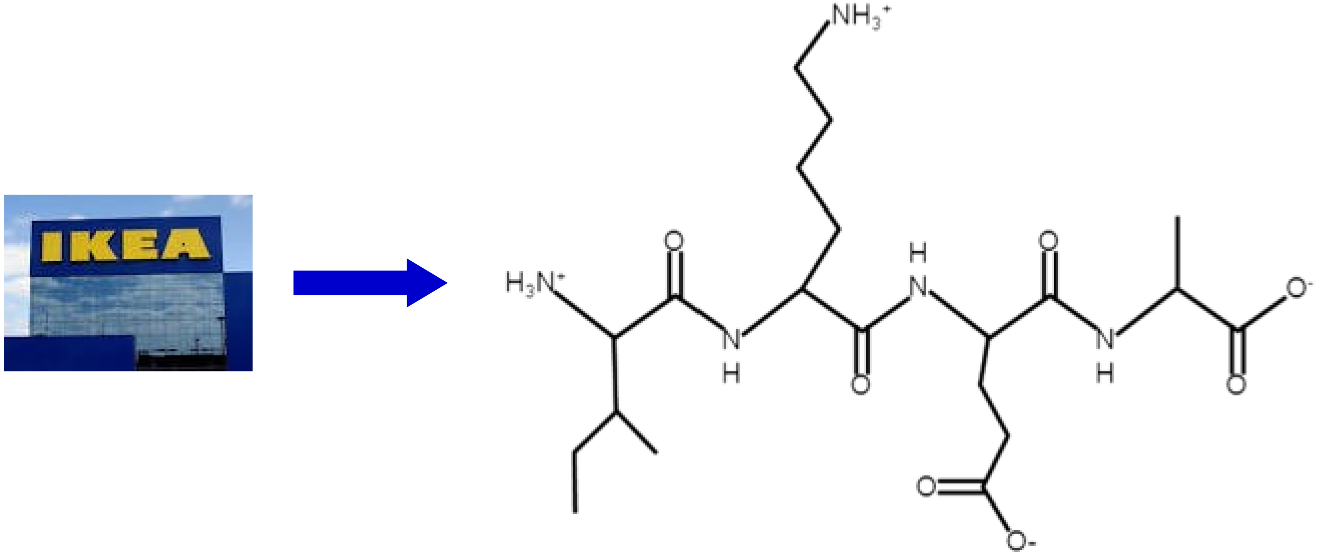
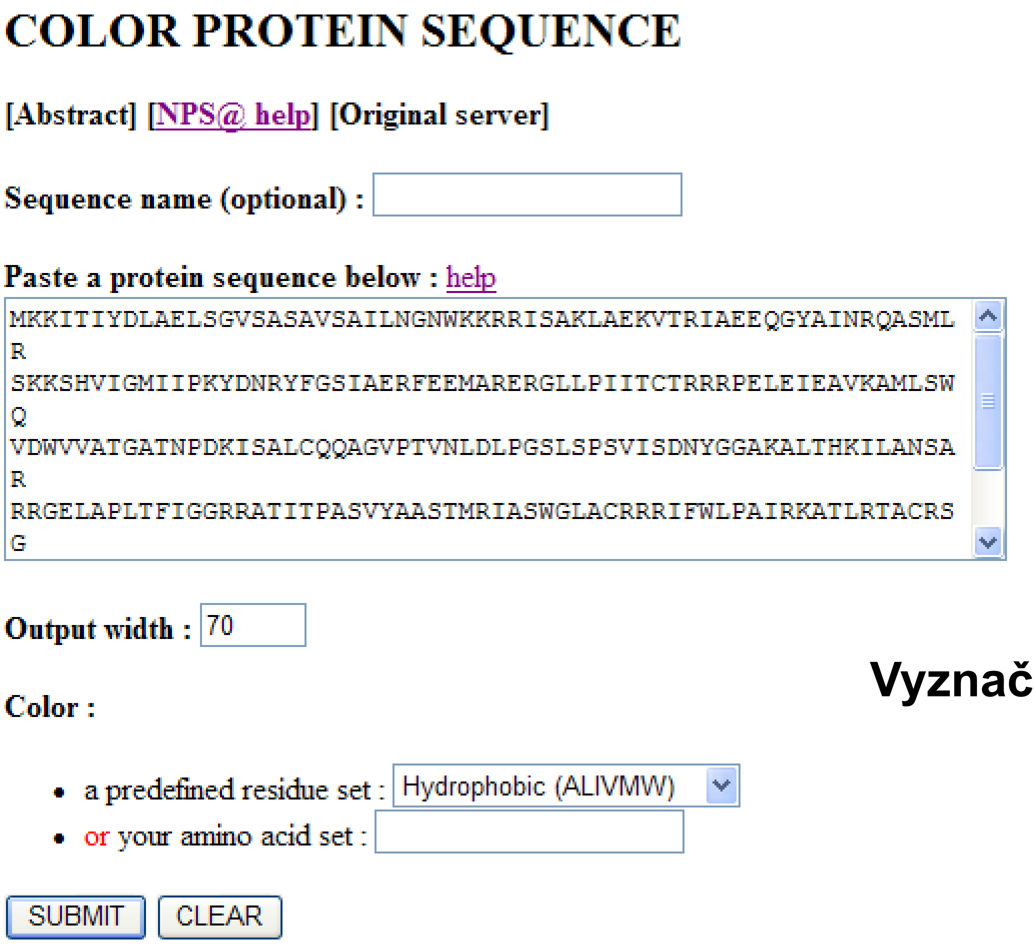


Nástroje pro zpracování a grafickou úpravu sekvenčních dat

Jaro 2015

ColorSeq

http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_color.html



COLOR PROTEIN SEQUENCE

[Abstract] [NPS@ help] [Original server]

Sequence name (optional) :

Paste a protein sequence below : help

```
MKKITIYDLAELSGVSASAVSAILNGNWKKRRISAKLAEKVTRIAEEQGYAINRQASML
R
SKKSHVIGMIIPKYDNRYFGSIAERFEEMARERGLLPIITCTRRRPELEIEAVKAMLSW
Q
VDWVVATGATNPDKISALCQQAGVPTVNLDLPGSLSPSVISDNYGGAKALTHKILANSA
R
RRGELAPLTFIGGRRATITPASVYAASTMRIASWGLACRRRIFWLPAIRKATLRTACRS
G
```

Output width : 70

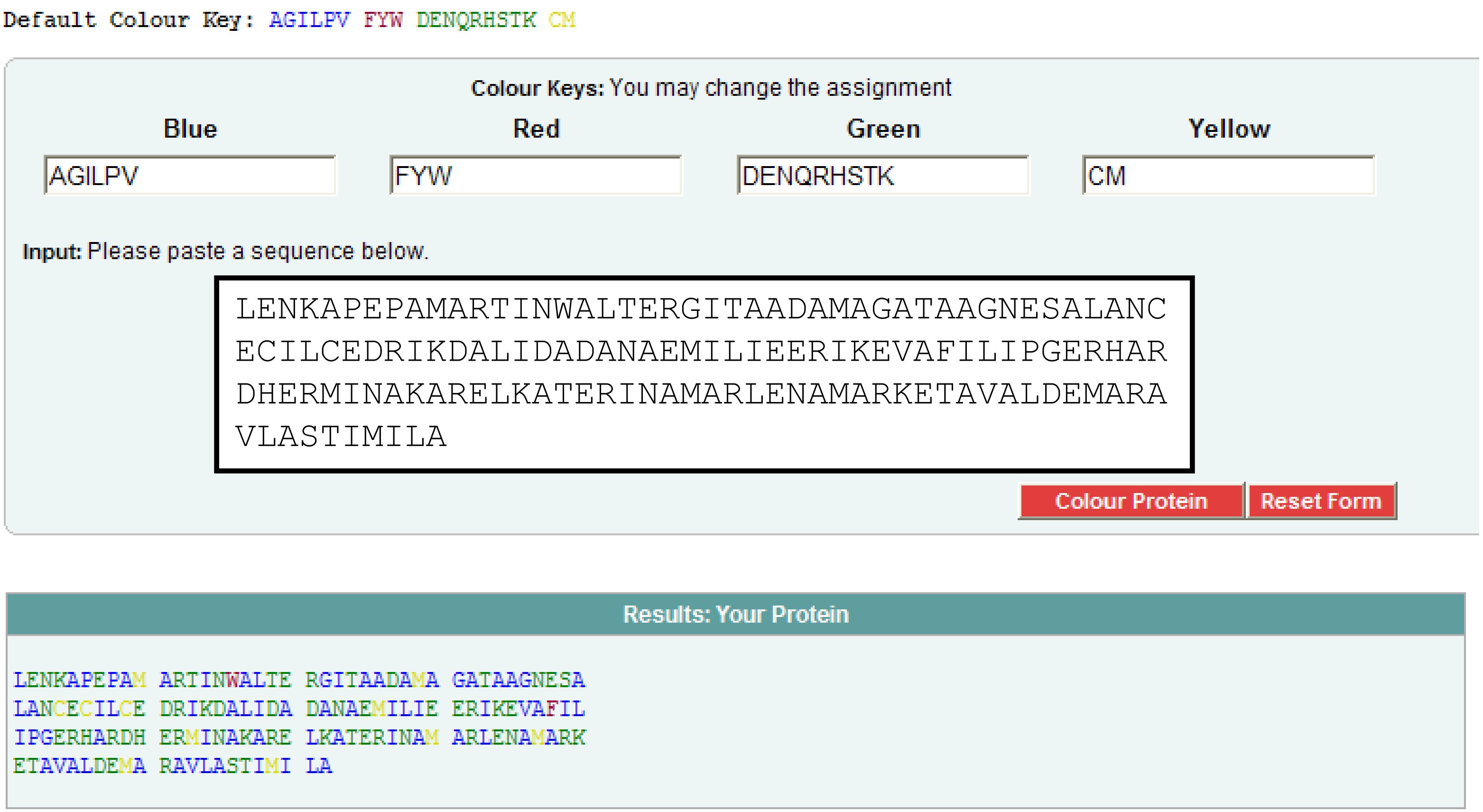
Color :

- a predefined residue set : Hydrophobic (ALIVMW)
- or your amino acid set :

Vyznačte v sekvenci všechny tryptofany, tyrosiny a cysteiny.

Protein Colourer

http://www.ebi.ac.uk/cgi-bin/proteincol/ProteinColourer.pl



Default Colour Key: AGILPV FYW DENQRHSTK CM

Colour Keys: You may change the assignment

Blue	Red	Green	Yellow
AGILPV	FYW	DENQRHSTK	CM

Input: Please paste a sequence below.

```
LENKAPEPAMARTINWALTERGITAADAMAGATAAGNESALANC
ECILCEDRIKDALIDADANAEMILIEERIKEVAFILIPGERHAR
DHERMINAKARELKATERINAMARLENAMARKETAVALDEMARA
VLASTIMILA
```

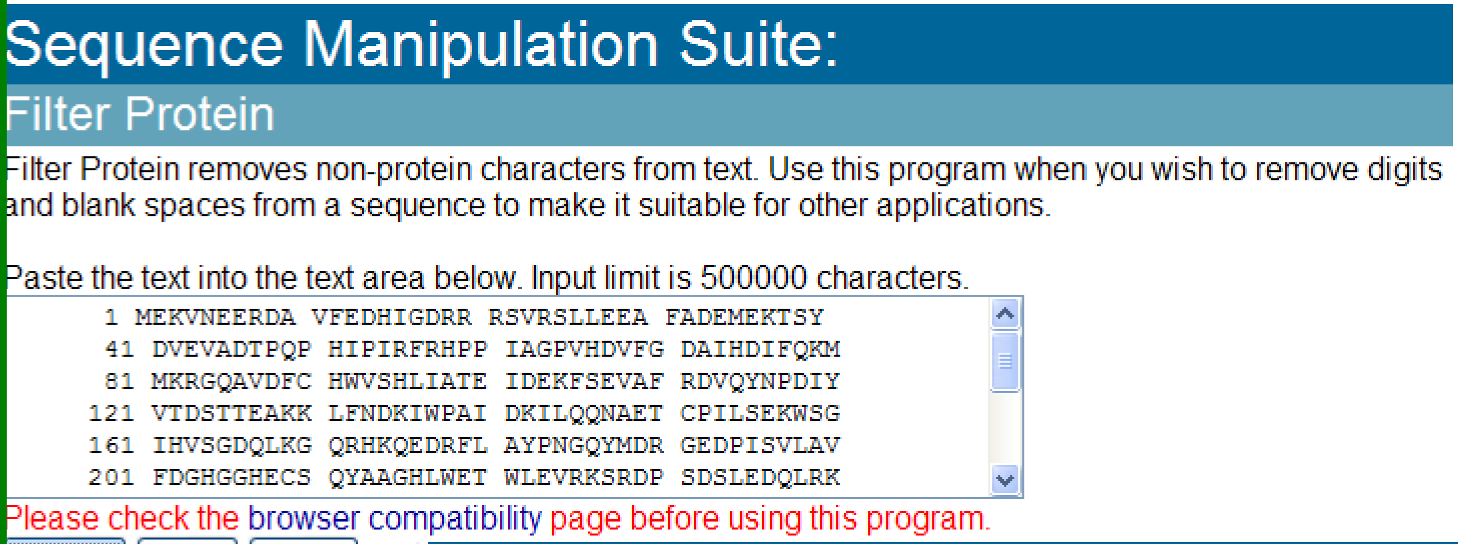
Results: Your Protein

```
LENKAPEPAM ARTINWALTE RGITAADAMA GATAAGNESA
LANCECILCE DRIKDALIDA DANAEMILIE ERIKEVAFIL
IPGERHARDH ERMINAKARE LKATERINAM ARLENAMARK
ETAVALDEMA RAVLASTIMI LA
```

Obarvěte si sekvenci...

Filter Protein/DNA

http://www.bioinformatics.org/sms2/filter_protein.html
http://www.bioinformatics.org/sms2/filter_dna.html



Sequence Manipulation Suite:

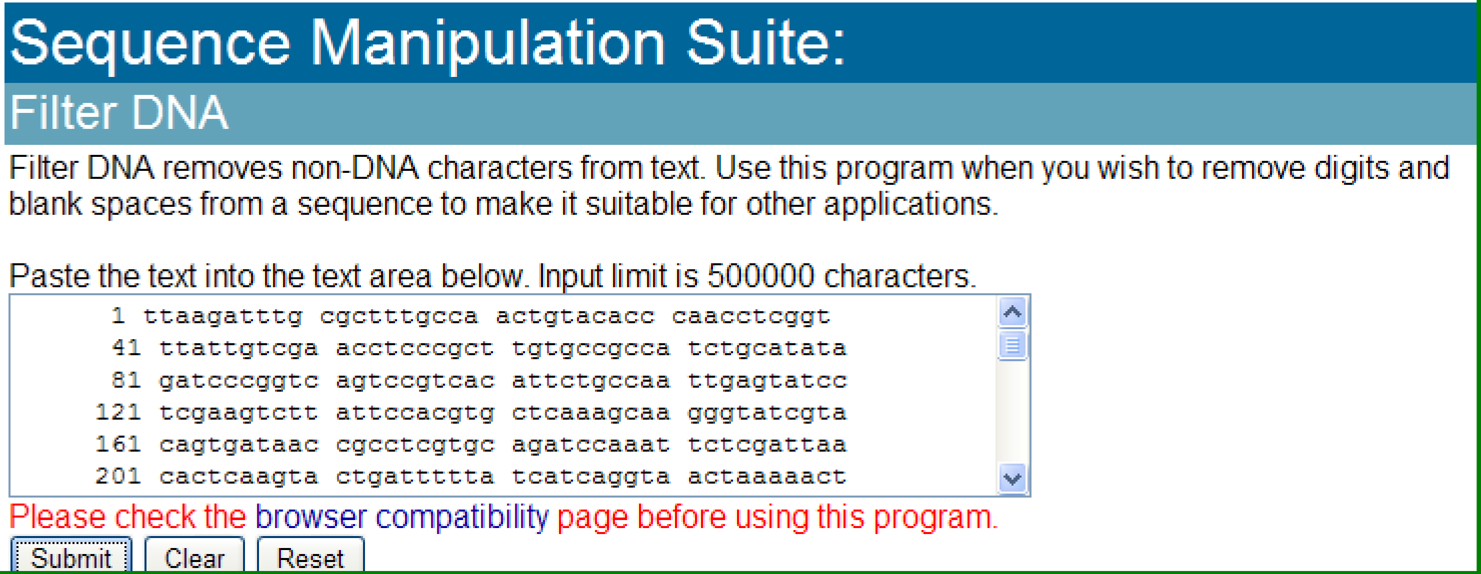
Filter Protein

Filter Protein removes non-protein characters from text. Use this program when you wish to remove digits and blank spaces from a sequence to make it suitable for other applications.

Paste the text into the text area below. Input limit is 500000 characters.

```
      1 MEKVNEERDA VFEDHIGDRR RSVRSLLEEA FADEMEKTSY
     41 DVEVADTPQP HIPIRFRHPP IAGPVHDVFG DAIHDIFQKM
     81 MKRGQAVDFC HWVSHLIATE IDEKFSEVAF RDVQYNPDIY
    121 VTDSTTEAKK LFNDKIWPAI DKILQQNAET CPILSEKWSG
    161 IHVSGDQLKG QRHKQEDRFL AYPNGQYMDR GEDPISVLAV
    201 FDGHGGHECS QYAAGHLWET WLEVRKSRDP SDSLEDQLRK
```

Please check the browser compatibility page before using this program.



Sequence Manipulation Suite:

Filter DNA

Filter DNA removes non-DNA characters from text. Use this program when you wish to remove digits and blank spaces from a sequence to make it suitable for other applications.

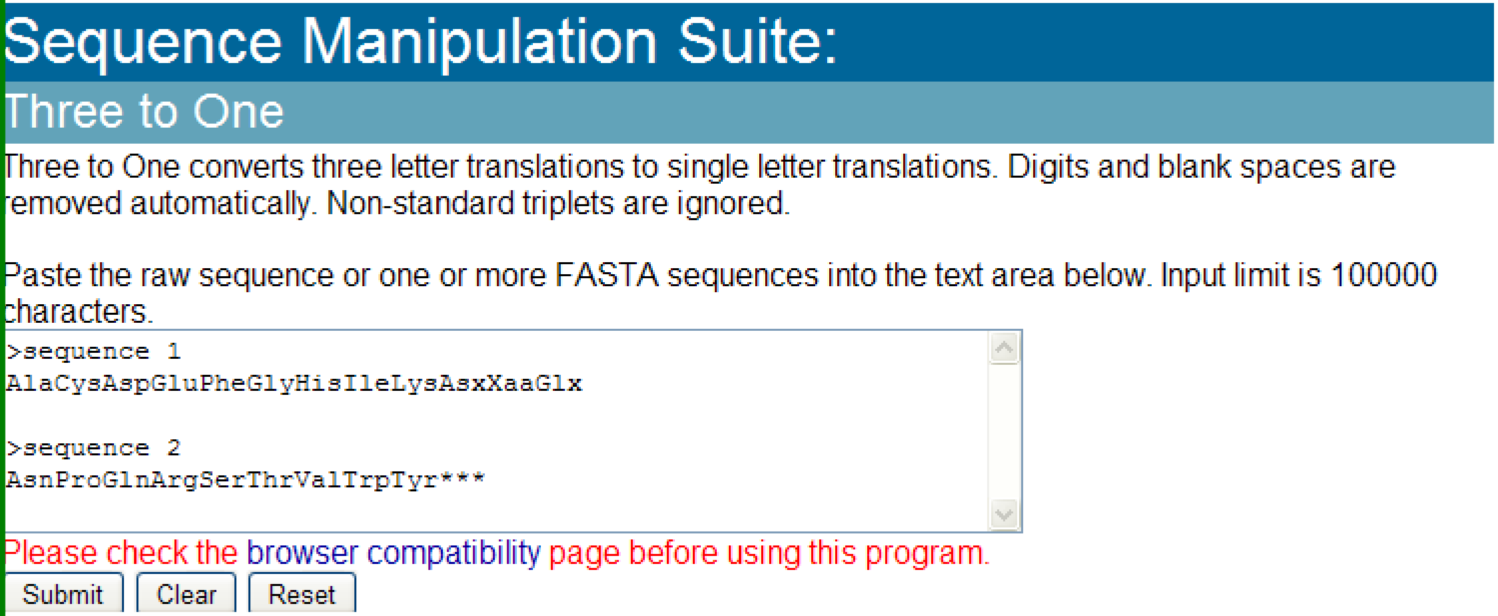
Paste the text into the text area below. Input limit is 500000 characters.

```
      1 ttaagatttg cgctttgcca actgtacacc caacctcggt
     41 ttattgtcga acctcccgct tgtgccgcca tctgcatata
     81 gatcccggtc agtccgtcac attctgccaa ttgagtatcc
    121 tcgaagtctt attccacgtg ctcaaagcaa gggtatcgta
    161 cagtgataac cgcctcgtgc agatccaaat tctcgattaa
    201 cactcaagta ctgattttta tcatcaggta actaaaaact
```

Please check the browser compatibility page before using this program.

Three to One

http://www.bioinformatics.org/sms2/three_to_one.html



Sequence Manipulation Suite:

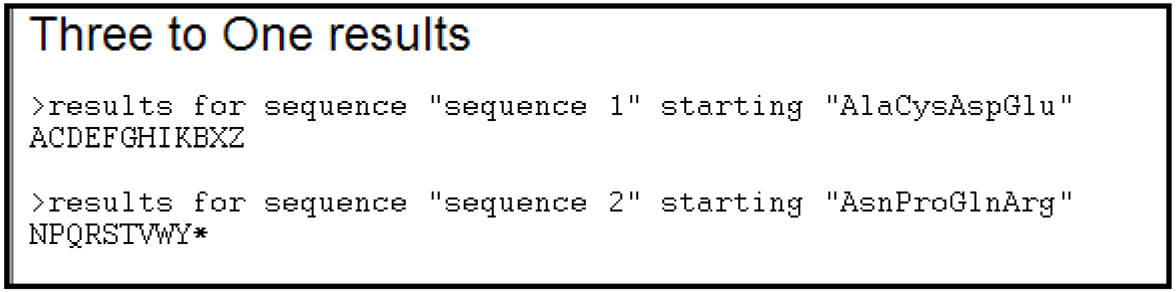
Three to One

Three to One converts three letter translations to single letter translations. Digits and blank spaces are removed automatically. Non-standard triplets are ignored.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 100000 characters.

```
>sequence 1
AlaCysAspGluPheGlyHisIleLysAsxXaaGlx

>sequence 2
AsnProGlnArgSerThrValTrpTyr***
```

Please check the browser compatibility page before using this program.

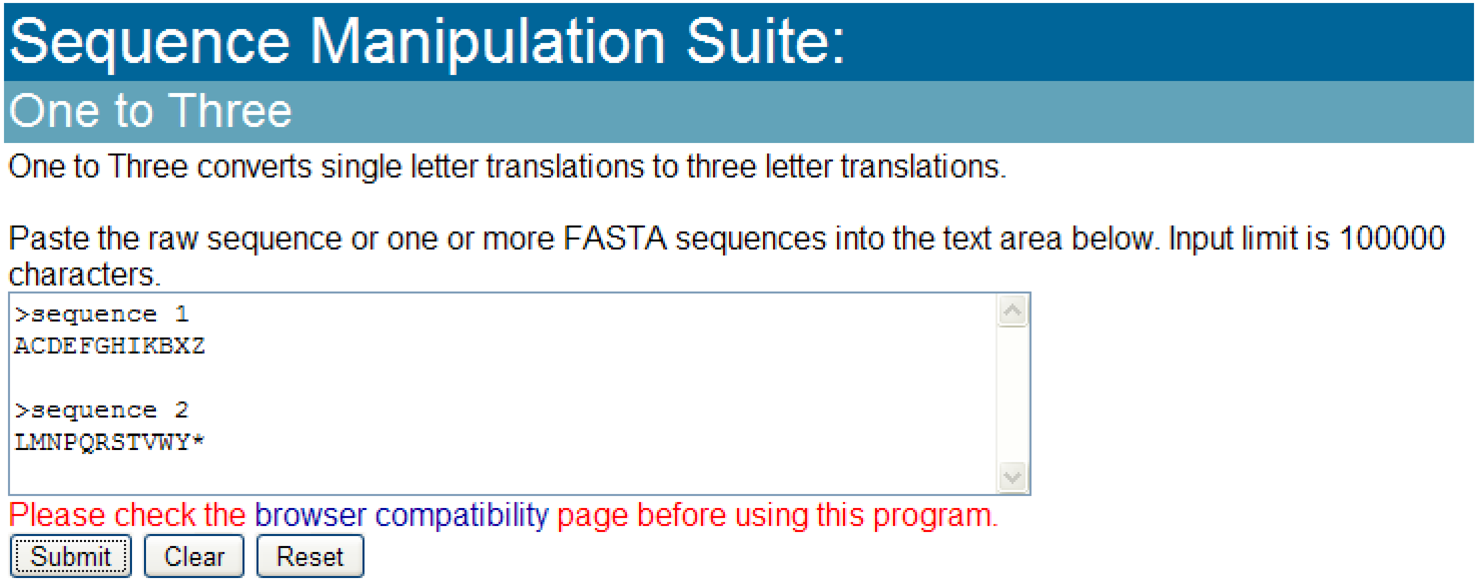


Three to One results

```
>results for sequence "sequence 1" starting "AlaCysAspGlu"
ACDEFGHIKBXZ

>results for sequence "sequence 2" starting "AsnProGlnArg"
NPQRSTVWY*
```

One to Three

http://www.bioinformatics.org/sms2/one_to_three.html



Sequence Manipulation Suite:

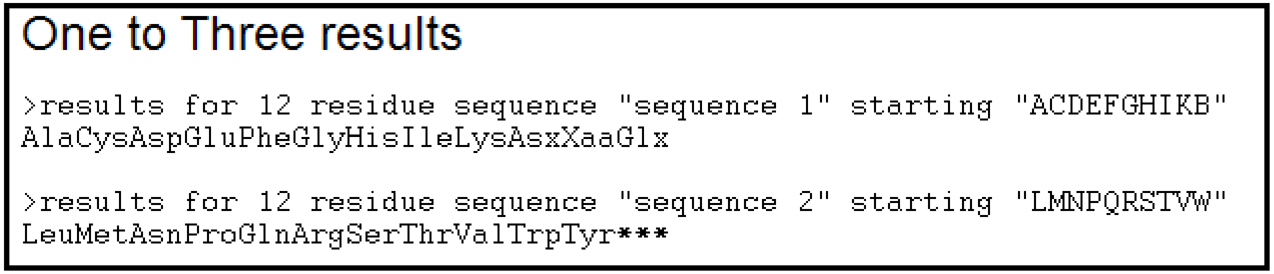
One to Three

One to Three converts single letter translations to three letter translations.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 100000 characters.

```
>sequence 1
ACDEFGHIKBXZ

>sequence 2
LMNPQRSTVWY*
```

Please check the browser compatibility page before using this program.

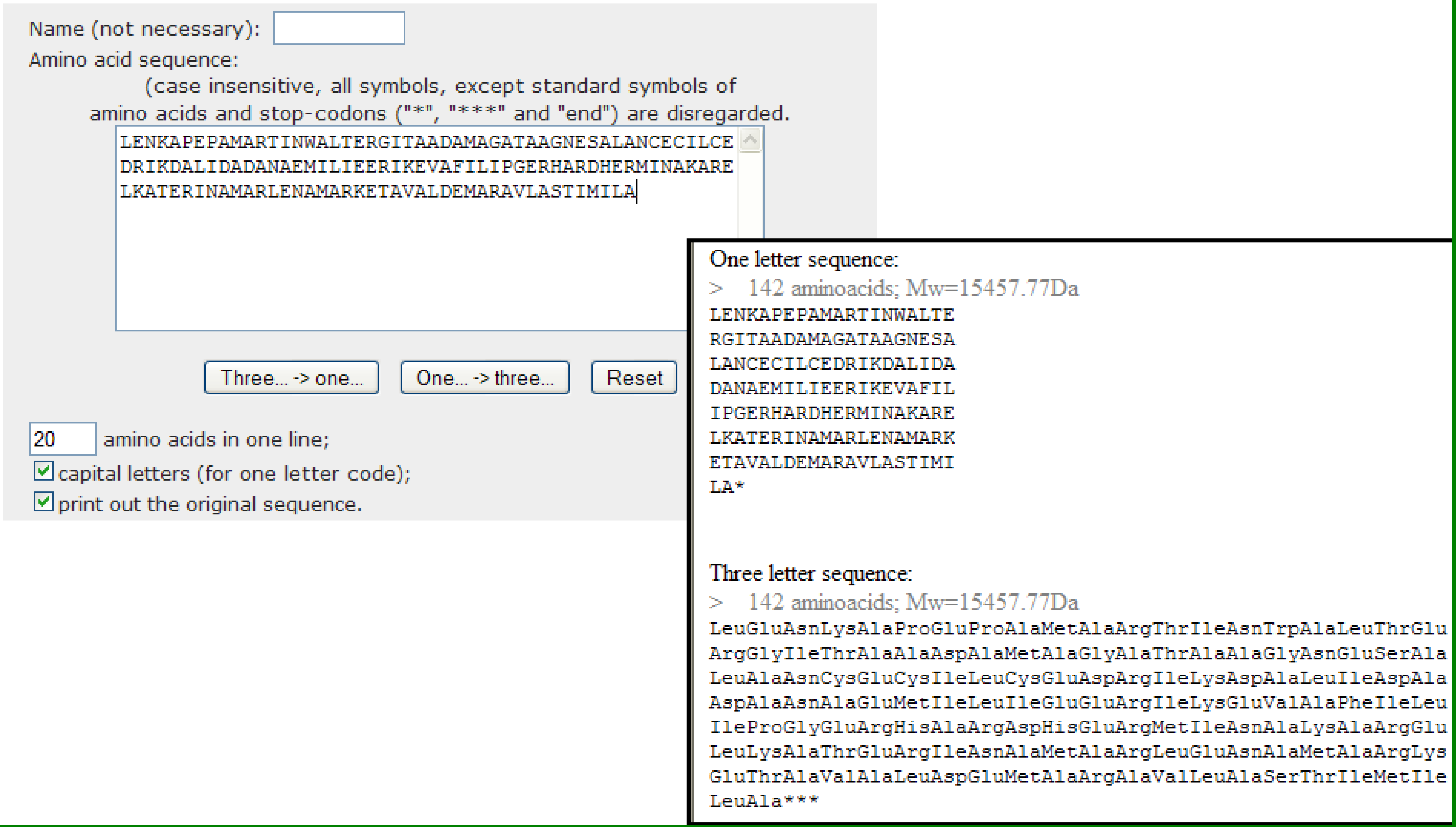


One to Three results

```
>results for 12 residue sequence "sequence 1" starting "ACDEFGHIKB"
AlaCysAspGluPheGlyHisIleLysAsxXaaGlx

>results for 12 residue sequence "sequence 2" starting "LMNPQRSTVW"
LeuMetAsnProGlnArgSerThrValTrpTyr***
```

Three- / one-letter code

http://molbiol.ru/eng/scripts/01_17.html



Name (not necessary):

Amino acid sequence:

(case insensitive, all symbols, except standard symbols of amino acids and stop-codons ("*", "***" and "end") are disregarded.

```
LENKAPEPAMARTINWALTERGITAADAMAGATAAGNESALANCECILCE
DRIKDALIDADANAEMILIEERIKEVAFILIPGERHARDHERMINAKARE
LKATERINAMARLENAMARKETAVALDEMARAVLASTIMILA
```

20 amino acids in one line;

☑ capital letters (for one letter code);

☑ print out the original sequence.

One letter sequence:

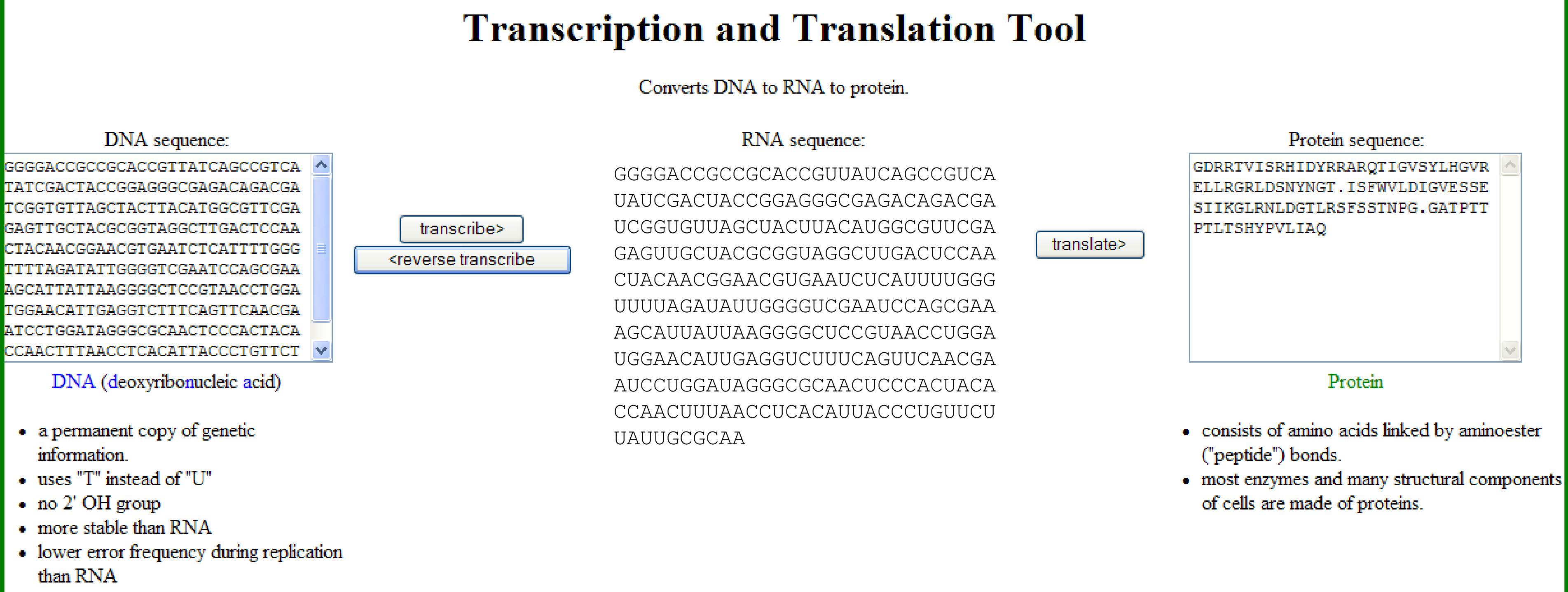
```
>   142 aminoacids; Mw=15457.77Da
LENKAPEPAMARTINWALTE
RGITAADAMAGATAAGNESA
LANCECILCEDRIKDALIDA
DANAEMILIEERIKEVAFIL
IPGERHARDHERMINAKARE
LKATERINAMARLENAMARK
ETAVALDEMARAVLASTIMI
LA*
```

Three letter sequence:

```
>   142 aminoacids; Mw=15457.77Da
LeuGluAsnLysAlaProGluProAlaMetAlaArgThrIleAsnTrpAlaLeuThrGlu
ArgGlyIleThrAlaAlaAspAlaMetAlaGlyAlaThrAlaAlaGlyAsnGluSerAla
LeuAlaAsnCysGluCysIleLeuCysGluAspArgIleLysAspAlaLeuIleAspAla
AspAlaAsnAlaGluMetIleLeuIleGluGluArgIleLysGluValAlaPheIleLeu
IleProGlyGluArgHisAlaArgAspHisGluArgMetIleAsnAlaLysAlaArgGlu
LeuLysAlaThrGluArgIleAsnAlaMetAlaArgLeuGluAsnAlaMetAlaArgLys
GluThrAlaValAlaLeuAspGluMetAlaArgAlaValLeuAlaSerThrIleMetIle
LeuAla***
```

Reverse Transcription and Translation Tool

http://www.attotron.com/cybertory/analysis/trans.htm



Transcription and Translation Tool

Converts DNA to RNA to protein.

DNA sequence:

```
GGGGACCGCCGCACCGTTATCAGCCGTCA
TATCGACTACCGGAGGGCGAGACAGACGA
TCGGTGTTAGCTACTTACATGGCGTTCGA
GAGTTGCTACGCGGTAGGCTTGACTCCAA
CTACAACGGAACGTGAATCTCATTTTGGG
TTTTAGATATTGGGGTCGAATCCAGCGAA
AGCATTATTAAGGGGCTCCGTAACCTGGA
TGGAACATTGAGGTCTTTCAGTTCAACGA
ATCCTGGATAGGGCGCAACTCCCACTACA
CCAACTTTAACCTCACATTACCCTGTTCT
```

DNA (deoxyribonucleic acid)

- a permanent copy of genetic information.
- uses "T" instead of "U"
- no 2' OH group
- more stable than RNA
- lower error frequency during replication than RNA

RNA sequence:

```
GGGGACCGCCGCACCGUUAUCAGCCGUCA
UAUCGACUACCGGAGGGCGAGACAGACGA
UCGGUGUUAGCUACUUACAUGGCGUUCGA
GAGUUGCUACGCGGUAGGCUUGACUCCAA
CUACAACGGAACGUGAAUCUCAUUUUGGG
UUUUAGAUAUUGGGGUCGAAUCCAGCGAA
AGCAUUAUUAAGGGGCUCCGUAACCUGGA
UGGAACAUUGAGGUCUUUCAGUUCAACGA
AUCCUGGAUAGGGCGCAACUCCCACUACA
CCAACUUUAACCUCACAUUACCCUGUUCU
UAUUGCGCAA
```

Protein sequence:

```
GDRRTVISRHIDYRRARQTIGVSYLHGVR
ELLRGRLDSNYNGT.ISFWVLDIGVESSE
SIIKGLRNLDGTLRSFSSTNPG.GATPTT
PTLTSHYPVLIAQ
```

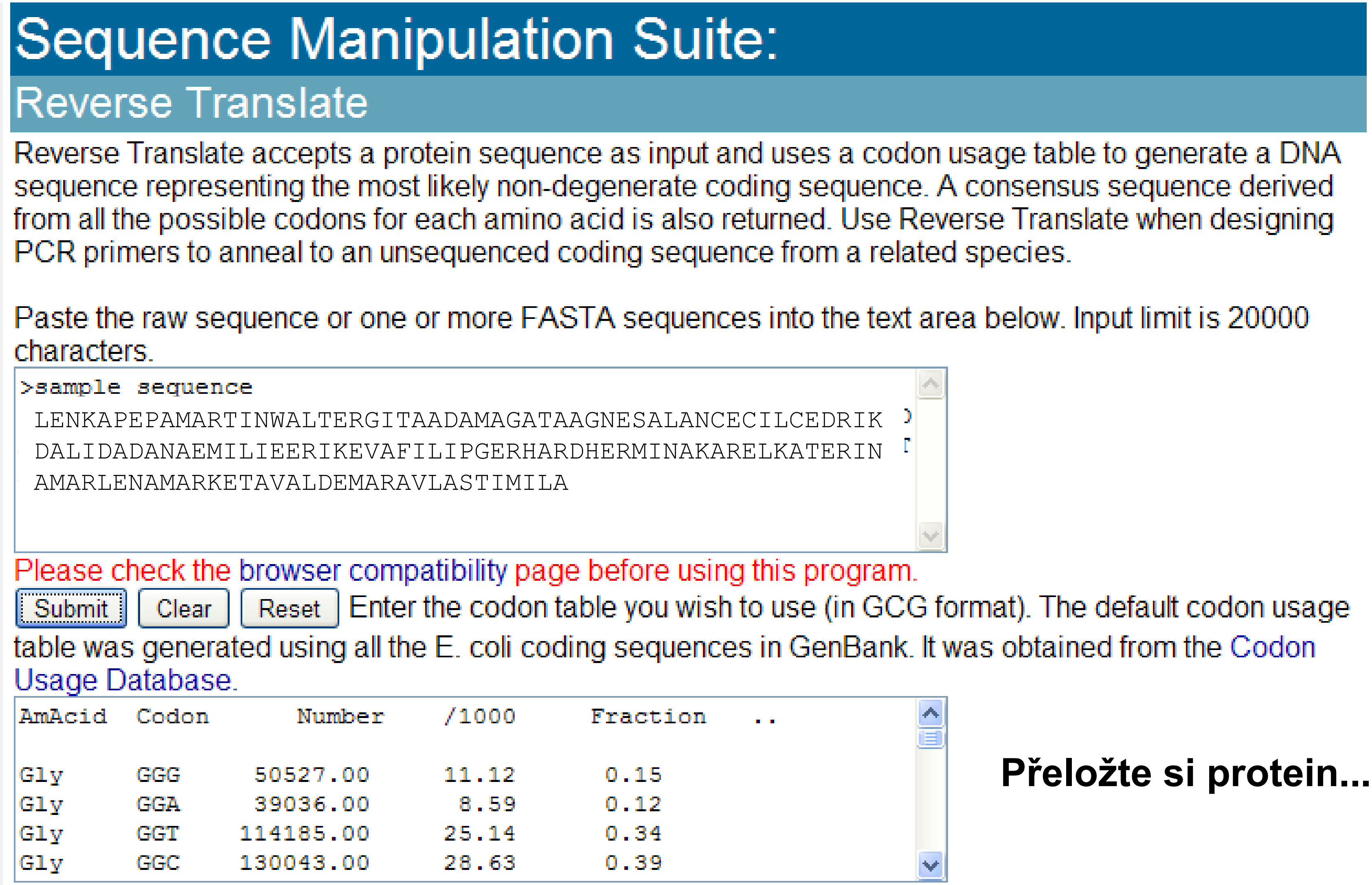
Protein

- consists of amino acids linked by aminoester ("peptide") bonds.
- most enzymes and many structural components of cells are made of proteins.

Pohrajte si s RNA...

Reverse Translate

http://www.bioinformatics.org/sms2/rev_trans.html



Sequence Manipulation Suite:

Reverse Translate

Reverse Translate accepts a protein sequence as input and uses a codon usage table to generate a DNA sequence representing the most likely non-degenerate coding sequence. A consensus sequence derived from all the possible codons for each amino acid is also returned. Use Reverse Translate when designing PCR primers to anneal to an unsequenced coding sequence from a related species.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 20000 characters.

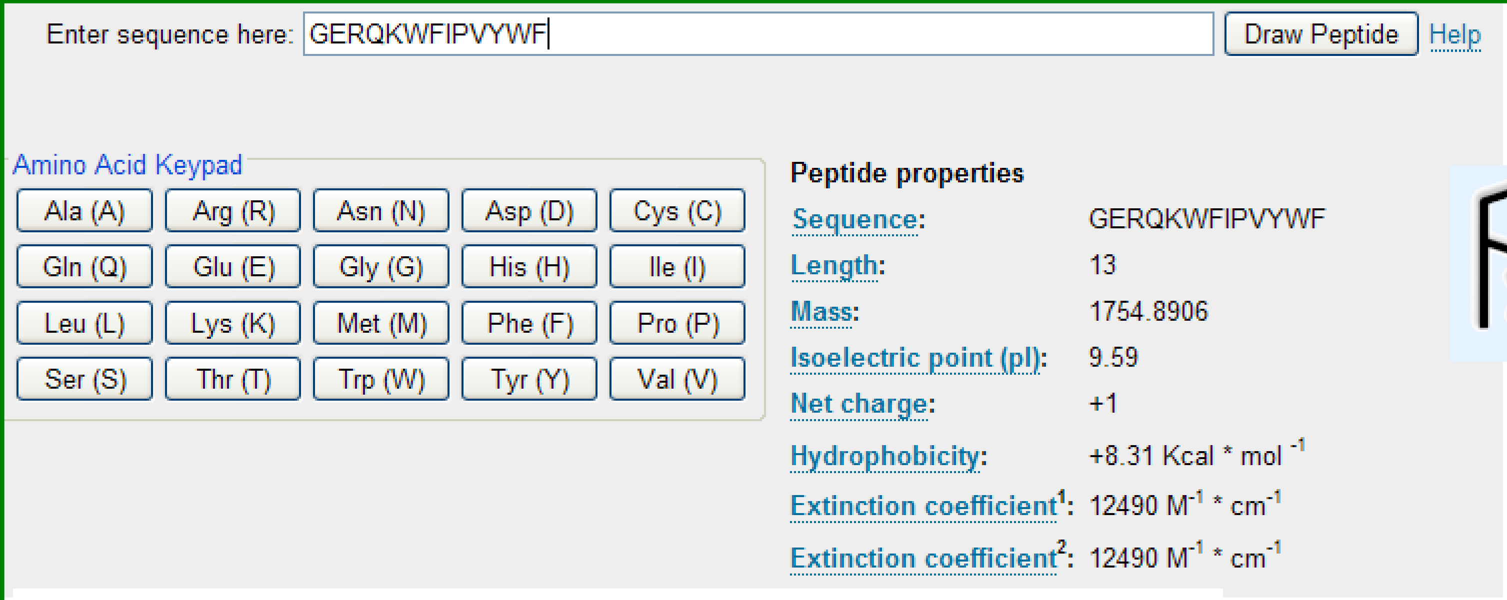
```
>sample sequence
 LENKAPEPAMARTINWALTERGITAADAMAGATAAGNESALANCECILCEDRIK
 DALIDADANAEMILIEERIKEVAFILIPGERHARDHERMINAKARELKATERIN
 AMARLENAMARKETAVALDEMARAVLASTIMILA
```

Please check the browser compatibility page before using this program.

Enter the codon table you wish to use (in GCG format). The default codon usage table was generated using all the E. coli coding sequences in GenBank. It was obtained from the Codon Usage Database.

```
AmAcid  Codon      Number    /1000     Fraction   ..

Gly     GGG     50527.00     11.12      0.15
Gly     GGA     39036.00      8.59      0.12
Gly     GGT    114185.00     25.14      0.34
Gly     GGC    130043.00     28.63      0.39
```

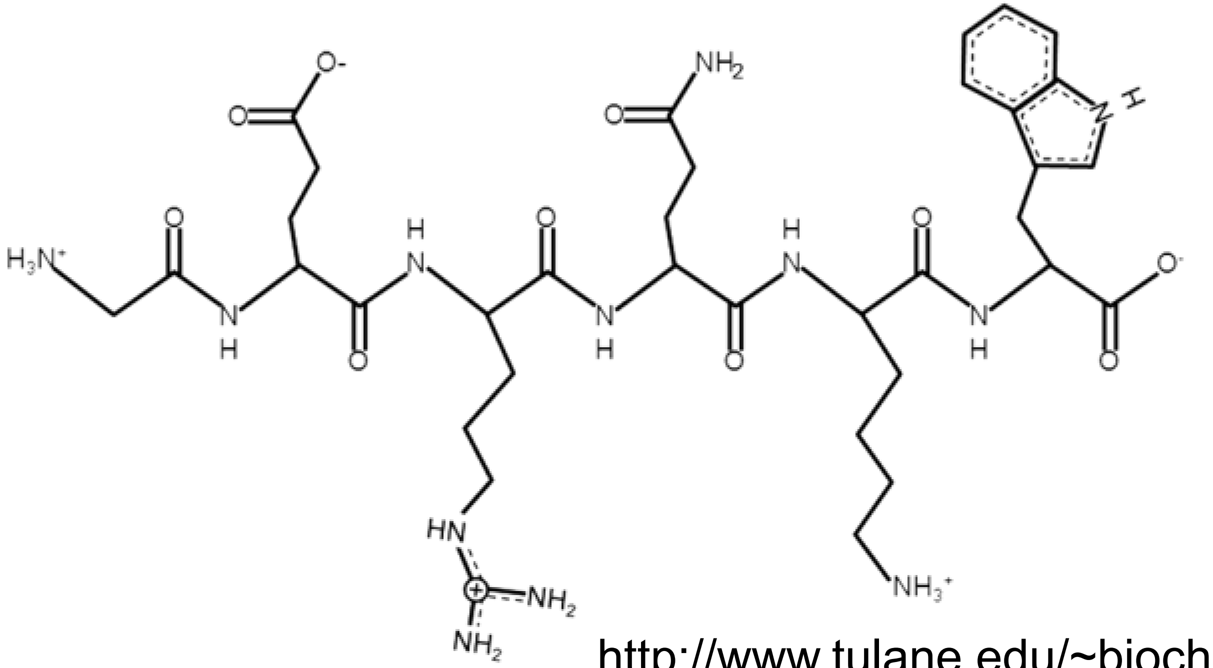
Přeložte si protein...



Enter sequence here: GERQKWFIPVYWF

Peptide properties

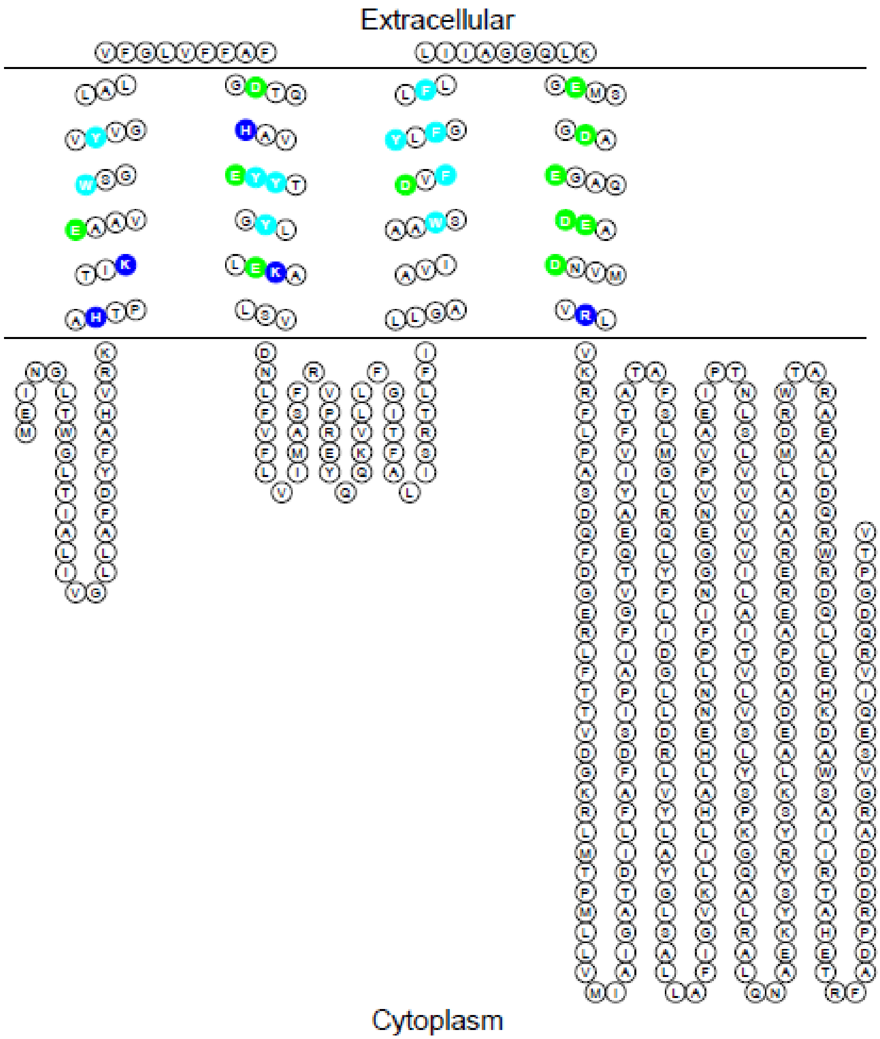
Sequence:	GERQKWFIPVYWF
Length:	13
Mass:	1754.8906
Isoelectric point (pI):	9.59
Net charge:	+1
Hydrophobicity:	+8.31 Kcal * mol $^{-1}$
Extinction coefficient[1]:	12490 M^{-1} * cm^{-1}
Extinction coefficient[2]:	12490 M^{-1} * cm^{-1}



Kolik aminokyselin program zvládne?

PepDraw

http://www.tulane.edu/~biochem/WW/PepDraw/index.html



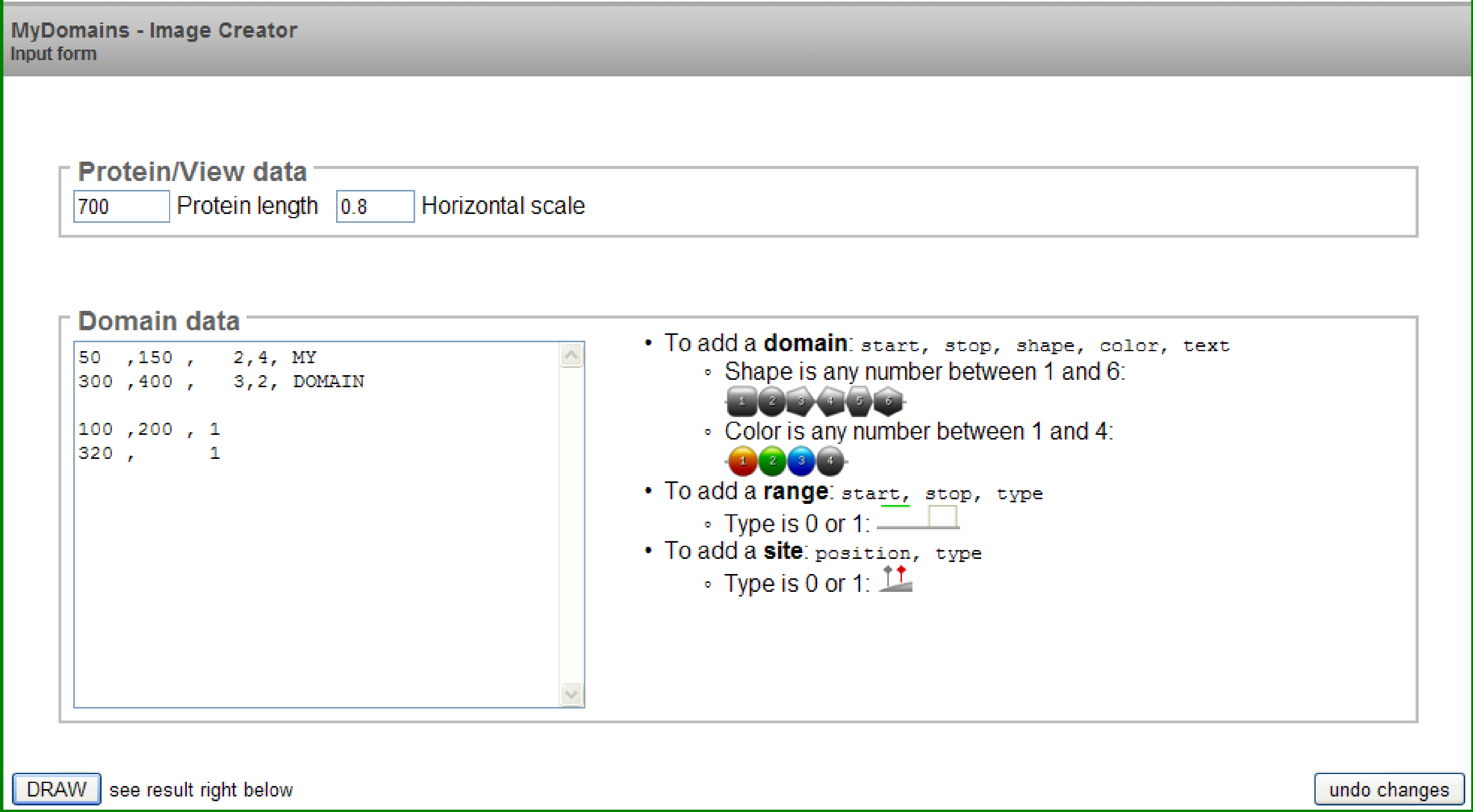
TOPO2

http://www.sacs.ucsf.edu/TOPO2/

TOPO2 software is for creating transmembrane protein 2D topology images. **It makes no attempt to predict the TMDs that it displays.** The user needs to supply that information. Residues of interest can be highlighted, if desired.

MyDomains

http://prosite.expasy.org/mydomains



MyDomains - Image Creator

Input form

Protein/View data

700 Protein length 0.8 Horizontal scale

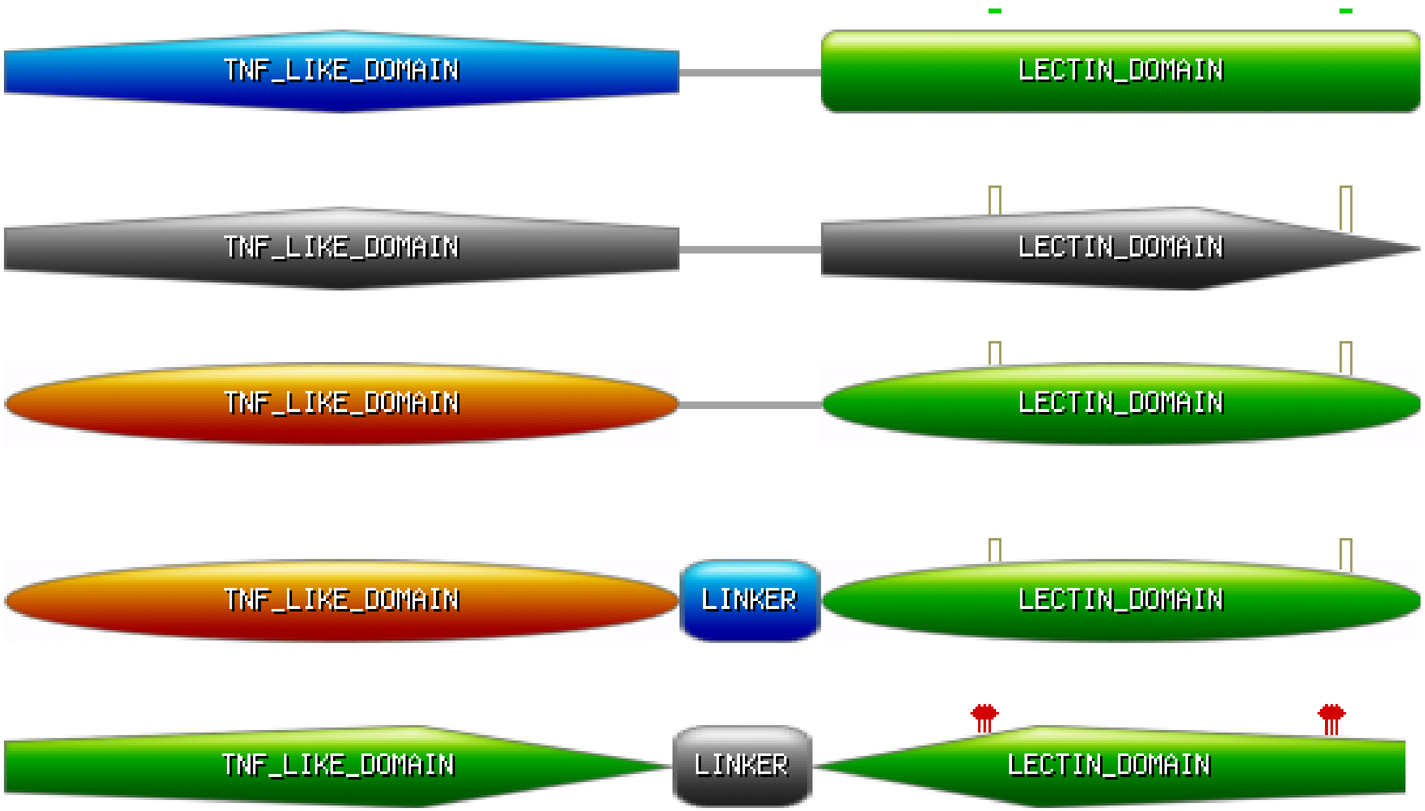
Domain data

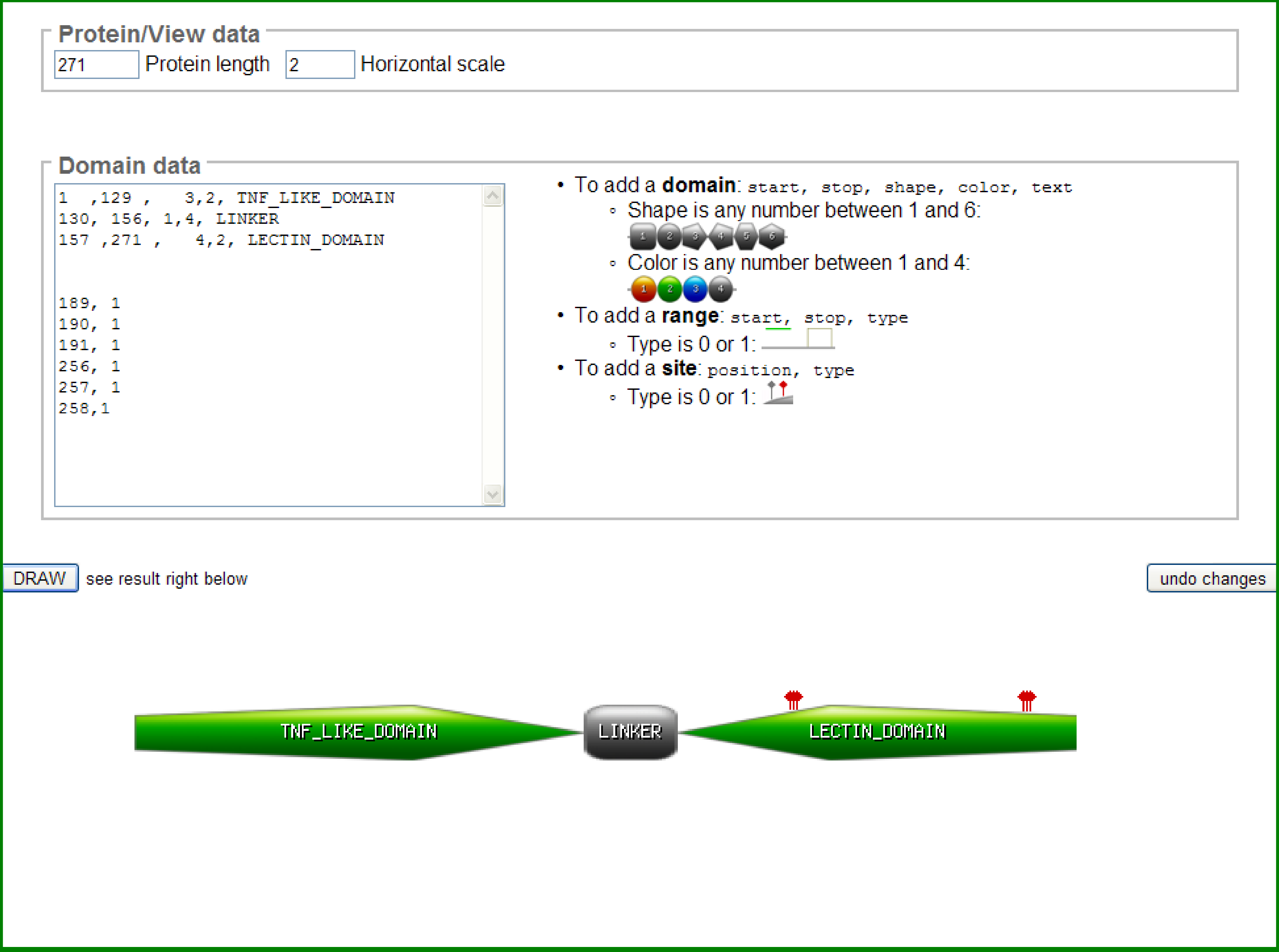
```
50  ,150 ,   2,4, MY
300 ,400 ,   3,2, DOMAIN

100 ,200 , 1
320 ,      1
```

- To add a **domain**: `start, stop, shape, color, text`
  - Shape is any number between 1 and 6:
  - Color is any number between 1 and 4:
- To add a **range**: `start, stop, type`
  - Type is 0 or 1:
- To add a **site**: `position, type`
  - Type is 0 or 1:

PLLSASIVSAPVVTSETYVDIPGLYLDVAKAGIRDGKLQVILNVPTPYATGNNFPGIYFAI
ATNQGVVADGCFTYSSKVPESTGRMPFTLVATIDVGSGVTFVKGQWKSVRGSAMHIDSYAS
LSAIWGTAAPSSQGSGNQGAETGGTGAGNIGGGG**ERDGTFNLPPHIKFGVTALTHAANDQT**
IDIYIDDDPKPAATFKGAGAQDQNLGTKVLDSGNGRVRVIVMANGRPSRLGSRQVDIFKKS
YFGIIGSEDGADDDYNDGIVFLNWPLG

271 aa, 1-129 TNF like domain, 157-271 Lectin domain,
189-191 vazebné místo, 256-258 vazebné místo





Protein/View data

Protein length: 271

Horizontal scale: 2

Domain data

```
1   ,129 ,    3,2, TNF_LIKE_DOMAIN
130, 156, 1,4, LINKER
157 ,271 ,    4,2, LECTIN_DOMAIN



189, 1
190, 1
191, 1
256, 1
257, 1
258,1
```

- To add a **domain**: `start, stop, shape, color, text`
  - Shape is any number between 1 and 6:
  - Color is any number between 1 and 4:
- To add a **range**: `start, stop, type`
  - Type is 0 or 1:
- To add a **site**: `position, type`
  - Type is 0 or 1:

```
FEATURES             Location/Qualifiers
     source          1..248
                     /organism="Candidatus Pelagibacter ubique HTCC1062"
                     /strain="HTCC1062"
                     /db_xref="taxon:335992"
     Protein         1..248
                     /product="multi-domain protein"
     Region          13..61
                     /region_name="SH3_3"
                     /note="Bacterial SH3 domain; pfam08239"
                     /db_xref="CDD:254683"
     Region          <129..220
                     /region_name="Spr"
                     /note="Cell wall-associated hydrolases
                     (invasion-associated proteins) [Cell envelope biogenesis,
                     outer membrane]; COG0791"
                     /db_xref="CDD:223862"
     Region          141..>213
                     /region_name="NLPC_P60"
                     /note="NlpC/P60 family; cl17555"
                     /db_xref="CDD:266746"
     CDS             1..248
                     /locus_tag="SAR11_0214"
                     /coded_by="CP000084.1:215418..216164"
                     /note="SH3 and NLP/P60 domains (Pfam)"
                     /transl_table=11
ORIGIN
        1 mkdnyfykkp lsniykkpna fsevtsqily gekfkiiskn knwikikvsf dnytgyiknk
       61 yytkdhqpth kiftlkaniy nkqknktknf lpfasrismi denkkfiefe knkwikkkdi
      121 kkinhiekdy lkvlkmflki kylwggktyr gidcsailql ffyynnkfyp rdtkdqikys
      181 akknkskvfk kgdiifwkgh vavcinaqkl ihaygpekkv limniketin riertakltv
      241 kkispiky
//
```

Vytvořte „doménové schéma“ proteinu.