

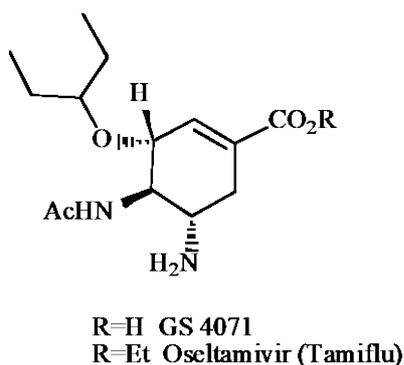
## MOLECULAR MODELLING EXERCISES

### ANTIVIRAL AGENTS

#### EXERCISE 20.5 Oseltamivir

##### INTRODUCTION

Oseltamivir (Fig. 1) is used for the treatment of influenza. In fact, it is a prodrug for the antiviral drug GS4071, which acts as an inhibitor of the viral neuraminidase enzyme (section 20.8.3 in the textbook). In this exercise, you will study the active conformation of GS4071 and identify how it interacts with amino acid residues in the binding site.



**Figure 1** GS4071 and oseltamivir.

##### INSTRUCTIONS

It is suggested that you attempt to carry out the following instructions yourself before following the more detailed **Procedures** that follow. You may find the file entitled **Common Operations for ChemBio3D/Chem3D** a useful guide on how to carry out various operations. The ChemDraw file for GS4071 can be accessed from the ChemDraw window.

##### PART A

- \*Create the energy-minimised structure of GS4071.
- \*Identify rotatable bonds and intramolecular H-bonds.
- \*Use molecular dynamics to identify the flexibility of the structure and the likelihood of different conformations.

##### PART B

- \*Download the crystal structure of the neuraminidase enzyme with bound GS4071 (pdb 2HU4).
- \*Identify the ligand and the binding site within the crystal structure.
- \*Identify the various components of the crystal structure.

##### PART C

- \*Access the PoseView image of possible binding interactions between the ligand and the binding site.
- \*Identify amino acids that are thought to interact with the ligands.
- \*Identify groups and substituents on the ligand that are proposed to form binding interactions.

##### PART D

Extract GS4071 from the crystal structure and compare the active conformation with the energy-minimised conformation generated in Part A.

##### PART E

\*Create a model binding site by extracting GS4071 and the closest amino acids in the binding site.

\*Ensure that GS40721 is displayed as a ball and stick model, and the amino acid residues are displayed as sticks.

\*Add residue labels and identify key amino acid residues.

#### **PART F**

\*Create a model binding site that includes hydrogen atoms on the ligand.

\*Ensure that GS40721 is displayed as a ball and stick model, and the amino acid residues are displayed as sticks.

\*Show polar hydrogens and intermolecular hydrogen bonds.

\*Remove any hydrogen atoms on the ligand which should not be present.

#### **PART G**

\*Identify and measure any hydrogen bonding interactions that are possible between the carboxylate group of GS4071 and the binding site.

\*Identify and measure any hydrogen bonding interactions that are possible between the aminium group of GS4071 and the binding site.

\*Identify and measure any hydrogen bonding interactions that are possible between the amide group of GS4071 and the binding site.

## PROCEDURES

There are various approaches that you can use to tackle these molecular modelling exercises. The following procedures illustrate how you might tackle this particular exercise, but they are not meant to be prescriptive. Note also that the results obtained may vary depending on the computer and the version of ChemBio3D/Chem3D used. For example, the specific conformations obtained from energy minimisation may differ, as may quantitative results such as steric energies. The ChemDraw file for GS4071 can be accessed from the ChemDraw folder.

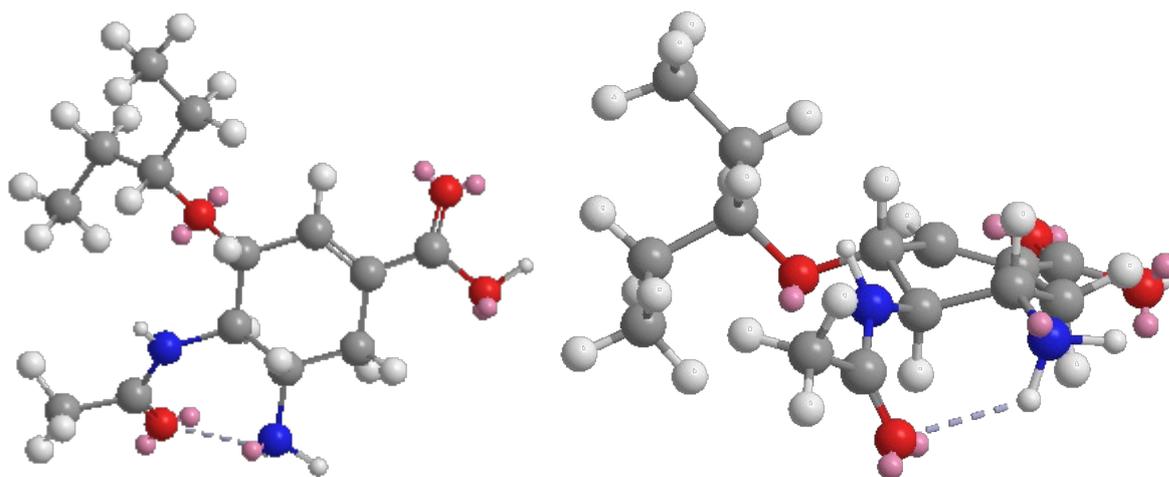
### PART A Conformations of GS4071

#### 1. Create the energy-minimised structure of GS4071 from its ChemDraw file.

\*Open **ChemBio3D** or **Chem3D**.

\*From the **File** menu, choose **Open**, then select the ChemDraw file for GS 4071. Click on **Open** to produce the 3D model.

\*Energy minimise the structure  (Fig. 2). The steric energy is given in the bottom window as 9.5 kcal/mol for the energy-minimised conformation. All three substituents are pseudo equatorial and the ring is a twist boat-like structure because of the double bond. There is an intramolecular hydrogen bond shown by the dotted line. If this is not visible go into the **View** menu to select **Model Display**. Click on **Show Hydrogen Bonds**, then select **All**.



**Figure 2** Energy-minimised structure of GS4071 from different viewpoints.

#### 2. Identify the number of rotatable bonds.

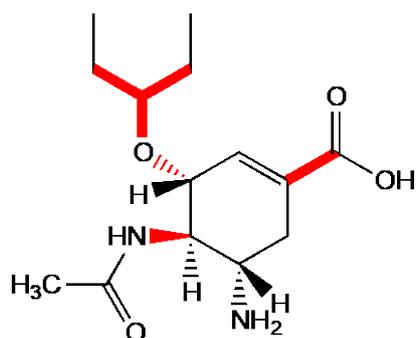
\*From the **Calculations** menu choose **Compute Properties**.

\***Expand Molecular Topology** and select **Num Rotatable Bonds**.

\*Click **OK**. The number of rotatable bonds appears in the bottom window as 7.

Note that rotatable bonds in this context are defined as those that result in distinct differences in conformation. Rotatable bonds that only alter the relative positions of hydrogen atoms (e.g. C-NH<sub>2</sub>, C-CH<sub>3</sub> etc) are not included in the total.

Six rotatable bonds in the structure are highlighted in red below (Fig. 3). The program also identified the amide bond as a rotatable bond, but this would not usually be counted as such.



**Figure 3** Rotatable bonds in GS4071 that would result in significant conformational changes.

**3. Carry out a molecular dynamics program to see if more stable conformations can be identified.**

\*From the **Calculations** menu, choose **MM2**, then **Molecular Dynamics** to open a table.

\*Keep the default values shown in the table and *click* **Run**. You will see the molecule vibrating and adopting various conformations. If you wish to curtail the molecular dynamics process at

any time, *click* on **Stop Calculation** on the menu bar . You should find that the molecule tends to stay very much as it is. Energy minimisation of the final structure results in it reverting back to the original conformation.

## PART B The crystal structure of GS4071 bound to neuraminidase (NA)

### 1. Download the crystal structure of GS4071 bound to the NA enzyme.

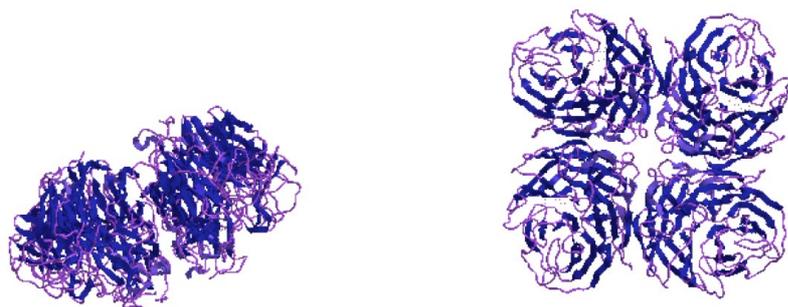
You will need to be connected to the internet in order to do this.

\*From the **Online** menu, choose **Find Structure from PDB ID** to open a dialogue box.

\*Enter the PDB code (**2HU4**).

\*Click on **Get File** and the protein will appear as a ribbon diagram, with ligands present in a ball and stick format.

Two tetrameric protein complexes are obtained (Fig. 5). The proteins are represented as ribbon structures with light purple sections corresponding to  $\alpha$ -helices, and dark blue sections corresponding to  $\beta$  sheets. The thin pink strands are regions that do not have a secondary structure.



**Figure 5** Crystal structure for 2hu4 obtained from the pdb website.

### 2. Recolour the helices and coils to make the distinction clearer between different secondary structures.

\*From the **File** menu, choose **Model Settings**.

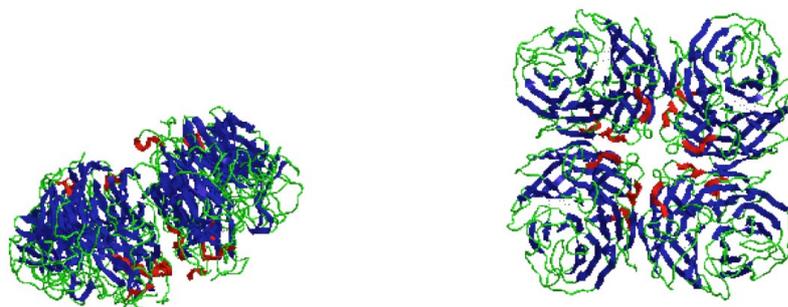
\*Click on the tab for **Colors and Fonts**.

\*Under the section on **Model Colors**, modify the colours used for the **Alpha helix**, **Beta sheet** and **Coil**.

Click on **Apply** to see the effect in the main window.

\*If satisfied, click **OK**.

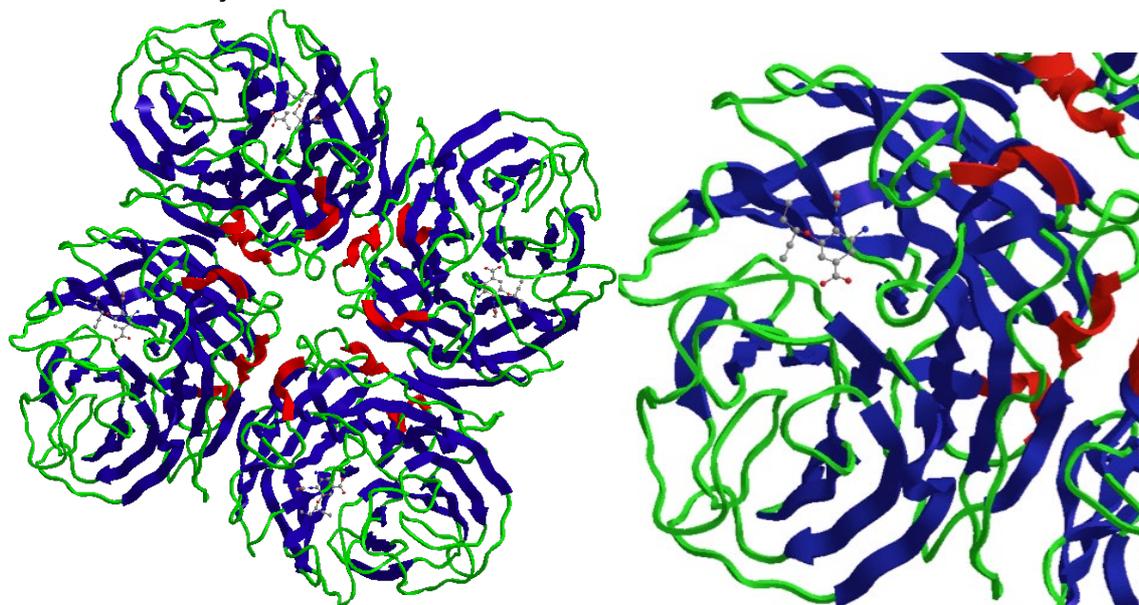
Figure 6 shows the helices in red, the sheets in blue and the coils in green.



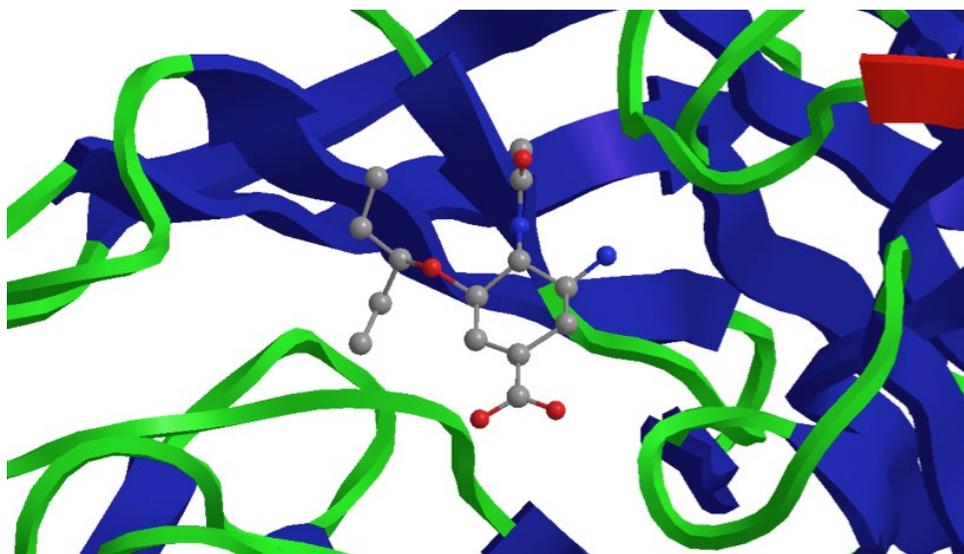
**Figure 6** Recoloured crystal structure for pdb 2hu4.

### 3. Identify the ligand in the crystal structure.

\*The mouse scroll allows you to *zoom* in on a specific tetrameric structure and a specific protein subunit (Fig.7). The GS4071 ligand is represented by a ball and stick model and can be seen if you enlarge the model further to see the binding site (Fig. 8).



**Figure 7** Enlargements A) of one of the tetramers and B) of one of the protein units with the ligand visible.



**Figure 8** Enlargement to show the ligand GS4071 in the binding site.

#### 4. Explore the crystal structure using the Model Explorer table

\*Click on the **Model Explorer** tab on the left margin of the screen to open the Model Explorer table (Fig. 9A). If the tab is not visible, go into the **View** menu and *choose Model Explorer*. If the table is still not visible, you may have to enlarge the window containing it by dragging the margins. The window should be to the left of the main window. The Table shows the presence of 8 Chains (Chains A-H).

\*Click on the various chains identified. The relevant chain will be highlighted in yellow in the main window. There are 8 chains in total, matching the number of proteins present in the two tetrameric complexes.

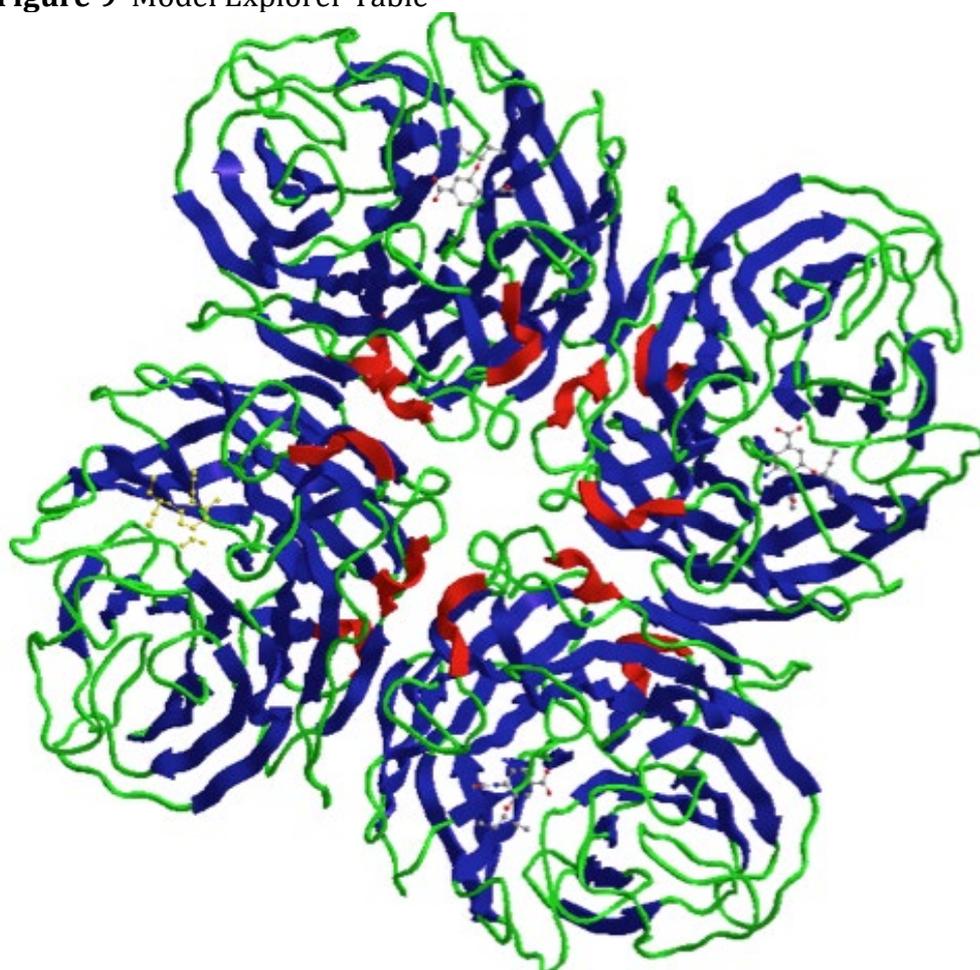
\*Click on the small box with the + sign to the left of each chain. This expands the entry and you will find that each of the 8 proteins has a ligand (Fig. 9B).

\*Click on the ligand label. The ligand will be highlighted in yellow in the main window (Fig. 10). This gives you a good indication of the relative size of the ligand to the protein. Figure 10

shows one of the tetrameric complexes with the ligand highlighted in yellow in the bottom left protein.

Model Explorer	
Model Explorer	Chain-A
A-nisr	F Fragment-83-468
B-nisr	F Ligand-84
C-nisr	Chain-B
D-nisr	Chain-C
E-nisr	Chain-D
F-nisr	Chain-E
G-nisr	Chain-F
H-nisr	Chain-G
H-nisr	Chain-H
Backbone	Backbone

**Figure 9** Model Explorer Table



**Figure 10** A tetrameric complex with one ligand highlighted in yellow within the bottom left protein.

### PART C The PoseView image of possible binding interactions

The PoseView image provided in the Protein Data Bank automatically generates possible binding interactions between the ligand and the binding site. This is a good starting point when investigating possible binding interactions.

#### 1. Find the PoseView image of proposed binding interactions.

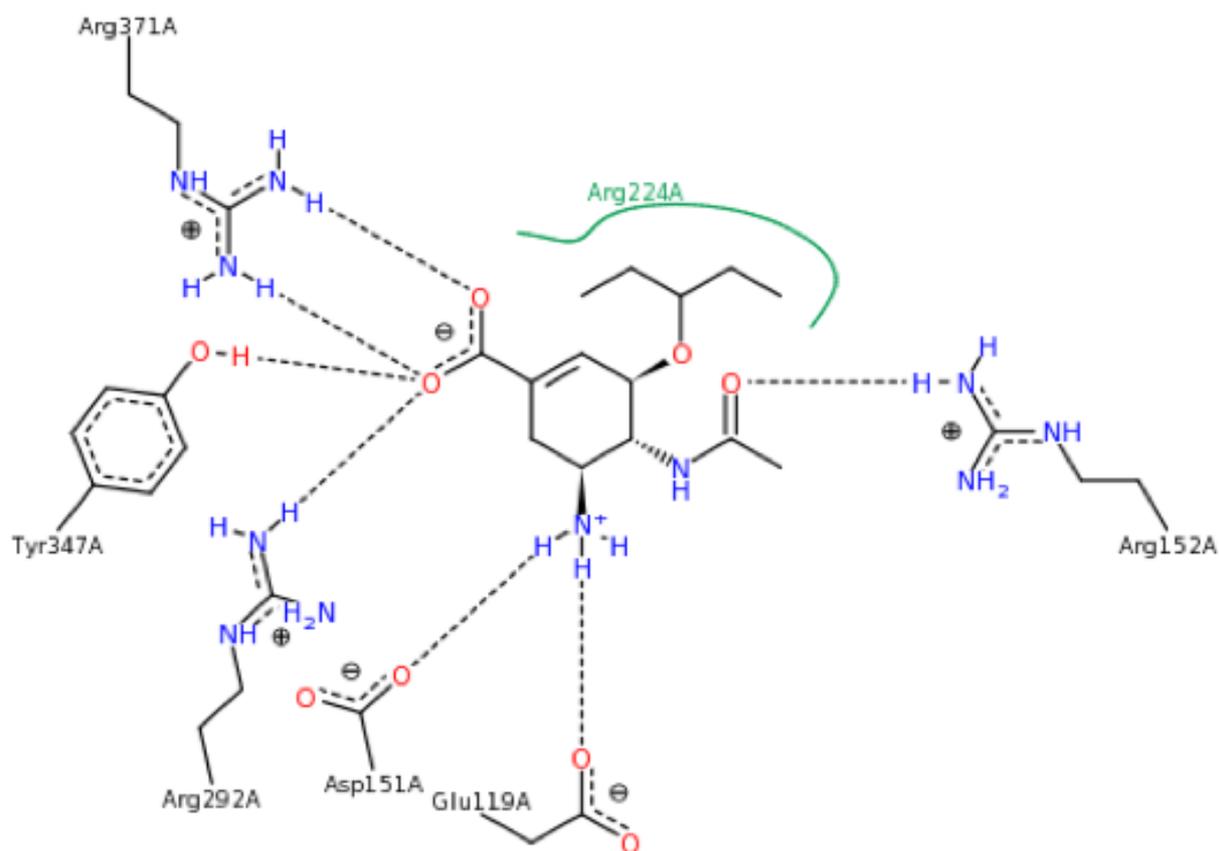
\*Access the home page for the protein data bank at [www.rcsb.org](http://www.rcsb.org).

\*In the search box provided on the home page, enter the PDB code (2HU4), then click on Go. You will now have the front page of that file.

\*Scroll down the page to a section headed **Small Molecules**, which refers to any ligands bound to the protein.

\*Under the section on ligands, there are two diagrams shown under **2D Diagram and Interactions**. The left-hand structure shows the structure of the ligand. The right-hand diagram shows intermolecular binding interactions between the ligand and key amino acids within the binding site.

\*Click on the diagram of the binding interactions to get a window showing an expanded view of the Poseview Image (Fig. 11). Hydrogen bonds are shown by dashed lines. Hydrophobic pockets are shown by green lines. The amino acids lining these pockets are also identified in green.



**Figure 11** Poseview image of likely binding interactions.

#### 2. Note the amino acid residues that form hydrogen bonds with GS4071.

The amino acids concerned are Arg-371, Tyr-347, Arg-292, Asp-151, Glu-119 and Arg-152.

**3. Note the functional groups in GS4071 that form hydrogen bonds with the binding site.**

The functional groups involved in hydrogen bonds are the carboxylate group, the aminium group, and the amide group.

## PART D The Active Conformation

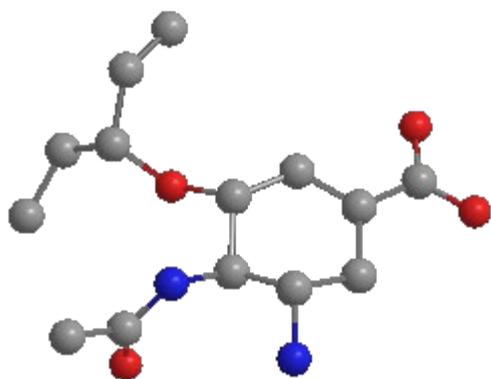
### 1. Extract GS4071 from the active site of neuraminidase.

\*Use the **zoom** tool  (or *scroll* the mouse) to zoom into one of the ligands.

\*Choose the **select** tool , then *double click* on any part of the ligand to select the whole molecule.

\*From the **Edit** file, choose **Copy**.

\**Open* a new Chem3D window and *paste* the structure into the window. A ball and sphere model for GS4071 will now appear in the new window (Fig. 12). Note that hydrogen atoms are absent from the copied structure. Do not energy minimise the structure as this will alter the conformation. Note also that it is not valid to measure the steric energy of the structure. There is an experimental error involved in measuring the position of the atoms by X-ray crystallography. Therefore, bond lengths, bond angles, and torsion angles will be less than optimal because of experimental error. Despite this, it is possible to observe the relative orientation of rings and substituents in the active conformation in a qualitative manner.



**Figure 12** Active conformation for GS4071.

### 2. Compare the active conformation with the energy-minimised conformation created in Part A.

\*From the **File** menu, choose **New** to open a new window.

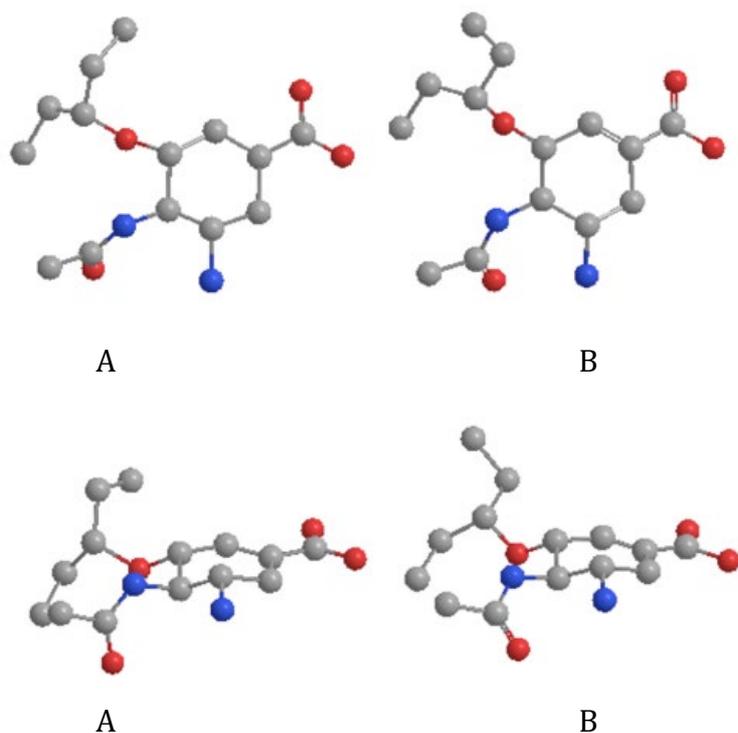
\**Copy* and *paste* the active conformation into the new window.

\**Copy* and *paste* the energy-minimised conformation generated in Part A into the same window.

\*From the **View** menu, choose **Model Display**, then **Show Hydrogen Atoms**. *Select Hide*.

\*From the **View** menu, choose **Model Display**, then **Show Lone Pairs**. *Select Hide*.

\*Position the two conformations side by side in similar orientations (Fig. 13). One structure can be manipulated selectively by *double clicking* on it with the **Select** tool, then carrying out the manipulation with the Shift key depressed.



**Figure 13** Comparison of A) the active conformation, and B) the initial energy-minimised conformation from two different perspectives.

The conformations are very similar - as one would expect with such a rigid structure. However, there are clear differences in the orientation of the carbon atoms in the ether substituent, as well as a more subtle difference in the orientation of the amide substituent.

## **PART E Creation of a model binding site.**

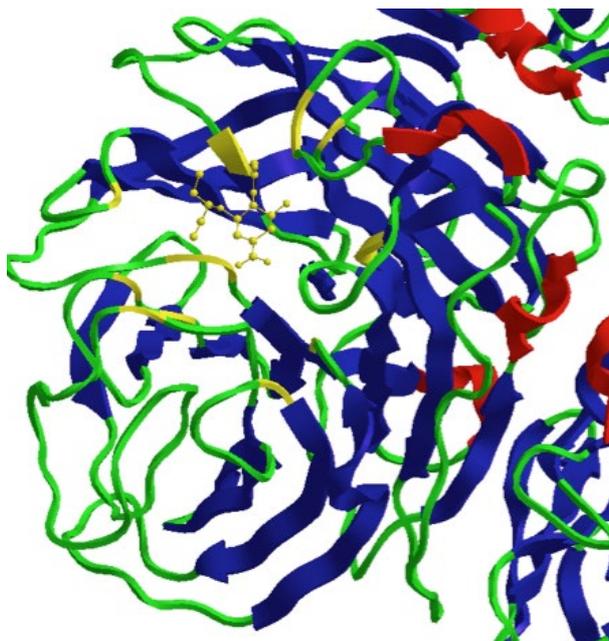
### **1. Select the ligand and amino acid residues within 4 Å.**

\*Zoom in on the ligand in the binding site.

\*Double click on any of the atoms to select it.

\*Position the mouse over one of the selected atoms, then *right click* the mouse to open a menu.

\*Choose **Select**, then choose **Select groups within Distance of Selection**. Choose **4 Angstrom** (Fig. 13).



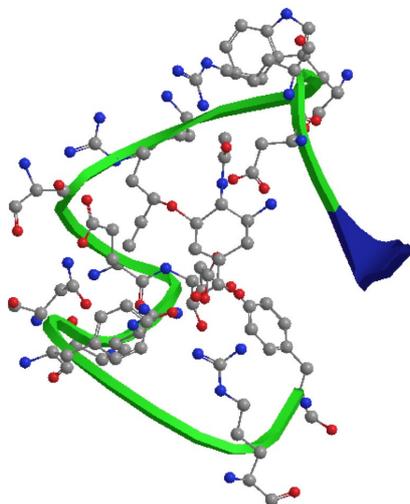
**Figure 13** Ligand highlighted with closest neighbouring amino acid residues.

### **2. Copy the selection into a new window.**

\*From the **Edit** menu, choose **Copy**.

\*From the **File** menu, choose **New** to open a new window.

\*From the **Edit** menu, choose **Paste**. The selected ligand and amino acid residues now appear in the new window (Fig. 14).



**Figure 14** Initial copy of ligand and surrounding amino acids.

**3. Modify the structure to distinguish between the ligand and the amino acid residues.**

The pasted model is a bit confusing and needs to be tidied up such that a distinction can be made between the ligand and the surrounding amino acid residues.

\*From the **View** menu choose **Model Display**. Choose **Display Mode**, then **Sticks**.

\*Go into the **Model Explorer** Menu and *expand Chain C*, then select **Ligand-86**. (This will vary depending on the protein subunit chosen).

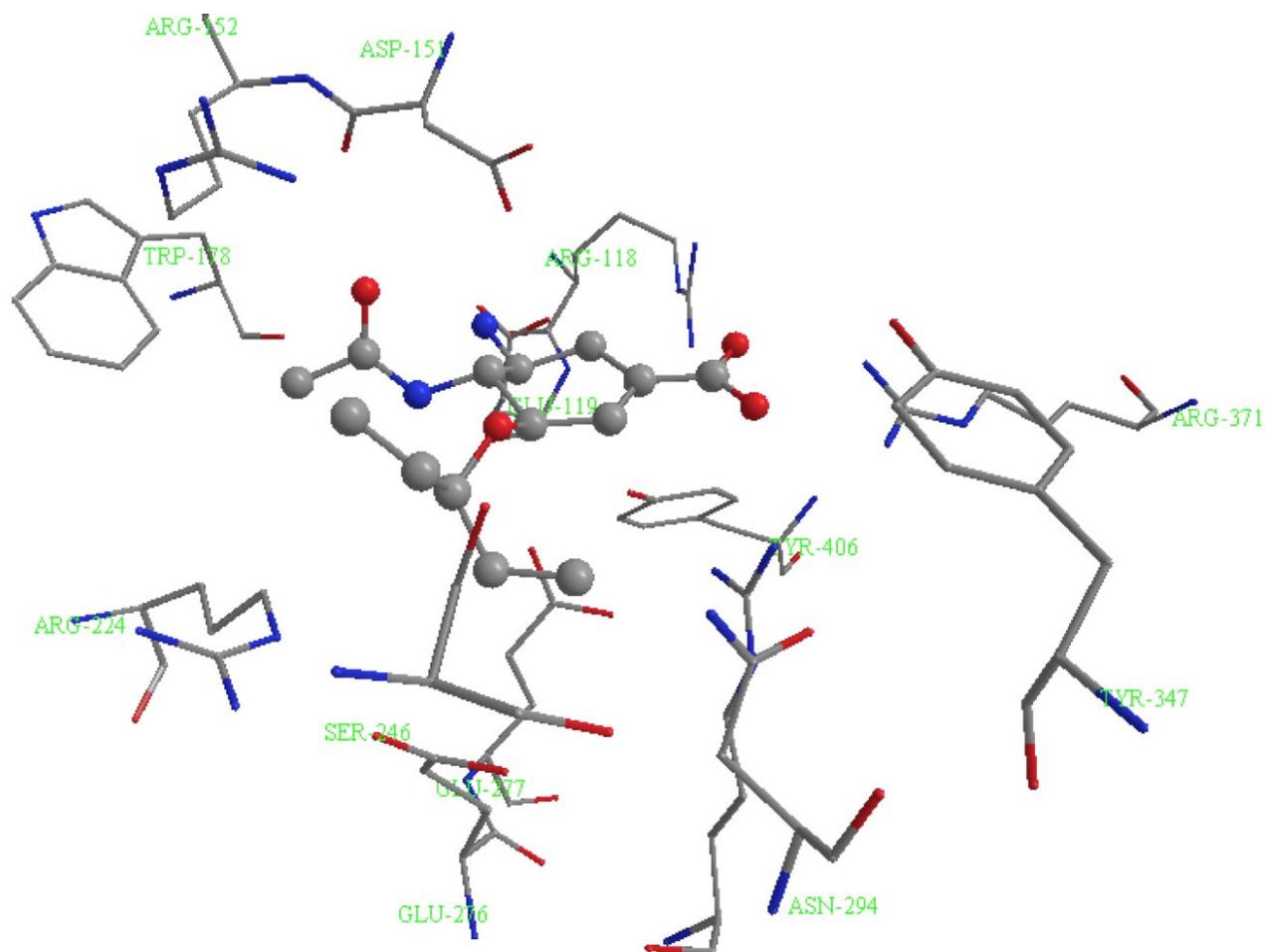
\*Position the mouse cursor over the selection, then *right click* the mouse to open a menu.

\*Choose **Display Mode**, then select **Ball and Stick**.

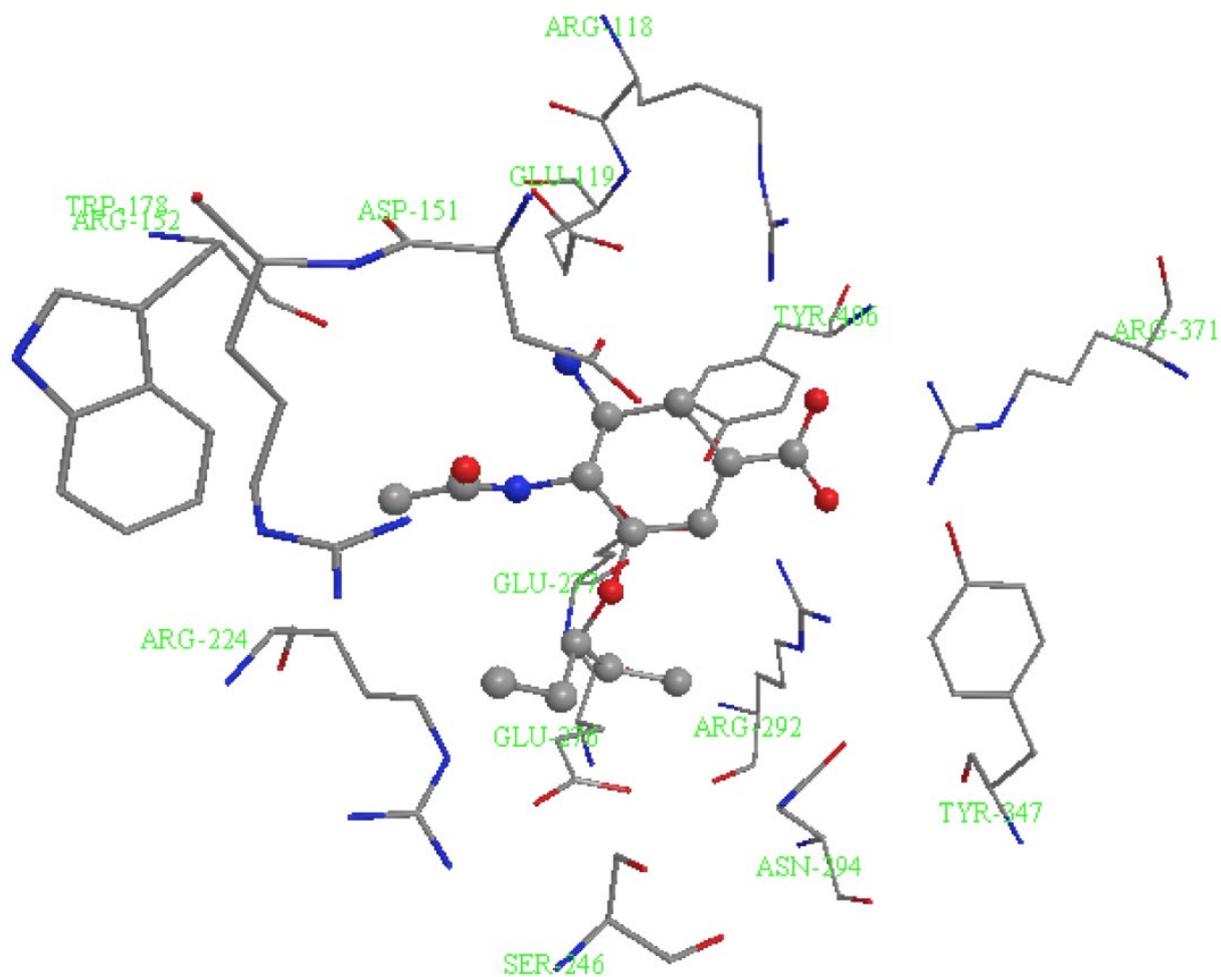
\*From the **View** menu, choose **Model Display**, then select **Show Residue Labels**.

\*From the **View** menu, choose **Display Mode**, then **Perspective**.

The resulting model binding site is shown in figures 15 and 16 from different viewpoints.



**Figure 15** Closest amino acid residues in the binding site (compare with figures 20 and 21).



**Figure 16** Identification of amino acid residues (Arg-292, Tyr-347, Arg-371, Glu-119 and Arg-118) interacting with the carboxylate group (compare Fig 22).

## **PART F Creation of a model binding site that includes hydrogen bonds.**

The model binding site in part E identifies the amino acid residues present, but not the hydrogen bonds. We will now create a binding site that includes hydrogen bonds. However, it has to be appreciated that the addition of hydrogen atoms to the ligand is not entirely accurate. The hybridization states of the ligand atoms are not defined properly in the copied structure, and so the program assumes that every atom is  $sp^3$  hybridised. This results in errors with respect to the number and orientations of hydrogen atoms added to  $sp$  and  $sp^2$  hybridised atoms. It is possible to work round this but you should be aware of the limitations involved.

### **1. Select the ligand and neighbouring amino acid residues.**

\*Return to the window that contains the full crystal structure. The selection of the ligand and surrounding amino acid residues should still be shown.

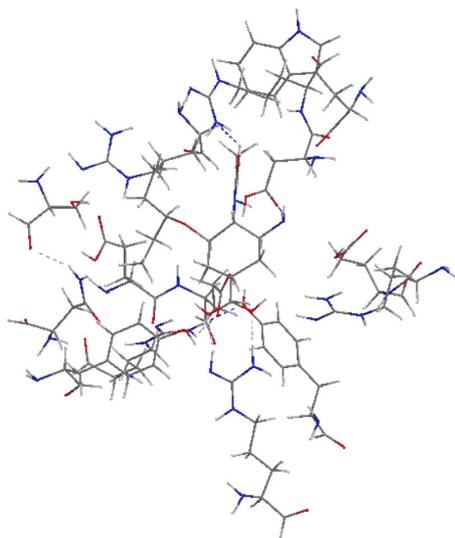
\*Choose the **rotate** tool  or the **translate** tool , then *click* in the main window to activate it. Do not use the **select** tool , or you will deselect the ligand and surrounding amino acid residues.

### **2. Copy the selection as a ChemDraw Structure and paste it into a new window.**

\*From the **Edit** menu choose **Copy As**. Choose **ChemDraw structure**.

\*From the **File** menu, choose **New** to open a new window.

\*From the **Edit** menu, choose **Paste** to paste the selection into the new window (Fig. 17).



**Figure 17** The initial pasted model of the selection.

### **3. Modify the model to distinguish between the amino acid residues and the ligand.**

The following instructions will produce a model where the amino acid residues are in a stick format and the ligand is in a ball and stick format. Only those hydrogen atoms likely to be involved in hydrogen bond interactions will be visible and only intermolecular hydrogen bonds will be displayed.

\*From the **View** menu choose **Model Display**, then **Display Mode** Choose **Sticks**.

\*Select the ligand in the **Model Explorer** Table (**Fragment 1**).

\*Position the mouse cursor over the selection, and *right click* the mouse to open a menu.

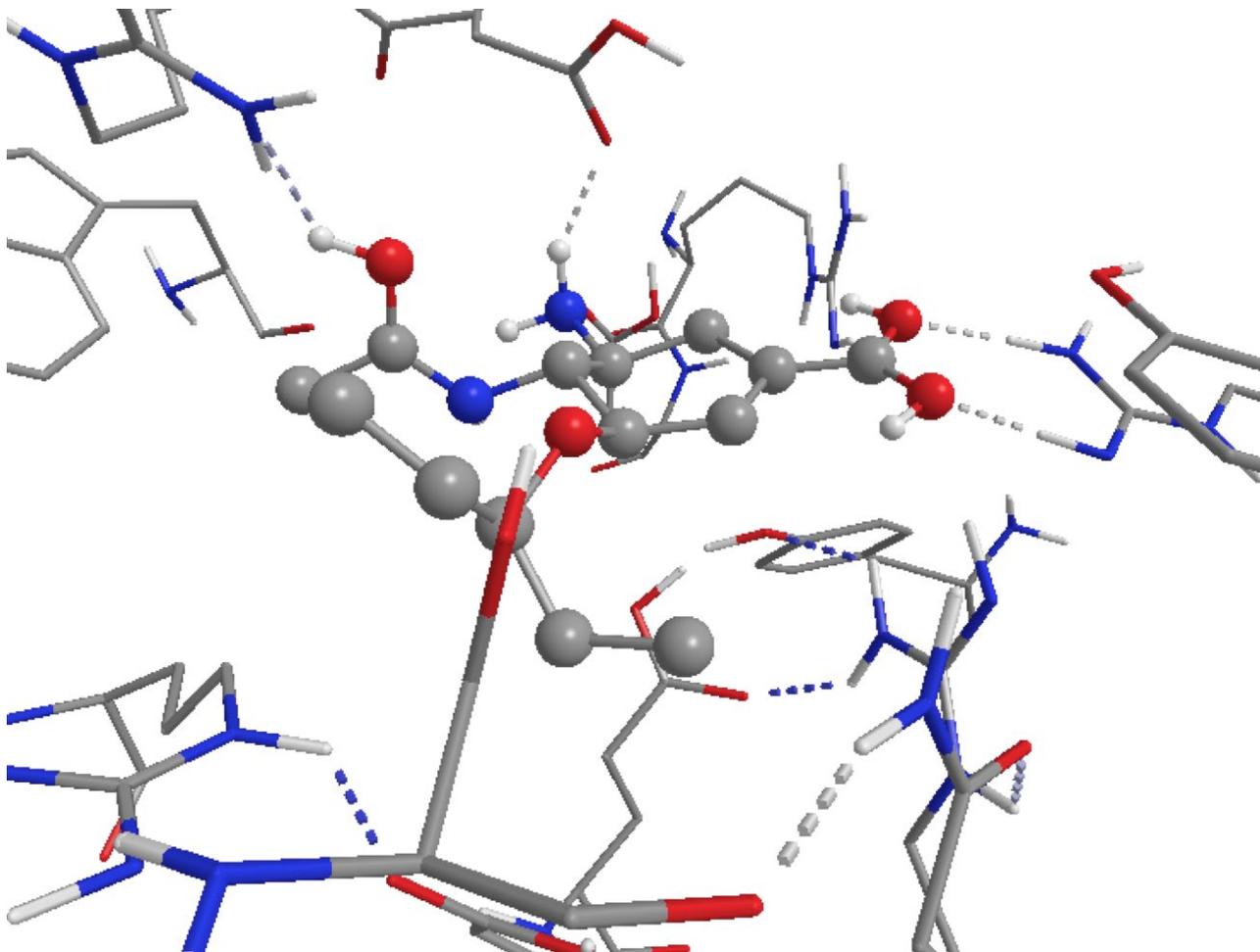
\*From the menu that appears, select **Display Mode**, then choose **Ball and stick**.

\*From the **View** menu, *select Model Display*, then **Show Hydrogen Atoms**. *Choose Show Polar*.

\*From the **View** menu, *select Model Display*, then **Show Hydrogen Bonds**. *Choose Intermolecular*.

\*From the **View** menu, *select Model Display*. *Choose Perspective*.

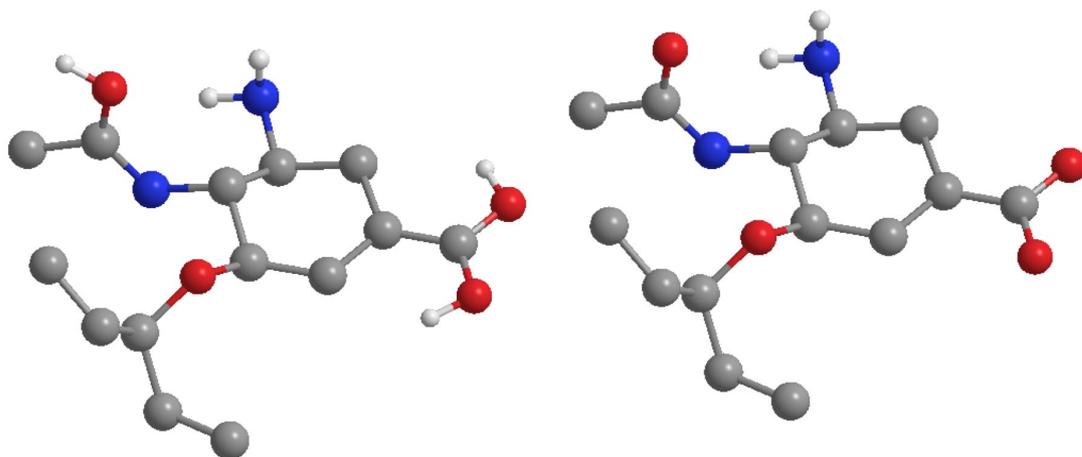
The resulting model is shown in figure 18.



**Figure 18** Ligand and closest amino acid residues with hydrogen bonds shown. Note that the number of hydrogen atoms on some ligand atoms is inaccurate.

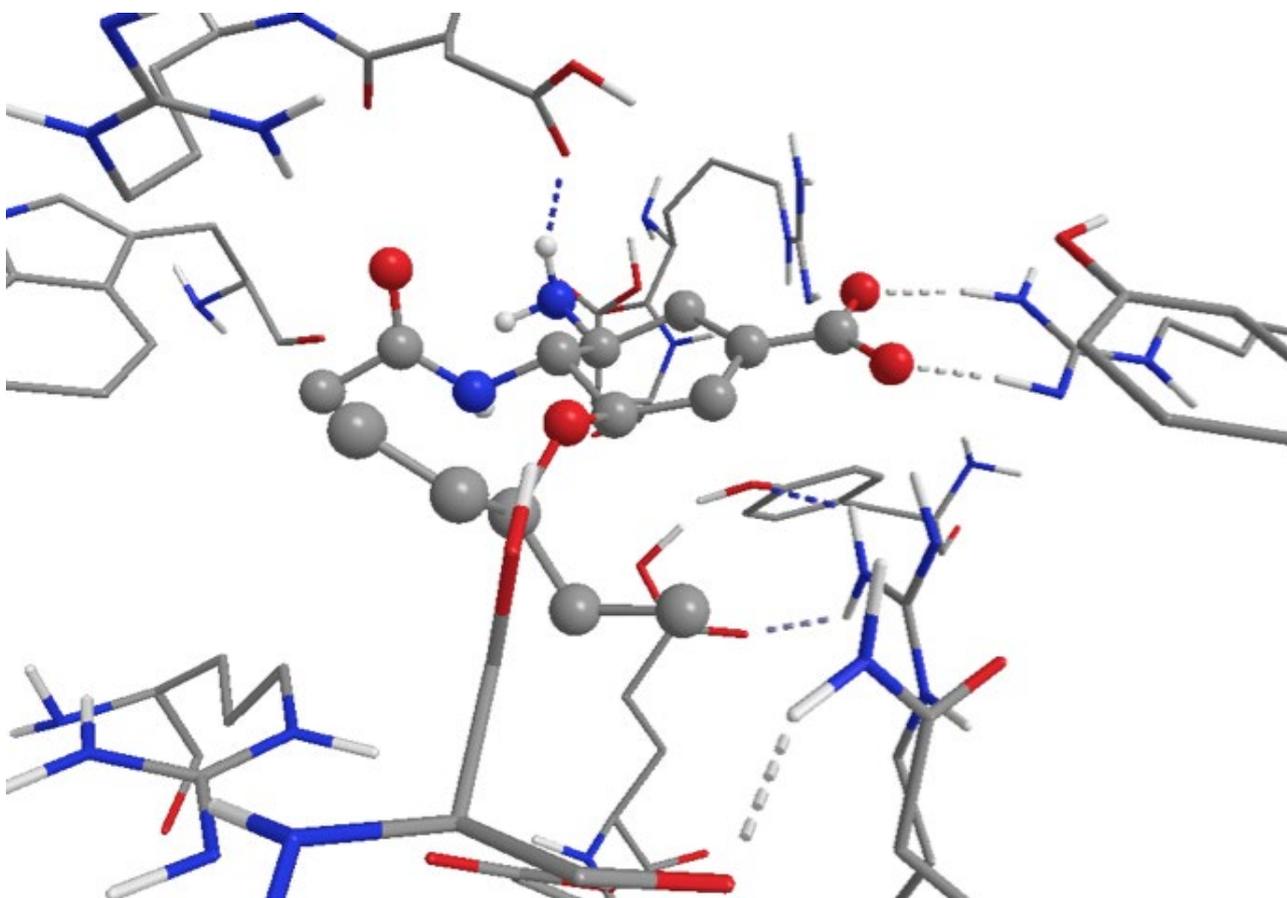
#### 4. Correct the structure of the ligand by removing erroneous hydrogen atoms

Hydrogen atoms are shown correctly on the amino acid residues. However, the addition of hydrogen atoms to the ligand is flawed since the software assumes that every atom in the ligand is  $sp^3$  hybridised. For example, a hydrogen atom is shown to be present on a carbonyl oxygen when clearly it should not (Fig. 19A). Such atoms can be hidden manually by *selecting* the offending hydrogen atom then *right clicking* the mouse as you position it over the atom. From the menu that appears, *choose Visibility*, then **Hide Atom**. In this model, you need to hide the hydrogen atoms that are on the oxygen atoms of the carboxylate group, and the carbonyl oxygen of the amide substituent to produce the corrected structure (Fig. 19B).



**Figure 19** A) Ligand with erroneous hydrogen atoms. B) Corrected structure.

When these hydrogen atoms have been hidden, the model binding site should now look like figure 20.



**Figure 20** Corrected structure to remove extraneous hydrogen atoms (see also Figs 15 & 16).

## **PART G Identification of hydrogen bonds between the ligand and the binding site**

It is not possible to label the amino acid residues in the model binding site created in Part F. Therefore, the model binding sites created in Parts E and F should be used together in order to identify hydrogen bonds and the amino acid residues involved.

### **1. Identify hydrogen bonds that are highlighted by ChemBio3D/Chem3D in the model binding site**

In figure 20, three hydrogen bonds are indicated between the ligand and surrounding amino acids. A bidentate interaction between the carboxylate group and the side chain of an arginine residue (Arg-371) is visible here. There is also a H-bond shown between the primary amine and the side chain of an aspartate residue (Asp-151). The identity of these amino acids can be determined from figures 15 & 16. Not all the possible hydrogen bonds have been picked up by the ChemBio3D/Chem3D program and we shall look more closely at these below. However, figure 20 emphasises the strong ionic and H-bonding interaction between the carboxylate ion and the arginine residue (Arg-371).

### **2. Measure H-bond lengths.**

The lengths of the various H-bonds can be measured as follows

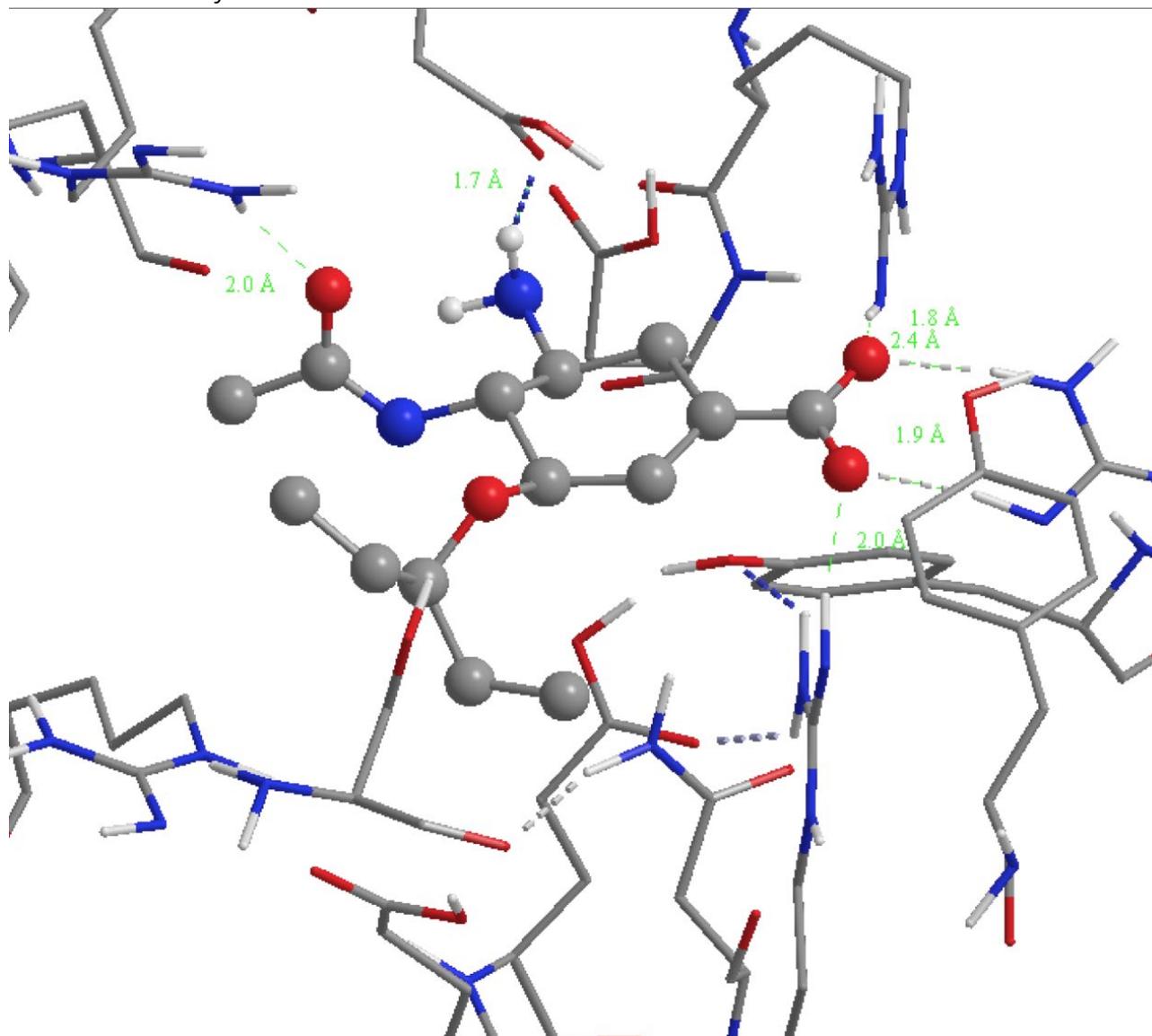
\*Choose the **Select** tool .

\*With the shift key depressed, *click* on the HBD and HBA involved in a hydrogen bond interaction.

\*From the **Structure** menu, *choose* **Display Distance Measurement**.

The distances are 1.8 Å and 1.9 Å for the carboxylate bidentate interaction with arginine-371, and 1.7 Å for the NH<sub>2</sub> interaction with the aspartate residue (Asp-151) (Fig. 21). There should also be an interaction with the side chain of a glutamate residue (Glu-119) visible behind the structure, but this is not shown because the group is represented as NH<sub>2</sub> rather than the aminium ion <sup>+</sup>NH<sub>3</sub>. In other words, the third hydrogen atom is not present to show that interaction.

There is another hydrogen bond measuring 2.0 Å between the primary amide carbonyl oxygen and the side chain of another arginine residue (Arg-152) (Fig. 21). There are also two possible hydrogen bond interactions involving the carboxylate group and the side chains of two further arginine residues above and below the carboxylate ion (Arg-118 and Arg-292). These would be 2.4 Å and 2.0 Å (Fig. 21).

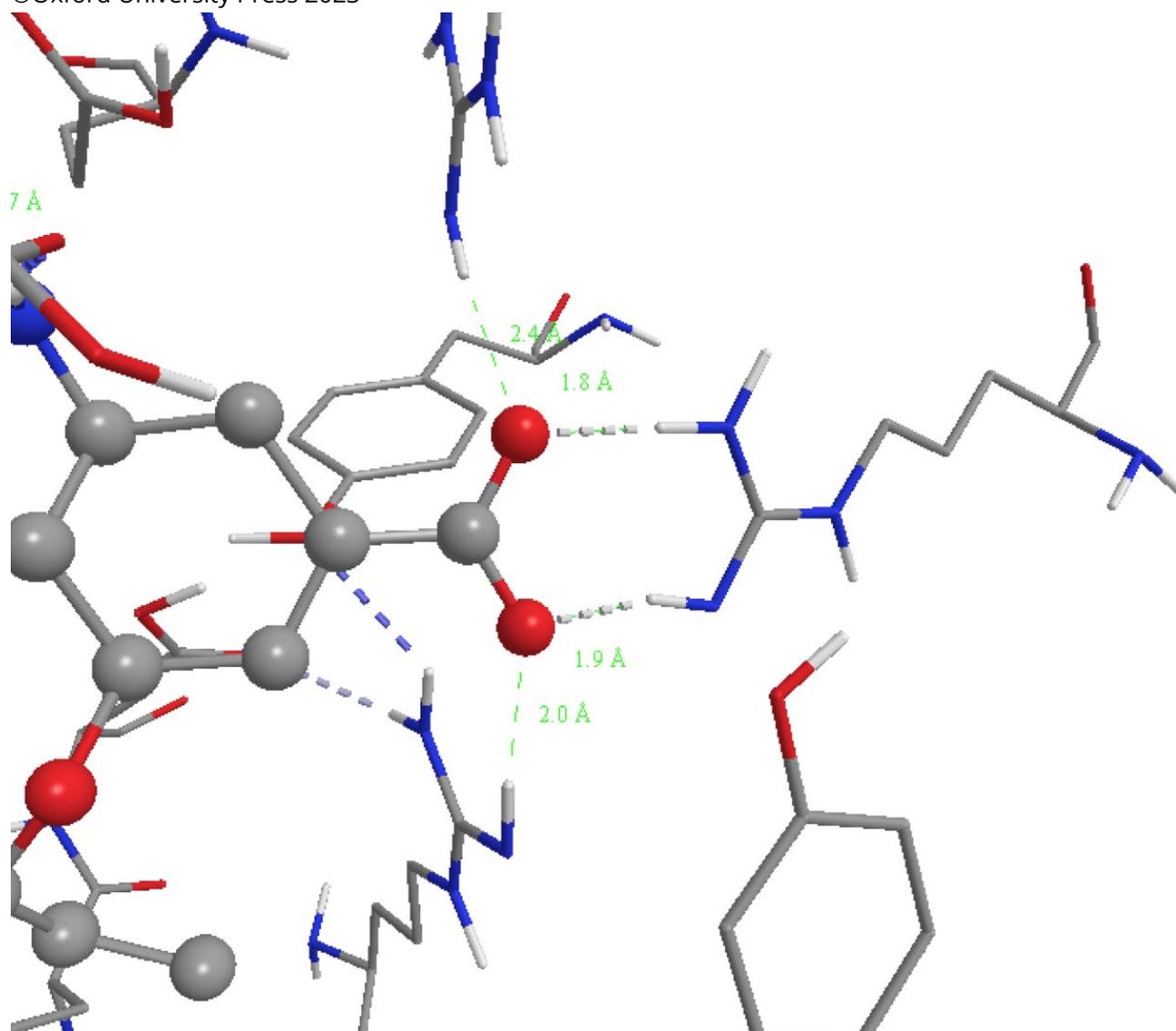


**Figure 21** Six hydrogen bond interactions (see also Fig 15).

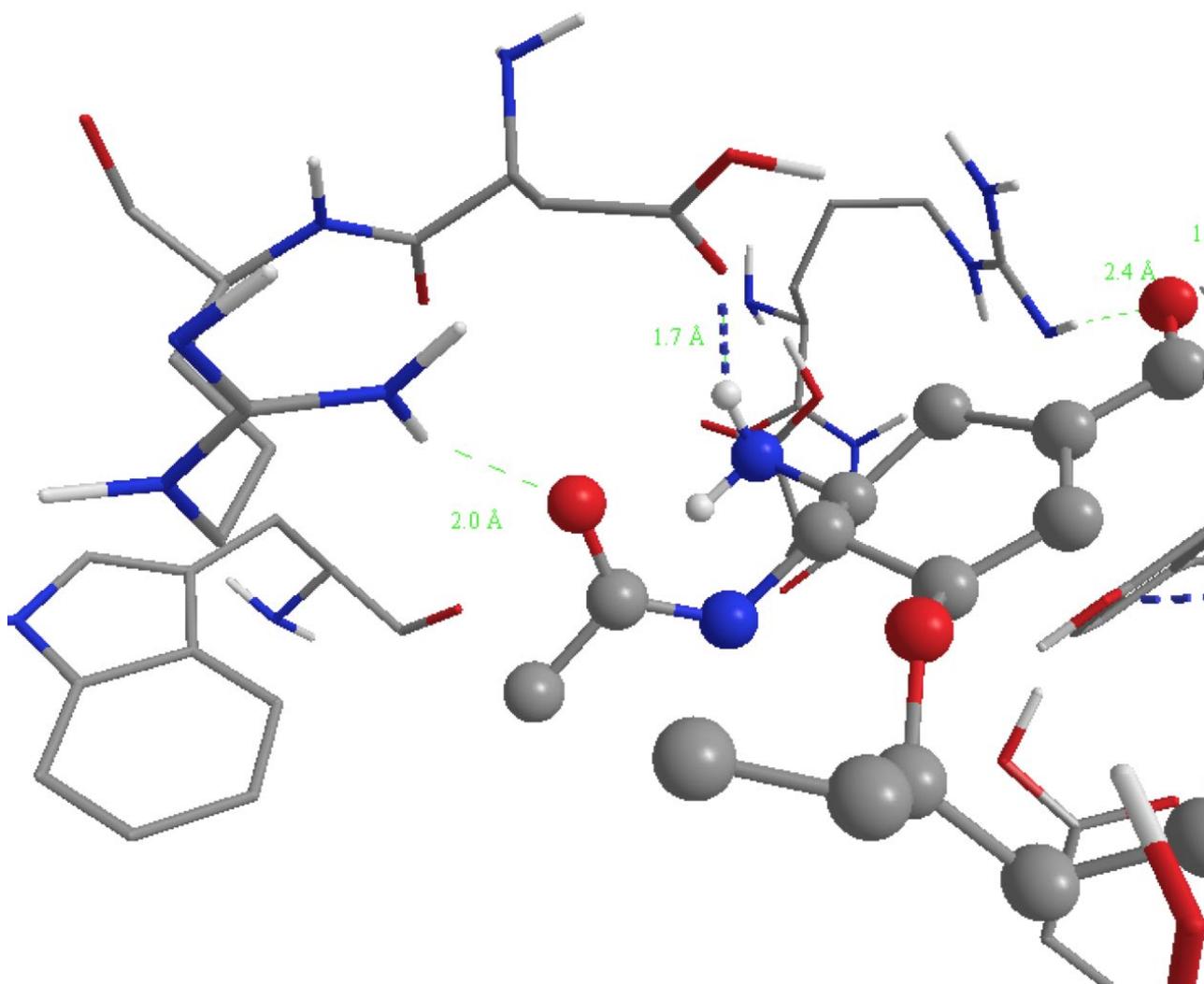
\*Zoom in on the relevant regions.

\*From the **View** menu, choose **Model Display**, then **Perspective**

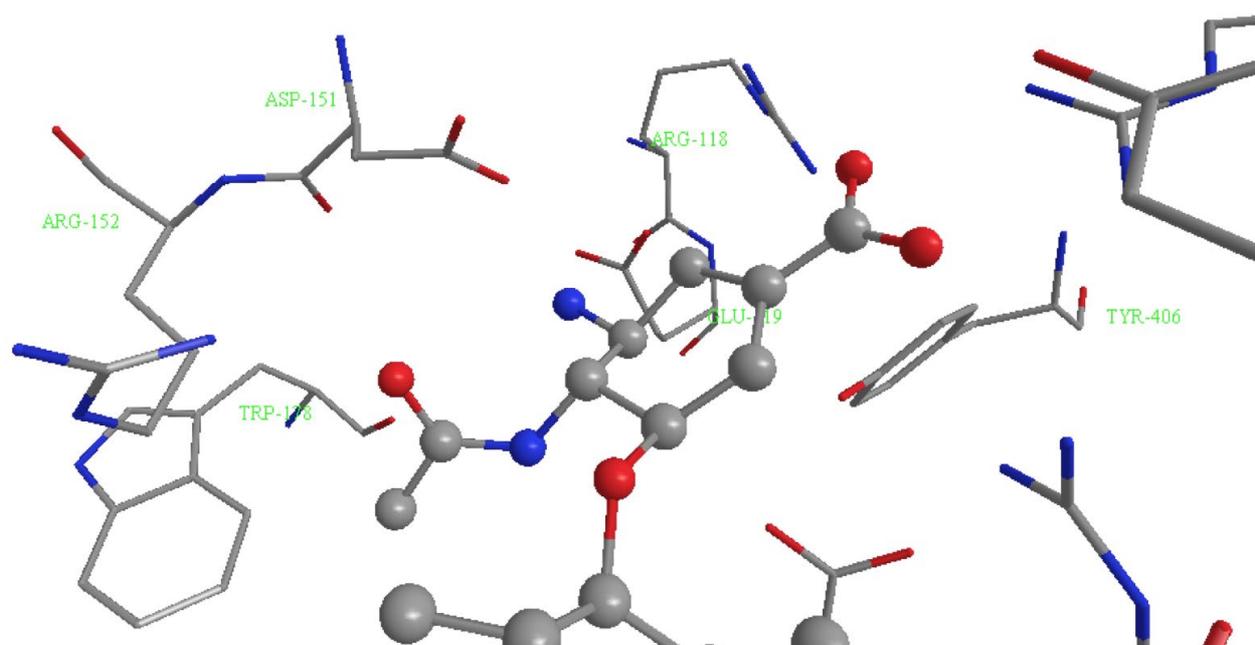
By zooming in and using Perspective, you should be able to reproduce the following images (Figs. 22-23). Note that a tyrosine residue (Tyr-347) is visible on the bottom right. Rotation of the phenolic group such that the hydrogen points towards the carboxylate group would permit another hydrogen bond interaction (as indicated in the PoseView image).



**Figure 22** The carboxylate interactions with three arginine residues (see also Fig 16).



**Figure 23** The hydrogen bond interactions involving the primary amine with Asp-151 and the amide with Arg-152 (see also Fig 24).



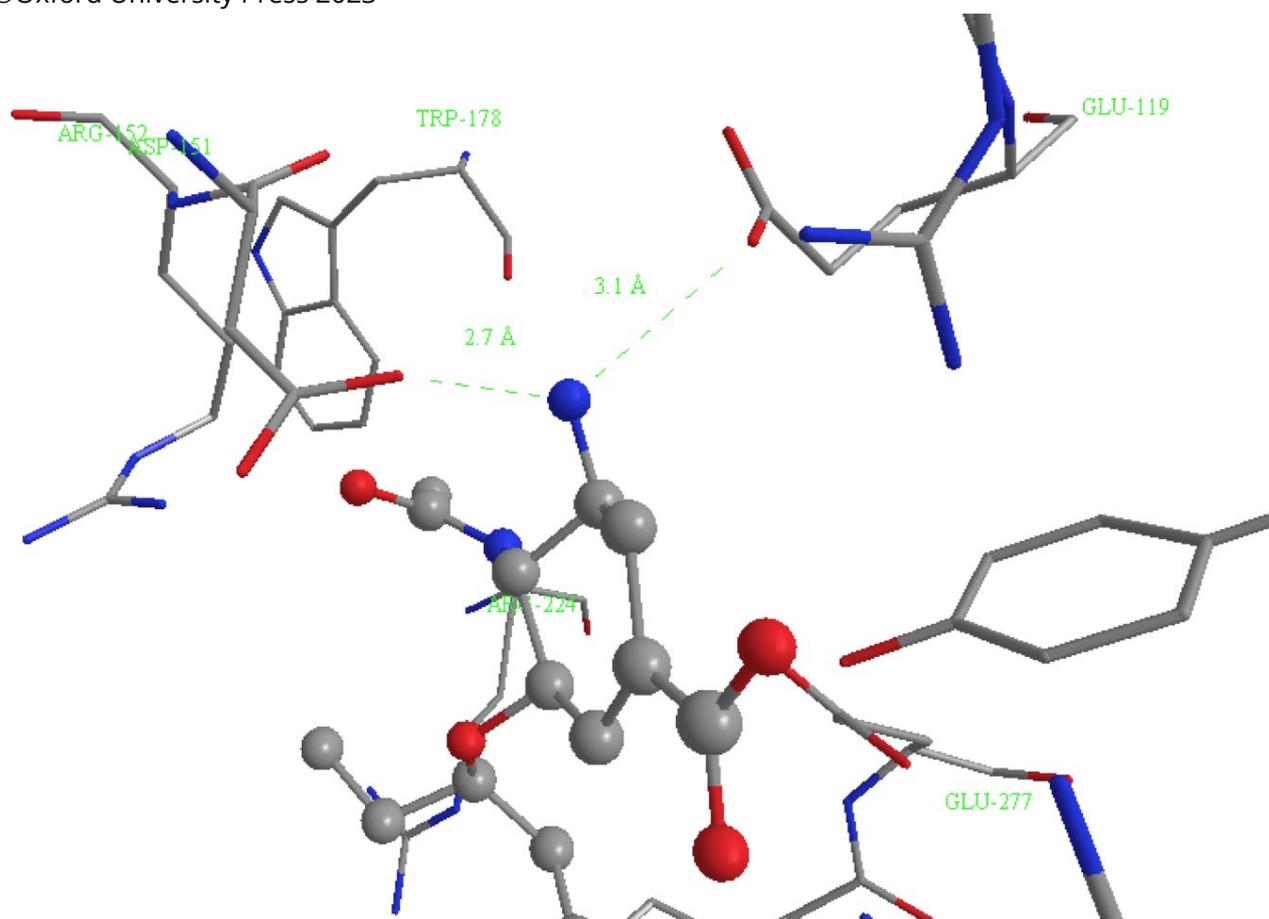
**Figure 24** Identification of amino acid residues (Asp-151 & Arg-152) in the vicinity of the primary amine and amide groups (compare fig 23).

**3. Identify the amino acids interacting with the carboxylate group of the ligand.**

The key arginine residue interacting with the carboxylate group is Arg-371 (Fig. 15). Other amino acid residues that can interact with the carboxylate group are Arg-292, Arg-118 and Tyr-347.

**4. Identify any interactions that the ligand might have with Asp-151 and Glu-119.**

These amino acids are interacting with the primary amine group (Fig. 25).



**Figure 25** Interactions involving Asp-151 and Glu-119.

