R3D Align User's Manual

Table of Contents

Installation

R3D Align was written in Matlab version 7.7 and has been incorporated into the **FR3D** suite of Matlab programs. For more information on FR3D, go to http://rna.bgsu.edu/FR3D. The easiest way to install **R3D Align** is to download the latest version of the zip file from http://r3dalign.googlecode.com. This will download a folder named **FR3D** and several subfolders. The **R3D Align** subfolder contains the files that are specific to its use. The first time **R3D Align** is asked to align a given PDB file, it reads the text, analyzes it, and saves a data file in the subfolder **PrecomputedData**. After that, it will not need to re-read the original PDB file. **PrecomputedData** contains data for four large PDB files upon download. The subfolder **PDBfiles** is where the user should place the PDB files of the structures to be aligned by **R3D Align**. See below if you have another folder on the computer with the PDB files to be used.

Initialization

Start Matlab and change the working directory to **FR3D** (you can use the cd command to change the directory). If you have another folder on your computer with PDB files, add that folder to Matlab's path (File, Set Path, Add Folder, Save).

Calling R3DAlign.m

To align two 3D structures, the user should call the function contained in **R3DAlign.m**, which has the following syntax:

```
[AlignedNTs1,AlignedNTs2] = R3DAlign(File1,NTList1,File2,NTList2,discCut,numNeigh,
                                     bandwidth,cliqueMethod,seed1,seed2)
```

The function R3DAlign accepts 9 parameters, of which the first 7 are required. A short description and the proper syntax for each is provided next.

Input Parameters

The function R3DAlign accepts 9 parameters, of which the first 7 are required. A short description and the proper syntax for each is provided.

File1 is the pdb id for the first RNA molecule. The parameter must be entered as a string variable, so the PDB id is to be placed in quotes (eg. `'1s72'`)

NTList1 specifies which nucleotides of the first RNA molecule are to be aligned. The user can specify the nucleotides in a variety of ways:

- To align all the nucleotides contained in File1, simply enter the string `'all'` for this parameter.
- To align all nucleotides contained in a specific chain of File1, enter the chain id place in quotes (eg. `'A'`).
- To align a specific set of nucleotides, a string of nucleotide numbers separated by commas can be entered. Any entry of the string can use the colon notation for a range of nucleotide numbers (eg. `121:125`). Nucleotide numbers can be followed by `(A)` or `_A` to indicate chain A. For example to align nucleotides 1 to 5, 7, and 8-10 of chain B in File1, enter `'1:5(B), 7(B), 8:10(B)'`. The quotes must be entered.

File2 is the PDB id for the second molecule.

NTList2 specifies which nucleotides of the second RNA molecule are to be aligned. The same options exist for *NTList2* as for *NTList1*. *NTList2* does not need to contain the same number of nucleotides as *NTList1*.

discCut is the geometric discrepancy cutoff value for which two neighborhoods are considered to be structurally similar. *discCut* is a positive real-valued number (eg. `0.5`). If discCut = 0.5, then two sets of nucleotides (neighborhoods) with a geometric discrepancy of less than or equal 0.5 are considered to be similar and any geometric discrepancy above 0.5 is implies dissimilarity.

numNeigh is positive, integer-valued, and gives the number of neighborhoods to retain for each nucleotide. If *numNeigh* = 7, then the 7 four-nucleotide neighborhoods with the smallest diameter will be retained for each nucleotide.

bandwidth is also positive and integer-valued. If *bandwidth* = *b*, then nucleotide *i* of the first structure will be compared with nucleotide *k* only if $j-b < k < j+b$, where *j* is the nucleotide to which *i* is aligned in the seed alignment.

cliqueMethod specifies whether the alignment will be produced by finding a clique using the greedy clique procedure or the full branch-and-bound maximum clique procedure. *cliqueMethod* is a string variable with two possible values: 1) `'greedy'` and 2) `'full'`.

seed1 and *seed2* are optional parameters which can be used to enter a seed alignment. They are cell arrays of the same size containing the nucleotides numbers aligned nucleotides in corresponding positions. That is, for each *i*, `seed1{i,1}` contains the number of the nucleotide in File1 that is aligned with the nucleotide in File 2 with the number contained in `seed2{i,1}`. The chain for each nucleotide can be identified in `seed1{i,2}` for each *i*. If *seed1* and *seed2* are omitted, then R3DAlign will internally compute a seed alignment using a sequence alignment algorithm. The output arguments described next can directly be used as *seed1* and *seed2*.

Output Arguments

The function returns two arguments:

AlignedNTs1 is an M x 3 cell array, where M is the number of nucleotides aligned. The first column of the cell array contains the numbers of the aligned nucleotides. The second column lists the chain for each aligned nucleotide. The third column contains the base type of each aligned nucleotide. Nucleotides in File1 aligned with a gap are not included in the cell array.

AlignedNTs2 is also an M x 3 cell array where the nucleotide represented in *AlignedNTs2{j}* is aligned with the nucleotide represented in *AlignedNTs1{j}*

AlignedNTs1 and *AlignedNTs2* can be used in a subsequent call to R3DAlign.m as *seed1* and *seed2*.

Output Files

R3DAlign.m also produces three additional files, each containing the alignment information in different formats. Each of the three files is produced automatically and saved in the current working directory. These three files include:

1) A FASTA file containing the alignment

2) An EXCEL spreadsheet that simultaneously displays the alignment and basepairing interactions of the structures
3) A 'bar diagram' that displays the alignment and indicates the structural similarity of aligned local neighborhoods. This 'bar diagram' is saved as a PDF file.

All three files are automatically saved with the filename in the following format:
"R3D Align Name_of_File1 Name_of File2 Current_Date Current_Time"

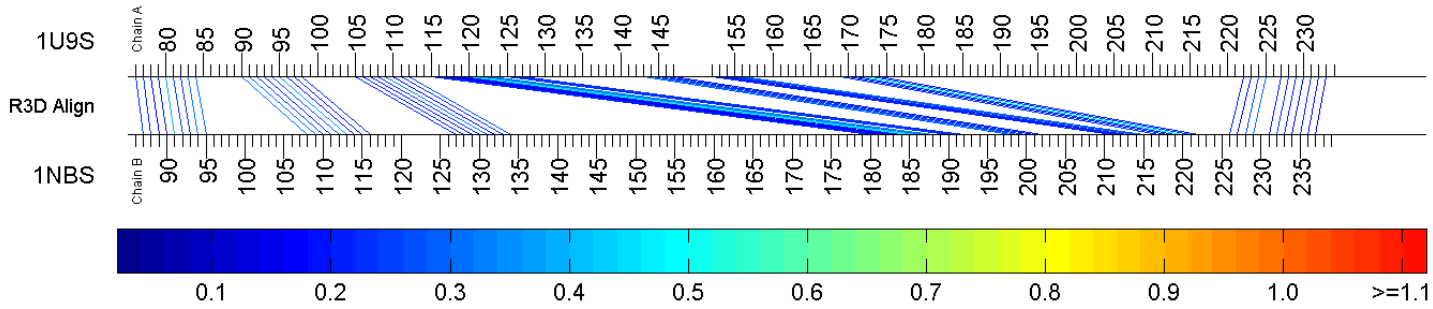
The FASTA file is saved with the extension '.FASTA', the Excel file with '.xlsx', and the bar diagram with '.pdf'.

Examples

Example 1: Aligning all the nucleotides of 1U9S with chain B of 1NBS; a discrepancy cutoff of 0.3 is to be used; 12 neighborhoods are to be retained for each nucleotide; a bandwidth of 60 is to be used on the internally computed seed alignment; the full maximum clique algorithm will be employed:

```
[AlNTs1u9s, AlNTs1nbs] = R3DAlign('1u9s','all', '1nbs','B',0.3,12,60,'full');
```

The produced bar diagram looks like the following:



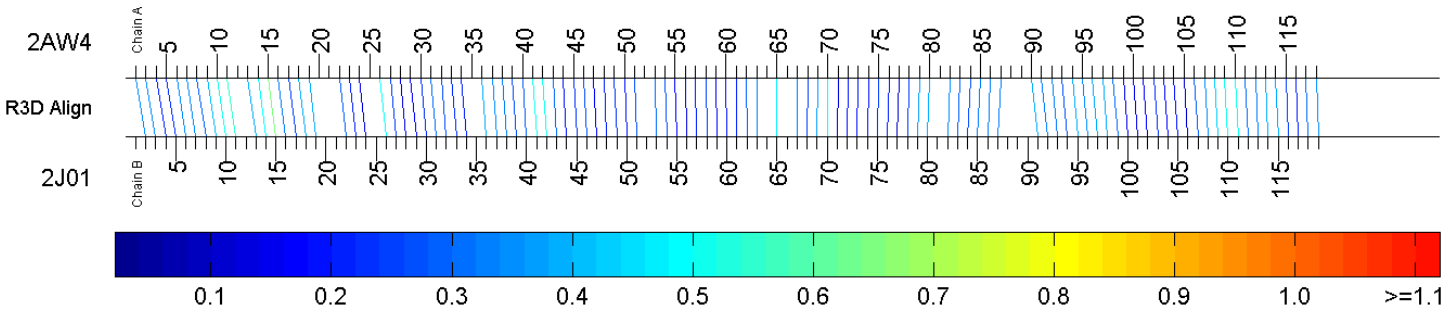
The FASTA file contains the following:

```
> 1U9S
-GGUGCCAGGUAACG-----------CCUGGGCGGGGUAAC------CCGACGGAAA----------------------
----------------------GUGCCACAGA-GAAGAGACCGCCAGCGGC-----CGGGG--------CCCCGGUGCG
GGCAAGG-GUGAAACGGCGGGGUAAGAGCCCACCGCCUGGCCUGGCAACAGGCCGGGGCAC---GGCAA-ACCCCACC--
> 1NBS B
GCGAGCCUAG-----CGAAGUCAUAAGCUAGGGCAG------UCUUGGCUGACGGC--AGGAAAAAAGCCUACGUCUUCG
GAUAUGGCUGAGUAUCCUUGAAAGUGCCACAG-UGAC--------------GAAGUCUCACUAGAAAUGGUGAG-----
-------AGUGGAA-------------------------------------------------CGCGGUA-AACCCCUC-GC
```

Example 2: Aligning the 5S rRNAs of E.Coli (chain A of PDB 2AW4) and T.Thermophilus (chain B of PDB 2J01); discrepancy cutoff 0.5; 7 neighborhoods; bandwidth = 50; full clique procedure used for final alignment; no seed alignment manually entered:

```
[AlNTs2aw4,AlNTs2j01] = R3DAlign('2aw4','A','2j01','B',0.5,7,50,'full');
```

Bar diagram:



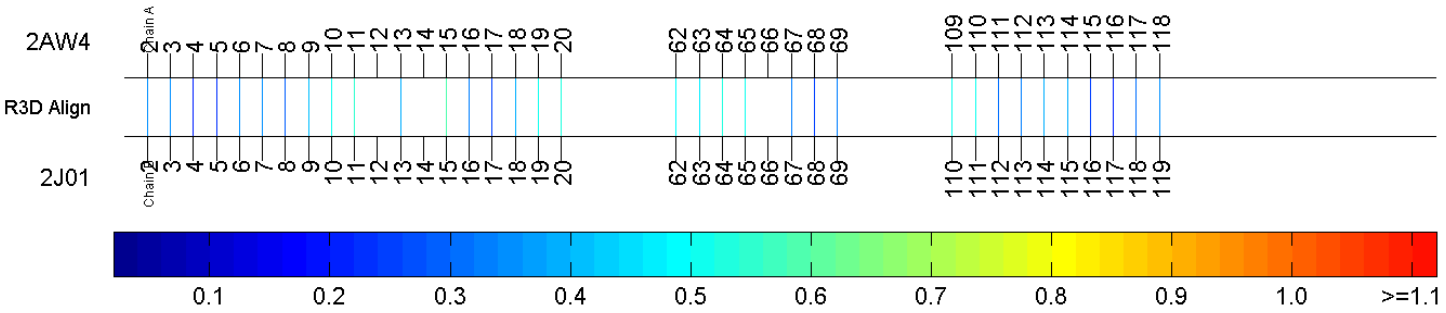
FASTA output:

```
> 2AW4 A
-GCCUGGCGGCC-GUAGCGCGG--UGGU-CCCACCUGAC-CCCAUGCCGAACUCAGA-AGUGAAACGCCG-UA-GCGCCG
AUGGUAGUG-UGGGGUCU---CCCCAUGCGAGAGUAGGGAACUGCCAGGC
> 2J01 B
UCCCCCGUGCC-CAUAGCGG--CGUGG-AACCACCCGU-UCCCAUUCCGAACACGG-AAGUGAAACGCG-CC-AGCGCCG
AUGGUACU-GGGCGGG--CGACCGCCUGGGAGAGUAGGUCGGUGCGGGGG
```

Example 3: Same as Example 2 but only aligning nucleotides 2-20, 62-69, 109-118 of 2AW4 and nucleotides 2-20, 62-69, 110-119 of 2J01:

```
[AlNTs2aw4,AlNTs2j01] = R3DAlign('2aw4','2:20(A),62:69(A),109:118(A)',
                                 '2j01','2:20(B),62:69(B),110:119(B)',
                                 0.5,7,50,'full');
```

Bar diagram:



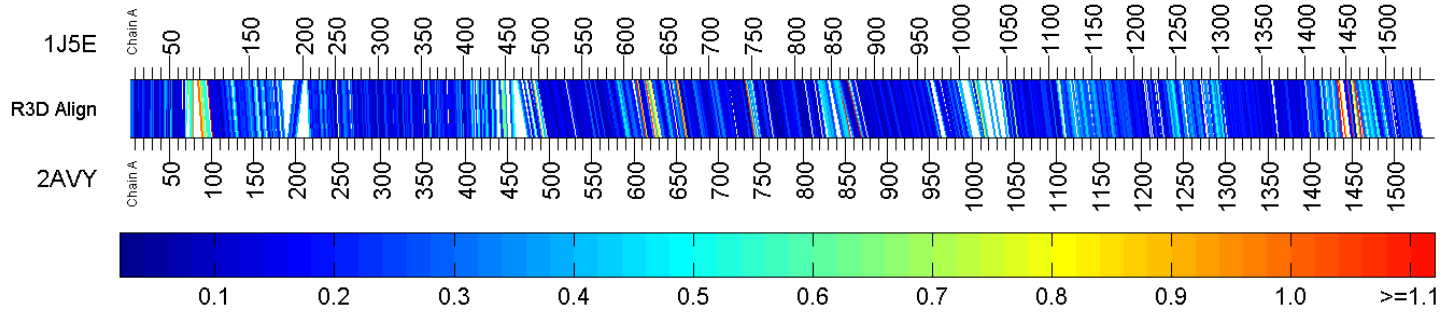
FASTA output:

```
> 2AW4 2:20(A),62:69(A),109:118(A)
GCCUGGCGGCC-GU-AGCGCGCCGUA-GCGACUGCCAGGC
> 2J01 2:20(B),62:69(B),110:119(B)
CCCCCGUGCC-CA-UAGCGGCCGCC-AGCGGGUGCGGGGG
```

Example 4: Align all nucleotides in 1J5E and all nucleotides in 2AVY; discrepancy cutoff=0.5; 3 neighborhoods; bandwidth=30; greedy clique procedure used for final alignment; no seed alignment manually entered

```
[AlNTs1j5e1, AlNTs2avy1] = R3DAlign('1j5e', 'all', '2avy', 'all',0.5,3,30,
                                    'greedy');
```

Bar diagram:



FASTA output:

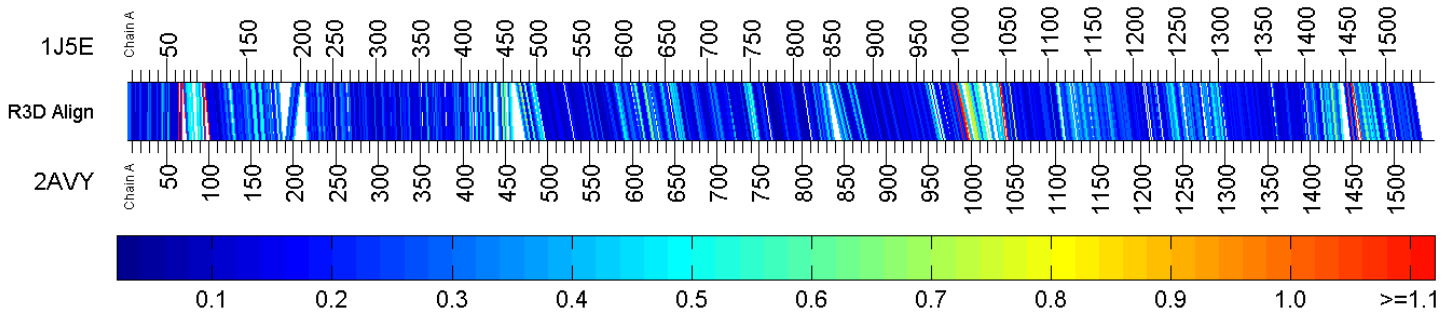
```
> 1J5E
UGGAGAGUUUGAUCCUGGCUCAGGGUGAACGCUGGCGGCGUGCCUAAGACAUGCAAGUCGU-GCG-----GGCCGCGGG-
GUUUUA---CUCCGUGG--UC----AGCGGCGGACGGGUGAGUAAC-GCGUGGGUGACCUACCCGGAAGAGGGGGACAAC
CCGGGGAAACUCGGGCUAAUCCCC-CAUGUG--GACCCGCCCCUUGGGGUGU---GUCCAAAGGGCUUU----------
---GCCCGCUUCCGGAUGGGCCCGCGUCCCAUCAGCUAG--UUGGUGGGGUAAUGG-CCCACCAAGGCGACGACGGGUAG
CCGGUCUGAGAGGAUGGCCGGCCACAGGGGCACUGAGACACGGGCCCCACUCCUACGGGAGGCA-GCAGUUAGGAAUCUU
CCGCAAUGGGCGC-AAGCCUGACGGAGCGACGCCGCUUGGAGGAA-GAA-GCCCUUCGGGGUGUAAACUCCUGAA---CC
CGGGACGAAAC--CCCCGACGA--------------GGGGACUGACGG-UACCGGGG-UAAU--AGCGCCGGCCAACUCC
GUGCCAGCAGCCGCGGUA-AUACGGAGGGCGCGAGCGUUACCCGGAUUCACUGGGCGUAAAGGGCGUGUAGGCGGCCUGG
GGCGUCCCAUGUGAAAG-ACCACGGCUCAACCGUGGGGGA-GCGUGGGAUACGCUCAGGCUA-GACGGUGGGAGAGGGUG
GUGGAAUUCCCGGAGUAGCGGUGAAAUGCGCAGAUACCGGGAGGAACGCCGAUGGCGAAGGCAGCCACCUGGUCCA-CCC
GUGACGCUGAGGCGCG-AAAGCGUGGGGAGCAAACCGGAUUAGAUACCCGGGUAGUCCACGCCC-UAAACGAUGCGCGCU
AGGUCUCUGGGUCU---------CCUGGGGGCCGAAGCUAACGCGUUAAGCGCGCCGCCUGGGGAGUACGGCCGCAAGGC
UGAAACUCAAAGGAAUUGACGGGGGCCCGCACAAGCGGUGGAGCAUGUGGUUUAAUUCGAAG-CAA--CGCGAAGA-ACC
UUACCAGGCCUUGACAUG-CUAGGGAA--CCCGGGUGAAAGCCUG--GGGUGC--CCCGCGA----GGGGAGCCCUAGC-
-ACAGGUGCUGCAUGGCCGUCGUCAGCUCGUGCCGUGAGGUGUUGGGUUAAGUCCCGCAACGAGCGCAA-CCCCCGCCGU
UAGU--UGCCAGCGGUUCGGCCGGGC-ACUCUAACGGGACUGCCCGCGA-AAGCGGGAGGAAGGAGGGGACGACGUCUGG
UCAGCAUGGCCCUU-AC-GGCCUGGGCGACA-CACGUGCUACAAUGCCCACUACAAAGCGAU--GCCACCCGGCAACGGG
GAGCUA--AUCGCAA-AAAGGUGGGCCCAGUU-CGGAUUGGGGUCUGCAACCC-GACCCCAUGAAGCCGGAAUCGCUAGU
AAUCGCGGAUCAGCCAUGCCGCGGUGAAUACGUUCCCGGGCCUUGUACACACC-GCCCGUCACGCCAUGGGAGCGGGCUC
UACCCGAAGUCGCCGGGAG-------CCUACGGGCA---GGCGCCGAGGGUAGGGCCCGUGACUGGGGCGAA-GUCGUAA
CA-AGGUAGCUGUACCGGAAGGUGCGGCUGGAUCACUUUCU--
> 2AVY
UGAAGAGUUUGAUCAUGGCUCAGAUUGAACGCUGGCGGCAGGCCUAACACAUGCAAGUCG-AACGGUAACAGG-AAGA-A
GC----UUGCUUCUU--UGCUGACGAGUGGCGGACGGGUGAGUAA-UGUCUGGGA-AACUGCCUGAUGGAGGGGGAUAAC
UACUGGAAACGGUAGCUAAUACC-GCAUA--ACGUCG--------------CAAGACCAAAGAGG---GGGACCUUCGG
GCCUCUUGCCAUCGGAUGUGCCCAGAUGGGAUUAGCU--AGUAGGUGGGGUAACG-GCUCACCUAGGCGACGAUCCCUAG
CUGGUCUGAGAGGAUGACCAGCCACACUGGAACUGAGACACGGUCCAGACUCCUACGGGAGGC-AGCAGUGGGGAAUAUU
GCACAAUGGGCG-CAAGCCUGAUGCAGCCAUGCCGCGUGUAUGA-AGA-AGGCCUUCGGGUUGUAAAGUACUU--UCAGC
GGGGAGGAA--GGGAGUA----AAGUUAAUACCUUUGCUCAUUGACG-UUACCCGC-AGAA-GAAGCACCGGCUAACUCC
GUGCCAGCAGCCGCGGU-AAUACGGAGGGUGCAAGCGUUAAUCGGAAUUACUGGGCGUAAAGCGCACGCAGGCGGUUUGU
UAAGUCAGAUGUGAAA-UCCCCGGGCUCAACCUGGGAAC-UGCAUCUGAUACUGGCAAGCU-UGAGUCUCGUAGAGGGGG
GUAGAAUUCCAGGUGUAGCGGUGAAAUGCGUAGAGAUCUGGAGGAAUACCGGUGGCGAAGGCGGCCCCCUGGACGAAG-A
```

```
CUGACGCUCAGGUGC-GAAAGCGUGGGGAGCAAACAGGAUUAGAUACCCUGGUAGUCCACGCC-GUAAACGAUGUCGACU
UGGAGGUUGUG---CCCUUGAGGCGUGGCUUCCGGAGCUAACGCGUUAAGUCGACCGCCUGGGGAGUACGGCCGCAAGGU
UAAAACUCAAAUGAAUUGACGGGGGCCCGCACAAGCGGUGGAGCAUGUGGUUUAAUUCGAU-GC--AACGCGAAG-AACC
UUACCUGGUCUUGACAUCCACGG----AAGUUUUCAGAGAUGAG-AAU-----GUGCC----UUCGGGAA----CCGUGA
GACAGGUGCUGCAUGGCUGUCGUCAGCUCGUGUUGUGAAAUGUUGGGUUAAGUCCCGCAACGAGCGCA-ACCCUUAUCCU
UUG-UUGCC-AGCGGUCCGGCCGGG-AACUCAAAGGAGACUGCCAGUGAUAAACUGGAGGAAGGUGGGGAUGACGUCAAG
UCAUCAUGGCCCU-UA-CGACCAGGGCUAC-ACACGUGCUACAAUGGCGCAUACAAAGAG--AAGCGACCUCGCGAGAGC
AAGC--GGACCUCA-UAAAGUGCGUCGUAGU-CCGGAUUGGAGUCUGCAACU-CGACUCCAUGAAGUCGGAAUCGCUAGU
AAUCGUGGAUCAG-AAUGCCACGGUGAAUACGUUCCCGGGCCUUGUACACAC-CGCCCGUCACACCAUGGGAGUGGGUUG
CAAAAGAAGUAGGU-----AGCUUAACCUUCGGGAGGGCGCUUACCACUUUGUGAUUCAUGACUGGGGUG-AAGUCGUAA
C-AAGGUAACCGUAGGGGAACCUGCGGUUGGAU--------CA
```

Example 5 (entering a seed alignment): Align all nucleotides in 1J5E and all nucleotides in 2AVY; discrepancy cutoff=0.5; 9 neighborhoods; bandwidth=10; greedy clique procedure used for final alignment; use output from Example 4 as the seed alignment:

```
[AlNTs1j5e2, AlNTs2avy2] = R3DAlign('1j5e', 'all', '2avy', 'all',0.5,9,10,
                                    'greedy',AlNTs1j5e1,AlNTs2avy1);
```

Bar diagram:




```
FASTA output:
> 1J5E
UGGAGAGUUUGAUCCUGGCUCAGGGUGAACGCUGGCGGCGUGCCUAAGACAUGCAAGUCGU-GCGGGCC---GCGGG-GU
UUU---------ACUC-C-GUGGUC-AGCGGCGGACGGGUGAGUAAC-GCGUGGGUGACCUACCCGGAAGAGGGGGACAA
CCCGG-GGAAACUC-GGGCUAAUCCCC-CAUGUG--GACCCGCCCCUUGGGGUGU---GUCCAAAGGGCUUU--------
------GCCCGCUUCCGGAUGGGCCCGCGUCCCAUCAGCUAG--UUGGUGGGGUAAUGG-CCCACCAAGGCGACGACGGG
UAGCCGGUCUGAGAGGAUGGCCGGCCACAGGGGCACUGAGACACGGGCCCCACUCCUACGGGAGGCA-GCAGUUAGGAAU
CUUCCGCAAUGGGCGC-AAGCCUGACGGAGCGACGCCGCUUGGAGGAA-GAA-GCCCUUCGGGGUGUAAACUCCUGAA--
--CCCGGGACGAAAC----CCCCGACGA--------------GGGGACU--GACGG--UACCGGGGU--AAU--AGCGCC
GGCCAACUCCGUGCCAGCAGCCGCGGUA-AUACGGAGGGCGCGAGCGUUACCCGGAUUCACUGGGCGUAAAGGGCGUGUA
GGCGGCCUGGGGCGUCCCAUGUGAAAGA-CCAC-GGCUCAACCG-UGGGGGA-GCGUGGGAUACGCUCAGGCUA-GACGG
UGGG-AGAGGGUGGUGGAAUUCCCGGAGUAGCGGUGAAAUGCGCAGAUACCGGGAGGAACGCCGAU-GGCGAAGGCAGCC
ACCUGGU-CC-ACCCGUGACGCUGAGGCGCG-AAAGCGUGGGGAGCAAACCGGAUUAGAUACCCGGGUAGUCCACGCCC-
UAAACGAUGCGC--GCUA---GGUCUCUGGGUCU-----CCUGGGGGCCG--AAGCUAACGCGUU-AAGCGC-GCCGCCU
GGGGAGUACGGCCGCAAGGCUGAAACUCAAAGGAAUUGACGG-GGGCCCGCACAAGCGGUGGAGCAUGUGGUUUAAUUCG
AAGCAA-----CGCGAAGA-ACCUUACCAGGCCUUGACAUG-CUAGGGAACCC---------GGGUGAAAGCCUGGGGUG
CC------CCGCGA----GGGGAGCCCU-----AGCA-CAGGUGCUGCAUGGCCGUCGUCAGCUCGUGCCGUGAGGUGUU
```

GGGUUAAGUCCCGCAACGAGCGCAA-CCCCCGCCGUUAGU--UGCCAGCGGUUCGGCCGGGC-ACUCUAACGGGACUGCC
CGCGA-AAGCGGGAGGAAGGAGGGGACGACGUCU-GGUCAGCAUGGCCCUU-AC-GGCCUGGGCG-ACA-CACGUGCUAC
AAUGCCCACUACAAAGCGAU--GCCACCCGGCAACGGGGAGCUA--AU-CGCAA-AAAGGUGGGCCCAGUU-CGGAUUGG
GGUCUGCAACCC-GACCCCAUGAAGCCGGAAUCGCUAGUAAUCGCGGAUCAGCC-AUGCCGCGGUGAAUACGUUCCCGGG
CCUUGUACAC-ACC-GCCCGUCACGCCAUGGGAGCGGGC-UCUACCCG-AAGUCG---CCGGGAG---CCUACGGGCA--
GGC--GCCGAG-GGUAGGGCCCGUGACUGGGGCGAA-GUCGUAACA-AGG-UAGCUGUACCGGAAGGUGCGGCUGGAUCA
CUUUCU--
> 2AVY
UGAAGAGUUUGAUCAUGGCUCAGAUUGAACGCUGGCGGCAGGCCUAACACAUGCAAGUCG-AACG-GU-AACAGGA-AG-
---AAGCUUGCUUCU-UUGCUGAC-GAGUGGCGGACGGGUGAGUAA-UGUCUGGGA-AACUGCCUGAUGGAGGGGGAUAA
CUAC-UGGAAACG-GUAGCUAAUACC-GCAUA--ACGUCG---------------CAAGACCAAAGAGG---GGGACCUU
CGGGCCUCUUGCCAUCGGAUGUGCCCAGAUGGGAUUAGCU--AGUAGGUGGGGUAACG-GCUCACCUAGGCGACGAUCCC
UAGCUGGUCUGAGAGGAUGACCAGCCACACUGGAACUGAGACACGGUCCAGACUCCUACGGGAGGC-AGCAGUGGGGAAU
AUUGCACAAUGGGCG-CAAGCCUGAUGCAGCCAUGCCGCGUGUAUGA-AGA-AGGCCUUCGGGUUGUAAAGUACUU--UC
AGCGGG-GAGG----AAGGGAGUA----AAGUUAAUACCUUUGCUCA--UUGAC--GUUAC-CCGC-AGAA-GAAGCACC
GGCUAACUCCGUGCCAGCAGCCGCGGU-AAUACGGAGGGUGCAAGCGUUAAUCGGAAUUACUGGGCGUAAAGCGCACGCA
GGCGGUUUGUUAAGUCAGAUGUGAAA--UCCCCGGGCUCAACCUGGGAA-C-UGCAUCUGAUACUGGCAAGCU-UGAG-U
CUCGUAGAGGGGGGUAGAAUUCCAGGUGUAGCGGUGAAAUGCGUAGAGAUCUGGAGGAAUACCGG-UGGCGAAGGCGGCC
CCCUGGACGAAGA-C-UGACGCUCAGGUGC-GAAAGCGUGGGGAGCAAACAGGAUUAGAUACCCUGGUAGUCCACGCC-G
UAAACGAUGU--CGACU-UGGAGGUUGUGCC---CUUGAGGCGUGGCUUCCGGAGCUAACGCGU-UAAGUC-GACCGCCU
GGGGAGUACGGCCGCAAGGUUAAAACUCAAAUGAAUUGACG-GGGGCCCGCACAAGCGGUGGAGCAUGUGGUUUAAUUCG
A-----UGCAACGCGAAG-AACCUUACCUGGUCUUGACA--UCCA---------CGGAAGUUUUCAG-AGA-UGAG-----
--AAUGUGCC----UUCGGGA-------ACCGUGAG-ACAGGUGCUGCAUGGCUGUCGUCAGCUCGUGUUGUGAAAUGUU
GGGUUAAGUCCCGCAACGAGCGCA-ACCCUUAUCCUUUG-UUGCC-AGCGGUCCGGCCGGG-AACUCAAAGGAGACUGCC
AGUGAUAAACUGGAGGAAGGUGGGGAUGACGUC-AAGUCAUCAUGGCCCU-UA-CGACCAGGGC-UAC-ACACGUGCUAC
AAUGGCGCAUACAAAGAG--AAGCGACCUCGCGAGAGCAAGC--GGA-CCUCA-UAAAGUGCGUCGUAGU-CCGGAUUGG
AGUCUGCAACU-CGACUCCAUGAAGUCGGAAUCGCUAGUAAUCGUGGAUCAG--AAUGCCACGGUGAAUACGUUCCCGGG
CCUUGUACA-CAC-CGCCCGUCACACCAUGGGAGUGGG-UUGCAAAA-GAAGUAGGUAGCU----UAACCUUCGGGA-GG
GCGCUUACCA-CUUUGUGAUUCAUGACUGGGGUGA-AGUCGUAAC-AAG-GUAACCGUAGGGGAACCUGCGGUUGGAU--
------CA