

# Razvoj metaprediktora neuređenosti proteina

Master rad iz oblasti  
Bioinformatika  
Matematički fakultet

Una Stanković  
una\_stankovic@yahoo.com

18. jul 2018.

## Sažetak

U ovom radu biće predstavljen proces kreiranja metaprediktora za analizu neuređenosti proteina. Najpre će biti objašnjeni svi pojmovi vezani za biološke osnove ovog problema, a potom i sam razvojni proces iz informatičkog ugla. Sav softver primenjen pri izradi rada je korišćen isključivo u akademske svrhe.

## Sadržaj

<b>1</b>	<b>Uvod</b>	<b>2</b>
<b>2</b>	<b>Biološke osnove</b>	<b>2</b>
2.1	Proteini . . . . .	2
2.1.1	Funkcije i osobine proteina . . . . .	2
2.1.2	Struktura proteina . . . . .	3
2.1.3	Savijanje proteina . . . . .	4
2.1.4	Denaturacija proteina . . . . .	4
<b>3</b>	<b>Prediktori</b>	<b>4</b>
3.1	SPOT-Disorder prediktor . . . . .	4
<b>4</b>	<b>Zaključak</b>	<b>4</b>
	<b>Literatura</b>	<b>4</b>
<b>A</b>	<b>Literatura</b>	<b>5</b>

# 1 Uvod

Neuređenost proteina

## 2 Biološke osnove

U ovoj sekciji biće ukratko predstavljene biološke osnove neophodne za razumevanje rada i motivacije koja stoji iza određenih njegovih elemenata.

Najpre, biće opisano šta su proteini koje su njihove osnovne funkcije i kakva im je struktura. Potom, biće opisana svaka od struktura ponao-sob, uz priložen grafički prikaz istih. Na kraju, posebno će biti opisani neuređeni proteini, njihova uloga i uzroci koji mogu dovesti do njihove pojave.

### 2.1 Proteini

Proteini (grč. *protos* - "zauzimam prvo mesto") su najvažniji sastojak žive materije i utiču na brojnost i raznolikost živih bića. Specifičnost proteina je tolika da svaka biljna i životinjska vrsta ima svoje proteine, a kod viših organizama oni se razlikuju i na individualnom nivou. Broj proteina u živim bićima je ogroman, a kao primer uzmimo *E.coli* sa 3000 i čoveka sa 5 miliona proteina. [4]

Proteini i peptidi su linearni polimeri izgrađeni od 22 *L* – *aminokiseline* <sup>1</sup> koje se javljaju u prirodi i povezani su peptidnim vezama. [1]

#### 2.1.1 Funkcije i osobine proteina

Proteini su biološki najaktivniji molekuli sa velikim brojem esencijal-nih uloga koje se dele na:

- dinamičke, od kojih su najvažnije:
  1. transportna - prenos molekula (kiseonika, gvožđa, lipida) i hor-mona od mesta sinteze do mesta delovanja,
  2. biološka - regulacija metaboličkih procesa u ćeliji, kontrola i regulacija transkripcije gena i translacija,
  3. katalizatorska - biološka katalizacija <sup>2</sup>,
  4. zaštitna - keratin, koagulacija krvi,
  5. održavanje zapremine tečnosti u organizmu,
- strukturne
  1. obezbeđivanje čvrstine i elastičnosti organa,
  2. davanje oblika organizmu,
  3. izgradnja strukturnih elemenata ćelije i
  4. bitna uloga u kontraktilnim i pokretnim elementima organizma.

Postoji nekoliko osobina koje karakterišu proteine, to su:

- izgradnja kompleksnih jedinjenja sa različitim supstancama po prin-cipu strukturne komplementarnosti i

---

<sup>1</sup>*L* – *aminokiseline* su one sa levom prostornom konfiguracijom, analogno, postoje i *D* – *aminokiseline*, sa desnom

<sup>2</sup>Katalizacija predstavlja proces povećavanja brzina reakcija

- visoka osetljivost na različite agense koji ih denaturišu <sup>3</sup>. Neki od agenasa su: visoka temperatura, pritisak, mehaničko tretiranje, dejstvo kiselina, baza, organskih rastvarača, materija, itd. [4]

### 2.1.2 Struktura proteina

Proteini imaju specifičnu trodimenzionalnu strukturu koja se formira presavijanjem polipeptidnog lanca na različite načine. Unutrašnjost takve strukture ima visoku gustinu, pa takav lanac ne dopušta promene u sastavu i zahteva prisustvo aminokiselina tačno određene veličine. Uobičajena raspodela aminokiselina u proteinima je daleko od ravnomerne. Neke aminokiseline se javljaju mnogo češće od ostalih, tako se, na primer, leucin pojavljuje devet puta više od triptofana.

Niska aminokiselina sadrži sve potrebne informacije kako bi se formirala trodimenzionalna struktura proteina, za koju se smatra da je najstabilnija. Proteinsku strukturu održavaju različite vrste kovalentnih i nekovalentnih interakcija između hemijskih skupina, npr. vodonične, jonske, elektrostatičke, dipolne, itd. Nabiranjem i uvijanjem lanaca nastaju različiti oblici proteina: vlaknasti, globularni ili eliptični. [2] Ako mutacija dovede do toga da aminokiselina sa malim bočnim lancem bude zamenjena aminokiselinom sa velikim, pojaviće se problem sa formiranjem trodimenzionalne strukture. Ako bi se, pak, velika aminokiselina zamenila sa malom pojavio bi se prazan prostor, što bi moglo dovesti do destabilizacije molekula proteina.

U molekulima proteina postoji hijerarhijska strukturalna organizacija u četiri nivoa:

1. primarna,
2. sekundarna,
3. tercijarna i
4. kvaternarna. [4]

**Primarna struktura** Primarna struktura proteina predstavlja samu sekvencu aminokiselina <sup>4</sup>. Ona je od ključnog značaja za funkciju proteina zbog interakcija između bočnih lanaca aminokiselina koji određuju trodimenzionalnu strukturu. Proteini koji imaju sličnu sekvencu aminokiselina su *homologi*, a poređenje sekvenci među takvim proteinima može ukazati na genetsku relaciju između različitih vrsta.

Mnoge genetske bolesti rezultuju u proteinima sa abnormalnim redosledom aminokiselina, što uzrokuje nepravilno presavijanje i gubitak ili nemogućnost normalnog funkcionisanja. Ukoliko su nam poznate strukture normalnih i mutiranih proteina te informacije možemo iskoristiti za dijagnostifikovanje ili proučavanje bolesti. [3]

### Sekundarna struktura

### Tercijarna struktura

<sup>3</sup>Denaturizacija proteina je proces koji izaziva promene u strukturi proteina menjajući i njihovo fiziološko dejstvo.

<sup>4</sup>Redosled kojim su aminokiseline poredane u nekom polipeptidu se zove sekvenca aminokiselina. [4]

## Kvaternarna struktura

### 2.1.3 Savijanje proteina

### 2.1.4 Denaturacija proteina

## 3 Prediktori

- SPINE-D/SPOT-Disorder
- PONDR
- s2D
- IUPred
- ESpritz
- SEG
- Disopred2

### 3.1 SPOT-Disorder prediktor

Abstract MOTIVATION: Capturing long-range interactions between structural but not sequence neighbors of proteins is a long-standing challenging problem in bioinformatics. Recently, long short-term memory (LSTM) networks have significantly improved the accuracy of speech and image classification problems by remembering useful past information in long sequential events. Here, we have implemented deep bidirectional LSTM recurrent neural networks in the problem of protein intrinsic disorder prediction. RESULTS: The new method, named SPOT-Disorder, has steadily improved over a similar method using a traditional, window-based neural network (SPINE-D) in all datasets tested without separate training on short and long disordered regions. Independent tests on four other datasets including the datasets from critical assessment of structure prediction (CASP) techniques and >10 000 annotated proteins from MobiDB, confirmed SPOT-Disorder as one of the best methods in disorder prediction. Moreover, initial studies indicate that the method is more accurate in predicting functional sites in disordered regions. These results highlight the usefulness combining LSTM with deep bidirectional recurrent neural networks in capturing non-local, long-range interactions for bioinformatics applications. AVAILABILITY AND IMPLEMENTATION: SPOT-disorder is available as a web server and as a standalone program at: <http://sparks-lab.org/server/SPOT-disorder/index.php> .

## 4 Zaključak

## Literatura

- [1] Gerhard Michal Dietmar Schomburg. *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2012.
- [2] Ivana Čepelak Dubravka Čvorišćec. *Štrausova medicinska biokemija*. Medicinska naklada, 2009.

- [3] Denise R. Ferrier Richard A. Harvey. *Lippincott Illustrated Reviews: Biochemistry, 5th edition*. Lippincott Williams & Wilkins, a Wolters Kluwer business, 351 West Camden Street, Baltimore, MD 21201, Two Commerce Square, 2001 Market Street, Philadelphia, PA 19103, 2011.
- [4] Vesna Spasojević-Kalimanovska Slavica Spasić, Zorana Jelić-Ivanović. *Opšta biohemija*. 2002.

## **A   Literatura**