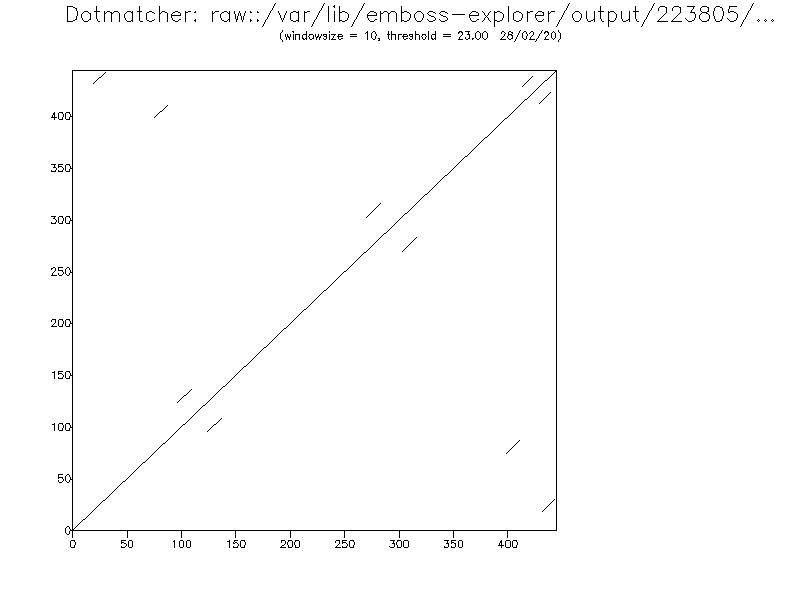
**Dot plot**

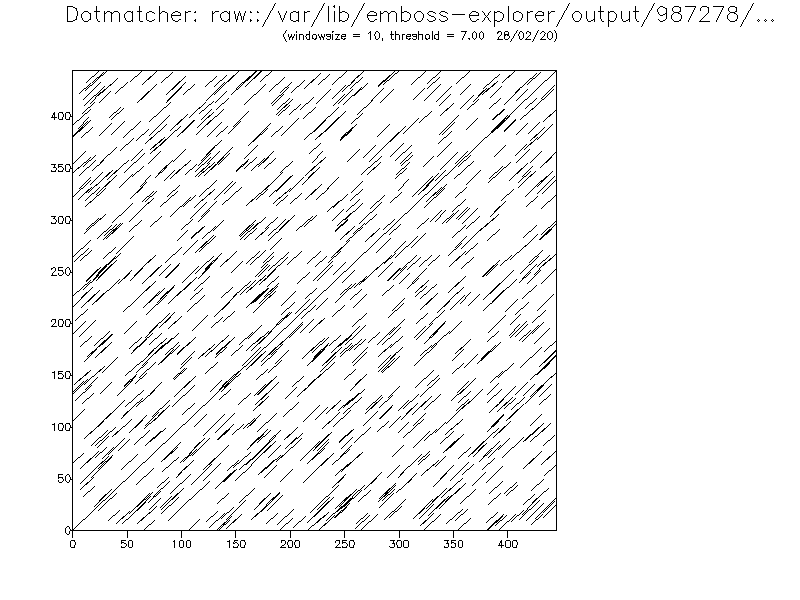
Подходът с точкова матрица, който имплицитно произвежда семейство от подравнения за отделните области на последователността, е качествен и концептуално прост, макар и да отнема много време за анализ в голям мащаб. При липса на шум може да бъде лесно визуално да се идентифицират определени характеристики на последователността - като вмъквания, изтривания, повторения или обърнати повторения - от точка-матрица. За да се конструира точково-матрична графика, двете последователности се записват по горния ред и в най-лявата колона на двуизмерна матрица и точка се поставя във всяка точка, където знаците в съответните колони съвпадат - това е типичен график за рецидивиране. Някои изпълнения варират размера или интензитета на точката в зависимост от степента на сходство на двата знака, за да се настанят консервативните замествания. Точковите графики на много тясно свързани последователности ще се появят като една линия по основния диагонал на матрицата.

[**http://www.bioinformatics.nl/cgi-bin/emboss/dotmatcher**](http://www.bioinformatics.nl/cgi-bin/emboss/dotmatcher)

[**https://bioinfo.cristal.univ-lille.fr/yass/yass.php**](https://bioinfo.cristal.univ-lille.fr/yass/yass.php)

****

Threshold: 7



**Подравняване по двойки**

Методите за подравняване на двойки по последователност се използват за намиране на най-добре съвпадащи на части (локални или глобални) подравнения на две последователности на заявки. Двойно подравняване може да се използва само между две последователности наведнъж, но те са ефикасни за изчисляване и често се използват за методи, които не изискват изключителна точност (като търсене в база данни за последователности с висока прилика с заявка). Трите основни метода за създаване на двойно подреждане са точково-матрични методи, динамично програмиране и методи с думи; [1], обаче, множество техники за подравняване на последователности могат също да подравнят двойки последователности. Въпреки че всеки метод има своите индивидуални силни и слаби страни, и трите двойки метода срещат трудности с много повтарящи се последователности с ниско информационно съдържание - особено там, където броят на повторенията се различава в двете последователности, които трябва да бъдат подравнени. Един от начините за количествено определяне на полезността на дадено двойно подравняване е „максималното уникално съвпадение“ (MUM), или най-дългото подреждане, което се среща и в двете последователности на заявката. По-дългите MUM последователности обикновено отразяват по-близката свързаност.

<https://www.ebi.ac.uk/Tools/psa/>

<https://www.ebi.ac.uk/Tools/psa/lalign/>

<https://www.ebi.ac.uk/Tools/psa/emboss_water/>

PAM250:

# /nfs/public/ro/es/appbin/linux-x86\_64/fasta-36.3.8h/lalign36 -m 9i lalign-I20200228-125740-0523-43799497-p2m.asequence lalign-I20200228-125740-0523-43799497-p2m.bsequence -p -s P250 -f -10 -g -2 -E 10.0 -m 0 -m "F11 lalign-I20200228-125740-0523-43799497-p2m.output.lav"

LALIGN finds non-overlapping local alignments

version 36.3.8h Aug, 2019

Please cite:

X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381

Query: lalign-I20200228-125740-0523-43799497-p2m.asequence

1>>>EMBOSS\_001 - 445 aa

Library: lalign-I20200228-125740-0523-43799497-p2m.bsequence

445 residues in 1 sequences

Statistics: (shuffled [500]) MLE statistics: Lambda= 0.1315; K=0.008394

statistics sampled from 1 (1) to 500 sequences

Threshold: E() < 10 score: 39

Algorithm: Smith-Waterman (SSE2, Michael Farrar 2006) (7.2 Nov 2010)

Parameters: PAM250 matrix (17:-8), open/ext: -10/-2

Scan time: 0.040

The best non-identical alignments are: ls-w bits E(1) %\_id %\_sim alen

EMBOSS\_001 ( 445) 2121 409.3 1.2e-118 0.991 1.000 445

+- 69 20.0 0.17 0.164 0.656 183

+- 69 20.0 0.17 0.149 0.657 181

+- 47 15.8 0.97 0.302 0.651 43

+- 47 15.8 0.97 0.302 0.651 43

+- 42 14.9 1 0.321 0.714 28

+- 41 14.7 1 0.286 0.714 28

+- 40 14.5 1 0.168 0.554 101

+- 40 14.5 1 0.321 0.714 28

>>>EMBOSS\_001, 445 aa vs lalign-I20200228-125740-0523-43799497-p2m.bsequence library

>>EMBOSS\_001 (445 aa)

Waterman-Eggert score: 2121; 409.3 bits; E(1) < 1.2e-118

99.1% identity (100.0% similar) in 445 aa overlap (1-445:1-445)

10 20 30 40 50 60

EMBOSS MISLNIEKTLGFISKENVFAYEAQVKAAQEALENGTGKGNDFLGWLHLPSSITKEHLADL

::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

EMBOSS MISLNIEKTLGFISKENVFAYEAQVKAAQEALENGTGKGNDFLGWLHLPSSITKEHLADL

10 20 30 40 50 60

70 80 90 100 110 120

EMBOSS KATAQVLRENCEAVVVAGIGGSYLGARAVIEALSNSFTWLQDKKSGPVILYAGHNIGEDY

:::::::::::::::::::::::::::::::::::::.::::::::::::::::::::::

EMBOSS KATAQVLRENCEAVVVAGIGGSYLGARAVIEALSNSFAWLQDKKSGPVILYAGHNIGEDY

70 80 90 100 110 120

130 140 150 160 170 180

EMBOSS LYELTEYLKDKKFGVINISKSGTTTETALAFRLLKKQCEDQRGKEMAKKVIVAVTDAKKG

::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

EMBOSS LYELTEYLKDKKFGVINISKSGTTTETALAFRLLKKQCEDQRGKEMAKKVIVAVTDAKKG

130 140 150 160 170 180

190 200 210 220 230 240

EMBOSS AARVTADNEGYKSFIIPDNVGGRFSVLTPVGLLPIAVAGFDIEKLVAGAAAMEKTCGKDV

::::::::::::::::::::::::::::::::::::::::::::::::::::::.:::::

EMBOSS AARVTADNEGYKSFIIPDNVGGRFSVLTPVGLLPIAVAGFDIEKLVAGAAAMEKACGKDV

190 200 210 220 230 240

250 260 270 280 290 300

EMBOSS PFAENPAAIYAATRNELYKNGKKIEILVNFNPKLHYVSEWWKQLYGESEGKENKGIFPAS

::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

EMBOSS PFAENPAAIYAATRNELYKNGKKIEILVNFNPKLHYVSEWWKQLYGESEGKENKGIFPAS

250 260 270 280 290 300

310 320 330 340 350 360

EMBOSS VDFSTDLHSMGQWIQEGERTIYETVISVEKTQYSLQVPSDEANLDGLNFLAEKHVDEVNK

::::::::::::::::::::::::::::::::::::::::.::::::::::.::::::::

EMBOSS VDFSTDLHSMGQWIQEGERTIYETVISVEKTQYSLQVPSDKANLDGLNFLAGKHVDEVNK

310 320 330 340 350 360

370 380 390 400 410 420

EMBOSS MAELGTQLAHVDGGVPNMRIVIPALNEESLGGLLYFFEKACGISGYLLGVNPFNQPGVEA

::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

EMBOSS MAELGTQLAHVDGGVPNMRIVIPALNEESLGGLLYFFEKACGISGYLLGVNPFNQPGVEA

370 380 390 400 410 420

430 440

EMBOSS YKKNMFALLNKPGYEEESKAIQARL

:::::::::::::::::::::::::

EMBOSS YKKNMFALLNKPGYEEESKAIQARL

430 440

BLOSUM62:

# /nfs/public/ro/es/appbin/linux-x86\_64/fasta-36.3.8h/lalign36 -m 9i lalign-I20200228-125934-0881-39188781-p2m.asequence lalign-I20200228-125934-0881-39188781-p2m.bsequence -p -s BL62 -f -7 -g -1 -E 10.0 -m 0 -m "F11 lalign-I20200228-125934-0881-39188781-p2m.output.lav"

LALIGN finds non-overlapping local alignments

version 36.3.8h Aug, 2019

Please cite:

X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381

Query: lalign-I20200228-125934-0881-39188781-p2m.asequence

1>>>EMBOSS\_001 - 445 aa

Library: lalign-I20200228-125934-0881-39188781-p2m.bsequence

445 residues in 1 sequences

Statistics: (shuffled [500]) MLE statistics: Lambda= 0.0956; K=0.002525

statistics sampled from 1 (1) to 500 sequences

Threshold: E() < 10 score: 41

Algorithm: Smith-Waterman (SSE2, Michael Farrar 2006) (7.2 Nov 2010)

Parameters: BL62 matrix (11:-4), open/ext: -7/-1

Scan time: 0.050

The best non-identical alignments are: ls-w bits E(1) %\_id %\_sim alen

EMBOSS\_001 ( 445) 2267 321.2 4e-92 0.991 0.998 445

+- 68 18.0 0.53 0.230 0.460 391

+- 64 17.5 0.67 0.238 0.452 323

+- 51 15.7 0.98 0.251 0.462 199

+- 51 15.7 0.98 0.232 0.542 142

+- 50 15.5 0.99 0.225 0.542 142

+- 47 15.1 1 0.246 0.462 199

>>>EMBOSS\_001, 445 aa vs lalign-I20200228-125934-0881-39188781-p2m.bsequence library

>>EMBOSS\_001 (445 aa)

Waterman-Eggert score: 2267; 321.2 bits; E(1) < 4e-92

99.1% identity (99.8% similar) in 445 aa overlap (1-445:1-445)

10 20 30 40 50 60

EMBOSS MISLNIEKTLGFISKENVFAYEAQVKAAQEALENGTGKGNDFLGWLHLPSSITKEHLADL

::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

EMBOSS MISLNIEKTLGFISKENVFAYEAQVKAAQEALENGTGKGNDFLGWLHLPSSITKEHLADL

10 20 30 40 50 60

70 80 90 100 110 120

EMBOSS KATAQVLRENCEAVVVAGIGGSYLGARAVIEALSNSFTWLQDKKSGPVILYAGHNIGEDY

:::::::::::::::::::::::::::::::::::::.::::::::::::::::::::::

EMBOSS KATAQVLRENCEAVVVAGIGGSYLGARAVIEALSNSFAWLQDKKSGPVILYAGHNIGEDY

70 80 90 100 110 120

130 140 150 160 170 180

EMBOSS LYELTEYLKDKKFGVINISKSGTTTETALAFRLLKKQCEDQRGKEMAKKVIVAVTDAKKG

::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

EMBOSS LYELTEYLKDKKFGVINISKSGTTTETALAFRLLKKQCEDQRGKEMAKKVIVAVTDAKKG

130 140 150 160 170 180

190 200 210 220 230 240

EMBOSS AARVTADNEGYKSFIIPDNVGGRFSVLTPVGLLPIAVAGFDIEKLVAGAAAMEKTCGKDV

::::::::::::::::::::::::::::::::::::::::::::::::::::::.:::::

EMBOSS AARVTADNEGYKSFIIPDNVGGRFSVLTPVGLLPIAVAGFDIEKLVAGAAAMEKACGKDV

190 200 210 220 230 240

250 260 270 280 290 300

EMBOSS PFAENPAAIYAATRNELYKNGKKIEILVNFNPKLHYVSEWWKQLYGESEGKENKGIFPAS

::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

EMBOSS PFAENPAAIYAATRNELYKNGKKIEILVNFNPKLHYVSEWWKQLYGESEGKENKGIFPAS

250 260 270 280 290 300

310 320 330 340 350 360

EMBOSS VDFSTDLHSMGQWIQEGERTIYETVISVEKTQYSLQVPSDEANLDGLNFLAEKHVDEVNK

::::::::::::::::::::::::::::::::::::::::.:::::::::: ::::::::

EMBOSS VDFSTDLHSMGQWIQEGERTIYETVISVEKTQYSLQVPSDKANLDGLNFLAGKHVDEVNK

310 320 330 340 350 360

370 380 390 400 410 420

EMBOSS MAELGTQLAHVDGGVPNMRIVIPALNEESLGGLLYFFEKACGISGYLLGVNPFNQPGVEA

::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

EMBOSS MAELGTQLAHVDGGVPNMRIVIPALNEESLGGLLYFFEKACGISGYLLGVNPFNQPGVEA

370 380 390 400 410 420

430 440

EMBOSS YKKNMFALLNKPGYEEESKAIQARL

:::::::::::::::::::::::::

EMBOSS YKKNMFALLNKPGYEEESKAIQARL

430 440

Global alignment:

########################################

# Program: needle

# Rundate: Fri 28 Feb 2020 13:03:17

# Commandline: needle

# -auto

# -stdout

# -asequence emboss\_needle-I20200228-130315-0091-56609162-p1m.asequence

# -bsequence emboss\_needle-I20200228-130315-0091-56609162-p1m.bsequence

# -datafile EBLOSUM62

# -gapopen 10.0

# -gapextend 0.5

# -endopen 10.0

# -endextend 0.5

# -aformat3 pair

# -sprotein1

# -sprotein2

# Align\_format: pair

# Report\_file: stdout

########################################

#=======================================

#

# Aligned\_sequences: 2

# 1: EMBOSS\_001

# 2: EMBOSS\_001

# Matrix: EBLOSUM62

# Gap\_penalty: 10.0

# Extend\_penalty: 0.5

#

# Length: 445

# Identity: 441/445 (99.1%)

# Similarity: 442/445 (99.3%)

# Gaps: 0/445 ( 0.0%)

# Score: 2267.0

#

#

#=======================================

EMBOSS\_001 1 MISLNIEKTLGFISKENVFAYEAQVKAAQEALENGTGKGNDFLGWLHLPS 50

||||||||||||||||||||||||||||||||||||||||||||||||||

EMBOSS\_001 1 MISLNIEKTLGFISKENVFAYEAQVKAAQEALENGTGKGNDFLGWLHLPS 50

EMBOSS\_001 51 SITKEHLADLKATAQVLRENCEAVVVAGIGGSYLGARAVIEALSNSFTWL 100

|||||||||||||||||||||||||||||||||||||||||||||||.||

EMBOSS\_001 51 SITKEHLADLKATAQVLRENCEAVVVAGIGGSYLGARAVIEALSNSFAWL 100

EMBOSS\_001 101 QDKKSGPVILYAGHNIGEDYLYELTEYLKDKKFGVINISKSGTTTETALA 150

||||||||||||||||||||||||||||||||||||||||||||||||||

EMBOSS\_001 101 QDKKSGPVILYAGHNIGEDYLYELTEYLKDKKFGVINISKSGTTTETALA 150

EMBOSS\_001 151 FRLLKKQCEDQRGKEMAKKVIVAVTDAKKGAARVTADNEGYKSFIIPDNV 200

||||||||||||||||||||||||||||||||||||||||||||||||||

EMBOSS\_001 151 FRLLKKQCEDQRGKEMAKKVIVAVTDAKKGAARVTADNEGYKSFIIPDNV 200

EMBOSS\_001 201 GGRFSVLTPVGLLPIAVAGFDIEKLVAGAAAMEKTCGKDVPFAENPAAIY 250

||||||||||||||||||||||||||||||||||.|||||||||||||||

EMBOSS\_001 201 GGRFSVLTPVGLLPIAVAGFDIEKLVAGAAAMEKACGKDVPFAENPAAIY 250

EMBOSS\_001 251 AATRNELYKNGKKIEILVNFNPKLHYVSEWWKQLYGESEGKENKGIFPAS 300

||||||||||||||||||||||||||||||||||||||||||||||||||

EMBOSS\_001 251 AATRNELYKNGKKIEILVNFNPKLHYVSEWWKQLYGESEGKENKGIFPAS 300

EMBOSS\_001 301 VDFSTDLHSMGQWIQEGERTIYETVISVEKTQYSLQVPSDEANLDGLNFL 350

||||||||||||||||||||||||||||||||||||||||:|||||||||

EMBOSS\_001 301 VDFSTDLHSMGQWIQEGERTIYETVISVEKTQYSLQVPSDKANLDGLNFL 350

EMBOSS\_001 351 AEKHVDEVNKMAELGTQLAHVDGGVPNMRIVIPALNEESLGGLLYFFEKA 400

|.||||||||||||||||||||||||||||||||||||||||||||||||

EMBOSS\_001 351 AGKHVDEVNKMAELGTQLAHVDGGVPNMRIVIPALNEESLGGLLYFFEKA 400

EMBOSS\_001 401 CGISGYLLGVNPFNQPGVEAYKKNMFALLNKPGYEEESKAIQARL 445

|||||||||||||||||||||||||||||||||||||||||||||

EMBOSS\_001 401 CGISGYLLGVNPFNQPGVEAYKKNMFALLNKPGYEEESKAIQARL 445

#---------------------------------------

#---------------------------------------

**Множественото подравняване**

**Множественото подреждане на последователността (MSA) обикновено е изравняване на три или повече биологични последователности (протеин или нуклеинова киселина) с подобна дължина. От изхода може да се направи извод за хомология и еволюционните връзки между изследваните последователности.**

**Множество подравняване на последователности (MSA) е подравняване на последователността от три или повече биологични последователности, обикновено протеин, ДНК или РНК. В много случаи се приема, че входният набор от последователности на заявки има еволюционна връзка, чрез която те споделят връзка и са произхождащи от общ прародител. От получената MSA може да се направи хомология на последователността и да се проведе филогенетичен анализ, за ​​да се оцени споделеният еволюционен произход на секвенциите. Визуалните изображения на подравняването, както на изображението вдясно, илюстрират мутационни събития, като точкови мутации (единични аминокиселини или нуклеотидни промени), които се появяват като различни знаци в една колона за подравняване и мутации за вмъкване или изтриване (индели или пропуски), които се появяват като тирета в една или повече последователности в подравняването. Многократно подравняване на последователности често се използва за оценка на запазването на последователността на протеинови домени, третични и вторични структури и дори отделни аминокиселини или нуклеотиди.**

**Множеството подравняване на последователности също се отнася до процеса на подравняване на такъв набор от последователности. Тъй като три или повече последователности с биологично значима дължина могат да бъдат трудни и почти винаги отнемат време за подравняване на ръка, изчислителните алгоритми се използват за създаване и анализ на подравняванията. MSA изискват по-сложни методологии, отколкото двойно подреждане, защото са по-сложно изчислително. Повечето множество програми за подравняване на последователности използват евристични методи, а не глобална оптимизация, тъй като идентифицирането на оптималното подравняване между повече от няколко последователности с умерена дължина е изчислително скъпо.**

<https://www.ebi.ac.uk/Tools/msa/>

<https://www.ebi.ac.uk/Tools/msa/clustalo/>

<https://www.genome.jp/tools-bin/clustalw>

Structural alignment attempts to establish homology between two or more polymer structures based on their shape and three-dimensional conformation. This process is usually applied to protein tertiary structures but can also be used for large RNA molecules. In contrast to simple structural superposition, where at least some equivalent residues of the two structures are known, structural alignment requires no a priori knowledge of equivalent positions. Structural alignment is a valuable tool for the comparison of proteins with low sequence similarity, where evolutionary relationships between proteins cannot be easily detected by standard sequence alignment techniques. Structural alignment can therefore be used to imply evolutionary relationships between proteins that share very little common sequence. However, caution should be used in using the results as evidence for shared evolutionary ancestry because of the possible confounding effects of convergent evolution by which multiple unrelated amino acid sequences converge on a common tertiary structure.

**Structural alignments**

Структурните подравнения могат да сравняват две или няколко последователности. Тъй като тези подравнявания разчитат на информация за триизмерните конформации на всички заявки, методът може да се използва само в последователности, където тези структури са известни. Те обикновено се откриват чрез рентгенова кристалография или ЯМР спектроскопия. Възможно е да се извърши структурно изравняване на структури, произведени чрез методи за предсказване на структурата. Всъщност оценката на такива прогнози често изисква структурно изравняване между модела и истинската известна структура, за да се оцени качеството на модела. Структурните подреждания са особено полезни при анализиране на данни от структурните геномични и протеомични усилия и те могат да се използват като сравнителни точки за оценка на подравняванията, произведени чрез чисто базирани на последователността методи на биоинформатика. [1]

Изходите от структурно подравняване са суперпозиция на атомните координатни набори и минимално средно квадратно отклонение на корен (RMSD) между структурите. RMSD на две подравнени структури показва тяхното разминаване една от друга. Структурното подравняване може да бъде усложнено от съществуването на множество протеинови домейни в една или повече от входните структури, тъй като промените в относителната ориентация на домейните между две структури, които трябва да бъдат подравнени, могат изкуствено да надуят RMSD..

<https://www.rcsb.org/pages/analyze_features#Sequence>

<https://en.wikipedia.org/wiki/Structural_alignment_software>

Sequence motif

Когато мотивът за последователност се появи в екзона на ген, той може да кодира "структурен мотив" на протеин; това е стереотипен елемент от общата структура на протеина. Независимо от това, мотивите не трябва да се свързват с отличителна вторична структура. Поредиците "некодиращи" не се превеждат в протеини и нуклеиновите киселини с такива мотиви не трябва да се отклоняват от типичната форма (напр. ДНК двойна спирала с "B-форма").

Извън генните екзони съществуват мотиви на регулаторна последователност и мотиви в рамките на "боклуците", като например сателитна ДНК. Смята се, че някои от тях влияят на формата на нуклеиновите киселини (вижте например РНК самосплискване), но това е понякога само така. Например много ДНК свързващи протеини, които имат афинитет към специфични ДНК свързващи места, свързват ДНК само в двойно-спиралната му форма. Те са в състояние да разпознаят мотиви чрез контакт с основния или втория канал на двойната спирала.

Кратките кодиращи мотиви, които изглежда нямат вторична структура, включват тези, които маркират протеини за доставка до определени части на клетката или ги маркират за фосфорилиране.

В рамките на последователност или база данни от поредици, изследователите търсят и намират мотиви, използвайки компютърно базирани техники за анализ на последователности, като BLAST. Такива техники принадлежат към дисциплината биоинформатика.

<https://prosite.expasy.org/scanprosite/>

https://www.genome.jp/tools/motif/

3D:

