Bioinformatics analysis of correlation between protein function and intrinsic disorder

Goran Vinterhalter

University of Belgrade Faculty of Mathematics

June 18, 2018

Inspiration

Reference work: Xie. et. al 2006

Functional anthology of intrinsic disorder. 1.

Biological processes and functions of proteins with long disordered regions

Inspiration

Reference work: Xie. et. al 2006

Functional anthology of intrinsic disorder. 1.

Biological processes and functions of proteins with long disordered regions

- *Swiss-Prot* v48 from 2005. (about 200 000 seq.)

Inspiration

Reference work: Xie. et. al 2006

Functional anthology of intrinsic disorder. 1.

Biological processes and functions of proteins with long disordered regions

- *Swiss-Prot* v48 from 2005. (about 200 000 seq.)
- UniprotKB Keywords

Inspiration

Reference work: Xie. et. al 2006

Functional anthology of intrinsic disorder. 1.

Biological processes and functions of proteins with long disordered regions

- *Swiss-Prot* v48 from 2005. (about 200 000 seq.)
- UniprotKB Keywords
- PONDR VSL3 disorder predictor

Used data

- *CAFA3* dataset (training set from CAFA3 challenge (2016))

Used data

- *CAFA3* dataset (training set from CAFA3 challenge (2016))
- GO terms

Used data

- *CAFA3* dataset (training set from CAFA3 challenge (2016))
- GO terms
- *CAFA3* dataset is a subset of *Swiss-Prot*

Used data

- *CAFA3* dataset (training set from CAFA3 challenge (2016))
- GO terms
- *CAFA3* dataset is a subset of *Swiss-Prot*
- We assume the *CAFA3* dataset is not statictially redundant

Predicting disorder

PONDR VSL2b disorder predictor

Predicting disorder

PONDR VSL2b disorder predictor

Definition

Protein is labled as **putatively disorderd** *if it contains a predicted IDPr at least 40 AA long, otherwise putatively orderd.*

-

$$S_L = \{s_i : |L - |s_i|| \le l\}, \quad l = 0.1 \cdot L \quad \text{and } |s_i| \text{ is sequence length}$$

-
  $$S_L = \{s_i : |L - |s_i|| \leq l\}, \quad l = 0.1 \cdot L \quad \text{and } |s_i| \text{ is sequence length}$$

- The probability a protein of length L is labeld putatively disorderd by chance is aproximated as:

  $$P_L = \frac{\sum_{s_i \in S_L} d(s_i)}{|S_L|}, \quad |S_L| \text{ is set cardinality}$$

  $$d(s_i) = \begin{cases} 1 & \text{if } s_i \text{ is } \textbf{putatively disorderd} \\ 0 & \text{otherwise} \end{cases}$$
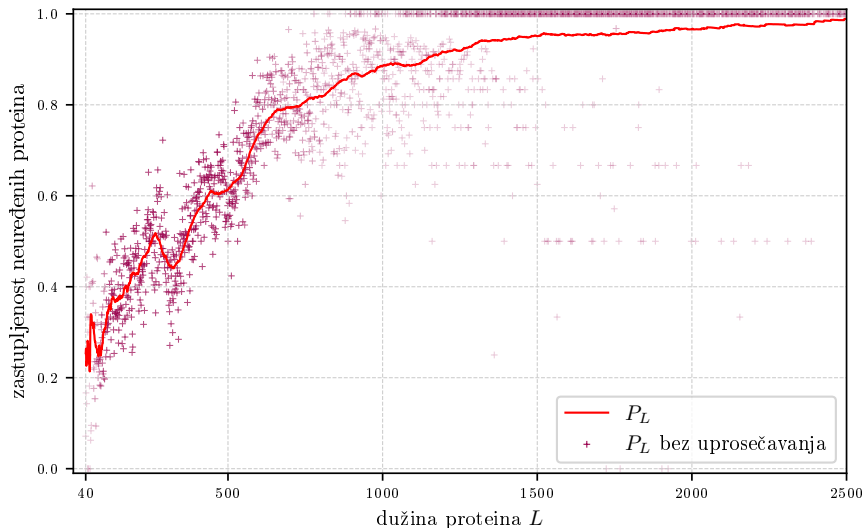
Zavisnost neuređenosti i dužine proteina iz CAFA3-skupa minimalne dužine 40 AK

- random model ($P_L random$)
- uniform model ($P_L uniform$)
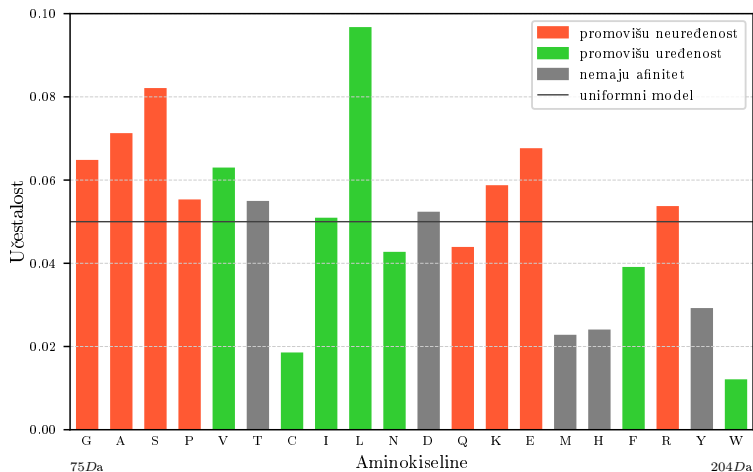
Amino acide frequency in CAFA3 dataset
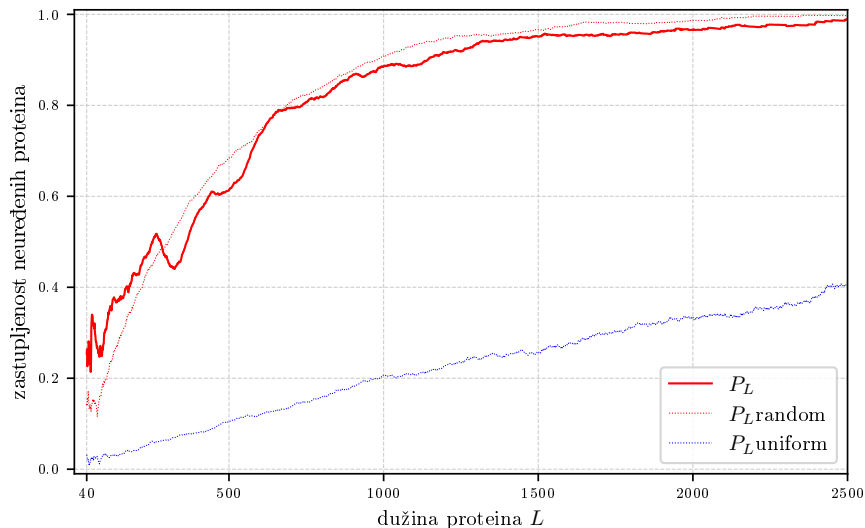
Comparing P_L, $P_L random$ and $P_L uniform$ on CAFA3 dataset

Correlation between function and (dis)order

- Let S_j be a set of proteins anotated by function j.
  Then by **disorder level** of function j we assume the frequency (F_j) of putatively disorderd proteins inside set S_j.

$$F_j = \frac{\sum_{s_i \in S_j} d(s_i)}{|S_j|}$$

Correlation between function and (dis)order

- Let S_j be a set of proteins anotated by function j.
  Then by **disorder level** of function j we assume the frequency (F_j) of putatively disorderd proteins inside set S_j.

$$F_j = \frac{\sum_{s_i \in S_j} d(s_i)}{|S_j|}$$

-

$$Y_j = \frac{\sum_{s_i \in S_j} X_{|s_j|}}{|S_j|} \qquad X_L : \begin{pmatrix} 0 & 1 \\ P_L & 1 - P_L \end{pmatrix}$$

Correlation between function and (dis)order

- Let S_j be a set of proteins anotated by function j.
  Then by **disorder level** of function j we assume the frequency (F_j) of putatively disorderd proteins inside set S_j.

$$F_j = \frac{\sum_{s_i \in S_j} d(s_i)}{|S_j|}$$

-

$$Y_j = \frac{\sum_{s_i \in S_j} X_{|s_j|}}{|S_j|} \qquad X_L : \begin{pmatrix} 0 & 1 \\ P_L & 1 - P_L \end{pmatrix}$$

  - $p > 0.95$: function is **disorder dependent** (correlated with ...)
  - $p < 0.05$: function is **order dependent**
  - otherwise: nothing can be said about correlation

Definitions

Definitions

Definition

Neuređenost *funkcije (GO-termina ili ključne reči) je mera učestalosti neuređenih proteina (F_j).*

Definitions

Definition

Neuređenost *funkcije (GO-termina ili ključne reči) je mera učestalosti neuređenih proteina (F_j).*

Definition

Function (GO term or keyword) is:

- **disorder dependent** *if it's correlated ($p < 0.05$) with p. dissorderd p.*
- **order dependent** *if it's correlated ($p > 0.95$) with p. orderd p.*

Definitions

Definition

Neuređenost *funkcije (GO-termina ili ključne reči) je mera učestalosti neuređenih proteina (F_j).*

Definition

Function (GO term or keyword) is:

- **disorder dependent** *if it's correlated ($p < 0.05$) with p. dissorderd p.*
- **order dependent** *if it's correlated ($p > 0.95$) with p. orderd p.*

Definition

Statistički najznačajnije *uređene/neuređene funkcije (GO-termini ili ključne reči) su one koje imaju najmanji/najviši Z-skor.*

1. Objedinjavanje baze *Swiss-Prot* i njenog CAFA3-podskupa

1. Objedinjavanje baze *Swiss-Prot* i njenog CAFA3-podskupa
   - Dobijeno je 66 530 validnih proteina (od ukupno 66 599)
   - Anotacije (term & kw) su preuzete iz baze *Swiss-Prot*
   - Sekvence su zadržane originalne iz CAFA3-skupa

1. Objedinjavanje baze *Swiss-Prot* i njenog CAFA3-podskupa
   - Dobijeno je 66 530 validnih proteina (od ukupno 66 599)
   - Anotacije (term & kw) su preuzete iz baze *Swiss-Prot*
   - Sekvence su zadržane originalne iz CAFA3-skupa
2. Grupisanje proteina po GO-terminima

1. Objedinjavanje baze *Swiss-Prot* i njenog CAFA3-podskupa
   - Dobijeno je 66 530 validnih proteina (od ukupno 66 599)
   - Anotacije (term & kw) su preuzete iz baze *Swiss-Prot*
   - Sekvence su zadržane originalne iz CAFA3-skupa
2. Grupisanje proteina po GO-terminima
   - Dobijeno je 1781 MF-termin koji anotira min. 20 proteina. Bez grupisanja bilo bi samo 1146 MF-termina.

1. Objedinjavanje baze *Swiss-Prot* i njenog CAFA3-podskupa
   - Dobijeno je 66 530 validnih proteina (od ukupno 66 599)
   - Anotacije (term & kw) su preuzete iz baze *Swiss-Prot*
   - Sekvence su zadržane originalne iz CAFA3-skupa
2. Grupisanje proteina po GO-terminima
   - Dobijeno je 1781 MF-termin koji anotira min. 20 proteina. Bez grupisanja bilo bi samo 1146 MF-termina.
3. Preslikavanje između ključnih MF-reči i MF-termina

1. Objedinjavanje baze *Swiss-Prot* i njenog CAFA3-podskupa
   - Dobijeno je 66 530 validnih proteina (od ukupno 66 599)
   - Anotacije (term & kw) su preuzete iz baze *Swiss-Prot*
   - Sekvence su zadržane originalne iz CAFA3-skupa

2. Grupisanje proteina po GO-terminima
   - Dobijeno je 1781 MF-termin koji anotira min. 20 proteina. Bez grupisanja bilo bi samo 1146 MF-termina.

3. Preslikavanje između ključnih MF-reči i MF-termina

	ukupno	nema pres.	pres. MF	pres. BP	pres. CC.
ključ. MF-reči	195	20	104	54	11

Metod sličnih anotacija

Metod sličnih anotacija

- Žakardov indeks (engl. *Jaccard index*), kraće **Ji**

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

$$J(A, B) \in [0, 1]$$

Metod sličnih anotacija

- Žakardov indeks (engl. *Jaccard index*), kraće **Ji**

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

$$J(A, B) \in [0, 1]$$

- U obzir su uzeti svi proteini iz baze *Swiss-Prot*

Metod sličnih anotacija

- Žakardov indeks (engl. *Jaccard index*), kraće **Ji**

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

$$J(A, B) \in [0, 1]$$

- U obzir su uzeti svi proteini iz baze *Swiss-Prot*
- Nije vršeno grupisanje po MF-terminima

Metod sličnih anotacija

- Žakardov indeks (engl. *Jaccard index*), kraće **Ji**

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

$$J(A, B) \in [0, 1]$$

- U obzir su uzeti svi proteini iz baze *Swiss-Prot*
- Nije vršeno grupisanje po MF-terminima
- Tačna predviđanja između 60 i 85 posto

Metod sličnih anotacija

- Žakardov indeks (engl. *Jaccard index*), kraće **Ji**

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

$$J(A, B) \in [0, 1]$$

- U obzir su uzeti svi proteini iz baze *Swiss-Prot*
- Nije vršeno grupisanje po MF-terminima
- Tačna predviđanja između 60 i 85 posto
- Preslikavanja su dopunjena sa 64 izvedena preslikavanja ($Ji > 0.1$).

ključna MF-reč	n_kw	Ji	n_go	MF-termin
Dermonecrotic toxin	148	0.96	142	phospholipase D activity
Ribosomal protein	49054	0.91	48096	structural constituent of ribosome
Complement sys. impairing toxin	160	0.81	142	phospholipase D activity
Hemagglutinin	397	0.75	299	host cell surface receptor binding
Mutator protein	255	0.75	288	damaged DNA binding
Antifreeze protein	10	0.7	7	ice binding
Light-harvesting polypeptide	90	0.68	61	bacteriochlorophyll binding
Cyclin	197	0.61	124	cyclin-dependent protein ...
Defensin	55	0.55	32	CCR6 chemokine receptor binding
Ribonucleoprotein	50698	0.54	28317	rRNA binding
Neurotoxin	2734	0.53	4145	toxin activity
Photoprotein	40	0.48	19	alkanal monooxygenase ...
Endorphin	48	0.45	32	opioid peptide activity
Protein synthesis inhibitor	150	0.43	67	rRNA N-glycosylase activity
Neuropeptide	561	0.42	267	neuropeptide hormone activity
Signal transduction inhibitor	157	0.38	158	GTPase activator activity
Mitogen	282	0.37	284	growth factor activity
Repressor	8177	0.33	7798	DNA binding
...	...	...	...	...
Vasoactive	243	0.17	489	hormone activity
Chromatin regulator	1939	0.12	861	chromatin binding
Developmental protein	6285	0.12	2464	sequence-specific DNA binding

Rezultati

	referentni rezultati	novi rezultati	
	Xie2007 kw	k.r. MF	MF-termini
ukupno ($br.prot \geq 20$)	143	186	1781
$p < 0.05$ (uređene)	37	53	699
$p > 0.95$ (neuređene)	51	44	616
suprotno (nepoznato)	55	89	576

Uopšteno poređenje rezultata

Rezultati

	referentni rezultati	novi rezultati	
	Xie2007 kw	k.r. MF	MF-termini
ukupno ($br.prot \geq 20$)	143	186	1781
$p < 0.05$ (uređene)	37	53	699
$p > 0.95$ (neuređene)	51	44	616
suprotno (nepoznato)	55	89	576

Uopšteno poređenje rezultata

Dve vrste rezultata:

- Tabelarno poređenje sa ključnim rečima
- Grafovsko predstavljanje

20 statistički najznačajnijih uređeneih ključnih MF-reči (Xie et. al. 2007)

#	name	n prot	n families	avg_len	Z-score	P-value
0	Oxidoreductase	14995	992	376.63	-29.54	0
1	Transferase	26525	1606	445.17	-24.25	0
2	Lyase	7262	347	377.92	-22.64	0
3	Hydrolase	20464	1995	430.68	-21.75	0
4	Isomerase	4487	220	383.98	-14.18	0
5	Glycosidase	1826	244	444.13	-13.98	0
6	Glycosyltransferase	2950	261	437.53	-12.51	0
7	Acyltransferase	2239	179	402.83	-10.85	0
8	Methyltransferase	3524	224	349.60	-10.53	0
9	Kinase	7017	322	448.29	-10.22	0
10	Ligase	8010	230	529.41	-10.06	0
11	Decarboxylase	1293	63	345.26	-9.66	0
12	Monooxygenase	1668	73	444.87	-9.26	0
13	Metalloprotease	1100	109	553.73	-7.89	0
14	Aminopeptidase	452	39	509.17	-7.55	0
15	Dioxygenase	360	66	433.20	-7.32	0
16	Aminoacyl-tRNA synthetase	3402	37	571.83	-7.15	0
17	Protease	4423	380	549.70	-7.1	0
18	Aminotransferase	955	28	420.27	-6.02	0

Uređene ključne MF-reči iz ovog rada (z-skor > -0.2)

#	name	n	avg_len	avg_dis	z	p
0	Oxidoreductase	4126	472.25	0.28	-41.35	0
1	Hydrolase	7564	614.81	0.51	-26.99	0
2	Lyase	1431	481.37	0.30	-23.62	0
3	Monooxygenase	555	503.36	0.20	-30.29	0
4	Transferase	8846	631.91	0.55	-19.72	0
5	Ligase	995	693.30	0.46	-18.05	0
6	Glycosyltransferase	1134	551.26	0.40	-17.04	0
7	Glycosidase	697	570.50	0.37	-16.81	0
8	Isomerase	931	422.72	0.35	-13.60	0
9	Protease	1863	674.42	0.54	-13.20	0
10	Transducer	1703	482.28	0.41	-12.56	0
11	G-protein coupled receptor	1385	465.62	0.39	-12.45	0
12	Acyltransferase	867	531.58	0.42	-11.28	0
13	Decarboxylase	195	488.21	0.25	-10.70	0
14	Aminotransferase	202	451.05	0.24	-10.23	0
15	Aminopeptidase	130	668.72	0.37	-9.10	0
16	Serine protease	460	700.07	0.50	-8.87	0
17	Metalloprotease	507	688.25	0.56	-8.43	0
18	Methyltransferase	874	611.22	0.47	-8.33	0
19	Carboxypeptidase	116	631.16	0.37	-8.25	0
20	Threonine protease	138	246.88	0.18	-7.50	0
21	Dioxygenase	366	622.32	0.48	-7.39	0
22	Serine esterase	141	423.09	0.28	-7.21	0
23	Receptor	3424	647.92	0.59	-7.09	0
24	Nucleotidyltransferase	600	969.68	0.63	-6.18	0
25	Peroxidase	452	651.47	0.40	-5.50	0
26	Serine protease inhibitor	182	497.66	0.38	-4.92	0
27	Porin	63	318.84	0.21	-4.55	0
28	Dipeptidase	22	522.27	0.23	-4.43	0
29	Integrin	71	1038.25	0.70	-4.20	0
30	Protease inhibitor	284	446.29	0.41	-4.15	0
31	RNA-directed RNA polymerase	62	2506.11	0.87	-4.11	0
32	Neurotoxin	66	209.62	0.17	-3.80	0
33	Hemagglutinin	23	281.22	0.13	-3.37	0
34	Retinal protein	47	384.06	0.28	-3.37	0
35	Myosin	185	1275.20	0.72	-3.29	0
36	Photoreceptor protein	77	540.68	0.47	-3.02	0.01
37	Prenyltransferase	39	376.21	0.31	-2.59	0.01
38	Elongation factor	106	467.68	0.48	-2.59	0.01
39	Endonuclease	443	728.14	0.60	-2.59	0.01
40	Toxin	214	412.51	0.39	-2.56	0
41	Bacteriolytic enzyme	35	434.66	0.31	-2.53	0.01
42	Nuclease	703	668.54	0.60	-2.49	0.01
43	Ion channel impairing toxin	60	74.92	0.15	-2.47	0
44	Voltage-gated sodium channe...	22	79.55	0.05	-2.42	0.01
45	Aspartyl protease	120	942.30	0.64	-2.39	0.01
46	Hemostasis impairing toxin	55	297.55	0.35	-2.10	0.02
47	Thiol protease inhibitor	63	411.06	0.35	-2.04	0.03
48	tRNA-binding	119	575.46	0.56	-2.01	0.03

Uređeni MF-termini dobijeni iz direktnog i izvedenog mapiranja

| #	name	n	avg_len	avg_dis	z	p
0	catalytic activity	28179	590.86	0.48	-52.25	0
1	oxidoreductase activity	4986	450.49	0.30	-39.62	0
2	hydrolase activity	11097	623.94	0.53	-27.76	0
3	lyase activity	1624	480.74	0.31	-24.47	0
4	monooxygenase activity	633	511.13	0.22	-21.40	0
5	transporter activity	5388	621.16	0.53	-19.15	0
6	transferase activity	10428	619.96	0.56	-18.80	0
7	transferase activity, trans...	1343	545.86	0.41	-18.49	0
8	ligase activity	1097	675.42	0.46	-16.68	0
9	hydrolase activity, acting ...	934	548.13	0.39	-16.32	0
10	isomerase activity	1046	425.93	0.35	-14.19	0
11	transferase activity, trans...	1235	540.97	0.43	-13.31	0
12	carboxy-lyase activity	292	494.47	0.26	-13.09	0
13	peptidase activity	2211	644.11	0.53	-13.05	0
14	G-protein coupled receptor ...	1473	465.64	0.40	-12.30	0
15	carboxylic ester hydrolase ...	628	482.21	0.37	-12.01	0
16	serine-type peptidase activity	708	614.43	0.46	-10.79	0
17	aminopeptidase activity	178	633.80	0.35	-10.74	0
18	transaminase activity	230	459.43	0.26	-10.46	0
19	methyltransferase activity	947	607.08	0.48	-8.77	0
20	carboxypeptidase activity	143	596.81	0.34	-8.54	0
21	metallopeptidase activity	632	663.90	0.56	-8.44	0
22	ATP binding	7585	801.90	0.69	-8.22	0
23	dioxygenase activity	419	604.98	0.48	-7.83	0
24	peroxidase activity	282	414.57	0.34	-7.11	0
25	antioxidant activity	499	344.42	0.36	-6.89	0
26	threonine-type endopeptidas...	139	247.87	0.19	-6.79	0
27	nucleotidyltransferase acti...	892	805.20	0.59	-6.07	0
28	porin activity	76	369.24	0.22	-5.22	0
29	serine-type endopeptidase i...	218	485.65	0.40	-4.99	0
30	phospholipase A2 activity	93	377.39	0.30	-4.93	0
31	phospholipase D activity	37	778.27	0.43	-4.87	0
32	prenyltransferase activity	128	369.37	0.32	-4.70	0
33	toxin activity	285	349.98	0.32	-4.59	0
34	dipeptidase activity	41	557.73	0.34	-4.23	0
35	metalloendopeptidase activity	369	711.73	0.63	-4.22	0
36	phospholipase A2 activity (...	70	365.39	0.30	-3.90	0
37	endonuclease activity	773	658.33	0.59	-3.85	0
38	antigen binding	277	345.87	0.38	-3.68	0
39	translation elongation fact...	116	464.12	0.46	-3.35	0
40	nuclease activity	1124	659.17	0.61	-3.20	0
41	G-protein coupled photorece...	44	390.23	0.30	-2.99	0
42	chloride channel activity	195	622.49	0.62	-2.67	0.01
43	RNA-directed 5'-3' RNA poly...	75	2341.31	0.87	-2.57	0.01
44	tRNA binding	247	566.35	0.57	-2.39	0.01
45	peptidase inhibitor activity	446	493.46	0.50	-2.38	0.01
46	photoreceptor activity	91	574.22	0.54	-2.30	0.01
47	glucan endo-1,3-beta-D-gluc...	28	522.18	0.43	-2.29	0.02
48	signal transducer activity	649	619.06	0.62	-2.17	0.02
49	aminoacyl-tRNA ligase activity	255	659.40	0.67	-1.93	0.04

20 statistički najznačajnijih neuređenih ključnih MF-reči (Xie et. al. 2007)

#	name	n prot	n families	avg_len	Z-score	P-value
0	Ribonucleoprotein	12236	412	150.55	22.13	1
1	Ribosomal protein	11692	330	140.58	20.63	1
2	Developmental protein	3260	721	149.87	19.28	1
3	Hormone	1187	161	141.13	15.58	1
4	Growth factor	785	84	255.70	11.16	1
5	Cytokine	899	110	213.28	10.21	1
6	Neuropeptide	268	209	95.08	9.65	1
7	Activator	3086	573	428.47	9.04	1
8	GAP protein	47	2	232.96	7.42	1
9	Antigen	1113	455	437.48	6.99	1
10	Repressor	2309	449	374.46	6.92	1
11	Chromatin regulator	334	100	801.24	6.70	1
12	Pyrogen	37	2	262.59	6.44	1
13	Vasoactive	125	39	160.39	5.56	1
14	Amphibian defense peptide	123	148	50.64	5.44	1
15	GTPase activation	311	70	831.03	5.36	1
16	Endorphin	42	4	226.68	5.35	1
17	Opioid peptide	24	4	216.96	5.14	1
18	Protein phosphatase inhibitor	47	8	366.51	5.07	1
19	Cyclin	182	25	430.58	4.88	1

Neuređene ključne MF-reči iz ovog rada (z-skor > 0.2)

#	name	n	avg_len	avg_dis	z	p
0	DNA-binding	6518	546.53	0.87	46.90	1
1	Developmental protein	3897	655.21	0.86	31.10	1
2	Activator	2574	600.51	0.88	28.12	1
3	Repressor	1988	589.29	0.85	22.63	1
4	RNA-binding	2728	575.76	0.76	16.62	1
5	Chromatin regulator	1038	847.06	0.90	13.91	1
6	Ribonucleoprotein	1886	272.29	0.60	13.39	1
7	Serine/threonine-protein ki...	1782	802.24	0.84	11.56	1
8	Chaperone	937	430.43	0.71	10.02	1
9	Ribosomal protein	1408	186.38	0.53	9.34	1
10	Growth factor	398	299.63	0.70	8.98	1
11	Protein kinase inhibitor	49	337.20	0.96	8.34	1
12	Calmodulin-binding	520	1229.00	0.90	7.57	1
13	Hormone	338	221.13	0.59	7.24	1
14	Cyclin	334	422.51	0.87	7.18	1
15	Signal transduction inhibitor	115	408.43	0.84	6.76	1
16	Guanine-nucleotide releasin...	319	1144.39	0.96	6.40	1
17	GTPase activation	424	867.35	0.88	6.28	1
18	Growth factor binding	50	593.98	1	6.09	1
19	Neuropeptide	105	234.96	0.68	5.99	1
20	Potassium channel	191	621.52	0.85	5.10	1
21	Calcium channel	193	1397.77	0.93	5.09	1
22	Protein phosphatase inhibitor	64	356.82	0.81	5.07	1
23	Tyrosine-protein kinase	376	863.00	0.89	5.04	1
24	Mitogen	137	830.00	0.74	4.80	1
25	Vasoactive	46	267.00	0.76	4.29	1
26	Heparin-binding	221	650.97	0.73	3.88	1
27	Muscle protein	193	920.01	0.73	3.81	1
28	Actin-binding	837	974.92	0.77	3.80	1
29	Amphibian defense peptide	49	85.80	0.53	3.34	1
30	Helicase	739	1086.05	0.87	3.29	1
31	Prion	22	497.05	0.91	3.06	1
32	Ion channel	1027	861.88	0.76	2.89	1
33	Voltage-gated channel	386	816.98	0.78	2.87	1
34	Viral nucleoprotein	39	1202.79	0.90	2.46	1
35	Tumor antigen	26	428.81	0.77	2.25	0.99
36	Exonuclease	239	725.99	0.75	2.19	0.99
37	Segmentation polarity protein	24	712.75	0.92	2.17	0.99

Neuređeni MF-termini dobijeni iz direktnog i izvedenog mapiranja

#	name	n	avg_len	avg_dis	z	p
0	DNA binding	8254	574.89	0.86	46.13	1
1	sequence-specific DNA binding	3836	536.95	0.92	41.77	1
2	chromatin binding	1538	713.64	0.93	23.54	1
3	RNA binding	5797	565.80	0.73	19.79	1
4	receptor binding	4109	529.68	0.68	14.04	1
5	protein serine/threonine ki...	1942	784.87	0.84	12.88	1
6	structural molecule activity	2884	558.88	0.64	10.84	1
7	structural constituent of r...	1350	182.95	0.53	9.25	1
8	calmodulin binding	710	1079.79	0.88	9.22	1
9	growth factor activity	458	312.38	0.70	8.99	1
10	protein kinase inhibitor ac...	237	401.60	0.82	8.94	1
11	hormone activity	452	218.09	0.61	8.84	1
12	cell adhesion molecule binding	1090	805.74	0.80	8.60	1
13	GTPase activator activity	633	820.60	0.87	8.15	1
14	actin binding	1228	1000.47	0.80	7.73	1
15	cyclin-dependent protein se...	142	357.65	0.77	6.58	1
16	protein tyrosine kinase act...	504	933.02	0.86	6.26	1
17	guanyl-nucleotide exchange ...	575	990.55	0.86	6.22	1
18	protein phosphatase inhibit...	89	425.83	0.83	6.19	1
19	heparin binding	403	652.76	0.76	6.01	1
20	neuropeptide hormone activity	117	158.64	0.65	5.88	1
21	microtubule motor activity	225	1394.78	0.97	5.65	1
22	enzyme inhibitor activity	1059	462.66	0.63	5.60	1
23	potassium channel activity	339	612.39	0.79	4.88	1
24	damaged DNA binding	284	680.80	0.80	4.31	1
25	growth factor binding	351	737.85	0.79	4.20	1
26	chloride channel inhibitor ...	25	827.64	0.96	3.96	1
27	calcium channel activity	327	1205.13	0.86	3.54	1
28	morphogen activity	23	331.57	0.83	3.20	1
29	helicase activity	766	1068.02	0.86	3.15	1
30	motor activity	447	1275.57	0.86	3.10	1
31	voltage-gated ion channel a...	501	767.73	0.76	3.06	1
32	rRNA activity	541	266.55	0.52	3.00	1
33	calcium ion binding	1878	859.00	0.70	2.80	1
34	RNA-directed DNA polymerase...	75	1427.97	0.97	2.72	1
35	kinase activity	3593	726.02	0.71	2.70	1
36	ATP-dependent helicase acti...	455	926.44	0.84	2.48	1
37	exonuclease activity	371	749.41	0.75	2.40	0.99
38	translation initiation fact...	268	482.02	0.65	1.99	0.97
39	acetylcholine-gated cation-...	60	504.17	0.75	1.99	0.99
40	metalloendopeptidase inhibi...	147	546.95	0.86	1.99	0.98
41	cysteine-type endopeptidase...	153	424.63	0.58	1.48	0.95
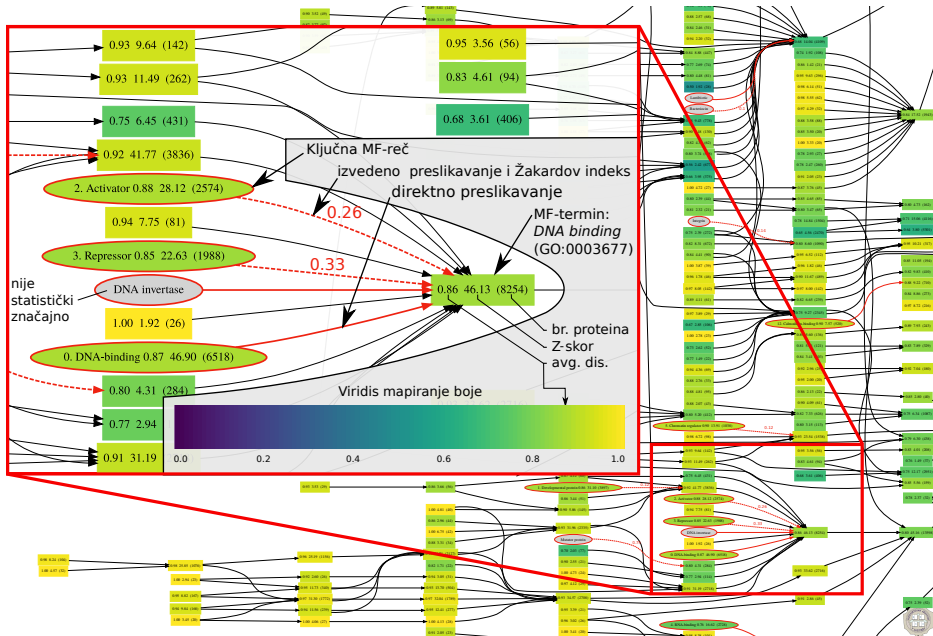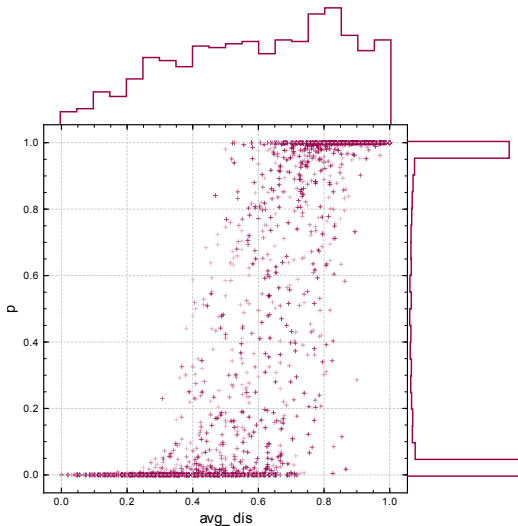
Odnos između neuređenosti (*avg_dis*) i p vrednosti MF-termina.

Hvala :)