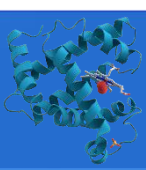


29.04.2026

Curs 10 – Predicția proprietăților unei proteine pe baza secvenței de aminoacizi

Relevanța structuri primare



Deși structura tridimensională a unei proteine oferă informații complete legate de proprietățile specifice, metodele de determinare a acesteia (*care?*) sunt relativ costisitoare și nu au un succes garantat (*proteinele transmembranare sunt dificil de cristalizat sau solubilizat*). Pe de altă parte, tehnicile de secvențiere a acizilor nucleici sunt relativ accesibile și au dus la acumularea unui număr uriaș de secvențe de ADN/ARN din care pot fi extrase ușor secvențele de aminoacizi.

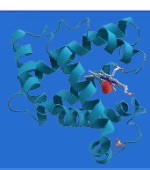
Dacă ținem cont de faptul că **structura primară este cea care dictează structura tridimensională** a unei proteine, atunci analiza structurii primare ar putea oferi informații legate de proprietățile proteinei de interes. **Tipurile de informații ce pot fi extrase relativ ușor din analiza structurii primare a unei proteine în acest moment se clasifică în:**

- 1. Proprietățile fizico-chimice simple: masă moleculară, pI, indice de stabilitate, indice GRAVY;**
- 2. Numărul și ordinea structurilor secundare;**
- 3. Identificarea zonelor trans-membranare;**
- 4. Modelarea structurii tridimensionale complete.**

Indiferent de tipul lor, toate aceste informații **au caracter predictiv** și oferă un punct de plecare ce trebuie verificat experimental. **Informațiile sunt însă extrem de utile deoarece permit prioritizarea și orientarea demersului experimental.**

Ex.: Identificarea genelor pentru transportul nicotinei. Paenarthrobacter nicotinovorans este o bacterie ce degradează nicotina datorită prezenței unui megaplasmid ce conține 165 gene. Pentru ca acest alcaloid să pătrundă în celulă trebuie să fie transportat de un sistem de transport transmembranar. Analiza secvenței de aminoacizi celor 165 de proteine codificate de megaplasmid arată că 8 proteine au zone transmembranare. Deci cel mai probabil, una dintre aceste gene este responsabilă de transportul nicotinei. De la 165 de proteine candidate ca transportori ai nicotinei, o simplă analiză ce durează câteva minute cu costuri 0 reduce lista de candidați la 8.

1. Calcularea proprietăților fizico-chimice simple: ProtParam



Calcularea proprietăților fizico-chimice simple are la bază o serie de operații matematice simple ce țin cont de numărul și tipurile de aminoacizi dintr-o secvență polipeptidică precum și de proprietăților cunoscute ale acestora. Site-ul **ExpASy** (<https://www.expasy.org>) al Swiss Institute of Bioinformatics oferă numeroase programe de investigare a secvențelor printre care și programul **ProtParam** (<https://web.expasy.org/protparam/>).

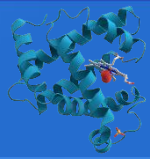
The screenshot shows the ExPASy Bioinformatics Resource Portal. At the top left is the SIB logo and the ExPASy logo. Below the logo is a search bar with the text 'Query all databases' and a search button. On the left side, there is a navigation menu with categories like 'Visual Guidance', 'Categories', 'proteomics', 'genomics', etc. The main content area is divided into 'Databases' and 'Tools' sections. The 'Databases' section lists various resources like UniProtKB, STRING, and SWISS-MODEL. The 'Tools' section lists various bioinformatics tools like 2ZIP, 3of5, AACompIdent, and BLAST.

 **ProtParam** • protein physical and chemical parameters • [\[more\]](#)

Protein Identification and Analysis Tools on the ExPASy Server;

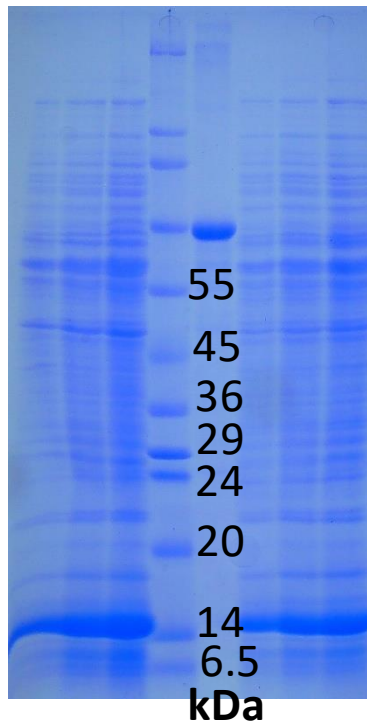
Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005); pp. 571-607

1. Calcularea proprietăților fizico-chimice simple: ProtParam



Parametrii calculați de **ProtParam**:

1. Masa moleculară Mw funcție de secvență – **suma maselor moleculare a tuturor aminoacizilor din secvența dată**; este exprimată în Dalton (Da); utilă pentru localizarea unei proteine pe un gel SDS-PAGE;



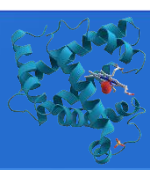
Unde mă aștept să fie localizată pe gelul SDS-PAGE din imagine proteina cu secvența:
MSSLALNENEDGCGLLAKLMATVRPEFRPEIVI PAPGSLVFNTEPCRI DECERQRTVK
GFCKSHYKRWLDQGRPEWESFLASPGPLVPGDGPLAACVTWCRFGSARRGLCVSHHG
FFSRSETDVIKWLATLSPEPVI ERPECRLAFCDLWAQGNSPLCLNHKSRWHAIGAPD
IDEFVARCESVGLDRFDFRCLADRPQLRLELQHAMQCRHDERRASTRS SVVTPVIRLV
SDSGVTSLLDWNLDQWAAFYIAGRVGRSHRHNGQLAFLRFAYTRLEDLVAGTGWSEY
SRDAWALHRLGYTDTRGTLRFDKIPQPWLRLAKRFIRWRLTSGREIIQGRVDILALN
RFAAFLAAPSPTPTARTASTAASWNASSPSSPGTNGQ

ProtParam: Molecular weight: 43263.17 Da
Theoretical pI: 8.93

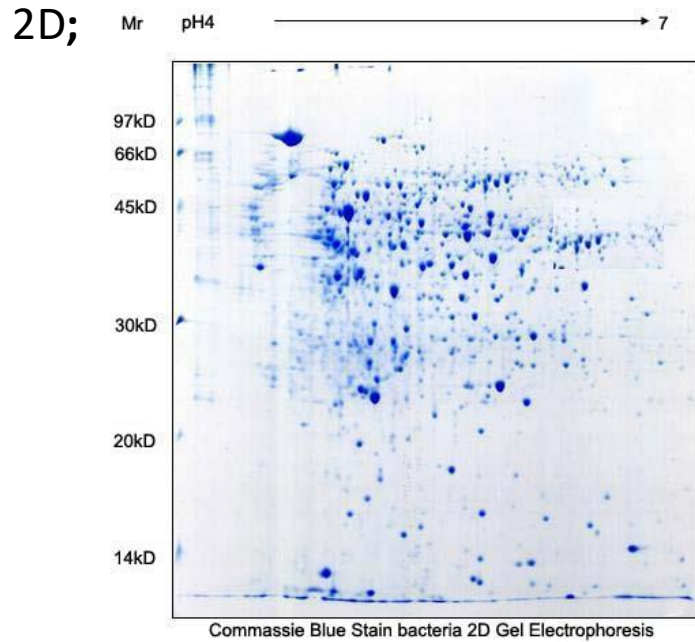
2. Compoziția atomică – calculează procentual ponderea fiecărui tip de atom din molecula proteică;

3. Compoziția de aminoacizi – calculează procentual ponderea fiecărui tip de aminoacid din molecula proteică;

1. Calcularea proprietăților fizico-chimice simple: ProtParam



4. pI – punctul izoelectric – valoarea de pH a unei soluții în care molecula proteică dată nu se deplasează în câmp electrostatic deoarece sarcina sa electrică de suprafață este zero; **ProtParam** calculează valoare pI luând în considerare valorile pK_a ale aminoacizilor determinate practic la o concentrație de 9.2M sau 9.8M uree și $15^\circ C$ sau $25^\circ C$; util pentru localizarea unei proteine pe un gel



Electrophoresis 1993, 14, 1023–1031

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

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

Prediction of focusing positions from amino acid sequences 1023

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 pI values. We show the advantages of being able to predict the position of a protein of known structure within a two-dimensional gel.

Electrophoresis 1994, 15, 529–539

Bengt Bjellqvist*
Bodil Basse
Eydfinnur Olsen
Julio E. Celis

Institute of Medical Biochemistry
and Danish Centre for Human
Genome Research, Aarhus
University, Aarhus

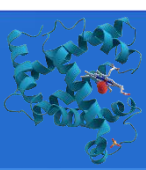
Reference points for comparisons of 2-D gel maps 529

Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions

A highly reproducible, commercial and nonlinear, wide-range immobilized pH gradient (IPG) was used to generate two-dimensional (2-D) gel maps of [^{35}S]methionine-labeled proteins from noncultured, unfractionated normal human epidermal keratinocytes. Forty one proteins, common to most human cell types and reported in the human keratome 2-D gel protein database

În calcularea M_w și pI programul ProtParam nu ține cont de eventualele modificări post-traducere ale proteinelor (PTM: acetylation, myristoylation; palmitoylation, prenylation, alkylation, glycosylation, hydroxylation, etc). Proteinele ce conțin aceste modificări vor avea mase molare și valori pI diferite față de cele teoretice calculate de ProtParam și deci vor migra diferite pe gelurile de poli-acrilamidă.

1. Calcularea proprietăților fizico-chimice simple: ProtParam



5. Coeficientul de extincție – cantitatea de lumină pe care proteina de interes o absoarbe la 280 nm; este utilă pentru a putea măsura direct cu precizie concentrația unei proteine purificate folosind un spectrofotometru;

ProtParam calculează acest parametru ținând cont de faptul că absorbția la 280 nm a unei catene polipeptidice este dependentă de numărul de resturi de tirozină (Y), triptofan (W) și cistină (două resturi de C legate printr-o punte disulfurică) după formula:

Coef Extincție Prot (280 nm) = $Nr(Y) * Ext(Y) + Nr(W) * Ext(W) + Nr(Cistină) * Ext(Cistină)$; unde

Coef Extincție Prot (280 nm) – coeficientul de extincție molară la 280 nm exprimat în $M^{-1} cm^{-1}$; **Nr(...)**- numărul de resturi de aminoacizi; **Ext(Y) = 1490**; **Ext(W) = 5500**; **Ext(Cistină) = 125**;

Exemplu de rezultate din ProtParam:

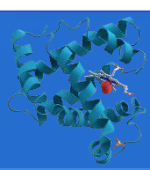
Extinction coefficients:

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

Ext. coefficient	79700
Abs 0.1% (=1 g/l)	1.842, assuming all pairs of Cys residues form cystines

Ext. coefficient	78950
Abs 0.1% (=1 g/l)	1.825, assuming all Cys residues are reduced

1. Calcularea proprietăților fizico-chimice simple: ProtParam



6. Timpul de înjumătățire estimat (*In vivo* half-life) – timpul necesar pentru ca jumătate din cantitatea de proteină de interes dintr-o celulă să dispară după ce a fost sintetizată;

ProtParam calculează acest parametru pornind de la așa numita **regulă a capătului N-terminal (N-end rule)** – timpul de înjumătățire a unei proteine depinde de **natura aminoacidului N-terminal**. Regula s-a stabilit pe baza unor observații experimentale ce au arătat că viteza de inactivare metabolică a unor beta-galactozidaze artificiale diferă funcție de aminoacidul terminal și de organismul în care sunt exprimate în limite foarte mari – 100 h până la mai puțin de 2 minute.

REVIEW

The N-end rule pathway of protein degradation

Alexander Varshavsky*

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

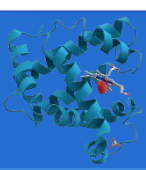
The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct versions of the N-end rule operate in all organisms examined, from mammals to fungi and bacteria. In eukaryotes, the N-end rule pathway is a part of the ubiquitin system. Ubiquitin 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.

Table 1 The N-end rule in *E. coli* and *S. cerevisiae*

Residue X	Half-life of X-βgal	
	<i>E. coli</i>	<i>S. cerevisiae</i>
Arg	2 min	2 min
Lys	2 min	3 min
Phe	2 min	3 min
Leu	2 min	3 min
Trp	2 min	3 min
Tyr	2 min	10 min
His	> 10 h	3 min
Ile	> 10 h	30 min
Asp	> 10 h	3 min
Glu	> 10 h	30 min
Asn	> 10 h	3 min
Gln	> 10 h	10 min
Cys	> 10 h	> 30 h
Ala	> 10 h	> 30 h
Ser	> 10 h	> 30 h
Thr	> 10 h	> 30 h
Gly	> 10 h	> 30 h
Val	> 10 h	> 30 h
Pro	?	> 5 h
Met	> 10 h	> 30 h

Approximate *in vivo* half-lives of X-βgal proteins in *E. coli* at 36 °C (Tobias *et al.* 1991) and in *S. cerevisiae* at 30 °C (Bachmair & Varshavsky 1989). A question mark at Pro indicates its uncertain status in the N-end rule (see the main text).

1. Calcularea proprietăților fizico-chimice simple: ProtParam



7. Indexul de instabilitate (instability index)- indică gradul de stabilitate al proteinei de interes *in-vitro*. Analiza statistică a **12 proteine labile și 32 proteine stabile** a demonstrat că există aproximativ 400 de dipeptide specifice a căror frecvență este semnificativ diferită în proteinele stabile *in-vitro* față de cele instabile. Pe baza tipului și numărului de dipeptide se calculează un index specific fiecărei proteine ce are semnificația:

- Dacă indexul este mai mic de **40**, proteina este cel mai probabil stabilă *in-vitro*;
- Dacă indexul este mai mare de **40**, proteina este cel mai probabil instabilă *in-vitro*;

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

Kunchur Guruprasad, B.V.Bhasker Reddy and Madhusudan W.Pandit¹

Centre for Cellular and Molecular Biology, Hyderabad - 500 007, India

¹To whom correspondence should be addressed

proteins. The statistical analytic dipeptide instability weight value possibility of its use in predicting would be unstable (or stable) in

8. Indexul alifatic (aliphatic index) – se definește ca fiind volumul relativ ocupat de catenele laterale alifatică ale aminoacizilor (alanină, valină, isoleucină și leucină); **cu cât acest index are o valoare mai mare, cu atât proteina va fi mai stabilă la temperatură;**

J. Mol. Biol. (1982) **157**, 105–132

9. GRAVY (grand average of hydropathicity) – un index ce ține cont de hidrofobicitatea și caracterul hidrofil al aminoacizilor din catena polipeptidică; **valori ale acestui index mai mici 0 de indică o proteină globulară, valori mai mari de 0 indică proteine membranare.**

A Simple Method for Displaying the Hydropathic Character of a Protein

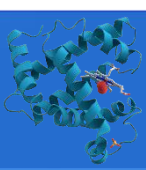
JACK KYTE AND RUSSELL F. DOOLITTLE

Department of Chemistry

University of California, San Diego, La Jolla, Calif. 92093, U.S.A.

(Received 8 August 1981, and in revised form 25 January 1982)

2. Predicția numărului și ordinii structurilor secundare



Scopul este de a stabili dacă un aminoacid dat dintr-o structură proteică face parte dintr-un α -helix, o structură β -pliată sau dintr-o buclă (**coil** sau **loop**).

Metodele de predicție a structurilor secundare se clasifică în:

1. Metode de primă generație – apărute în anii 60-70; au la bază alinieri de secvențe înrudite cu structuri cunoscute și **stabilirea predilecției** (propensity) **unor aminoacizi de a face parte dintr-o structură secundară specifică** sub forma unei probabilități;

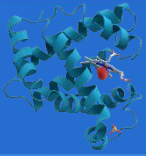
2. Metode de generația a doua – au fost utilizate intens până în anii 90; au dezvoltat conceptul de predilecție a aminoacizilor spre o anumită structură secundară și l-au extins către **secțiuni de aminoacizi cu lungime cuprinsă între 3-51 resturi; în calculul probabilității ca un anumit aminoacid să facă parte dintr-un structură secundară anume se ține cont de predilecția pe care o are întreg segmentul pentru acea structură;**

Indiferent de modalitatea practică de implementare, cele 2 tipuri de metode au o acuratețe de până la **60%** (**apartenența la o SS specifică este indicată corect pentru 6 din 10 aminoacizi și greșit pentru 4 din 10 aminoacizi**). Această acuratețe redusă se explică prin:

-interacțiunile locale (dintre aminoacizii vecini) contribuie cu aproximativ 65% la generarea structurii secundare a unei proteine, restul de 35% fiind realizat de interacțiuni dintre structuri secundare;

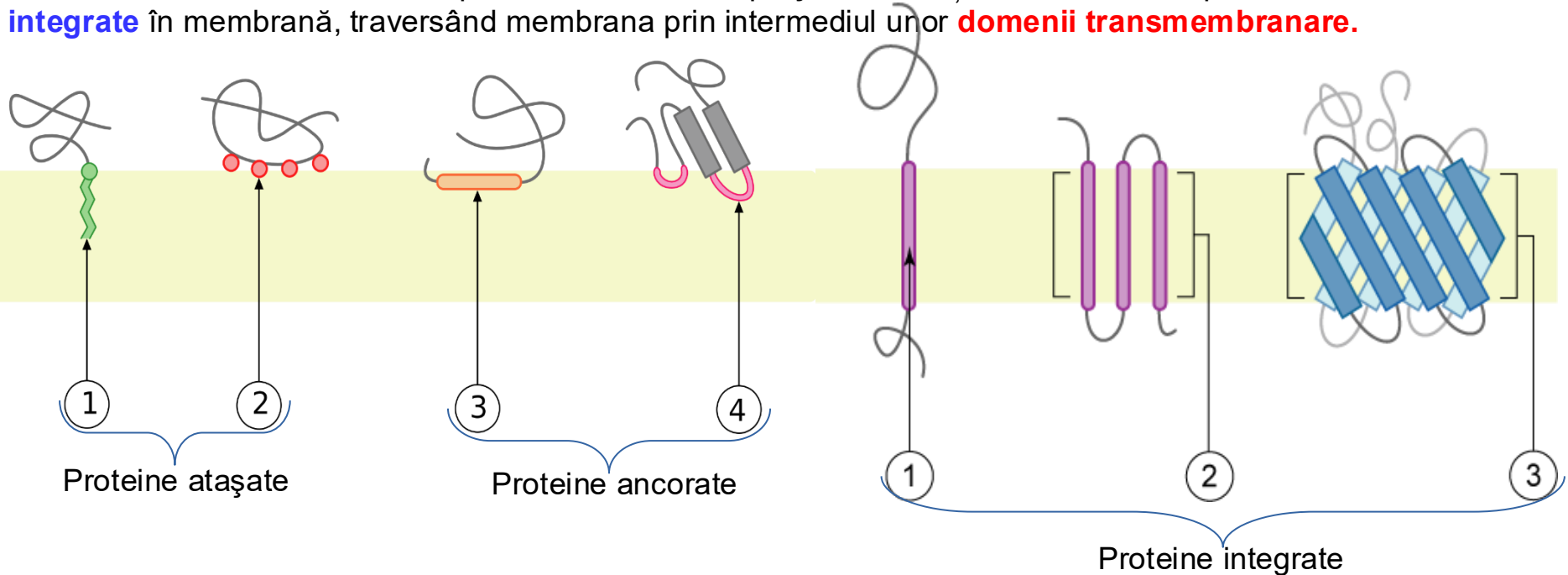
-Legăturile de H implicate în formarea structurilor β -pliate se realizează între aminoacizi amplasați la distanțe mari în catenă.

3. Identificarea zonelor trans-membranare



Proteinele membranare pot fi:

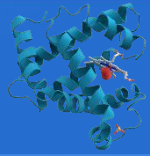
- a. **atașate** pe suprafața membranei celulare prin intermediul unui fosfolipid;
- b. **ancorate** în membrana celulară prin intermediul unei porțiuni ce conține aminoacizi nepolari;
- c. **integrate** în membrană, traversând membrana prin intermediul unor **domenii transmembranare**.



Funcție de structura secundară a zonei transmembranare, proteinele integrate se clasifică în două clase:

- TMH** (trans-membrane helices) – proteine integrate prin domenii alcătuite exclusiv din una sau mai multe structuri helicale (1-2 din figura). În cazul proteinelor **TMH cu multiple helix-uri** acestea sunt **grupate într-un manunchi și unite prin bucle intra- sau extra-citoplasmatică** ;
- TMB** (trans membrane β -strands barrels) – pori transmembranari ce **conțin exclusiv structuri β -pliate** grupate într-o structură asemănătoare unui **butoi** și unite prin bucle intra- sau extra-citoplasmatică (3 din figură).

3. Identificarea zonelor trans-membranare



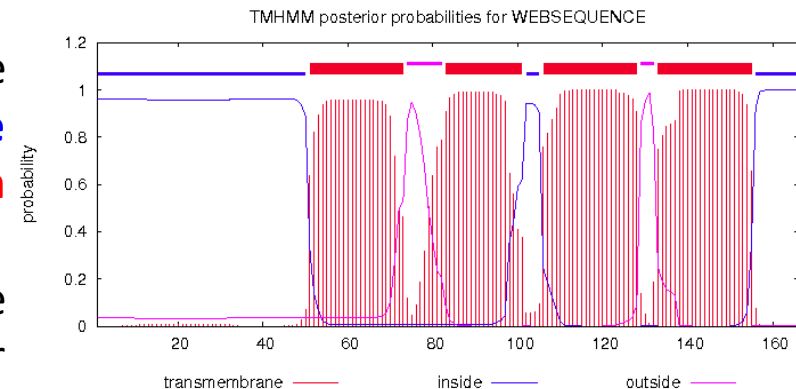
Datorită condițiilor specifice deosebit de uniforme (în raport cu citosolul) din membrana celulară, procesul de identificare și predicție a zonelor transmembranare este deosebit de robust și se bazează pe următoarele observații experimentale:

- **zonele transmembranare conțin predominant aac. nepolari hidrofobi;**
- **helix-urile transmembranare au între 12-35 aminoacizi lungime (grosimea membranei celulare);**
- majoritatea proteinelor TMH respectă așa numita regulă a "**sarcinilor pozitive în interior**" ce se referă la distribuția aminoacizilor pozitivi (K și R) – buclele ce conectează două helix-uri transmembranare amplasate spre interiorul celulei conțin mai multe sarcini pozitive decât cele amplasate în exterior;

- în cazul în care între două helix-uri transmembranare există **zone globulare** acestea au în general **mai puțin de 60 de** resturi și respectă regula "**sarcinilor pozitive în interior**";
- zonele globulare dintre două helix-uri transmembranare **mai mari de 60 resturi** nu respectă regula "sarcinilor pozitive în interior";

Exemple de programe pentru identificarea zonelor transmembranare:

- Topred2 - http://sbc.bioch.ox.ac.uk/TM_noj/TM_noj.html
- TMPred - https://embnet.vital-it.ch/software/TMPRED_form.html
- HMMTOP - <http://www.enzim.hu/~tusi/hmmtop/>
- TMHMM - <http://www.cbs.dtu.dk/services/TMHMM-2.0/>



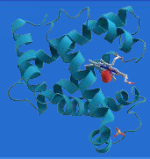


TABLE 29.1. Availability of Prediction Methods^a

Method	Type	Server	Program
ACCpro	acc	www.ics.uci.edu/~baldig/scratch/index.html	Gianluca Pollastri
FKNNacc	acc	hinja.ssu.ac.kr/FKNNacc	Julian Lee
PHDacc	acc	www.predictprotein.org/	Guy Yachdav
PROFacc	acc	www.predictprotein.org/	Guy Yachdav
SABLE	acc	sable.cchmc.org	Jarek Meller
ASP	sec +	www.predictprotein.org/	Malin Young
COILS	sec	www.predictprotein.org/	Andrei Lupas
HMMSTR	sec +	www.bioinfo.rpi.edu/~bystrc/hmmstr/server.php	Chris Bystroff
Porter	sec	distill.ucd.ie/porter	Gianluca Pollastri
PROFphd	sec	www.predictprotein.org/	Guy Yachdav
PSIPRED	sec	insulin.brunel.ac.uk/psiform.html	David Jones
SABLE	sec	sable.cchmc.org	Jarek Meller
SAM-T99sec	sec	www.cse.ucsc.edu/research/compbio/HMM-apps/T99-query.html	Kevin Karplus
SSpro2	sec	promoter.ics.uci.edu/BRNN-PRED	Pierre Baldi Gianluca Pollastri
YASPIN	sec	zeus.cs.vu.nl/programs/yaspinwww	Jaap Heringa
B2TMR	tmb	gpcr.biocomp.unibo.it/biodec	Rita Casadio
BOMP	tmb	www.bioinfo.no/tools/bomp	Ingvar Eidhammer
PROFtmb	tmb	www.predictprotein.org/	Henry Bigelow
TMB-HUNT	tmb	www.bioinformatics.leeds.ac.uk/~andy/betaBarrel/AACompPred/aaTMB_Hunt.cgi	Andrew Garnon DR Westhead
HMMTOP	tmh	www.enzim.hu/hmmtop	Gábor E. Tusnady
MEMSAT	tmh	insulin.brunel.ac.uk/psipred	David Jones
MINNOU	tmh	minnou.cchmc.org	Jarek Meller
PHDhtm	tmh	www.predictprotein.org/	Guy Yachdav
Phobius	tmh	phobius.cgb.ki.se	Erik Sonnhammer
SOSUI	tmh	sosui.proteome.bio.tuat.ac.jp/sosuiframe0.html	Takatsugu Hirokane
SPLIT	tmh	www.mbb.ki.se/tmap/index.html	Davor Juricic
TMAP	tmh	www.mbb.ki.se/tmap/index.html	Bengt Persson
TMHMM	tmh	www.cbs.dtu.dk/services/TMHMM-1.0	Anders Krogh
TMpred	tmh	www.ch.embnet.org/software/TMPRED_form.html	Philipp Bucher
TopPred2	tmh	www.sbc.su.se/~erikw/TopPred22	Gunnar von Heijne

^aAcronyms: prediction of: acc—solvent accessibility; sec—secondary structure; tmb—transmembrane beta strands; tmh—transmembrane helices.