRRYQKSTELLIR
KLPFQRLVREIAQDFKTDLRFQSSAVMA
LQEASEAYLVGLFEDTNLCAIHAKR
```

→ **?**

Lenka Malinovská

Proteiny



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

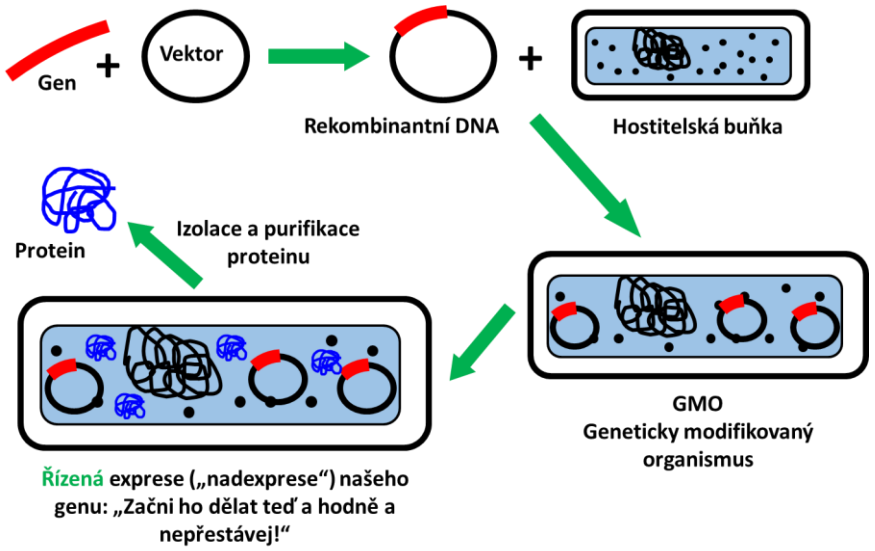
3) Proteiny. Složení, struktura, funkce proteinů. Proteinogenní aminokyseliny. Nestandardní proteinogenní aminokyseliny. Názvosloví aminokyselin, zkratky. Translace. Proteinové sekvence, databáze proteinů. Určování proteinové struktury, strukturní databáze proteinů.

Jak se chovají proteiny (v laboratoři)?

- **Limitované** množství proteinu (cena, dostupnost)

- Izolace **nativních** proteinů z přirozeného zdroje
- **Kom€rčně** dostupné proteiny
- Produkce rekombinantních proteinů v **hostitelském organismu**

Buňky různých tkání a orgánů



Price and Availability

SKU-Pack Size	Availability	Price (EUR/CZK)	Quantity
A2153-10G	Estimated Delivery 09.04.2013 - FROM	1,708.20	0
A2153-50G	Estimated Delivery 09.04.2013 - FROM	6,383.00	0
A2153-100G	Estimated Delivery 09.04.2013 - FROM	11,024.00	0
A2153-500G	Estimated Delivery 09.04.2013 - FROM	43,940.01	0
A2153-1KG	Estimated Delivery 09.04.2013 - FROM	66,040.00	0

Bulk orders?

ADD TO CART

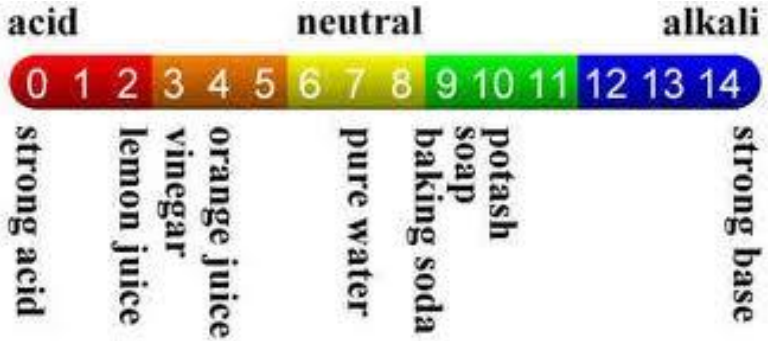
Jak se chovají proteiny (v laboratoři)?

- 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 (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).

...zlobí.

Jak se chovají proteiny (v laboratoři)?

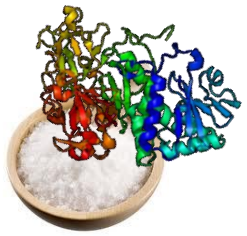
- Proteiny jsou aktivní (a stabilní) v určitém rozmezí pH.
  A to může být pro některé proteiny velmi úzké… Fyziologické pH pro většinu proteinů je cca 7,2-7,4. Silně kyselé nebo zásadité prostředí proteiny denaturuje.

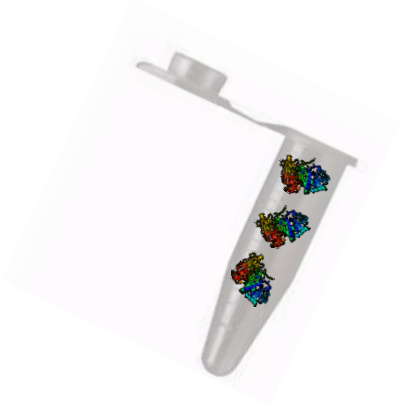


Pufr, tlumivý roztok, ústojný roztok, ústoj:
látka (směs látek) schopná udržovat stabilní pH po přídavku silné kyseliny nebo zásady do systému.
Příklad: slabá kyselina/její sůl, HA/A^-.
Pufry nesmí interagovat s proteiny nebo interferovat s jejich funkcí!

Jak se chovají proteiny (v laboratoři)?

- Proteiny vyžadují pro svou aktivitu (a stabilitu) určitou koncentraci „solí“ (iontová síla). 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!**

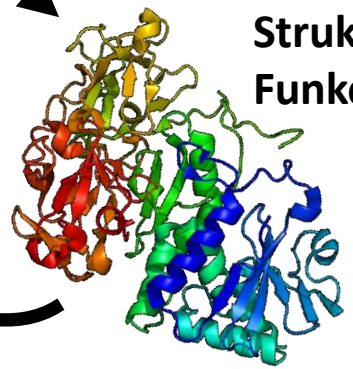
…opatrně.

Práce s proteiny

Protein je správně sbalený, aktivní, v dostatečném množství a koncentraci

Optimální podmínky

Gen → Protein

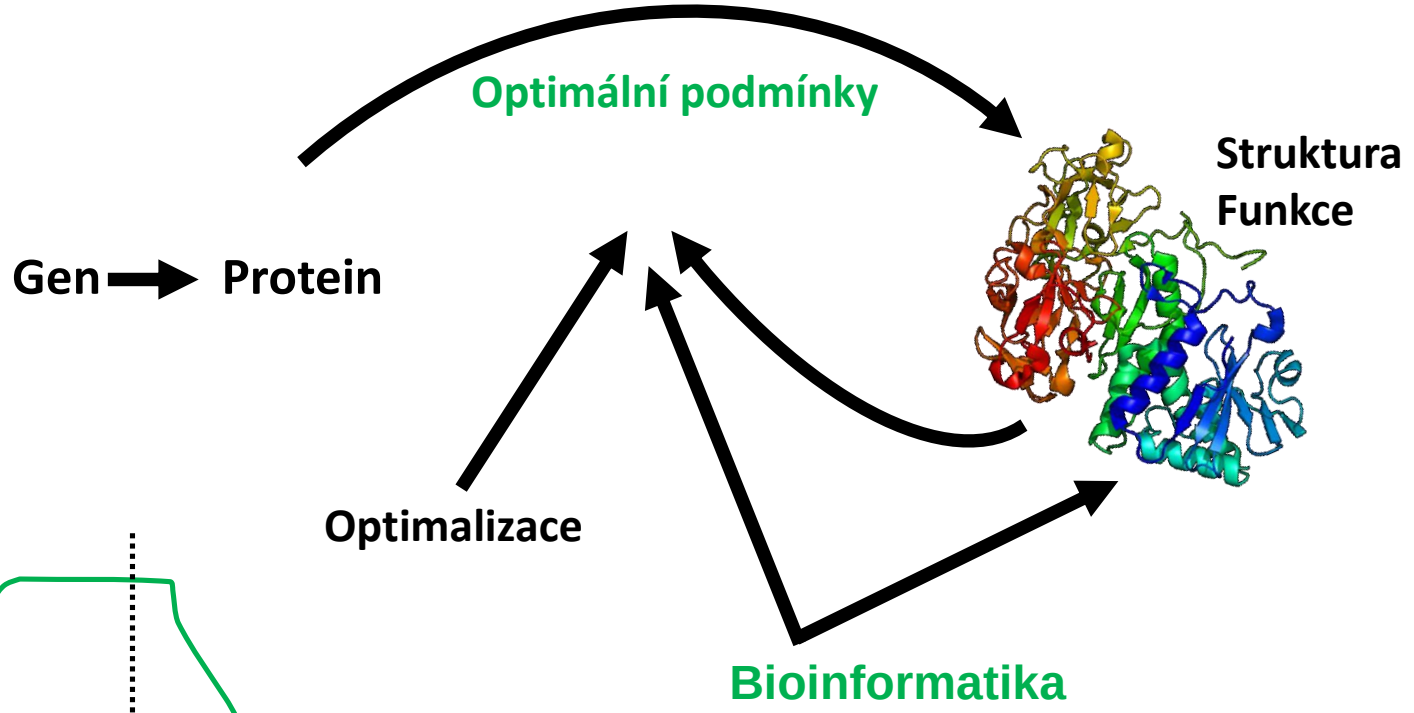


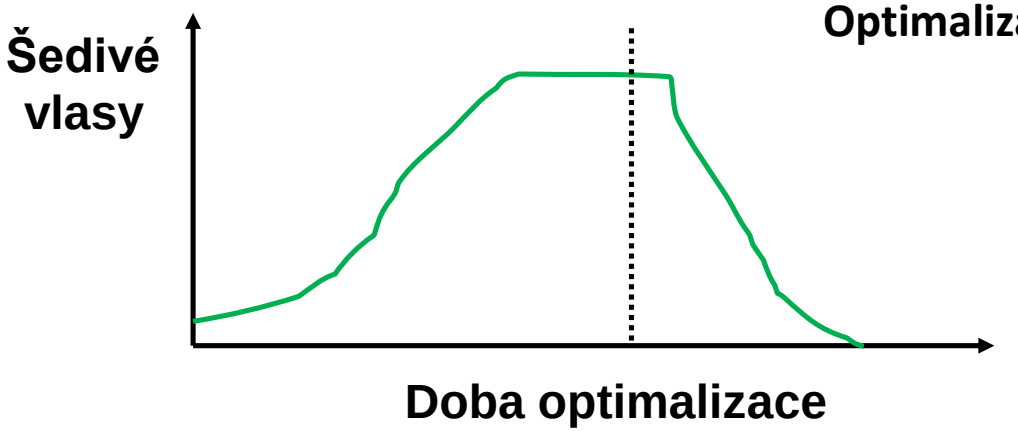
Struktura
Funkce

Hydrofobní interakce?
Sirné (disulfidové) můstky?
Oligomerizace?
Nutné kofaktory?
Přirozené prostředí (kompartment) v buňce?

Práce s proteiny

Protein je správně sbalený, aktivní, v dostatečném množství a koncentraci

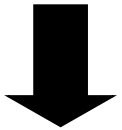





```
PLLSASIVSAPVVTSETYVD
IPGLYLDVAKAGIRDGKLQV
ILNVPTPY
```

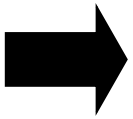
→ **Predikce vlastností a struktury**

CCTTTATTATCCGCTTCCATTGTTTCCGCTCCTGTTGTTACTTCCGAAACTTATGTTGATATTCCTGGTTTATATTTAGA
TGTTGCTAAAGCTGGTATTCGCGATGGTAAATTACAAGTTATTTTAAATGTTCCTACTCCTTATGCTACTGGTAATAATT
TTCCTGGTATTTATTTTGCTATTGCTACTAATCAAGGTGTTGTTGCTGATGGTTGTTTTACTTATTCCTCCAAAGTTCCT
GAATCCACTGGTCGCATGCCTTTTACTTTAGTTGCTACTATTGATGTTGGTTCCGGTGTTACTTTTGTTAAAGGTCAATG
GAAATCCGTTCGCGGTTCCGCTATGCATATTGATTCCTATGCTTCCTTATCCGCTATTTGGGGTACTGCTGCTCCTTCCT
CCCAAGGTTCCGGTAATCAAGGTGCTGAAACTGGTGGTACTGGTGCTGGTAATATTGGTGGTGGTGGTGAACGCGATGGT
ACTTTTAATTTACCTCCTCATATTAAATTTGGTGTTACTGCTTTAACTCATGCTGCTAATGATCAAACTATTGATATTTA
TATTGATGATGATCCTAAACCTGCTGCTACTTTTAAAGGTGCTGGTGCTCAAGATCAAAATTTAGGTACTAAAGTTTTAG
ATTCCGGTAATGGTCGCGTTCGCGTTATTGTTATGGCTAATGGTCGCCCTTCCCGCTTAGGTTCCCGCCAAGTTGATATT
TTTAAAAAATCCTATTTTGGTATTATTGGTTCCGAAGATGGTGCTGATGATGATTATAATGATGGTATTGTTTTTTTAAA



PLLSASIVSAPVVTSETYVDIPGLYLDVAKAGIRDGKLQVILNVPTPYATGNNFPGIYFA
IATNQGVVADGCFTYSSKVPESTGRMPFTLVATIDVGSGVTFVKGQWKSVRGSAMHIDSY
ASLSAIWGTAAPSSQGSGNQGAETGGTGAGNIGGGGERDGTFNLPPHIKFGVTALTHAAN
DQTIDIYIDDDPKPAATFKGAGAQDQNLGTKVLDSGNGRVRVIVMANGRPSRLGSRQVDI
FKKSYFGIIGSEDGADDDYNDGIVFL

Nukleotidová a proteinová sekvence hypotetických genů/proteinů



Predikce vlastností a struktury

Identifikován v transkriptomu klíštěte

Příklad

Predikce posttranslačních modifikací, lokalizace

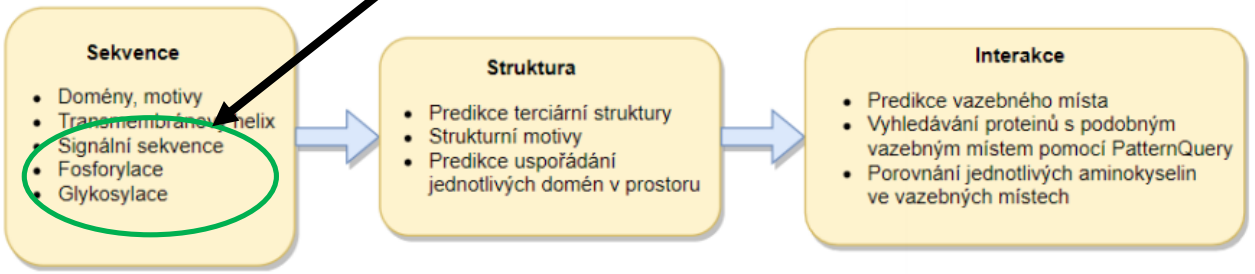
Charakterizace nového hypotetického lektinu z Ixodes ricinus

Characterization of new hypothetical lectin from Ixodes ricinus

Bc. Kateřina Bezděková, učo 433289

Abstrakt

Lektiny patří mezi významnou skupinu proteinů, které specificky interagují s různými monosacharidy či oligosacharidy. Jsou schopny rozpoznat i nepatrné rozdíly v sacharidové struktuře a hrají tak klíčovou roli v mnoha biologických procesech. Klíšťata se řadí mezi významné parazity, kteří přenáší různá onemocnění. Předpokládá se, že lektiny jsou u klíšťat součástí jejich vrozené imunity, zároveň však mohou být zapojeny do přenosu patogenů klíšťaty. Tato práce se věnuje charakterizaci hypotetického lektinu IrCLec z klíštěte *Ixodes ricinus*, jehož blízký homolog HlCLec hraje významnou roli v obraně vůči gramnegativním bakteriím. Pomocí bioinformatických nástrojů proběhla analýza sekvence IrCLec, byla predikována jeho struktura, funkce a potenciální ligandy jednotlivých vazebných míst.



Obrázek 7: Postup celkové analýzy hypotetického lektinu IrCLec.

SignalP (http://www.cbs.dtu.dk/services/SignalP)	Signální sekvence
Signal-3L (http://www.csbio.sjtu.edu.cn/bioinf/Signal-3L/)	Signální sekvence
NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc)	*N*-glykosylace
NetOGlyc (http://www.cbs.dtu.dk/services/NetOGlyc)	*O*-glykosylace
NetPhos 3.1 (http://www.cbs.dtu.dk/services/NetPhos/)	Fosforylace
SPDB (https://proline.bic.nus.edu.sg/spdb/)	Databáze signálních sekvencí
SignalPeptide (http://www.signalpeptide.de)	Databáze signálních sekvencí

10) Sacharidy a lipidy. Struktura, význam a funkce. Bioinformatický potenciál sacharidů. Glykoproteiny, jejich kódování v genomu.

Základní predikované vlastnosti proteinů

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

+

Lokalizace proteinů
Predikce posttranslačních modifikací

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 plně 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:

```
SKEPLRPRCRPINATLAVEKEGCPVCITVNTTICAGY
CPTMTRVLQGVLPALPQVVCNYRDVRFESIRLPGCPR
GVNPVVSYAVALSCQCALCRRSTTDCGGPKDHPLTCD
DPRFQDSSSSKAPPPSLPSPSRLPGPSDTPILPQ
```

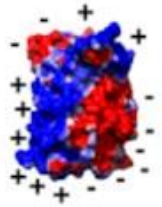
RESET | Compute parameters

ProtParam nebere v úvahu možné **posttranslační modifikace (PTM)** a **oligomerizaci** proteinů.

```
Number of amino acids: 145

Molecular weight: 15531.97

Theoretical pI: 8.65

Amino acid composition:  CSV format
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:

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

Formula: C668H1090N196O203S13
Total number of atoms: 2170
```

```
Number of amino acids: 936

Molecular weight: 69936.76

Theoretical pI: 5.09

Amino acid composition:  CSV format
Ala (A) 220      23.5%
Arg (R)   0       0.0%
Asn (N)   0       0.0%
Asp (D)   0       0.0%
Cys (C) 193      20.6%
Gln (Q)   0       0.0%
Glu (E)   0       0.0%
Gly (G) 420      44.9%
His (H)   0       0.0%
Ile (I)   0       0.0%
Leu (L)   0       0.0%
Lys (K)   0       0.0%
Met (M)   0       0.0%
Phe (F)   0       0.0%
Pro (P)   0       0.0%
Ser (S)   0       0.0%
Thr (T) 103      11.0%
Trp (W)   0       0.0%
Tyr (Y)   0       0.0%
Val (V)   0       0.0%
Pyl (O)   0       0.0%
Sec (U)   0       0.0%

 (B)   0          0.0%
 (Z)   0          0.0%
 (X)   0          0.0%
```

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

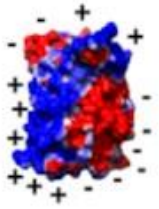
The focusing positions in narrow range immobilized pH gradients of 29 polypeptides of known amino acid sequence were determined under denaturing conditions. The isoelectric points of the proteins calculated from their amino acid sequences matched with good accuracy the experimentally determined p*I* 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



- **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)

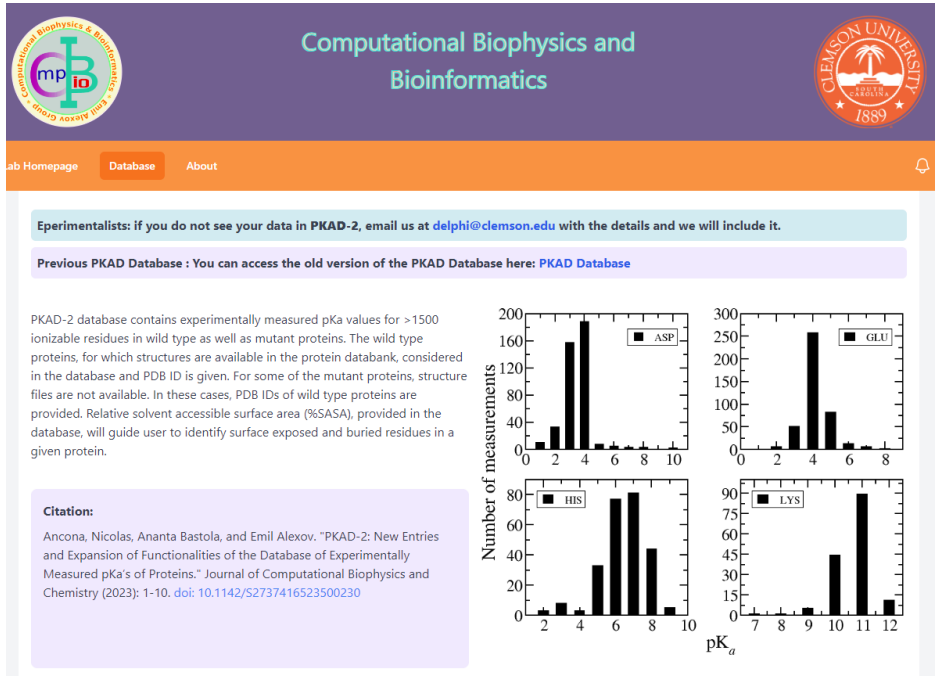


Table 1. Summary of experimentally measured 1350 residue-specific pKa values for wild-type proteins collected from literature

Residue ID	No. of measurements	Average pK_a	Lowest pK_a	Highest pK_a
ASP	408	3.43	0.5	9.9
GLU	417	4.14	2.1	7.2
HIS	253	6.45	<2.3	9.19
LYS	155	10.68	6.5	12.12
TYR	47	10.98	6.08	12.5
CYS	20	6.25	2.88	11.1
C-term	23	3.16	2.4	4.03
N-term	21	7.64	6.91	9.14

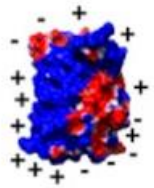
http://compbio.clemson.edu/PKAD-2/

Izoelektrický bod - pI

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

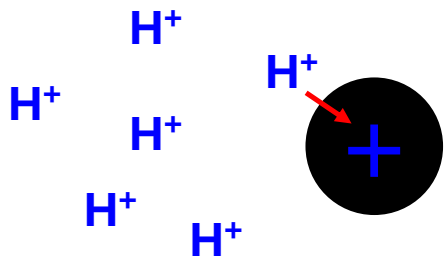
4.6 Isoelectric Point Precipitation

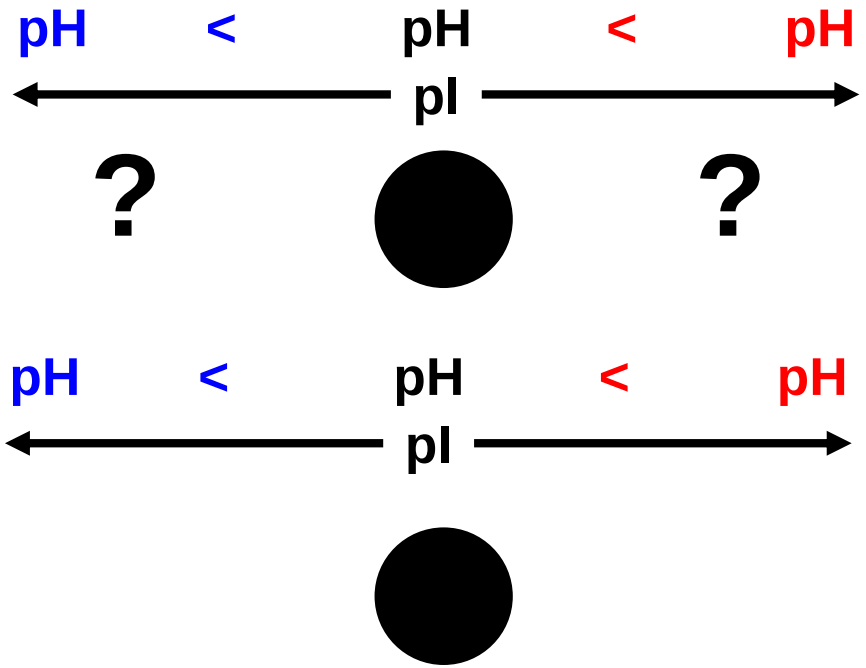
The isoelectric point (pI) is the pH of a solution at which the net charge of a protein becomes zero. At solution pH that is above the pI, the surface of the protein is predominantly negatively charged, and therefore like-charged molecules will exhibit repulsive forces. Likewise, at a solution pH that 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].

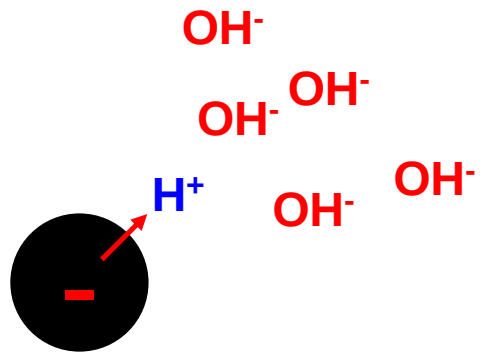


pH=2

protein protonated
positively charged



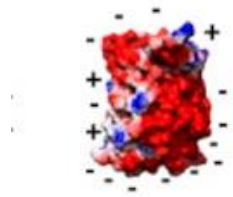




Using isoelectric point to determine the pH for initial protein crystallization trials

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.

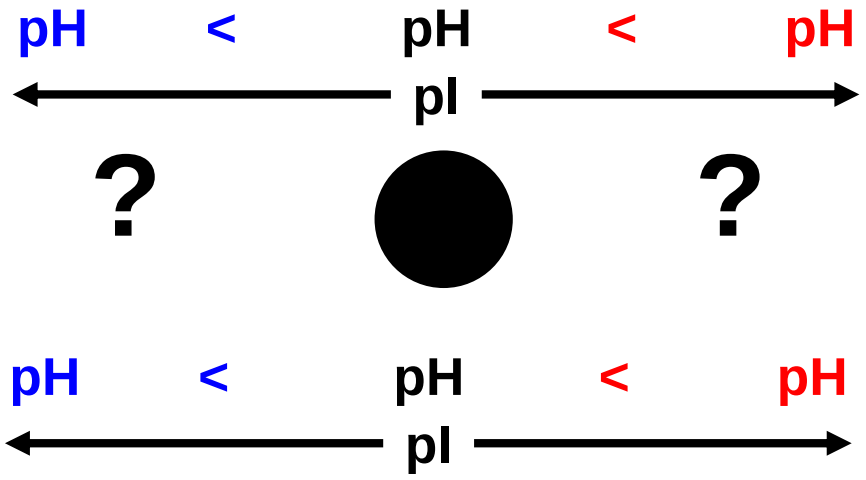


pH=10

protein deprotonated
negatively charged

Izoelektrický bod - pI

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



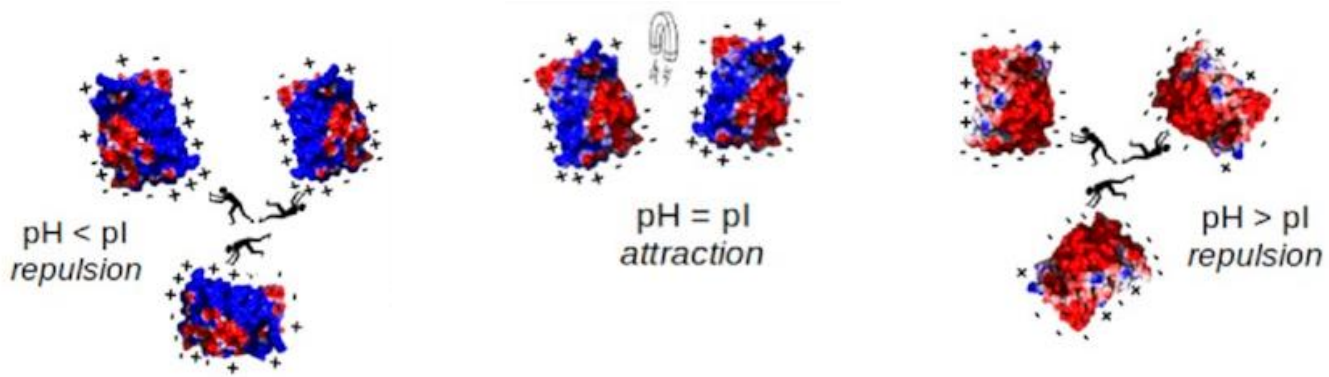
4.6 Isoelectric Point Precipitation

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].

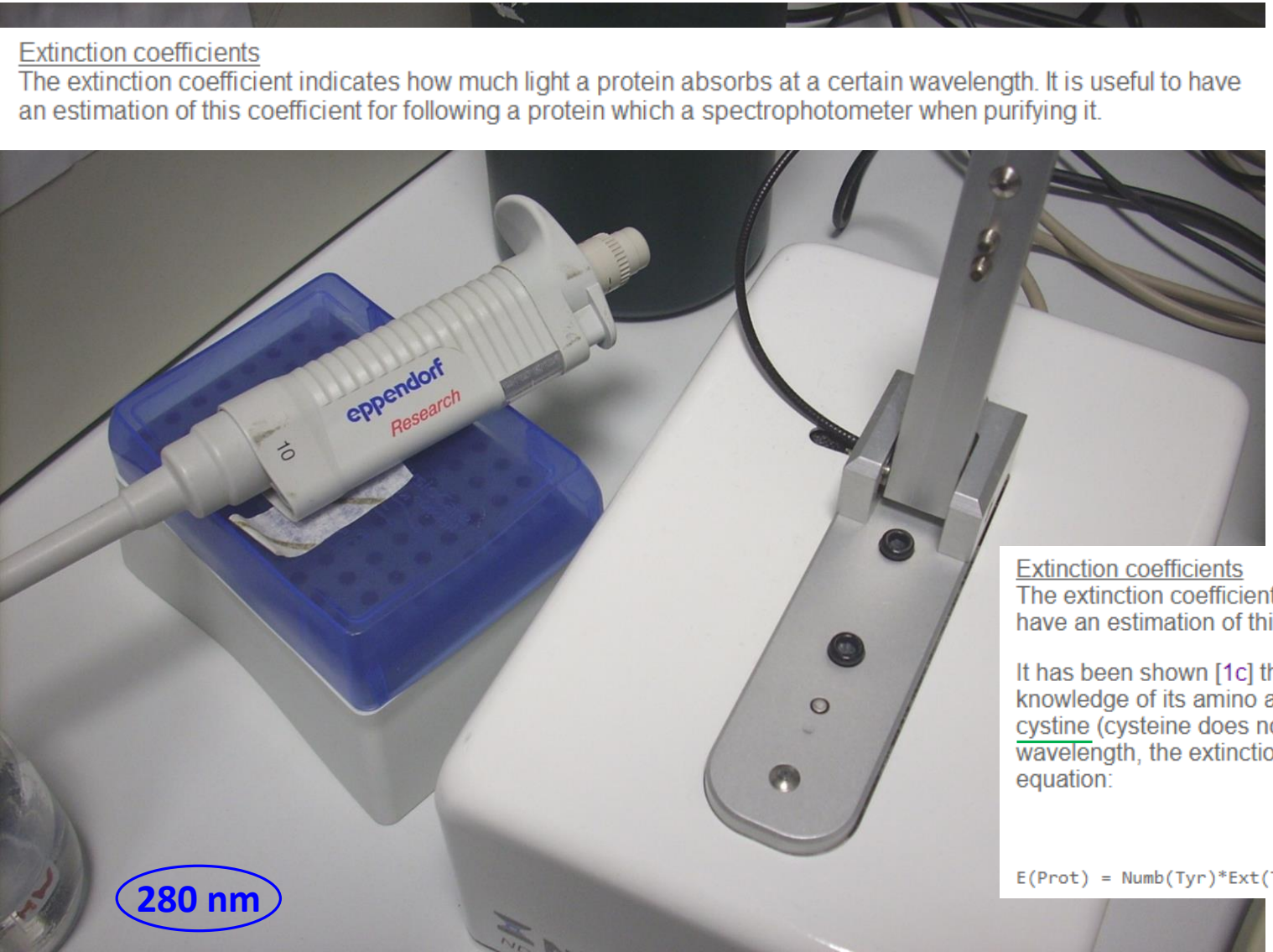
Using isoelectric point to determine the pH for initial protein crystallization trials

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.



Extinkční koeficient



- Extinkční (molární absorpční) 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.

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)
```

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.

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 Parameter Calculator

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.

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.

LENKALENKALENKALENKALENKALENKA

submit reset

Reference (for further information and for citation):

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. pubmed pdf

http://nickanthis.com/tools/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.

Amino acid sequence:
LENKALENKALENKALENKALENKALENKA

Number of residues:
30
```

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

Molar absorptivity (extinction coefficient) at 205 nm =
83020 M-1 cm-1
```

Protein Parameter Calculator

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.

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.

LENKALENKALENKALENKALENKALENKA

submit reset

Reference (for further information and for citation):

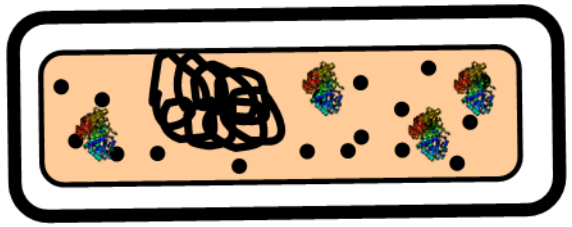
Anthis N.J. & Clore G. M. (2013) Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm, *Protein Science* 22, 851-8. pubmed pdf

http://nickanthis.com/tools/a205.html

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***

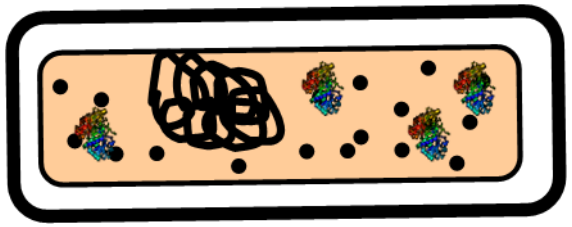
- Stabilita proteinu ve zkumavce

- **„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?

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

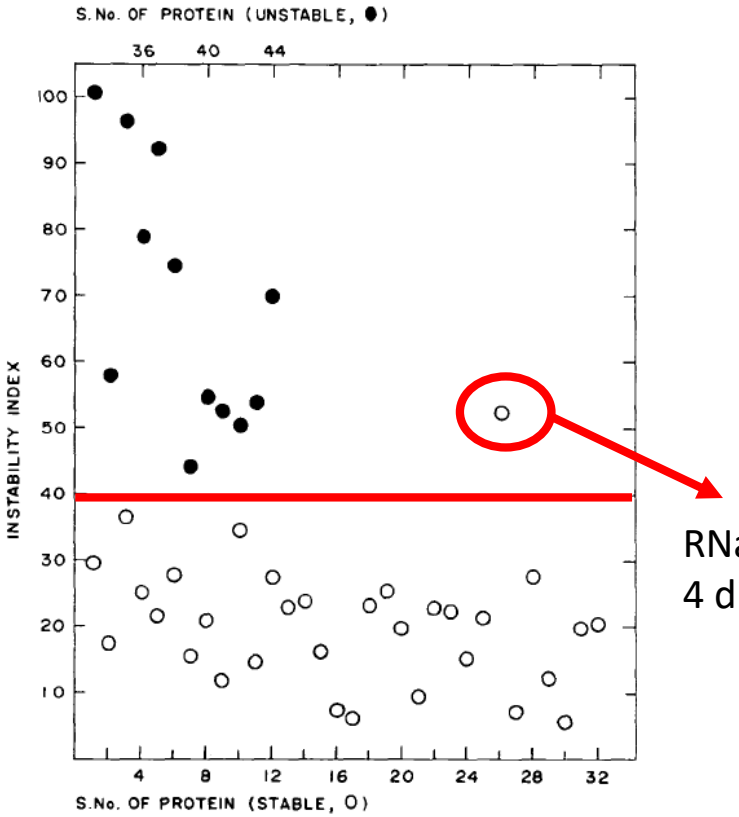
- Stabilita ***in vitro***
- Stabilita proteinu ve zkumavce

- **„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



RNasa A, stabilní protein,
4 disulfidické můstky

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 protein from its primary sequence

- **„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

- **„Instability index“**

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.

„Aliphatic index“/GRAVY

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.

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

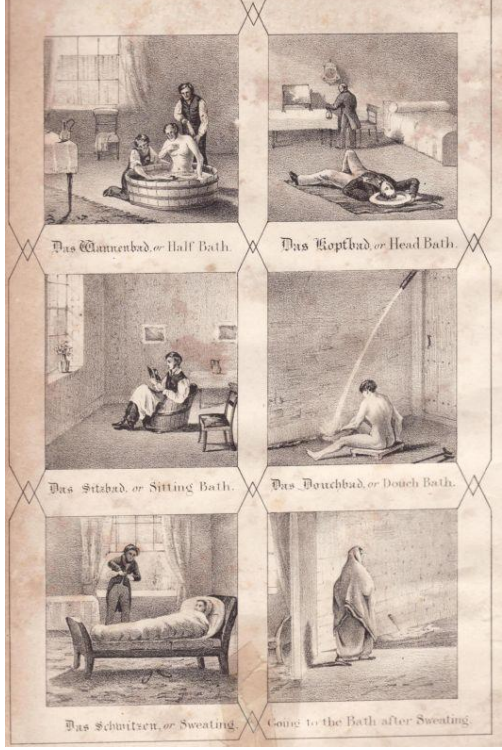
Hydrofobní/hydrofilní proteiny?

Membránové proteiny?

A Simple Method for Displaying the Hydropathic Character of a Protein

JACK KYTE AND RUSSELL F. DOOLITTLE

„Aliphatic index“/GRAVY



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.

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

Hydrofobní/hydrofilní proteiny?

Membránové proteiny?

A Simple Method for Displaying the Hydropathic Character of a Protein

JACK KYTE AND RUSSELL F. DOOLITTLE

GRAVY

Protein 1:

DPIALTAAVGADLLGDGRPETLWLGIGTLLMLIGTFYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVGSEMLDIYYARYADWLFTTPLLLLDLALLAKVDRVSIGTLVGV
DALMIVTGLVGALSHTPLARYTWWLFSTICMIVVLYFLATSLRAAAKERGPEVASTFNTLTALVLVLWTAYPILWIIGTEGAGVVGLGIETLLFMVLDVTAKVGFGFILLRSRAILGDTEAPEPS
AGAEASAAD **Membránový protein**

Protein 2:

KLAVYSTKQYDKKYLQQVNESFGFELEFFDFLLTEKTAKTANGCEAVCIFVNDDGSRPVLEELKKHGVKYIALRCAGFNNVDLDAAKELGLKVVRVPAYDPEAVAEHAIGMMMTLNRRIHRAYQR
TRDANFSLEGLTGFTMYGKTAGVIGTGKIGVAMLHILKGFGMRLLAFDPYPSAAALELGVEYVDLPTLFSESDVISLHCPLTPENYHLLNEAAFDQMKNGVMIVNTSRGALIDSQAAIEALKNQK
IGSLGMDVYENERDLFFEDKSNDVIQDDVFRRLSACHNVLFTGHQAFLTAEALTSISQTTLQNLSNLEKGETCPNELV **Cytoplasmatický protein**

Protein 1

Aliphatic index: 126.46

Grand average of hydropathicity (GRAVY): 0.816

Amino acid composition:

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:

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: 91.03

Grand average of hydropathicity (GRAVY): -0.097

GRAVY

RESEARCH Open Access

Murine colon proteome and characterization of the protein pathways

Sameh Magdeldin[1,2*], Yutaka Yoshida[1], Huiping Li[1,3], Yoshitaka Maeda[4], Munesuke Yokoyama[4], Shymaa Enany[1,5], Ying Zhang[1], Bo Xu[1], Hidehiko Fujinaka[1], Eishin Yaoita[1], Sei Sasaki[6] and Tadashi Yamamoto[1]

Abstract

Background: Most of the current proteomic researches focus on proteome alteration due to pathological disorders (*i.e.*: colorectal cancer) rather than normal healthy state when mentioning colon. As a result, there are lacks of information regarding normal whole tissue- colon proteome.

Results: We report here a detailed murine (mouse) whole tissue- colon protein reference dataset composed of 1237 confident protein (FDR < 2) with comprehensive insight on its peptide properties, cellular and subcellular localization, functional network GO annotation analysis, and its relative abundances. The presented dataset includes wide spectra of p*I* and Mw ranged from 3–12 and 4–600 KDa, respectively. Gravy index scoring predicted 19.5% membranous and 80.5% globularly located proteins. GO hierarchies and functional network analysis illustrated proteins function together with their relevance and implication of several candidates in malignancy such as Mitogen- activated protein kinase (Mapk8, 9) in colorectal cancer, Fibroblast growth factor receptor (Fgfr 2), Glutathione S-transferase (Gstp1) in prostate cancer, and Cell division control protein (Cdc42), Ras-related protein (Rac1,2) in pancreatic cancer. Protein abundances calculated with 3 different algorithms (NSAF, PAF and emPAI) provide a relative quantification under normal condition as guidance.

Conclusions: This highly confidence colon proteome catalogue will not only serve as a useful reference for further experiments characterizing differentially expressed proteins induced from diseased conditions, but also will aid in better understanding the ontology and functional absorptive mechanism of the colon as well.

Keywords: Colon, Proteome, Mass spectrometry, HPLC

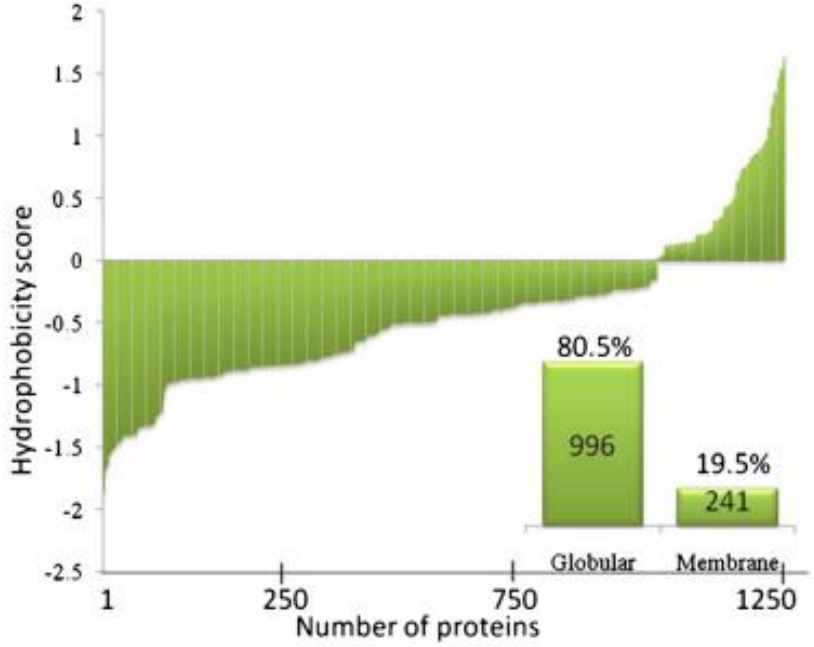


Figure 3 Gravy index score (average hydrophobicity and hydrophilicity) of colon proteins measured by Kyte-Doolittle and Hopp Woods formula. Hydrophobicity score (arbitrary unit) below 0 are more likely globular (hydrophilic protein), while scores above 0 are more likely membranous (hydrophobic). Inserted bar panel represents number and percentage of each group in colon proteome.

Predikce lokalizace

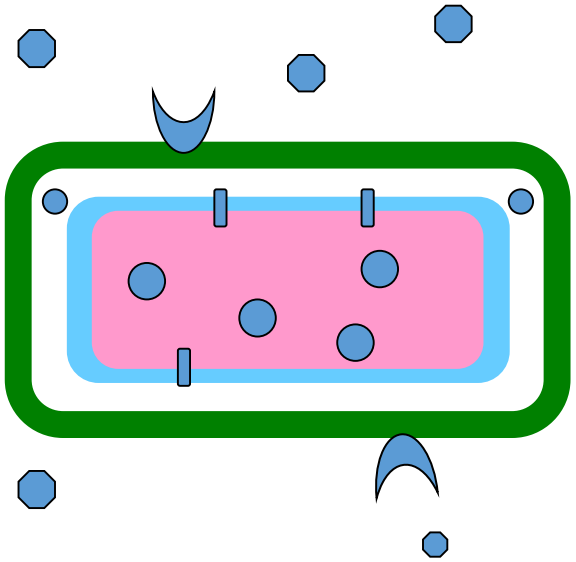
- Predikce **lokalizace** proteinů v buňce (prokaryotická, eukaryotická).
- Lokalizace proteinů napomáhá určení (ověření) jejich **funkce**.
- Vypovídá o předpokládaných **vlastnostech** proteinů (cytoplasmatické x membránové).

The computational prediction of the subcellular localization of bacterial proteins is an important step in genome annotation and in the search for novel vaccine or drug targets. Since the 1991 release of PSORT I--the first comprehensive algorithm to predict bacterial protein localization--many other localization prediction tools have been developed. These methods offer significant improvements in predictive performance over PSORT I and the accuracy of some methods now rivals that of certain high-throughput laboratory methods for protein localization identification.

Gardy et al., 2006

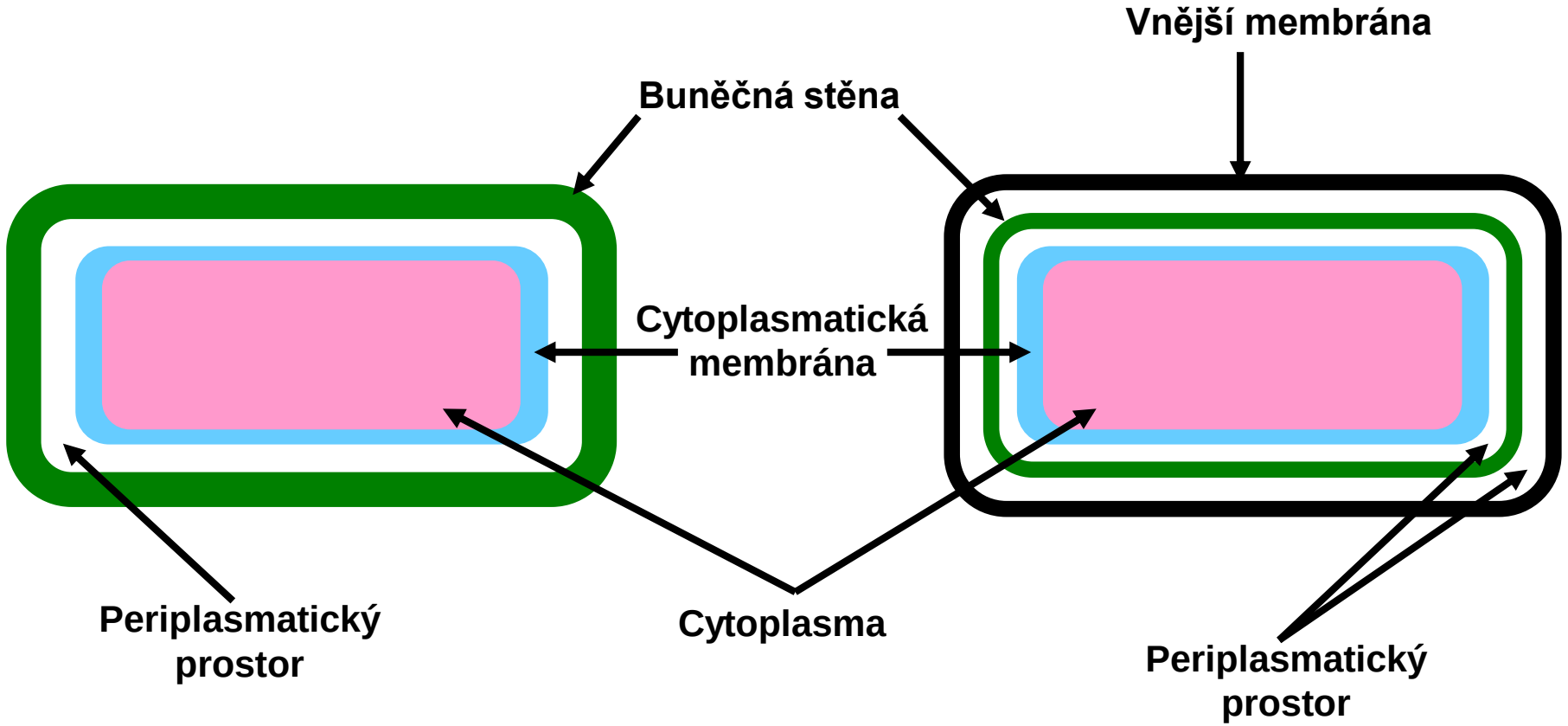
Computational prediction of bacterial protein subcellular localization (SCL) provides a quick and inexpensive means for gaining insight into protein function, verifying experimental results, annotating newly sequenced bacterial genomes, detecting potential cell surface/secreted drug targets, as well as identifying biomarkers for microbes. In recent years, this area of computational research has achieved an impressive level of precision (Gardy and Brinkman, 2006), allowing SCL prediction tools to be reliably integrated into automated proteome annotation pipelines and to complement analyses of high-throughput proteomics experiments.

Yu et al., 2010



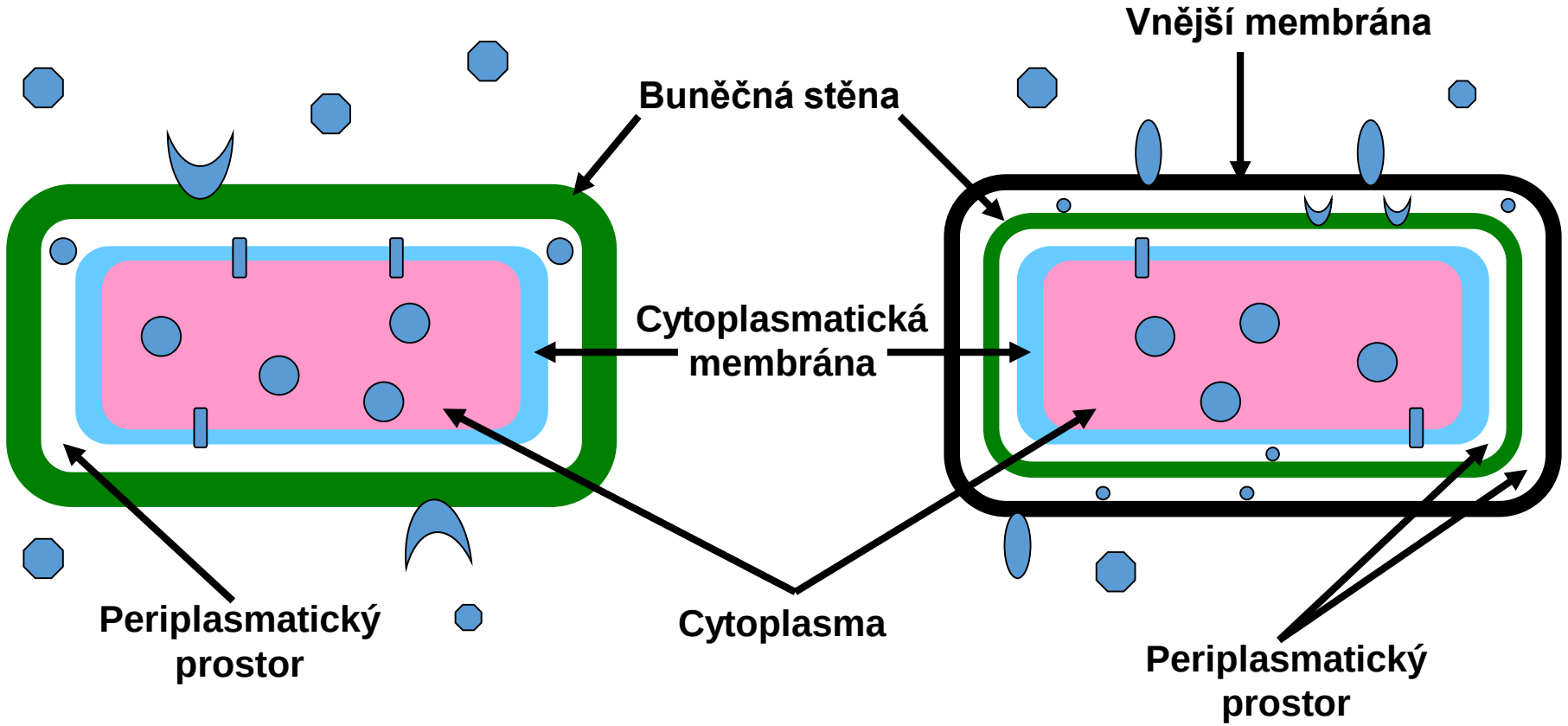
Predikce lokalizace

- Predikce **lokalizace** proteinů v buňce (prokaryotická, eukaryotická).
- Lokalizace proteinů napomáhá určení (ověření) jejich **funkce**.
- **Grampozitivní**/**Gramnegativní** bakterie/eukaryotické buňky



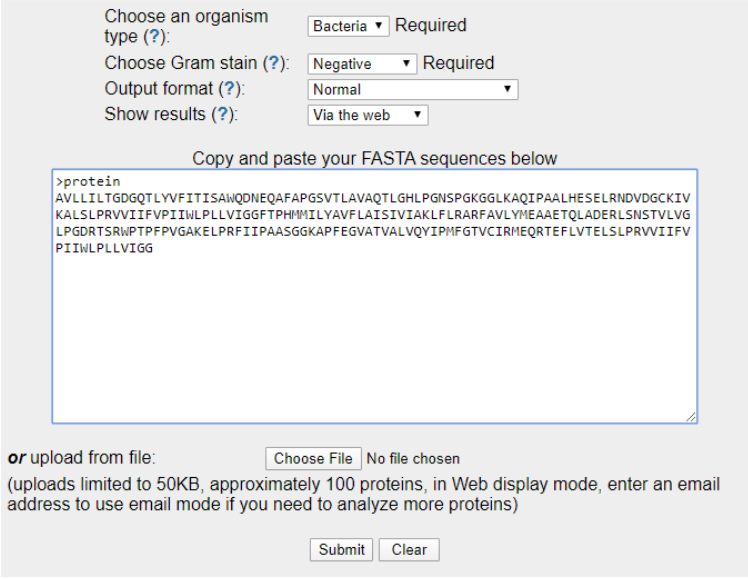
Predikce lokalizace

- Predikce **lokalizace** proteinů v buňce (prokaryotická, eukaryotická).
- Lokalizace proteinů napomáhá určení (ověření) jejich **funkce**.
- **Grampozitivní**/**Gramnegativní** bakterie/eukaryotické buňky



Predikce lokalizace

- Predikce **lokalizace** proteinů v buňce (prokaryotická, eukaryotická).
- Lokalizace proteinů napomáhá určení (ověření) jejich **funkce**.
- **Grampozitivní**/**Gramnegativní** bakterie/eukaryotické buňky



BIOINFORMATICS **ORIGINAL PAPER** *Vol. 26 no. 13 2010, pages 1608–1615*
doi:10.1093/bioinformatics/btq249

Sequence analysis Advance Access publication May 13, 2010

PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes

Nancy Y. Yu[1], James R. Wagner[2,†], Matthew R. Laird[1], Gabor Melli[2], Sébastien Rey[1], Raymond Lo[1], Phuong Dao[2], S. Cenk Sahinalp[2], Martin Ester[2], Leonard J. Foster[3] and Fiona S. L. Brinkman[1,*]

https://psort.org/

psortb

Submit Sequences | Documentation | Resources | Contact | Updates

PSORTb Results (Click here for an explanation of the output formats)

```
SeqID: cokoliv
  Analysis Report:
    CMSVM-              CytoplasmicMembrane          [No details]
    CytoSVM-            Unknown                      [No details]
    ECSVM-              Unknown                      [No details]
    ModHMM-             CytoplasmicMembrane          [7 internal helices found]
    Motif-              Unknown                      [No motifs found]
    OMPMotif-           Unknown                      [No motifs found]
    OMSVM-              Unknown                      [No details]
    PPSVM-              Unknown                      [No details]
    Profile-            Unknown                      [No matches to profiles found]
    SCL-BLAST-          Unknown                      [No matches against database]
    SCL-BLASTe-         Unknown                      [No matches against database]
    Signal-             Unknown                      [No signal peptide detected]
  Localization Scores:
    Cytoplasmic             0.00
    CytoplasmicMembrane    10.00
    Periplasmic             0.00
    OuterMembrane           0.00
    Extracellular           0.00
  Final Prediction:
    CytoplasmicMembrane    10.00
-------------------------------------------------------------------------------
```

psortb

Submit Sequences | Documentation | Resources | Contact | Updates

PSORTb Results (Click here for an explanation of the output formats)

```
SeqID: cokoliv
  Analysis Report:
    CMSVM-              Unknown                      [No details]
    CytoSVM-            Cytoplasmic                  [No details]
    ECSVM-              Unknown                      [No details]
    ModHMM-             Unknown                      [No internal helices found]
    Motif-              Unknown                      [No motifs found]
    OMPMotif-           Unknown                      [No motifs found]
    OMSVM-              Unknown                      [No details]
    PPSVM-              Unknown                      [No details]
    Profile-            Unknown                      [No matches to profiles found]
    SCL-BLAST-          Cytoplasmic                  [matched 27461218: Cytoplasmic protein]
    SCL-BLASTe-         Unknown                      [No matches against database]
    Signal-             Unknown                      [No signal peptide detected]
  Localization Scores:
    Cytoplasmic             9.97
    CytoplasmicMembrane     0.01
    Periplasmic             0.01
    OuterMembrane           0.00
    Extracellular           0.00
  Final Prediction:
    Cytoplasmic             9.97
-------------------------------------------------------------------------------
```

- Sekvenční **podobnost** proteinům se známou funkcí (lokalizací)
- Sekundární **struktura** – transmembránové helixy
- Analýza specifických **motivů** v sekvenci
- **Signální peptidy**

Choose an organism type (?): Bacteria ▾ Required
Choose Gram stain (?): Negative ▾ Required
Output format (?): Normal ▾
Show results (?): Via the web ▾

Copy and paste your FASTA sequences below

```
>protein
AVLLILTGDGQTLYVFITISAWQDNEQAFAPGSVTLAVAQTLGHLPGNSPGKGGLKAQIPAALHESELRNDVDGCKIV
KALSLPRVVIIFVPIIWLPLLVIGGFTPHMMILYAVFLAISIVIAKLFLRARFAVLYMEAAETQLADERLSNSTVLVG
LPGDRTSRWPTPFPVGAKELPRFIIPAASGGKAPFEGVATVALVQYIPMFGTVCIRMEQRTEFLVTELSLPRVVIIFV
PIIWLPLLVIGG
```

or upload from file: Choose File No file chosen
(uploads limited to 50KB, approximately 100 proteins, in Web display mode, enter an email address to use email mode if you need to analyze more proteins)

Submit Clear

About WoLF PSORT

WoLF PSORT predicts the subcellular localization sites of proteins based on their amino acid sequences. The method, which is a major extension to the venerable PSORTII program, makes predictions based on both known sorting signal motifs and some correlative sequence features such as amino acid content. Like PSORT and PSORTII, WoLF PSORT displays some information about detected sorting signals which is useful in helping users determine the reliability of the prediction in specific cases. Our experiments (paper in preparation) show that the overall prediction accuracy of WoLF PSORT is over 80%. For common localization sites (e.g. cytosol, nucleus, mitochondria, etc) WoLF PSORT makes better than majority classifier predictions even for queries that do not have strong sequence similarity to any sequence in the dataset. Thus WoLF PSORT is a useful complement to tools such as BLAST. The current dataset used to train WoLF PSORT contains over 12,000 animal sequences and more than 2,000 plant and fungi sequences respectively. It was gathered mainly from Uniprot but several hundred *Arabidopsis thaliana* sequences from the Gene Ontology database were also included.

https://psort.org/

What's in a name

"WoLF" does not necessarily stand for anything. A rather dramatic mnemonic would be "Where Life Functions". Originally it was going to be "Learned Weight Features" but I wanted the acronym to be a pronouncable English word. Waldo Lives Forever.

- Predikce lokalizace proteinů v **eukaryotických** buňkách (živočichové, rostliny, houby).
- Mnohem více možných lokalizací proteinů!

What's in a name

"WoLF" does not necessarily stand for anything. A rather dramatic mnemonic would be "Where Life Functions". Originally it was going to be "Learned Weight Features" but I wanted the acronym to be a pronouncable English word. Women only Love Fools.

Abbrev	Localization Site
chlo	chloroplast
cyto	cytosol
cysk	cytoskeleton
E.R.	endoplasmic reticulum
extr	extracellular
golg	Golgi apparatus
lyso	lysosome
mito	mitochondria
nucl	nuclear
pero	peroxisome
plas	plasma membrane
vacu	vacuolar membrane

About WoLF PSORT

WoLF PSORT predicts the subcellular localization sites of proteins based on their amino acid sequences. The method, which is a major extension to the venerable PSORTII program, makes predictions based on both known sorting signal motifs and some correlative sequence features such as amino acid content. Like PSORT and PSORTII, WoLF PSORT displays some information about detected sorting signals which is useful in helping users determine the reliability of the prediction in specific cases. Our experiments (paper in preparation) show that the overall prediction accuracy of WoLF PSORT is over 80%. For common localization sites (e.g. cytosol, nucleus, mitochondria, etc) WoLF PSORT makes better than majority classifier predictions even for queries that do not have strong sequence similarity to any sequence in the dataset. Thus WoLF PSORT is a useful complement to tools such as BLAST. The current dataset used to train WoLF PSORT contains over 12,000 animal sequences and more than 2,000 plant and fungi sequences respectively. It was gathered mainly from Uniprot but several hundred *Arabidopsis thaliana* sequences from the Gene Ontology database were also included.

https://psort.org/

What's in a name

"WoLF" does not necessarily stand for anything. A rather dramatic mnemonic would be "Where Life Functions". Originally it was going to be "Learned Weight Features" but I wanted the acronym to be a pronouncable English word. Waldo Lives Forever.

- Predikce lokalizace proteinů v **eukaryotických** buňkách (živočichové, rostliny, houby).
- Mnohem více možných lokalizací proteinů!

What's in a name

"WoLF" does not necessarily stand for anything. A rather dramatic mnemonic would be "Where Life Functions". Originally it was going to be "Learned Weight Features" but I wanted the acronym to be a pronouncable English word. Women only Love Fools.

WoLF PSORT

Protein Subcellular Localization Prediction

about WoLF PSORT　WoLF PSORTについて　links　Example Output

Please select an organism type:
- Animal
- Plant
- Fungi

Please select input method:
- From Text Area
- From File

Input Filename:
Choose File　No file chosen

Text Area: Enter multifasta format protein sequence(s) here.

Submit　clear

(↑Select organism type to activate the submit button)

Predikce – signální sekvence/peptidy

SignalP-5.0 Server

DeepLoc Remember, the presence or absence of a signal peptide is not the whole story about the localization of a protein! If you want to find out more about the sorting of your eukaryotic proteins, try the protein subcellular localization predictor.

Submit data

The SignalP 5.0 server predicts the presence of signal peptides and the location of their cleavage sites in proteins from Archaea, Gram-positive Bacteria, Gram-negative Bacteria and Eukarya. In Bacteria and Archaea, SignalP 5.0 can discriminate between three types of signal peptides:

- Sec/SPI: "standard" secretory signal peptides transported by the Sec translocon and cleaved by Signal Peptidase I (*Lep*)
- Sec/SPII: lipoprotein signal peptides transported by the Sec translocon and cleaved by Signal Peptidase II (*Lsp*)
- Tat/SPI: Tat signal peptides transported by the Tat translocon and cleaved by Signal Peptidase I (*Lep*)

https://services.healthtech.dtu.dk/service.php?SignalP-5.0

https://services.healthtech.dtu.dk/services/SignalP-6.0/

SignalP - 6.0

Prediction of Signal Peptides and their cleavage sites in all domains of life

The SignalP 6.0 server predicts the presence of signal peptides and the location of their cleavage sites in proteins from Archaea, Gram-positive Bacteria, Gram-negative Bacteria and Eukarya. In Bacteria and Archaea, SignalP 6.0 can discriminate between five types of signal peptides:

Sec/SPI: "standard" secretory signal peptides transported by the Sec translocon and cleaved by Signal Peptidase I (*Lep*)
Sec/SPII: lipoprotein signal peptides transported by the Sec translocon and cleaved by Signal Peptidase II (*Lsp*)
Tat/SPI: Tat signal peptides transported by the Tat translocon and cleaved by Signal Peptidase I (*Lep*)
Tat/SPII: Tat lipoprotein signal peptides transported by the Tat translocon and cleaved by Signal Peptidase II (*Lsp*)
Sec/SPIII: Pilin and pilin-like signal peptides transported by the Sec translocon and cleaved by Signal Peptidase III (*PilD/PibD*)

Proteom a proteomika

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

- **Proteomika** – studium proteomů.
- **Proteomika** – věda zabývající se komplexní analýzou proteinů (identifikace, exprese, charakterizace).
- **Proteomika** – analýza proteinů ve velkém rozsahu (struktura, funkce, interakce).

- **Proteomika** – vyžaduje separační techniky, hmotnostní spektrometrii, bioinformatiku, databáze genů a 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.

- 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.

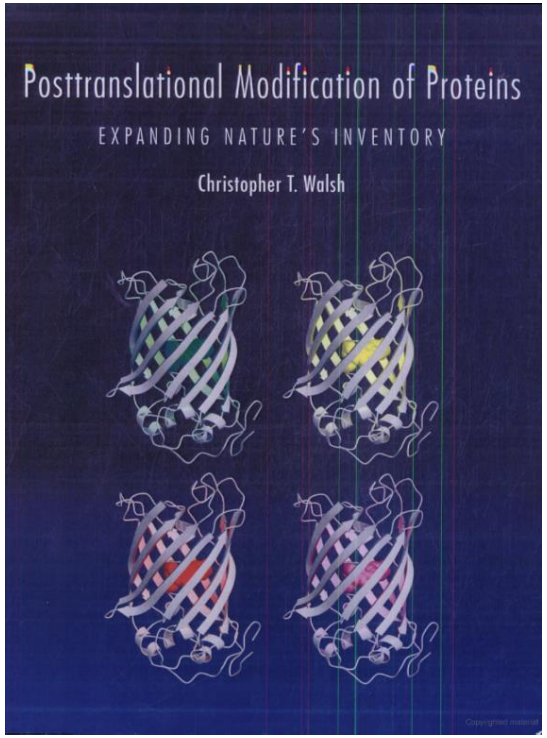
- **Klasifikace posttranslačních modifikací** – typ modifikované aminokyseliny, podle modifikujícího enzymu, reverzibilita modifikací.

Posttranslační modifikace - typy

Table 1: Posttranslational protein modifications at the side chains.[a]

Residue	Reaction	Example
Asp	phosphorylation	protein tyrosine phosphatases; response regulators in two-component systems
	isomerization to isoAsp	
Glu	methylation	chemotaxis receptor proteins
	carboxylation	Gla residues in blood coagulation
	polyglycination	tubulin
	polyglutamylation	tubulin
Ser	phosphorylation	protein serine kinases and phosphatases
	O-glycosylation	notch O-glycosylation
	phosphopantetheinylation	fatty acid synthase
	autocleavages	pyruvamidyl enzyme formation
Thr	phosphorylation	protein threonine kinases/phosphatases
	O-glycosylation	
Tyr	phosphorylation	tyrosine kinases/phosphatases
	sulfation	CCR5 receptor maturation
	ortho-nitration	inflammatory responses
	TOPA quinone	amine oxidase maturation
His	phosphorylation	sensor protein kinases in two-component regulatory systems
	aminocarboxypropylation	diphthamide formation
	N-methylation	methyl CoM reductase
Lys	N-methylation	histone methylation
	N-acylation by acetyl, biotinyl, lipoyl, ubiquityl groups	histone acetylation; swinging-arm prosthetic groups; ubiquitin; SUMO (small ubiquitin-like modifier) tagging of proteins
	C-hydroxylation	collagen maturation
Cys	S-hydroxylation (S-OH)	sulfenate intermediates
	disulfide bond formation	protein in oxidizing environments
	phosphorylation	PTPases
	S-acylation	Ras
	S-prenylation	Ras
	protein splicing	intein excisions
Met	oxidation to sulfoxide	Met sulfoxide reductase
Arg	N-methylation	histones
	N-ADP-ribosylation	$G_{s\alpha}$
Asn	N-glycosylation	N-glycoproteins
	N-ADP-ribosylation	eEF-2
	protein splicing	intein excision step
Gln	transglutamination	protein cross-linking
Trp	C-mannosylation	plasma-membrane proteins
Pro	C-hydroxylation	collagen; HIF-1α
Gly	C-hydroxylation	C-terminal amide formation

[a] No modifications of Leu, Ile, Val, Ala, Phe side chains are known. A more extensive list can be found in reference [3].



Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications

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

DOI: 10.1002/anie.200501023

Posttranslační modifikace - příklady

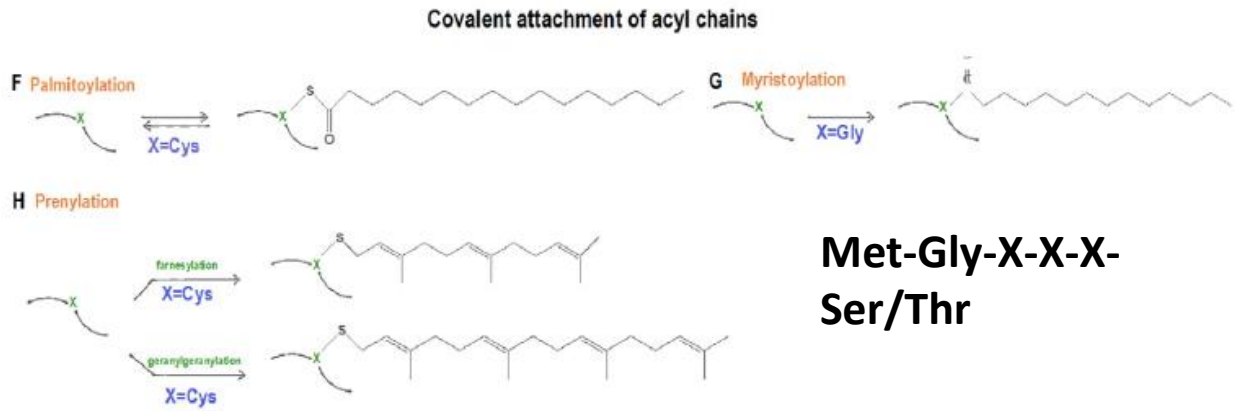
- **Kovalentní připojení malé molekuly (funkční skupiny):**

- **Fosforylace (serin, threonin, tyrosin, histidin, arginin, lysin), aktivace/inhibice enzymů. Nejstudovanější PTM.**

- **Glykosylace (*N*-glykosylace, *O*-glykosylace, GPI kotva, C-mannosylace).**

- **S-nitrosylace (cystein). Neenzymatické připojení NO. Vazba je labilní, náročná experimentální identifikace!**

- **Kovalentní připojení acylových řetězců:**



Met-Gly-X-X-X-Ser/Thr

- **Kovalentní připojení malých proteinů:**
- **Ubikvitinace (lysin). Regulace odbourávání proteinů, regulace funkce proteinů.**

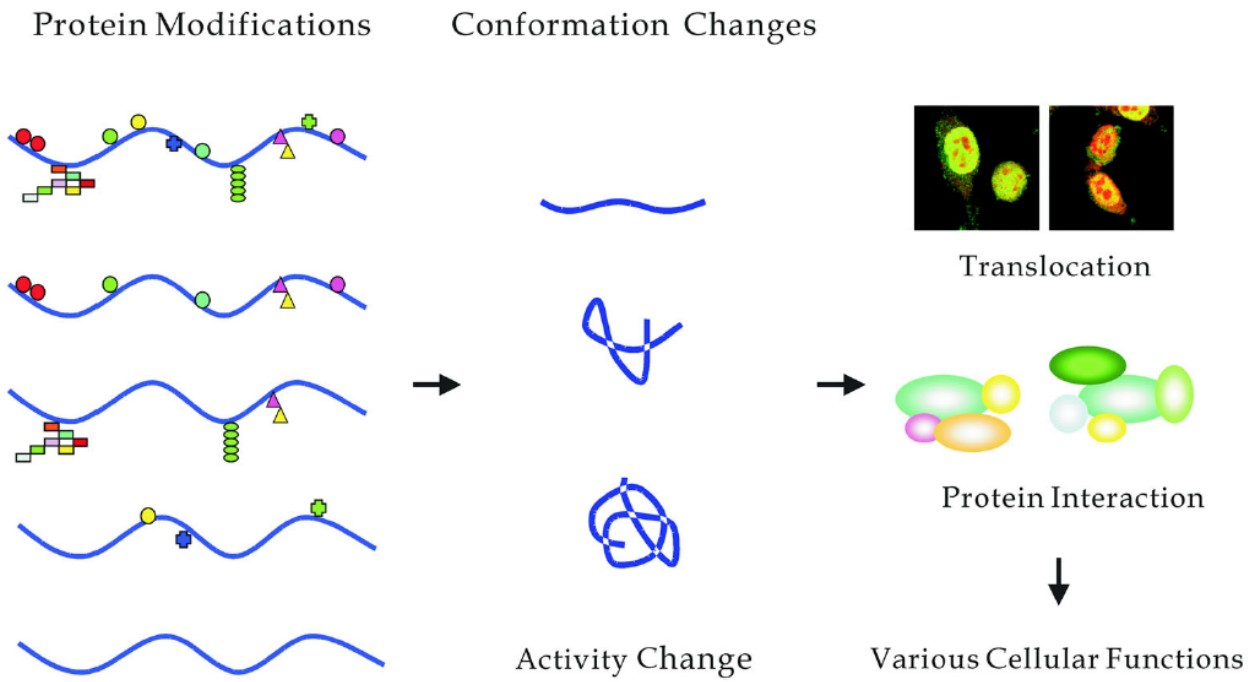
Protein post-translational modifications: *In silico* prediction tools and molecular modeling

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

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

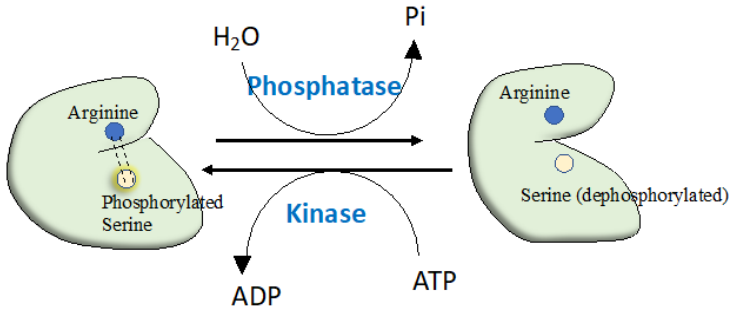
Jawon Seo and Kong-Joo Lee*

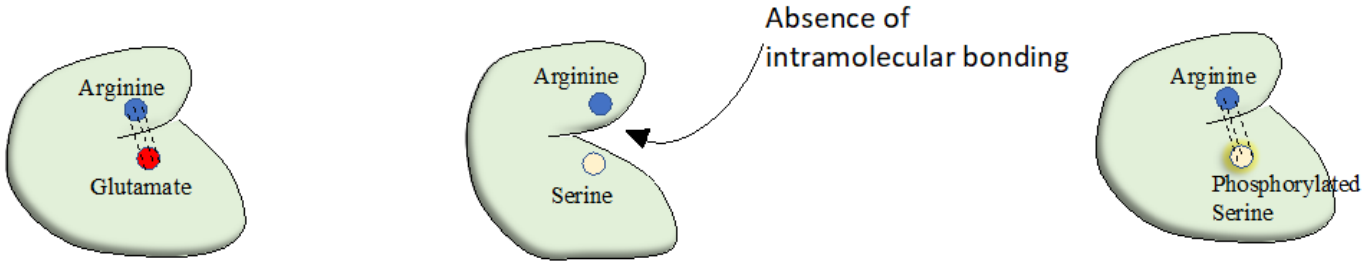
Posttranslační modifikace - příklady

- **Kovalentní připojení malé molekuly (funkční skupiny):**

- **Fosforylace (serin, threonin, tyrosin, histidin, arginin, lysin), aktivace/inhibice enzymů. Nejstudovanější PTM.**

- **Ovlivňují 3D a 4D strukturu proteinů, aktivitu a funkci (rozpustnost, stabilita, interakce, vypnuto/zapnuto).**





1
Ion pair formation between Arginine and Glutamate stabilizes the fold

2
Substitution of Glutamate with Serine leads to loss of intra-molecular bonding and loss of functional fold

3
Phosphorylated Serine can substitute for Glutamate and ion pair forms again stabilizing the proteins conformation

https://info.gbiosciences.com/blog/protein-phosphorylation-natures-favorite-way-of-decorating-proteins

Posttranslační modifikace - význam

Table 1 Golgi PTMs and associated diseases. This table summarizes all the human disorders linked to Golgi PTMs presented in this review

Affected gene	Protein	Affected PTM	Disease	Major clinical manifestations	Gene OMIM entry	Disease OMIM entry
MAN1B1	α 1,2 mannosidase	glycosylation	MAN1B1-CDG	severe mental retardation, delayed speech	604346	614202
SLC35A1	CMP-sialic acid transporter	glycosylation	SLC35A1-CDG	seizures, intellectual disability, ataxia, bleeding	605634	603585
SLC35A2	UDP-galactose transporter	glycosylation	SLC35A2-CDG	intellectual disability, seizures, skeletal abnormalities	314375	300896
SLC35A3	UDP-GlcNAc transporter	glycosylation	Arthrogryposis, mental retardation and seizures	autism spectrum disorder, hypotonia, epilepsy, and arthrogryposis	605632	615553
SEC23B	Sec23 homolog B	glycosylation	Dyserythropoietic anemia, congenital, type II	erythroblastic anemia: splenomegaly, gallstones, and iron overload potentially with liver cirrhosis or cardiac failure.	610512	224100
TRIP11	Golgi microtubule associated protein 210	glycosylation	Achondrogenesis type 1A	severe chondrodysplasia, lethal before or shortly after birth	604505	200600
UBE3A	Ubiquitin ligase E3A	glycosylation	Angelman syndrome	intellectual disability, seizures, lack of speech, and characteristic abnormal behavior	601623	105830
COG2	Component of oligomeric Golgi complex 2	glycosylation	COG2-CDG	microcephaly, developmental delay, intellectual disability, seizures, facial dysmorphism, liver dysfunction	606974	no entry yet
SLC33A1	Solute carrier family 33 (acetyl-CoA transporter), member 1	acetylation	Spastic paraplegia-42	spastic gait, increased lower limb tone, weakness and atrophy of the lower limb muscles, pes cavus	603690	612539
CHST3	Chondroitin 6-O-sulfotransferase	sulfation	Spondylo-epiphyseal dysplasia with joint dislocations	unusual skeletal dysplasia	603799	143095
CHST6	Corneal N-acetylglucosamine-6-O-sulfotransferase	sulfation	Macular corneal distrophy type II	progressive corneal opacification and reduced corneal sensitivity	605294	217800
CHST8	GalNAc-4-O sulfotransferase I	sulfation	Peeling skin syndrome	general skin peeling	610190	270300
CHST14	Dermatan sulfate GalNAc-4-O sulfotransferase I	sulfation	Ehlers-Danlos syndrome musculocontractural type 1	craniofacial dysmorphism, congenital contractures of thumbs and fingers, clubfeet, severe kyphoscoliosis	608429	601776
ARSE	Arylsulfatase E	sulfation	Chondrodysplasia punctata 1	stippled epiphyses, brachytelephalangy, nasomaxillary hypoplasia	300180	302950
PAPPS2	PAPS synthase	sulfation	Brachyolmia type 4	short-trunk stature, rectangular vertebral bodies, precocious calcification of rib cartilages, short femoral neck. Early death for severe cases.	603005	612847
SLA26A2	Sulfate anion transporter	sulfation	Achondrogenesis type 1B	severe chondrodysplasia, early death of respiratory failure	606718	600972
			Atelosteogenesis type 2	pulmonary hypoplasia, lethal in infants	606718	256050
			Epiphyseal dysplasia multiple 4	joint pain, scoliosis, malformations of the hands, feet, and knees	606718	226900
			Diastrophic dysplasia	scoliosis, clubfeet, malformed pinnae with calcification of the cartilage, cleft palate in some cases	606718	222600
GNTPG	N-acetylglucosamine-1-phosphotransferase gamma subunit	phosphorylation	Mucolipidosis III gamma	short stature, skeletal abnormalities, cardiomegaly, and developmental delay	607838	252605

Golgi post-translational modifications and associated diseases

Sven Potelle[1] · André Klein[2] · François Foulquier[1]

DOI 10.1007/s10545-015-9851-7

Posttranslační modifikace - význam

Table 1 (continued)

Affected gene	Protein	Affected PTM	Disease	Major clinical manifestations	Gene OMIM entry	Disease OMIM entry
GNTPAB	N-acetylglucosamine-1-phosphotransferase alpha and beta subunits	phosphorylation	Mucolipidosis II and III	Hip dislocation, gingival hyperplasian, thoracic deformities and hernia soon after birth. Delayed psychomotor development. Same clinical features for mucolipidosis III as described just above.	607840	252500 252600
IMPAD1	Golgi-resident PAP phosphatase	phosphorylation	Chondrodysplasia with joint dislocations	short stature, chondrodysplasia with brachydactyly, congenital joint dislocations, micrognathia, cleft palate, and facial dysmorphism	614010	614078
INPP5E	Inositol polyphosphate-5-phosphatase	phosphorylation	Morm syndrome	Mental retardation, truncal obesity, retinal dystrophy, and micropenis	613037	610156
		phosphorylation	Joubert syndrome 1	Heterogenous: hypoplasia of the cerebellar vermis with the characteristic neuroradiologic molar tooth sign , dysregulation of breathing pattern and developmental delay.	613037	213300
AKAP9	A-kinase anchor protein 9	phosphorylation	Long QT syndrome-11	recurrent syncope, seizure, or sudden death	604001	611820
FAM20C	Golgi kinase (family with sequence similarity 20, member C)	phosphorylation	Raine syndrome	neonatal osteosclerotic bone dysplasia, increased ossification of the skull	611061	259775
CAMKMT	Calmodulin-lysine N-methyltransferase	methylation	2p21 deletion syndrome	cystinuria, neonatal seizures, hypotonia, severe somatic and developmental delay, facial dysmorphism	609559	606407
MBTPS2	Site-2 protease	proteolytic cleavage	IFAP syndrome with or without BRESHECK syndrome	ichthyosis follicularis, atrichia, and photophobia	300294	308205
			Olmsted syndrome, X-linked	periorificial keratotic plaques and bilateral palmoplantar transgredient keratoderma	300294	300918
			keratosis follicularis spinulosa decalvans, X-linked	keratosis pilaris, progressive cicatricial alopecia of the scalp, eyebrows, and eyelashes	300294	308800
ZDHHC8	Zinc finger, DHHC-type containing 8	palmitoylation	Schizophrenia susceptibility	hallucinations and delusions, inappropriate emotional responses, disordered thinking and concentration, erratic behavior	608784	181500
ZDHHC9	Zinc finger, DHHC-type containing 9	palmitoylation	X-linked mental retardation (Raymond type)	general intellectual limitations associated with impairments in adaptive behavior	300646	300799
ZDHHC15	Zinc finger, DHHC-type containing 15	palmitoylation	X-linked mental retardation-91	general intellectual limitations associated with impairments in adaptive behavior	300576	300577
PPT1	Palmitoyl-protein thioesterase 1	palmitoylation	Neuronal ceroid lipofuscinosis 1	Heterogenous: progressive dementia, seizures, and progressive visual deficiency. The cellular phenotype includes intracellular accumulation of autofluorescent lipopigment storage material.	600722	256730

Golgi post-translational modifications and associated diseases

Sven Potelle[1] · André Klein[2] · François Foulquier[1]

DOI 10.1007/s10545-015-9851-7

Posttranslační modifikace - význam

Affected gene	Protein	Affected PTM	Major clinical manifestations	Gene OMIM entry	Disease OMIM entry
MAN1B1	α 1,2 mannosidase	glycosylation	severe mental retardation, delayed speech	604346	614202
SLC35A1	CMP-sialic acid transporter	glycosylation	seizures, intellectual disability, ataxia, bleeding	605634	603585
SLC35A2	UDP-galactose transporter	glycosylation	intellectual disability, seizures, skeletal abnormalities	314375	300896
SLC35A3	UDP-GlcNAc transporter	glycosylation	autism spectrum disorder, hypotonia, epilepsy, and arthrogryposis	605632	615553
SEC23B	Sec23 homolog B	glycosylation	erythroblastic anemia: splenomegaly, gallstones, and iron overload potentially with liver cirrhosis or cardiac failure.	610512	224100
TRIP11	Golgi microtubule associated protein 210	glycosylation	severe chondrodysplasia, lethal before or shortly after birth	604505	200600
UBE3A	Ubiquitin ligase E3A	glycosylation	intellectual disability, seizures, lack of speech, and characteristic abnormal behavior	601623	105830
COG2	Component of oligomeric Golgi complex 2	glycosylation	microcephaly, developmental delay, intellectual disability, seizures, facial dysmorphism, liver dysfunction	606974	no entry yet

OMIM® - Online Mendelian Inheritance in Man®

Welcome to OMIM®, Online Mendelian Inheritance in Man®. OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. The full-text, referenced overviews in OMIM contain information on all known mendelian disorders and over 15,000 genes. OMIM focuses on the relationship between phenotype and genotype. It is updated daily, and the entries contain copious links to other genetics resources.

This database was initiated in the early 1960s by Dr. Victor A. McKusick as a catalog of mendelian traits and disorders, entitled Mendelian Inheritance in Man (MIM). Twelve book editions of MIM were published between 1966 and 1998. The online version, OMIM, was created in 1985 by a collaboration between the National Library of Medicine and the William H. Welch Medical Library at Johns Hopkins. It was made generally available on the internet starting in 1987. In 1995, OMIM was developed for the World Wide Web by NCBI, the National Center for Biotechnology Information.

OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada Hamosh.

NLM's Profiles in Science -- The McKusick Papers

Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes

Joanna S. Amberger[1] and **Ada Hamosh**
McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, Tel. 410-955-0313, Fax. 410-955-4999

Abstract

Online Mendelian Inheritance in Man (OMIM) at OMIM.org is the primary repository of comprehensive, curated information on genes and genetic phenotypes and the relationships between them. This unit provides an overview of the types of information in OMIM and optimal strategies for searching and retrieving the information. OMIM.org has links to many related and complementary databases providing easy access to exploring more information on a topic. The relationship between genes and genetic disorders is highlighted in this unit. The basic protocol explains searching OMIM both from a gene then clinical features perspective. Two alternate protocols provide strategies for viewing gene-phenotype relationships as a gene map table and clinical features as a Quick View or Side-by-Side format. OMIM.org is updated nightly and the MIMmatch service, described in the Support Protocol, provides a convenient way to follow updates to entries, gene-phenotype relationships, and collaborate with other researchers.

OMIM databáze – geny a geneticky podmíněné choroby

50 YEARS OMIM
Human Genetics Knowledge for the World

https://omim.org

https://omim.org

Posttranslační modifikace - význam

603585

CONGENITAL DISORDER OF GLYCOSYLATION, TYPE IIf; CDG2F

Alternative titles; symbols

CDG IIf; CDGIIf

Clinical Features

Willig et al. (2001) reported a 4-month-old boy who presented with a spontaneous massive bleed in the posterior chamber of the right eye along with cutaneous hemorrhages. Laboratory studies showed marked thrombocytopenia and neutropenia. The patient experienced multiple episodes of bleeding over the next 30 months, including severe pulmonary hemorrhage. He also had multiple recurrent bacterial infections. Bone marrow transplantation was performed at age 34 months, but the patient died of complications at age 37 months.

Macrothrombocytopenia with abnormal demarcation membranes in megakaryocytes and neutropenia with a complete lack of sialyl-Lewis-X antigen in leukocytes--a new syndrome?

Willig TB, Breton-Gorius J, Elbim C, Mignotte V, Kaplan C, Mollicone R, Pasquier C, Filipe A, Miélot F, Cartron JP, Gougerot-Pocidalo MA, Debili N, Guichard J, Dommergues JP, Mohandas N, Tchernia G.

Blood. 2001 Feb 1;97(3):826-8. doi: 10.1182/blood.v97.3.826.

PMID: 11157507 Free article.

PubMed.gov

* 606672

GLYCOPROTEIN Ib, PLATELET, ALPHA POLYPEPTIDE; GP1BA

Alternative titles; symbols

GP Ib, ALPHA SUBUNIT
PLATELET GLYCOPROTEIN Ib, ALPHA POLYPEPTIDE
CD42B

Biochemical Features

By detailed laboratory analysis of a patient with thrombocytopenia and recurrent infections, Willig et al. (2001) found markedly decreased amounts of platelet membrane GP Ib (see GP1BA, 606672) and undetectable sialyl-Lewis-X on the surface of neutrophils, suggesting a defect in the posttranslational modification of glycoproteins. Martinez-Duncker et al. (2005) noted that the plasma of the patient reported by Willig et al. (2001) showed a normal sialylation pattern of transferrin (TF; 190000) and other major serum glycoproteins. The phenotype was due to the lack of sialyl-Lewis-X, which has considerable roles in cell-to-cell interactions, such as infections and megakaryocytic immaturity, that were defective in this patient.

Molecular Genetics

In a patient originally described by Willig et al. (2001), Martinez-Duncker et al. (2005) identified compound heterozygosity for 2 mutations in the SLC35A1 gene (605634.0001; 605634.0002). Martinez-Duncker et al. (2005) referred to this disorder as CDG type IIf.

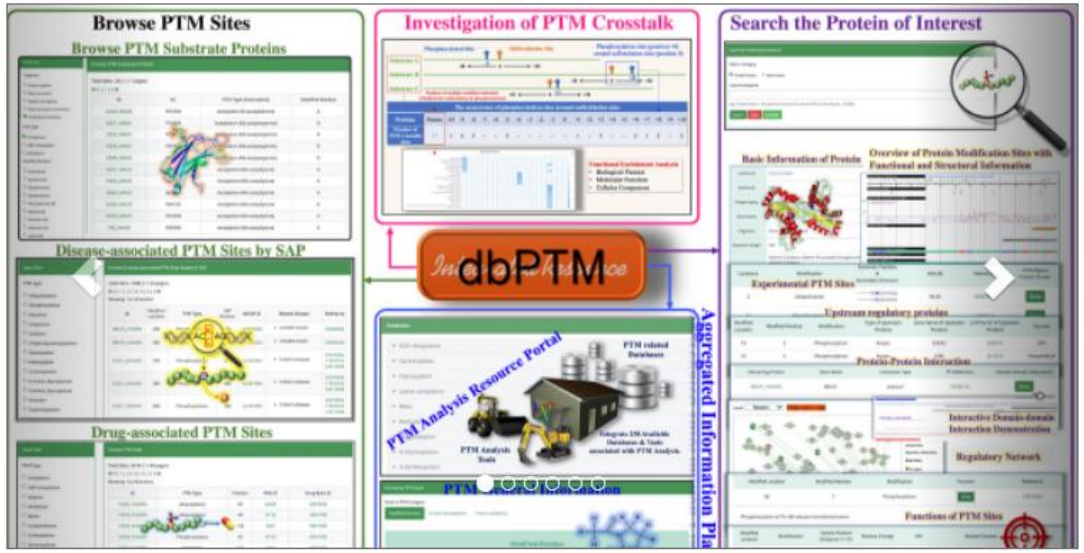
* 605634

SOLUTE CARRIER FAMILY 35 (CMP-SIALIC ACID TRANSPORTER), MEMBER 1; SLC35A1

The SLC35A1 gene encodes a CMP-sialic acid transporter located within the membrane of the Golgi apparatus. The transporter moves nucleotide sugars across the membrane for use in glycosylation reactions that take place within the Golgi department (Eckhardt et al., 1996).

50 YEARS OMIM
Human Genetics Knowledge for the World

Posttranslační modifikace - databáze



dbPTM: an information repository of protein post-translational modification

Tzong-Yi Lee[1], Hsien-Da Huang[1,2,*], Jui-Hung Hung[1], Hsi-Yuan Huang[1], Yuh-Shyong Yang[2,3] and Tzu-Hao Wang[4]

https://awi.cuhk.edu.cn/dbPTM/index.php

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 (LC, MS)

Predikce

Predikce PTM

Vyhodnocování experimentálních dat

Experimentální identifikace/ ověření PTM.
Databáze PTM.
Databáze modifikovaných proteinů.

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*

Methodology article

Open Access

Coronavirus 3CL^{pro} proteinase cleavage sites: Possible relevance to SARS virus pathology

Lars Kiemer, Ole Lund, Søren Brunak and Nikolaj Blom*

Address: Center for Biological Sequence Analysis BioCentrum-DTU, Building 208 Technical University of Denmark DK-2800 Lyngby, Denmark

Email: Lars Kiemer - lars@cbs.dtu.dk; Ole Lund - lund@cbs.dtu.dk; Søren Brunak - brunak@cbs.dtu.dk; Nikolaj Blom* - nikob@cbs.dtu.dk

* Corresponding author

Published: 06 June 2004

BMC Bioinformatics 2004, **5**:72

Received: 23 January 2004

Accepted: 06 June 2004

This article is available from: http://www.biomedcentral.com/1471-2105/5/72

Abstract

Background: Despite the passing of more than a year since the first outbreak of Severe Acute Respiratory Syndrome (SARS), efficient counter-measures are still few and many believe that reappearance of SARS, or a similar disease caused by a coronavirus, is not unlikely. For other virus families like the picornaviruses it is known that pathology is related to proteolytic cleavage of host proteins by viral proteinases. Furthermore, several studies indicate that virus proliferation can be arrested using specific proteinase inhibitors supporting the belief that proteinases are indeed important during infection. Prompted by this, we set out to analyse and predict cleavage by the coronavirus main proteinase using computational methods.

NetCGlyc - 1.0

C-mannosylation sites in mammalian proteins

The NetCGlyc 1.0 produces neural network predictions of C-mannosylation sites in mammalian proteins.

```
##gff-version 2
##source-version netCglyc-1.0b
##date 2007-03-14
##Type Protein
# seqname               source         feature  start    end   score +/-  ?
# ------------------------------------------------------------------------
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno     21     21   0.269  .   .
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno     38     38   0.459  .   .
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno     41     41   0.639  .   W
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno     44     44   0.484  .   .
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno     72     72   0.221  .   .
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno    115    115   0.285  .   .
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno    207    207   0.228  .   .
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno    244    244   0.246  .   .
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno    299    299   0.160  .   .
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno    317    317   0.203  .   .
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno    410    410   0.243  .   .
Q6ZMM2_ATL5_HUMAN       netCglyc-1.0b  C-manno    454    454   0.227  .   .
# ------------------------------------------------------------------------
```

Only the residues with scores higher than **0.5**, marked with "**W**" are predicted as C-mannosylated.

DTU Health Tech

DTU - Technical University of Denmark

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

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 Health Tech

DTU - Technical University of Denmark

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

Expasy

Swiss Bioinformatics Resource Portal

PeptideCutter

Potential cleavage sites in a protein

GlycoMod

Possible oligosaccharide structures on proteins from masses

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.

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

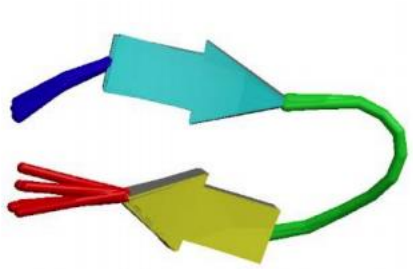
Základní predikované vlastnosti proteinů

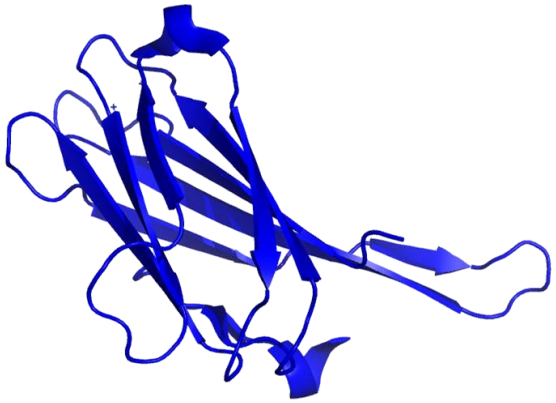
```
RKSTGGKAPRKQLATKAARKSAPATGGV
KKPHRYRPGTVALREIRRYQKSTELLIR
KLPFQRLVREIAQDFKTDLRFQSSAVMA
LQEASEAYLVGLFEDTNLCAIHAKR
```

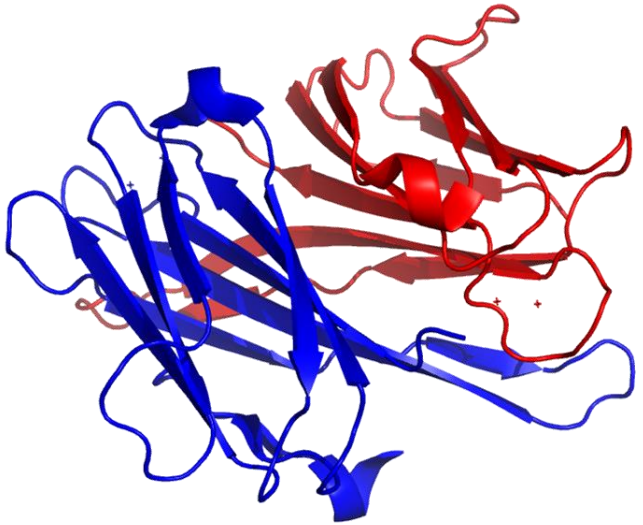
- Počet aminokyselin
- Pozitivně (Arg + Lys)/záporně (Asp + Glu) nabitá rezidua
- Molekulová hmotnost
- Izoelektrický bod
- Extinkční koeficient
- „Instability index“
- Poločas života
- GRAVY (Grand Average of Hydropathy) index
- „Aliphatic index“

+ **Lokalizace proteinů**
Predikce posttranslačních modifikací

9) Predikce 2D, 3D a 4D struktury proteinů. Souřadnice. Formáty. Vizualizační nástroje.







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

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.

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

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

Jawon Seo and Kong-Joo Lee*

Methods for predicting bacterial protein subcellular localization

Jennifer L. Gardy and Fiona S. L. Brinkman*

Sequence analysis Advance Access publication May 13, 2010

PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes

Nancy Y. Yu[1], James R. Wagner[2,†], Matthew R. Laird[1], Gabor Melli[2], Sébastien Rey[1], Raymond Lo[1], Phuong Dao[2], S. Cenk Sahinalp[2], Martin Ester[2], Leonard J. Foster[3] and Fiona S. L. Brinkman[1,*]

WoLF PSORT: protein localization predictor

Paul Horton[1], Keun-Joon Park[1,2], Takeshi Obayashi[3], Naoya Fujita[1,3], Hajime Harada[1], C.J. Adams-Collier[4] and Kenta Nakai[3,*]

Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications

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

dbPTM: an information repository of protein post-translational modification

Tzong-Yi Lee[1], Hsien-Da Huang[1,2,*], Jui-Hung Hung[1], Hsi-Yuan Huang[1], Yuh-Shyong Yang[2,3] and Tzu-Hao Wang[4]