Supplemental methods

1. Instructions

1.1. Installation

4MRNA can be installed by sequentially executing the following commands (Script S4) in a command-line interface (CLI) that can run shell scripts, such as Windows Subsystem for Linux (WSL) on Windows, Terminal on macOS, or a Linux terminal. In addition, the external programs *3DNA*, *Phenix*, and *CCP4* are required. These programs should be installed separately from their official websites according to the instructions provided there, and they should be executable from the command line.

Script S4: Installation commands of 4MRNA.

```
cd $HOME
git clone https://github.com/S-Ando-Biophysics/4MRNA-Install.git
bash 4MRNA-Install/install.sh
echo 'export PATH="$HOME/4MRNA-Install/bin:$PATH"' >> ~/.bashrc
source ~/.bashrc
```

1.2. Usage

First, enter a command in the CLI to move to the directory containing the reflection file (MTZ format). It is also recommended to create a new directory for 4MRNA and copy the reflection file into this directory. Next, enter "4MRNA" in the CLI. These steps are summarized in Script S5.

Script S5: Commands to start 4MRNA.

```
cd <path_to_directory_containing_the_reflection_file>
4MRNA
```

After 4MRNA is launched, the user is first asked how many types of models will be used for MR. This question corresponds to how many stem regions the target molecule should be divided into, as described in detail in the Tutorials section below. For each model, the user is then asked to enter three types of information: the nucleic acid type, selected from A-form DNA, B-form DNA, and A-form RNA; the sequence; and the number of copies to be searched for in the asymmetric unit (ASU) of the crystal. After all this information has been entered, the setup of the scripts running in the background of 4MRNA is completed.

Finally, the user selects either the default mode or the customize mode. In the default mode, all steps of 4MRNA are performed automatically, and the user only needs to wait for the results. This mode is recommended for all users. For expert users, the customize mode is also provided. When the customize mode is selected, the background scripts are downloaded and the program then terminates, allowing the user to edit the scripts manually and test more advanced options.

1.3. Checking the results

After 4MRNA has finished, a directory named "4MRNA-Results" is created. This directory contains three to seven MR results, which are automatically selected from the top-ranked solutions based on the MR statistics LLG and TFZ. These statistics are summarized in the file "4MRNA-results.txt".

The user should first check the statistical values listed in "4MRNA-results.txt". Based on our experience, for nucleic acid structures, solutions with LLG values greater than 100 and TFZ values greater than 8 may indicate a possible MR solution. Using these values as practical criteria, the user should visually inspect the agreement between the electron density map and the model, starting from the solutions with better statistics. If the model agrees well with the electron density map, the user can proceed to refinement. In some cases, multiple solutions may show good statistics exceeding the above criteria and reasonable agreement with the electron density map. In such cases, our recommendation is to select the solution with the highest TFZ value first. If multiple solutions have the same TFZ value, the solution with the highest LLG value among them should be chosen.

2. Tutorials

2.1. Basic example (PDB ID 257D)

In this section, PDB ID 257D is used as a basic example to demonstrate how to run 4MRNA. 4MRNA was performed according to the following steps. The commands and required inputs are summarized in Script S6.

1. Launch the CLI
2. Move to the directory containing the reflection file.
3. Execute the command "4MRNA".
4. Enter the requested information in the CLI. In this example, the model information is as follows.
    - The number of models for MR: Since the entire molecule forms one duplex (it is treated as one stem region), the number of models for MR is entered as 1.
    - Nucleic acid type: A-form DNA.
    - Sequence: GCCGGC. (Only one strand of the duplex needs to be entered.)
    - How many copies to be searched in MR: Based on the unit cell parameters and space group, the ASU was estimated to contain one molecule.
5. Select the default mode. (In general, the default mode is recommended.)

Script S6: Commands and required inputs for a basic example. (white: questions displayed in the CLI)

```
cd <path_to_directory_containing_the_reflection_file>
4MRNA
The number of models for molecular replacement [1,2,3]: 1
Nucleic acid type [A-DNA,B-DNA,A-RNA]: A-DNA
Sequence (one strand of duplex, 5'→3'): GCCGGC
How many copies of the model to be searched in molecular replacement: 1
Please choose execution mode. [default,customize]: default
```

After 4MRNA had finished, the results were checked in the following steps.

1. Open the file “4MRNA-Results.txt” in the directory “4MRNA-Results” (Figure S16 and S17).
2. Check the statistics. In this example, all selected solutions exceeded the practical criteria of LLG ≧ 100 and TFZ ≧ 8. Among these solutions, the highest TFZ value was 11.8. Among the solutions with this TFZ value, the highest LLG value was 223. Therefore, based on the MR statistics, the MR result labeled “3DNAAD-12-7-14” was judged to be the best solution.
3. Open the electron density map (MTZ format) and the model coordinate file (PDB format) contained in the directory labeled “phaser-3DNAAD-12-7-14” using a molecular modeling software such as *Coot* (Figure S18 and S19). Then, visually check whether the model agrees well with the electron density map.

Since the MR statistics were good and the model agreed well with the electron density map, MR using 4MRNA was considered successfully completed. The selected solution was then used for subsequent refinement by the preferred method.

Figure S16: Contents of the directory “4MRNA-Results”.

```
3DNAAD-14-5-14 254 11.7
3DNAAD-12-5-14 245 11.6
3DNAAD-12-7-14 223 11.8
3DNAAD-21-14-14 220 11.8
3DNAAD-14-14-14 241 11.7
```

Figure S17: Contents of the file “4MRNA-Results.txt”.

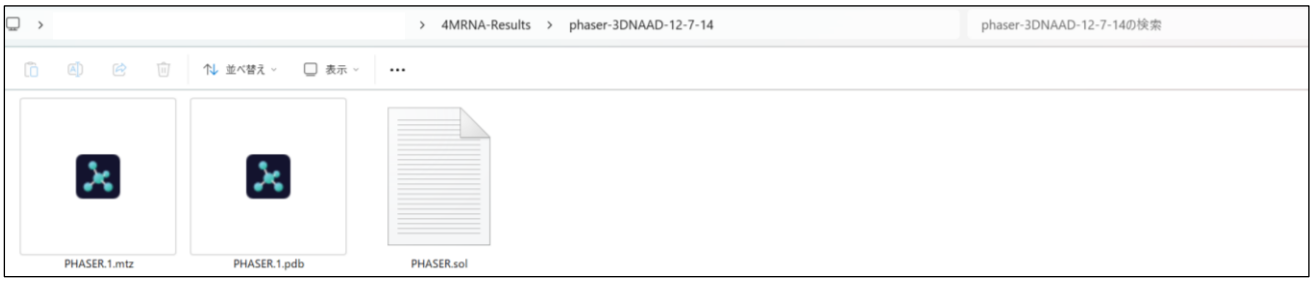

Figure S18: Contents of the directory “phaser-3DNAAD-12-7-14”.

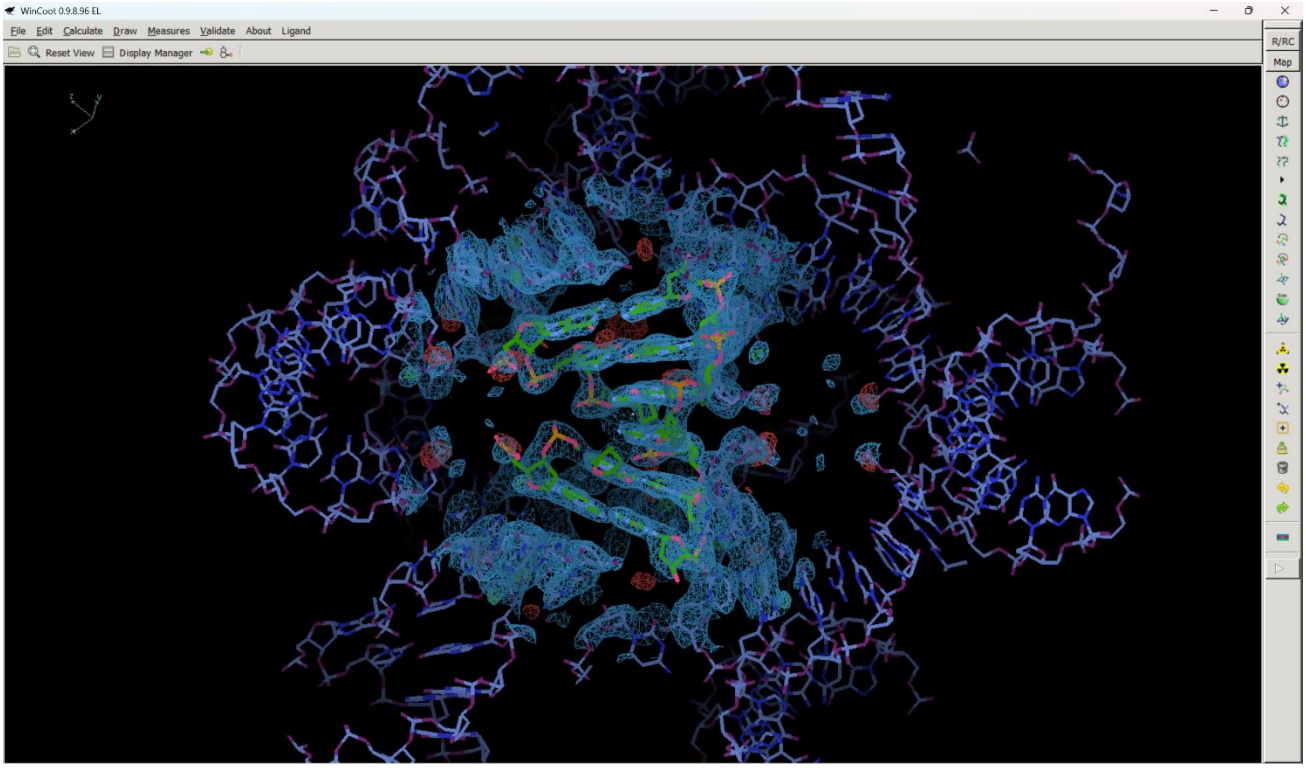

Figure S19: The electron density map and model opened in *Coot*.

2.2. Example using multiple models (PDB ID 3TD0)

In this section, PDB ID 3TD0 is used as an example to demonstrate how to enter the required information when multiple models are used.

This nucleic acid was designed so that the self-complementary sequence hybridizes to form double helical structure containing two internal loops, as shown in Figure S20. When the resulting duplex is considered from one side, it can be divided into a 5-bp stem, an internal loop, an 8-bp stem, an internal loop, and another 5-bp stem. The two terminal 5-bp stems have the same sequence because of self-complementarity. Therefore, in 4MRNA, MR is performed using two types of models: an 8-bp stem model and 5-bp stem models. In other words, the target molecule contains two types of stem regions. Accordingly, when asked for "The number of models for molecular replacement", the user should enter 2.

The information for the first model (8-bp stem model) is as follows.

- Nucleic acid type: A-form RNA.
- Sequence: CGUCGACG. (Only one strand of the duplex needs to be entered.)
- Number of copies to be searched for in MR: The ASU was estimated to contain one entire double-helical molecule based on the unit cell parameters and space group. Since the 8-bp stem occurs at one position in this molecule, the number of copies to be searched for is 1 × 1 = 1.

The information for the second model (5-bp stem model) is as follows.

- Nucleic acid type: A-form RNA.
- Sequence: GCGUC. (Only one strand of the duplex needs to be entered.)

- Number of copies to be searched for in MR: The ASU was estimated to contain one entire double-helical molecule based on the unit cell parameters and space group. Since the 5-bp stem occurs at two positions in this molecule, the number of copies to be searched for is 1 × 2 = 2.

After all required inputs had been provided, the default mode was selected to start the 4MRNA workflow. The procedures after completion of 4MRNA can be the same as those described in Tutorial 2.1.

The commands and required inputs are summarized in Script S7.

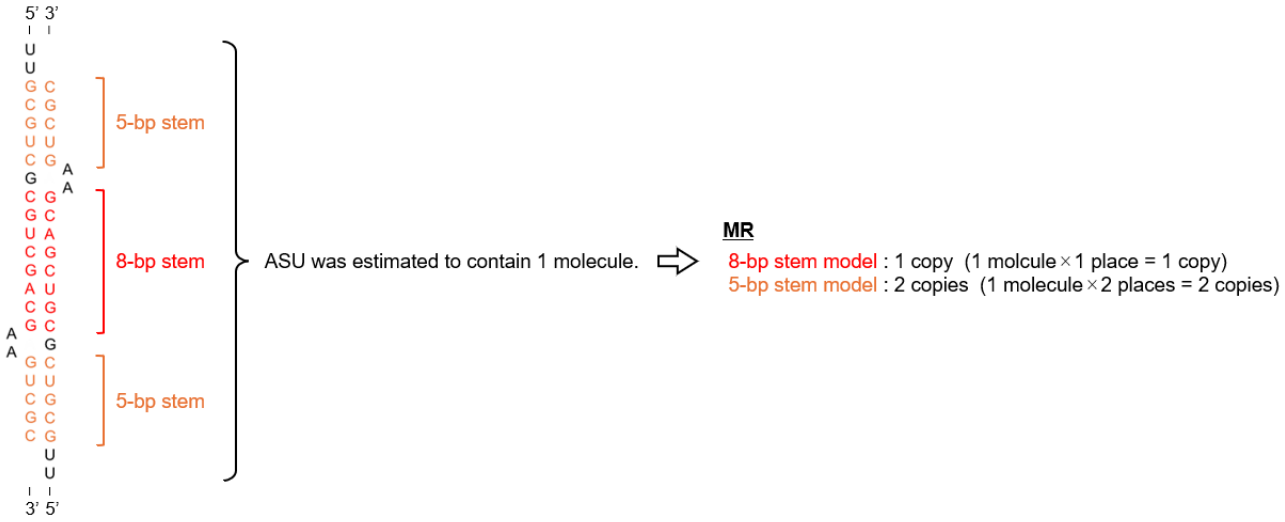

Figure S20: Sequence and the number of molecules in the asymmetric unit.

Script S7: Commands and required inputs for an example using multiple models.

```
cd <path_to_directory_containing_the_reflection_file>
4MRNA
The number of models for molecular replacement [1,2,3]: 2

Enter information of model #1.
Nucleic acid type [A-DNA,B-DNA,A-RNA]: A-RNA
Sequence (one strand of duplex, 5'→3'): CGUCGACG
How many copies of the model to be searched in molecular replacement: 1

Enter information of model #2.
Nucleic acid type [A-DNA,B-DNA,A-RNA]: A-RNA
Sequence (one strand of duplex, 5'→3'): GCGUC
How many copies of the model to be searched in molecular replacement: 2

Please choose execution mode. [default,customize]: default
```

2.3. Example of a structure with only one strand in the asymmetric unit (PDB ID 118D)

In some cases, only one strand is present in the ASU, and this strand forms a double helix with a symmetry-related strand generated by crystallographic symmetry. In such cases, a single-stranded model should be prepared for MR. Although 4MRNA is basically designed to use duplex models composed of two strands, it also supports cases in which only one strand is present in the ASU. In this case, the user can enter 0.5 as the number of copies to be searched for in MR. This value "0.5" means half of one double-helical molecule, corresponding to one strand.

Here, PDB ID 118D (Bingman et al. 1992) is used as an example. Based on the unit cell parameters and space group, the ASU was estimated to contain one strand (0.5 of a duplex molecule). The commands and inputs in the CLI for this example are shown in Script S8. In addition, the best MR result displayed in *Coot* is shown in Figure S21. This figure shows that the strand in the ASU forms a duplex with a symmetry-related strand.

Script S8: Commands and required inputs for an example of a structure with only one strand in the ASU.

```
cd <path_to_directory_containing_the_reflection_file>
4MRNA
The number of models for molecular replacement [1,2,3]: 1
Nucleic acid type [A-DNA,B-DNA,A-RNA]: A-DNA
Sequence (one strand of duplex, 5'→3'): GTGCGCAC
How many copies of the model to be searched in molecular replacement: 0.5
Please choose execution mode. [default,customize]: default
```

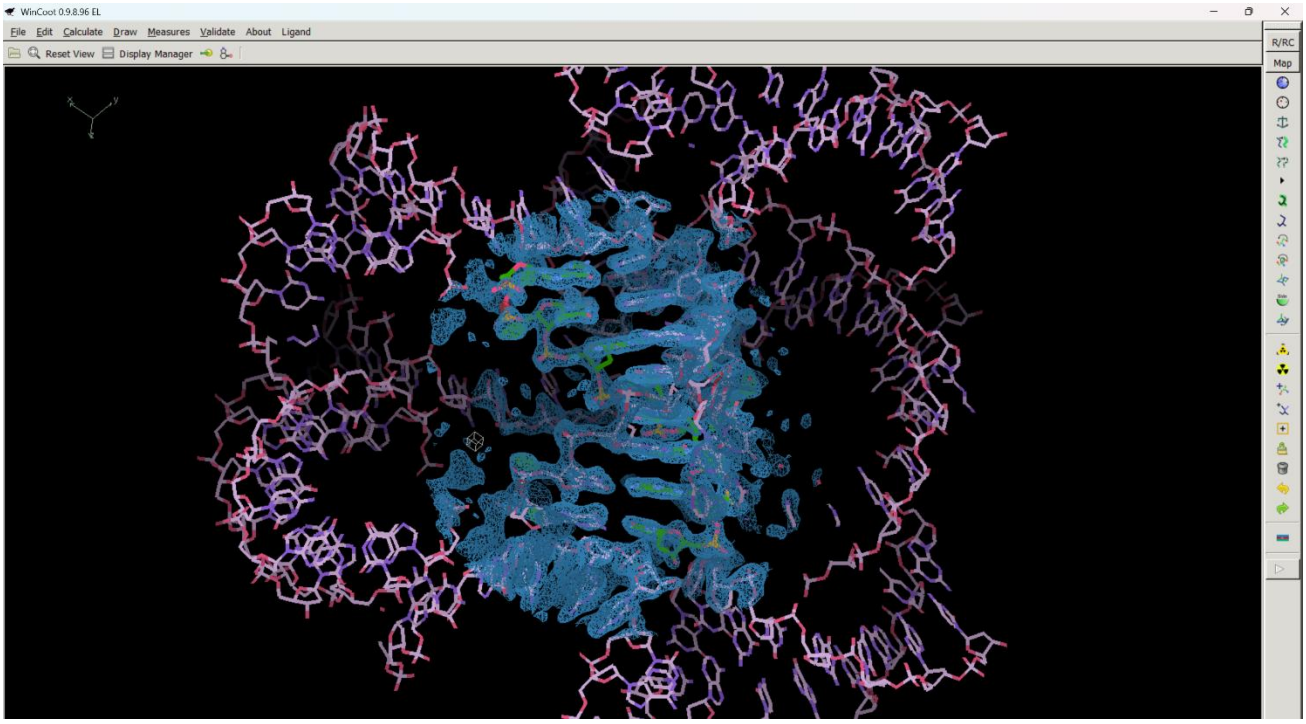

Figure S21: The best MR result opened in *Coot*. (Green: strand in the ASU; pink: symmetry-related strand)

2.4. Example of a structure with an odd number of strands in the asymmetric unit (PDB ID 5E36)

In some cases, an odd number of strands are present in the ASU. Some of these strands can form duplexes within the ASU, whereas the remaining strand can form a duplex with a symmetry-related strand generated by crystallographic symmetry. For example, when three strands are present in the ASU, two strands can form a duplex within the ASU, while the remaining one strand forms a duplex with a symmetry-related strand.

In such cases, the duplexes formed within the ASU should first be determined. The remaining strand can then be placed manually if the corresponding electron density is clearly visible. Alternatively, an additional MR search can be performed using a single-stranded model with the already determined duplexes fixed.

Here, PDB ID 5E36 (Wang et al. 2019) is used as an example. Based on the unit cell parameters and space group, the ASU was estimated to contain three strands (two duplex molecules and one strands). The commands and inputs in the CLI for this example are shown in Script S9.

Script S9: Commands and required inputs for an example of a structure with three strands in the ASU.

```
cd <path_to_directory_containing_the_reflection_file>
4MRNA
The number of models for molecular replacement [1,2,3]: 1
Nucleic acid type [A-DNA,B-DNA,A-RNA]: A-DNA
Sequence (one strand of duplex, 5'→3'): GTGTACAC
How many copies of the model to be searched in molecular replacement: 1
Please choose execution mode. [default,customize]: default
```

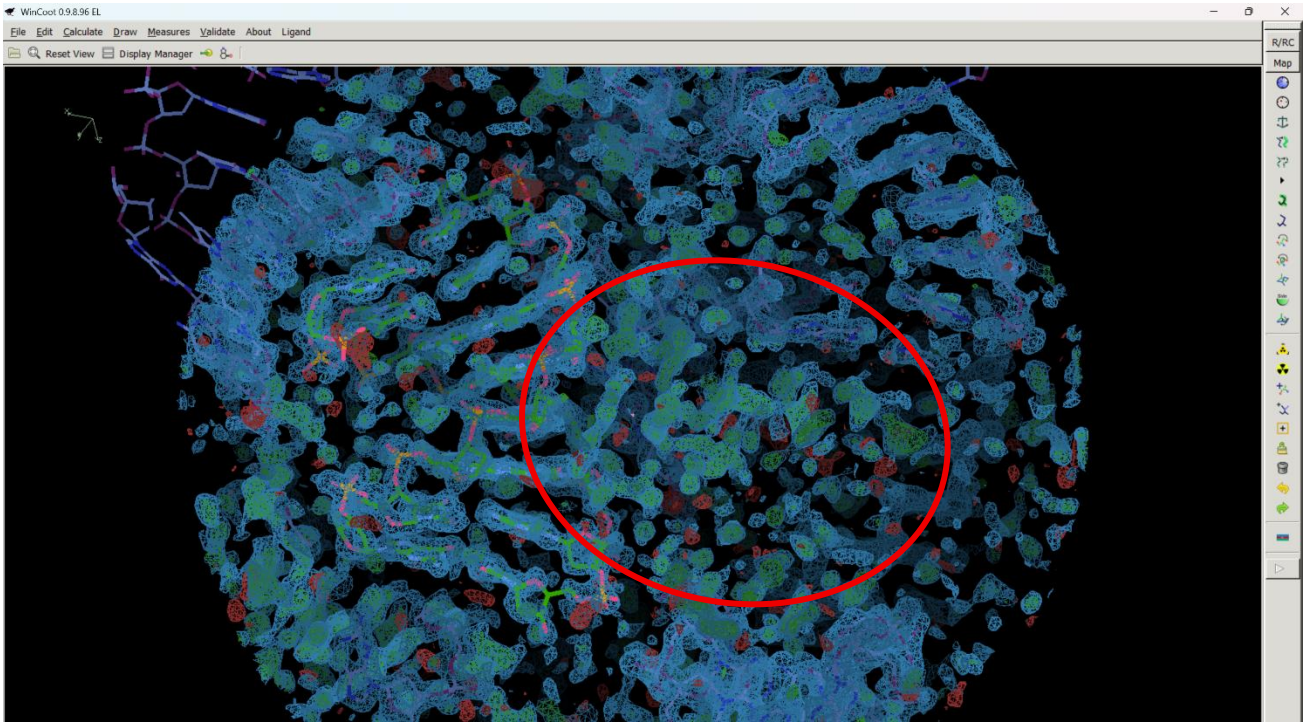

Figure S22: The best MR result obtained by 4MRNA opened in *Coot*.

The best MR result obtained by 4MRNA was displayed in *Coot*, as shown in Figure S22. This figure confirmed that the electron density corresponding to the unplaced strand was clearly visible. The enlarged view of this region is shown in Figure S23A. Since the third G was most clearly visible in the electron density map, this base was first placed manually (Figures S23B and S23C). Starting from this G, the surrounding residues were placed manually (Figure S23D).

After the remaining strand had been built and refinement was performed, the resulting model was obtained as shown in Figure S24. The remaining strand placed by manually forms a duplex with a symmetry-related strand.

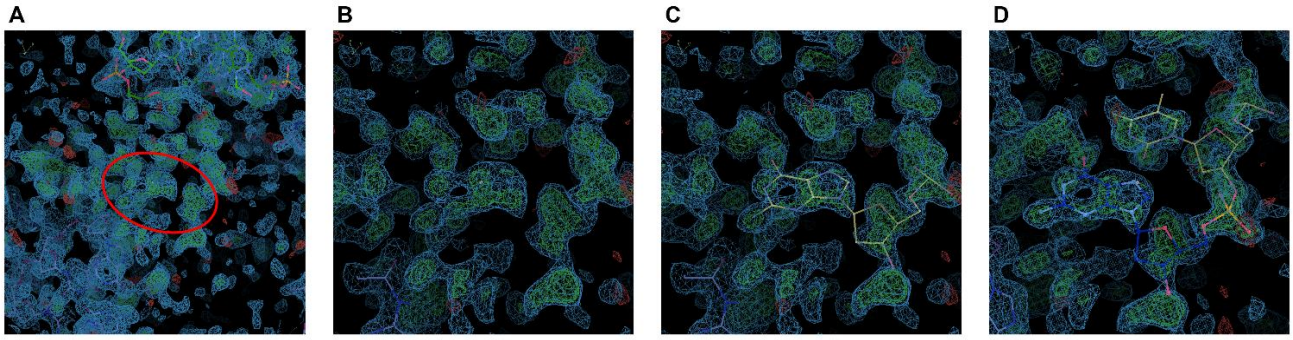

Figure S23: Manual modeling of the remaining strand based on the electron density map.

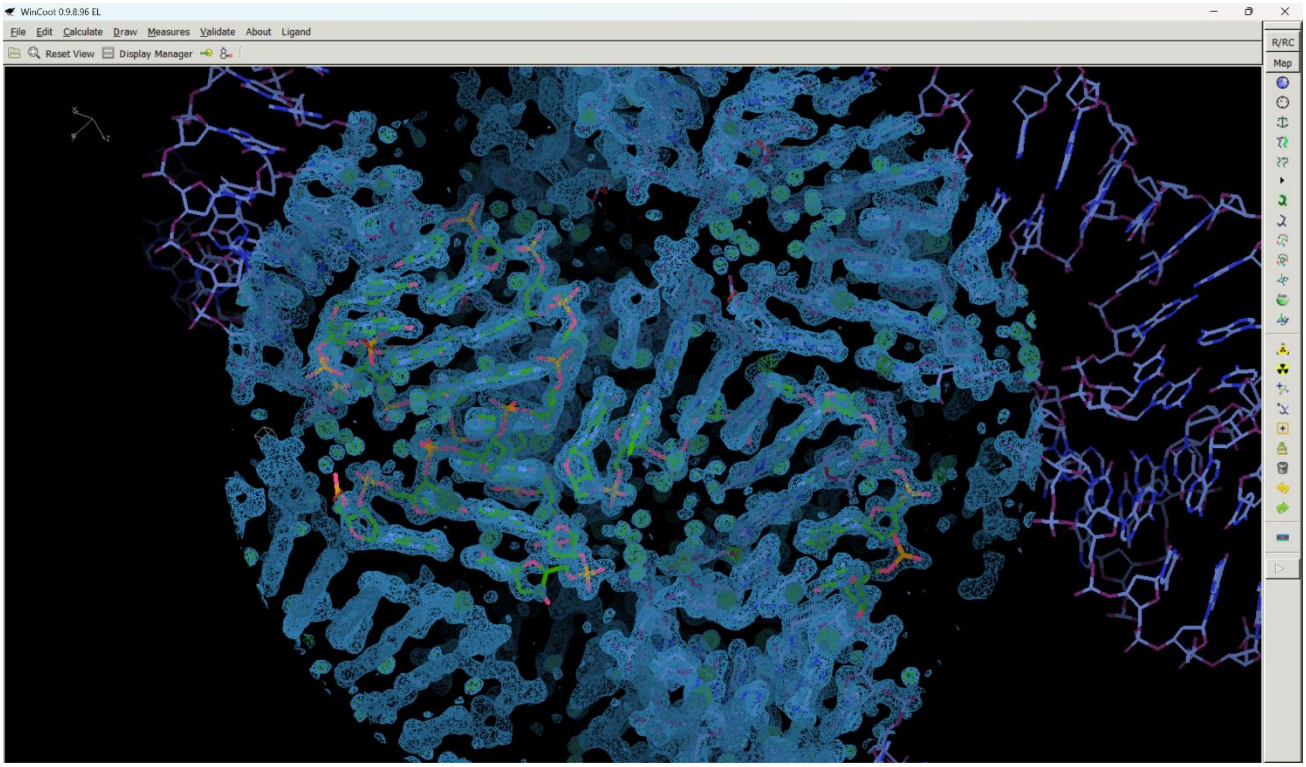

Figure S24: Refined model after manual modeling of the remaining strand.