GrAfSS Help Page

Table of Contents

Browser Compatibility

OS	Version	Chrome	Firefox	Microsoft Edge	Safari
Linux	CentOS 7	71.0	61.0	n/a	n/a
MacOS	Mojave	71.0	61.0	n/a	12.0
Windows	10.0	73.0	38.0	42.17134.1.0	n/a
Android	8.0	73.0	66.0	n/a	n/a
iOS	12.2	n/a	n/a	n/a	5.2

Flowchart

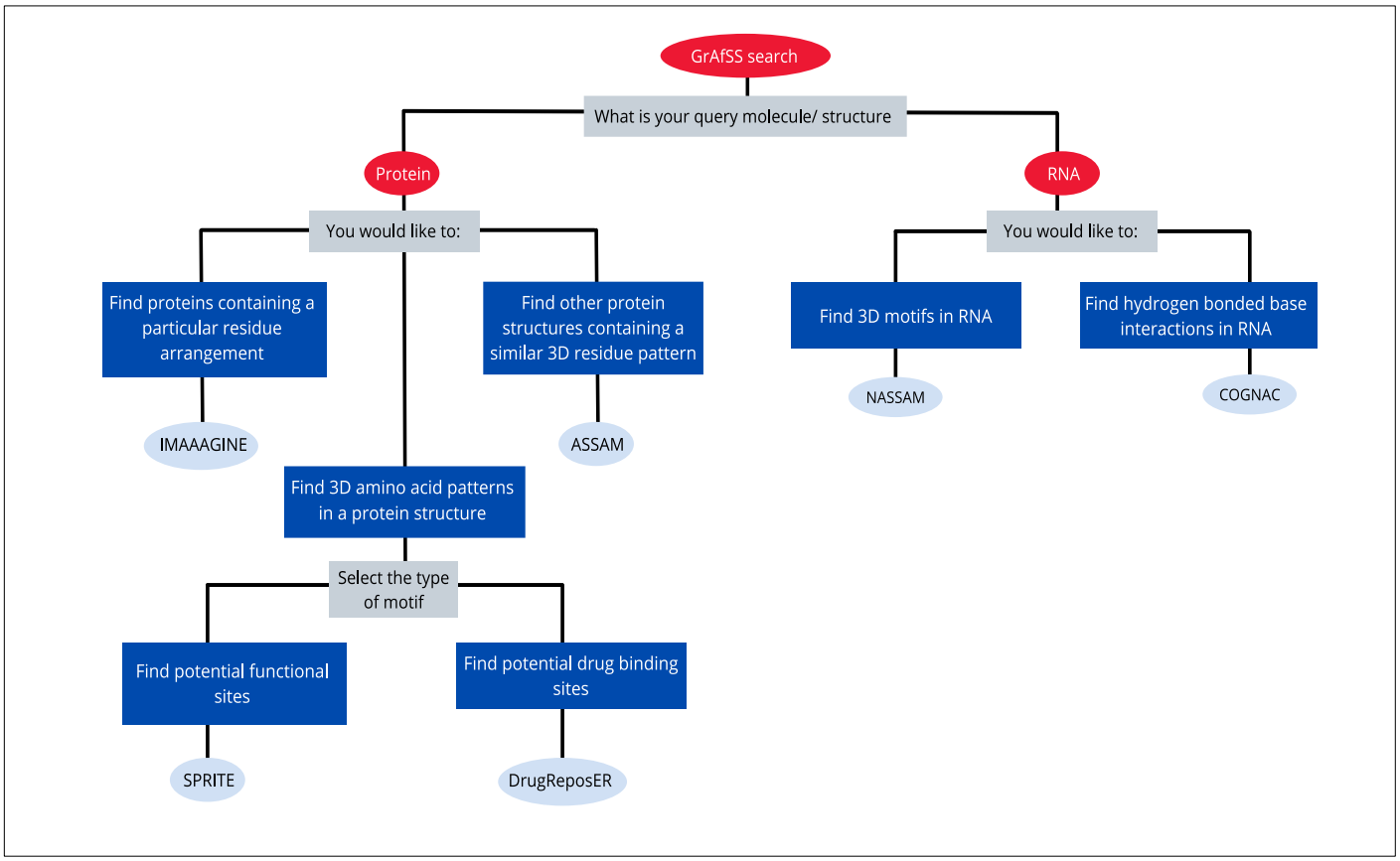

Figure 1 is an overview of the five different algorithms in GrAfSS: (1) SPRITE (2) Drug ReposER (3) ASSAM (4) IMAAAGINE for protein query; whereas (5) NASSAM and (6) COGNAC for RNA query.

Please follow through the selections to go to the desired algorithm's interface. If you wish to reset your search inputs, click on the *Reset search* button, or click on *Home* below the GrAfSS banner.

This help page is made to guide you on how to use and how to interpret the results.

Searching for Known Motifs or 3D Arrangements in a Query Protein Structure Using SPRITE

About SPRITE

- By using the SPRITE algorithm, you can find whether known motifs or 3D arrangements of amino acids are present in a query structure coordinate file. This would allow for the structural annotation of new structures that have been experimentally or computationally generated, such as by AlphaFold2. Moreover, this can be used for proteins of uncharacterized functions with no detectable sequence or fold similarity to any examples available in databases such as in UniProt or PDB.
- Summary of SPRITE algorithm:

Input query: A protein structure coordinate file.

Database: The query is searched against a database of known motifs such as catalytic sites, binding sites or other motifs that have been curated and added to the database.

Output: A list of potential 3D patterns found in the query protein structure.

SPRITE Job Submission Page

- To go to SPRITE job submission page, go to the GrAfSS home page and choose the options as shown in the flowchart.
- In the SPRITE job submission page, you can enter a four-letter protein PDB ID in the blank box or upload a protein PDB file. Let's type *4cha* in the blank box. Excluding 2-residue patterns is recommended. Click *Submit*.

SPRITE Result Page

- The 4cha result viewing option page like this will appear.
- The output can be viewed in three ways: (1) Right-handed superposition (RHS) list, (2) Left-handed superposition (LHS) list, and (3) One text file.
- There are three display options to view the RHS and LHS lists. At the top of each page, you can click on the respective button if you wish to change the display option:
    - List of hits
    - Full details

    - Arranged by sites

- **'List of hits'** result page:
    - Click here to go to RHS list of hits for 4cha.
    - To import the summary of SPRITE hits for 4cha to your machine, click on the dynamic *Download text version of the SPRITE output.*
    - **Select**
        - Click on the circle of an entry and click *Submit*. This will bring you to the visualization page where you can visualize the query (i.e., 4cha) and hit 3D amino acid arrangements. Click here to learn more about the visualization page.
    - **Hit**
        - E.g., 1ds2_c00. The first four characters represent the hit's PDB ID. The last three characters represent the SPRITE database motif ID.
    - **Source PDB ID**
        - Shows the PDB ID of the hit.
    - **Description**
        - Briefly describes what the PDB ID is.
    - **RMSD**
        - The hit-query superposition score in Angstrom (Å).
- **'Full details'** result page:
    - Click here to go to RHS full details hits for 4cha.
    - **Select**
        - Check the box and click *Submit* will bring you to another visualization page. Click here to learn more about the visualization page.
    - **Hit**.
        - E.g., 1ds2_c00. The first four characters represent the hit's PDB ID. The last three characters represent the SPRITE database motif ID.
    - **Description**
        - Briefly describes what the PDB ID is.
    - **RMSD**
        - The hit-query superposition score in Angstrom (Å).
    - **Number of residues**
        - The number of hit-query matching residues found.
    - **Send for ASSAM search**
        - Click on the dynamic *Send* of an entry to submit an ASSAM job using the respective matching query residues (i.e., 4cha) as the input. The ASSAM job will search for 3D amino acid arrangements similar to the input. You will be brought to an ASSAM result page like this.
    - **CCP4 TRANSFORM**
        - This transform command for CCP4 will be executed when you click on the *Superposed motif* button of the first NGL window in the visualization page. Read below.
    - **Matches**

        - List the hit residues that match with the corresponding query residues
- '**Arranged by sites**' result page:
    - **Select**
        - Check the box and click *Submit* will bring you to a visualization page. Click here to learn more about the visualization page.
    - **Site in input/ query structure**
        - Shows all residues in the query protein in ascending order that are matched to respective motifs.
    - **Hit(s) to sites in SPRITE database**
        - Shows the sites in hit proteins that are matched to the respective site in the query structure.
    - **Residues of hit(s)**
        - Shows the residues that make up the respective sites the hit proteins.
- **Download text version of the SPRITE output:**
    - Click here to see a sample.

SPRITE Visualization Page

Directed from 'List of hits' Result Page

- Click here to see a sample.
- The query residues (i.e., 4cha) are in CPK color.
- Some of the useful visualization features:
    - **Show pattern match**
        - To show the hit residues in green.
    - **Side chain**
        - To show the query side chain atoms in yellow.
    - **Label**
        - To show label the residue type and residue number of the query (i.e., 4cha)

Directed from 'Full details' and 'Arranged by sites' Result Pages

- Click here to see a sample.
- The first NGL window shows the query motif in yellow (i.e., 4cha).
- Some of the useful visualization features:
    - **Superposed motifs**
        - To show hit-query superposition. Hit residues are shown in green.
        - This is feature performs the transform command for CCP4.
    - **Backbone**
        - To show the backbone of the query protein (i.e., 4cha)
- In the second NGL window, you can better locate the potential motifs (shown as green balls) in the query structure. Click on the *Backbone* button will show the query (i.e., 4cha) backbone to further aid your visualization.

Search for Amino Acid 3D Arrangements in a Query Protein that are Similar to Approved Drug Binding Sites Using SPRITE- Drug ReposER

About Drug ReposER

- By using Drug ReposER, you can search and identify if the query amino acid side chain arrangements are similar to those found in drug binding sites. Hence, you can use Drug ReposER as a starting point to identify and explore potential targets for drug repurposing.
- Summary of Drug ReposER algorithm:

Input / Query file: PDB file or PDB ID.

Database: Precomputed database of sites that have a similar amino acid 3D arrangement to that of known drug binding sites.

Output: A list of potential 3D patterns similar to known drug binding sites found in the query protein structure.

Drug ReposER Job Submission Page

- To go to Drug ReposER job submission page, go to the GrAfSS home page and choose the options as shown in the flowchart.
- Please read the guides before uploading an input file on the job submission page.
- In the Drug ReposER job submission page, you can enter a four-letter protein PDB ID in the blank box or upload a protein PDB file. Let's type *5gkx* in in the blank box and click *Submit*.

Drug ReposER Result Page

- You will see a result viewing option page like this will appear. There are five ways to view your outputs:
  - List of both hits (LHS and RHS side by side)
  - List of all hits (combinations of RHS and LHS hits)
  - List of hits for lefthanded superpositions
  - List of hits for righthanded superpositions
  - One text file.
- Please read the *quick guide to understanding left-handed and right-handed superpositions*.
- While analyzing the result table on the webserver, you can click on the 'All hits' dropdown to filter the outputs by RMSD or residue number.
- The elements in the result table are:

- **Select**
  - Check the box of an entry and click *Submit* will show the potential drug binding site in the query structure as yellow residues in the NGL window on the same page. Some of the useful visualization features:
    - **Background**
      - To change the background of the NGL from black to white.
    - **Backbone**
      - To show the backbone of the query structure.
    - **Heteroatom**
      - To show the heteroatom present in the PDB structure.
    - **Superposed motifs**
      - To show the hit drug binding residues as green structure in the hit superposing the potential drug binding residues as yellow structure in the query.
    - **Hit pocket**
      - To show the binding pocket of the hit protein.
    - **Hit ligand**
      - To show the drug that targets the hit in CPK colors.
    - **Centre**
      - To centre the structure in the NGL window.
    - **Spin**
      - To spin the structure.
    - **Screenshot**
      - To save a screenshot of the NGL window in png format.
- **Hit**
  - E.g., 1bzm_1 (1BZM_A_MZMA262_1). The first four characters represent the hit PDB ID. Click on the dynamic characters in the parentheses (i.e., Drug ReposER ID) to view the hit residues in the visualization page.
- **Description**
  - Briefly describes what the hit PDB ID is.
- **RMSD**
  - The hit-query superposition score in Angstrom (Å).
- **Residue**
  - The number of matching hit-query residues.
- **Matches (Hit-Query)**
  - Shows the chain identifier, residue number, and residue type of matching hit and query respectively.
- **Superposition**
  - Shows the hit as either RHS or LHS.

Drug ReposER Visualization Page

- Click on the Drug ReposER ID, to go to the visualization page of the selected hit. Click on the Drug ReposER ID on the top center page to refresh the NGL window and display the hit (i.e., known drug binding site) as grey structures.

- Summary of the hit displayed:
  - **PDB ID**
    - Click on the dynamic PDB ID will bring you to its respective page on RCSB PDB.
  - **Macromolecule**
    - Briefly describes the macromolecule present in the PDB structure.
  - **Source Organism**
    - The name of organism where the PDB structure is present in.
  - **Pfam Annotation**
    - Shows the Pfam annotation.
  - **Ligand ID**
    - Click on the dynamic PDB ligand ID to go to its respective page on the RCSB PDB.
  - **Drugbank ID**
    - Click on the dynamic Drugbank ID to go to its respective page on Drugbank.
  - **HETATM residue**
    - Shows where the ligand is stored in the PDB file (as HETATM record).
  - **Number of binding residues**
    - Shows the total count of residues in the hit binding site.
- Some of the useful visualization features are:
  - **Pocket**
    - To show the binding pocket of the hit protein.
  - **Backbone**
    - To show the backbone of the hit structure.
  - **Ligand**
    - To show the ligand as magenta structure present the hit PDB file.
  - **Sidechain**
    - To show the side chain atoms in yellow.
  - **Label**
    - To label the binding site residues of the hit protein.
  - **Screenshot**
    - To save a screenshot of the NGL window in png format.
- To enlarge your NGL window, click on the dynamic *Click here for a larger view.*
- Integrated ASSAM search results are displayed in the lower part of the visualization page.

Similar Patterns of Amino Acids Derived from ASSAM Search

- The ASSAM search is performed by using the known drug binding site residues from the hit protein (i.e., The one whose information is being displayed in the visualization page) which produces a list of similar patterns to the known drug binding site.
- You can click on the ‘All hits’ dropdown to filter the ASSAM search:
  - All hits

- All hits from Homo Sapiens
- Sort hits by RMSD
- Only hits with RMSD < 1.0 Å
- Only hits with Dali Z-score < 2.0 Å
- Only hits with SeqID < 30%
- Only hits with exact matches

- The ASSAM result table lists hits of 3D amino acids pattern similar to the currently displayed known drug binding site on the same page:
- The elements in the ASSAM result table are:
  - **Drug repurposing remarks**

    Crosshair aim sign: Potential target for repurposing.

    i.e., matched protein structure has less than 30% sequence identity to known drug target and may potentially interact with the same drug molecule based on local structural similarity of binding site.

    Hollow aim sign: Available target for annotated drug.

    i.e., matched protein structure has more than 30% sequence identity to known drug target.

  - **View**
    - Checking the circle of an entry will display its drug-binding-like residue pattern as green structures in the NGL window on the same page.
  - **Hit**
    - Shows the PDB ID which contains the hit drug-binding-like residue pattern.
  - **Macromolecule/ Organism**
    - Briefly describes the macromolecule and which organism does the protein with drug-binding-like pattern come from.
  - **Pfam**
    - Clicking on an entry will bring you to its Pfam page.
  - **Residue**
    - The number of residues in the protein with drug-binding-like pattern that are matched to the known drug binding site.
  - **Interface**
    - Shows the type of residue, chain identifier, and residue number of each residue that makes up the drug-binding-like pattern.
  - **HETATM**
    - Shows the HETATM record of the ligand present in the hit PDB file.
  - **RMSD**
    - Shows the superposition score in Angstrom (Å) of the known drug binding site and its hit drug-binding-like pattern.

- **Dali Z-score**
  - Shows the pre-computed Dali Z-score of the structural alignment between the currently displayed known drug binding site and the respective hit drug-binding-like residue pattern. Z-score above 2.0 normally corresponds to the two aligned structures having similar folds.
- **Sequence identity (%)**
  - Shows the sequence similarity in percentage between the known drug binding site and its hit drug-binding-like pattern.
- **Dock**
  - Clicking on ‘Script’ will download a python docking script

Searching for an Amino Acid Arrangement of Interest (e.g., Ligand Binding Site, 3D Motif) against a Protein Structure Dataset Using ASSAM

About ASSAM

- If you have encountered an amino acid arrangement or a series of residues of known functional or structural importance, you can use the ASSAM algorithm to search for the residues in other proteins. ASSAM allows you to identify whether a similar 3D side chain arrangement or conserved residues are present in different protein structures, including those that have no fold similarities. Because the search is independent of fold or sequence similarity, it can allow very small functional regions to be identified that cannot otherwise be identified by fold similarity searching. Examples of such motifs include convergently evolved catalytic sites. Furthermore, you can also search for a drug binding site against human protein structures to identify whether a drug could have off-target sites that could cause potential side effects or toxicity.
- Summary of ASSAM algorithm:

Input query: A 3D arrangement of amino acid residues in PDB format containing 12 residues or less.

Database: The query can be searched against several types of datasets:

- Non-redundant (NR) PDB at 30% sequence identity cut-off (excluding mutant structures).
- NR-PDB at 35% sequence identity cut-off (excluding mutant structures).
- NR-PDB at 30% sequence identity cut-off (including mutant structures).
- NR-PDB at 35% sequence identity cut-off (including mutant structures).
- Alphafold human proteome.
- Alphafold *E. coli* proteome.
- Alphafold S. cerevisiae proteome.
- PDB excluding near duplicate structures.
- Structure keywords = Protease (ALL).

Output: A list of 3D protein structures containing the query pattern.

ASSAM Job Submission Page

- To go to ASSAM job submission page, go to the GrAfSS Home page and choose the options as shown in the flowchart.

- Please read the notes to guide you on uploading an input file on the job submission page. As an example, prepare a 2atq PDB-formatted input file containing only GLU 747, ARG 750, ARG 751, GLU 755 on your machine.
  To see a sample, click on *[Download sample input (2atqRERE.pdb)]*. You can review the file by opening it with your favorite text editor. To upload it to ASSAM, click Choose File and choose the downloaded file. Choose *Non-redundant (NR) PDB at 30% sequence identity cut-off (excluding mutant structures)* as the database to be searched against the input. Click *Submit.*

ASSAM Result Page

- A result page like this will appear. The output is displayed and classified into two tables: (1) Right-handed superposition (RHS), (2) Left-handed superposition (LHS).
- You can import the text version RHS and LHS results by clicking on the respective dynamic *Download text version of the ASSAM output*.
- When viewing the results on the table, you can sort by the table's element by clicking on the header of the respective column.
- Elements in the RHS and LHS tables:
  - **Matches found in query**
    - Shows the PDB ID of the hit protein.
  - **Description**
    - Briefly describes what the PDB ID is.
  - **Residues**
    - The type of query amino acid residues.
  - **Residue Matches > Query**
    - Shows the chain identifier and the residue number of the input file submitted by user.
  - **Residue Matches > Database Hits**
    - Shows the chain identifier and the residue number of hits matched with the query.
  - **Heteroatoms Notes in Database hit**
    - Shows the distance in Angstrom between the ligand and the respective hit residue.
  - **RMSD**
    - The hit-query superposition score in Angstrom (Å).
  - **Viewer**
    - Click on the dynamic *Submit* of an entry to go to the respective visualization page where you can visualize the hit residues.

ASSAM Visualization Page

- Click here for a sample page.
- The hit residues are shown in CPK color in the NGL window.
- Some of the useful visualization features are:
  - **Background**

    - To change the background of the NGL from black to white.
- **Backbone**
    - To show the backbone of the hit protein.
- **Sidechain**
    - To show the side chain atoms in yellow.
- **Heteroatom**
    - To show the heteroatoms present in the PDB structure.
- **Label**
    - To label the residue type and residue number.
- **Transform**
    - To show the superimposition with query residues.

Designing a Theoretical 3D Amino Acid Arrangement and Searching for Its Presence in the PDB Protein Structures Using IMAAAGINE

About IMAAAGINE

- By using IMAAAGINE, you can design your own 3D amino acid arrangement query of eight residues or less, and then search whether it can be found in certain PDB and Alphafold protein structure datasets. The input design interface has menu-based selection and wild-card capability. When designing the query, you should utilize your understanding of the biochemical requirements for a particular function so you to specify the query's residue types as well as the inter-residue distances accordingly.
- This search allows you to explore the PDB and Alphafold protein structures for the presence of a 3D amino acid arrangement of a specific chemistry (e.g., find where the region that contains a cluster of basic amino acids in certain proteins). Moreover, you can also use IMAAAGINE to identify novel motifs that are only detectable in a 3D context (e.g., identify a single Histidine from different monomeric subunits that function as an interface for quaternary structure assembly).
- Summary of IMAAAGINE algorithm:

Input query: Self-designed 3D amino acid arrangement of eight residues or less.

Database: The query can be searched against several types of datasets:

- NR-PDB at 30% sequence identity cut-off.
- Biological assemblies at 30% sequence identity cut-off.
- Aplhafold human proteome.
- Alphafold *E. coli* proteome.
- Alphafold *S. cerevisiae* proteome.
- Biological assembly (PISA)
- Structure keywords = Protease (ALL)

Output: A list of protein structures and respective 3D amino acid arrangements that conform to the query.

IMAAAGINE Job Submission Page

- To go to the IMAAAGINE job submission page, go to the GrAfSS Home page and choose the options as shown in the flowchart.
- Please read the guides before you follow these steps to design your query:

Step 1: Select the number of amino acid residues and the pattern type (ie. General or one residue surrounded by other residues).

Step 2: The inter-residue distance tolerance is set at 1.5 Å by default. Specify a different value, if you wish.

Step 3: Select the database to be searched against.

Step 4: Click the *Continue* button. You will then see a query model that you need to design further.

Step 5: Click on the drop-down button in each residue to select your desired residue type.

Step 6: Fill in the distance boxes to specify the inter-residue distances. Make sure that all residues are connected.

Step 7: Click the *Submit* button to submit your job.

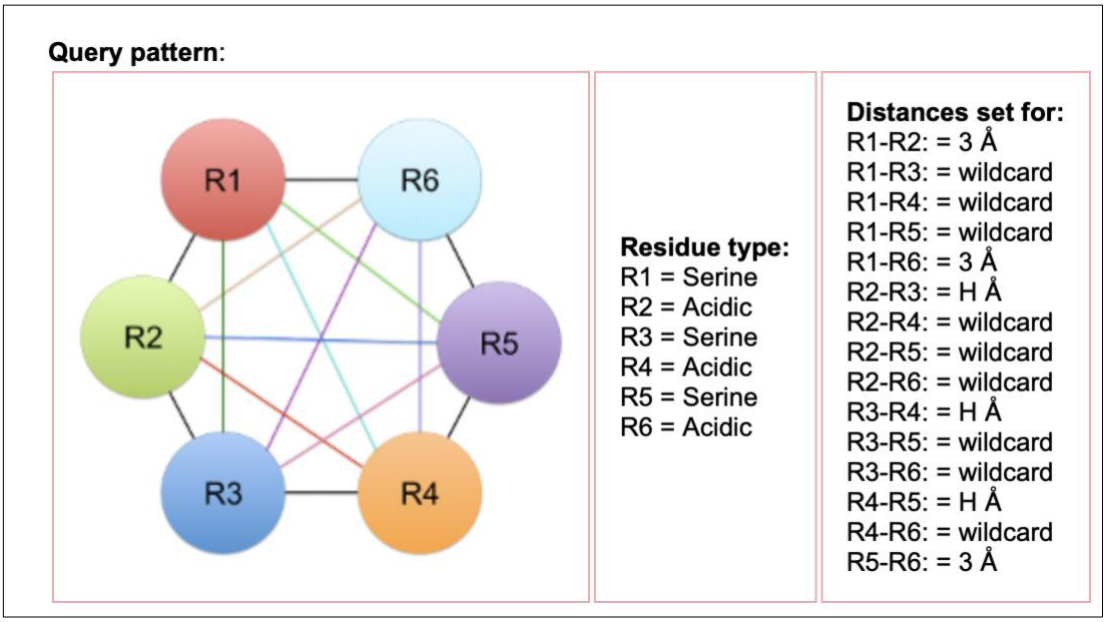

Figure 2 shows a sample IMAAAGINE query consisting of 6-residue of alternating Serine and Acidic residues.

- As an example, let's design a query like in Figure 2 and searched it against NR-PDB at 30% sequence identity cut-off.

IMAAAGINE Result Page

- A result page like this will appear.

- You can analyze the outputs on the webserver via the result table. To sort by the table's element, click on the header of the respective column.
- By default, the result table displays all hits. You can choose to display only hits on the same chain by clicking on the dropdown button under the 'Filter hits by'.
- Elements in the result table:
    - **PDB ID**
        - Shows the PDB ID of the hit protein.
    - **Structure Description**
        - Briefly describes what the PDB ID is.
    - **RMSD**
        - The hit-query superposition score in Angstrom (Å)
    - **Resolution (Å)**
        - Shows the structure resolution of the hit.
    - **Residues**
        - Shows the chain identifier, residue number, and residue type of the hit e.g., A 550 GLN
    - **Distances to HETATMs**
        - Shows the distance from the hit residue to the corresponding nearby heteroatom
    - **Nearby HETATMs**
        - The nearest heteroatom presents near the corresponding hit residue, within 5 (Å).
    - **Viewer**
        - Click on the dynamic *Submit* to go to the respective hit visualization page.

IMAAAGINE Visualization Page

- You can visualize the 3D arrangement of the hit amino acid residues of your selected entry in the NGL window.
- Some of the useful visualization features are:
    - **Background**
        - To change the background of the NGL from black to white.
    - **Backbone**
        - To show the backbone of the hit protein.
    - **Sidechain**
        - To show the side chain atoms in yellow.
    - **Heteroatom**
        - To show the heteroatoms present in the PDB structure.
    - **Label**
        - To label the residue type and residue number.
    - **Spin**
        - To spin the structure.
    - **Screenshot**
        - To save a screenshot of the NGL window in png format.

Searching for 3D Motifs and Patterns of Base Arrangements in RNA (and RNA associated) Using NASSAM

About NASSAM

- Find whether known motifs or 3D arrangements of RNA bases are present in a query structure coordinate file. This would allow for the structural annotation of new structures that were either experimentally or computationally generated. The Nucleic Acid Search for Substructures and Motifs (NASSAM) program searches can annotate a newly solved structure for the presence of known base motifs.
- NASSAM program accepts three-dimensional RNA crystallographic structures formatted as PDB files as input queries to search against a database of 3D base arrangements that include base pairings, base triple arrangements, A-minor motifs, kink-turns, ribose-zippers, tetraloops and T-loop motifs. The motif database consists of graph representations of base arrangements designed from patterns reported in literature.
- Summary of NASSAM algorithm:

Input query: PDB formatted RNA structure file.

Database: The query is searched against a database of RNA base arrangements and 3D motifs.

Output: A list of potential 3D motifs in the query RNA structure.

NASSAM Job Submission Page

- To go to NASSAM job submission page, go to the GrAfSS home page and choose the options as shown in the flowchart.
- Please read the guides before uploading an input file on the job submission page. As an example, prepare your input by downloading the PDB-formatted structure file of 6tna from the RCSB PDB website to your machine. To upload it to NASSAM, click *Choose file* and choose the downloaded 6tna.pdb. You will then see an interface like in Figure 3.

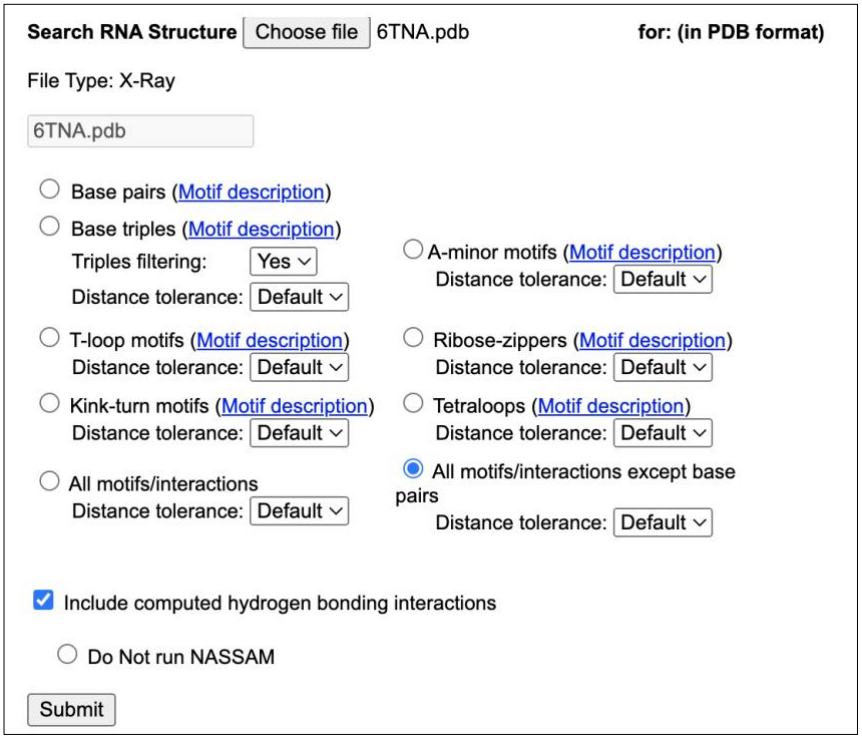

Figure 3 shows an interface where you can customize your NASSAM search.

- Select the motifs or interactions in the input query that you want to search for:
  - **Base pairs**
    - A base-pair consists of two nitrogenous bases interacting with each other via hydrogen bonds. It can be divided into two types: canonical base-pair and non-canonical base pair. In the canonical RNA Watson Crick pairing, adenine (A) forms a base-pair with uracil (U) while guanine (G) forms a base-pair with cytosine (C). Alternate hydrogen bonding patterns between bases lead to the formation of non-canonical base pairs such as wobble GU base pair and Hoogsteen base-pair (1,2) .
  - **Base triples triple filtering**
    - A base-triple is a widespread tertiary interaction that consists of three bases interacting via hydrogen bonds in a planar orientation. For triples in the NASSAM database, each nucleotide base is interacting with other bases by at least two hydrogen bonds (3,4).
  - **T-loop motifs**
    - The T-loop, originally characterized in tRNA is a five-nucleotide motif composed of a U-turn flanked by a non-canonical base pair (5).
  - **Kink-turn motifs**
    - The K-turn is a ~15 nucleotides motif which is formed by two strands in a helix-internal loop-helix arrangement (6). The vector patterns for kink-turns were designed using the three base pairs flanking the internal-loop (-1b, -1n, 1b, 1n, 2b, and 2n). Therefore, the hits returned represent the six nucleotides and not the entire motif.

    - **A-minor motifs**
        - The A-minor motif is formed when an adenosine interacts with the minor-groove of a receptor base pair, which is usually a Watson-Crick base pair. The A-minor motif can be divided into 4 types based on the position of the inserted adenosine relative to the receptor base pair (7). Currently NASSAM will only search for Type 0 and Type I A-minor motifs. The representations of other A-minor motifs are in progress.
    - **Ribose-zippers**
        - The ribose-zipper interaction is characterized by consecutive hydrogen-bonding interactions between ribose 2'-hydroxyls from different regions of an RNA chain or between RNA chains (8).
    - **Tetraloops**
        - A tetraloop is a four-nucleotide hairpin loop in RNA secondary structure that caps many helices. They can generally be divided into three sequence motifs: the 'UNCG', the 'GNRA' and the 'CUYG' tetraloops. In these sequences, R stands for a purine base (R= A or G), Y stands for a pyrimidine base (Y = U or C) and N indicates the position that can be any base. (N = A, U, G, C) (9–11).
    - **All-motifs/interactions**
    - **All-motifs/interactions except base pairs**
        - This is the default setting.
- Select the distance tolerance of your search:
    - Default
    - 10%
    - 15%
    - 20%
    - 40%
    - 50%
    - 60%
- By default, NASSAM will search for all motifs or interactions (except base pairs) with an optimized value of distance tolerance in your input query.
- The results will *include computed hydrogen bonding interactions*. Uncheck the box if you wish otherwise.
- Once ready, click *Submit*

NASSAM Result Page

- A result page like this will appear.
- Summary of NASSAM hits for 6tna is shown in a table. You can import the text file results to your machine by clicking *Download text version of the NASSAM output*.
- Elements in **Summary of NASSAM hits for 6tna** table are:
    - **Select**
        - Checking the box and click *Submit* will bring you to the visualization page. You can check more than one boxes and click *Submit*.

    - **Hits**
        - Shows the chain identifier in parentheses, residue type and residue number.
    - **Pattern/ Motif Information**
        - Describes the hit pattern or motif in the query.
- If you previously checked the box to include computed hydrogen bonding interactions, summary of hydrogen bonding interactions for 6tna will be displayed. To import the text file results to your machine, click *Download text version of the HBPRED output*.

NASSAM Visualization Page

- You can visualize the 3D arrangement of the hit residues in the NGL window.
- If you selected more than one hit previously, each hit will be colored differently.
- Click 'Please choose an interaction' dropdown to select the hits that you want to view in detail.
- Some of the useful visualization features are:
    - **Backbone**
        - To hide the RNA backbone.
    - **Label**
        - To label the hit residues.
    - **Hydrogen bonds**
        - To display the hydrogen bond present in the structure.
    - **Bases**
        - To emphasize the bases of the hit.
    - **Background**
        - To change the background of the NGL from black to white.
    - **Spin**
        - To spin the structure.
    - **Screenshot**
        - To save a screenshot of the NGL window in png format.
- If present, hydrogen bond will be shown in the other window

References

1. Xin Y, Olson WK. BPS: a database of RNA base-pair structures. Nucleic acids research [Internet]. 2009 [cited 2022 Mar 27];37(Database issue). Available from: https://pubmed.ncbi.nlm.nih.gov/18845572/

2. Tinoco JrI, Gesteland RF, Atkins JF. In The RNA World. Cold Spring Harbor Laboratory Press. 1993;603–7.

3. Firdaus-Raih M, Harrison AM, Willett P, Artymiuk PJ. Novel base triples in RNA structures revealed by graph theoretical searching methods. BMC Bioinformatics

[Internet]. 2011 Nov 30 [cited 2022 Mar 27];12(SUPPL. 13):1–13. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-S13-S2

4. Walberer BJ, Cheng AC, Frankel AD. Structural diversity and isomorphism of hydrogen-bonded base interactions in nucleic acids. Journal of molecular biology [Internet]. 2003 Apr 4 [cited 2022 Mar 27];327(4):767–80. Available from: https://pubmed.ncbi.nlm.nih.gov/12654262/

5. Krasilnikov AS, Mondragón A. On the occurrence of the T-loop RNA folding motif in large RNA molecules. RNA [Internet]. 2003 Jun 1 [cited 2022 Mar 27];9(6):640. Available from: /pmc/articles/PMC1370430/

6. Klein DJ, Schmeing TM, Moore PB, Steitz TA. The kink-turn: a new RNA secondary structure motif. The EMBO Journal [Internet]. 2001 Aug 1 [cited 2022 Mar 27];20(15):4214. Available from: /pmc/articles/PMC149158/

7. Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. Proceedings of the National Academy of Sciences of the United States of America [Internet]. 2001 Apr 24 [cited 2022 Mar 27];98(9):4899–903. Available from: www.pnas.orgcgidoi10.1073pnas.081082398

8. Tamura M, Holbrook SR. Sequence and Structural Conservation in RNA Ribose Zippers. Journal of Molecular Biology. 2002 Jul 12;320(3):455–74.

9. Correll CC, Swinger K. Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 Å resolution. RNA [Internet]. 2003 Mar 1 [cited 2022 Mar 27];9(3):355–63. Available from: http://rnajournal.cshlp.org/content/9/3/355.full

10. Jucker FM, Pardi A. Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. Biochemistry [Internet]. 1995 [cited 2022 Mar 27];34(44):14416–27. Available from: https://pubmed.ncbi.nlm.nih.gov/7578046/

11. Cheong C, Varani G, Tinoco I. Solution structure of an unusually stable RNA hairpin, 5GGAC(UUCG)GUCC. Nature 1990 346:6285 [Internet]. 1990 [cited 2022 Mar 27];346(6285):680–2. Available from: https://www.nature.com/articles/346680a0

Identifying the Presence of Hydrogen Bonds Within Specific Cluster of RNA bases Using COGNAC

About COGNAC

- By using COGNAC, you can find whether a specific cluster of RNA bases that are connected by hydrogen bonds are present or not in structure coordinate data containing RNA chains, such as the ribosomal subunits or ribozymes. COGNAC allows you to identify minute conformational differences in RNA structures that are a result of differences in hydrogen bonding. Moreover, you can also use COGNAC to identify highly stable clusters of bases that are held together in a specific arrangement by hydrogen bonds.
- Summary of COGNAC algorithm:

Input query: Is designed by defining how RNA bases can be interconnected by hydrogen bonds to form a cluster of bases from base pairs to sextuples.

Database: The query is searched against high resolution subsets of RNA structures from the PDB.

Output: Hydrogen bonded base connections in query RNA structure.

COGNAC Job Submission Page

- To go to COGNAC job submission page, go to the GrAfSS home page and choose the options as shown in the flowchart.
- Please read the guides before uploading an input file on the job submission page. Follow these steps to submit a job.

Step 1: Choose one of the options:

- search for a specific set of hydrogen bonded base connections in a database of RNA structures (sourced from the PDB), or
- upload an RNA structure (in PDB format) to annotate the hydrogen bonded base interactions present in the query structure, or
- upload two RNA structures and compare the hits

Step 2: Select the number of residues and the arrangement of the interactions for the query patterns from the list provided (2-6 bases or select ‘All patterns’ for arrangement types).

Step 3: For specific patterns selected in Step 2, enter the specific residue or general type (for all base possibilities).

For searches using Option 3; the difference is mainly in the requirement to have two structures uploaded. Please note that the program does not filter out highly dissimilar structures which are not useful comparisons.

- As an example, let's check the circle of *Upload an RNA structure*, click *Choose file* and upload a 6tna.pdb file. Click *Submit*. Choose Type I, 2 residues. Click *Continue*. Click on the dropdown menu to choose *Any Base* for the two residues. Click *Submit*.

COGNAC Result Page

- A result display option page like this will appear. Click on the dynamic *View hits* to view the result table of your COGNAC job. You will see a page like this.
- The elements in COGNAC result table:
  - **PDB ID**
    - Shows the PDB ID of the hit RNA.
  - **Structure description**
    - Describes what the PDB ID is.
  - **Resolution (Å)**
    - Shows the PDB structure resolution in Angstrom (Å).
  - **(Chain) Residue Number + Residue Type**
    - Shows the chain identifier in parentheses, residue number and residue type of the donor atom and acceptor atom.
  - **View**
    - Check the box of an entry and click *Submit to view in NGL* to go to the respective hit in the visualization page.

COGNAC Visualization Page

- You can visualize the 3D arrangement of the hit residues in the NGL window.
- Click the *Please choose an interaction* dropdown to select the hit that you want to view in detail.
- Some of the useful visualization features are:
  - **Backbone**
    - To hide the RNA backbone.
  - **Label**
    - To label the hit residues.
  - **Hydrogen bonds**
    - To display the hydrogen bond present in the hit.
  - **Bases**
    - To emphasize the bases of the hit.
  - **Background**
    - To change the background of the NGL from black to white.

    - **Spin**
        - To spin the structure.
    - **Screenshot**
        - To save a screenshot of the NGL window in png format.
- If present, hydrogen bond will be shown in the other window.
- 2D structure of the hit is displayed in the lower part of the page.