Aug 23, 2022

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

DOI
https://dx.doi.org/10.17504/protocols.io.ewov1nr4ogr2/v1

Akshay Uttarkar[1], Mansi Babu[2], Vidya Niranjan[1]

[1]Department of Biotechnology, R V College of Engineering;
[2]Research Intern, Centre of Excellence Computational Genomics, R V College of Engineering

Vidya Niranjan
R V College of Engineering

DOI: https://dx.doi.org/10.17504/protocols.io.ewov1nr4ogr2/v1

Protocol Citation: Akshay Uttarkar, Mansi Babu, Vidya Niranjan 2022. Protocol for In-silico Design, Docking and Molecular Dynamic Simulation of Antisense Oligonucleotides. **protocols.io** https://dx.doi.org/10.17504/protocols.io.ewov1nr4ogr2/v1

Protocol status: Working
We use this protocol and it's working

Created: August 02, 2022

Last Modified: August 23, 2022

Protocol Integer ID: 68045

Keywords: ASO design, mRNA docking, MD simulation, antisense oligonucleotides insilico drug design, protocol for aso design, molecular dynamic simulation, aso drug, docking, developing novel chemical, aso design, md simulation study, silico design, fda

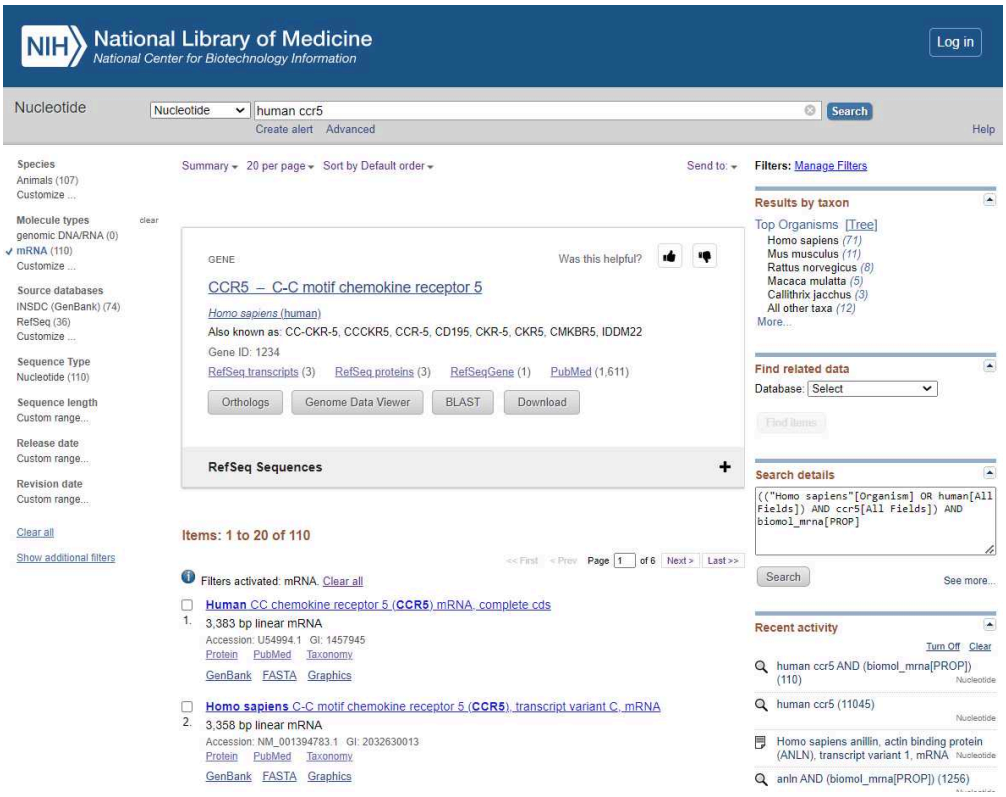
Abstract

Insilico 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

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.



b.After entering the search statement, select the Limits link to limit the search to a particular molecule type and database.

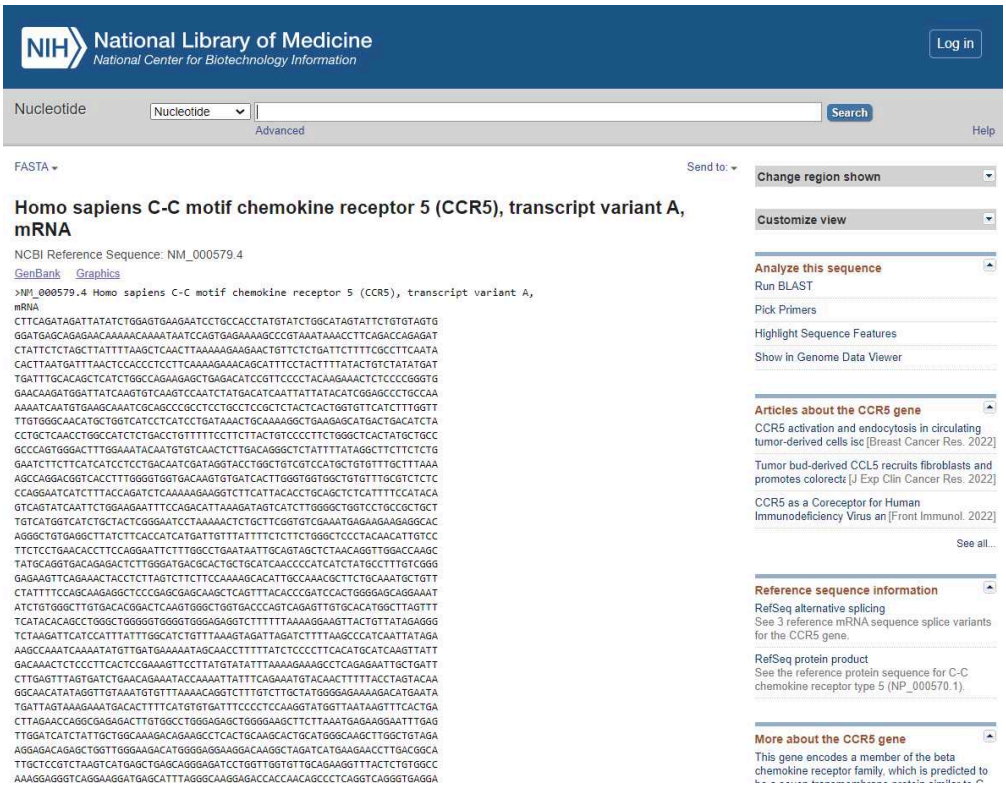
c.Selecting mRNA from the Molecule drop-down menu will retrieve only mRNA sequences.

d.Each of these records is a variant form of the mRNA for the gene.

e.To see how these variants differ from one another one needs to open each record and read about the variant in the COMMENT section of the record.

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

g.Retrieve the mRNA sequence in FASTA format.



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 *S*fold program (**http://sfold.wardsworth.org/index.pl**), which is available in the public domain.

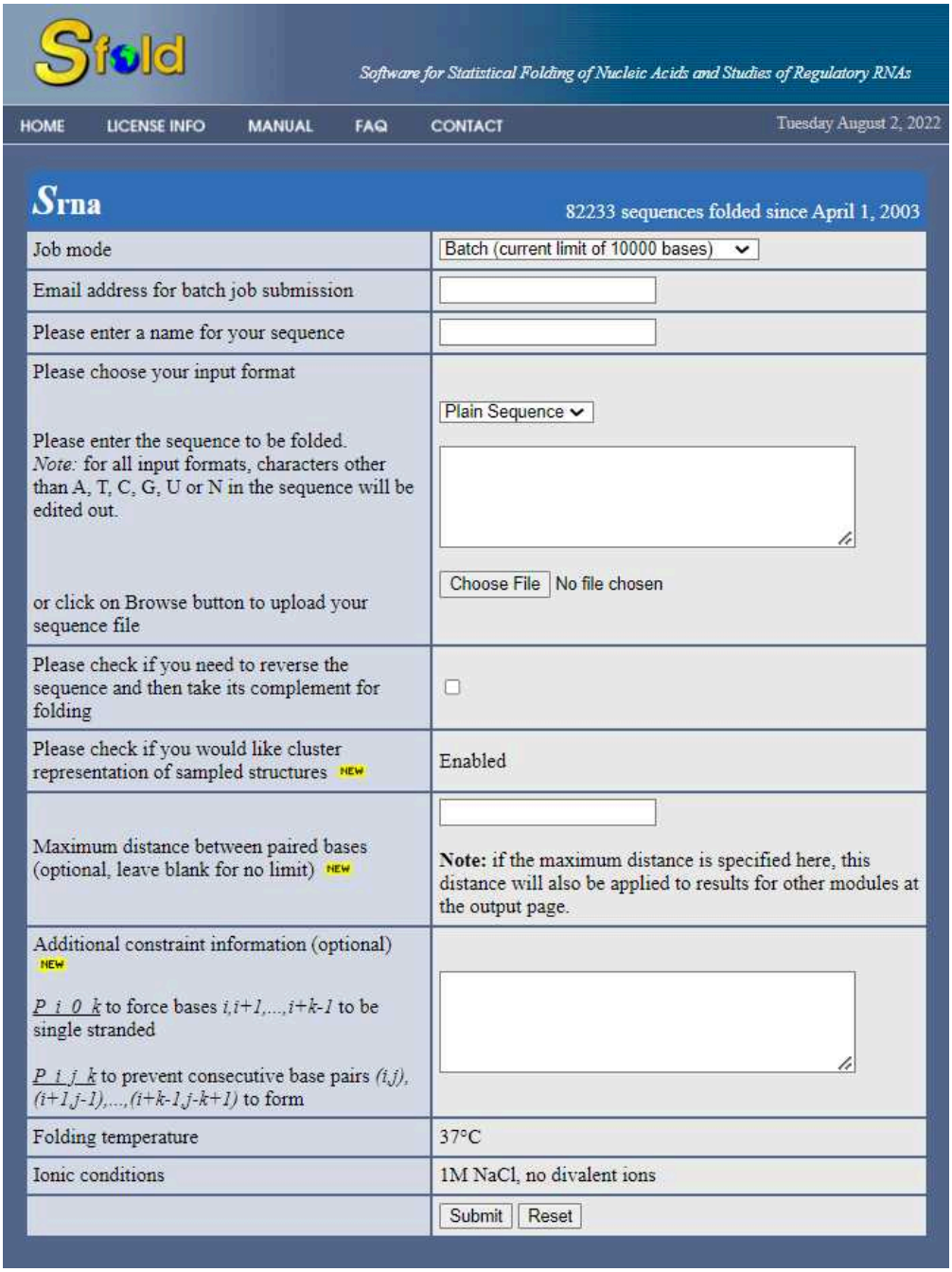
i.The Srna tool on the Sfold web server is used to predict the general features of RNA folding.

ii.The algorithm used by Srna predicts only the best secondary structure of the target transcript by a statistical sampling algorithm

iii.The tool allows upto 10,000 bases in its batch mode.

iv.The algorithm performs the folding at a temperature of 37°C and assumes the ionic conditions to be 1M NaCl with no divalent ions.

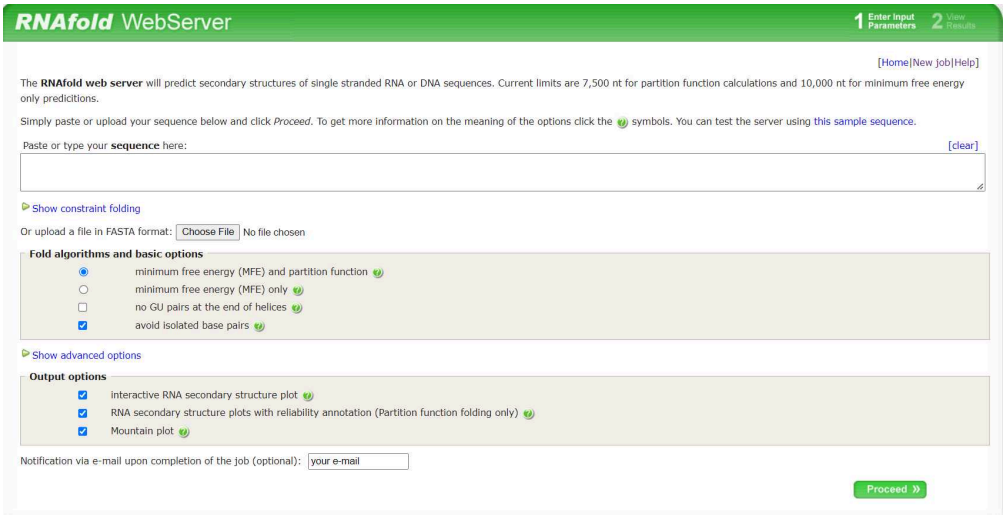
Sfold Reference : Ding, Y., Chan, C. Y., & Lawrence, C. E. (2004). S fold web server for statistical folding and rational design of nucleic acids.*Nucleic acids research*,*32*(suppl_2), W135-W141.



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.

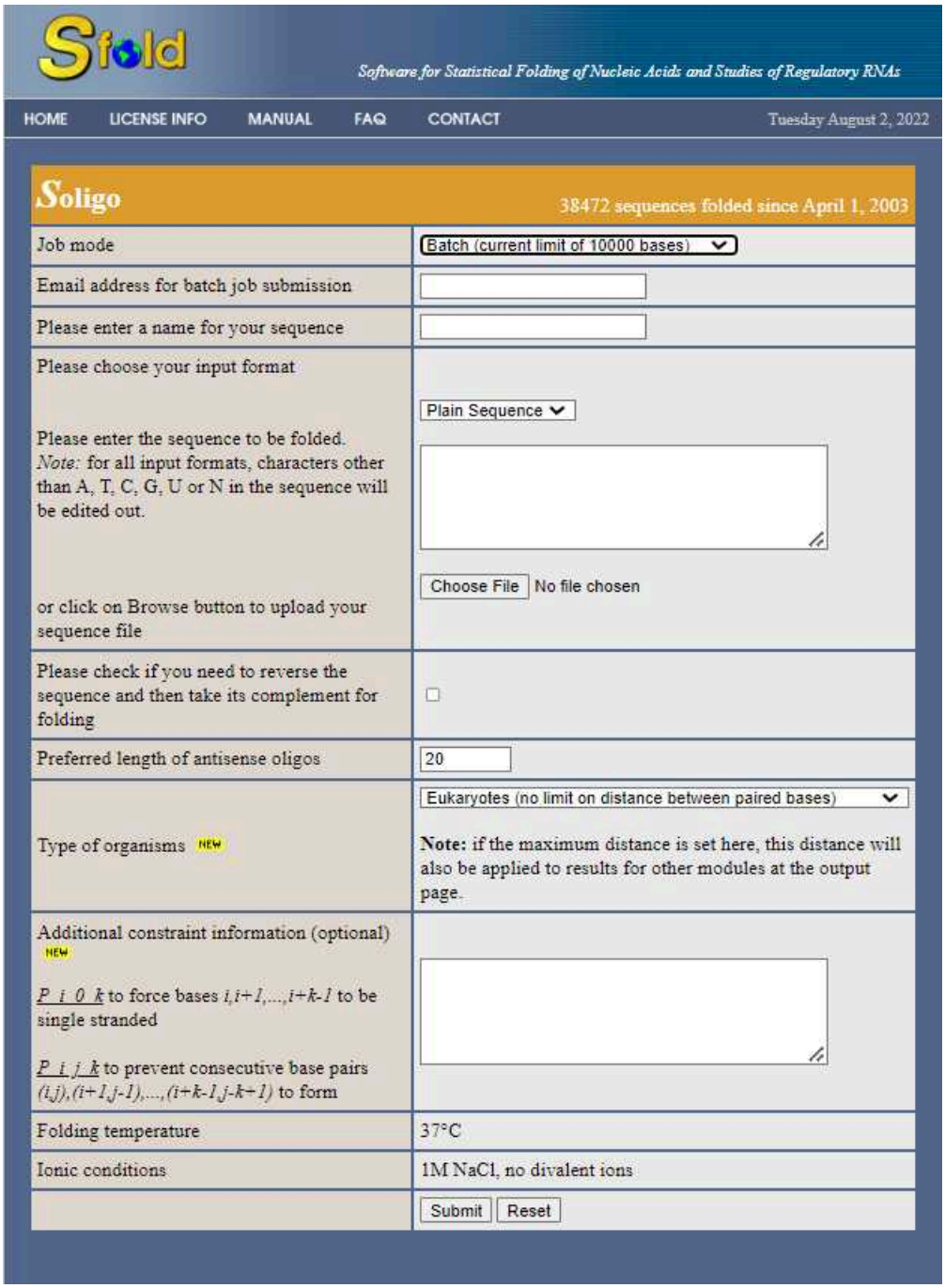
d.Studies have found a positive correlation between ASO-mediated mRNA knockdown and the presence of CCAC, TCCC, ACTC, GCCA and CTCT motifs in the ASOs. Conversely, the presence of GGGG (G-quartets formation), ACTG, AAA and TAA motifs in ASOs weakened ASO activity. Studies also showed strong ASO effects 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 oligonuceotides



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:

- 40% ≤ GC % ≤ 60%;
- Antisense oligo binding energy ≤ -8 kcal/mol.
- 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: ("<=": less than or equal to)

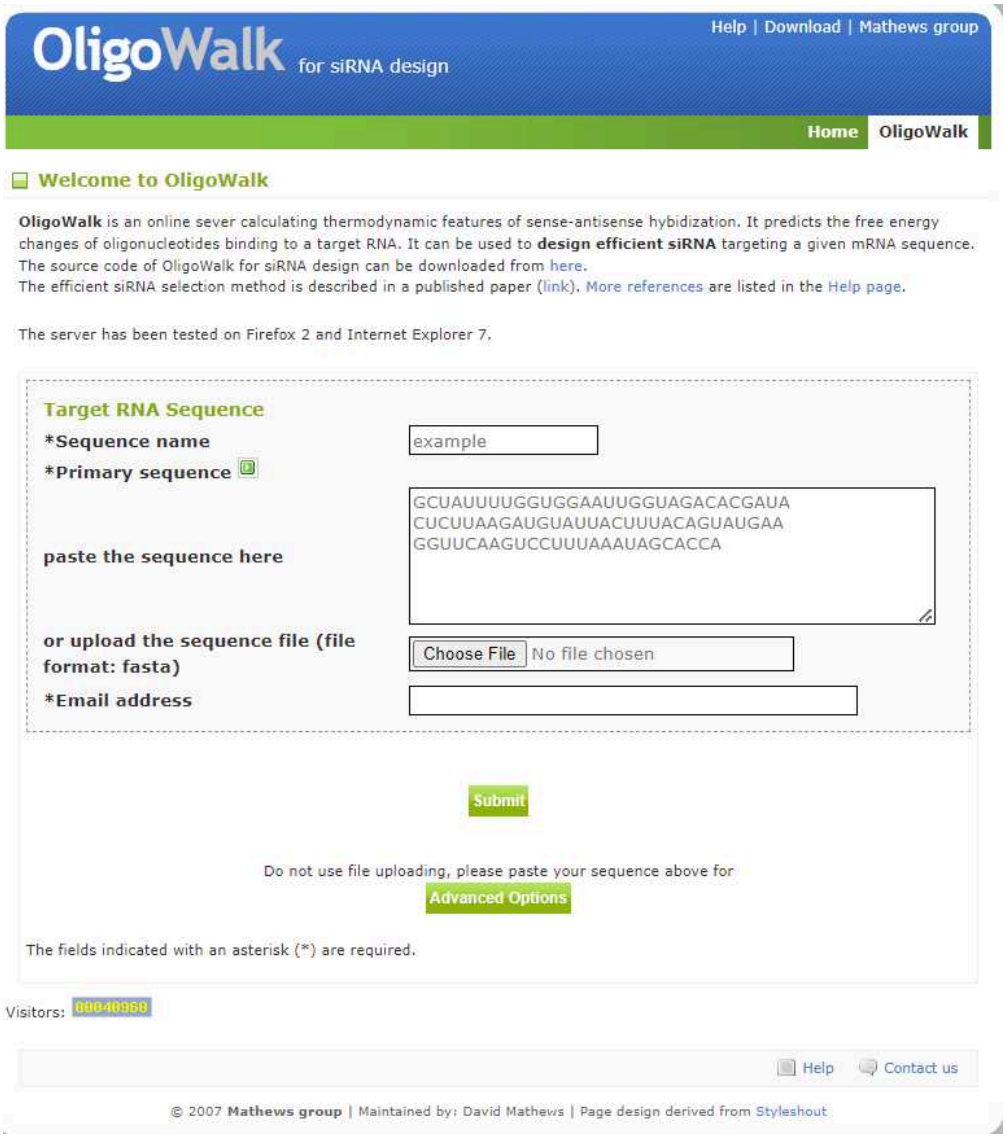
A) 40% <= GC % <= 60%;
B) Antisense oligo binding energy <= -8 kcal/mol;
C) No GGGG in the target sequence.

2- 21	AGUCCGUCACUGGAAGCCGA	TCGGCTTCCAGTGACGGACT	60.0%	-8.4
182- 201	AGAAACUGCUGGAGCGAACC	GGTTCGCTCCAGCAGTTTCT	55.0%	-8.3
183- 202	GAAACUGCUGGAGCGAACCC	GGGTTCGCTCCAGCAGTTTC	60.0%	-8.3
342- 361	AUCUUGUACAAAACCAUCGC	GCGATGGTTTTGTACAAGAT	40.0%	-8.8
343- 362	UCUUGUACAAAACCAUCGCC	GGCGATGGTTTTGTACAAGA	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	GGCGCCGUUGGGAUAAUGAU	ATCATTATCCCAACGGCGCC	55.0%	-11.1
624- 643	GCGCCGUUGGGAUAAUGAUG	CATCATTATCCCAACGGCGC	55.0%	-11.4
625- 644	CGCCGUUGGGAUAAUGAUGA	TCATCATTATCCCAACGGCG	50.0%	-9.0
697- 716	GAAAAGGCUGCUUCCCCUCC	GGAGGGGAAGCAGCCTTTTC	60.0%	-8.7
698- 717	AAAAGGCUGCUUCCCCUCCC	GGGAGGGGAAGCAGCCTTTT	60.0%	-10.7
699- 718	AAAGGCUGCUUCCCCUCCCA	TGGGAGGGGAAGCAGCCTTT	60.0%	-11.1
711- 730	CCCUCCCAGACCUCUGCUUU	AAAGCAGAGGTCTGGGAGGG	60.0%	-13.0
712- 731	CCUCCCAGACCUCUGCUUUC	GAAAGCAGAGGTCTGGGAGG	60.0%	-11.1
713- 732	CUCCCAGACCUCUGCUUUCA	TGAAAGCAGAGGTCTGGGAG	55.0%	-9.5
714- 733	UCCCAGACCUCUGCUUUCAA	TTGAAAGCAGAGGTCTGGGA	50.0%	-8.9
1032-1051	CUCUGUGAAAGCUACUUCUC	GAGAAGTAGCTTTCACAGAG	45.0%	-8.0
1033-1052	UCUGUGAAAGCUACUUCUCC	GGAGAAGTAGCTTTCACAGA	45.0%	-8.1
1045-1064	ACUUCUCCAGUGAAAUCUAC	GTAGATTTCACTGGAGAAGT	40.0%	-9.1
1046-1065	CUUCUCCAGUGAAAUCUACU	AGTAGATTTCACTGGAGAAG	40.0%	-8.1
1048-1067	UCUCCAGUGAAAUCUACUAC	GTAGTAGATTTCACTGGAGA	40.0%	-9.5
1049-1068	CUCCAGUGAAAUCUACUACA	TGTAGTAGATTTCACTGGAG	40.0%	-9.3
1051-1070	CCAGUGAAAUCUACUACAUC	GATGTAGTAGATTTCACTGG	40.0%	-9.4
~~~~~~~~~~~~~~~~~~

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

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



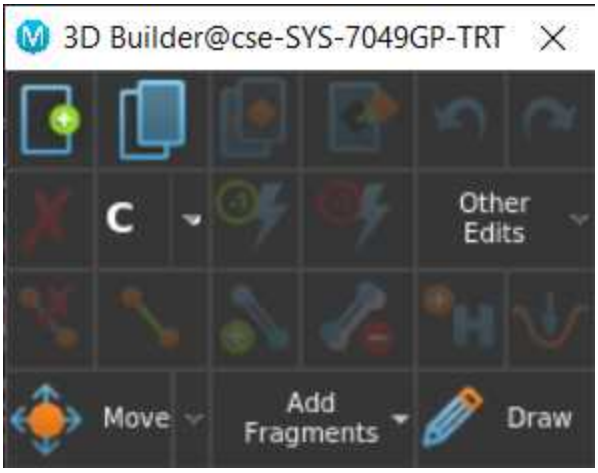
OligoWalk Reference : Lu, Z. J., & Mathews, D. H. (2008). OligoWalk: an online siRNA design tool utilizing hybridization thermodynamics.*Nucleic acids research*,*36*(suppl_2),

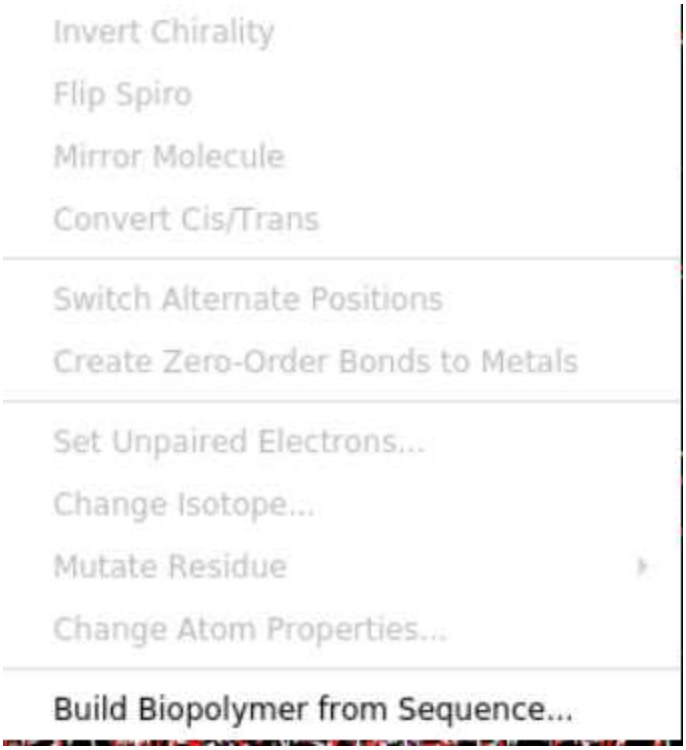
W104-W108.

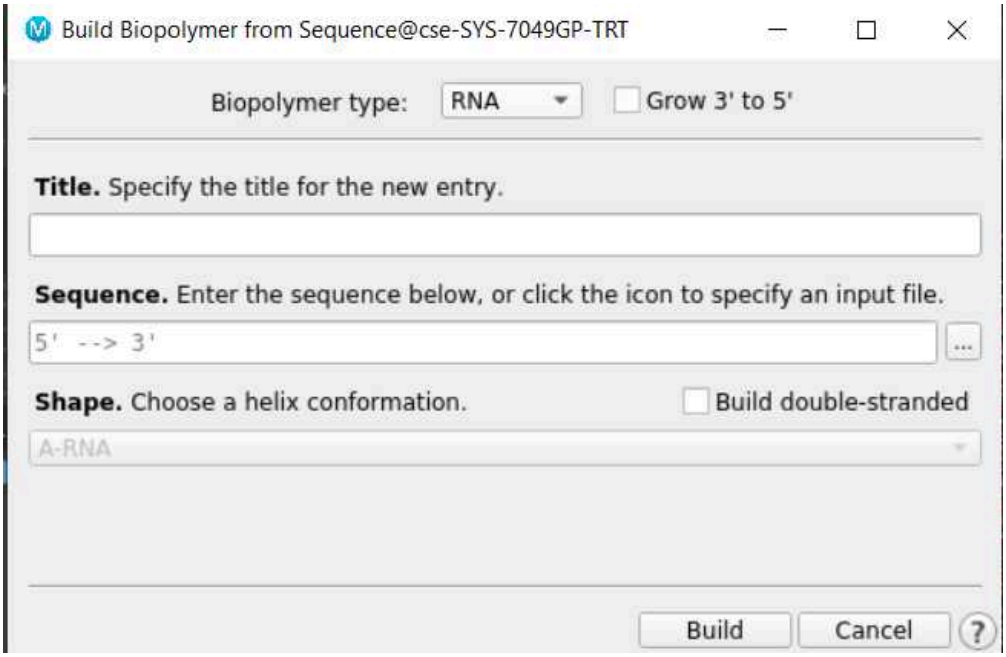
4 **Constructing pdb structure of Nucleic Acids using Desmond**(The Desmond software used is free for academic use)

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′.

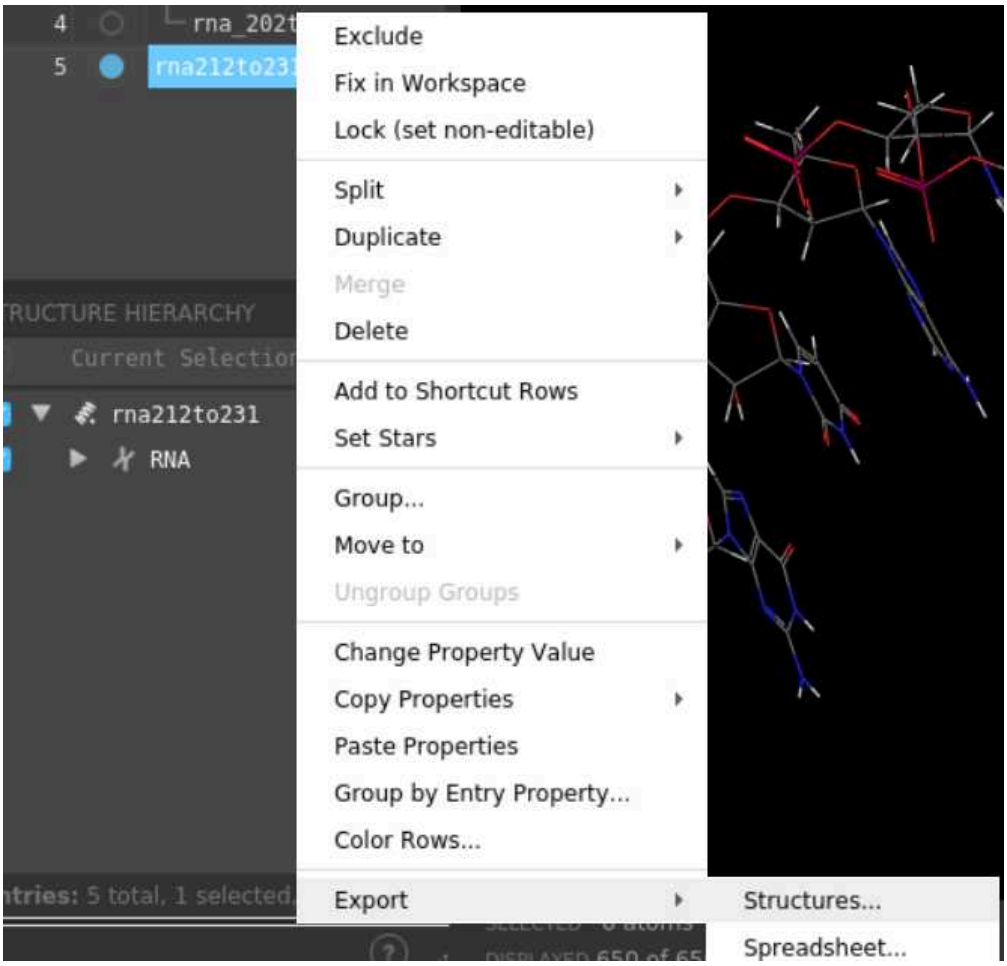
b.Click on the Build option.

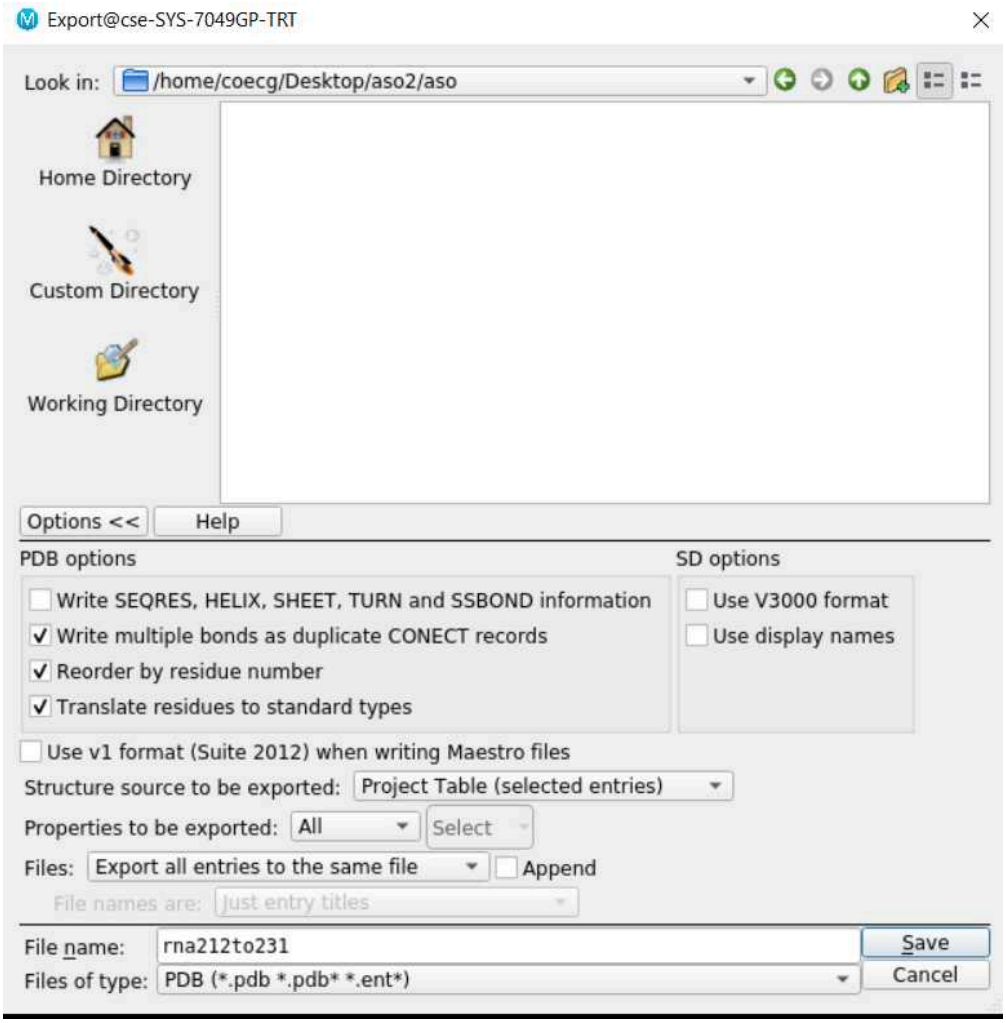






c.A structure is created in the workspace, left click on the structure in the project table and select the Export option and Structures under it. A tab will open, Specify the name of the structure and select the option pdb from the Extention dropdown box.


4
5
rna212to231
RUCTURE HIERARCHY
Current Selection
rna212to231
RNA
Exclude
Fix in Workspace
Lock (set non-editable)
Split
Duplicate
Merge
Delete
Add to Shortcut Rows
Set Stars
Group...
Move to
Ungroup Groups
Change Property Value
Copy Properties
Paste Properties
Group by Entry Property...
Color Rows...
Export
Structures...
Spreadsheet...
tries: 5 total, 1 selected



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.

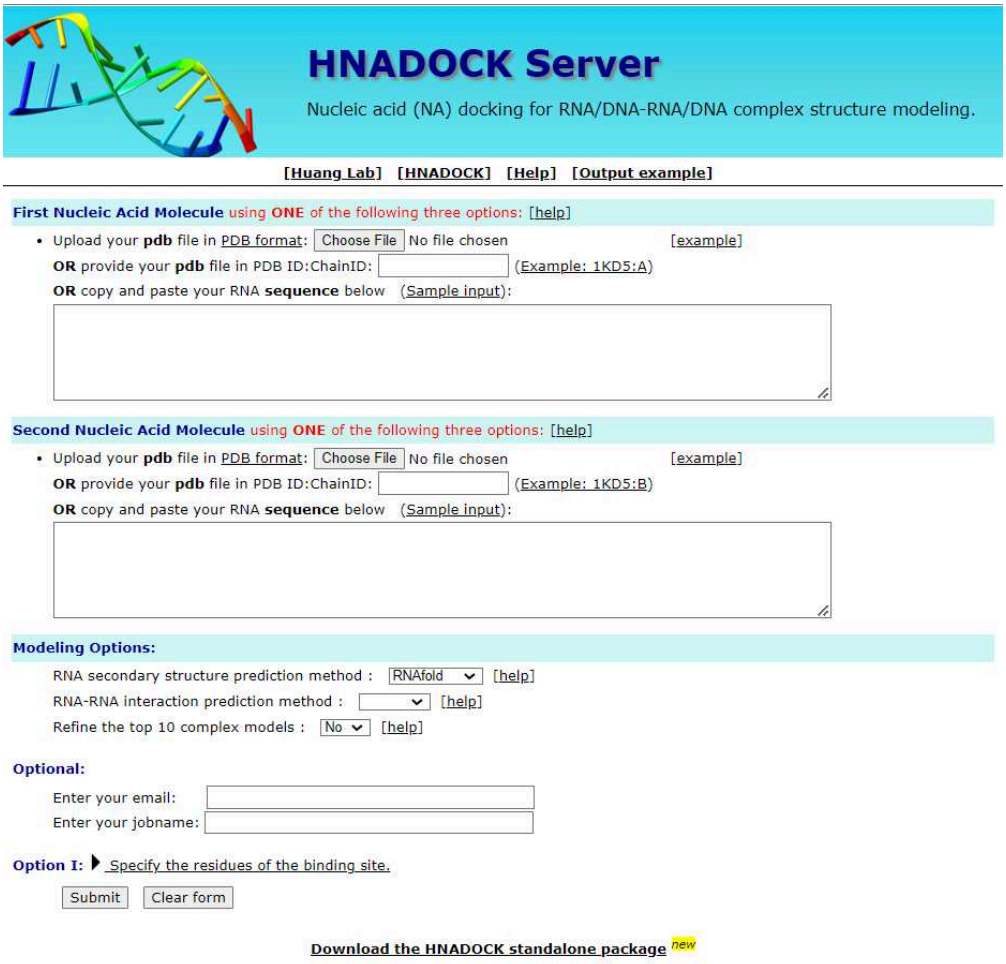
ii.**RNA-ASO Interaction Prediction:** By default, HNADOCK server builds the complex structures between two RNAs without prior binding information through ab initio docking. However, the server also offers users an option to choose a method (RNAup or RactIP) for RNA-RNA interaction predictions. The predicted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

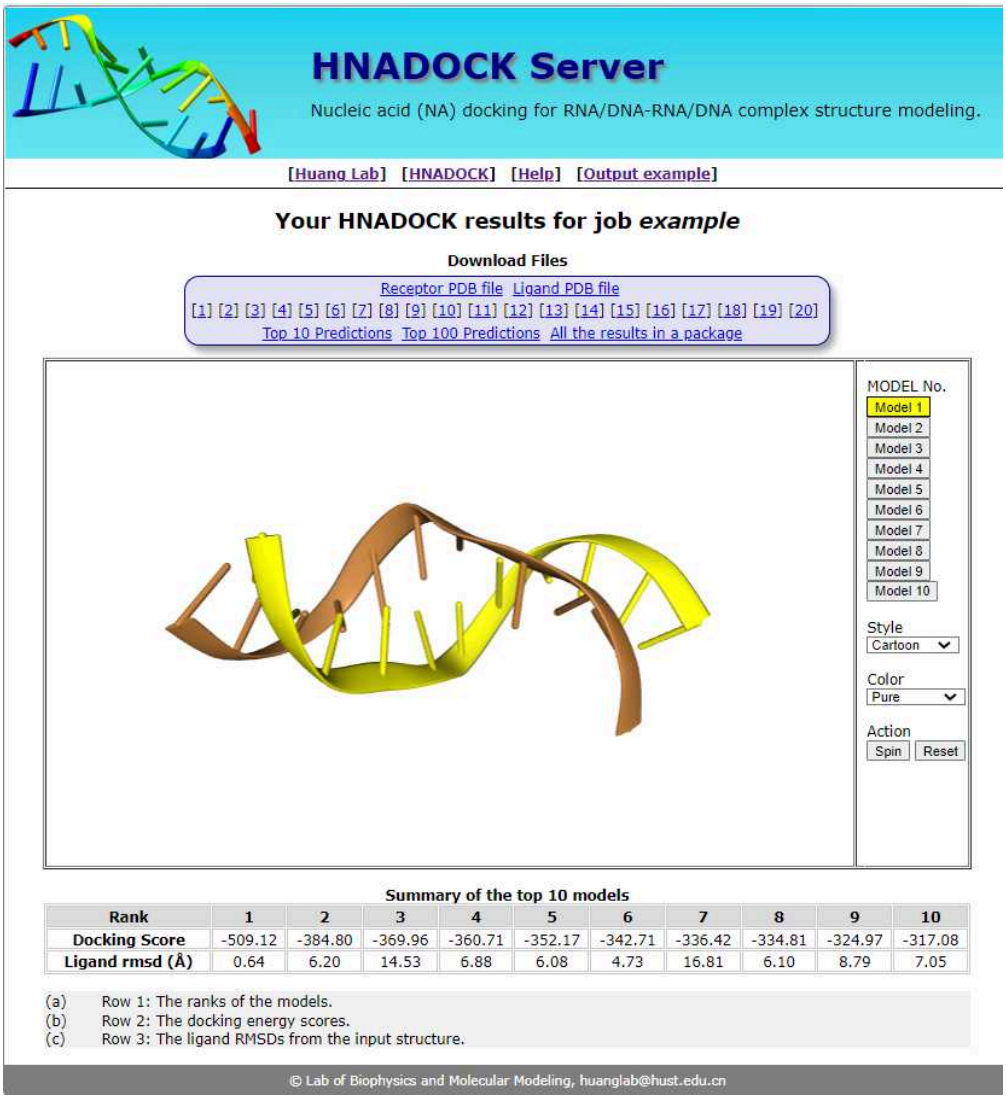
RNAup Reference : Mückstein, U., Tafer, H., Bernhart, S. H., Hernandez-Rosales, M., Vogel, J., Stadler, P. F., & Hofacker, I. L. (2008, July). Translational control by RNA-RNA interaction: Improved computation of RNA-RNA binding thermodynamics. In*International Conference on Bioinformatics Research and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



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.



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

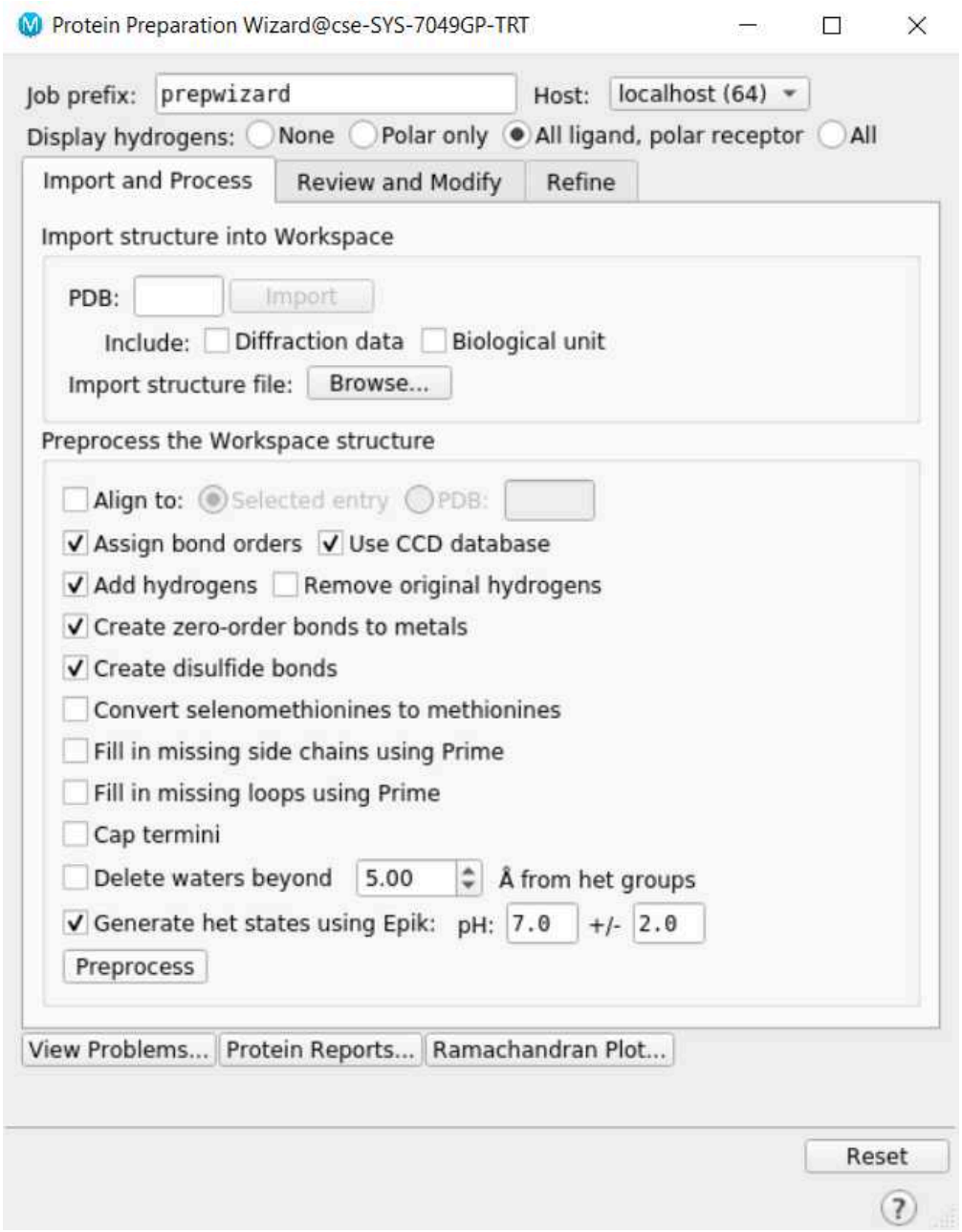
6 **Molecular Dynamic Simulation of the Docked Structure of RNA-ASO**

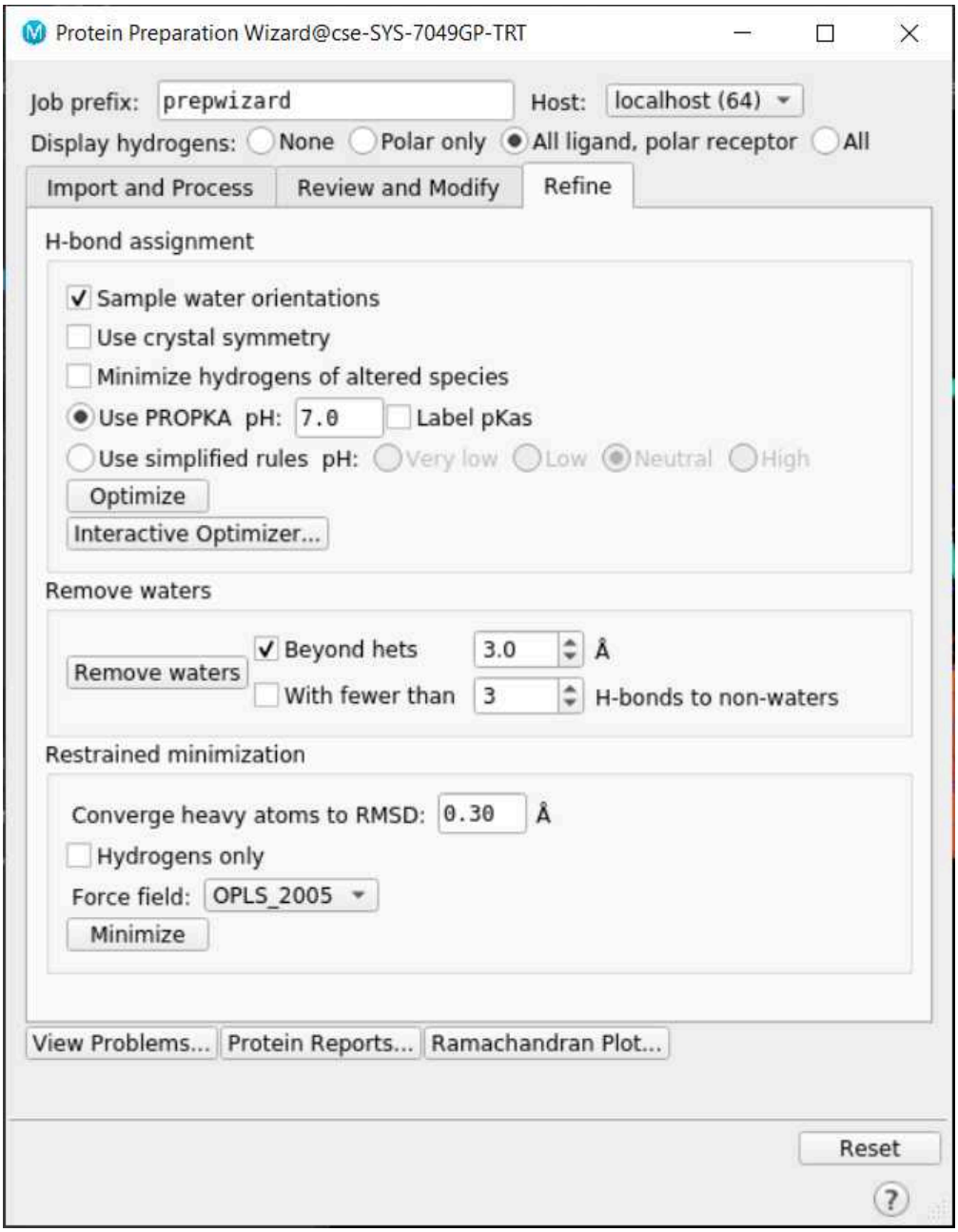
Note: The following steps have to be performed using the Desmond software that is free for academic use, The Maestro GUI was used for visualization for this example.

a.Create a new project in the Maestro GUI workspace.

b.Import the docked structure of RNA-ASO into the workspace

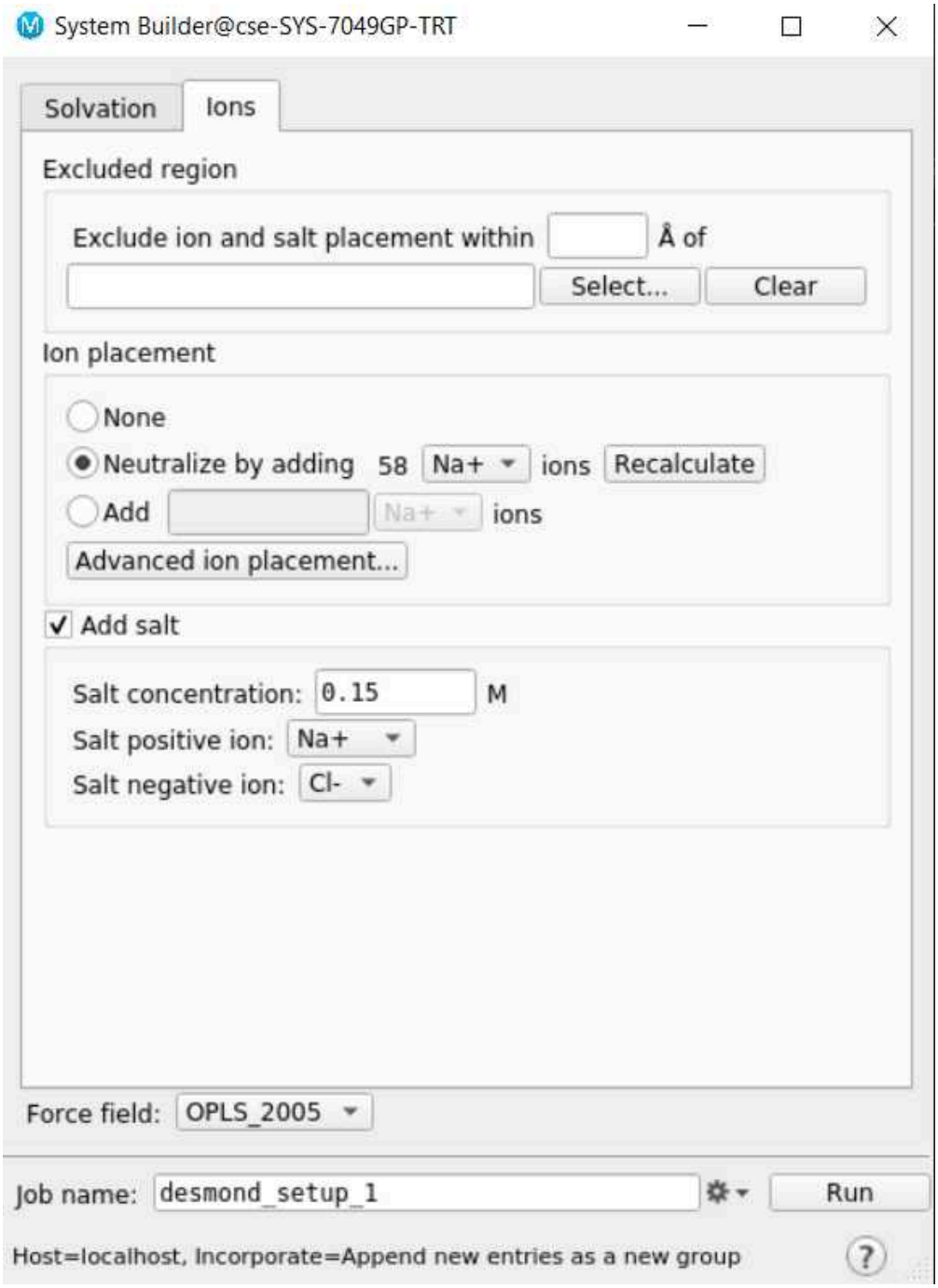
c.Select the Protein Preparation Wizard tool to preprocess the structure and also minimize the preprocessed structure.


Protein Preparation Wizard@cse-SYS-7049GP-TRT
Job prefix: prepwizard
Host: localhost (64)
Display hydrogens: None Polar only All ligand, polar receptor All
Import and Process
Review and Modify
Refine
Import structure into Workspace
PDB:
Import
Include: Diffraction data Biological unit
Import structure file: Browse...
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 5.00 Å from het groups
Generate het states using Epik: pH: 7.0 +/- 2.0
Preprocess
View Problems...
Protein Reports...
Ramachandran Plot...
Reset

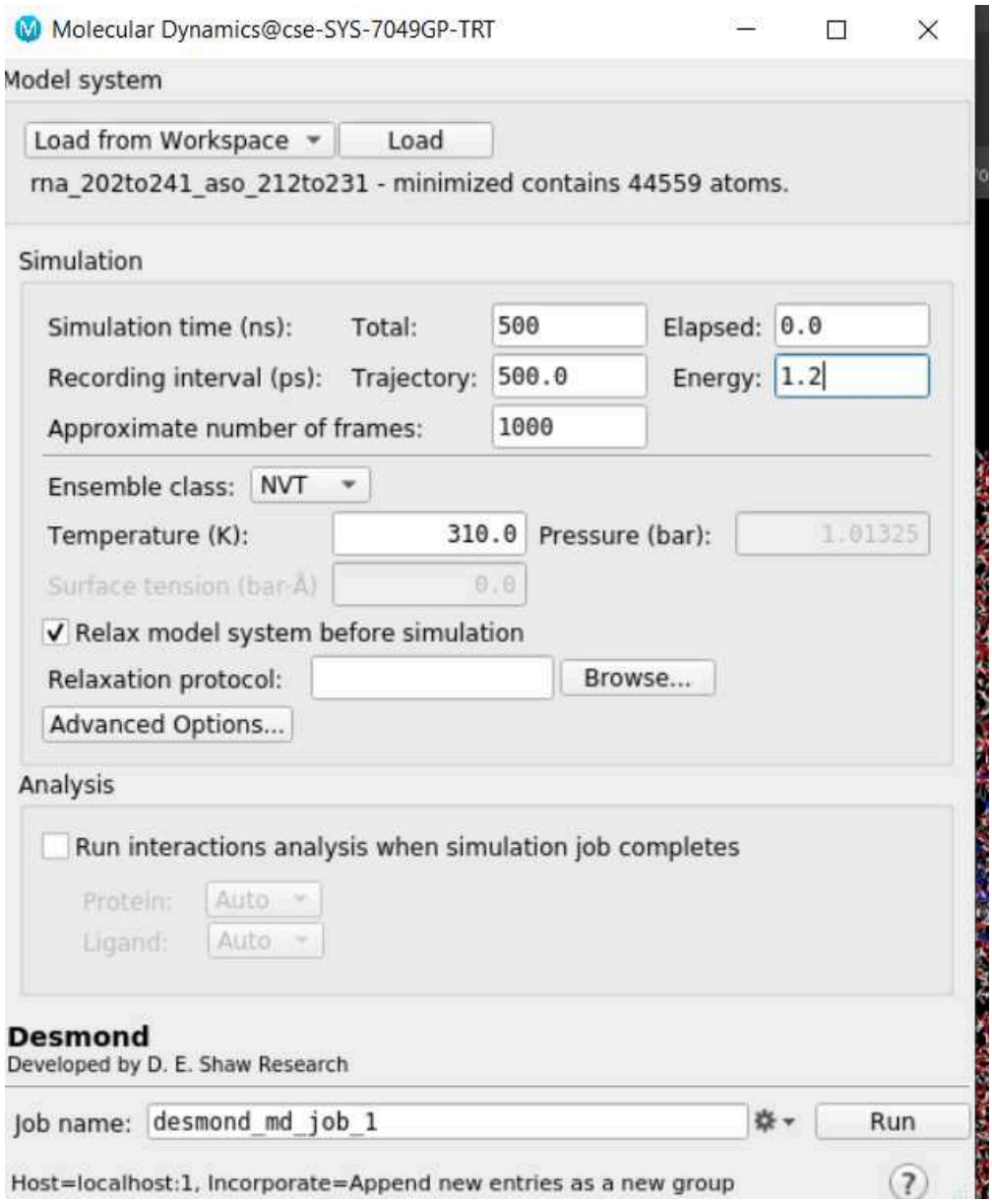


d.Build the system using the tool, System Builder. In the Solvation tab, select the predefined Solvent Model, SPC and define the boundary conditions: Box Shape>>Orthorhombic; Box Size Calculation Method>>Buffer; Distances>>a)10A°, b)10A°, c)10A°. Select the Force Field, OPLS_2005. Under the ions tab, select the option Neutralize by adding ions, add the salt ions and select the option Minimize Volume.

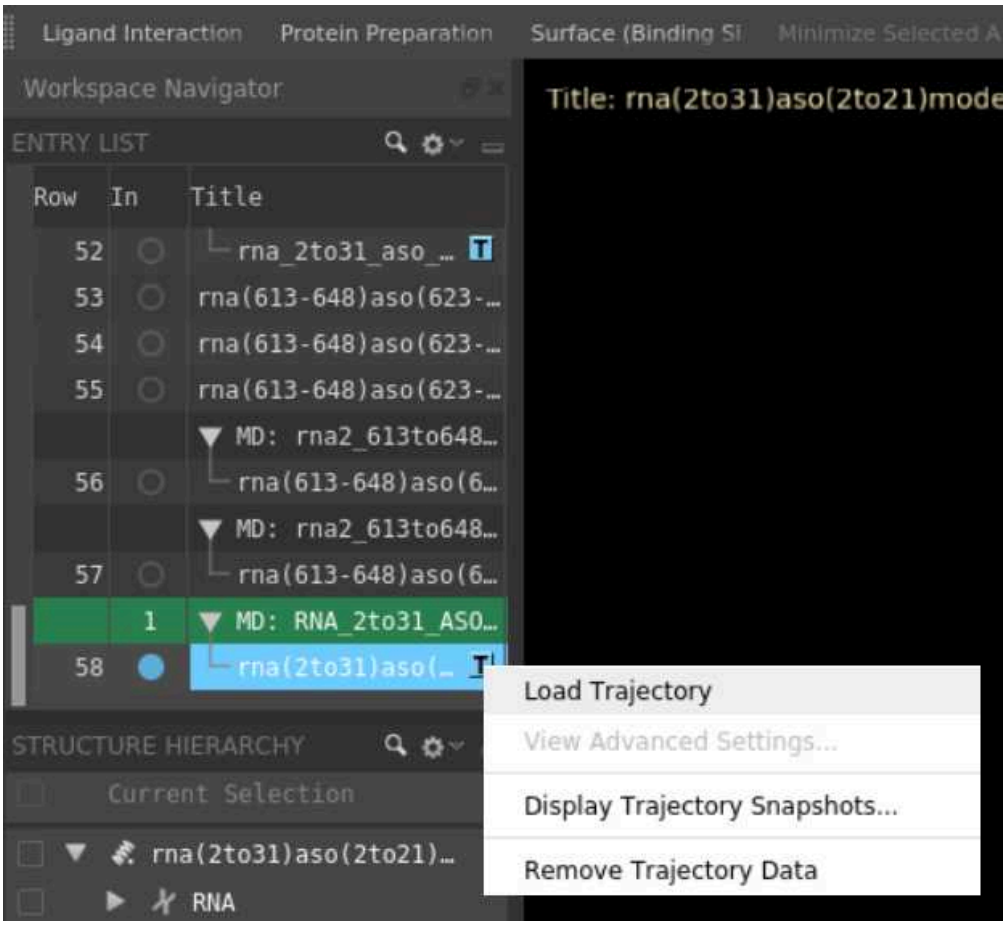

System Builder@cse-SYS-7049GP-TRT
Solvation
Ions
Set Up Membrane...
Delete Membrane
Solvent model
None
Predefined:
SPC
Custom:
Browse...
Boundary conditions
Box shape:
Orthorhombic
Box size calculation method:
Buffer
Absolute size
Distances (Å):
a:
10.0
b:
10.0
c:
10.0
Angles (°):
α:
90.0
β:
90.0
γ:
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



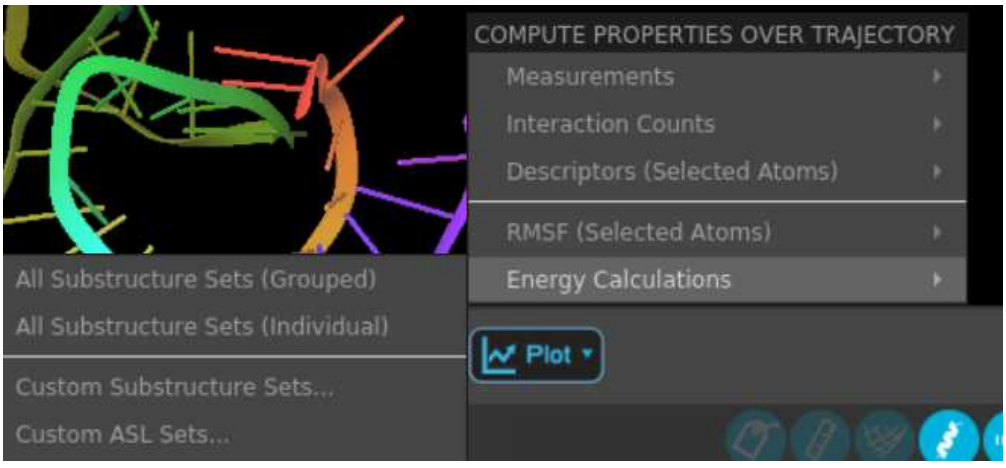
e.Open the Molecular Dynamics tool and load the minimized system obtained from the previous step from the workspace; Define the Simulation time. Select the Ensemble Class as NVT and change the temperature to 310.0K. Let the default values remain for the other parameters.



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