FR3D User's Manual

Table of Contents

Installation

Running FR3D from Matlab

Installation: FR3D was written in Matlab version 7.1 and has been run successfully on PC, Macintosh, and UNIX platforms. The easiest way to install it is to download the latest version of the Starter Package from http://rna.bgsu.edu/FR3D . This will create a folder named **FR3D** and several subfolders. It has pre-computed data for four large PDB files, results of a few **FR3D** searches, and the **FR3D** source files, data files, and PC executables.

If you already have a **FR3D** installation, download the current version of the Matlab code, unzip, and copy the new program files over the old ones in the **FR3DSource** folder.

If you have another folder on your computer with PDB files, add that folder to Matlab's path (File, Set Path, Add Folder, Save). The first time **FR3D** is asked to search a given PDB file, it reads the text, analyzes it, and saves a data file in the subfolder **PrecomputedData**. After that, it will not need to re-read the original PDB file. New versions of **FR3D** may need to re-analyze the 3D structure data stored in **PrecomputedData** to take advantage of new classifications.

Launching the Graphical User Interface (GUI): Start Matlab and change the working directory to **FR3D** (you can use the cd command to change the directory).   At the command prompt >>, type FR3D to launch the graphical user interface

Matlab 6 users: The program has been lightly tested with Matlab 6. The **.mat** data files distributed with the programs are saved in a Matlab 7.1 format. In the **FR3DSource** folder, delete **PDBInfo.mat** and rename **PDBInfo_version_6.mat** as **PDBInfo.mat**, and similarly with **PairExemplars.mat**. Delete the **.mat** files in **PrecomputedData**, and download the **.pdb** files from the Protein Data Bank and place in **PDBFiles**.

Using the compiled version of FR3D for the PC

Installation: First download and install the Matlab MCR installer (100 MB). MCR stands for Matlab Component Runtime. It lets you run compiled Matlab programs without purchasing Matlab. You can read about it at the Matlab website. Then download the latest Starter Package and follow the instructions above. This will create a folder named FR3D and several subfolders. The executable **FR3D.exe** will be in the **FR3D** folder.

Running FR3D: Double click the executable file **FR3D.exe** to launch the graphical user interface to **FR3D**.

Step-By-Step Tutorial on Performing Motif Searches using FR3D

In this tutorial, we will take you through a step-by-step approach to performing a Sarcin/ricin motif search, using purely geometric, symbolic and mixed parameters. Each step will have a red arrow (➡), which will direct the user to the step which is being explained. For each search we will be focusing on a sub-motif of the Sarcin/ricin shown below. The six nucleotides are 2701, 2702, 2703, 2691, 2693, and 2694 from PDB file 1s72. The interactions involved are G2701/A2694 – trans Sugar Edge/Hoogsteen, A2702/U2693 – trans Hoogsteen/Watson-Crick, and A2703/A2691 – trans Hoogsteen/Hoogsteen.
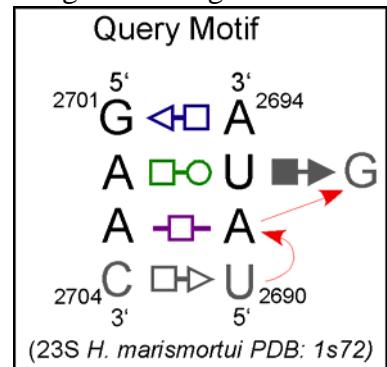

(Leontis, et al., 2002; Leontis and Westhof, 2001)

## *Performing a Purely Geometric Search with FR3D*

1. The first step is to click on the radio button which says Geometric + symbolic search.
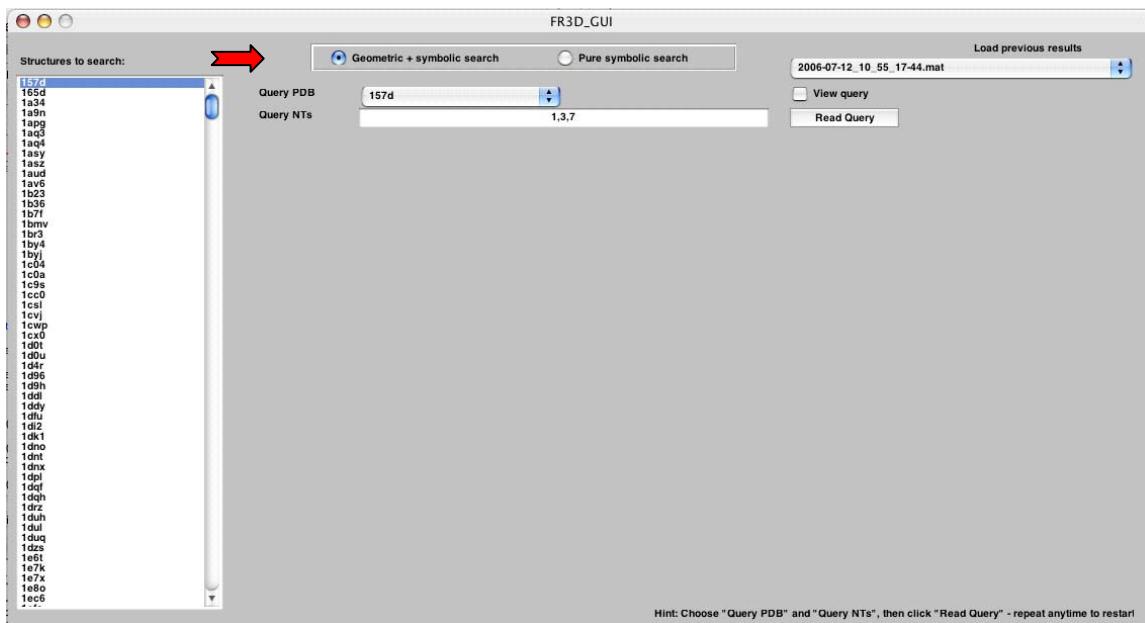
2. The user must specify the PDB file which contains the known motif from the drop-down menu labeled Query PDB. (e.g. 1s72)
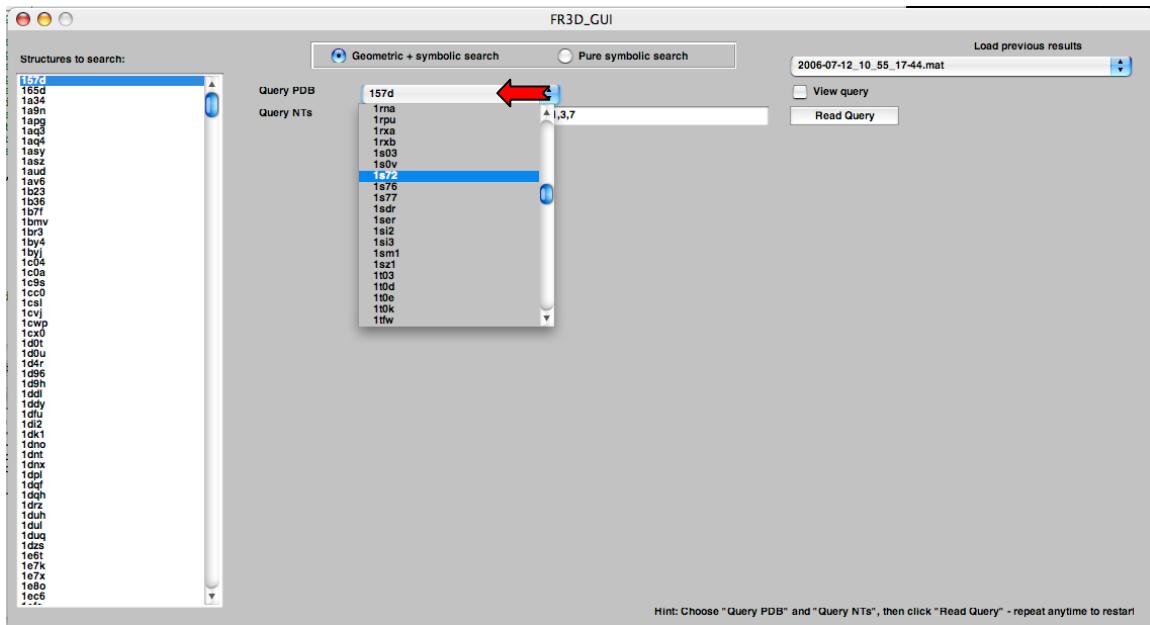
3. The query nucleotides are entered into the text-box, labeled by Query NTs. (e.g. 2701, 2702, 2703, 2691, 2693, 2694. Nucleotide numbers may be separated by commas, spaces, or semicolons. A range of nucleotide numbers may be indicated with a colon, as in 2701:2703. Ranges may be increasing or decreasing. The chain may be indicated with the syntax 2701(0) or 2701_0, or it may be specified later; see below.
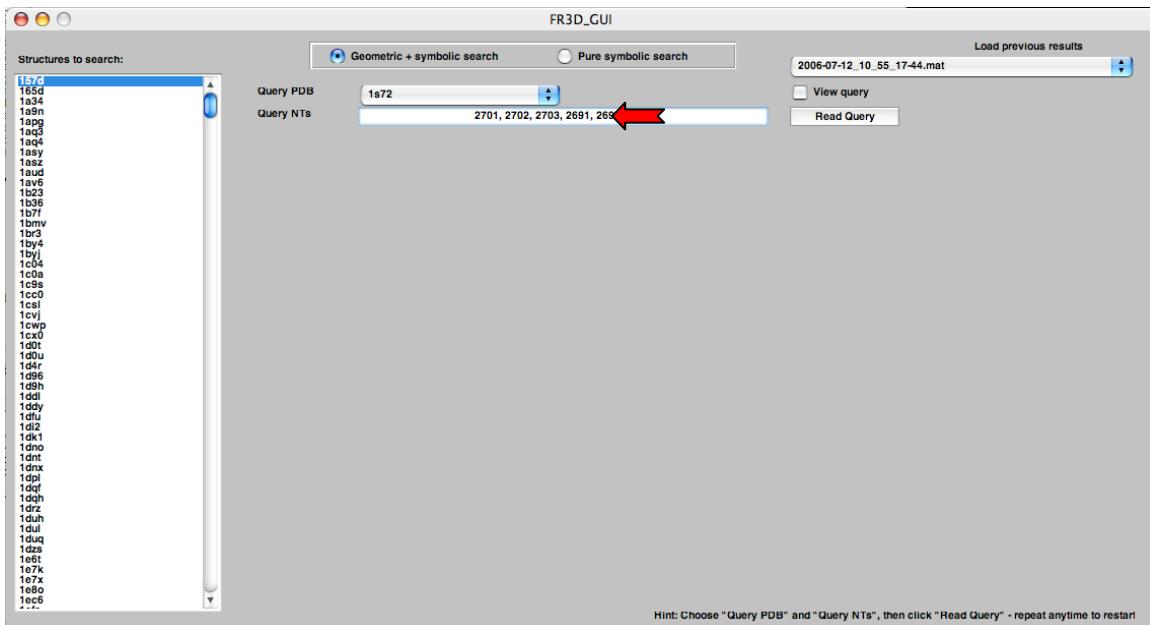
4. The user tells FR3D to read the crystal structure information about the Query motif by pressing the button labeled Read Query. If the user would like to view the motif they inputted, they may check the check-box labeled View query and then press the Read Query button and new figure will pop-up to show the user the Query motif. The interactions present will be displayed in the console window; this will be explained in more detail below.
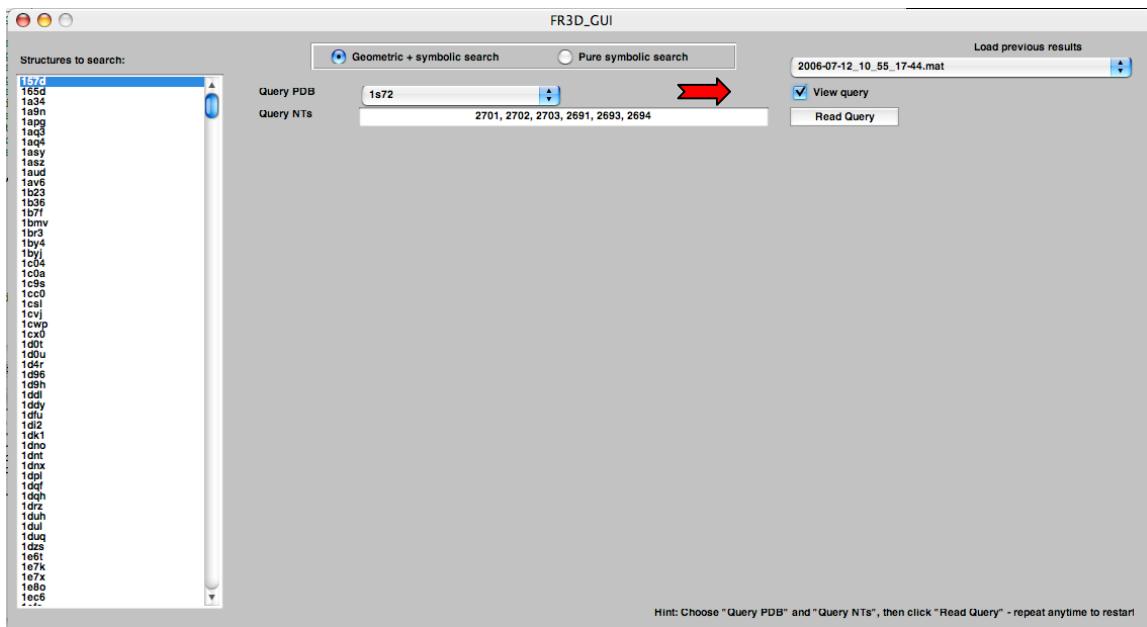
The figure below shows the Query motif, which is displayed by checking the Read query checkbox. It may be rotated in the figure window.
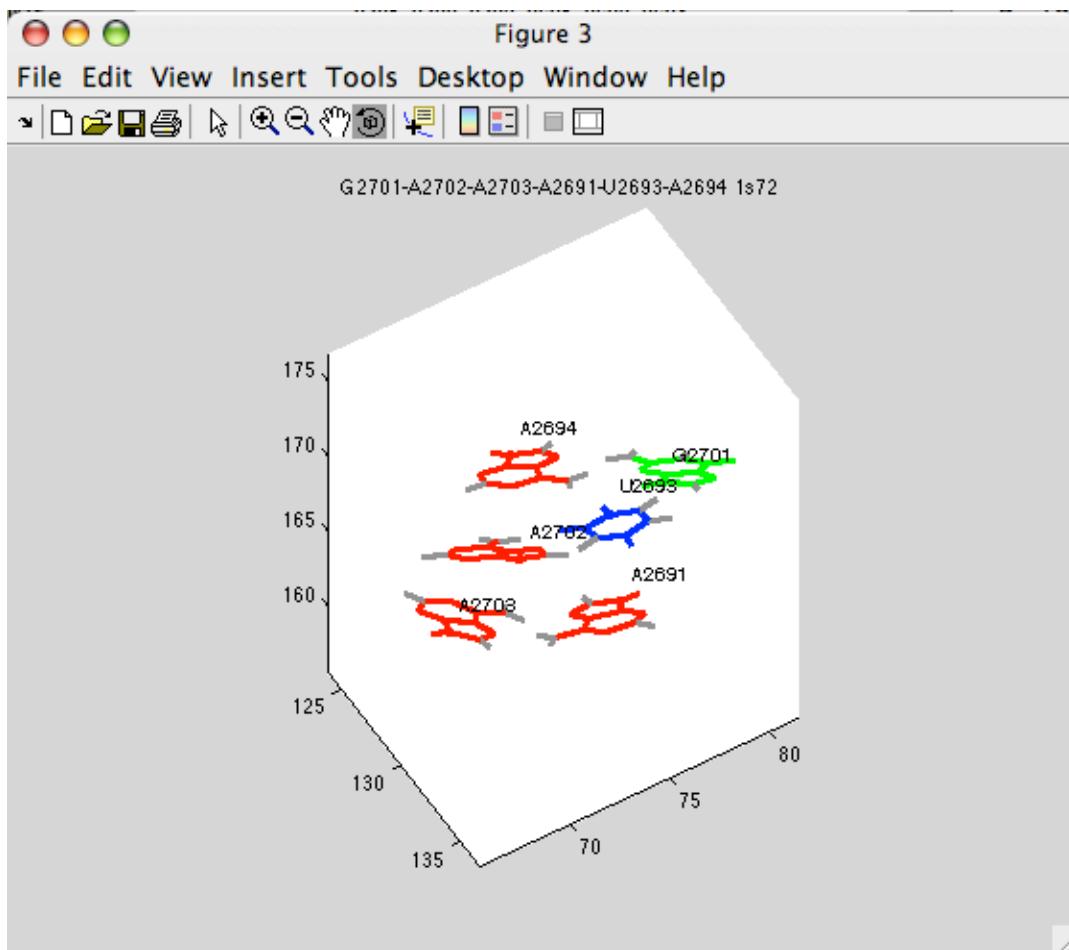
5. Some PDB files have multiple RNA chains.   For instance, 1s72.pdb contains a 5S chain and a 23S chain, and both chains have some of the same nucleotide numbers. If there is any ambiguity in the chain for the Query nucleotides, the user must specify the chain which contains the Query motif. This may be selected using the drop-down menu labeled Query Chains. The order of the drop-down menus corresponds to the order of the nucleotides supplied by the user. In this case we are using chain '0'.
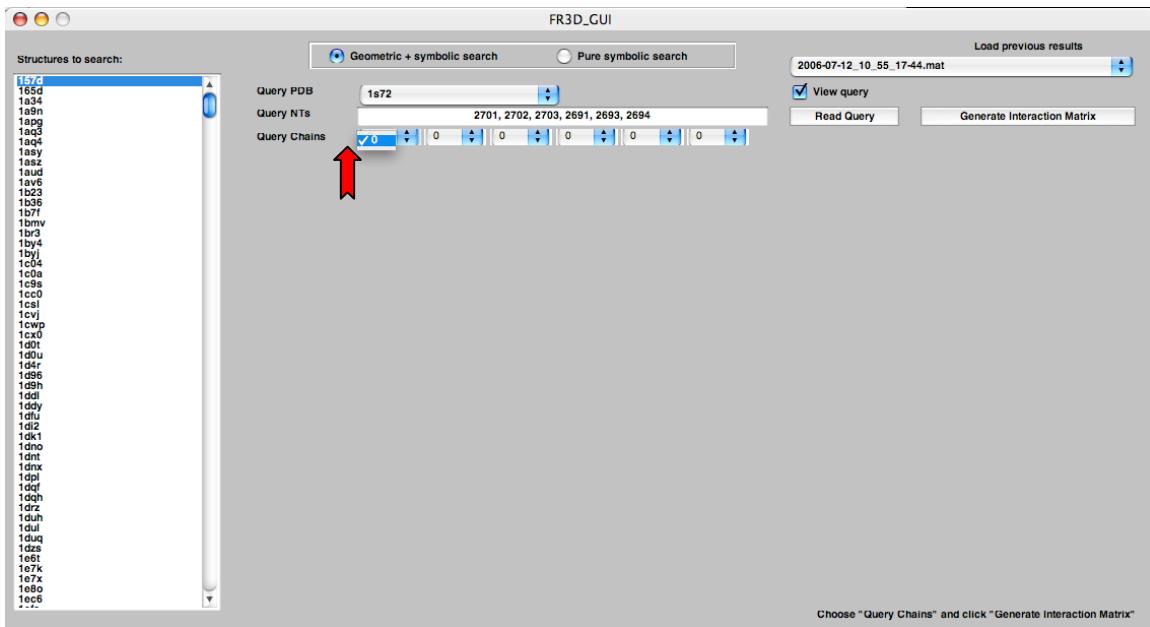
6. Once the chains have been selected the user should press the Generate Interaction Matrix button. The Interaction Matrix allows the user to impose certain types of constraints on the search. A purely geometric search makes no such constraints. Below we describe mixed geometric and symbolic searches and purely symbolic searches.
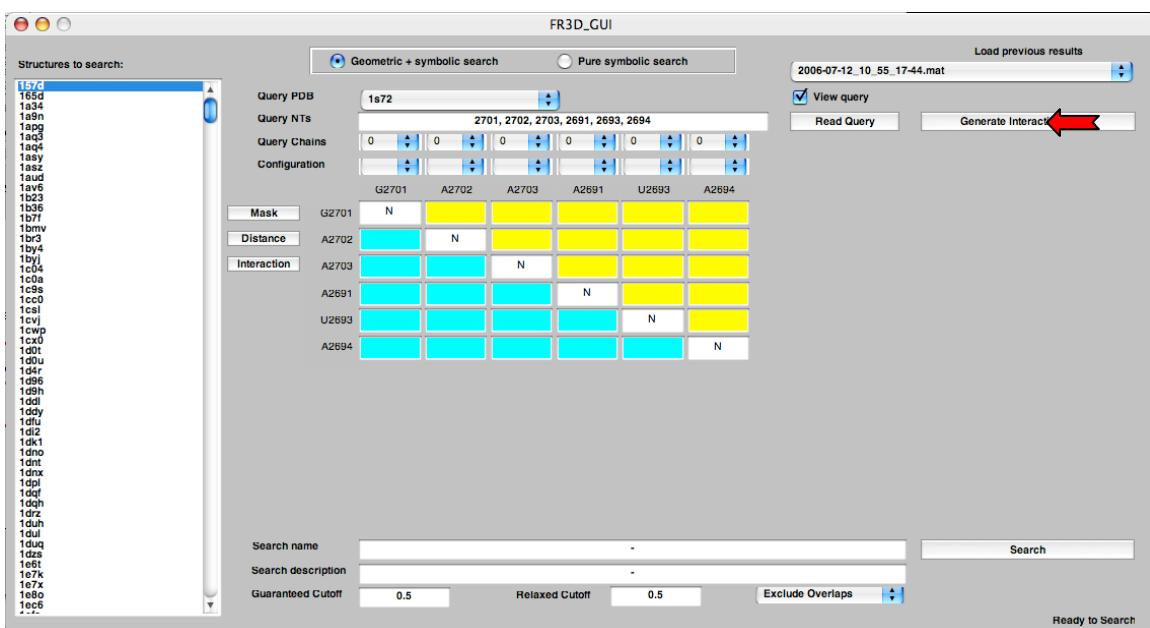
7. The user should give a name to the search in the text-box labeled Search name (e.g., Sarcin-ricin Motif). This will become part of a filename, so the name should not use characters such as ":", "?", "/" or "\", because these have meanings in filenames and paths.
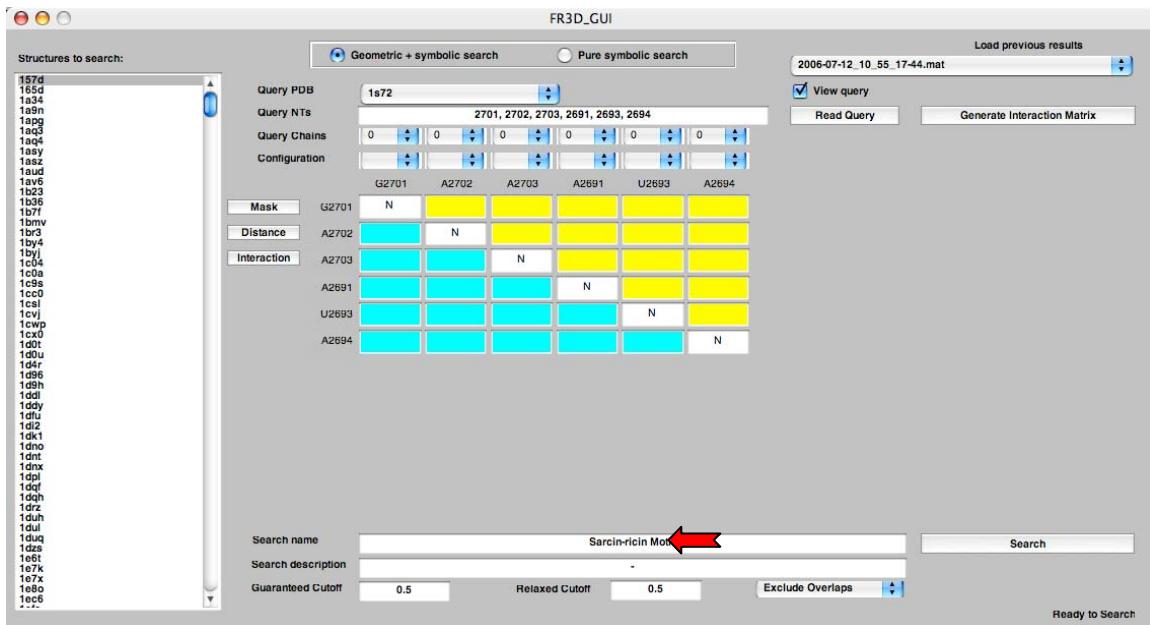
8. The user can add more descriptive information about their search in the text-box labeled Search description (e.g., Geometric – Sarcin/ricin sub-motif comprised of 6 NTs from 23S Haloarcula marismortui). Other comments about the search can be added here as well.
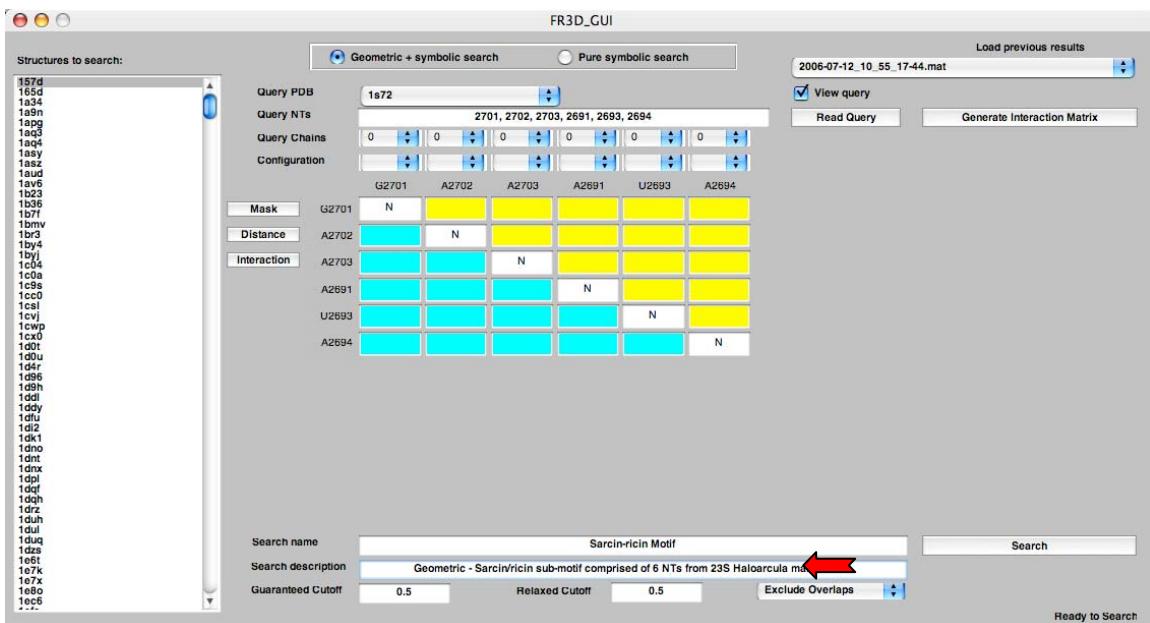
9. The user sets the Guaranteed Cutoff discrepancy in the text-box labeled Guaranteed Cutoff (i.e. 0.5). The search algorithm is guaranteed to find all candidates whose geometric discrepancy with the Query motif is less than this number. The discrepancy is roughly comparable to RMS discrepancy. Increasing the value of the guaranteed cutoff will rapidly increase the running time of the program. Values above 1.0 are often impractical.
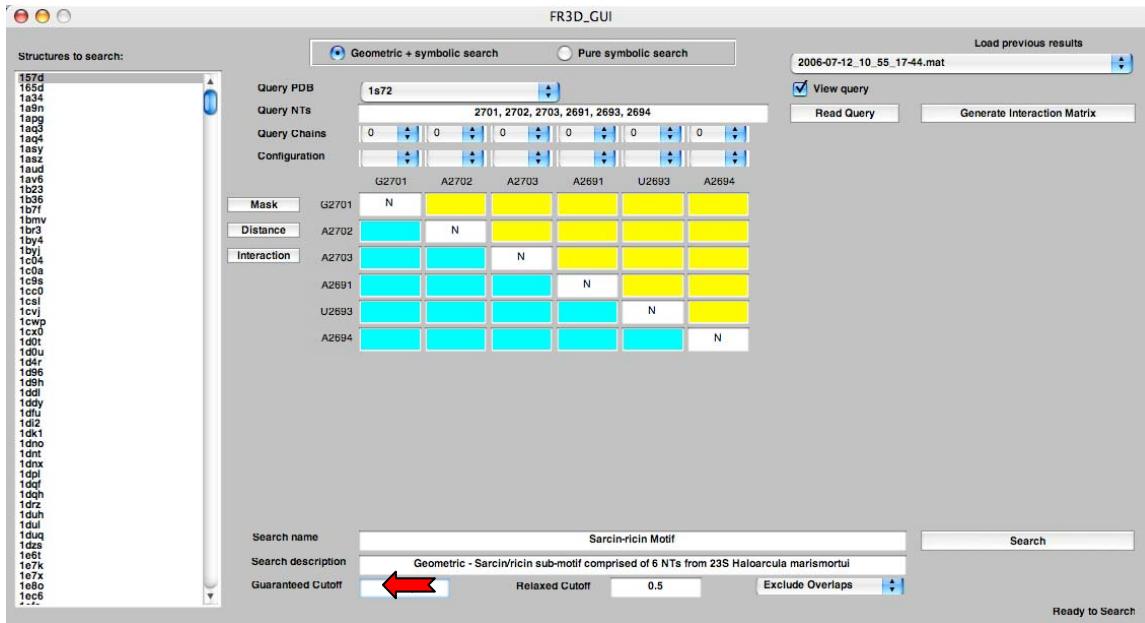
10. The user must specify the Relaxed Cutoff discrepancy, using the text-box labeled Relaxed Cutoff (e.g., 1.0). This number must be equal to or greater than the Guaranteed Cutoff. Making the relaxed cutoff larger than the guaranteed cutoff will retain some candidates which are similar to the Query motif without greatly increasing the running time. The algorithm is not guaranteed to find all candidates whose discrepancy from the Query motif is between the guaranteed cutoff and the relaxed cutoff.
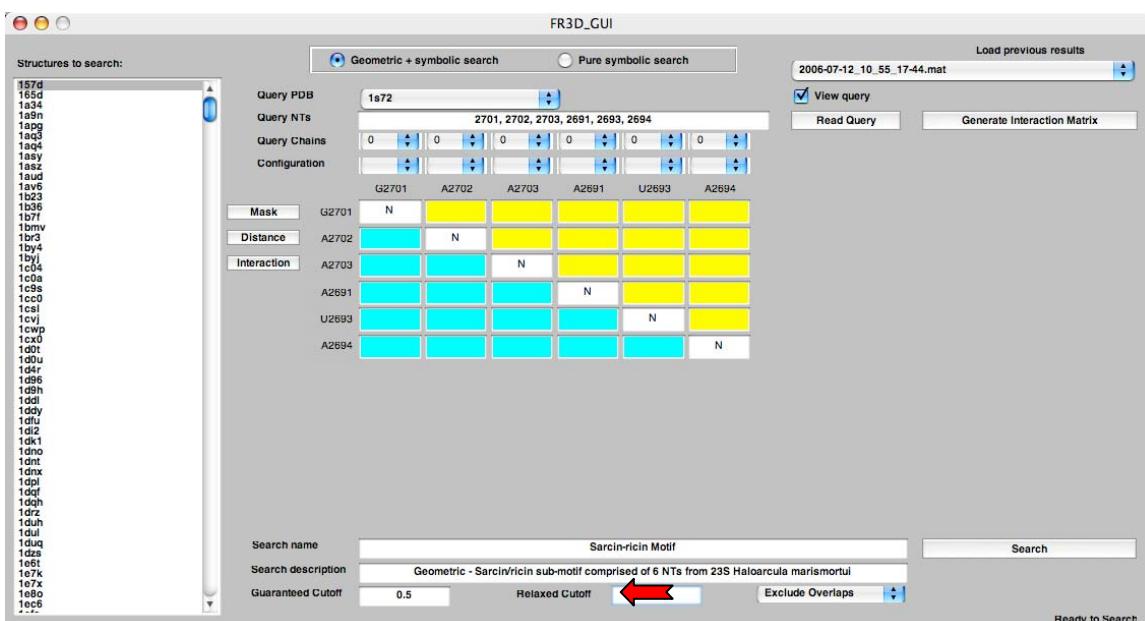
11. Using the drop-down menu to the right of the Relaxed Cutoff text-box, the user can specify whether to Exclude Overlaps or Include Overlaps. An example of this is when performing a search using nucleotides 10, 11, 12, 13, 14, 15 from some PDB file. The algorithm will certainly return the Query motif, but it may also return slight variations of the same motif such as nucleotides 9, 11, 12, 13, 14, 15. These are referring to the same motif, just one nucleotide is different, and so we consider this an overlap, or redundant version of the motif. The option Include Overlaps would keep this candidate, while Exclude Overlaps will remove candidates which have more than half of their nucleotides in common with another candidate having lower discrepancy from the Query motif. In this search we are Excluding Overlaps.

12. Select the PDB files which you would like to search in for your particular motif (i.e. 1qrs, 1qvf, 1rc7, 1s72). On a PC, by holding down the Control key on your keyboard one can select multiple files, which are not consecutive in the list-menu. To do this on a Mac, the user must hold down the Command (or Open-Apple) key. It is possible to make user-defined lists of PDB files to facilitate specifying the PDB files to search, see below.

13. Perform the Search by pressing the Search Button. Information about the progress of the search is displayed in the bottom right corner of the GUI. Often, the slowest part of the search is loading PDB data. If a PDB file has not already been analyzed by FR3D, it will need to be analyzed, which is rather slow. Even loading pre-computed data may be slow. The length of the search itself will vary depending on the number of nucleotides in the Query motif and the guaranteed discrepancy cutoff.
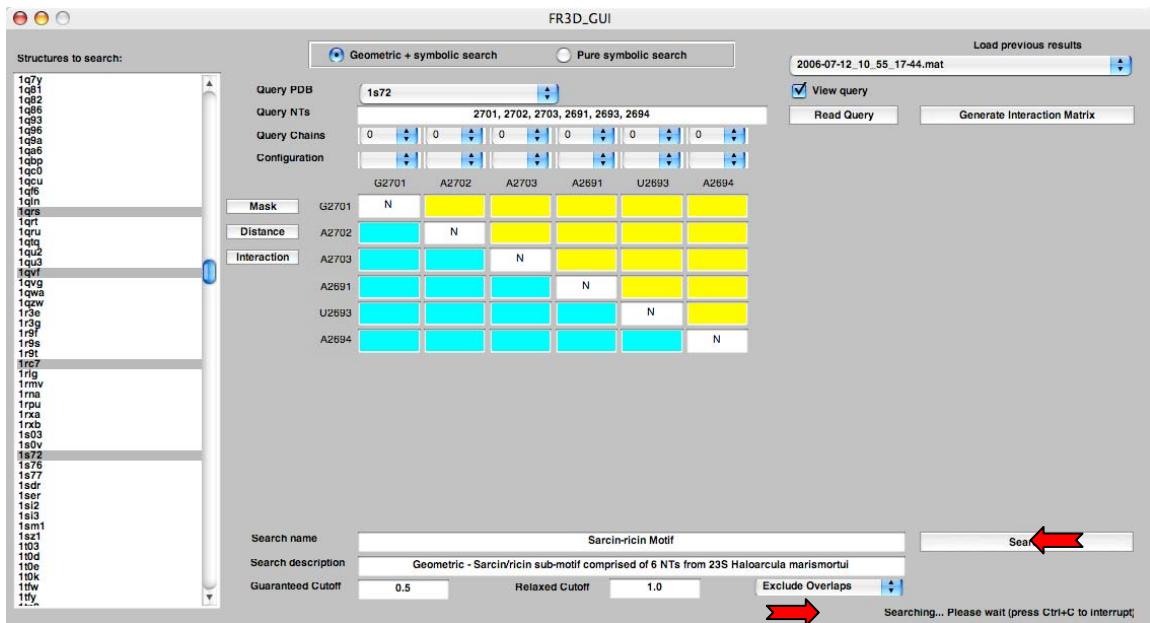
14. Once the search is complete the total number of Candidates found will be displayed in the bottom right corner of the GUI (e.g., 400 Candidates found). To learn about Displaying or Listing Candidates refer to those sections within this manual.

Performing Mixed Geometric and Symbolic searches with FR3D

We assume the reader has read the previous section on purely geometric searches, and so we focus on what is new in a mixed search.

1. The first step is to click on the radio button which says Geometric + symbolic search.

2. The user must specify the PDB file (e.g., 1s72) which contains the known motif from the drop-down menu labeled Query PDB.
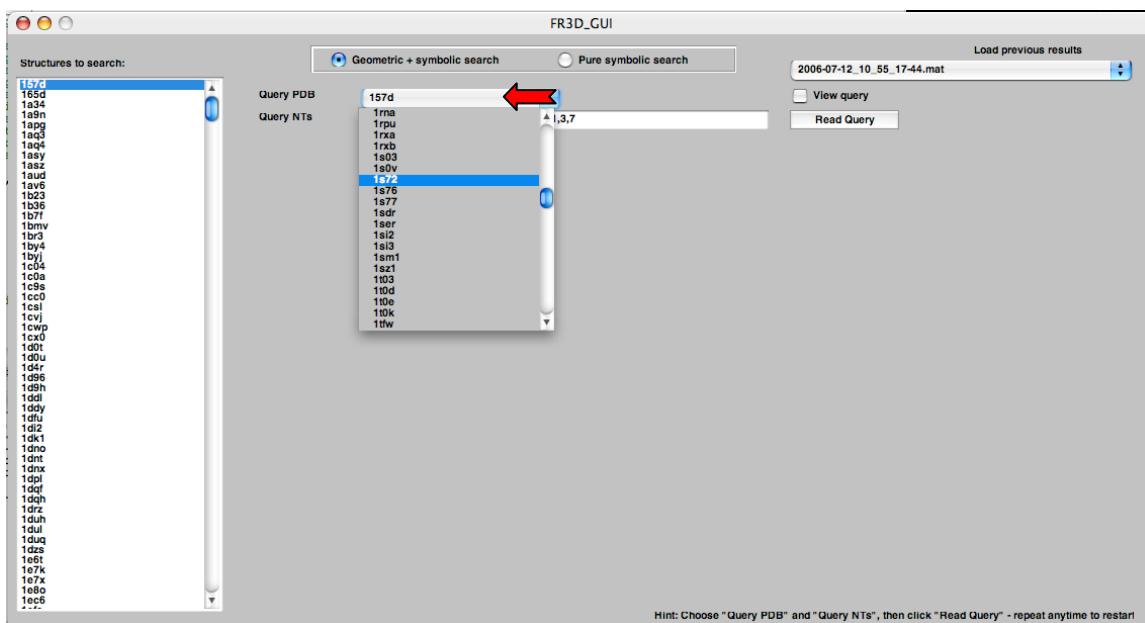
3. The query nucleotides are entered into the text-box, labeled by Query NTs. (e.g., 2701, 2702, 2703, 2691, 2693, 2694).
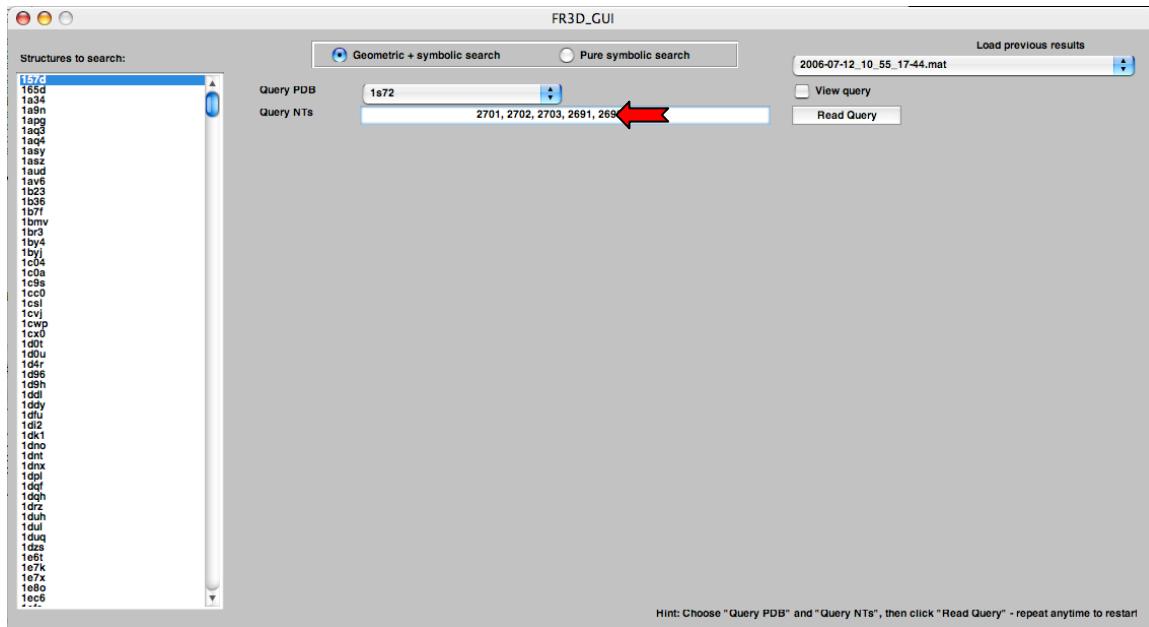
4. Press Read Query. To see the Query motif, check View query before pressing Read Query.
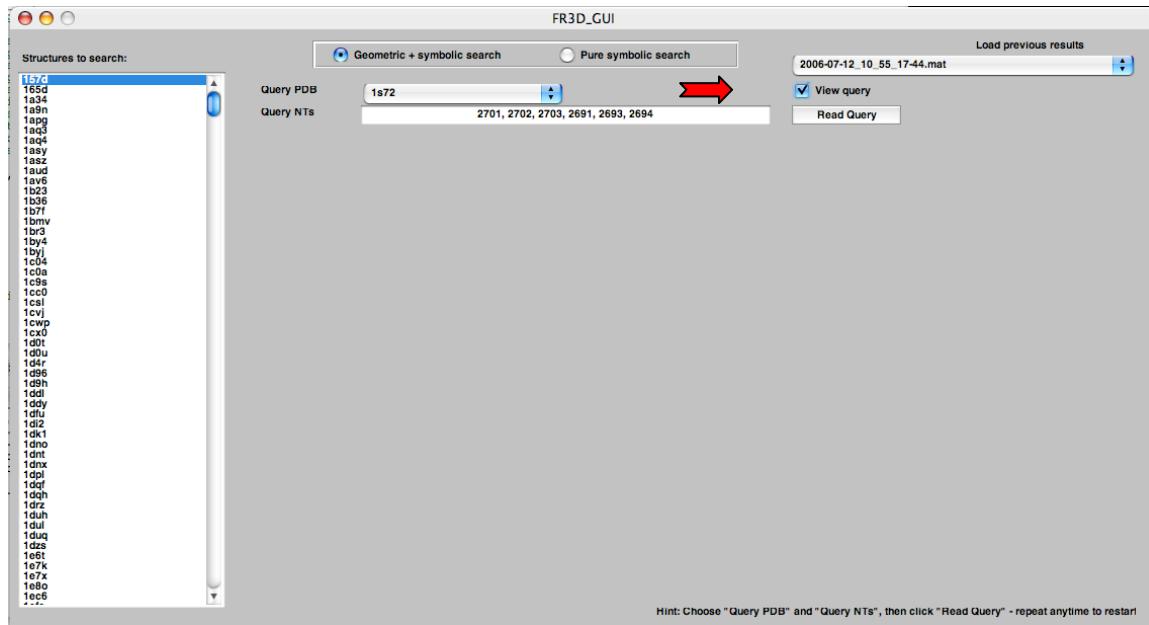
5. The user may use the drop-down menus to select the chain, in case of ambiguity.
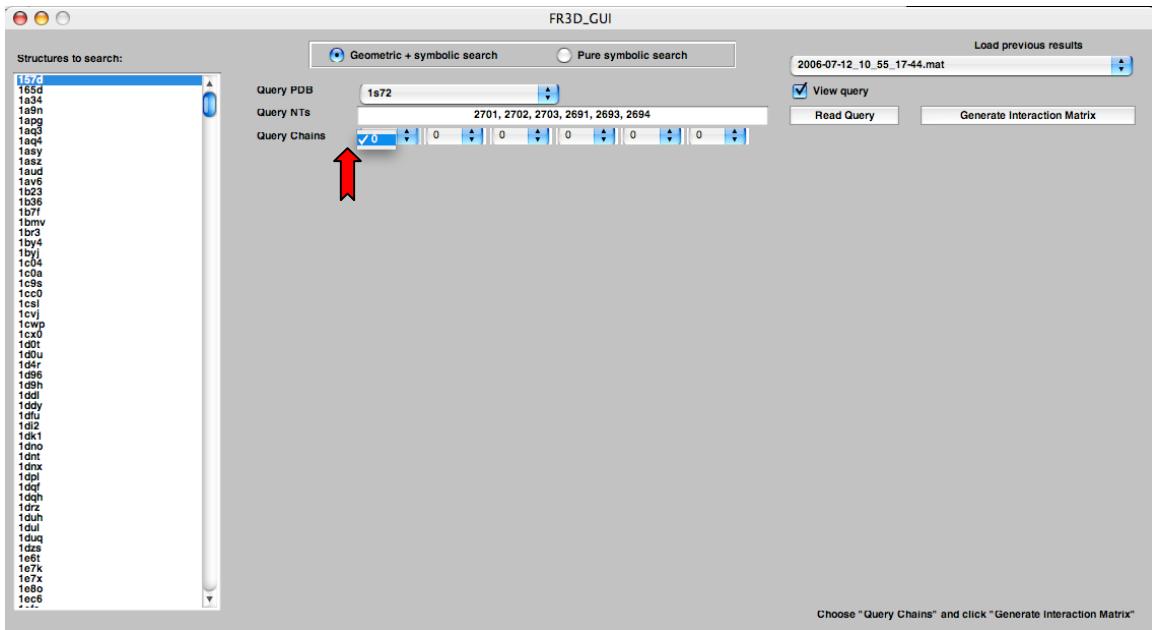
6. The user should press the Generate Interaction Matrix button. When this is pressed an Interaction matrix will appear on the GUI. Now we describe how to focus the search by specifying symbolic constraints which must be met by each candidate. Adding symbolic constraints shortens the running time of the search algorithm.

7. The user can specify the glycosidic bond conformation (anti or syn) for each base in their search using the drop-down menu labeled Configuration. The order of the drop-down menus corresponds to the same order as the Query nucleotides. To allow both conformations, leave the selection(s) blank.

8. The user can impose a basepair identify constraint (nucleotide mask) for their search by putting in nucleotide constraints in the text-boxes on the diagonal in the Interaction Matrix, which has a white background. Typing "A," for instance, means that only candidate motifs with an A in the corresponding position will be kept. Typing "AG" allows either A or G, etc. The program uses these standard abbreviations for other combinations:

M for A or C
R for A or G
W for A or U
S for C or G
Y for C or U
K for G or U
V for A, C, or G
H for A, C, or U
D for A, G, or U
B for C, G, or U
N for A, C, G, or U

Note that N is the default. It is not necessary to use these abbreviations, however. One may also exclude a given base using the syntax "~G" for instance, to exclude candidates with a G in the corresponding position.

The diagonal boxes are also the place to specify certain parameters that modify the definition of the geometric discrepancy. These are described in the pop-up window concerning the mask and in the article Sarver et al. 2007.

If the user would like to know more information on using masks, they can press the help button labeled Mask.
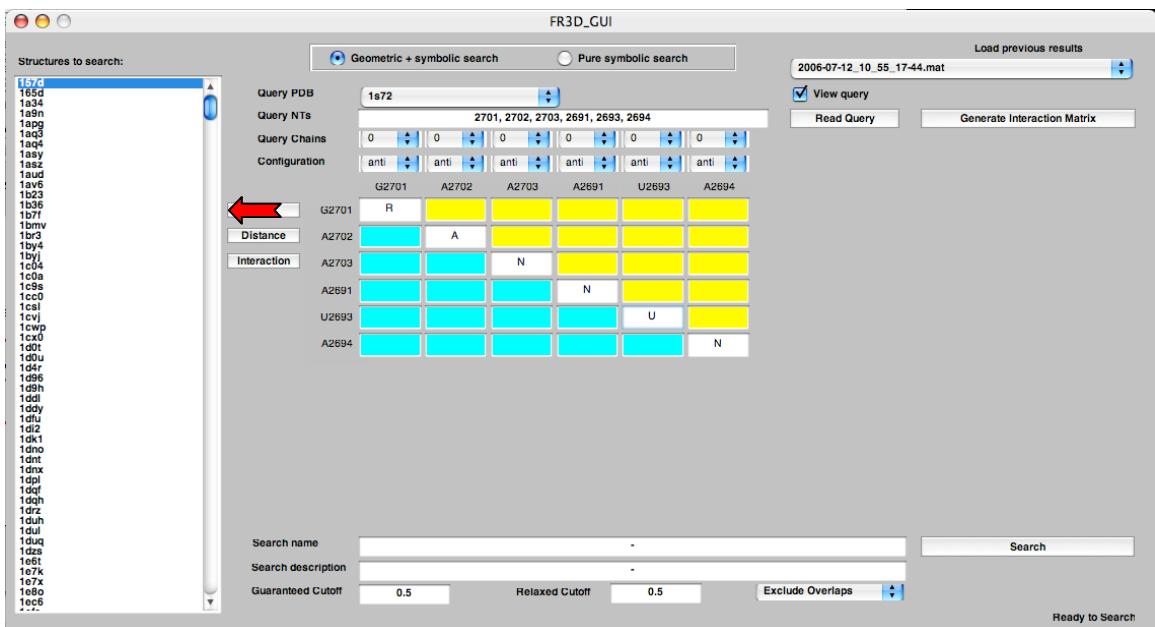
This is the pop-up help menu when the Mask button is pressed.

Mask help

Nucleotide mask, location weight, and angle weight are specified in the boxes on the diagonal.

For the mask, type A to allow only A's, similarly with C, G, U. Type AC to allow only A or C, or AGU to allow only A, G, or U. Type ~G to exclude G (or type ACU). You may use one of these standard abbreviations: M = AC; R = AG; W = AU; S = CG; Y = CU; K = GU V = ACG; H = ACU; D = AGU; B = CGU; N = ACGU Using no mask is the same as specifying ACGU or N. Self-contradictory specifications (such as A ~A) may give unpredictable results or empty searches.

Specify the location weight for the given nucleotide by typing something of the form LW2.5 in the box. Location weights are normalized automatically, so making them all equal to 5.0 has no effect on the search. The default value, if nothing is specified, is 1.0.

Specify the angle weight for the given nucleotide by typing something of the form AW2.0 in the box. Angle weights are not automatically normalized; the default is 1.0. Making angle weights larger than 1.0 increases the orientation error. Thus, candidates with low discrepancy will have bases more closely aligned than when using the default value of 1.0 for angle weight.

Separate the mask, location weight, and angle weight by spaces.

9. The user can impose basepair, base stacking, and base-phosphate constraints using the text-boxes, which are colored yellow and in the upper right half of the diagonal in the Interaction Matrix (i.e. G2701 (row) forms a trans Sugar Edge/Hoogsteen basepair (tSH) with A2794 (column), A2702-U2693 form trans Hoogsteen/Watson-Crick basepair (tHW), and A2703-A2691 forms a trans Hoogsteen/Hoogsteen basepair (tHH)).
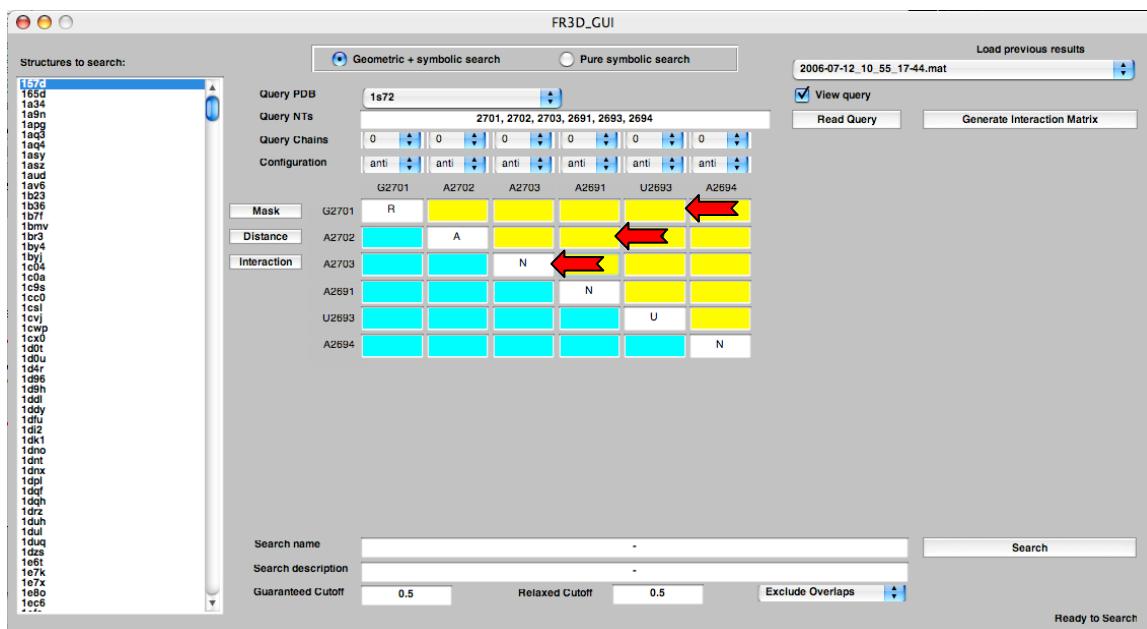
For more information on using the basepair constraints, the user can press the Interaction Button to the left of the Interaction Matrix.

This is the pop-up help menu for the Basepair interactions.

Basepair, base stacking, base phosphate, or letter pair constraint is specified above the diagonal.

To specify that all candidate motifs must have a tWH basepair between the nucleotides corresponding to the first and second nucleotides in the query motif, type tWH in the first row, second column. This means that the nucleotide in the first row must use its Watson-Crick edge, and the nucleotide in the second column must use its Hoogsteen edge.

Valid basepair specifications are cWW, tWW, cWH, cHW, tWH, tHW, cWS, cSW, tWS, tSW, cHH, tHH, cHS, cSH, tHS, tSH, cSS, tSS. Note, however, that the cSS and tSS interactions are not, in fact, symmetric, because each base can use the sugar edge differently. Follwing Leontis, Stombaugh, Westhof (NAR 2002), type cSs to specify that the first base has priority, csS for the second, or cSS for either.

Specifying multiple interactions allows more ways a candidate can satisfy the constraints; for example, typing cWH cHW requires a cis Watson-Crick/Hoogsteen basepair, but either base can use the Watson-Crick edge, and the other uses the Hoogsteen edge.

The abbreviation "trans" gives all trans categories, "cis" for cis.

Type "bif" for bifurcated basepairs (see LSW 2002).

Type ~cWW to exclude candidates having a cWW basepair.

Some pairs of bases are close to, say, cWW, but do not meet the strict criteria for membership in the cWW classification. Type ncWW ("near cWW") to get basepairs that are not classified into any category, but for which the cWW category is the closest match, up to a certain fairly generous limit. Type "cWW ncWW" to get cWW and near cWW pairs, cWW. Type ntrans to get all pairs nearest to a trans pair.

Type s35 for stacking in which the first base uses its 3 face, and the second base uses its 5 face. Similarly, type s53, s33, or s55. Type "stack" to allow all stacking interactions. The prefixes "n" and "~" work with stacking, as above.

To specify that the nucleotides must match a certain pattern, type, for example, "cWW CG GC" to get only CG or GC cWW pairs.

To require that two nucleotides make a base-phosphate interaction, enter BPh in the corresponding yellow box. This will select pairs of nucleotides in which the first nucleotide's base is a hydrogen bond donor and the second nucleotide's phosphate is an acceptor. To reverse the roles, type PhB. To specify particular base-phosphate categories, type 0BPh, 1BPh, 2BPh, ..., 9BPh, or 0PhB, 1PhB, etc. For near base-phosphate interactions, type nBPh, nPhB, n1BPh, n1PhB, etc.

One can restrict to pairs that play a certain role in the secondary and tertiary structure. For pairs that are nested, type "N" or "nested". For pairs that cross nested interactions but involve nucleotides in the same branch of the RNA, type "local" or "L". For long-range or distant interactions, between different branches of the RNA, type "long-range", "distant", "D", or "LR". Note that "nested", "local", and "distant" are mutually exclusive. They can be negated with ~, but ~local only returns distant interactions, not nested ones.

To specify bases on the same strand that form cWW pairs that flank a hairpin, internal, or junction loop, type "flank" or "F". Note: for internal and junction loops, flanking nucleotides will be on the same strand, one on each side of the internal or junction loop. Such flanking nucleotides usually do not interact with one another.

Base-phosphate interactions. Specify that nucleotides i and j should make a base-phosphate interaction by putting the text BPh in row i and column j. This requires that the nucleotide in row i be a hydrogen donor and that the phosphate of the nucleotide in column j be an acceptor in a hydrogen bond between the two nucleotides. Specify a particular base-phosphate interaction with text such as 1BPh, 2BPh, … 9BPh, 0BPh. To reverse the roles of i and j, use PhB, 1PhB, etc. Near base-phosphate interactions can be obtained with nBPh, n1BPh, n3PhB, etc.

11. The user can put sequence constraints on the search using the text-boxes colored in cyan and are located on the bottom-left of the diagonal. For this search we are only using two constraints which include '> <2' and '> <4'. The first greater-than sign in each example represents that the row nucleotide should be after the column nucleotide sequentially. The second less than signs represent the number of bulged base are allowed between the row and column nucleotides.
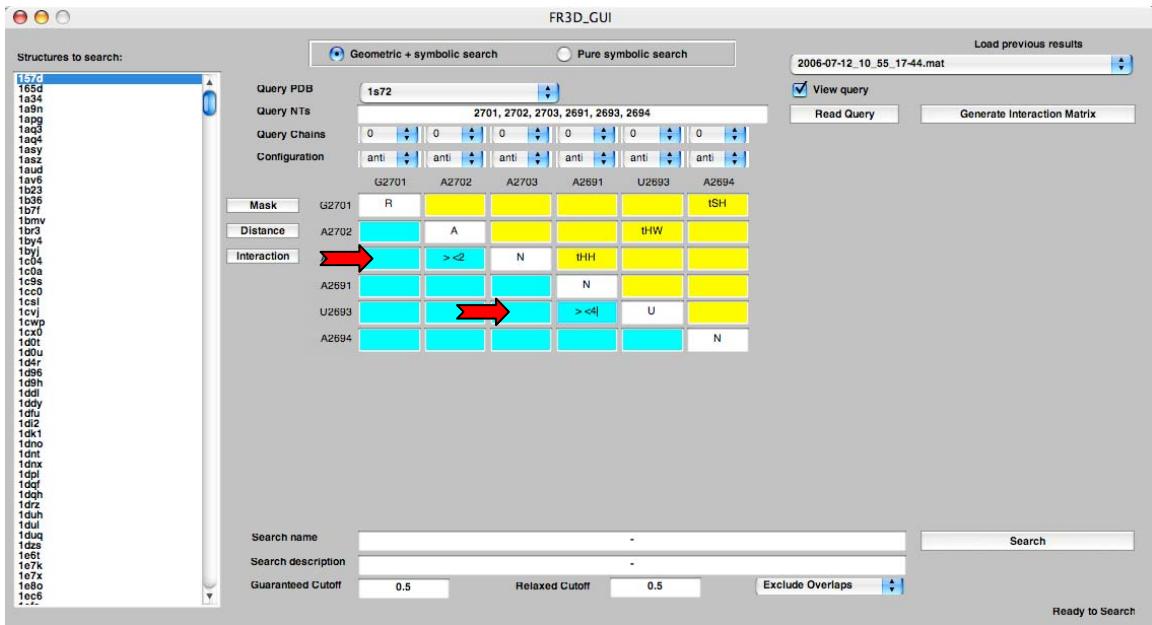
For more information on using the sequence constraints the user can press the Distance buttons located to the left of the Interaction Matrix.
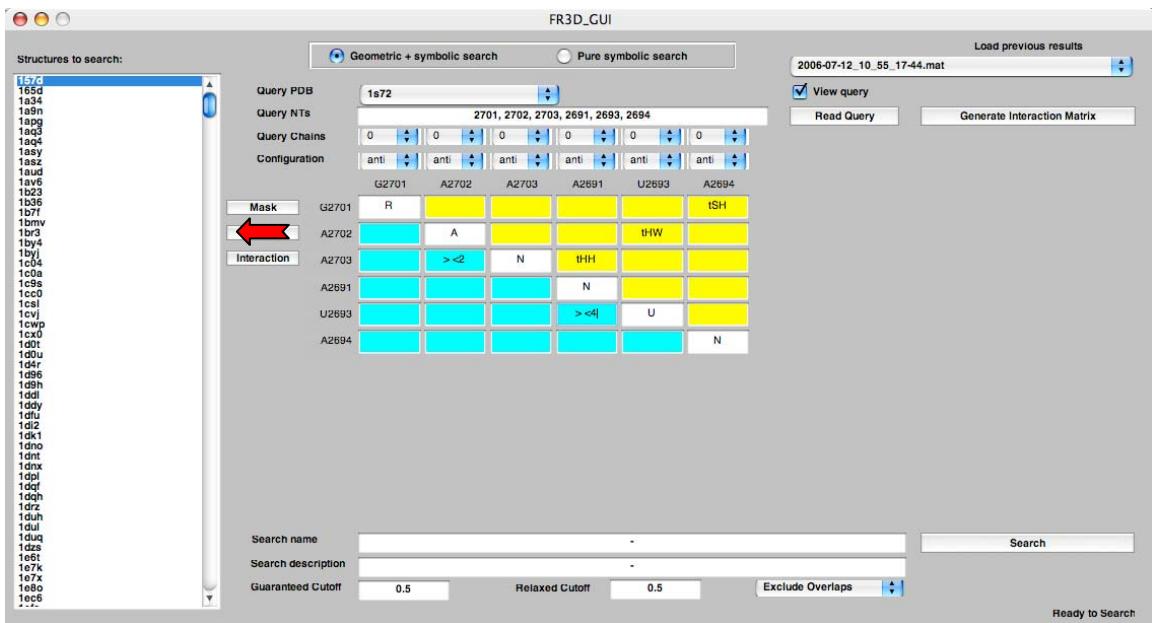

This is the pop-up menu which is displayed after the user presses the Distance button.

Distance help

Set limits on the difference between nucleotide numbers using the boxes below the diagonal. (Actually, what is used is the difference between the index of nucleotides in the file, not NDB nucleotide number.) To put an upper limit on the difference, type something like <5 or <=5. To put a lower limit on the difference, type something like >5 or >=5. To put both limits at once, type something like >5 <=12. To insist that the nucleotide in the given row have a lower nucleotide number than the nucleotide in the given column, type <, separated by a space from other specifications. For greater, type >.

12. The user should enter a name the type of search that they will be performing in the text-box, labeled Search name (i.e. Sarcin-ricin Motif). This will be the name used to recall a previous search, so the name should not use "/" or "\", because these refer to folders. The user can add more descriptive information about their search in the text-box labeled Search description (i.e. Geometric – Sarcin/ricin sub-motif comprised of 6 NTs from 23S Haloarcula marismortui). Other comments about the search can be added here as well.
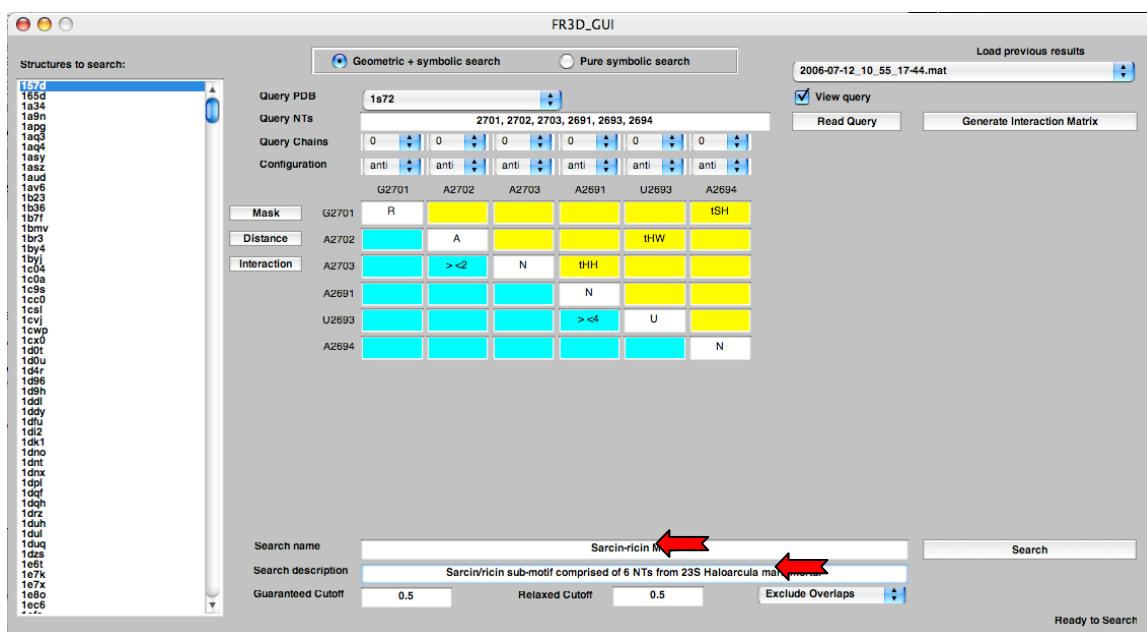
13. The user must set the Guaranteed Cutoff discrepancy, using the text-box labeled Guaranteed Cutoff (i.e. 0.5)

21
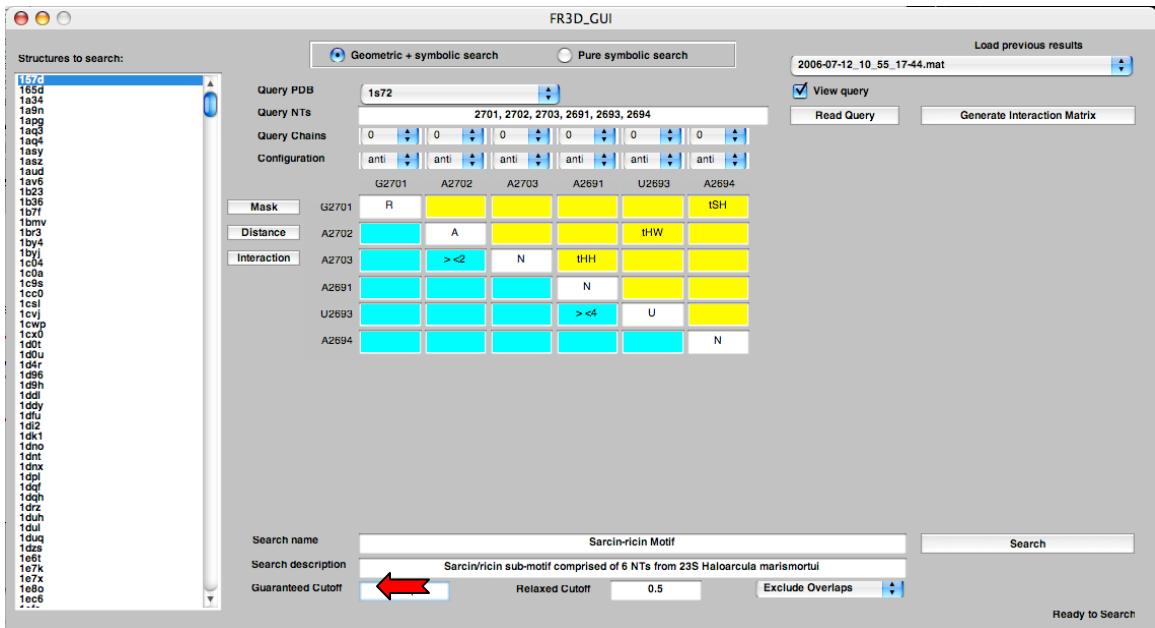
14. The user can specify the Relaxed Cutoff discrepancy, using the text-box labeled Relaxed Cutoff (i.e. 1.0). This number must be equal to or greater than the Guaranteed Cutoff. Using a larger value can allow for additional flexibility in searches without increasing runtime as much as increasing the guaranteed cutoff would do.
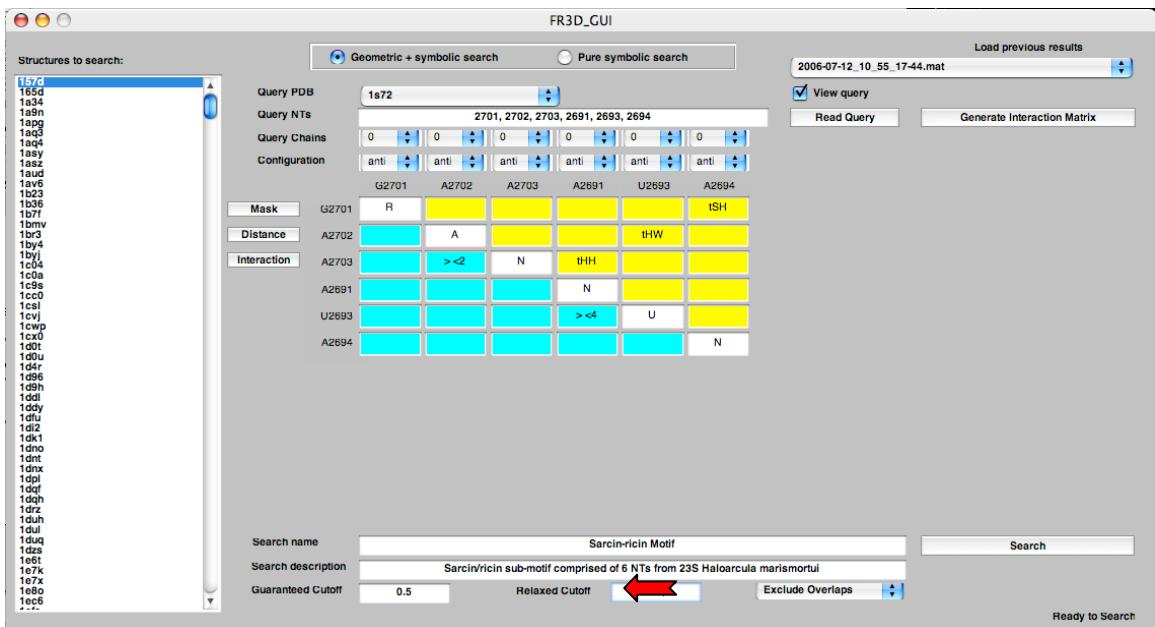

15. Using the drop-down menu to the right of the Relaxed Cutoff text-box, the user can specify whether to Exclude Overlaps or Include Overlaps.
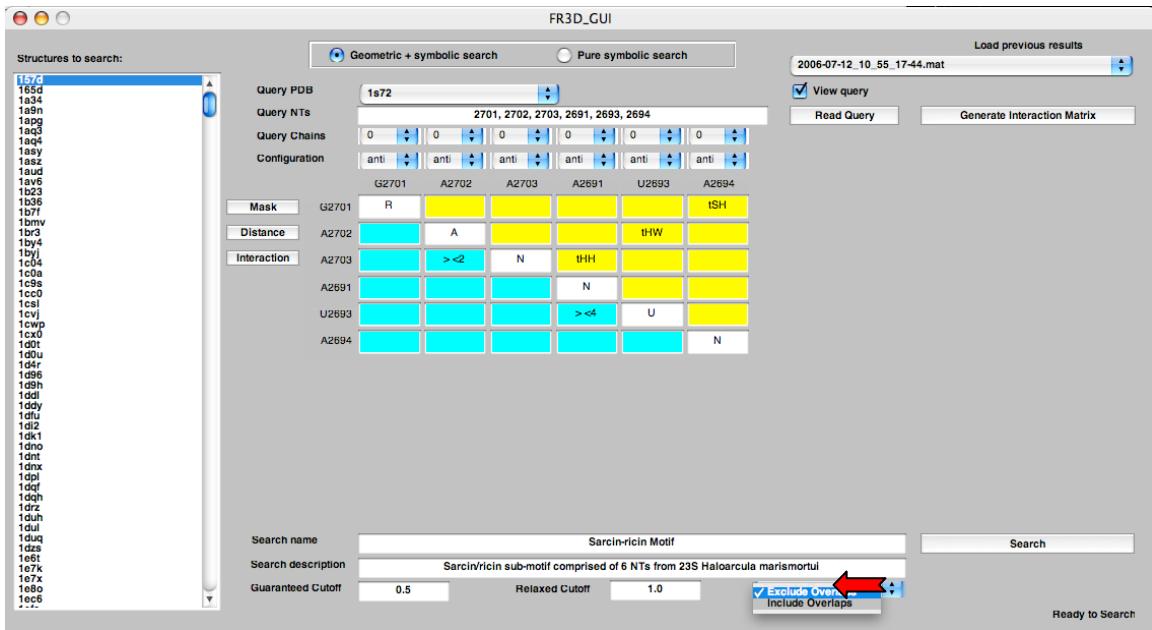
16. Select the PDB files which you would like to search in for your particular motif (i.e. 1qrs, 1qvf, 1rc7, 1s72). On a PC, by holding down the Control key on your keyboard one can select multiple files, which are not consecutive in the list-menu. To do this on a Mac, the user must hold down the Command (or Open-Apple) key. If the user would like to select consecutive file in the list-menu, they should hold down the Shift Key (PC/Mac) and select their files.
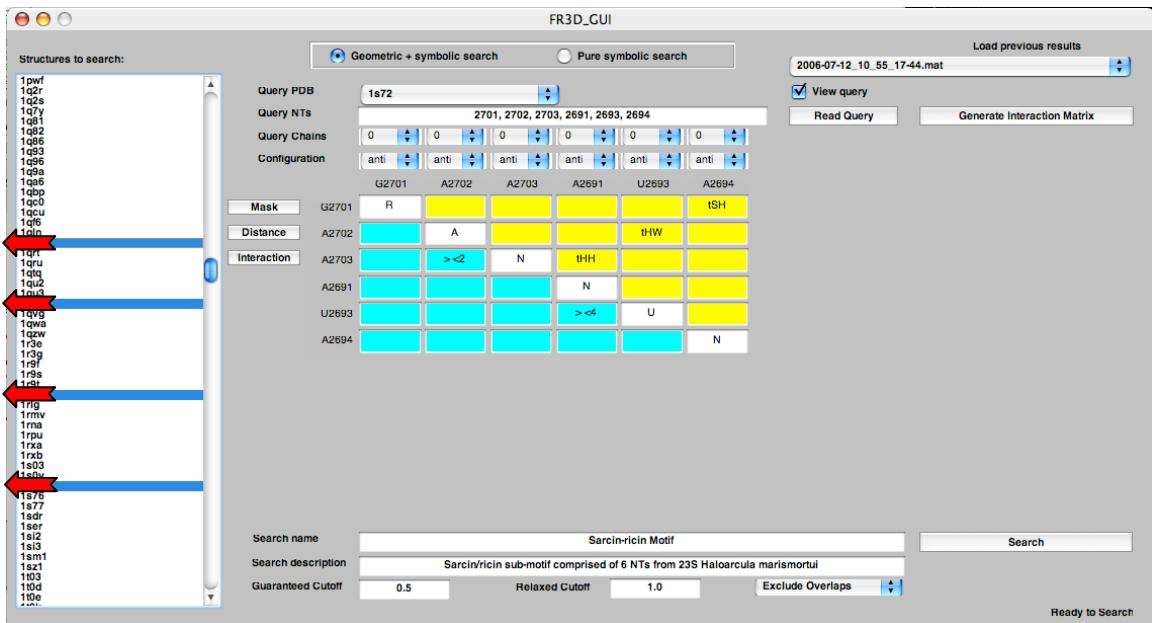
17. Perform the Search by pressing the Search Button. Information about the progress of the search is displayed in the bottom right corner of the GUI.
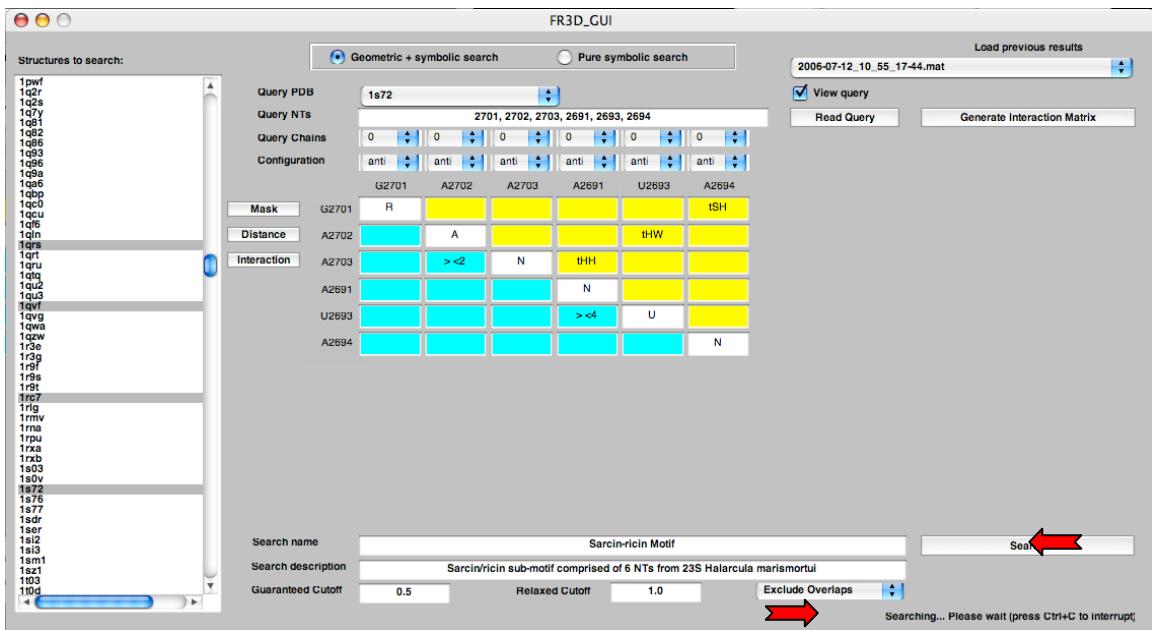
18. Once the search is complete the total number of Candidate will be displayed in the bottom right corner of the GUI (i.e. 400 Candidates found). To learn about Displaying or Listing Candidates refer to those sections within this manual.
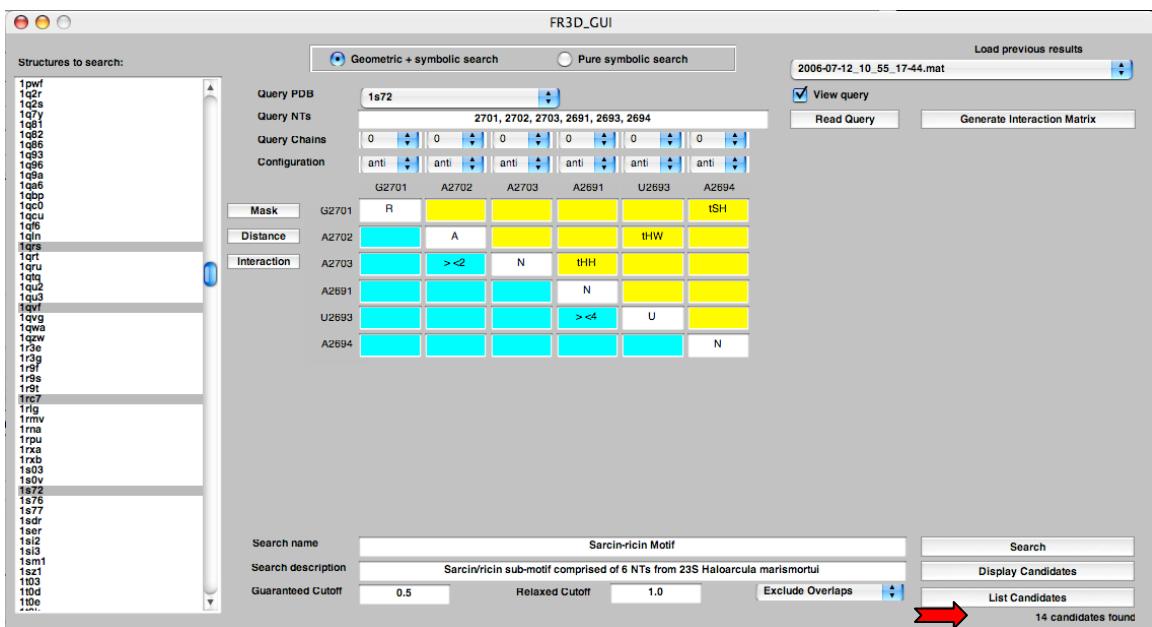
Conducting Symbolic searches with FR3D

Here we illustrate the ability of FR3D to search for motifs based only on symbolic criteria such as desired basepairing, base stacking, nucleotide identity, and sequential continuity constraints. We assume the reader has read the previous sections and focus only on what is new to purely symbolic searches.

1. Start by selecting Pure symbolic search. Then, the user must enter the number of nucleotides in the motif for which they want to search (e.g., 6).
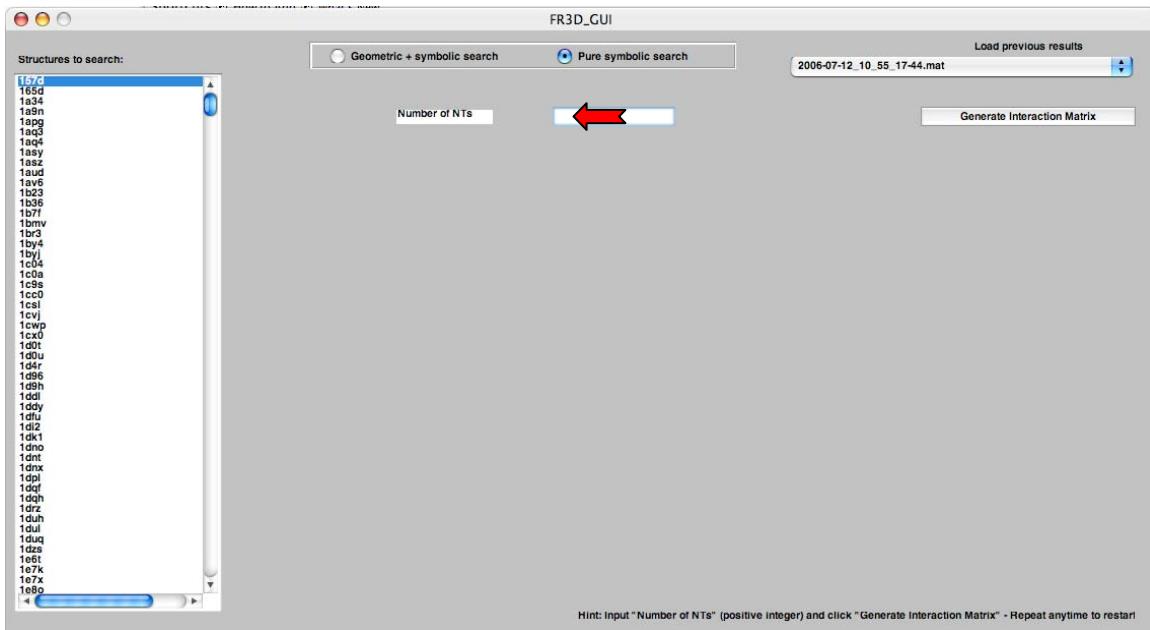
2. The user should now press the Generate Interaction Matrix button. When this is pressed an Interaction matrix will appear on the GUI.
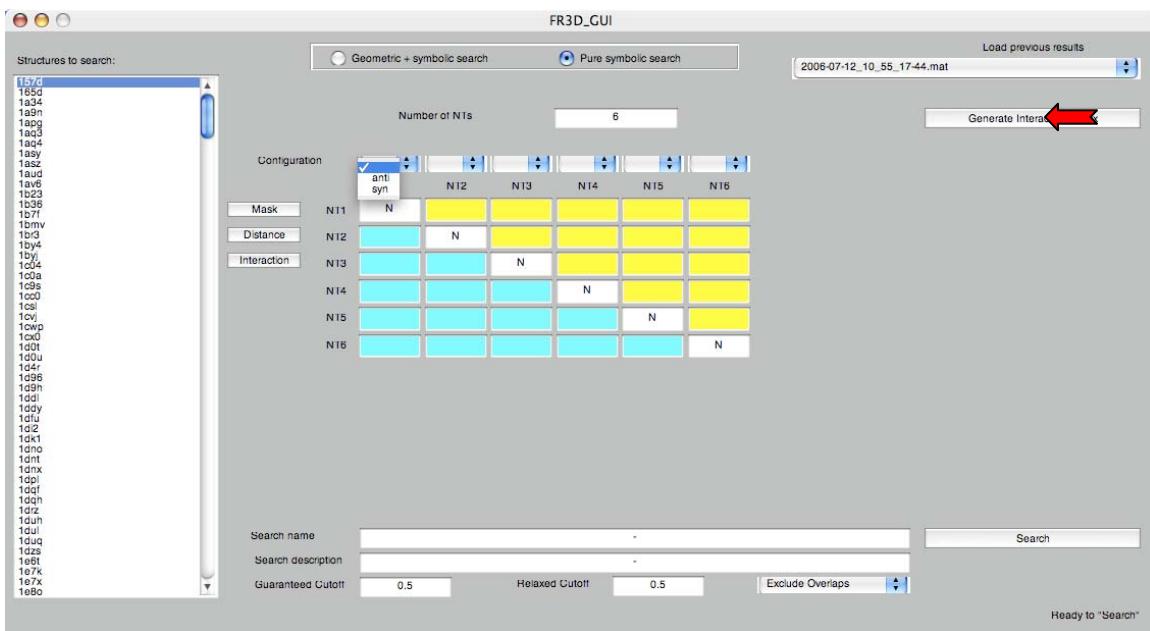
3. The user can specify the glycosidic bond conformation (anti or syn) for each base in their search using the drop-down menu labeled Configuration. The order of the drop-down menus corresponds to the same order as the Query nucleotides. If the user does not want to restrict the conformation to either anti or syn, they can leave the selection blank, which means both conformations are allowed.
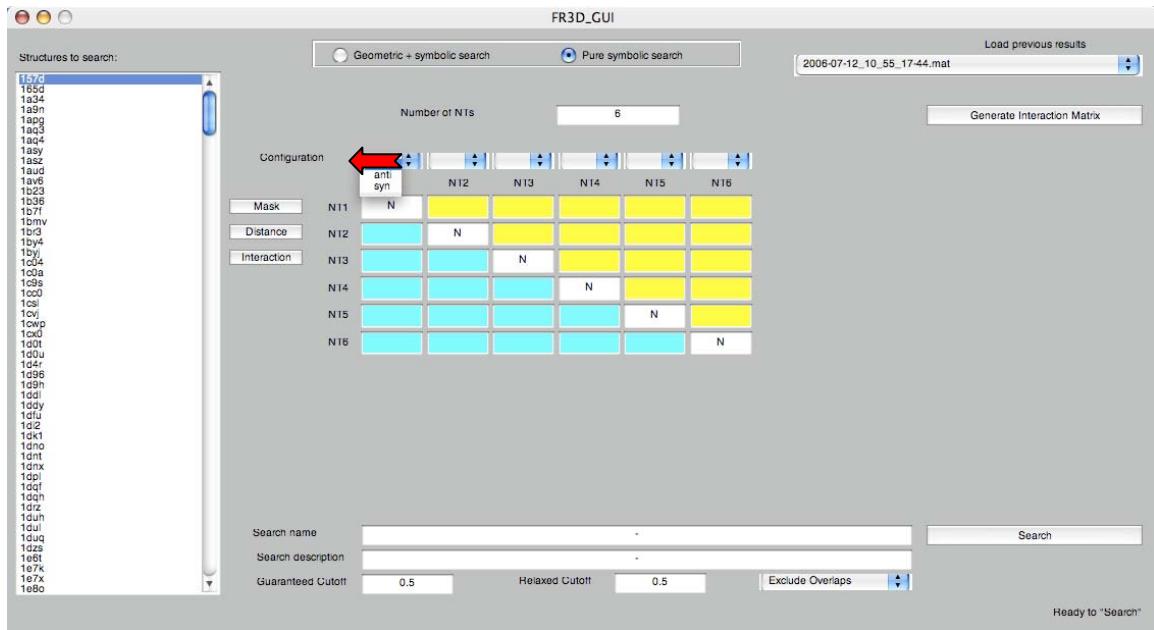
4. The user can impose a mask for their search by putting in nucleotide constraints in the text-box in the Interaction Matrix, which has a white background. The program will take many types of masking letters (i.e. A, C, G, U, R, Y, etc.).
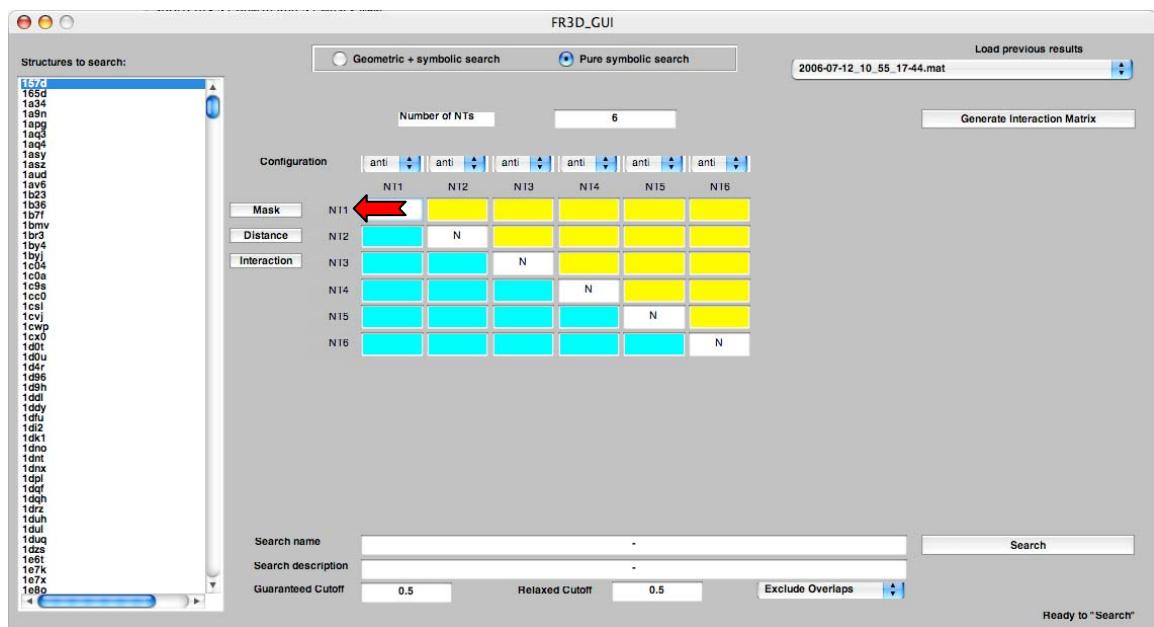
5. The user can impose basepair constraints using the text-boxes, which are colored yellow and in the upper right half of the diagonal in the Interaction Matrix (i.e. G2701 (row) forms a trans Sugar Edge/Hoogsteen basepair (tSH) with A2794 (column), A2702-U2693 form trans Hoogsteen/Watson-Crick basepair (tHW), and A2703-A2691 forms a trans Hoogsteen/Hoogsteen basepair (tHH)).

Base-phosphate interactions. Specify that nucleotides i and j should make a base-phosphate interaction by putting the text BPh in row i and column j. This requires that the nucleotide in row i be a hydrogen donor and that the phosphate of the nucleotide in column j be an acceptor in a hydrogen bond between the two nucleotides. Specify a particular base-phosphate interaction with text such as 1BPh, 2BPh, … 9BPh, 0BPh. To reverse the roles of i and j, use PhB, 1PhB, etc. Near base-phosphate interactions can be obtained with nBPh, n1BPh, n3PhB, etc.

6. The user can put sequence constraints on the search using the text-boxes colored in cyan and are located on the bottom-left of the diagonal. For this search we are only using two constraints which include '> <2' and '> <4'. The first greater-than sign in each example represents that the row nucleotide should be after the column nucleotide sequentially. The second less than signs represent the number of bulged base are allowed between the row and column nucleotides.
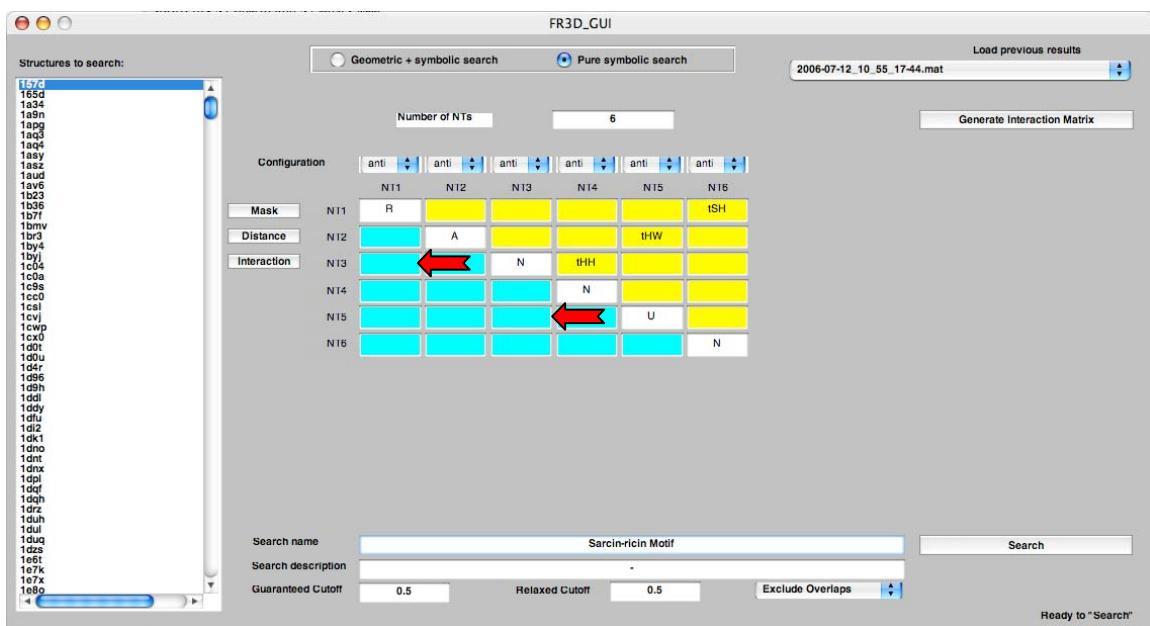
7. The user should enter a name the type of search that they will be performing in the text-box, labeled Search name (i.e. Sarcin-ricin Motif). This will be the name used to recall a previous search, so the name should not use "/" or "\", because these refer to folders. The user can add more descriptive information about their search in the text-box labeled Search description (i.e. Geometric – Sarcin/ricin sub-motif comprised of 6 NTs from 23S Haloarcula marismortui). Other comments about the search can be added here as well.
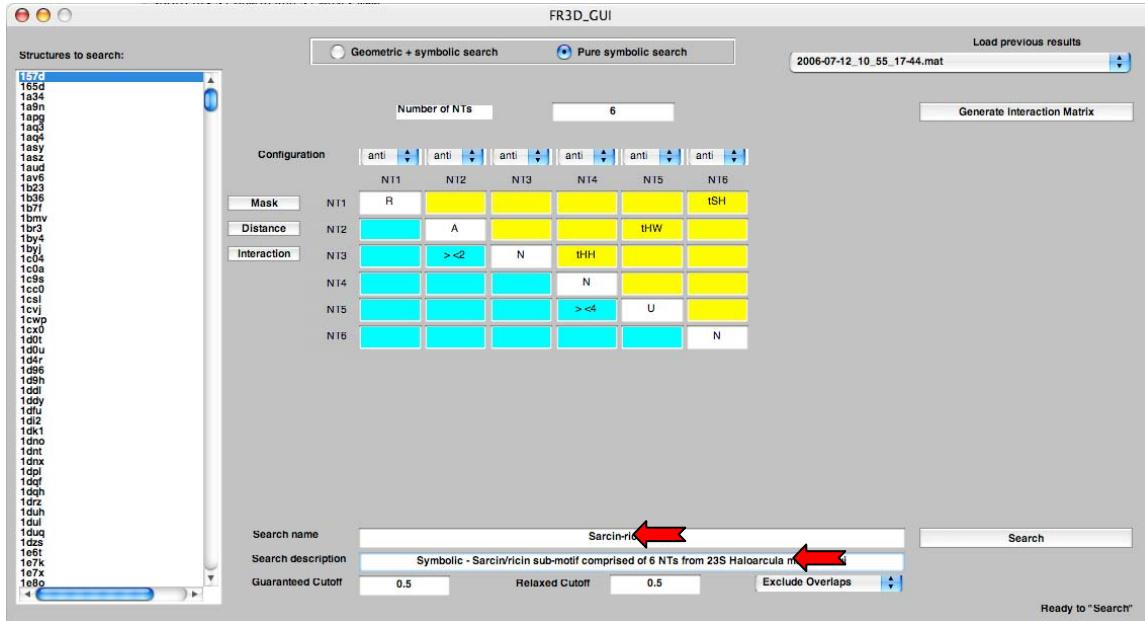
8. For a Symbolic search, the user does not need to specify a Guaranteed or Relaxed Cutoff.
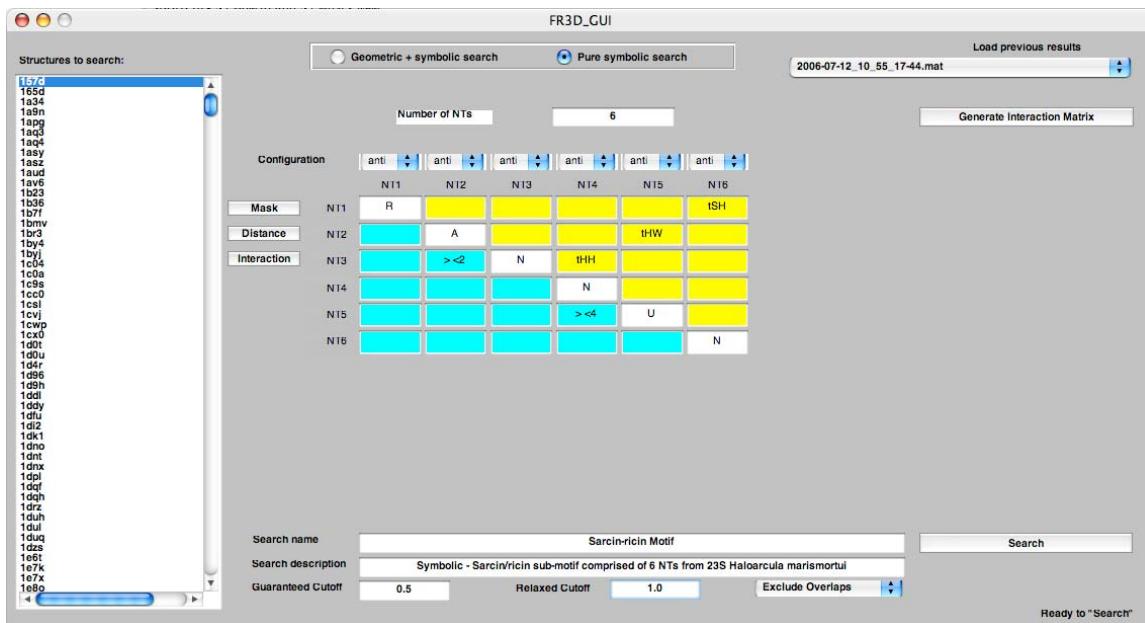

9. Using the drop-down menu to the right of the Relaxed Cutoff text-box, the user can specify whether to Exclude Overlaps or Include Overlaps.

10. Select the PDB files which you would like to search in for your particular motif (i.e. 1qrs, 1qvf, 1rc7, 1s72). On a PC, by holding down the Control key on your keyboard one can select multiple files, which are not consecutive in the list-menu. To do this on a Mac, the user must hold down the Command (or Open-Apple) key. If the user would like to select consecutive file in the list-menu, they should hold down the Shift Key (PC/Mac) and select their files.

11. Perform the Search by pressing the Search Button. Information about the progress of the search is displayed in the bottom right corner of the GUI.

12. Once the search is complete the total number of Candidate will be displayed in the bottom right corner of the GUI (i.e. 400 Candidates found). To learn about Displaying or Listing Candidates refer to those sections within this manual.
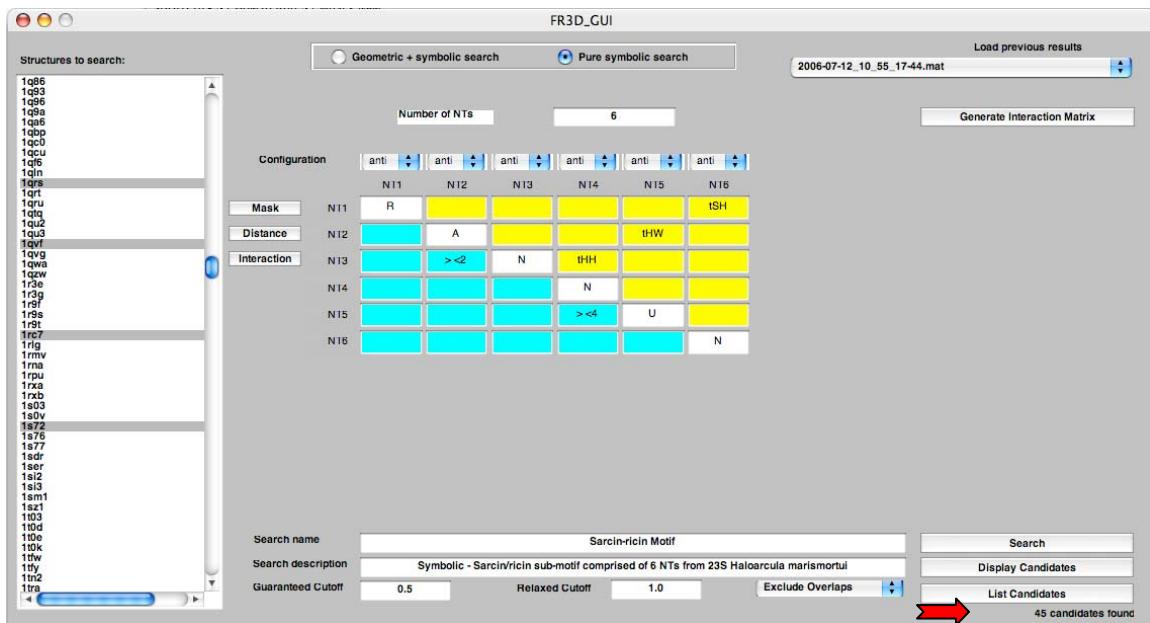

Viewing candidates

After performing a Geometric, Symbolic, or mixed search, the user can view the candidate motifs by pressing the Display Candidates button in the bottom-right corner of the GUI. The green arrows will direct the user to places in the figures where changes may have occurred, while the red arrows direct the attention to user actions.
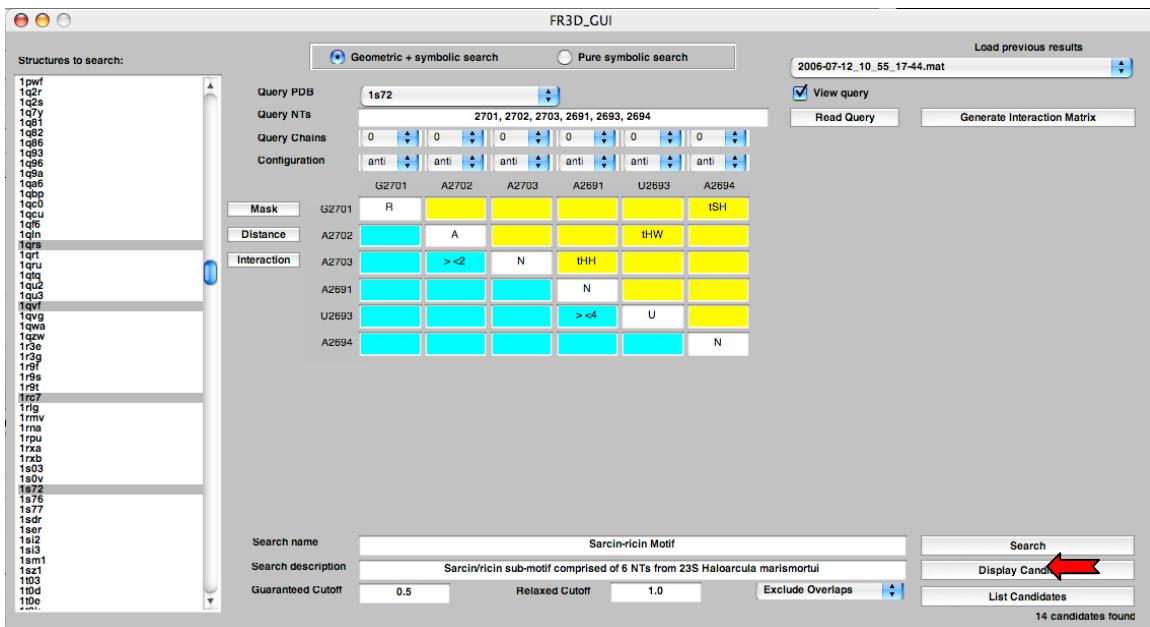
This pop-up figure and menu appear after the user presses the Display Candidates button. The order of the motifs is arranged according to the lowest discrepancy, so the first Candidate should always be the Query motif.
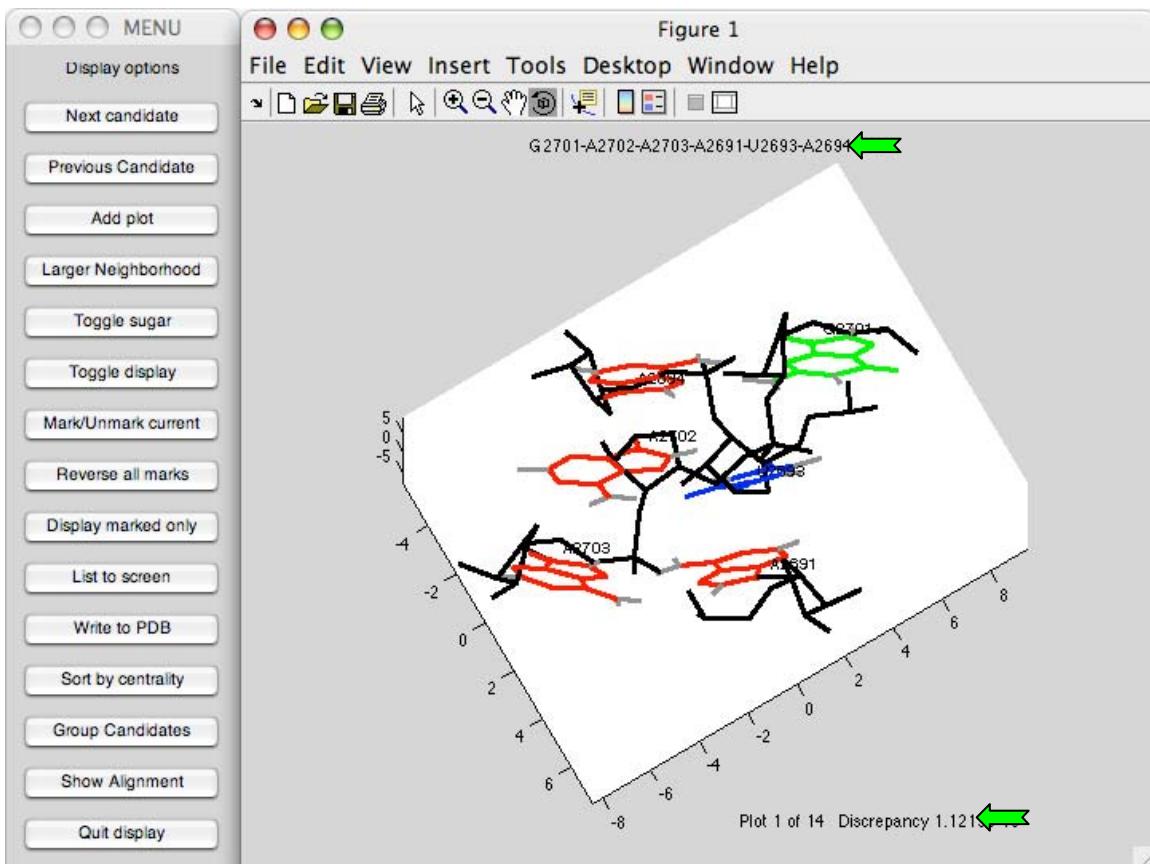

To look at the next Candidate in the user should press the Next Candidate button on the menu. In the figure you can see the next lowest scoring Candidate motif.
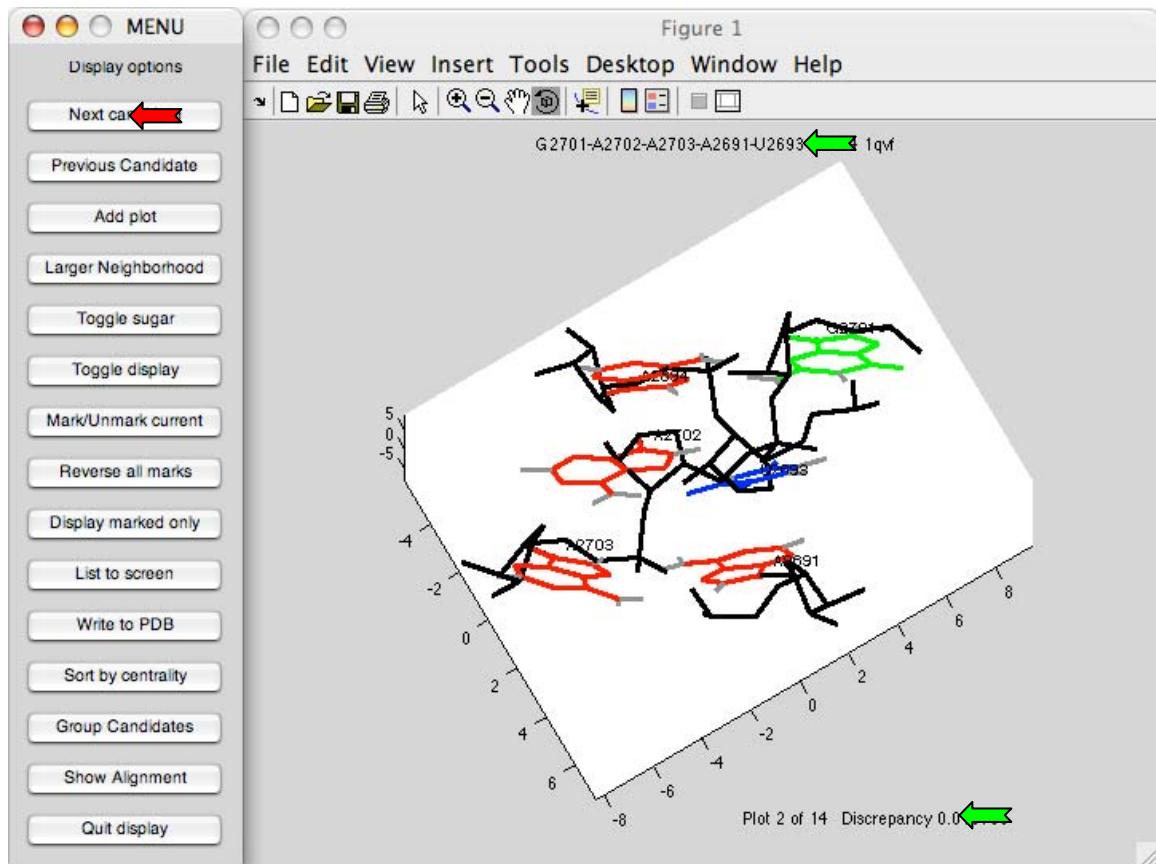
To go back to a previous Candidate, the user should press the Previous Candidate button on the menu. Now the figure refers back to our first Candidate, which is the Query motif.
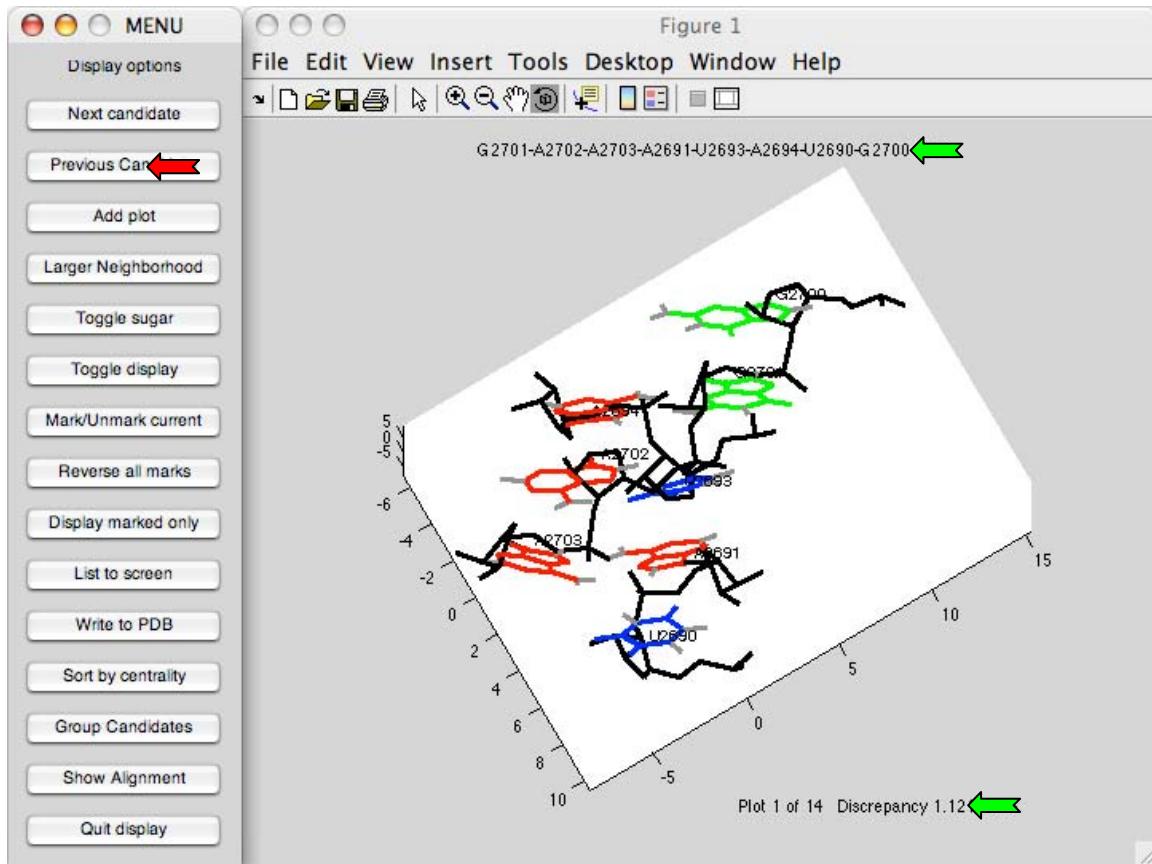
If the user would like to view more than one Candidate a time, they can add more figures, by pressing the Add plot button on the menu. When selecting a particular figure, the user can press Next Candidate or Previous Candidate and the selected figure will change.
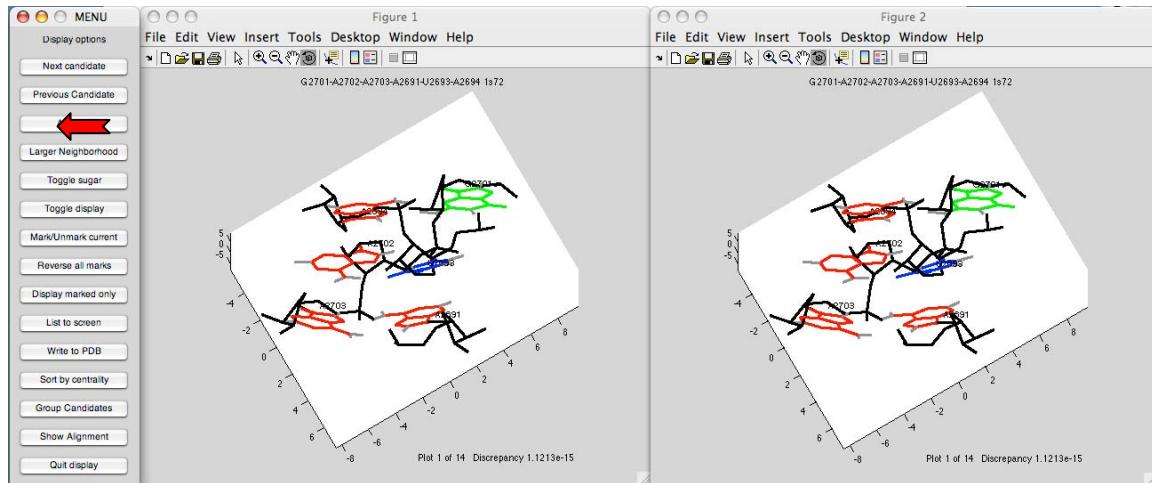
If the user would like to look at the surrounding bases of one of the Candidate motifs, they can press the Larger Neighborhood button. In the figure it should be noted that the nucleotide list now includes the

neighboring bases. By pressing the Large Neighborhood button several times 2 or 3 times, the motif will go back to the original size of the Candidate motif.
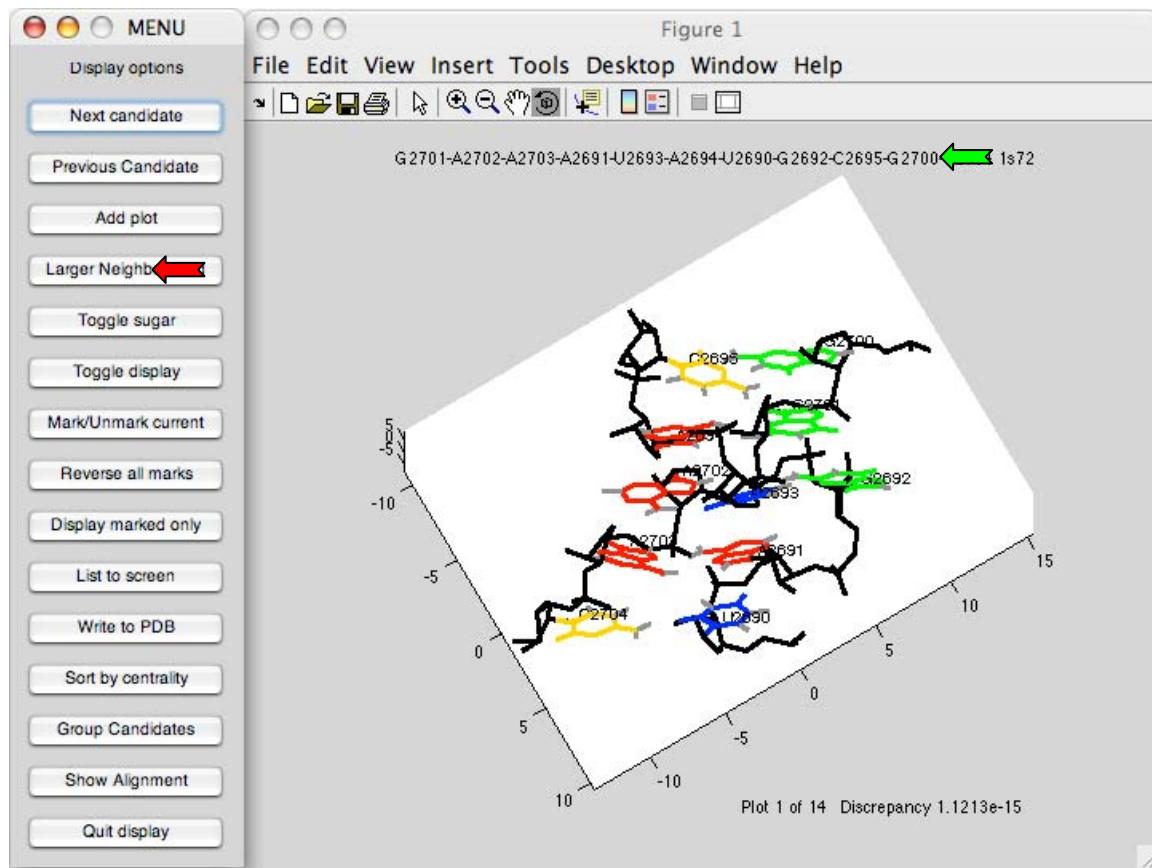
If the sugars are impeding your visualization, the user can press the Toggle sugar button to turn-on or turn-off the sugars as shown in the figure below.
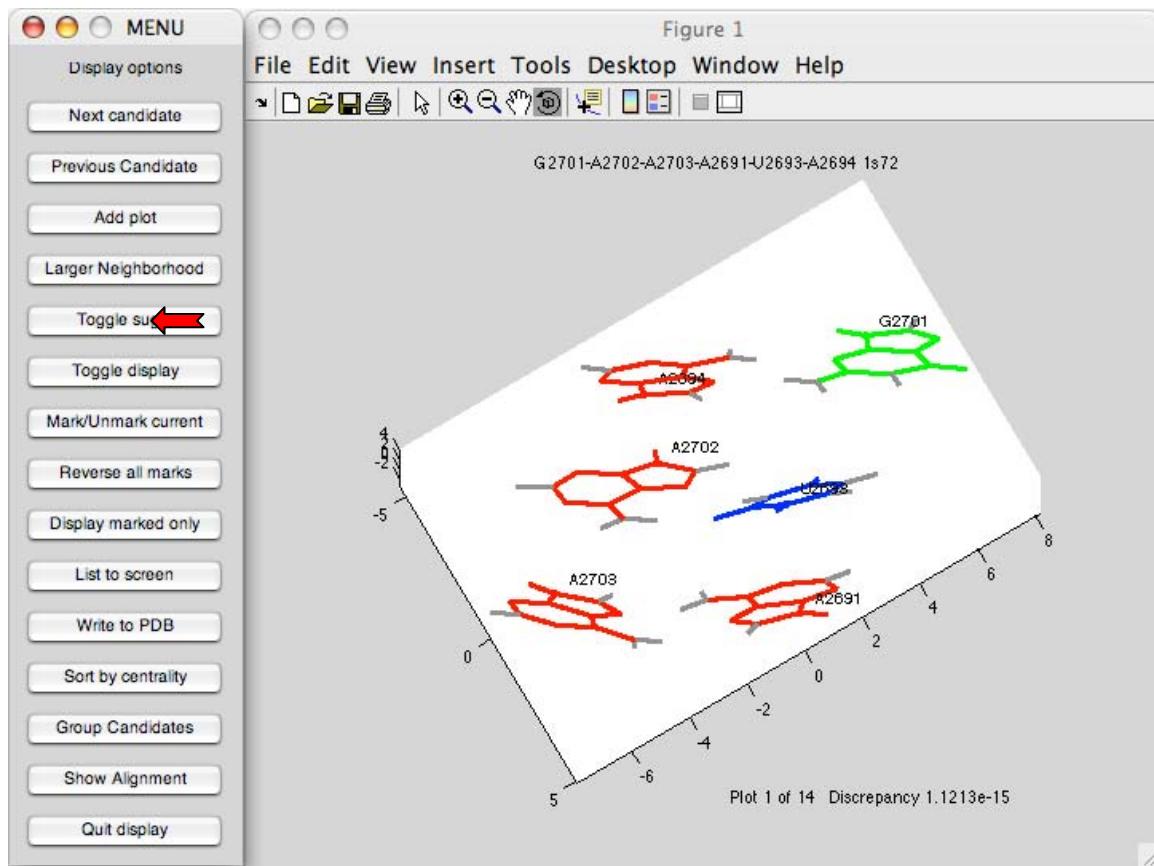
When analyzing the structures of the Candidate motifs, the user can mark each candidate they feel is what they are looking for. Once marked, they can list out just the marked candidates or view just the marked candidates. The default is unmarked, but the figure below, shows that the query motif is marked. This is a very useful when correlating it with some of the other tools, such as writing pdb files, sorting by centrality, grouping candidates and showing an alignment. For example, the user marks 5 of 7 Candidate motifs and then wants to write them out to a pdb file, only the marked Candidates will be written. This idea applies to the other tools in the menu as well.
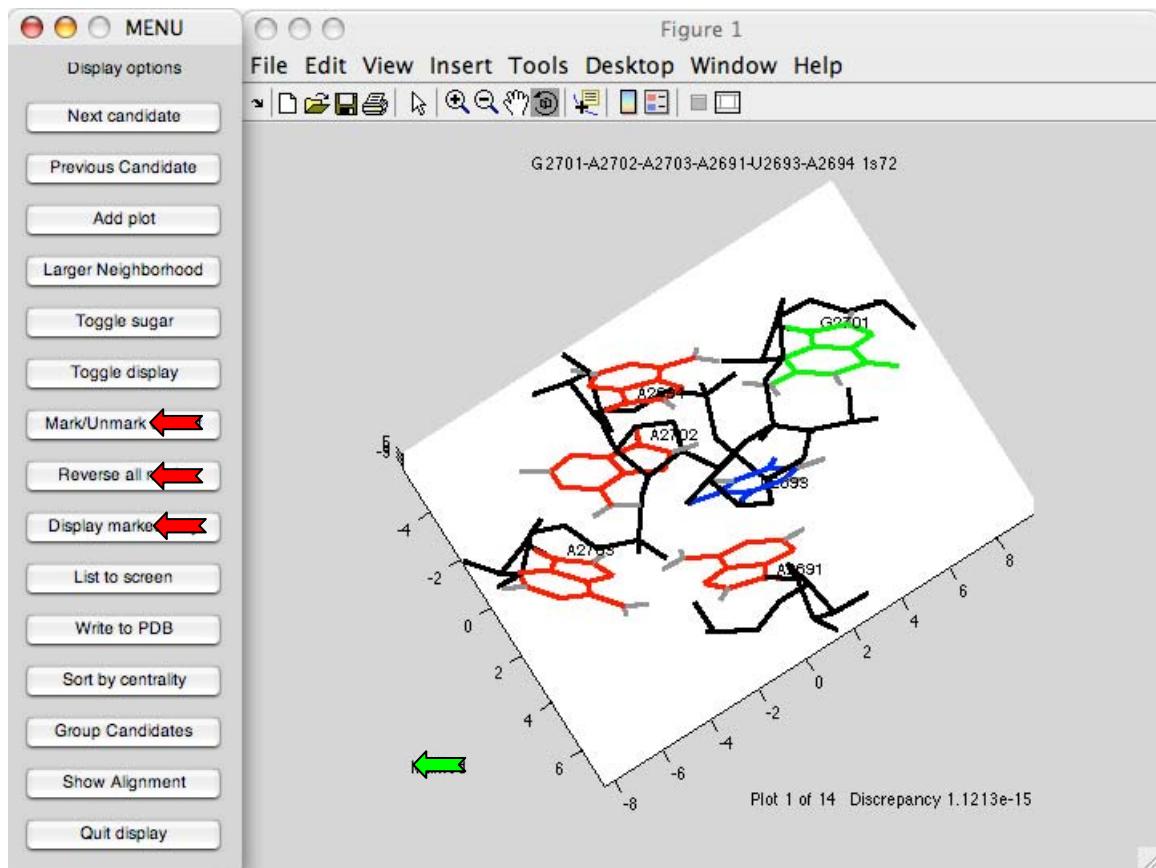
Listing Candidates

There are two methods to list the Candidates:

1) Using the FR3D GUI after performing a search and Candidates are found the user can press the List Candidates button located in the bottom-right corner of the GUI.
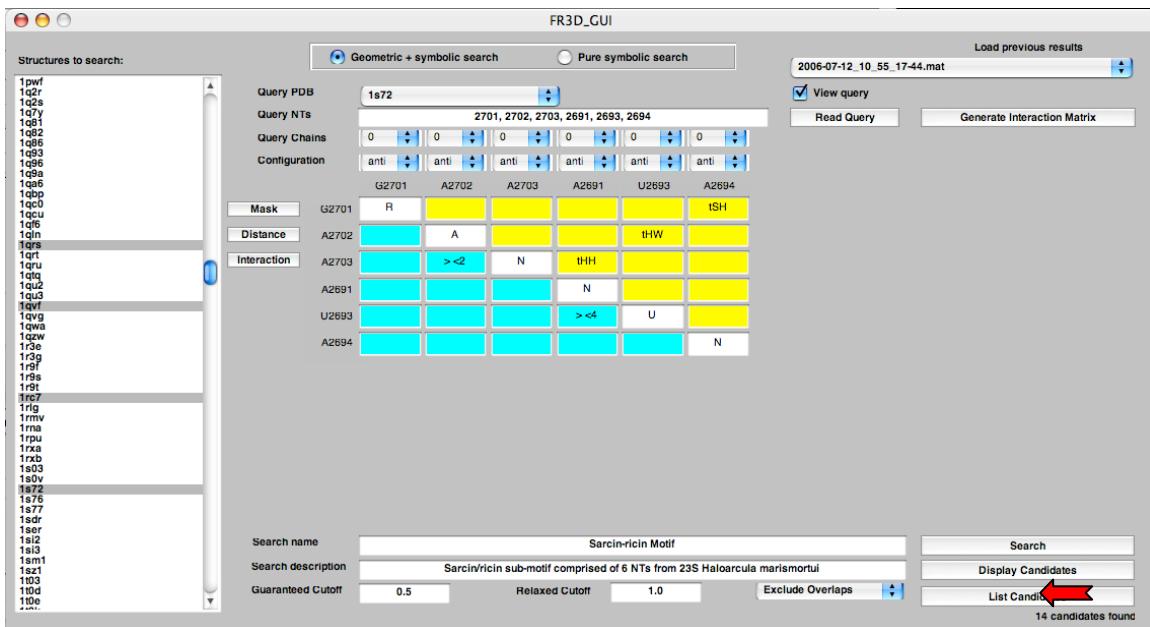
2) When the user is Displaying the Candidates they can press the List Candidates button located on the menu.
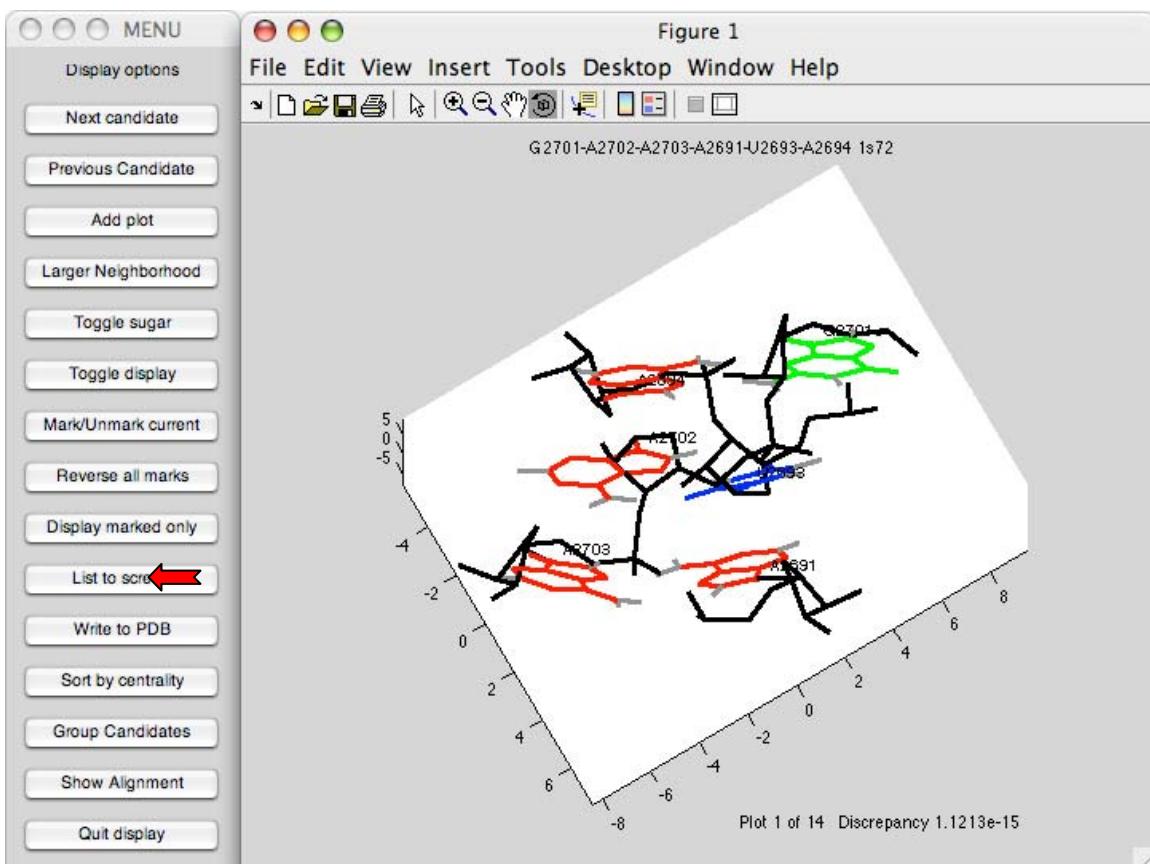
The output is displayed in the Matlab command window or, with the PC executable, in two pop-up windows. The first columns of the output look like this:

```
Query Sarcin 5 nucleotide geometric: Sarcin/ricin motif with 5 nucleotides, geometric search
```

```
Found  137457 possibilities from        1s72 in   17.109 seconds
Calculating discrepancy
Seconds remaining: 62 56 49 42 36 29 22 14 7
Found 14 candidates in the desired discrepancy range
Removed highly overlapping candidates, kept 14
Entire search took  87.8125 seconds, or   1.4635 minutes
  Filename Discrepancy      1        2        3        4        5 Chain
      1s72     0.0000  G 2692  U 2693  A 2694  G 2701  A 2702 00000
      1s72     0.0712  G 1370  U 1371  A 1372  G 2053  A 2054 00000
      1s72     0.1080  G  381  U  382  A  383  G  406  A  407 00000
      1s72     0.1275  G  588  U  589  A  590  G  568  A  569 00000
      1s72     0.1784  G  175  U  176  A  177  G  159  A  160 00000
      1s72     0.1844  G  464  U  465  A  466  G  475  A  476 00000
      1s72     0.1976  G  358  U  359  A  360  G  292  A  293 00000
      1s72     0.2284  G  213  U  214  A  215  G  225  A  226 00000
      1s72     0.2374  G   78  U   79  A   80  G  102  A  103 99999
      1s72     0.2391  G 1971  U 1972  A 1973  G 2009  A 2010 00000
      1s72     0.2491  G 1292  U 1293  A 1294  G  911  A  912 00000
      1s72     0.2714  G  953  U  954  A  955  A 1012  A 1013 00000
      1s72     0.4395  G 1543  U 1544  C 1545  C 1640  A 1641 00000
      1s72     0.4644  G  706  C  707  A  708  G  720  A  721 00000
```

The first lines tell details about the search process. FR3D screens out possible candidates to reduce the number of candidates it has to consider in detail. In this example, it found 137457 five-nucleotide motifs which could not be rejected based on the pairwise distances between their constituent nucleotides alone. This took 17 seconds. For each of these, it calculated the geometric discrepancy from the Query motif; this took an additional 70 seconds. Only 14 candidates had discrepancy less than 0.5, the default cutoff discrepancy. These candidates are listed in order of increasing discrepancy. The Query motif is listed first, with discrepancy 0.0000. Each of the five nucleotides is listed, followed by a brief listing of the chains in which the nucleotides are found. Note that the candidate with discrepancy 0.2374 was found in the 5S chain, chain 9.

The full display format includes information about the pairwise interactions between the nucleotides in each candidate, and other information. The output is quite wide, so we use a very small font here:

```
Query Sarcin 5 nucleotide geometric: Sarcin/ricin motif with 5 nucleotides, geometric search
Found  137457 possibilities from        1s72 in   17.109 seconds
Calculating discrepancy
Seconds remaining: 62 56 49 42 36 29 22 14 7
Found 14 candidates in the desired discrepancy range
Removed highly overlapping candidates, kept 14
Entire search took  87.8125 seconds, or   1.4635 minutes
  Filename Discrepancy      1        2        3        4        5 Chain  1-2  1-3  1-4  1-5  2-3  2-4  2-5  3-4  3-5  4-5 Confi  1-2  1-3  1-4  1-5  2-3  2-4  2-5  3-4  3-5  4-5
      1s72     0.0000  G 2692  U 2693  A 2694  G 2701  A 2702 00000  cSH  ntSH s33 ncSH s35  s33 tWH  tHS  s55 ncSH AAAAA  1    2    9    10   1    8    9    7    8    1
      1s72     0.0712  G 1370  U 1371  A 1372  G 2053  A 2054 00000  cSH  ntSH s33 ncSH s35  s33 tWH  tHS  s55 ncSH AAAAA  1    2   670  671   1   669  670  668  669   1
      1s72     0.1080  G  381  U  382  A  383  G  406  A  407 00000  cSH  ntSH s33 ncSH s35  s33 tWH  tHS  s55 ncSH SAAAA  1    2    25   26   1    24   25   23   24   1
      1s72     0.1275  G  588  U  589  A  590  G  568  A  569 00000  cSH  ntSH s33 ncSH s35  s33 tWH  tHS  s55 ncSH SAAAA  1    2    20   19   1    21   20   22   21   1
      1s72     0.1784  G  175  U  176  A  177  G  159  A  160 00000  cSH  ntSH s33 ncSH s35  s33 tWH  tHS  s55 ncSH AAAAA  1    2    16   15   1    17   16   18   17   1
      1s72     0.1844  G  464  U  465  A  466  G  475  A  476 00000  cSH  ntSH s33 ncSH s35  s33 tWH  tHS  s55 ncSH AAAAA  1    2    11   12   1    10   11    9   10   1
      1s72     0.1976  G  358  U  359  A  360  G  292  A  293 00000  cSH  ntSH s33 ncSH s35  s33 tWH  tHS  s55 ncSH SAAAA  1    2    66   65   1    67   66   68   67   1
      1s72     0.2284  G  213  U  214  A  215  G  225  A  226 00000  cSH  ntSH s33 ncSH s35  s33 tWH  tHS  s55 ncSH AAAAA  1    2    12   13   1    11   12   10   11   1
      1s72     0.2374  G   78  U   79  A   80  G  102  A  103 99999  cSH  ntSH s33 ncSH s35  s33 tWH  tHS  s55  -   SAAAA  1    2    24   25   1    23   24   22   23   1
      1s72     0.2391  G 1971  U 1972  A 1973  G 2009  A 2010 00000  cSH  ntSH s33 ncSH s35  s33 tWH  tHS  s55  -   AAAAA  1    2    38   39   1    37   38   36   37   1
      1s72     0.2491  G 1292  U 1293  A 1294  G  911  A  912 00000  cSH  ntSH s33 ncSH s35  s33 tWH  tHS  s55 ncSH AAAAA  1    2   353  352   1   354  353  355  354   1
      1s72     0.2714  G  953  U  954  A  955  A 1012  A 1013 00000  cSH  ncSH s33 ncSH s35 ncHW tWH  tHS  s55 ns35 AAAAA  1    2    31   32   1    30   31   29   30   1
      1s72     0.4395  G 1543  U 1544  C 1545  C 1640  A 1641 00000  ncSH  -   s33 ntSH s35 ns33 tWH  tHS  s55 ns35 AAAAA  1    2    96   97   1    95   96   94   95   1
      1s72     0.4644  G  706  C  707  A  708  G  720  A  721 00000  ns35 ntSH s33 ntSH s35  s33 tWH  tHS  s55 ncSH AAAAA  1    2    13   14   1    12   13   11   12   1
```

The columns following the Chain column indicate the basepairing or base stacking interactions between the nucleotides noted at the top of the column. For instance, in each of the candidates, nucleotides 2 and 5 form a tWH (*trans* Watson-Crick / Hoogsteen) basepair. The column headed Configuration indicates the configuration of each base, whether anti (A) or syn (S). The final columns indicate the differences in nucleotide numbers between the indicated nucleotides. This makes it easy to spot local versus composite motifs. In this case, all of the candidates consist of two strands, one corresponding to 2692:2694, the other corresponding to 2701:2702.

Writing Candidates to a PDB File

To write the Candidate motifs to a PDB file, which can be view using 3D visualization tools, the user can press the Write to PDB button in the menu.
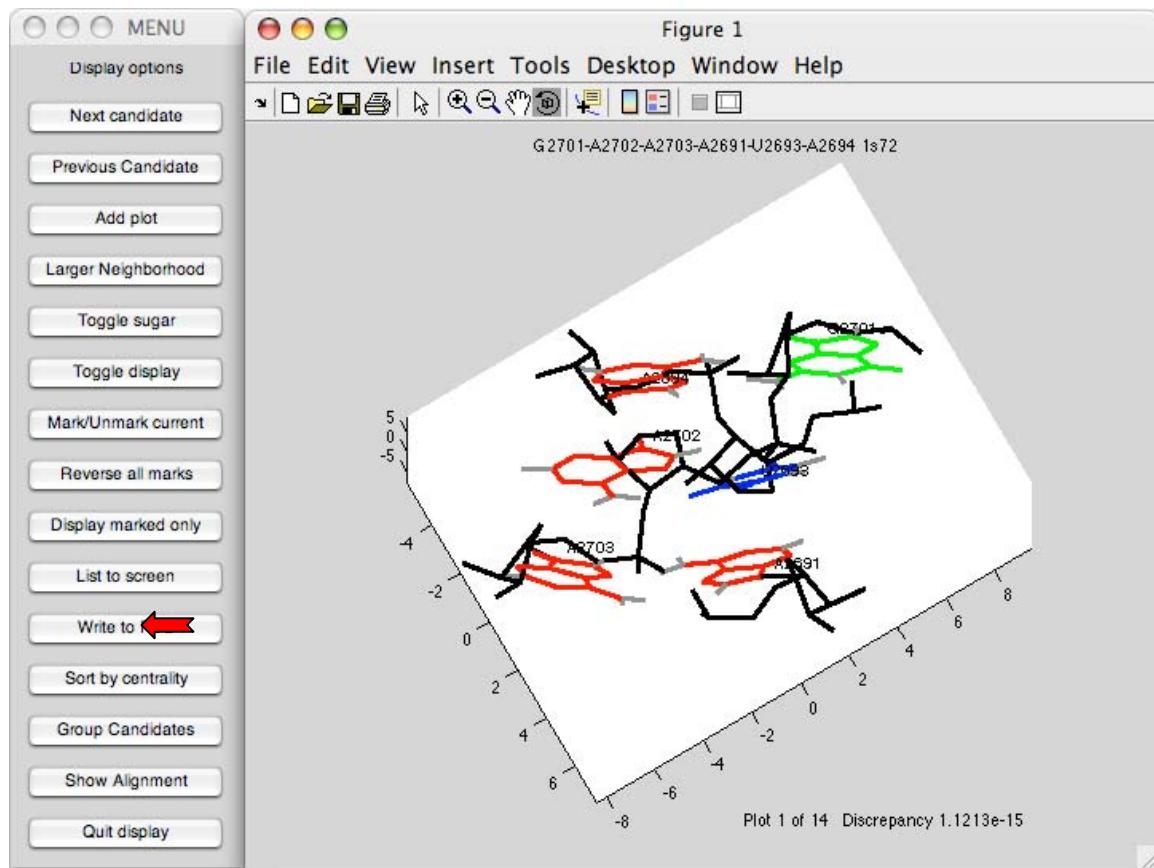
After pressing the Write to PDB but, the pdb filename will be displayed in the command window. The following is an example of an output printed to the command window. The file should be stored in the user's local working directory.

Wrote 2007-01-25_12_12_47-Sarcin-Cand.pdb

New Display Candidates Menu as of April 2009

The Display Candidates menu has been updated as of April 2009. Most features work the same as described above, but there are some new features, which are explained below.

MENU

Display options

- Next candidate
- Previous Candidate
- Add plot
- Larger Neighborhood
- Toggle sugar
- Toggle display
- Mark/Unmark current
- Reverse all marks
- Display marked only
- List to screen
- Write to PDB
- Sort by centrality
- Order by Similarity
- Show Alignment
- Show Scatterplot
- Navigate with Fig 99
- Quit display

Graphical display of discrepancies between candidates; Figure 99

When the user presses the buttons "Sort by centrality", "Order by similarity", or "Navigate using Fig 99", FR3D calculates the geometric discrepancies between the candidates and displays them in Figure 99. The candidates are listed in the same order on the vertical and horizontal axes. The discrepancy between a candidate and itself is zero, so the discrepancies down the diagonal are zero, which are displayed as red boxes. All pairs of candidates are considered and displayed.
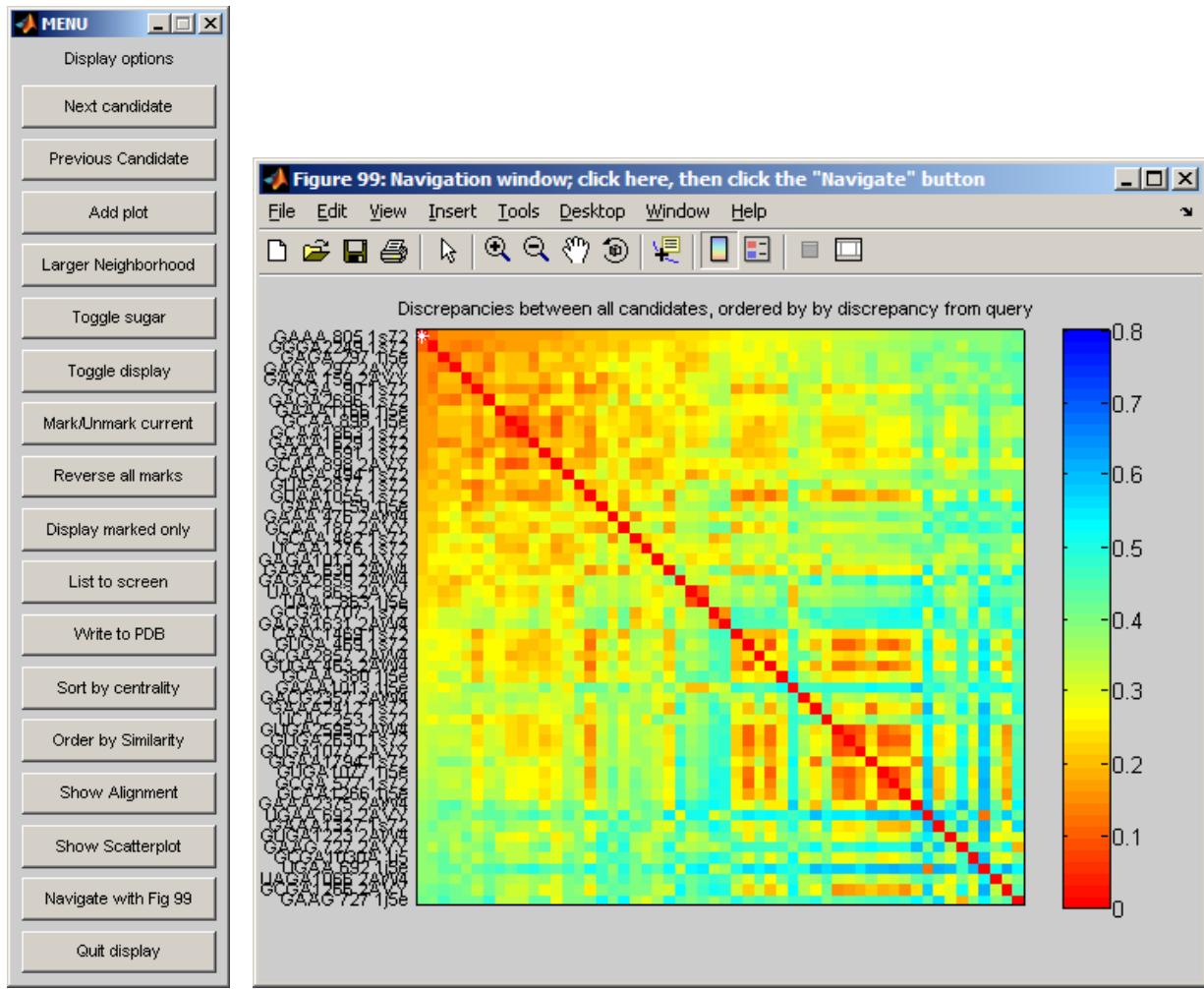
(Initially, FR3D only calculates discrepancies between the first 300 candidates; even this takes time. To increase the limit, go to the first candidate, press Previous Candidate, then Next Candidate. Each time you do this, it doubles the number of candidates whose discrepancies will be calculated. Be patient as FR3D calculates more discrepancies.)

The candidate currently being displayed is indicated with a white star on the diagonal. If multiple plot windows are open to show candidates, the other candidates being displayed are indicated by other white symbols. If some candidates have been marked, this is shown with white dots on the diagonal.

Navigating with Figure 99

To jump to a candidate of interest, click on the diagonal in Figure 99 in the row and column corresponding to the candidate, then click "Navigate with Fig 99" in the menu. FR3D will jump to the specified candidate. To mark a group of candidates that are contiguous in Figure 99, click off the diagonal at the upper right or lower left corner of the square containing the discrepancies between the candidates you want to mark, then click "Navigate with Fig 99" in the menu. FR3D will mark the candidates you indicated.

Sorting by Centrality

Pressing the "Sort by Centrality" button on the menu will re-order the candidates. The candidate whose average discrepancy with all other candidates is lowest will appear first, and so on. In this way, the user can find the centroid for their Candidate motifs. The centroid is often a good choice as the query motif for the next search. In the figure below, the centroid candidate is from 1s72 and is listed at the top of the graph. "List to Screen" will list the candidates in this order.
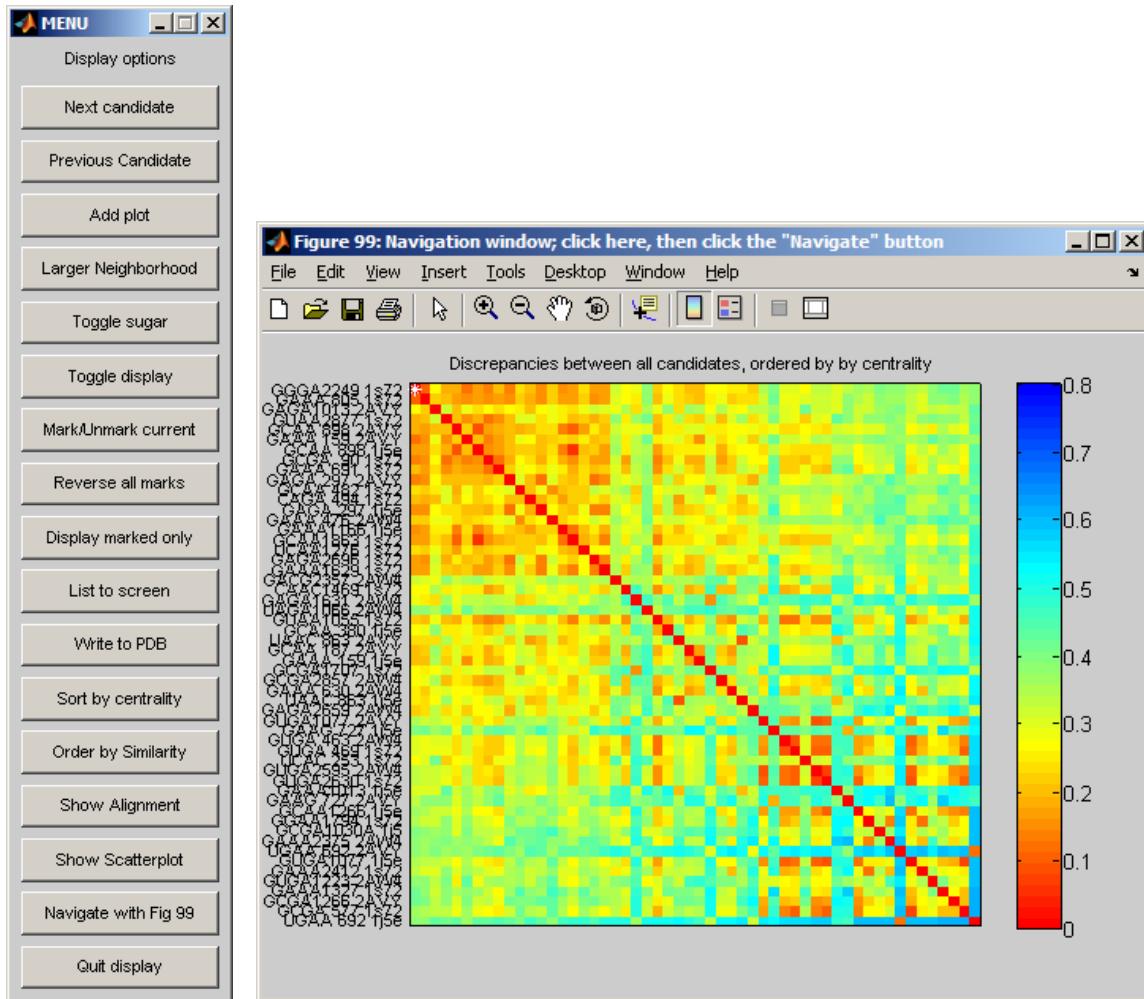

Order by Similarity

Pressing the "Order by Similarity" button will re-order the candidates in such a way that candidates which are geometrically similar to one another appear near each other in the list. This makes it possible to identify clusters of candidates which share geometric features.
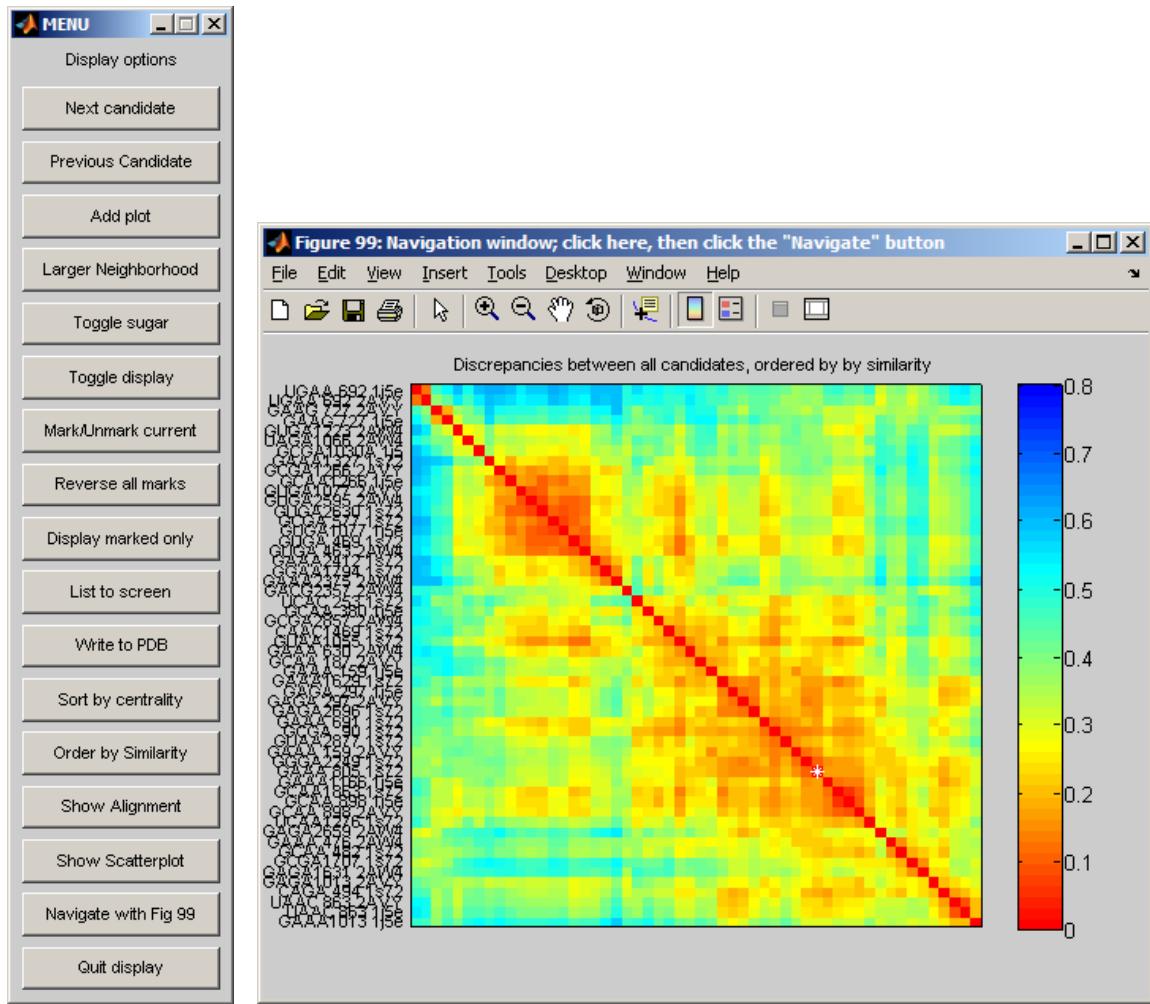
Aligning candidates

The user can look at the sequence alignment of each of the Candidates by pressing the Show Alignment button in the menu.
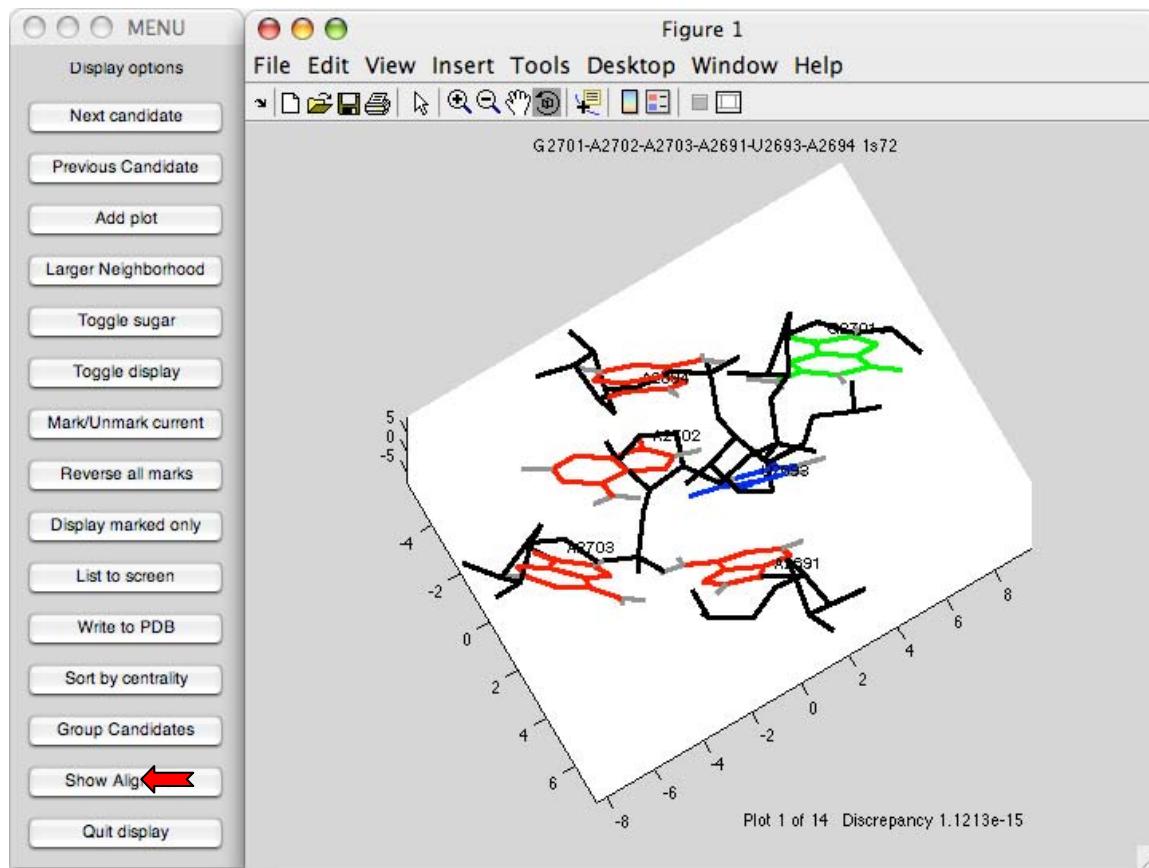
The Candidate motifs along with their alignment will be printed in the command window, as shown by the example below.

```
                                                             1 23    456
1s72    0.0000  A2691  U2693  A2694  G2701  A2702  A2703   AGUA....GAA
1qvf    0.0138  A2691  U2693  A2694  G2701  A2702  A2703   AGUA....GAA
1qvf    0.0954  A1369  U1371  A1372  G2053  A2054  A2055   AGUA....GAA
1s72    0.1046  A1369  U1371  A1372  G2053  A2054  A2055   AGUA....GAA
1s72    0.1451  A 380  U 382  A 383  G 406  A 407  A 408   AGUA....GAA
1qvf    0.1518  A 380  U 382  A 383  G 406  A 407  A 408   AGUA....GAA
1s72    0.1742  A  77  U  79  A  80  G 102  A 103  A 104   AGUA....GAA
1s72    0.1790  A 212  U 214  A 215  G 225  A 226  A 227   AGUA....GAA
1qvf    0.1804  A3077  U3079  A3080  G3102  A3103  A3104   AGUA....GAA
1s72    0.1841  A 174  U 176  A 177  G 159  A 160  A 161   AGUA....GAA
1qvf    0.1861  A 212  U 214  A 215  G 225  A 226  A 227   AGUA....GAA
1qvf    0.1872  A 463  U 465  A 466  G 475  A 476  A 477   AGUA....GAA
1qvf    0.1875  A 174  U 176  A 177  G 159  A 160  A 161   AGUA....GAA
1s72    0.1937  A 463  U 465  A 466  G 475  A 476  A 477   AGUA....GAA
```

Retrieving the results of previous searches

To retrieve previous search results the user can select one of their previous searches using the Load previous search drop-down menu located at the top-right corner of the GUI.

User-maintained lists of PDB files

To facilitate searching a subset of the entire collection of PDB files in the Matlab search path, the user may maintain lists of PDB files. To do so, create a text file with a name ending with "_list.pdb". For example, the file **Nonredundant_list.pdb** has these lines:

```
2AW4
2AVY
1s72
1j5e
```

Note that case does not matter. In the FR3D GUI, the list **Nonredundant_list** will appear in the list of PDB files to search. Selecting this list will include all named files in the search. When the results of the search are saved, all files in the list are saved by name, so that when the results of the search are loaded again later, the individual files that were searched will be highlighted so that it is clear which files were searched. When using xSpecifyQuery to specify searches, names of lists can appear for the PDB files to be searched.

Appendix

FR3D includes additional programs that may be of interest. These are a little harder to use, however. From the Matlab command prompt, >>, load PDB data this way:

```
>> File = zAddNTData('Nonredundant_list');
```

Specify searches in the Matlab program file xSpecifyQuery, then execute the search using:

```
>> FR3D
```

Searches are saved as usual and may be retrieved later using FR3D.

References

Leontis, N.B., Stombaugh, J. and Westhof, E. (2002) Motif prediction in ribosomal RNAs Lessons and prospects for automated motif prediction in homologous RNA molecules, *Biochimie*, **84**, 961-973.

Leontis, N.B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs, *RNA*, **7**, 499-512.

Sarver, M., Zirbel, C. L., Stombaugh, J., Mokdad, A., Leontis, N. B. (2007) FR3D: Finding Local and Composite Recurrent Structural Motifs in RNA 3D Structures. To appear in the Journal of Mathematical Biology.