

# Appendix A

## Motifs discovered by NestedMICA

Figures in this appendix shows the motifs discovered by NestedMICA (Chapter 2) in multiple runs on different eukaryotic localisation datasets (see 3.2.1). NestedMICA was run using the entire sequences as well as chunks of certain length from the N or C-terminal regions. Not all of these motifs, shown as sequence logos here, have been used in the development of Lokum.

The “Notes” columns in the below figures imply the sequence region (N-terminus, whole sequence, or C-terminus) a particular shown motif was discovered from. “First 20aa”, for example, indicates that the corresponding motif has been discovered within the first 20 N-terminal amino acid chunks of a particular localisation dataset.

---














#	Dataset	Motif	Notes
1	Nuclear		
2	Nuclear		
3	Nuclear		
4	Nuclear		
5	Nuclear		
6	Nuclear		
7	Nuclear		
8	Nuclear		
9	Nuclear		
10	Nuclear		
11	Nuclear		
12	Nuclear		
13	Nuclear		

Figure A.1: “Nuclear motifs” discovered by NestedMICA

---












#	Dataset	Motif	Notes
1	Plasma membrane		
2	Plasma membrane		
3	Plasma membrane		
4	Plasma membrane		
5	Plasma membrane		
6	Plasma membrane		
7	Plasma membrane		
8	Plasma membrane		
9	Plasma membrane		
10	Plasma membrane		
11	Plasma membrane		

Figure A.2: “Plasma membrane motifs” discovered by NestedMICA

---











#	Dataset	Motif	Notes
1	Cytoplasmic		
2	Cytoplasmic		
3	Cytoplasmic		
4	Cytoplasmic		
5	Cytoplasmic		
6	Cytoplasmic		
7	Cytoplasmic		
8	Cytoplasmic		
9	Cytoplasmic		
10	Cytoplasmic		

Figure A.3: “Cytoplasmic motifs” discovered by NestedMICA















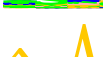








#	Dataset	Motif	Notes
1	Mitochondrial		
2	Mitochondrial		
3	Mitochondrial		
4	Mitochondrial		
5	Mitochondrial		
6	Mitochondrial		
7	Mitochondrial		
8	Mitochondrial		
9	Mitochondrial		
10	Mitochondrial		
11	Mitochondrial		
12	Mitochondrial		
13	Mitochondrial		
14	Mitochondrial		
15	Mitochondrial		
16	Mitochondrial		
17	Mitochondrial		
18	Mitochondrial		
19	Mitochondrial		
20	Mitochondrial		
21	Mitochondrial		

Figure A.4: “Mitochondrial motifs” discovered by NestedMICA

---

#	Dataset	Motif	Notes
1	ER		
2	ER		
3	ER		
4	ER		
5	ER		
6	ER		
7	ER		
8	ER		
9	ER		
10	ER		

Figure A.5: “Endoplasmic reticulum motifs” discovered by NestedMICA

---


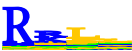


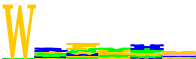










#	Dataset	Motif	Notes
1	Golgi		
2	Golgi		
3	Golgi		
4	Golgi		
5	Golgi		
6	Golgi		
7	Golgi		
8	Golgi		
9	Golgi		
10	Golgi		
11	Golgi		
12	Golgi		
13	Golgi		
14	Golgi		
15	Golgi		

Figure A.6: “Golgi motifs” discovered by NestedMICA

---









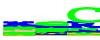


#	Dataset	Motif	Notes
1	Extracellular		
2	Extracellular		
3	Extracellular		
4	Extracellular		
5	Extracellular		
6	Extracellular		
7	Extracellular		
8	Extracellular		
9	Extracellular		
10	Extracellular		
11	Extracellular		

Figure A.7: “Extracellular motifs” discovered by NestedMICA










#	Dataset	Motif	Notes
1	Lysosome		
2	Lysosome		
4	Lysosome		
4	Lysosome		
5	Lysosome		
6	Lysosome		
7	Lysosome		
8	Lysosome		
9	Lysosome		

Figure A.8: “Lysosome motifs” discovered by NestedMICA








#	Dataset	Motif	Notes
1	Peroxisomal		
2	Peroxisomal		
4	Peroxisomal		
4	Peroxisomal		
5	Peroxisomal		
6	Peroxisomal		
7	Peroxisomal		

Figure A.9: “Peroxisomal motifs” discovered by NestedMICA

---



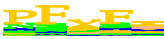


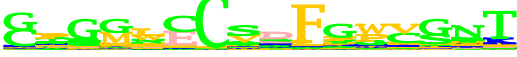





#	Dataset	Motif	Notes
1	Vacuolar		
2	Vacuolar		
4	Vacuolar		
4	Vacuolar		
5	Vacuolar		
6	Vacuolar		
7	Vacuolar		
8	Vacuolar		
9	Vacuolar		
10	Vacuolar		
11	Vacuolar		

Figure A.10: “Vacuolar motifs” discovered by NestedMICA

## Appendix B

# Amino acid composition rates in different localisations

Figures in this appendix show frequency distributions for each of the 20 amino acids in 11 eukaryotic subcellular localisation classes. The provided amino acid composition rates were obtained from the redundancy-reduced protein sequence datasets used in the training and testing of Multiloc (Höglund *et al.*, 2006), a eukaryotic localisation predictor. All the given composition values sum up to 1.0 for each localisation class.

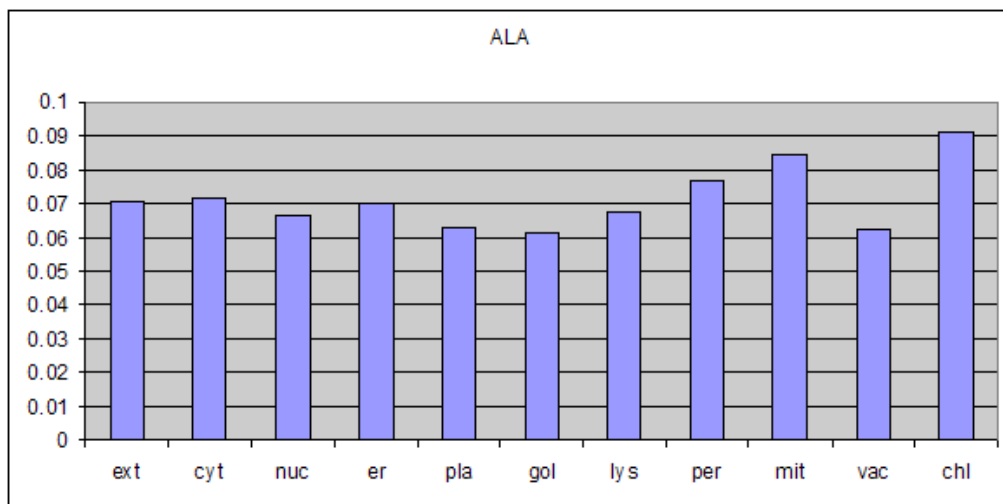


Figure B.1: Amino acid composition for Alanine (ALA / A)

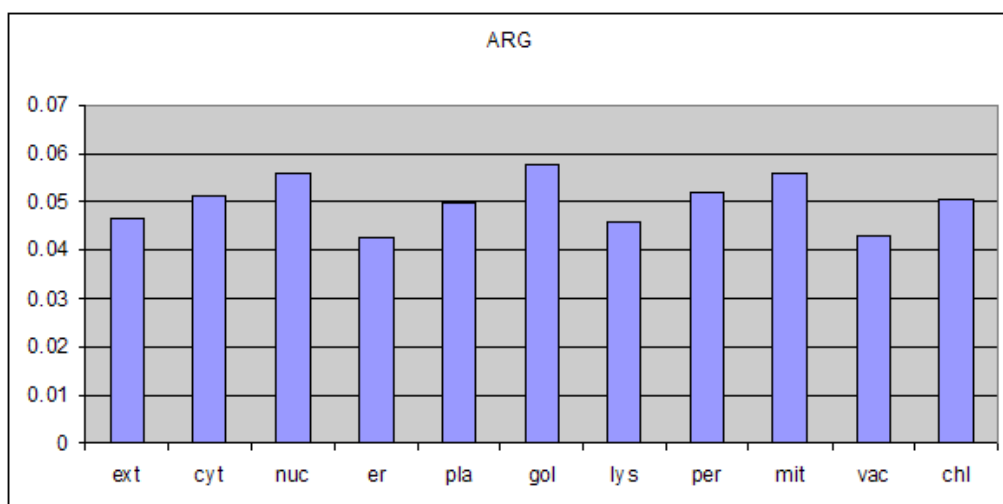


Figure B.2: Amino acid composition for Arginine (ARG / R)



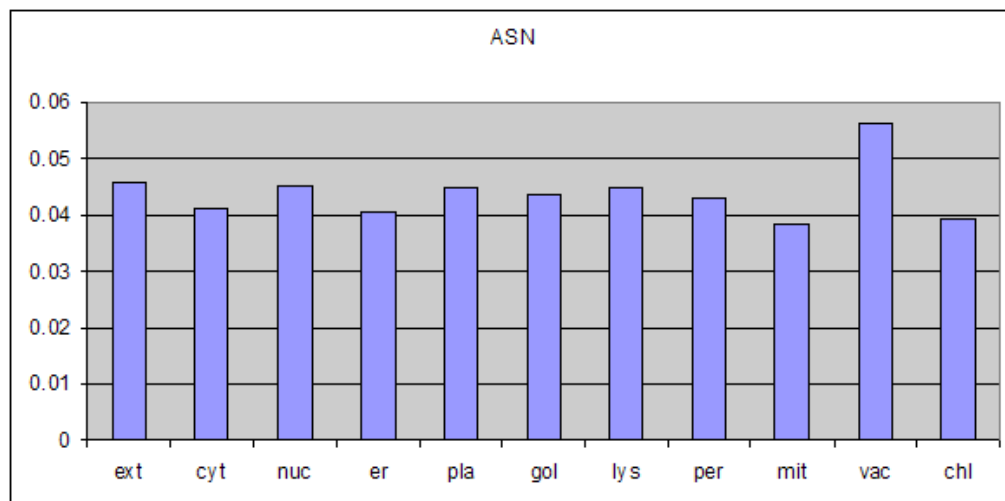


Figure B.3: Amino acid composition for Asparagine (ASN / N)

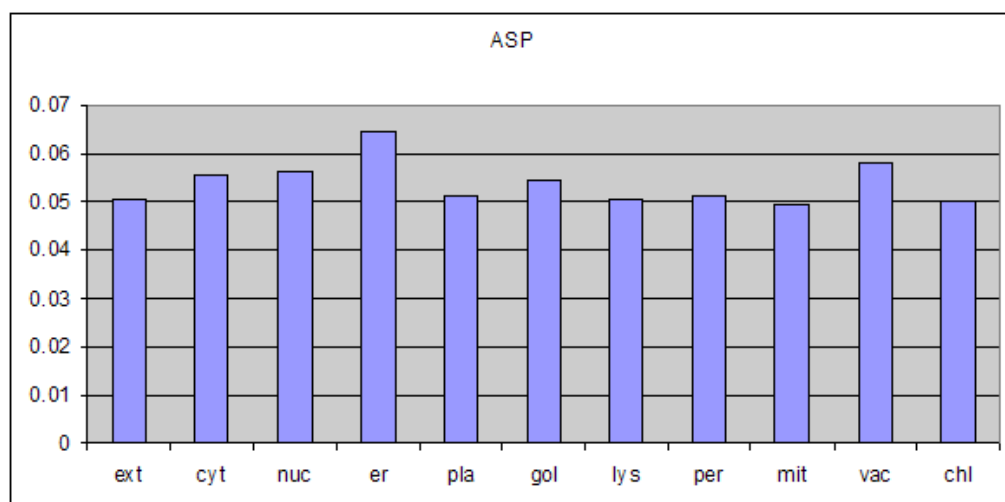


Figure B.4: Amino acid composition for Aspartic Acid (ASP / D)

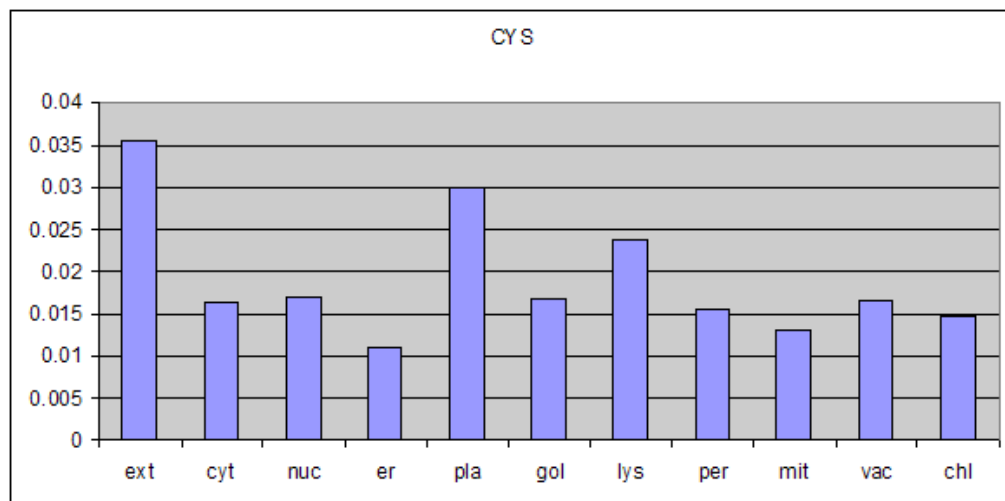


Figure B.5: Amino acid composition for Cysteine (CYS / C)

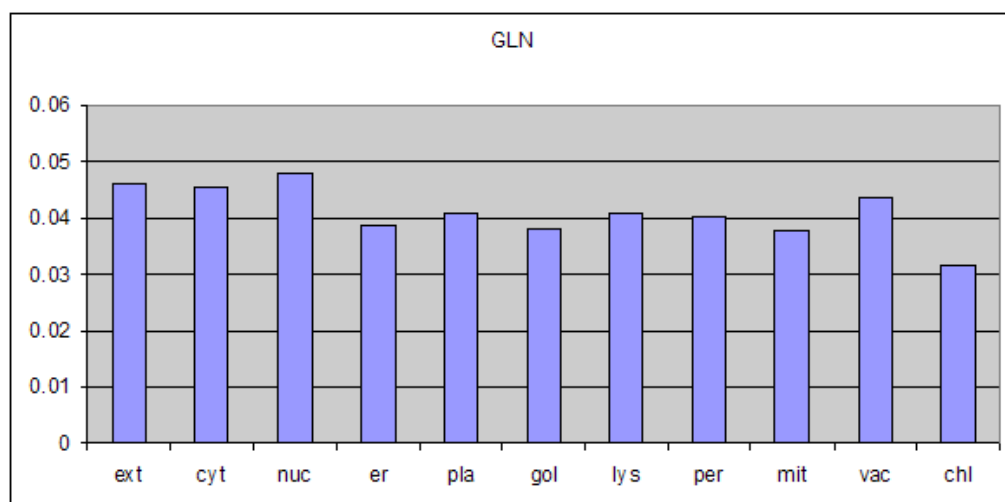


Figure B.6: Amino acid composition for Glutamine (GLN / Q)

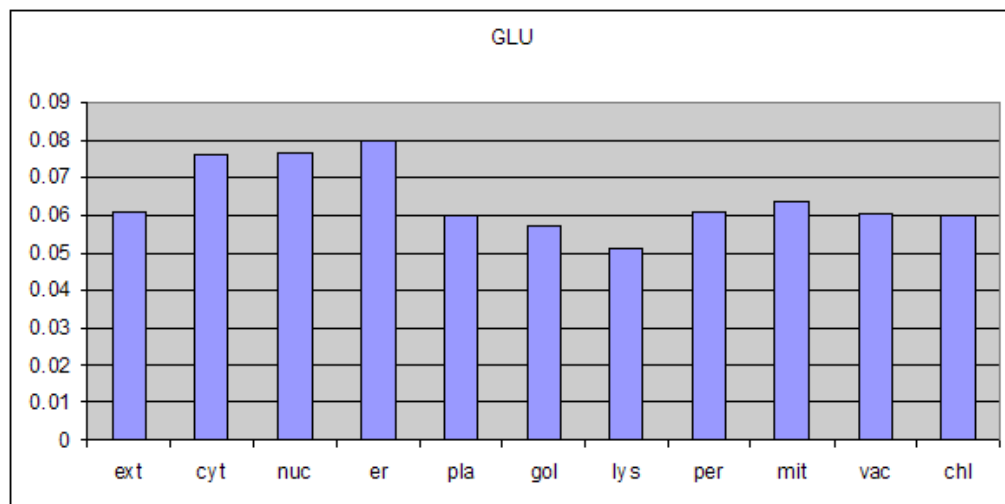


Figure B.7: Amino acid composition for Glutamic Acid (GLU / E)

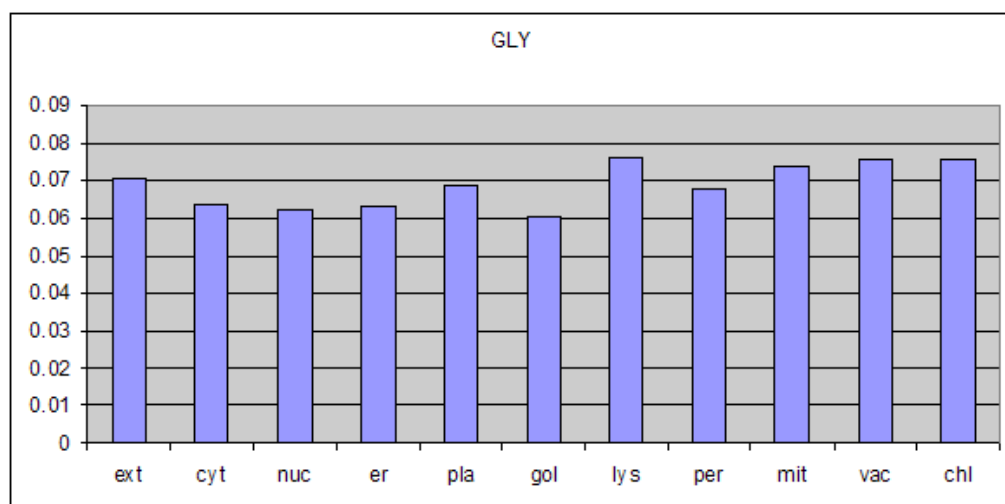


Figure B.8: Amino acid composition for Alanine (GLY / G)

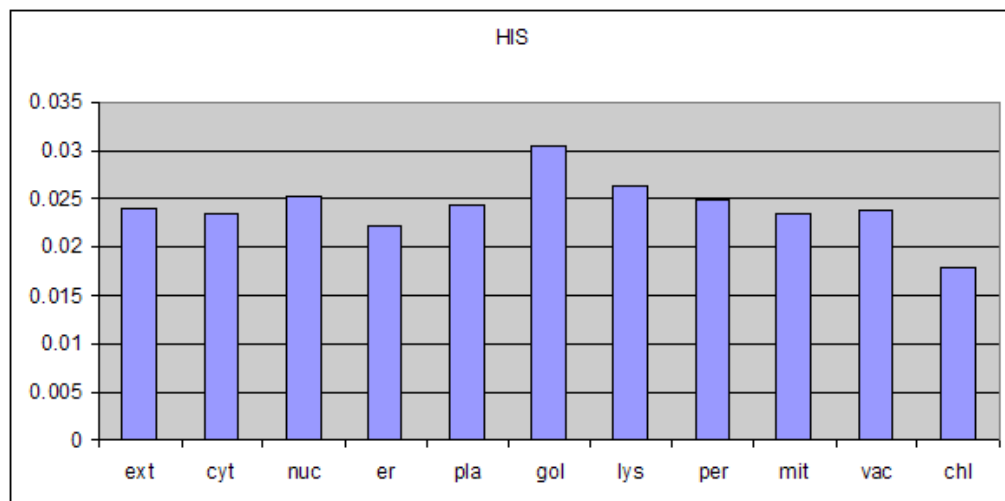


Figure B.9: Amino acid composition for Histidine (HIS / H)

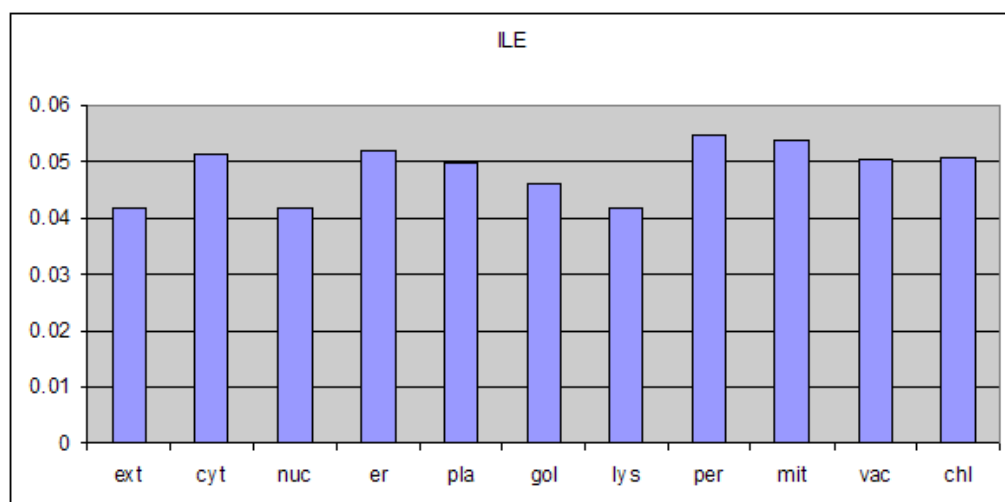


Figure B.10: Amino acid composition for Isoleucine (ILE / I)

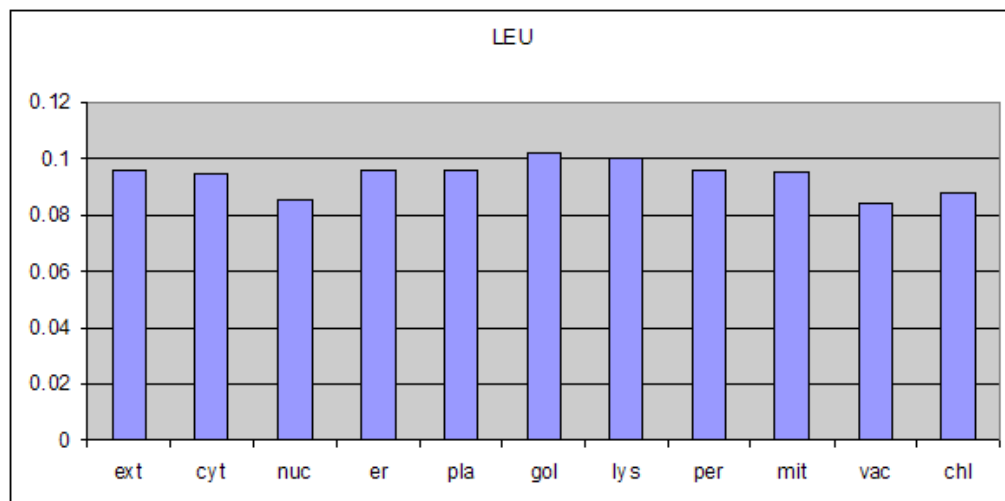


Figure B.11: Amino acid composition for Leucine (LEU / L)

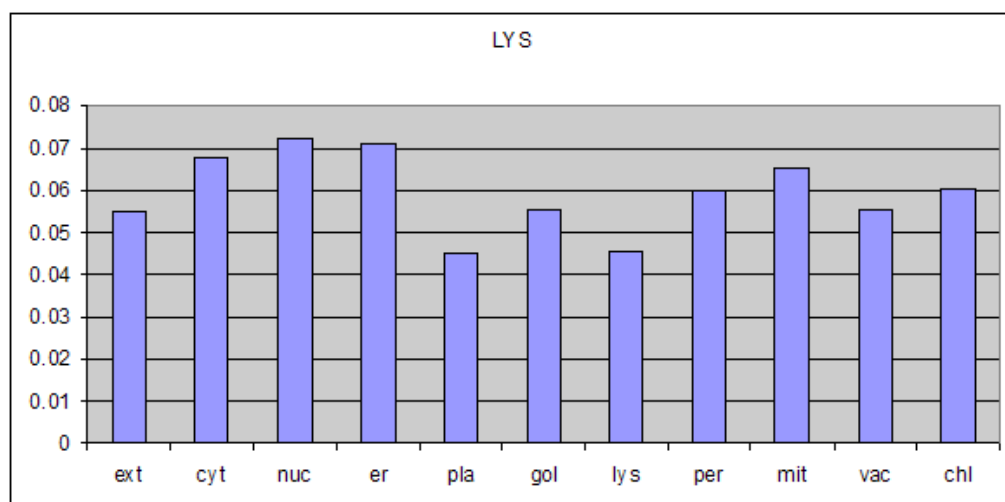


Figure B.12: Amino acid composition for Lysine (LYS / K)

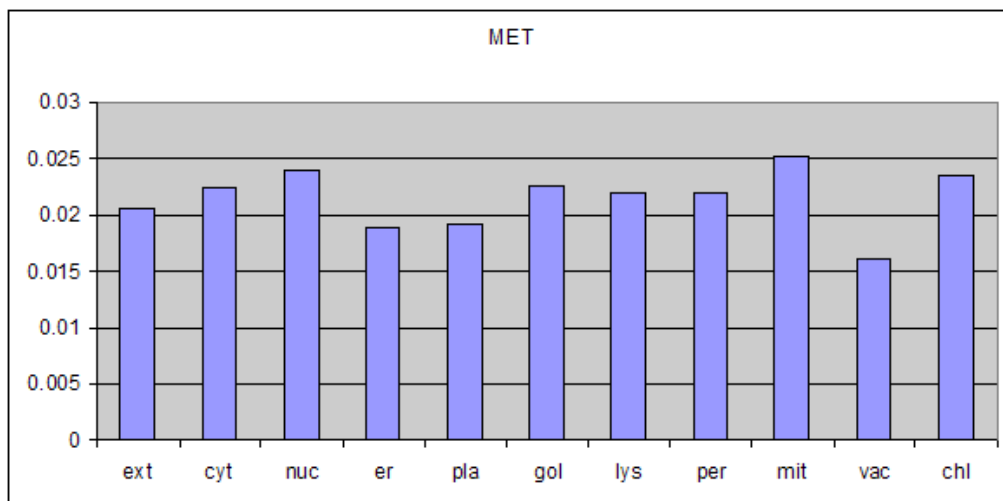


Figure B.13: Amino acid composition for Methionine (MET / M)

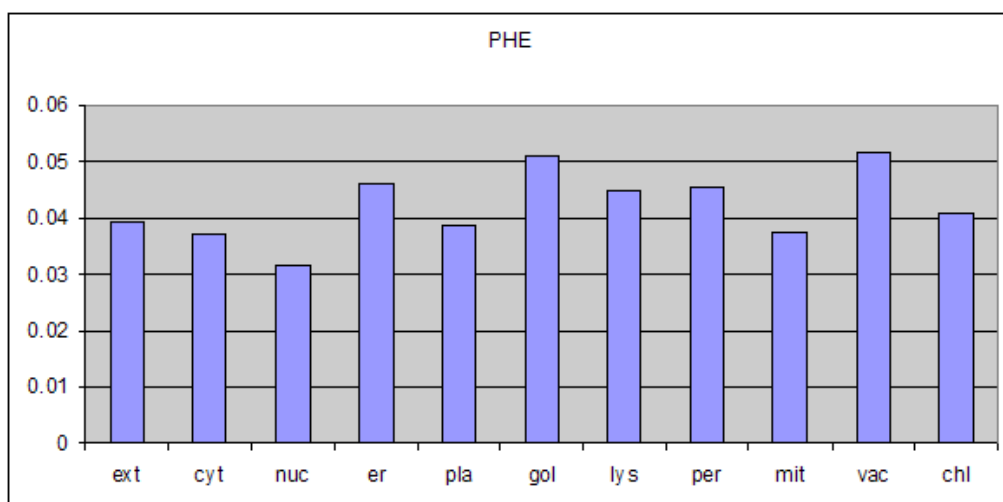


Figure B.14: Amino acid composition for Phenylalanine (PHE / F)

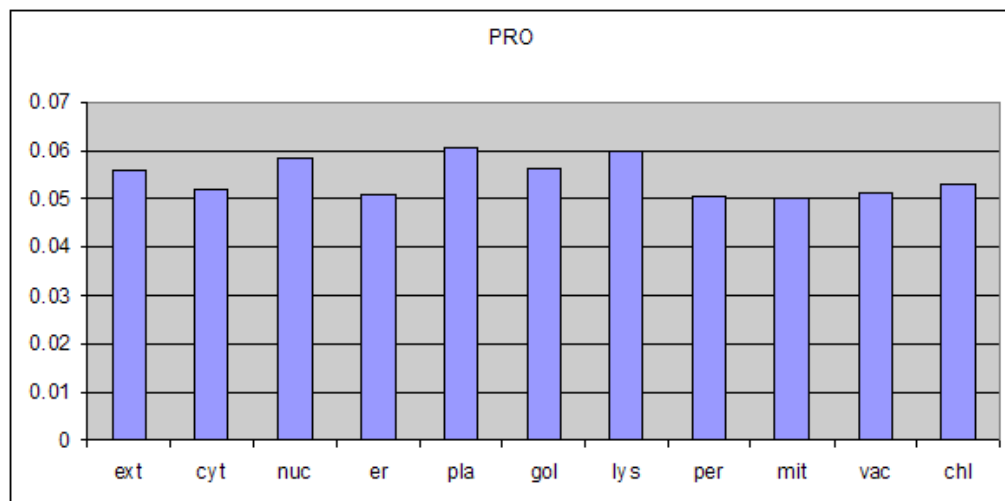


Figure B.15: Amino acid composition for Proline (PRO / P)

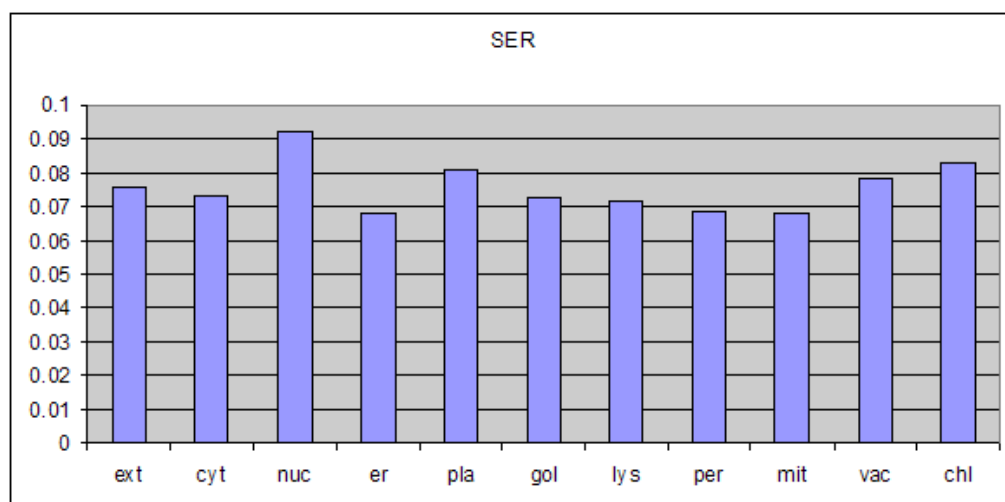


Figure B.16: Amino acid composition for Serine (SER / S)

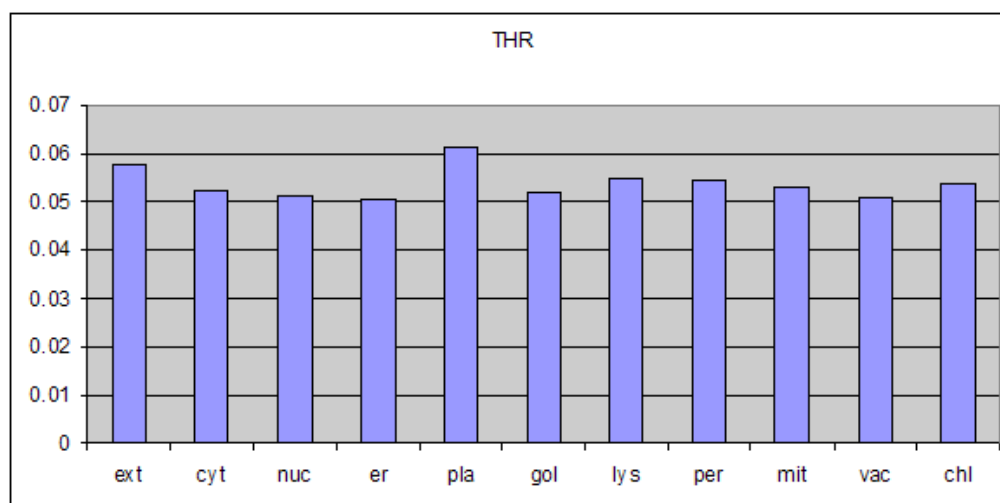


Figure B.17: Amino acid composition for Threonine (THR / T)

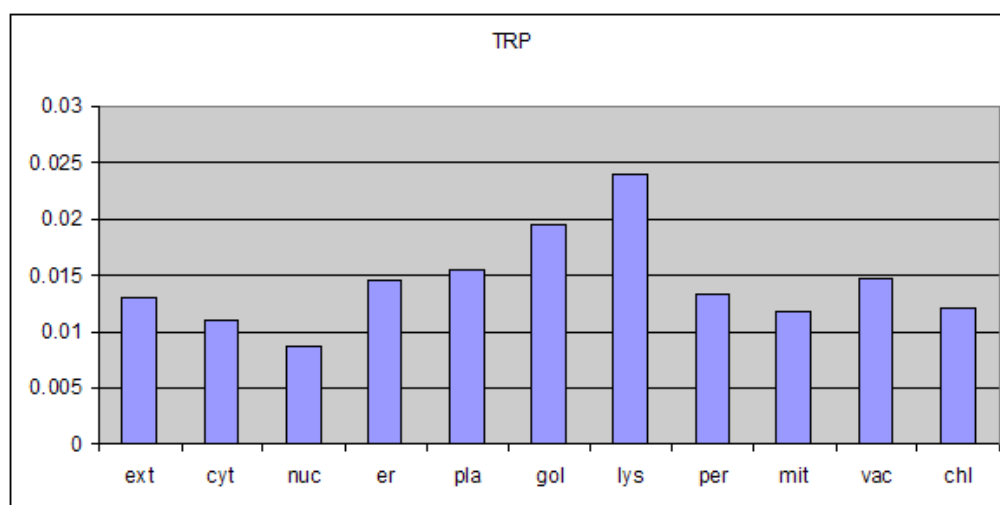


Figure B.18: Amino acid composition for Tryptophan (TRP / W)



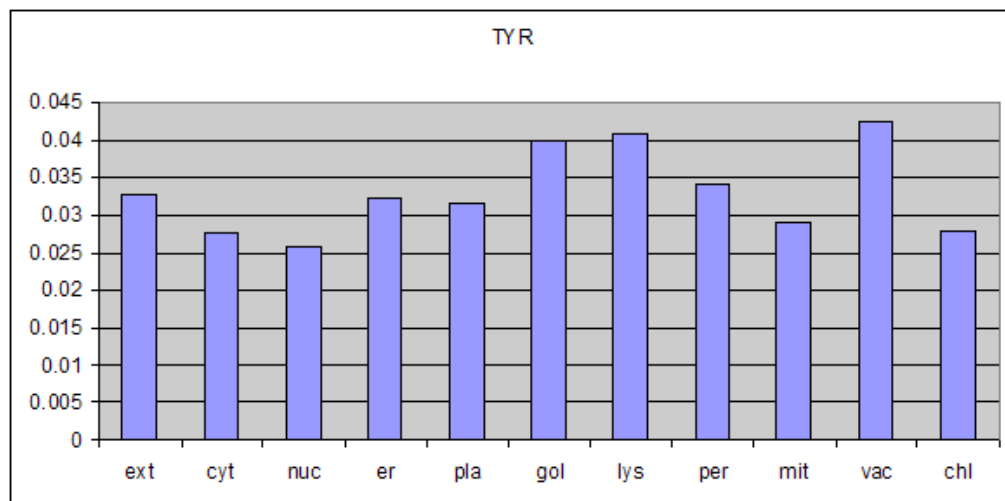


Figure B.19: Amino acid composition for Tyrosine (TYR / Y)

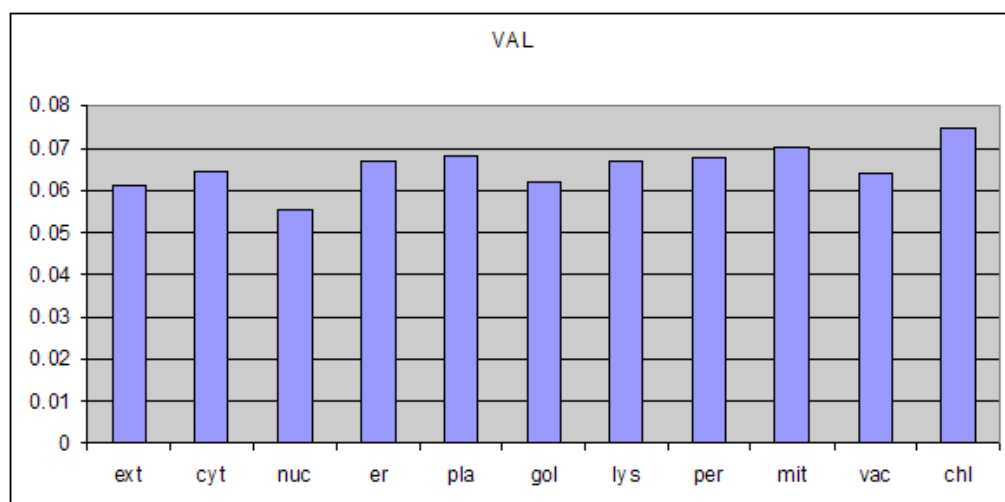


Figure B.20: Amino acid composition for Valine (VAL / V)

# Appendix C

## Sequence IDs of nuclear and nucleolar proteins filtered from the LOCATE database

### C.1 Proteins in nucleoli

5832447M01	8030477B02	AAH20037	B130024I17
0610007L03	9030008E11	AAH21402	B230113I11
0610010G24	9030015I21	AAH21438	B230341P13
0610010K23	9030404K10	AAH21497	B230345A13
0610010L07	9430023O10	AAH21646	B230384I08
0610012B16	9430042M18	AAH21922	B430205I06
0610037N12	9430068A02	AAH22656	B430304I01
0610041G09	9430070M15	AAH23108	BAA13139
0610043M01	9630032J03	AAH23495	BAA19479
0710005E17	9630058I18	AAH23755	BAA88301
1110007P10	9830141H16	AAH24049	BAA95050
1110017C15	9930013P05	AAH24718	BAB68541
1110017O22	9930036K22	AAH24730	C130053F21
1190002L16	A130086A08	AAH24881	C130060J12
1190005P17	A230078D05	AAH25074	C130087M08

## C.1 Proteins in nucleoli

---

1200003I18	A330071M14	AAH26492	C230037L02
1500005E20	A530027J07	AAH27220	C230071K24
1600021G09	A530056M01	AAH27223	C430045D17
1700010I21	A630008G24	AAH27357	C730016M05
1700020D05	A730016J17	AAH27399	C920029A19
1700026C17	A830025P17	AAH28246	CAA31278
1810029B16	AAA40067	AAH28305	CAA32372
1810063O22	AAA64248	AAH28640	CAA40012
1810073C22	AAB01504	AAH28860	CAA43091
2010206B19	AAB03664	AAH29834	CAA50196
2010300E13	AAB08894	AAH29892	CAA59260
2200007C21	AAB22970	AAH30169	CAB09797
2210401D21	AAB48630	AAH30493	CAD59182
2310002H12	AAB50013	AAH31127	CAE11688
2310039I18	AAB63526	AAH31531	D030042B21
2310040C05	AAB63915	AAH32932	D130027G07
2310057C03	AAB91426	AAH34506	D130070F09
2310057K05	AAB94491	AAH34516	D130072G11
2310061O04	AAB96870	AAH37634	D330049F08
2400004F19	AAC08435	AAH37681	D430026L04
2400011D10	AAC32982	AAH39185	D430043E23
2410041L12	AAC37664	AAH39648	D630003B12
2410089D17	AAC40061	AAH42502	D830050A13
2410115I17	AAC53171	AAH42708	E130104C03
2410130M07	AAC62511	AAH42940	E230013K19
2510038A11	AAC79683	AAH43014	E230019A18
2510039P04	AAD02877	AAH43017	E330001M23
2610204M17	AAD08676	AAH46977	E330016H10
2610507A14	AAD15718	AAH48190	E330019F09
2700027I18	AAD26855	AAH48412	E330028F04
2700052B17	AAD32094	AAH48685	E430003J02
2700066J21	AAF25951	AAH48709	E430005G16
2700067M10	AAF80246	AAH49118	E430008H02
2810004E23	AAH02004	AAH49166	E430012M21
2810012N22	AAH02014	AAH49245	E430014C08
2810017J07	AAH02025	AAH49565	E430014G22
2810026E11	AAH02027	AAH49928	E430014N21
2810037I08	AAH02044	AAH51673	E430019K12
2810453C09	AAH02079	AAH52386	E430020A18
2810473M21	AAH02108	AAH52401	E430031K14
2810486E17	AAH02306	AAH52482	E430031M22

---

## C.1 Proteins in nucleoli

2900001K19	AAH03244	AAH52790	E860029H08
3010025E17	AAH03261	AAH53333	F630017O19
3100001N19	AAH03709	AAH53404	F630021E13
3200001N24	AAH03775	AAH53453	F630048K01
4432409G09	AAH03885	AAH54085	F630105J12
4732414G15	AAH04028	AAH54541	F630222J08
4831429D18	AAH05547	AAH54723	F630223G06
4832420E07	AAH05734	AAH54778	F730043N02
4833436C12	AAH05776	AAH55393	F830044L17
4833442I16	AAH06631	AAH55484	F830213J22
4930408P03	AAH06684	AAH55787	G270004D20
4930417F03	AAH06805	AAH55860	G270124L20
4930429N24	AAH07174	AAH56232	G430020C23
4930512K19	AAH07487	AAH56383	G430046N16
4930528I04	AAH08161	AAH56650	G430074J02
4930558P17	AAH08270	AAH56992	G430138A13
4930563C04	AAH09100	AAH57033	G430146M18
4931421E07	AAH09142	AAH57054	G630007K23
4932409F19	AAH10987	AAH57156	G830049J11
4932434G09	AAH11213	AAH57342	G930019G02
4933403E10	AAH11248	AAH57645	G930027I02
4933431P07	AAH11484	AAH59089	I0C0003A04
5330437I08	AAH12276	AAH59822	I0C0030N23
5430425F10	AAH12281	AAH60072	I0C0040N18
5730405D16	AAH12433	AAH60147	I1C0027A21
5730406H19	AAH12641	AAH60375	I1C0031M12
5730419M09	AAH13165	AAH60959	I420019J01
5730436C18	AAH13618	AAH62146	I420024J09
5730470K22	AAH14688	AAH63100	I530003K08
5730563P06	AAH14703	AAH63748	I730026M06
5730589J07	AAH16194	AAH63755	I730039C23
5830405E04	AAH16489	AAH64712	I730045C06
5830465M17	AAH16569	AAK01204	I730051L04
6030446B09	AAH16676	AAK49787	I830034K17
6030461M12	AAH17637	AAK70403	I830055B02
6330401E03	AAH18321	AAL27006	I920011D05
6430407C24	AAH18373	AAL62331	I920020L14
6430528J02	AAH18399	AAL74402	I920030N07
6430603C09	AAH18545	AAO15605	I920037I11
6430628D06	AAH19218	AAO18683	I920056H18
6720463L11	AAH19418	AAR87796	I920065D03

---

6720473C09	AAH19535	B020030J01	I920089B02
7330416M24	AAH19693	B130012H23	I920089E19
			K530012F21

Table C.1: **Nucleolar proteins.** FANTOM3 IDs of nucleolar proteins which are filtered from LOCATE database (Chapter 4).

## C.2 Proteins in nuclei

0610010G04	6430529J03	AAH19520	C330049H01
0710005K22	6430549H08	AAH20099	C430003H13
1110003H09	6430598F23	AAH20990	C730024K17
1110021J02	6820408J04	AAH21306	C920026J05
1110067L22	7120441D04	AAH21750	CAA31138
1110069I04	7120476M05	AAH21839	CAA31808
1200009L24	7420438E06	AAH22600	CAA31957
1300002I11	8430431N14	AAH22628	CAA32372
1500010M05	9130009B16	AAH22681	CAA33096
1500017I02	9130019G03	AAH22733	CAA33373
1600032G08	9130211K13	AAH23110	CAA43091
1700003P16	9130217P20	AAH23324	CAA43723
1700019E19	9330101O11	AAH23775	CAA55350
1700028K03	9330177B18	AAH23815	CAA56450
1700030B17	9430072B20	AAH23915	CAA63733
1700030G05	9530046I22	AAH23961	CAA70213
1700067K01	9630025P05	AAH24341	CAA72404
1810037C20	9630027H07	AAH24521	CAA76637
1810046K07	9630029K22	AAH25073	CAB60732
2010001O09	9630045G21	AAH25602	CAB86873
2010300P09	9630050P12	AAH26841	CAC83967
2210002J07	9630054L03	AAH28871	D030011P09
2310008J22	9830147C21	AAH29834	D030020D13
2310014B11	9830160I16	AAH30915	D030053M22
2310042K15	9830169A11	AAH31168	D030054H07
2310043K02	A030005M07	AAH31463	D130019F14
2310047L21	A230011H19	AAH31769	D130064J02
2400002C23	A230039L04	AAH34855	D130084E01
2400011D10	A230054A08	AAH35298	D230016D14

## C.2 Proteins in nuclei

---

2410003J06	A230057M07	AAH36287	D230040E22
2410012M07	A230078I01	AAH37187	D330006L06
2410046L22	A230084K08	AAH37695	D430003F23
2410089D17	A230106F01	AAH40370	D530036A19
2410141M05	A330080J22	AAH43086	D630039M16
2510005J23	A430043P11	AAH46286	D830005I23
2510049I19	A530086E13	AAH47152	D930007K17
2610021E10	A730013C09	AAH48503	D930030M14
2700099C19	A730020L03	AAH48779	D930033J18
2810004A21	A730063C17	AAH50803	E030029E20
2810021G24	A730096N15	AAH51049	E030029N02
2810039M17	A830038H15	AAH51261	E130012E07
2810417H13	A830097I09	AAH51631	E130013F06
2810457D07	A930007L12	AAH51967	E130118K14
2900074D10	A930041F19	AAH52030	E130303A03
3000002C10	A930104E21	AAH52173	E230001H19
3100002L24	AAA20039	AAH52468	E330032L15
3110007F17	AAA37184	AAH52672	E330034H06
3110030B08	AAA37291	AAH52856	E430001M22
3200002N09	AAA97500	AAH53409	E430007C11
3732413B21	AAA98977	AAH54456	E430007F09
4432404N24	AAB24330	AAH54768	E430014D12
4631408J24	AAB40892	AAH56922	E970008A17
4632401G08	AAB41327	AAH57096	F420011N06
4632404G05	AAB65839	AAH57165	F530014L05
4632415C14	AAB70094	AAH57205	F630011I01
4632417G13	AAB81245	AAH57453	F630205L24
4732403I07	AAC02226	AAH58103	F730014I05
4732424P06	AAC36358	AAH60072	F730216I21
4732458H05	AAC40148	AAH60234	F830002J06
4732467A04	AAC52994	AAH60613	F830007J16
4833406K08	AAD00238	AAH61493	F830017M05
4833413D08	AAD13139	AAH64018	F830022C10
4921510H08	AAD39396	AAH64757	F830108M10
4921531G14	AAF27311	AAH65165	F830211J08
4922501K05	AAF27551	AAK07621	G430032E03
4930433I11	AAF63757	AAK35053	G430037K05
4930538K15	AAF72874	AAK39099	G430090A08
4931400M17	AAF86375	AAK39438	G530118I12
4931400O07	AAG01633	AAK60496	G630048M14
4931408L03	AAG29950	AAL09305	G730014O11

---

4931409I21	AAG34081	AAL40860	G730050K01
4932416N14	AAG34793	AAL47577	G830045E23
4932441K08	AAG40809	AAL67834	I0C0040N18
4933400A06	AAG50171	AAL69526	I0C0048H21
4933415E13	AAH03259	AAL71902	I0C0048J01
4933439J20	AAH03266	AAM33069	I0C0048L09
5330404L13	AAH03292	AAM64199	I1C0020M01
5330418E10	AAH03330	AAM77216	I1C0033H16
5430425F10	AAH04738	B020012J09	I420001D20
5430434J22	AAH05426	B130019L12	I420006B09
5530401L07	AAH05516	B230111C05	I420014I19
5730407F12	AAH05620	B230120H23	I420025C06
5730438N18	AAH05694	B230213G02	I420033H08
5730548J20	AAH05744	B230309O17	I530008I17
5730592N24	AAH06016	B230375D17	I530014J18
5930431H10	AAH06939	BAA05885	I530027I10
6330405E07	AAH09004	BAA21725	I530028C19
6330414C15	AAH10496	BAA23648	I830025I02
6330417C18	AAH10841	BAA95075	I830031G17
6330437A14	AAH11091	BAA96361	I830037L15
6330503C03	AAH11131	BAB79232	I830043E20
6330513G01	AAH12715	BAC53845	I830128O08
6330541F16	AAH12953	BAC75669	I920021I20
6330562H21	AAH13718	C130032B15	I920062F22
6430402E12	AAH14828	C130083B17	I920087J04
6430519P13	AAH19168	C230004D03	K230011N15
K230305H19	K230320F07		

Table C.2: **Nuclear proteins.** FANTOM3 IDs of nuclear proteins which are filtered from LOCATE database (Chapter 4).

## Appendix D

Kullback-Leibler divergence for  
transmembrane helix cap  
positions in terms of amino acid  
composition



---

Amino acid	1 - 1	1 - 2	1 - 3	1 - 4	1 - 5
GLY	-0.016	0.003	-0.031	-0.017	-0.028
PHE	0.008	-0.051	-0.047	-0.049	-0.039
SER	-0.024	0.034	0.001	-0.004	0.009
PRO	-0.019	-0.014	-0.008	0.001	0.012
TYR	-0.006	-0.020	-0.017	-0.017	-0.019
ARG	0.163	0.505	0.544	0.575	0.531
TRP	-0.011	-0.020	-0.023	-0.021	-0.020
ALA	-0.006	-0.053	-0.050	-0.041	-0.057
LYS	0.155	0.530	0.636	0.547	0.610
ASN	0.004	0.061	0.068	0.071	0.103
GLN	-0.019	0.024	0.022	0.032	0.014
CYS	0.008	0.034	0.016	0.014	0.012
MET	-0.011	-0.018	-0.021	-0.023	-0.018
THR	-0.014	0.008	0.008	0.000	0.017
VAL	-0.002	-0.052	-0.044	-0.056	-0.058
ILE	-0.004	-0.049	-0.044	-0.050	-0.052
ASP	-0.017	0.108	0.130	0.141	0.167
LEU	-0.013	-0.085	-0.082	-0.077	-0.086
HIS	-0.002	0.025	0.029	0.019	0.027
GLU	-0.034	0.062	0.062	0.076	0.121
TOTAL KL	0.137	1.032	1.148	1.122	1.245

Table D.1: **KL deviations of cytoplasmic side helix positions from the non-cytoplasmic side helix cap positions in relative amino acid abundance rates.** The “1-1” column lists the KL distances for the first position of the cytoplasmic side cap and the first position of the non-cytoplasmic side, and so on.

---

Amino acid	2 - 1	2 - 2	2 - 3	2 - 4	2 - 5
GLY	-0.027	-0.012	-0.041	-0.028	-0.038
PHE	0.086	-0.027	-0.021	-0.025	-0.005
SER	-0.031	0.021	-0.009	-0.012	-0.001
PRO	-0.022	-0.017	-0.012	-0.003	0.007
TYR	0.030	0.008	0.013	0.013	0.009
ARG	-0.036	0.013	0.019	0.023	0.017
TRP	0.021	-0.006	-0.014	-0.007	-0.005
ALA	0.095	-0.003	0.004	0.022	-0.012
LYS	-0.031	0.013	0.025	0.015	0.022
ASN	-0.028	-0.008	-0.005	-0.004	0.007
GLN	-0.022	-0.009	-0.009	-0.006	-0.012
CYS	-0.006	0.006	-0.002	-0.003	-0.004
MET	0.054	0.029	0.021	0.014	0.029
THR	-0.021	-0.001	-0.001	-0.009	0.007
VAL	0.134	0.008	0.028	-0.002	-0.007
ILE	0.176	0.003	0.022	-0.003	-0.009
ASP	-0.028	-0.008	-0.004	-0.002	0.002
LEU	0.295	0.039	0.049	0.069	0.035
HIS	-0.016	-0.007	-0.006	-0.009	-0.007
GLU	-0.030	-0.005	-0.005	-0.001	0.011
TOTAL KL	0.594	0.037	0.052	0.040	0.045

Table D.2: **KL deviations of cytoplasmic side helix positions from the non-cytoplasmic side helix cap positions in relative amino acid abundance rates.** The “2-1” column lists the KL distances for the second position of the cytoplasmic side cap and the first position of the non-cytoplasmic side, and so on.

---

Amino acid	3 - 1	3 - 2	3 - 3	3 - 4	3 - 5
GLY	-0.021	-0.004	-0.036	-0.022	-0.033
PHE	0.098	-0.022	-0.016	-0.020	0.002
SER	-0.036	0.008	-0.018	-0.021	-0.011
PRO	-0.024	-0.019	-0.015	-0.007	0.002
TYR	-0.003	-0.018	-0.015	-0.015	-0.017
ARG	-0.036	0.010	0.016	0.020	0.014
TRP	0.033	0.001	-0.007	0.000	0.002
ALA	0.145	0.028	0.036	0.057	0.016
LYS	-0.032	0.015	0.028	0.017	0.024
ASN	-0.028	-0.010	-0.008	-0.007	0.003
GLN	-0.025	-0.003	-0.004	0.001	-0.008
CYS	-0.004	0.011	0.001	0.000	-0.001
MET	0.036	0.014	0.008	0.002	0.014
THR	-0.022	-0.002	-0.002	-0.009	0.005
VAL	0.114	-0.003	0.016	-0.012	-0.017
ILE	0.179	0.004	0.023	-0.002	-0.008
ASP	-0.028	-0.008	-0.004	-0.002	0.002
LEU	0.277	0.029	0.039	0.058	0.025
HIS	-0.016	-0.002	0.000	-0.005	-0.001
GLU	-0.035	0.002	0.002	0.008	0.025
TOTAL KL	0.572	0.031	0.044	0.040	0.040

Table D.3: **KL deviations of cytoplasmic side helix positions from the non-cytoplasmic side helix cap positions in relative amino acid abundance rates.** The “3-1” column lists the KL distances for the third position of the cytoplasmic side cap and the first position of the non-cytoplasmic side, and so on.

---

Amino acid	4 - 1	4 - 2	4 - 3	4 - 4	4 - 5
GLY	-0.011	0.009	-0.027	-0.012	-0.024
PHE	0.109	-0.018	-0.011	-0.015	0.007
SER	-0.034	0.014	-0.013	-0.017	-0.007
PRO	-0.017	-0.011	-0.005	0.004	0.016
TYR	0.000	-0.016	-0.012	-0.012	-0.015
ARG	-0.036	0.010	0.016	0.020	0.014
TRP	0.006	-0.015	-0.021	-0.016	-0.015
ALA	0.095	-0.003	0.004	0.022	-0.012
LYS	-0.032	0.018	0.032	0.020	0.029
ASN	-0.028	-0.010	-0.008	-0.007	0.003
GLN	-0.025	-0.003	-0.004	0.001	-0.008
CYS	0.006	0.031	0.015	0.013	0.011
MET	0.039	0.017	0.010	0.004	0.017
THR	-0.019	0.002	0.002	-0.006	0.010
VAL	0.116	-0.002	0.017	-0.011	-0.016
ILE	0.227	0.027	0.049	0.020	0.013
ASP	-0.030	-0.006	-0.002	0.000	0.005
LEU	0.224	0.001	0.010	0.027	-0.002
HIS	-0.015	-0.001	0.001	-0.004	0.000
GLU	-0.030	-0.005	-0.005	-0.001	0.011
TOTAL KL	0.545	0.040	0.047	0.030	0.038

Table D.4: **KL deviations of cytoplasmic side helix positions from the non-cytoplasmic side helix cap positions in relative amino acid abundance rates.** The “4-1” column lists the KL distances for the fourth position of the cytoplasmic side cap and the first position of the non-cytoplasmic side, and so on.

---

Amino acid	5 - 1	5 - 2	5 - 3	5 - 4	5 - 5
GLY	-0.007	0.014	-0.024	-0.008	-0.021
PHE	0.141	-0.003	0.005	0.000	0.026
SER	-0.014	0.054	0.015	0.010	0.024
PRO	-0.026	-0.022	-0.018	-0.011	-0.003
TYR	0.018	-0.001	0.003	0.003	0.000
ARG	-0.034	0.003	0.008	0.011	0.006
TRP	0.012	-0.012	-0.018	-0.013	-0.011
ALA	0.079	-0.012	-0.006	0.011	-0.021
LYS	-0.030	0.009	0.021	0.011	0.018
ASN	-0.028	-0.009	-0.006	-0.005	0.006
GLN	-0.025	0.001	0.000	0.006	-0.005
CYS	0.009	0.036	0.018	0.016	0.014
MET	0.027	0.008	0.002	-0.003	0.008
THR	-0.017	0.004	0.004	-0.004	0.013
VAL	0.132	0.007	0.027	-0.003	-0.008
ILE	0.230	0.028	0.050	0.022	0.014
ASP	-0.008	-0.005	-0.005	-0.004	-0.004
LEU	0.180	-0.021	-0.013	0.002	-0.024
HIS	-0.016	-0.007	-0.006	-0.009	-0.007
GLU	-0.030	-0.005	-0.005	-0.001	0.011
TOTAL KL	0.595	0.068	0.052	0.031	0.037

Table D.5: **KL deviations of cytoplasmic side helix positions from the non-cytoplasmic side helix cap positions in relative amino acid abundance rates.** The “5-1” column lists the KL distances for the fifth position of the cytoplasmic side cap and the first position of the non-cytoplasmic side, and so on.

---

Amino acid	1 - 2	1 - 3	1 - 4	1 - 5
GLY	0.017	0.007	-0.005	-0.009
PHE	-0.036	-0.040	-0.042	-0.049
SER	0.010	0.024	0.017	-0.012
PRO	0.005	0.010	-0.003	0.017
TYR	-0.025	-0.003	-0.006	-0.019
SEC	0.000	0.000	0.000	0.000
ARG	0.412	0.428	0.428	0.474
TRP	-0.018	-0.021	-0.013	-0.016
ALA	-0.051	-0.064	-0.051	-0.046
LYS	0.421	0.412	0.395	0.441
ASN	0.083	0.093	0.093	0.086
GLN	0.052	0.030	0.030	0.022
CYS	0.021	0.014	0.001	-0.001
MET	-0.027	-0.023	-0.024	-0.021
THR	0.009	0.011	0.006	0.003
VAL	-0.055	-0.051	-0.051	-0.055
ILE	-0.049	-0.050	-0.055	-0.055
ASP	0.153	0.153	0.141	0.314
LEU	-0.096	-0.094	-0.086	-0.078
HIS	0.048	0.029	0.027	0.048
GLU	0.080	0.056	0.080	0.080
TOTAL KL	0.953	0.921	0.880	1.122

Table D.6: **KL deviations of cytoplasmic side helix cap positions.** The “1-2” column lists the KL distances for the first position of the cytoplasmic side cap and the second position of the cytoplasmic side, and so on.

---

Amino acid	2 - 3	2 - 4	2 - 5	3 - 4	3 - 5	4 - 5
GLY	-0.008	-0.019	-0.022	-0.011	-0.015	-0.004
PHE	-0.006	-0.011	-0.025	-0.005	-0.020	-0.015
SER	0.012	0.006	-0.020	-0.005	-0.027	-0.024
PRO	0.005	-0.007	0.011	-0.011	0.006	0.021
TYR	0.035	0.030	0.009	-0.003	-0.017	-0.014
SEC	0.000	0.000	0.000	0.000	0.000	0.000
ARG	0.002	0.002	0.009	0.000	0.006	0.006
TRP	-0.007	0.014	0.007	0.024	0.016	-0.005
ALA	-0.026	0.000	0.010	0.031	0.043	0.010
LYS	-0.001	-0.003	0.002	-0.002	0.004	0.006
ASN	0.003	0.003	0.001	0.000	-0.002	-0.002
GLN	-0.007	-0.007	-0.009	0.000	-0.004	-0.004
CYS	-0.003	-0.009	-0.010	-0.008	-0.009	-0.002
MET	0.012	0.009	0.019	-0.002	0.006	0.008
THR	0.001	-0.003	-0.005	-0.004	-0.006	-0.002
VAL	0.011	0.010	0.001	-0.001	-0.010	-0.009
ILE	-0.001	-0.020	-0.021	-0.019	-0.020	-0.001
ASP	0.000	-0.002	0.027	-0.002	0.027	0.033
LEU	0.009	0.038	0.066	0.028	0.055	0.024
HIS	-0.006	-0.007	0.000	-0.001	0.010	0.011
GLU	-0.006	0.000	0.000	0.009	0.009	0.000
TOTAL KL	0.019	0.025	0.049	0.017	0.051	0.037

Table D.7: **KL deviations of cytoplasmic side helix cap positions.** The “2-3” column lists the KL distances for the second position of the cytoplasmic side cap and the third position of the cytoplasmic side, and so on.

---

Amino acid	1 - 2	1 - 3	1 - 4	1 - 5
GLY	0.022	-0.018	-0.001	-0.015
PHE	-0.051	-0.048	-0.049	-0.040
SER	0.078	0.033	0.027	0.044
PRO	0.008	0.016	0.029	0.046
TYR	-0.015	-0.012	-0.012	-0.015
SEC	0.000	0.000	0.000	0.000
ARG	0.160	0.178	0.192	0.171
TRP	-0.018	-0.023	-0.019	-0.017
ALA	-0.051	-0.048	-0.038	-0.056
LYS	0.171	0.219	0.178	0.207
ASN	0.054	0.061	0.063	0.094
GLN	0.065	0.062	0.077	0.050
CYS	0.020	0.007	0.005	0.004
MET	-0.012	-0.016	-0.019	-0.012
THR	0.027	0.027	0.017	0.038
VAL	-0.052	-0.043	-0.056	-0.058
ILE	-0.050	-0.044	-0.052	-0.053
ASP	0.153	0.180	0.193	0.226
LEU	-0.086	-0.082	-0.076	-0.087
HIS	0.028	0.033	0.023	0.031
GLU	0.172	0.172	0.197	0.277
TOTAL KL	0.622	0.652	0.681	0.834

Table D.8: **KL deviations of non-cytoplasmic side helix cap positions.** The “1-2” column lists the KL distances for the first position of the non-cytoplasmic side cap and the second position of the non-cytoplasmic side, and so on.



---

Amino acid	2 - 3	2 - 4	2 - 5	3 - 4	3 - 5	4 - 5
GLY	-0.033	-0.019	-0.030	0.020	0.004	-0.014
PHE	0.008	0.003	0.029	-0.004	0.020	0.025
SER	-0.023	-0.025	-0.017	-0.004	0.008	0.013
PRO	0.007	0.019	0.034	0.011	0.024	0.011
TYR	0.005	0.005	0.001	0.000	-0.003	-0.003
SEC	0.000	0.000	0.000	0.000	0.000	0.000
ARG	0.004	0.006	0.002	0.002	-0.001	-0.003
TRP	-0.008	-0.001	0.001	0.008	0.011	0.002
ALA	0.007	0.025	-0.010	0.018	-0.016	-0.029
LYS	0.007	0.001	0.005	-0.003	-0.001	0.004
ASN	0.003	0.005	0.020	0.001	0.015	0.013
GLN	-0.001	0.005	-0.006	0.006	-0.005	-0.009
CYS	-0.005	-0.005	-0.006	-0.001	-0.002	-0.001
MET	-0.005	-0.010	0.000	-0.005	0.006	0.012
THR	0.000	-0.007	0.008	-0.007	0.008	0.017
VAL	0.019	-0.010	-0.015	-0.025	-0.029	-0.005
ILE	0.019	-0.005	-0.012	-0.021	-0.027	-0.006
ASP	0.006	0.009	0.017	0.002	0.008	0.005
LEU	0.009	0.026	-0.003	0.016	-0.012	-0.026
HIS	0.002	-0.003	0.001	-0.005	-0.001	0.005
GLU	0.000	0.005	0.021	0.005	0.021	0.013
TOTAL KL	0.021	0.023	0.042	0.013	0.029	0.024

Table D.9: **KL deviations of non-cytoplasmic side helix cap positions.** The “2-3” column lists the KL distances for the second position of the non-cytoplasmic side cap and the third position of the non-cytoplasmic side, and so on.