

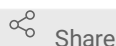


Aug 23, 2022

Protocol for In-silico Design, Docking and Molecular Dynamic Simulation of Antisense Oligonucleotides

Akshay Uttarkar¹, Mansi Babu², Vidya Niranjana¹¹Department of Biotechnology, R V College of Engineering;²Research Intern, Centre of Excellence Computational Genomics, R V College of Engineering

2 Works for me



Share

dx.doi.org/10.17504/protocols.io.ewov1nr4ogr2/v1

Centre of Excellence in Computational Genomics (Vidya Lab)

Vidya Niranjana

ABSTRACT

In-silico drug design has been a catalyst in developing novel chemicals, Phyto-actives, nanoparticles, and anti-sense oligonucleotides. A robust screening setup increases the chances of success of every project. In recent years, anti-sense oligonucleotides (ASO) have provided a major breakthrough in developing 10 ASO drugs that are FDA-approved. This prompts us to develop a protocol for ASO design, docking, and MD simulation studies. The protocol has been divided into 6 sub-sections inclusive of 30 steps. The tools used in the protocol are either open-source or academic free version. The protocol is developed and validated on GPU workstation using NVIDIA A100

DOI

dx.doi.org/10.17504/protocols.io.ewov1nr4ogr2/v1

PROTOCOL CITATION

Akshay Uttarkar, Mansi Babu, Vidya Niranjana 2022. Protocol for In-silico Design, Docking and Molecular Dynamic Simulation of Antisense Oligonucleotides.

protocols.io<https://protocols.io/view/protocol-for-in-silico-design-docking-and-molecular-dynamic-simulation-of-antisense-oligonucleotides>

KEYWORDS

ASO design, mRNA docking, MD simulation

LICENSE

_____ This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

CREATED

Aug 02, 2022

LAST MODIFIED

Aug 23, 2022

PROTOCOL INTEGER ID

68045


1 Retrieve the whole mRNA sequence of target gene from GenBank.

URL: <http://www.ncbi.nlm.nih.gov/nucleotide/>

a. Open the home page and search with keyword i.e. the name of the gene.

The screenshot shows the NIH Nucleotide search interface. The search term 'human ccr5' is entered in the search bar. The results page displays a list of items, with the first item being 'Human CC chemokine receptor 5 (CCR5) mRNA, complete cds'. The page also shows a 'RefSeq Sequences' section with a '+' button. The 'Results by taxon' section lists top organisms, including Homo sapiens (71), Mus musculus (11), Rattus norvegicus (8), Macaca mulatta (5), Callithrix jacchus (3), and All other taxa (12). The 'Search details' section shows the search query: '(["Homo sapiens"[Organism] OR human[All Fields]) AND ccr5[All Fields]) AND biomol_mrna[PROP]'. The 'Recent activity' section shows recent searches, including 'human ccr5 AND (biomol_mrna[PROP]) (110)' and 'human ccr5 (11045)'.

- f. After examining these records, check for the longest, most complete mRNA reference sequence for the gene.



National Library of Medicine
National Center for Biotechnology Information

Log in

Nucleotide

Advanced

GenBank
Send to:

Change region shown

Customize view

Homo sapiens C-C motif chemokine receptor 5 (CCR5), transcript variant A, mRNA

NCBI Reference Sequence: NM_000579.4

[FASTA](#)
[Graphics](#)

[Go to:](#)

LOCUS	NM_000579	3661 bp	mRNA	linear	PRI 31-JUL-2022
DEFINITION	Homo sapiens C-C motif chemokine receptor 5 (CCR5), transcript variant A, mRNA.				
ACCESSION	NM_000579	NM_001105536	XM_001125981	XM_005264854	
VERSION	NM_000579.4				
KEYWORDS	RefSeq.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 3661) Hombouisse F, Nardi G, Colin P, Hery M, Cordeiro N, Blachier S, Schwartz O, Arenzana-Seisdedos F, Sauvonnnet N, Olivo-Marin JC, Lagane B, Lagache T and Brelot A. Tracking receptor motions at the plasma membrane reveals distinct effects of ligands on CCR5 dynamics depending on its dimerization status Elife 11, e76281 (2022) JOURNAL PUBLISHED 35866628				
REMARK	GeneRIF: Tracking receptor motions at the plasma membrane reveals distinct effects of ligands on CCR5 dynamics depending on its dimerization status. Publication Status: Online-Only				
REFERENCE	2 (bases 1 to 3661) van Eekeren LE, Matzaraki V, Zhang Z, van de Wijer L, Blaauw MJT, de Jonge MI, Vandekerckhove L, Trypsteen H, Joosten LAB, Netea MG, de Mast Q, Koenen HJPM, Li Y and van der Ven AJAM. People with HIV have higher percentages of circulating CCR5+ CD8+ T cells and lower percentages of CCR5+ regulatory T cells Sci Rep 12 (1), 11425 (2022) JOURNAL PUBLISHED 35794176				

Analyze this sequence
Run BLAST
Pick Primers
Highlight Sequence Features
Find in this Sequence
Show in Genome Data Viewer

Articles about the CCR5 gene
CCR5 activation and endocytosis in circulating tumor-derived cells [Breast Cancer Res. 2022]
Tumor bud-derived CCL5 recruits fibroblasts and promotes colorectal [J Exp Clin Cancer Res. 2022]
CCR5 as a Coreceptor for Human Immunodeficiency Virus an [Front Immunol. 2022]
See all...

Reference sequence information

RefSeq alternative splicing
See 3 reference mRNA sequence splice variants for the CCR5 gene.
RefSeq protein product
See the reference protein sequence for C-C chemokine receptor type 5 (NP_000570.1).

More about the CCR5 gene

- g.Retrieve the mRNA sequence in FASTA format.

National Library of Medicine
National Center for Biotechnology Information

Log in

Nucleotide [Advanced](#) [Help](#)

FASTA

Homo sapiens C-C motif chemokine receptor 5 (CCR5), transcript variant A, mRNA

NCBI Reference Sequence: NM_000579.4

[GenBank](#) [Graphics](#)

>NM_000579.4 Homo sapiens C-C motif chemokine receptor 5 (CCR5), transcript variant A, mRNA

```

CTTCAGATAGATTATCTGGAGTGAAGATCTGCCACCTATGTATCTGGCATAGTATCTGTAGTG
GGATGAGCAGAGAACAAAAACAAATATCCAGTGAAGAAAGCCGTAAATAACCTTCAGACAGAGAT
CTATCTCTAGCTATTTTAAGCTCAACTTAAAAAGAGAGACTGTCTCTGATCTTTTCGCTTCAATA
CACTTAATGATTTAACTCCACCTCTCTCAAAAAGAGAGACTTTCCCTACTTTTATAGCTCTATAGT
TGATTTGACAGCTCATCTGGCCAGAGAGCTGAGACATCCGTCCCTACAGAAACTCTCCCGGGTG
GAACAGATGGATTACAAGTGTCAAGTCAATCTATGACATCAATTATTATACATCGAGCCCTGCCAA
AAATCAATGTGAAGCAATCGAGCCCGCTCTGCTCCGCTCACTACTGCTGTTCTCTTTGGTT
TTGTGGGCACTGCTGGTCACTCTCATCTGATAAAGTGAAGAGCTGGAAGCATGACTGACATCTA
CTGCTCAACTGGCCATCTGACCTGTTTTCCTCTTACTGTCCTCTGCGCTCACTACTGCTGCTG
GCCAGTGGGACTTTGGAAATACAATGTGTCAACTCTTGAAGGGCTCTATTTATAGGCTCTCTCTG
GAATCTTCTCATCTCTCTGACATCGATAGGTACCTGGCTGCTCCATGCTGTTGTTGCTTAAAG
AGCCAGGACGCTCACCTTTGGGGTGGTGAAGGTGTGATCACTTGGGTGGTGGCTGTTTGGCTCTC
CCAGGAATCATCTTTACAGACTCTCAAAAGAGGTCTTCACTACACCTGCAAGCTCTCACTTTCCATACA
GTCAATCATCTTGAAGAAATTTGAGAGATTAAGATAGTCACTTGGGGCTGGCTCTGCGCTGCT
TGTCAATGCTCATCTGCTGCTGAGGATCCAAAACTCTGCTGCGTGTGAAGTGAAGAGAGGAC
AGGGCTGTGAGGCTTATCTCACCATCATGATTGTTATTTCTCTCTGGGCTCCCTACCAAGTGTCC
TTCTCTGACACCTCTCAGGAATCTTTGGGCTGAATATGCAAGTCTCAACAGGTGGAGCAAGC
TATGACAGTGAACAGACTCTTGGGATGACGACTGCTGATCAACCCCATCATCTATGCTTTGTGGG
GAGAGTTGAGAACTACCTTATGCTCTCTTCCAAAGACATTCGCAAGCTCTGCAATGCTGTT
CTATTTCCAGCAAGAGGCTCCCGAGGAGCAAGTCAAGTTTACCCGATCACTGGGAGCAGGAAT
ATCTGTGGGCTTGTGACAGGACTCAAGTGGGCTGGTGAAGGCTCTTTTAAAGGAAGTACTGTATAGAGG
TCTAAGATTCATCCTATTTATGATCTGTTTAAAGTATAGTATCTTTAAAGCCATCAATATAGA
AAGCAAAATCAAAATATGTTGATGAAAAATAGCAACTTTTATCTCCCTTCACTGATCAAGTATT
GACAACTCTCCCTCACTCGAAAGTCTCTATGATATTTAAAGAAAGCTCAGAGAAATGCTGATT
CTTGAGTTTAGTGATCTGAACAGAAATACCAAAATATTTGAGAAATGTACAATTTTACCTAGTACAA
GGCAACATATAGTTGTGTAATGTTTAAACAGGCTTTGCTCTGCTGATGGGAGAGAGACATGAATA
TGATAGTAAAGAAATGACACTTTTCTGTTGATTTCCCTCCAAGGTATGGTAAATAGGTTTCACTGA
CTTAGAACAGGAGGAGACTTGTGGCTGGGAGGCTGGGGAAGCTCTTAAATGAAGGAATTTGAG
TTGGATCATCTATTGCTGGCAAGACAGAGGAGGCAAGGCTAGATCATGAAGAACTTGGCTGAGCA
AGGAGACAGAGCTGTTGGGAGAGATGGGAGGAGGAGCAAGGCTAGATCATGAAGAACTTGGCTGAGCA
TTGCTCCGTCTAAGTCAAGTCAAGTGAAGGAGGATCTGTTGGTGTGAGAGAGGTTTACTCTGTGGCC
AAAGGAGGCTCAGGAAGGATGAGCATTTAGGGCAAGGAGACCAACACGCTCTCAGTCAAGGAGGAG

```

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Show in Genome Data Viewer

Articles about the CCR5 gene

CCR5 activation and endocytosis in circulating tumor-derived cells [Breast Cancer Res. 2022]

Tumor bud-derived CCL5 recruits fibroblasts and promotes colorectal [J Exp Clin Cancer Res. 2022]

CCR5 as a Coreceptor for Human Immunodeficiency Virus an [Front Immunol. 2022]

See all...

Reference sequence information

RefSeq alternative splicing

See 3 reference mRNA sequence splice variants for the CCR5 gene.

RefSeq protein product

See the reference protein sequence for C-C chemokine receptor type 5 (NP_000579.1).

More about the CCR5 gene

This gene encodes a member of the beta chemokine receptor family, which is predicted to be a transmembrane protein with 7 transmembrane domains.


GenBank Reference : Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2010). GenBank. *Nucleic acids research*, 39(suppl_1), D32-D37.

2 Predict the secondary structure of the mRNA with minimum free energy.

a. A commonly used computer algorithm to predict the secondary structure of the mRNA is the Sfold program (<http://sfold.wardsworth.org/index.pl>), which is available in the public domain.

- The Srna tool on the Sfold web server is used to predict the general features of RNA folding.
- The algorithm used by Srna predicts only the best secondary structure of the target transcript by a statistical sampling algorithm
- The tool allows up to 10,000 bases in its batch mode.
- The algorithm performs the folding at a temperature of 37°C and assumes the ionic conditions to be 1M NaCl with no divalent ions.

Sfold Reference : Ding, Y., Chan, C. Y., & Lawrence, C. E. (2004). S fold web server for statistical folding and rational design of nucleic acids. *Nucleic acids research*, 32(suppl_2), W135-W141.


Software for Statistical Folding of Nucleic Acids and Studies of Regulatory RNAs

HOMELICENSE INFOMANUALFAQCONTACT
Tuesday August 2, 2022

Srna

82233 sequences folded since April 1, 2003

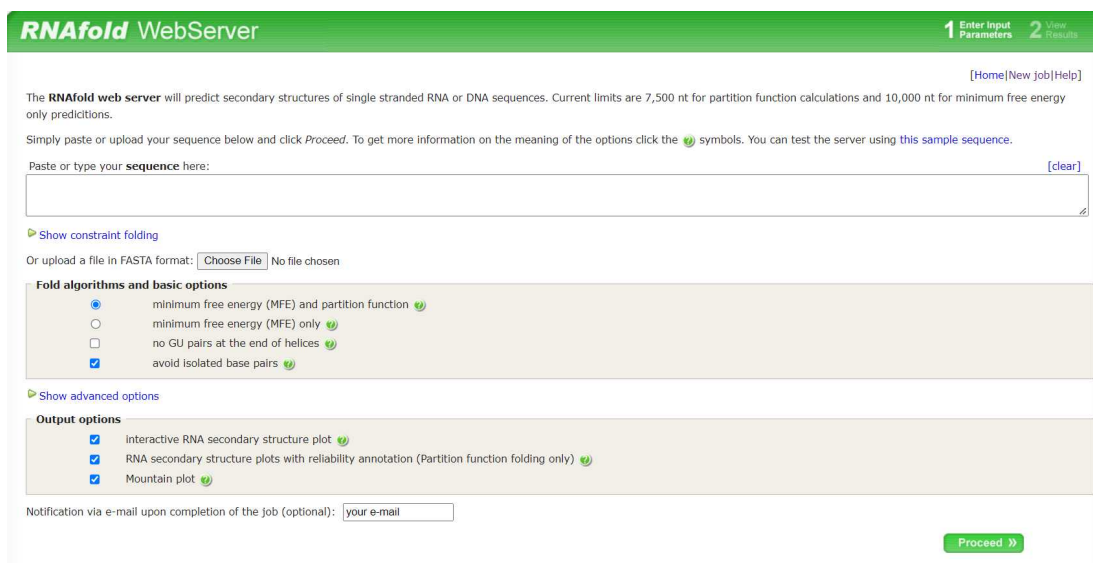
Job mode	Batch (current limit of 10000 bases) ▾
Email address for batch job submission	<input type="text"/>
Please enter a name for your sequence	<input type="text"/>
Please choose your input format Please enter the sequence to be folded. <i>Note:</i> for all input formats, characters other than A, T, C, G, U or N in the sequence will be edited out. or click on Browse button to upload your sequence file	Plain Sequence ▾ <div><input type="text"/></div> <input type="button" value="Choose File"/> No file chosen
Please check if you need to reverse the sequence and then take its complement for folding	<input type="checkbox"/>
Please check if you would like cluster representation of sampled structures <small>NEW</small>	Enabled
Maximum distance between paired bases (optional, leave blank for no limit) <small>NEW</small> Additional constraint information (optional) <small>NEW</small> <i>P i 0 k</i> to force bases <i>i, i+1, ..., i+k-1</i> to be single stranded <i>P i j k</i> to prevent consecutive base pairs (<i>i,j</i>), (<i>i+1,j-1</i>), ..., (<i>i+k-1,j-k+1</i>) to form	<div><input type="text"/></div> <i>Note:</i> if the maximum distance is specified here, this distance will also be applied to results for other modules at the output page. <div><input type="text"/></div>
Folding temperature	37°C
Ionic conditions	1M NaCl, no divalent ions
<input type="button" value="Submit"/> <input type="button" value="Reset"/>	

b. The RNAfold algorithm can also be used for the prediction of the secondary structure of the target mRNA.

i. RNAfold predicts the minimum free energy (MFE) structure of the RNA sequence by using a dynamic programming algorithm. As an RNA secondary structure can be uniquely decomposed into loops and external bases the loop-based energy model treats the free energy $F(s)$ of an RNA secondary structure s as the sum of the contributing free energies F_L of the loops L contained in s . According to the chosen energy parameter set and a given

temperature (defaults to 37 °C) the secondary structure s that minimizes $F(s)$ is computed.

Note: The RNAfold server does not take into account pseudoknots.



RNAfold WebServer Reference : Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic acids research*, 31(13), 3429-3431.

3 Identification of preferable mRNA local secondary structures for ASO binding.

a. An effective ASO should be designed at the regions where mRNA is accessible for hybridization.

b. The accessible sites are identified using the secondary structure of the RNA. Local structures accessible to ASOs are those usually located at the terminal end, internal loops, joint sequences, hairpins and bulges of 10 or more consecutive nucleotides.

c. After confirmation of the accessible conserved local secondary structures and the corresponding sequences of ASOs (approximately 20 bp), one can settle on some well-defined activity enhancing motifs and discard those activity decreasing motifs in the ASOs.


d. Studies have found a positive correlation between ASO-mediated mRNA knockdown and the presence of CCAC, TCCC, ACTC, GCCA and CTCT motifs in the ASOs. Conversely, the presence of GGGG (G-quartets formation), ACTG, AAA and TAA motifs in ASOs weakened ASO activity. Studies also showed strong ASO effects with a minimum of 11 G or C residues in 20 bp ASOs, whereas poor inhibition was observed by ASOs having nine or fewer G or C residues.

e. The tool *Soligo* present in the server, *Sfold* predicts accessible sites on the target RNA for ASO binding.

i. The user can directly enter the target mRNA sequence in the server or upload a file containing the sequence in FASTA format.

ii. The batch mode allows the entry of up to 10,000 bases.

iii. The user can specify the desired length of the antisense oligonucleotides


Software for Statistical Folding of Nucleic Acids and Studies of Regulatory RNAs

HOMELICENSE INFOMANUALFAQCONTACT
Tuesday August 2, 2022

Soligo

38472 sequences folded since April 1, 2003

Job mode	Batch (current limit of 10000 bases) ▼
Email address for batch job submission	<input type="text"/>
Please enter a name for your sequence	<input type="text"/>
Please choose your input format Please enter the sequence to be folded. <i>Note:</i> for all input formats, characters other than A, T, C, G, U or N in the sequence will be edited out. or click on Browse button to upload your sequence file	Plain Sequence ▼ <div><input type="text"/></div> <input type="button" value="Choose File"/> No file chosen
Please check if you need to reverse the sequence and then take its complement for folding	<input type="checkbox"/>
Preferred length of antisense oligos	20
Type of organisms <small>NEW</small>	Eukaryotes (no limit on distance between paired bases) ▼ <i>Note:</i> if the maximum distance is set here, this distance will also be applied to results for other modules at the output page.
Additional constraint information (optional) <small>NEW</small> <i>P I O k</i> to force bases $i, i+1, \dots, i+k-1$ to be single stranded <i>P I j k</i> to prevent consecutive base pairs $(i,j), (i+1,j-1), \dots, (i+k-1,j-k+1)$ to form	<div><input type="text"/></div>
Folding temperature	37°C
Ionic conditions	1M NaCl, no divalent ions
<input type="button" value="Submit"/> <input type="button" value="Reset"/>	

iv. An accessible site can be targeted by a number of antisense oligonucleotides, selection of the "optimal" one can be based on binding energy, together with other empirical rules such as GC content, avoidance of GGGG (or more stringent GGG) motifs, etc.

v. Stronger binding is indicated by smaller binding energy (stacking energies are *negatively valued*). To design a potent ASO, the binding energy between the ASO and mRNA should be -8 kcal/mol.

vi. The antisense oligo binding energy is a weighted sum of the DNA/RNA stacking

energies for the hybrid formed by the antisense oligo and the targeted sequence. For a base-pair stack, the weight for the sum is calculated by the probability of the unpaired dinucleotide in the target sequence that is involved in the stack. This weighting scheme accounts for the structural variation at the target site among the structures in the sample.

vii. The output has the following filter criteria:

- o $40\% \leq \text{GC \%} \leq 60\%$;
- o Antisense oligo binding energy ≤ -8 kcal/mol.
- o No GGGG in the target sequence.

~~~~~Filtered output for design of antisense oligos~~~~~

Column 1: target position (starting - ending)  
 Column 2: target sequence (5p --> 3p)  
 Column 3: antisense oligo (5p --> 3p)  
 Column 4: GC content  
 Column 5: oligo binding energy (kcal/mol)

FILTER CRITERIA: (" $\leq$ ": less than or equal to)

- A)  $40\% \leq \text{GC \%} \leq 60\%$ ;
- B) Antisense oligo binding energy  $\leq -8$  kcal/mol;
- C) No GGGG in the target sequence.

```

2- 21 AGUCCGUCACUGGAAGCCGA TCGGCTTCCAGTGACGGACT 60.0% -8.4
182- 201 AGAAACUGCUGGAGCGAACC GGTTCGCTCCAGCAGTTTCT 55.0% -8.3
183- 202 GAAACUGCUGGAGCGAACC GGGTTCGCTCCAGCAGTTTC 60.0% -8.3
342- 361 AUCUUGUACAAAACCAUCGC GCGATGGTTTTGTACAAGAT 40.0% -8.8
343- 362 UCUUGUACAAAACCAUCGCC GGCATGGTTTTGTACAAGA 45.0% -8.8
344- 363 CUUGUACAAAACCAUCGCCA TGGCGATGGTTTTGTACAAG 45.0% -8.8
345- 364 UUGUACAAAACCAUCGCCAU ATGGCGATGGTTTTGTACAA 40.0% -8.8
346- 365 UGUACAAAACCAUCGCCAUC GATGGCGATGGTTTTGTACA 45.0% -9.0
347- 366 GUACAAAACCAUCGCCAUCA TGATGGCGATGGTTTTGTAC 45.0% -8.4
349- 368 ACAAAACCAUCGCCAUCAAA TTTGATGGCGATGGTTTTGT 40.0% -8.3
607- 626 AAACUUGCAGAGCAACGGCG CGCCGTTGCTCTGCAAGTTT 55.0% -9.0
608- 627 AACUUGCAGAGCAACGGCGC GCGCCGTTGCTCTGCAAGTT 60.0% -10.6
619- 638 CAACGGCGCCGUUGGGAUAA TTATCCCAACGGCGCCGTTG 60.0% -9.8
620- 639 AACGGCGCCGUUGGGAUAAU ATTATCCCAACGGCGCCGTT 55.0% -10.1
621- 640 ACGGCGCCGUUGGGAUAAUG CATTATCCCAACGGCGCCGT 60.0% -10.6
622- 641 CGGCGCCGUUGGGAUAAUGA TCATTATCCCAACGGCGCCG 60.0% -10.8
623- 642 GCGCGCCGUUGGGAUAAUGAU ATCATTATCCCAACGGCGCC 55.0% -11.1
624- 643 GCGCGCCGUUGGGAUAAUGAUG CATCATTATCCCAACGGCGC 55.0% -11.4
625- 644 CGCCGUUGGGAUAAUGAUGA TCATCATTATCCCAACGGCG 50.0% -9.0
697- 716 GAAAGGCGUGCUUCCCUCC GGAGGGGAAGCAGCCTTTTC 60.0% -8.7
698- 717 AAAAGGCGUGCUUCCCUCC GGGAGGGGAAGCAGCCTTTT 60.0% -10.7
699- 718 AAAGGCGUGCUUCCCUCCA TGGAGGGGAAGCAGCCTTT 60.0% -11.1
711- 730 CCCUCCAGACCUCUGCUUU AAAGCAGAGGTCTGGGAGGG 60.0% -13.0
712- 731 CCUCCAGACCUCUGCUUUC GAAAGCAGAGGTCTGGGAGG 60.0% -11.1
713- 732 CUCCAGACCUCUGCUUUA TGAAGCAGAGGTCTGGGAG 55.0% -9.5
714- 733 UCCAGACCUCUGCUUUA TGAAGCAGAGGTCTGGGA 50.0% -8.9
1032-1051 CUCUGUGAAAGCUACUUCUC GAGAAGTAGCTTTCACAGAG 45.0% -8.0
1033-1052 UCUGUGAAAGCUACUUCUC GAGAAGTAGCTTTCACAGA 45.0% -8.1
1045-1064 ACUUCUCCAGUGAAAUCUAC GTAGATTTCACTGGAGAAGT 40.0% -9.1
1046-1065 CUUCUCCAGUGAAAUCUACU AGTAGATTTCACTGGAGAAG 40.0% -8.1
1048-1067 UCUCUCCAGUGAAAUCUACU GTAGTAGATTTCACTGGAGA 40.0% -9.5
1049-1068 CUCCAGUGAAAUCUACUACA TGTAGTAGATTTCACTGGAG 40.0% -9.3
1051-1070 CCAGUGAAAUCUACUACAUC GATGTAGTAGATTTCACTGG 40.0% -9.4
1450-1469 ACUCCAGUUAUUGGAGAGG TGGTGGGATTAAGTTGAGT 40.0% -8.4

```

f. The software OligoWalk can be used to calculate the thermodynamic properties between ASO and target mRNA sequence.

URL : [https://rna.urmc.rochester.edu/cgi-bin/server\\_exe/oligowalk/oligowalk\\_form.cgi](https://rna.urmc.rochester.edu/cgi-bin/server_exe/oligowalk/oligowalk_form.cgi)



i. The output of the software gives the probability of a sequence being an effective ASO in descending order.

The screenshot shows the OligoWalk web interface. At the top is a blue header with the 'OligoWalk' logo and 'for siRNA design'. Navigation links include 'Help | Download | Mathews group', 'Home', and 'OligoWalk'. A green banner reads 'Welcome to OligoWalk'. Below this, a paragraph describes the tool: 'OligoWalk is an online server calculating thermodynamic features of sense-antisense hybridization. It predicts the free energy changes of oligonucleotides binding to a target RNA. It can be used to design efficient siRNA targeting a given mRNA sequence. The source code of OligoWalk for siRNA design can be downloaded from here. The efficient siRNA selection method is described in a published paper (link). More references are listed in the Help page. The server has been tested on Firefox 2 and Internet Explorer 7.'

The main form is titled 'Target RNA Sequence' and contains the following fields:

- \*Sequence name: A text box containing 'example'.
- \*Primary sequence: A large text area containing the sequence: GCUAUUUUGGUGGAAUUGGUAGACACGAUA, CUCUUAAGAUGUAUUACUUUACAGUAUGAA, GGUUCAAGUCCUUUAAAUAGCACCA. Below the text area is the instruction 'paste the sequence here'.
- or upload the sequence file (file format: fasta): A button labeled 'Choose File' next to the text 'No file chosen'.
- \*Email address: An empty text box.

A green 'Submit' button is located below the form. Below the submit button, a message states: 'Do not use file uploading, please paste your sequence above for Advanced Options'. A note at the bottom of the form says: 'The fields indicated with an asterisk (\*) are required.'

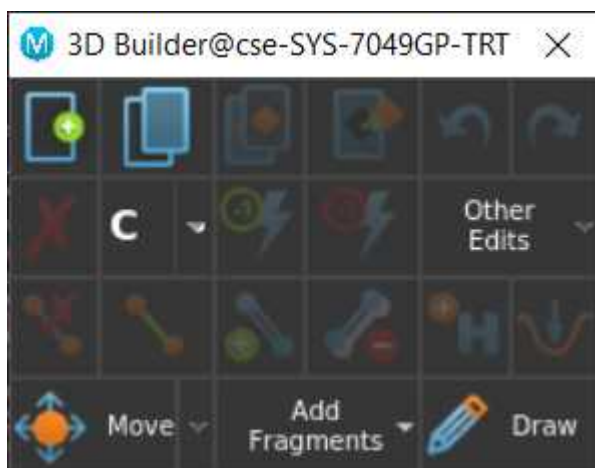
At the bottom of the page, it shows 'Visitors: 69048968', a 'Help' link, a 'Contact us' link, and a footer: '© 2007 Mathews group | Maintained by: David Mathews | Page design derived from Styleshout'.

**OligoWalk Reference :** Lu, Z. J., & Mathews, D. H. (2008). OligoWalk: an online siRNA design tool utilizing hybridization thermodynamics. *Nucleic acids research*, 36(suppl\_2), W104-W108.

#### 4 Constructing pdb structure of Nucleic Acids using Desmond (The Desmond software used is free for academic use)

a. Open the Build Panel on the Maestro GUI, select the option Other Edits, under which select the option, Build Biopolymer from Sequence. In the panel that opens, select the Biopolymer type as RNA. Specify the Title and enter the Sequence of mRNA or ASO from 5' to 3'.

b. Click on the Build option.



Build Biopolymer from Sequence@cse-SYS-7049GP-TRT

Biopolymer type: RNA ☐ Grow 3' to 5'

**Title.** Specify the title for the new entry.

**Sequence.** Enter the sequence below, or click the icon to specify an input file.

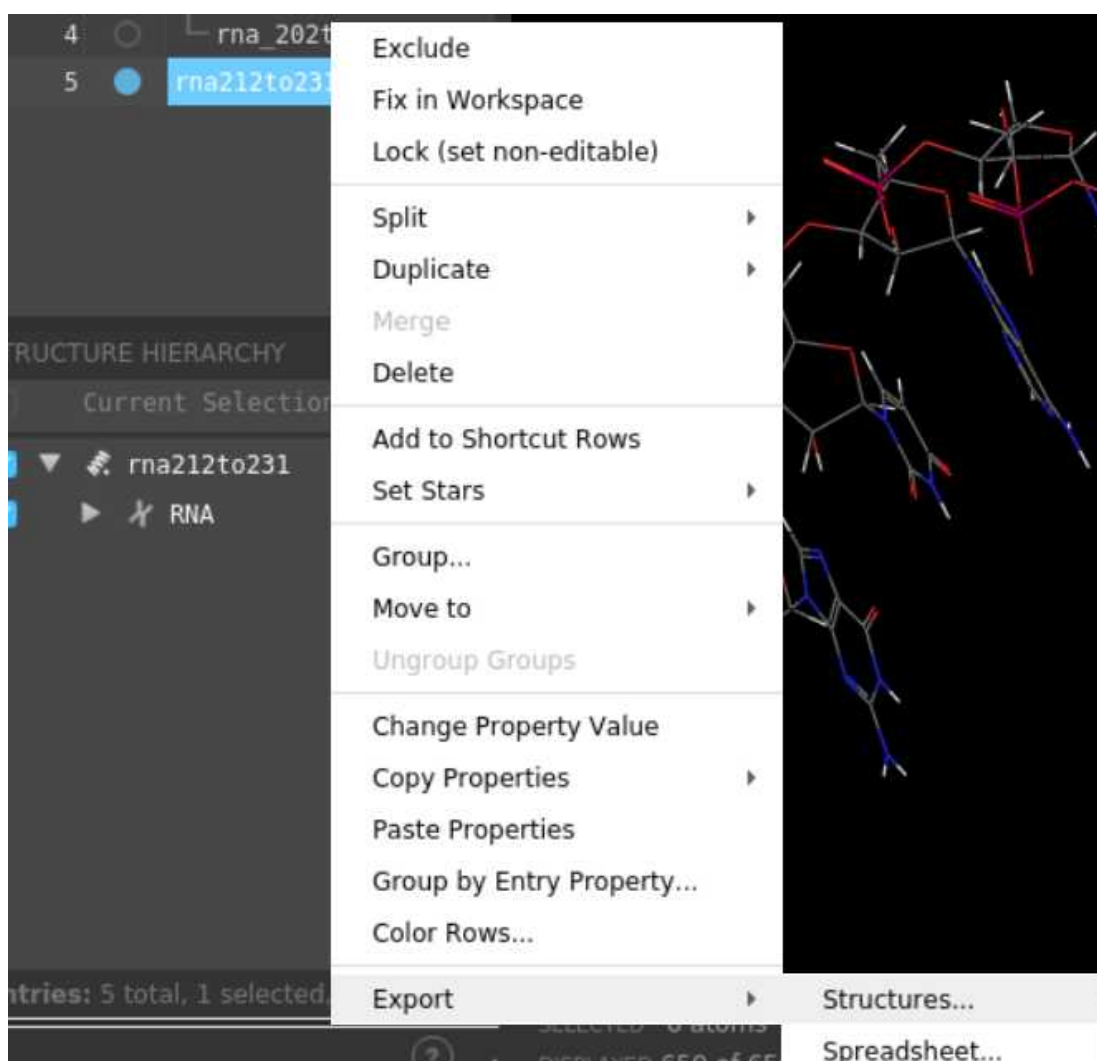
5' --> 3' ...

**Shape.** Choose a helix conformation. ☐ Build double-stranded

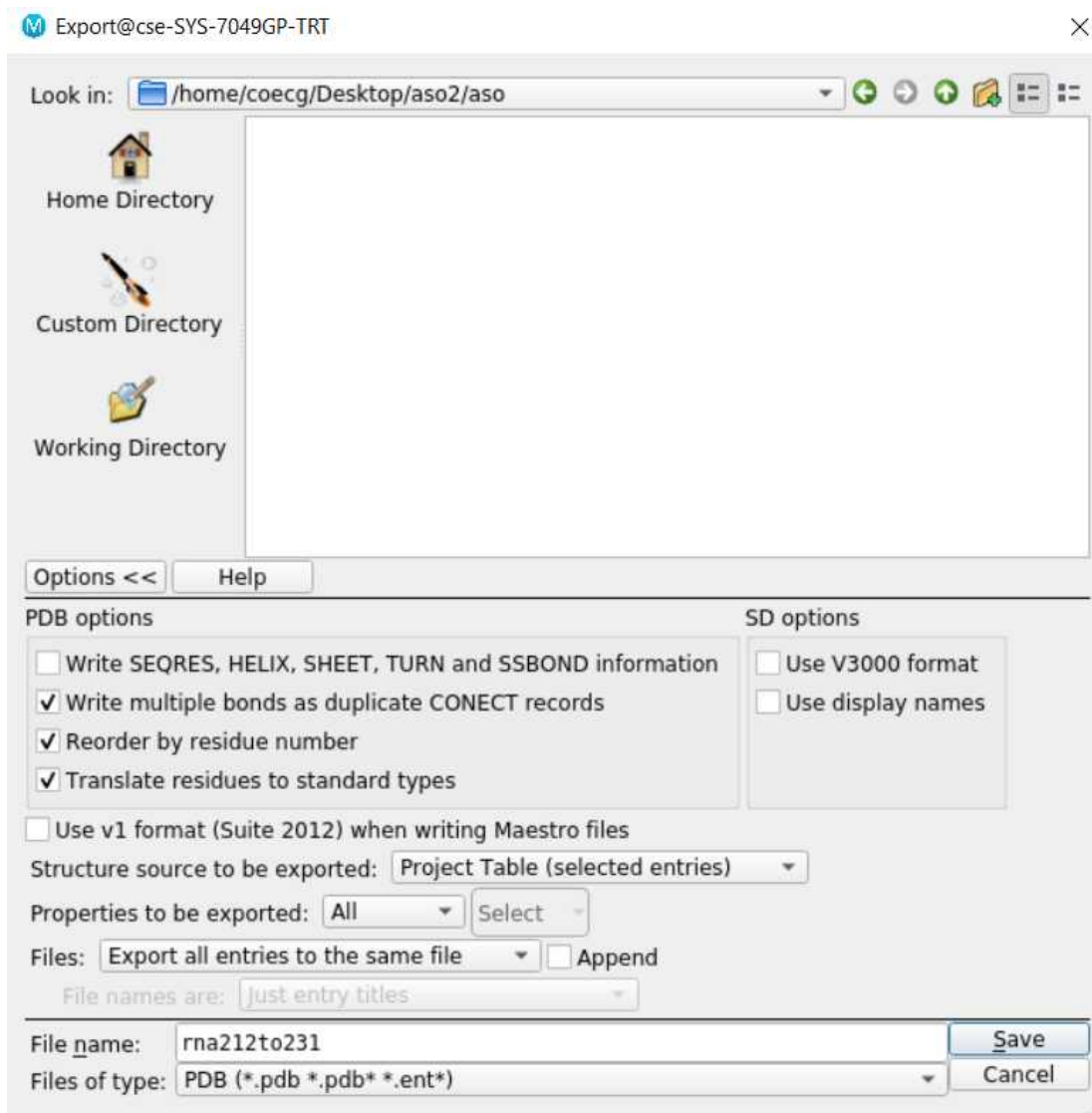
A-RNA

Build Cancel ?

c. A structure is created in the workspace, left click on the structure in the project table and select the Export option and Structures under it. A tab will open, Specify the name of the structure and select the option pdb from the Extension dropdown box.







d. The structure can then be downloaded.

**Desmond Reference :** Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw, "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters," *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, Tampa, Florida, 2006, November 11-17

## 5 Docking of target mRNA and ASO.

a. Select the best predicted ASOs based on the steps above for further docking and

simulation.

b. Nucleic acid-Nucleic acid docking can be done using the HNADOCK Server web interface.

**i. Secondary Structure Prediction:** By default, the server uses the "RNAfold" to predict the secondary structure of an RNA sequence for ab-initio three-dimensional structural modeling if no template is found for the sequence. Users may also choose other secondary structure prediction method like Fold, MaxExpect, ProbKnot, and IPKnot, where Fold, MaxExpect, and ProbKnot are taken from the RNAstructure package.


**ii. RNA-ASO Interaction Prediction:** By default, HNADOCK server builds the complex structures between two RNAs without prior binding information through ab initio docking. However, the server also offers users an option to choose a method (RNAup or RactIP) for RNA-RNA interaction predictions. The predicted interaction information will be then used as inter-RNA distance constraints during docking.

**Note:** A) RNAup calculates the thermodynamics of RNA-ASO interactions, by decomposing the binding into two stages. (1) First the probability that a potential binding sites remains unpaired (equivalent to the free energy needed to open the site) is computed. (2) Then this accessibility is combined with the interaction energy to obtain the total binding energy. All calculations are done by computing partition functions over all possible conformations. B) RactIP is a prediction method for RNA-ASO interaction of general type using integer programming. RactIP can integrate approximate information on an ensemble of equilibrium joint structures into the objective function of integer programming using posterior internal and external base-pairing probabilities.

**HNADOCK Reference :** He J, Wang J, Tao H, Xiao Y, Huang SY. HNADOCK: a nucleic acid docking server for modeling RNA/DNA-RNA/DNA 3D complex structures. *Nucleic Acids Res.* 2019 May 22. pii: gkz412.

**RNAup Reference :** Mückstein, U., Tafer, H., Bernhart, S. H., Hernandez-Rosales, M., Vogel, J., Stadler, P. F., & Hofacker, I. L. (2008, July). Translational control by RNA-RNA interaction: Improved computation of RNA-RNA binding thermodynamics. In *International Conference on Bioinformatics Research and Development* (pp. 114-127). Springer, Berlin, Heidelberg.

**RactIP Reference :** Kato, Y., Sato, K., Hamada, M., Watanabe, Y., Asai, K., & Akutsu, T. (2010). RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming. *Bioinformatics (Oxford, England)*, 26(18), i460–i466.



# HNADOCK Server

Nucleic acid (NA) docking for RNA/DNA-RNA/DNA complex structure modeling.

[\[Huang Lab\]](#)
[\[HNADOCK\]](#)
[\[Help\]](#)
[\[Output example\]](#)

---

**First Nucleic Acid Molecule** using **ONE** of the following three options: [\[help\]](#)

- Upload your **pdb** file in PDB format:  No file chosen [\[example\]](#)
- OR provide your **pdb** file in PDB ID:ChainID:  (Example: 1KD5:A)
- OR copy and paste your RNA **sequence** below [\(Sample input\)](#):

**Second Nucleic Acid Molecule** using **ONE** of the following three options: [\[help\]](#)

- Upload your **pdb** file in PDB format:  No file chosen [\[example\]](#)
- OR provide your **pdb** file in PDB ID:ChainID:  (Example: 1KD5:B)
- OR copy and paste your RNA **sequence** below [\(Sample input\)](#):

**Modeling Options:**

RNA secondary structure prediction method :  [\[help\]](#)

RNA-RNA interaction prediction method :  [\[help\]](#)

Refine the top 10 complex models :  [\[help\]](#)

**Optional:**

Enter your email:


Enter your jobname:

**Option I:** [Specify the residues of the binding site.](#)

[Download the HNADOCK standalone package](#) new

iii. **Results Page:** the result page provides an interactive visualization of the top 10 models using the NGL viewer. Users can choose to view any of the top 10 models or all together by different colors, representations and styles. The result page also gives a summary of the rankings and docking scores for the top 10 complex models, where the score is based the scoring function for RNA-RNA interactions DITScoreRR.

**DITScoreRR Reference :** Yan, Y., Wen, Z., Zhang, D., & Huang, S. Y. (2018). Determination of an effective scoring function for RNA-RNA interactions with a physics-based double-iterative method. *Nucleic acids research*, 46(9), e56.



# HNADOCK Server

Nucleic acid (NA) docking for RNA/DNA-RNA/DNA complex structure modeling.

[\[Huang Lab\]](#)
[\[HNADOCK\]](#)
[\[Help\]](#)
[\[Output example\]](#)

---

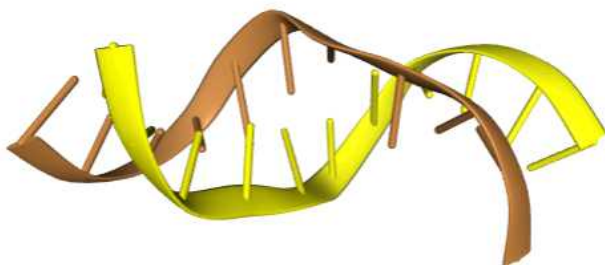
**Your HNADOCK results for job *example***

Download Files

[Receptor PDB file](#)
[Ligand PDB file](#)

[\[1\]](#) [\[2\]](#) [\[3\]](#) [\[4\]](#) [\[5\]](#) [\[6\]](#) [\[7\]](#) [\[8\]](#) [\[9\]](#) [\[10\]](#) [\[11\]](#) [\[12\]](#) [\[13\]](#) [\[14\]](#) [\[15\]](#) [\[16\]](#) [\[17\]](#) [\[18\]](#) [\[19\]](#) [\[20\]](#)

[Top 10 Predictions](#)
[Top 100 Predictions](#)
[All the results in a package](#)



MODEL No.

**Model 1**

Model 2

Model 3

Model 4

Model 5

Model 6

Model 7

Model 8

Model 9

Model 10

Style

Cartoon

Color

Pure

Action

Spin   Reset

**Summary of the top 10 models**

| Rank                   | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| <b>Docking Score</b>   | -509.12 | -384.80 | -369.96 | -360.71 | -352.17 | -342.71 | -336.42 | -334.81 | -324.97 | -317.08 |
| <b>Ligand rmsd (Å)</b> | 0.64    | 6.20    | 14.53   | 6.88    | 6.08    | 4.73    | 16.81   | 6.10    | 8.79    | 7.05    |

(a) Row 1: The ranks of the models.

(b) Row 2: The docking energy scores.

(c) Row 3: The ligand RMSDs from the input structure.

© Lab of Biophysics and Molecular Modeling, huanglab@hust.edu.cn

iv. The resulting docked structures can be downloaded in pdb format.

## 6 Molecular Dynamic Simulation of the Docked Structure of RNA-ASO

**Note:** The following steps have to be performed using the Desmond software that is free for academic use, The Maestro GUI was used for visualization for this example.

- a. Create a new project in the Maestro GUI workspace.
- b. Import the docked structure of RNA-ASO into the workspace
- c. Select the Protein Preparation Wizard tool to preprocess the structure and also minimize the preprocessed structure.



Protein Preparation Wizard@cse-SYS-7049GP-TRT

Job prefix:  Host:  — □ ×

Display hydrogens: ☐ None ☐ Polar only ☒ All ligand, polar receptor ☐ All

Import and Process Review and Modify Refine

Import structure into Workspace

PDB:

Include: ☐ Diffraction data ☐ Biological unit

Import structure file:

Preprocess the Workspace structure

☐ Align to: ☒ Selected entry ☐ PDB:

☒ Assign bond orders ☒ Use CCD database

☒ Add hydrogens ☐ Remove original hydrogens

☒ Create zero-order bonds to metals

☒ Create disulfide bonds

☐ Convert selenomethionines to methionines


☐ Fill in missing side chains using Prime

☐ Fill in missing loops using Prime

☐ Cap termini

☐ Delete waters beyond  Å from het groups

☒ Generate het states using Epik: pH:  +/-



Protein Preparation Wizard@cse-SYS-7049GP-TRT

Job prefix:  Host: 
Display hydrogens: ☐ None ☐ Polar only ☒ All ligand, polar receptor ☐ All

Import and Process Review and Modify Refine

H-bond assignment


☒ Sample water orientations  
☐ Use crystal symmetry  
☐ Minimize hydrogens of altered species  
☒ Use PROPKA pH:  ☐ Label pKas  
☐ Use simplified rules pH: ☐ Very low ☐ Low ☒ Neutral ☐ High

Remove waters

☒ Beyond hets  Å  
☐ With fewer than  H-bonds to non-waters

Restrained minimization

Converge heavy atoms to RMSD:  Å  
☐ Hydrogens only  
Force field:



d. Build the system using the tool, System Builder. In the Solvation tab, select the predefined Solvent Model, SPC and define the boundary conditions: Box Shape>>Orthorhombic; Box Size Calculation Method>>Buffer; Distances>>a)10Å, b)10Å, c)10Å. Select the Force Field, OPLS\_2005. Under the ions tab, select the option Neutralize by adding ions, add the salt ions and select the option Minimize Volume.

System Builder@cse-SYS-7049GP-TRT

Solvation Ions

Set Up Membrane... Delete Membrane

Solvent model

☐ None

☒ Predefined: SPC

☐ Custom: Browse...

Boundary conditions

Box shape: Orthorhombic

Box size calculation method: ☒ Buffer ☐ Absolute size

Distances (Å): a: 10.0 b: 10.0 c: 10.0

Angles (°): α: 90.0 β: 90.0 γ: 90.0

Box volume: 306517 Å³ Minimize Volume

☐ Show boundary box

☐ Use custom charges

☒ Do not use

☐ Partial charges from structure

☐ Custom: Apply to: Select...

Force field: OPLS\_2005

Job name: desmond\_setup\_1 Run

Host=localhost, Incorporate=Append new entries as a new group

System Builder@cse-SYS-7049GP-TRT

Solvation Ions

Excluded region

Exclude ion and salt placement within  Å of

Select... Clear

Ion placement

☐ None

☒ Neutralize by adding 58 Na+ ions Recalculate

☐ Add  Na+ ions

Advanced ion placement...

☒ Add salt

Salt concentration: 0.15 M

Salt positive ion: Na+

Salt negative ion: Cl-

Force field: OPLS\_2005

Job name: desmond\_setup\_1 Run

Host=localhost, Incorporate=Append new entries as a new group ?

e. Open the Molecular Dynamics tool and load the minimized system obtained from the previous step from the workspace; Define the Simulation time. Select the Ensemble Class as NVT and change the temperature to 310.0K. Let the default values remain for the other parameters.



Molecular Dynamics@cse-SYS-7049GP-TRT

### Model system

Load from Workspace ▾ Load

rna\_202to241\_aso\_212to231 - minimized contains 44559 atoms.

### Simulation

Simulation time (ns): Total: 500 Elapsed: 0.0

Recording interval (ps): Trajectory: 500.0 Energy: 1.2

Approximate number of frames: 1000

Ensemble class: NVT ▾

Temperature (K): 310.0 Pressure (bar): 1.01325

Surface tension (bar-Å): 0.0

☒ Relax model system before simulation

Relaxation protocol: Browse...

Advanced Options...

### Analysis

☐ Run interactions analysis when simulation job completes

Protein: Auto ▾

Ligand: Auto ▾

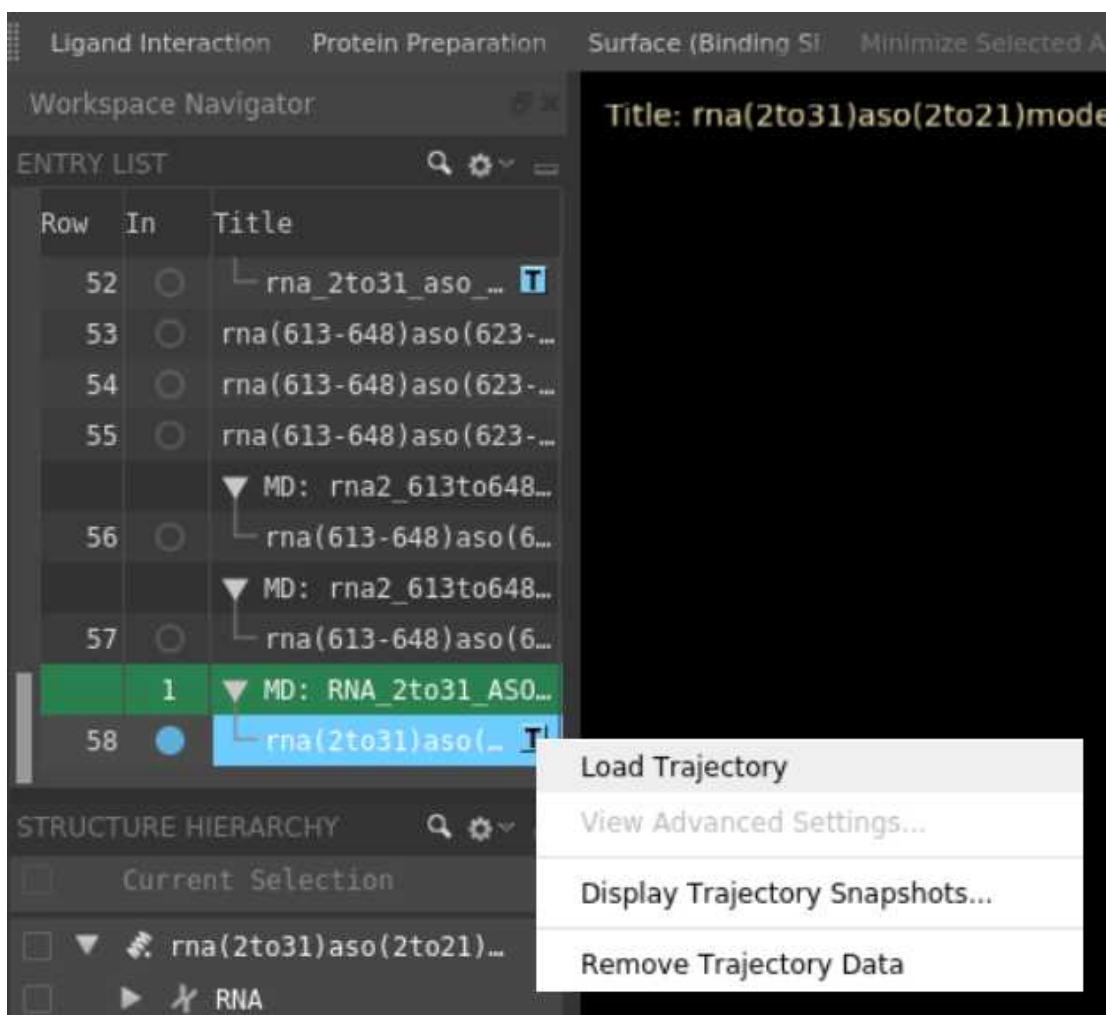
### Desmond

Developed by D. E. Shaw Research

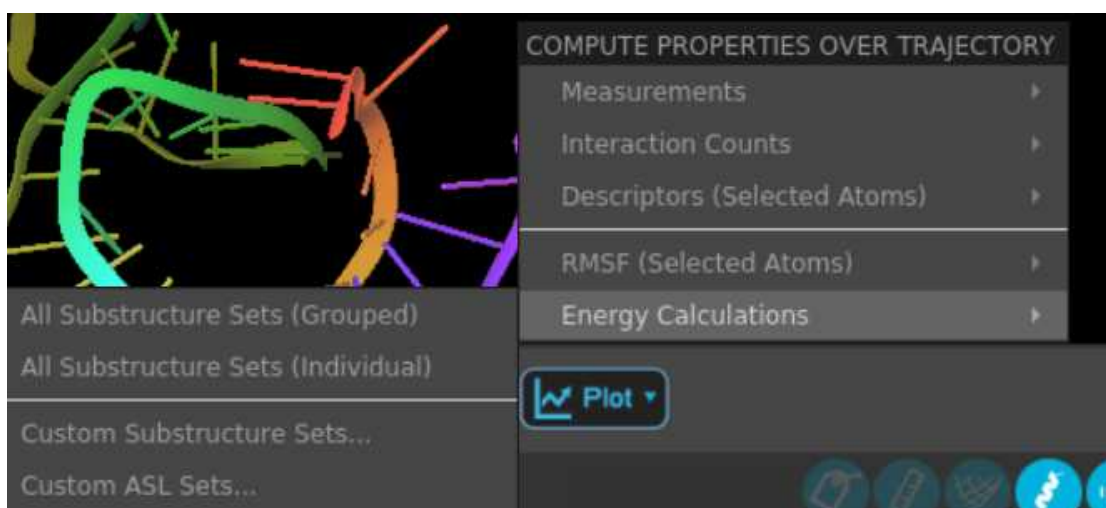
Job name: desmond\_md\_job\_1 ⚙ ▾ Run

Host=localhost:1, Incorporate=Append new entries as a new group ?

f. Once the job is completed, load the trajectory and visualize the results.



g. Energies of the system can be found out by using the Plot tool located on the lower right corner. Select the option, Energy Calculations. An Energy Plot is generated that can be downloaded for further analysis.



**Desmond Reference :** Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico

D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw, "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters," Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, 2006, November 11-17