Protein Structure and Design

(6BBB0333)

Practical Worksheet #2 – Computer Modelling

Name: Jude Popham

Student Number: K20041606

NOTE: When you have completed all questions,
convert this into a single PDF document before submission.
Please only submit PDF files!

Molecular Graphics Practical

Introduction

The molecular graphics program you will be using in this practical is called YASARA (www.yasara.org, Krieger E, Koraimann G, Vriend G (2002) *Proteins* **47**, 393-402). It is a sophisticated graphics package that enables you to build, edit, model and simulate protein structures, and in this practical we will use some of the more basic features to study protein structure.

It is available in all the SCRs in New Hunt's House; the supervised practical help sessions will take place in one of these rooms (refer Course Module Book). **If Yasara is not installed on your computer, search for Applications Catalogue in Windows 10 (Software Centre if Windows 8); find Yasara and install it.** The data files that you will require are installed on their servers, and you can save and print the screen images. If you wish to save any of the files that you create for use at a later date, you can save these on the desktop area of the computer, or on your own USB memory stick. **To save a structure as a PDB file, you must replace the star/asterisk at the end of the character string that appears on the screen (……./*.PDB) by a filename (……./filename.PDB).** (We have a licence for the full version installed at King's; a free, downloadable version unfortunately does not permit energy calculation or simulations and will not enable you to complete the practical).

Log onto the computer. To start the program, either type YASARA at the Start screen, or it can be located by clicking on the down arrow (bottom left corner), and then scrolling across to the subsection YASARA (typically the last group).

It is recommended that you save your work as you progress through the practical so you can come back to it later if necessary.

1. Getting to know YASARA

The main display shows the structure(s) you have loaded, with information about the atom properties on the left, and information on the structures you currently have open on the right (scene content). A console window also appears at the bottom of the screen when you do calculations, and you can make this appear at any time by pressing **SPACE**. The **HELP** command is good, so use it if you get stuck – there is documentation and movies to help you out. You can also **UNDO** your last commands if you make a mistake – either click **EDIT**, **UNDO** or use the forward and backward arrows on the top left of the menu bar.

To open a molecule, from the menu bar select **FILE, LOAD, PDB** file. These files are in the YASARA/PDB directory which is where the program looks by default. To save a file select **FILE, SAVE, SAVE AS**, and you can then choose PDB or other formats, for example a YASARA scene, which will retain all the structures, atom styles and colours as they appear on your screen. To delete, you can either select **EDIT**, **DELETE**, **OBJECT** (you can select which object to delete), or **FILE**, **NEW** which erases everything.

Open 1crn.PDB – this is a small protein called crambin and you can use it to familiarise yourself with the mouse controls. When you open it, it is displayed as a CPK model with no hydrogens. You can change this under **VIEW**, **STYLE SCENE** – explore the options. **STICK** is a good general option, as you can see all the atoms and **RIBBON** is good for looking at secondary structures.

Here is a table (taken from the **HELP** documentation) of the mouse commands for moving the molecules around on the screen.

LeftClick in the 'Name' column of the HUD	Grab the selected object for movement/rotation (YASARA Model and above)
LeftButton + MouseMovement	Rotate object or scene around X/Y-axes or move marked label
Ctrl+LeftButton + MouseMovement	Rotate object or scene around X/Y-axes, centered on the marked atom

LeftButton + RightButton + MouseMovement	Move object or scene along the X/Y-axes
MiddleButton + MouseMovement	Move object or scene along the X/Y-axes
RightButton + MouseMovement	Move object, scene or label along the Z-axis
LeftClick on label	Mark label and move it around

Table 1: Mouse commands to move molecules on the screen in Yasara

Familiarise yourself with these movements. You can click on individual atoms and information appears under ATOM PROPERTIES on the left hand side of the screen. You can select more than one atom at once by holding **CTRL** when clicking, and information about angles and distances is displayed.

2. Construction of a free amino acid residue

You can build structures in YASARA using a chemical language called SMILES which is a standard language used by professional chemists. It is a simple method of entering a string of letters, which the program will turn into a structure. There are some instructions about the syntax of SMILES in Appendix 3.

To enter a SMILES string, go to **EDIT**, **BUILD**, **MOLECULE FROM SMILES STRING**. Below are the strings for L- and D-forms of alanine. Build both of these, and determine which is which, using the CORN rule you used in the model building practical.

Question 1:

Which of these SMILES string is the L, which is the D form?

O=C(O)C(N)C **ANSWER:** This is the L isomer because the CORN rule is followed in an counter clockwise rotation when the hydrogen is pointed toward the screen.

O=C(O)C(C)N **ANSWER:** This is the D isomer because the CORN rule will can only be followed in a clockwise rotation when the hydrogen is pointed towards the screen.

Question 2:

Create your own smiles strings for:

L-lysine **ANSWER:** O=C(O)C(N)C(CCCN)

D-isoleucine **ANSWER:** O=C(O)C(C(C)CC)N

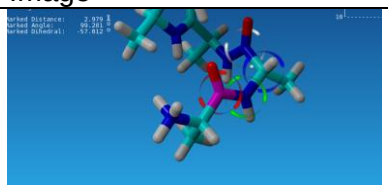
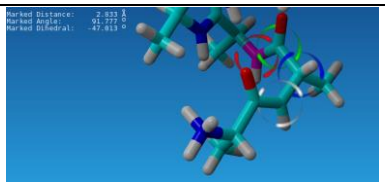
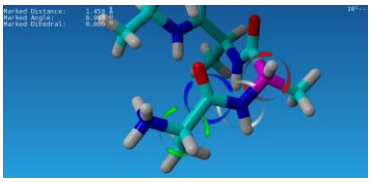
L-phenylalanine **ANSWER:** O=C(O)C(N)C(C1ccccC1)

3. Secondary structure: the α-helix

Open the file called **HELIX.PDB** in the pdb directory: this contains a right-handed polyalanine α-helix. Measure the ϕ, ψ and ω torsion angles. You can do this by selecting the appropriate atoms (hold down ctrl) and the dihedral (torsion) angle appears at the bottom of the ATOM PROPERTIES on the left of the screen.

Question 3:
What are the values for the torsion angles ϕ, ψ and ω in a right-handed α-helix?

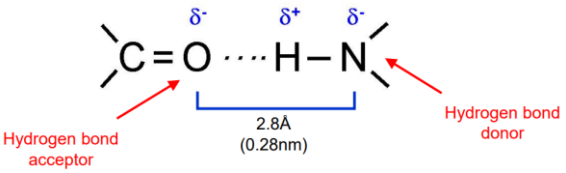
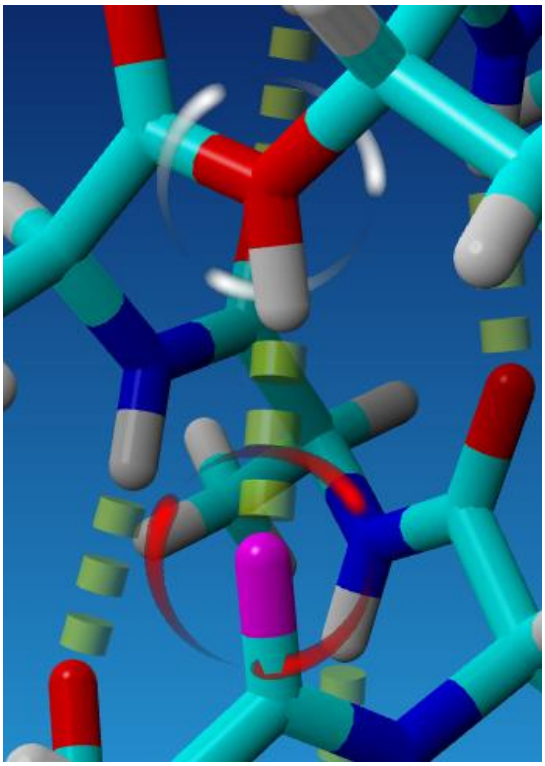
Question 3 Answer:

Angle	Image	Value
The ϕ angle value in the right-handed alpha helix		-57.012
The ψ angle value in the right-handed alpha helix		-47.013
The ω angle value in the right-handed alpha helix		This is a trans peptide bond as the omega angle is -179 (close to 180 degrees).

You can display the hydrogen bonds by selecting **VIEW**, **SHOW INTERACTIONS**, **HYDROGEN BONDS OF**, **ALL**, then pick the bottom option, which will display all H-bonds. To measure a distance between 2 atoms, select both atoms and the value is displayed at the bottom of the ATOM PROPERTIES on the left-hand side of the screen.

Question 4:
Measure the lengths of the hydrogen bonds (distance between donor and acceptor atom) and compare them with the values shown in Appendix 1.

Question 4 Answer:

The length of hydrogen bonds in appendix 1 varies between 2.70 and 3.4 angstroms, depending on the atoms within the bond. The hydrogen bonds within a right handed polyalanine alpha helix are between the CO group of each amino acid forms a hydrogen bond with the NH group of amino acid four residues earlier in the sequence.	
The length of the hydrogen bonds in Yasara is **3.092-3.094Å** (measure from nitrogen to hydrogen). This is slightly longer than the hydrogen bonds in *Appendix 1*: they are nitrogen oxygen bonds which are typically 2.90 ± 0.1 Å and 2.80 ± 0.1Å between GC base pairs and peptide bonds, respectively. In the case of **N-H—O=C H-bonds** in the helix they measured around 3.1Å .	

Now calculate the energy of the right-handed α-helix with L-amino acids. Select **SIMULATION**, **DEFINE SIMUATION CELL,** click **OK** (do not change any of the parameters!). This defines the space in which energy calculations will be made. Also select **SIMULATION**, **FORCEFIELD, AMBER99, OK, and if a forcefield is selected above, also set its default parameters**. This selects the forcefield for the calculation, and therefore the "rules" and parameters by which the energy is calculated. AMBER99 is appropriate for proteins. To measure the energy, click **ANALYSE**, **ENERGY**, **POTENTAL ENERGY – ALL**. Select the defaults (calculate energy of all components). The value for energy appears in the console at the bottom of the screen.

Now calculate the energy for a right-handed α-helix with D-amino acids (this is energetically equivalent to a left-handed helix with L-amino acids). Click on **EDIT**, **SWAP**, **RESIDUE**. This opens a box, and you need to select all 22 of the residues (hold down shift to select them all), and click **OK**. Another box opens, and you need to select **Ala** and **D-amino acid,** and click **OK**. All of the amino acids in the helix have been swapped for D-alanine. Recalculate the energy for this helix.

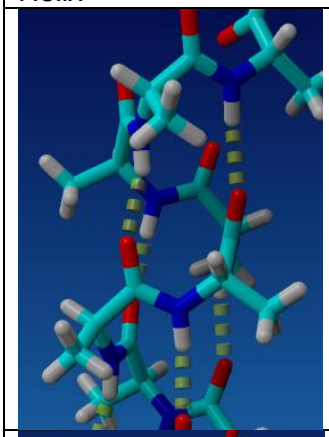
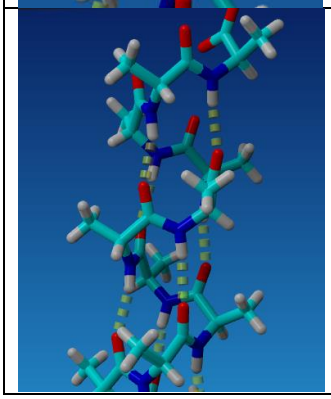
Question 5:

What are the values of the energy for both helices?

Energy values for the right-handed alpha helix with L amino acids:

Energy values for the right-handed alpha helix with D amino acids (swapping the amino acids to unnatural D chirality):

Question 5 answer:

Helix	Energy
	Energy for the right-handed alpha helix with L-amino acids: **1520.54 KJ/mol**
	Energy for the right-handed alpha helix with D-amino acids: **1979.93 KJ/mol**

Question 6:

Why is the right-handed α-helical conformation more stable than the left-handed conformation, which has never been observed in proteins?

Question 6 Answer

What is the right handed helix?

The right-handed alpha helical conformation (clockwise) is more stable than its left-handed (counter-clockwise) form (the right handed alpha helix with D amino acids is equivalent to a left handed alpha helix with L amino acids) [1].

Why is it more stable?

This is because the right-handed alpha helix has fewer steric clashes between the side chains and main chain which makes it lower energy and thus a thermodynamically more favourable conformation for amino acids to form in a helix. As a result, less than one percent of residues are found in contiguous left-handed turns or helices of length three amino acids or greater [2].

[1] Annavarapu, S. and Nanda, V. (2009) *Mirrors in the PDB: Left-handed α-turns guide design with D-amino acids - BMC structural biology*, *BioMed Central*. BioMed Central. Available at: https://bmcstructbiol.biomedcentral.com/articles/10.1186/1472-6807-9-61 (Accessed: December 7, 2022).

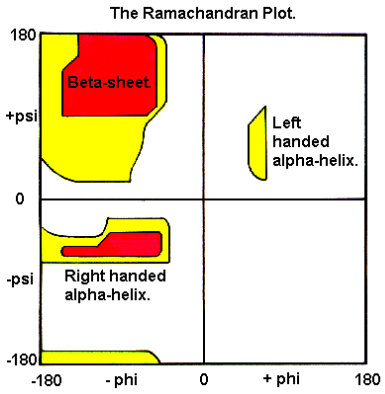
[2] *Alpha Helix* (no date) *Alpha Helix - an overview | ScienceDirect Topics*. Available at: https://www.sciencedirect.com/topics/medicine-and-dentistry/alpha-helix#:~:text=Despite%20the%20fact%20that%2C%20based,chains%20and%20the%20main%20chain (Accessed: December 7, 2022).

4. Secondary Structure: β-sheet

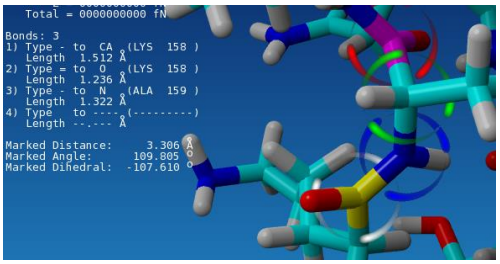
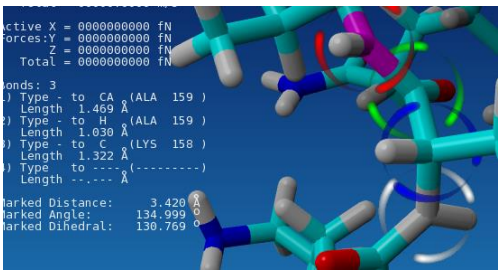
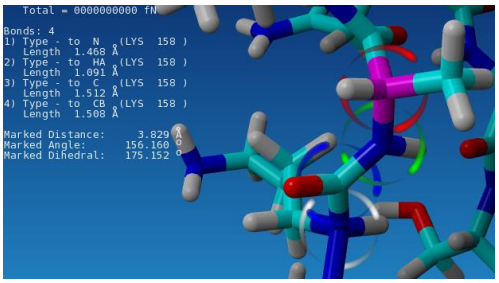
Delete the previous molecule and open the file **BETA.PDB** containing two strands of anti-parallel β-sheet, taken from a real protein.

Question 7:

What are the values for the ϕ, ψ and ω angles?

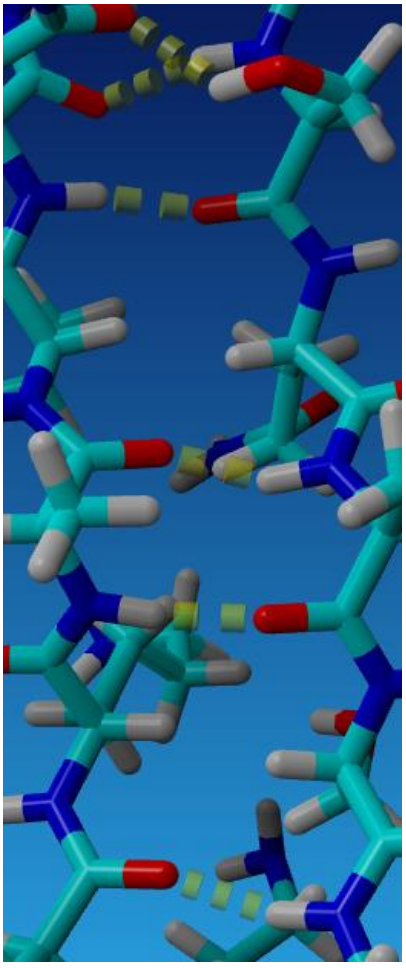


Question 7 Answer:

Angle	Yasara image
The ϕ angle value in the right-handed B-sheet is: **-107.610** Range: -155 to -107	
The ψ angle value in the right-handed B-sheet is: **130.679** Range: 127 to 162	
The ω angle value in the right-handed B-sheet is: **175.152** Range: around 180	

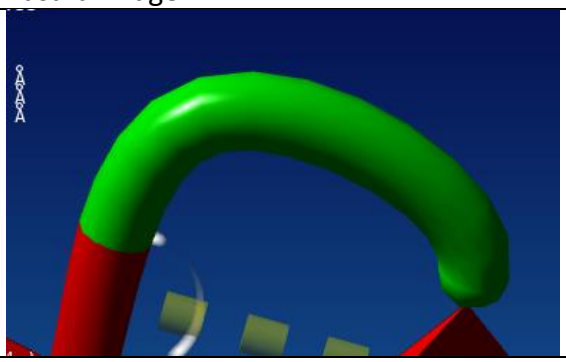
Question 8:
Identify the hydrogen bonds in the β-sheet and measure the lengths of a few of them.

Question 8 Answer

Explanation	Hydrogen Bonds
The hydrogen bonds in the B-sheet occur between the antiparallel strands. They vary in length due to the misaligned structure of the antiparallel strands. This is intermolecular hydrogen bonding between two antiparallel sheets rather than intramolecular hydrogen bonding that occurs in the alpha helix. The bonds are, however, shorter and thus stronger than hydrogen bonds in parallel beta sheets because antiparralel beta sheets are arranged so that C=O and N-H from adjacent strands face each other (hydrogen bonds in parallel beta sheets can reach 3.5 angstroms) Hydrogen bonds toward N and C termini are longer whereas towards the core of the beta sheet are shorter. Toward the N and C termini, hydrogen bonds are 3.183 Å, 3.024Å, 2.985Å, 3.081Å etc. In the middle of the beta sheet, hydrogen bonds are 2.810Å, 2.786Å and 2.839Å. Thus, overall, hydrogen bonds range from 2.786 to 3.183 Å within the beta sheet.	

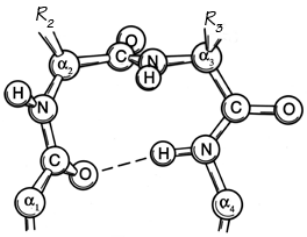
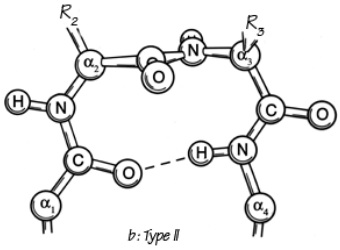
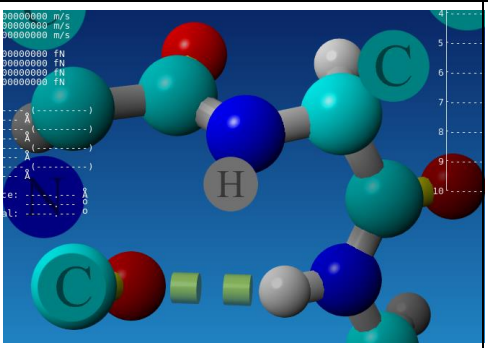
Question 9:
Which 4 residues are involved in the reverse turn?

Question 9 Answer

Residues	Yasara Image
The reverse turn involves 4 specific residues including: ASN 161, ARG 162, VAL 163 and SER 164.	

Question 10:
What type of β-turn is found between the two strands?

Question 10 Answer

Different types of β-turns (appendix)	The Yasara Turn	Explanation
 a: Type I  b: Type II		The reverse B-turn found between the two strands is a **type 1 turn**. This is because there is no glycine residue at position 3 which is required to avoid the carbonyl oxygen of residue 2 in a type 2 turn. As with both kinds of turn, there is a hydrogen bond between the C=O of residue 1 and N-H of residue 4. The yasara image matches the type 1 diagram (the oxygen is behind the carbon and hydrogen is in front of the nitrogen)

5. Conformational analysis of an amino acid residue:

Construct the following molecule:

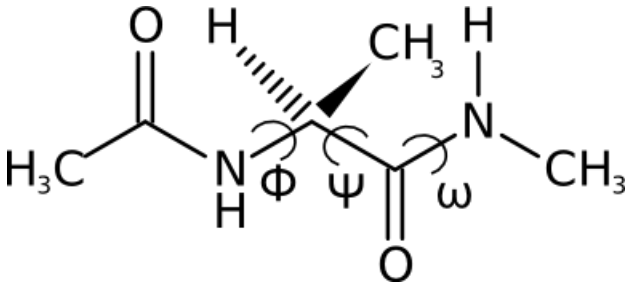


Figure 2: N-acetylalanyl-N'-methylamide

The SMILES code for this molecule is: CC(=O)NC(C)C(=O)NC

This represents a single alanine residue in a polypeptide chain, blocked on either side by methyl groups. Save your molecule, so that you can return to your initial conformation after exploring different conformations, if necessary. An analysis of the conformations accessible to this molecule will reflect the possible conformations for a polypeptide chain in a protein at an alanine residue, as defined by the two torsion angles ϕ and ψ. The results of such an analysis, first performed by Ramachandran and Sasisekharan (*Adv. Prot. Chem*., **23**, 283, 1968), are represented by the familiar plot of ϕ against ψ, shown below.

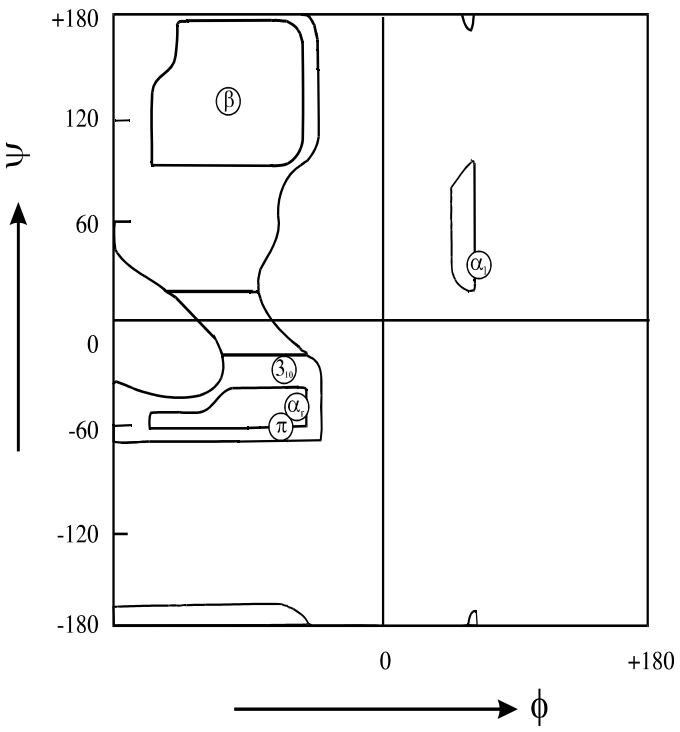


Figure 3: Ramachandran Plot

You can now investigate these 'allowed' and 'disallowed' conformations in more detail. Before you start, select the simulation cell and forcefield as before (section 3). To change a dihedral (torsion) angle, select the atoms for that angle, right hand click on the flashing white selected atom, select **GEOMETRY**, **DIHEDRAL**, which brings up a box with a slider bar which you can use to change the angle. Click **OK** when you are done.

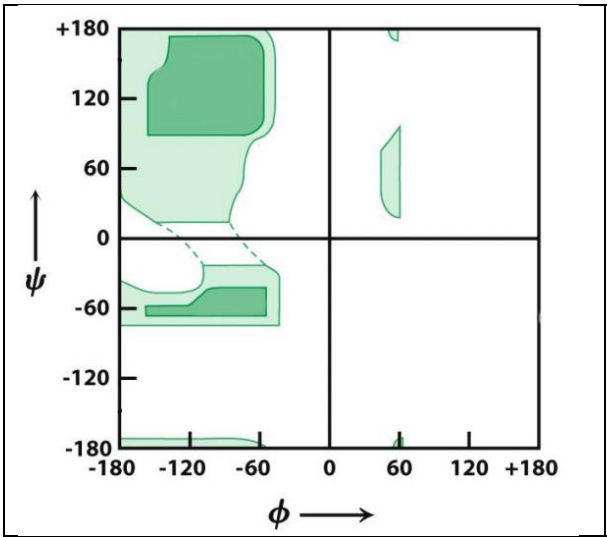
Set ϕ and ψ to any desired pair of values, then calculate the energy of the molecule. Investigate 3 or 4 different pairs of values including both allowed and disallowed regions of the plot.

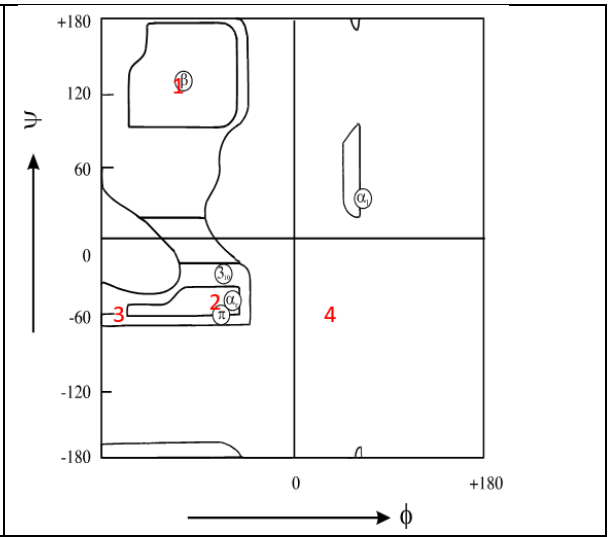
Question 11:

Explore different regions of the Ramachandran Plot, and confirm the principal regions of low energy, and other regions of high energy. Either annotate the Ramachandran Plot above indicating your values, insert an annotated scanned image, or list the coordinates of the points that you selected together with the energy values.

Baseline Energy of the Molecule: **16.18kj/mol**

Question 11 Answer





Region	Phi	Psi	Energy
1 No steric clashes region (dark green): B sheet – between antiparallel and parallel	-90	120	**106.58 Kj/mol** (far less favourable than base)
2 No steric clashes region (dark green): right-handed alpha helix – very narrow	-65	-55	**130.81 Kj/mol**
3 Allowed if slightly shorter van der Waals radii are used (light green): right handed helix	-169	-60	**15.03 Kj/mol**
4 Disallowed region (white)	30	-60	**20181.34 Kj/mol**

6. Energy minimisation of a molecule: Ethane

Construct ethane:

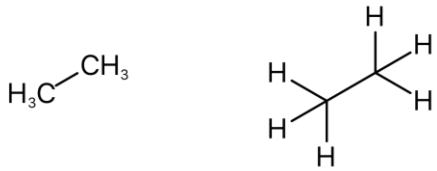


Figure 4: The structure of ethane

You should be able to work out the SMILES code for this yourself (remember there are instructions in Appendix 3).

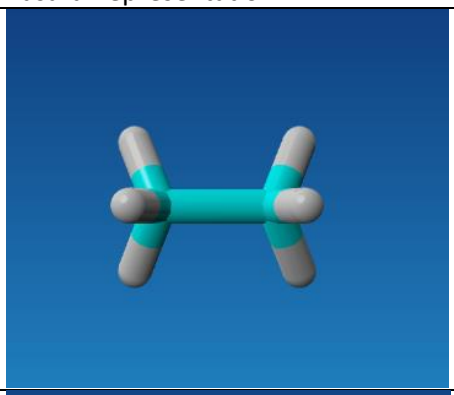
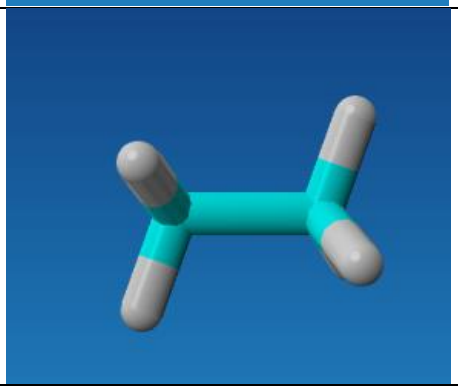
Set the dihedral (torsion) angle to zero. Calculate the energy (don't forget to define the simulation cell and forcefield, Amber99). Now you are going to run an energy minimiser to drive the structure to an energy minimum. Before you do this, click **WINDOW**, **HEAD UP DISPLAY**, **SIMULATION PARAMETERS**. This changes the display on the right of the screen from SCENE CONTENT to SIMULATION PARAMETERS. Then select the type of simulation you want by pressing **F10** repeatedly until you get to STEEPEST DESCENT, which appears under TEMP CTRL in the middle of the SIMULATION PARAMETERS.

To turn on the simulator, click **SIMULATION**, **SIMULATOR INITIALISE**, then **SIMULATION**, **SIMULATOR ON.** (**F12** also turns the simulator on and off). The molecule will change conformation and adopt a more energetically favourable structure. Stop the simulator and calculate the energy of the new structure.

(If the molecule doesn't change conformation, but remains at 0°, set the torsion angle to 1° and try again).

Question 12:
What are the energy values for the eclipsed and staggered versions of the ethane molecule?
The default energy: **7.71 Kj/mol**

Question 12 Answer

Molecule Type	Yasara Representation	Energy
Eclipsed: This is the default (torsion 0)		**20.15**kJ/mol
Staggered: This is the eclipsed form (torsion)		**7.56** kJ/mol
Explanation:	The energy of the eclipsed conformation is approximately 3 kcal/mol (12 kJ/mol) higher than that of the staggered conformation. The 0 dihedral angle formation is the eclipsed form which is around **20** Kj/mol The simulator adopts a new staggered conformation that moves around the Hydrogen atoms in space. This creates less repulsion and a far lower energy.	

Question 13:
Why do you think the energy is lower in the staggered molecule?

Question 13 Answer

A staggered ethane molecule is a form where the bonds to the axis of rotation are not aligned so there are two atoms and/or groups whose dihedral angle is ~60 degrees; thus, all the atoms are equally spaced from each other [1].

An eclipsed ethane molecule is one where the carbons are aligned so that the hydrogens are not lined up with each other which creates steric hindrance between them.

As a result, the lack of steric hindrance in a staggered molecule renders this a lower energy conformation to form by the amino acids. It is thus the favoured conformation. Staggered is less repulsion between the hydrogen atoms [2].

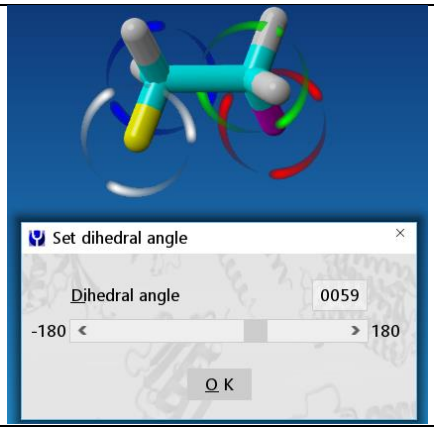
[1] (no date) *Illustrated glossary of organic chemistry - staggered*. Available at: http://www.chem.ucla.edu/~harding/IGOC/S/staggered.html#:~:text=Illustrated%20Glossary%20of%20Organic%20Chemistry,of%20rotation%20are%20not%20aligned. (Accessed: December 7, 2022).

[2] (no date) *Conformation*. Available at: https://www.sas.upenn.edu/~badoyle/Conformation.html (Accessed: December 7, 2022).

Question 14:
What is the torsion angle in the staggered molecule?

Question 14 Answer

It is 59 degrees in Yasara.
The torsion angle in a staggered molecule is 60 degrees; the 3 hydrogen groups attached to the carbon front and back atoms are at a maximum distance from each other.



7. Looking at protein structures

In the YASARA/PDB directory there are a variety of other protein structures you can look at. Filenames and the structures are listed in the table below - all structures are from the Protein Data Bank (www.rcsb.org). Have a look at a selection of the samples to familiarise yourselves with their tertiary structures and identify the secondary structure elements. Experiment with the various display options. If you prefer, you may download a structure of your choice from the PDB – powerful search tools are available on the website, and there is the molecule of the month, etc.

Question 15:

Download one of the structures and label the main features: secondary structures, disulphide bonds and co-factors if present. Insert the annotated image.

Protein structures were originally stored as legacy PDB files, then CIFs (Crystallographic Information Files) to archive small molecule crystallographic experiments. A lot of files now are in the PDBx/mmCIF format which utilizes the ASCII character set. Protein data bank (PDB) files are textual files which described 3D structures of protein: atomic coordinates, secondary structure assignments and atomic connectivity. The structures on the protein data bank website are obtained by X-ray crystallography, NMR spectroscopy and cryo-electorn microscopy.

Homology modelling is a multi-step process used to predict protein structures.. This begins with the template identification and alignment, followed by backbone generation, loop and side chain modelling and then model optimization. The pdb structures provided are from protein databank

Table 2: A selection of protein structures from the protein data bank (PDB)

Protein	Filename	Notes
Crambin	**1CRN.PDB**	A small hydrophobic protein from Abyssinian cabbage seeds (*Crambe abyssinica*).
DNA-CRO repressor complex	**3CRO.PDB**	File contains the structure of the CRO repressor and also a stretch of DNA. Molecules A and B are the two DNA chains.
Insulin hexamer	**1ZNJ.PDB**	You should be able to see the Zn^{2+}, and the dimer interface between monomers.
Triose phosphate isomerase	**1WYI.PDB**	Classical beta barrel flanked by alpha helices. Dimer of two identical chains (A & B).
DNA polymerase I-Klenow fragment	**2KZZ.PDB**	This is a large structure with the DNA polymerase bound to single-stranded DNA.
Neuraminidase	**4HZV.PDB**	Viral coat protein: note the "super-barrel".
IgE Fab complexed with a grass pollen allergen	**2VXQ.PDB**	Try identifying the V and C type domains and the heavy and light chain
GCN4 leucine zipper	**4DMD.PDB**	This is the leucine zipper from the yeast DNA-binding protein GCN4. Note the H-bonding patterns in the helices.

Question 15 Answer: DNA-CRO repressor complex

DNA-CRO repressor – What is it?

The **DNA CRO-repressor** is a family of repressor proteins. They are found in the bacteriophage lambda that includes the Cro repressor [1].

The Cro-repressor's dimer is held together by weak interactions between the two proteins within: molecule 1 and molecule 2 in the figure below. The interactions are colour coded based on a yasara legend: HBonds=Red,Hydrophobic=Green,Ionic=Blue. Interestingly, most of the interactions are hydrophobic and ionic within the dimer [2].

The secondary structure of the protein consists of 5 alpha helices, which from a compact bundle within the hydrophobic interior. Helices 2 and 3 form a helix-turn-helix motif. This secondary structure did not change between complexed and free DNA-CRO repressor proteins so there is no major change in conformation upon binding (**fig.2**).

The main change in the protein upon DNA binding is within the torsion angles of side chains that lose/acquire important interactions upon binding. In helix 3, this includes residues 27, 28, 29 and 32 which change conformation and contact the DNA. Phenylalanine 46 also changes conformation to fit the dimer interface and both arginine 43 and glutamic acid 47 move to form a salt bridge between monomers. The dimer interaction as seen in Yasara is represented below [2].

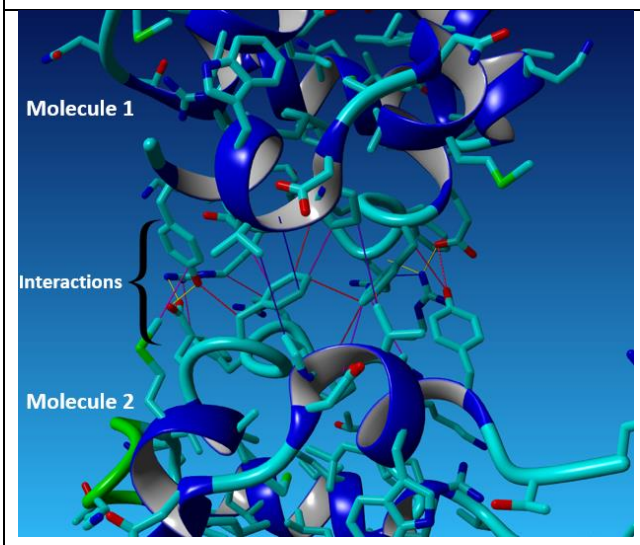


Figure 1: CRO-Repressor dimer form with interactions shown between the two molecules (molecule 1 and 2), colour coded according to the legend below.

Labelled Structure: CRO Dimer visualised [2]

Bond	Colour
Hb (hydrophobic)	Red
Hyd	Green
Hyd + HB	Yellow
Ion	Blue
Ion + HB	Magenta
Ion + Hyd	Cyan
Ion + Hyd + HB	Gray

Labelled Structure: interactions between CRO dimers (**fig.1**)

The interactions, between the Cro dimers include 4(**Hyd+Hb**). The interactions are mainly facilitated by hydrophobic interactions (**Hb**) between the rings of tyrosine and phenylalanine groups within the two components of the dimer.

There are some ionic interactions: two C-N bonds and one C-C bond which form. These will facilitate most of the force that hols the dimer components together. The ionic bonds are important to hold the protein together: hydrogen bonds would not be enough force withstand external force and pressure.

The dimer interface includes bulky side chains of phenylalanine 46 and tyrosine 61 to create a hydrophobic core [2].

However, the weak interactions provide the flexibility needed in the dimer to bind DNA effectively. [3]

Helix-Turn-Helix Motif and additional Beta Sheets (secondary structure)

The repressor is a member of the **helix-turn-helix protein superfamily**. It binds to DNA as a highly flexible dimer. The contacts in the major groove between side-chains of the helix-turn-helix and groups on base-pairs 1 through 4 determine recognition of the 5'-ACAA element. The protein consists of five cr-helices, which form a compact. bundle with a hydrophobic interior. Helices 2 and 3 form the helix-turn-helix motif [4].

The structure of the dimer is rather unique compared to other Cro family members: most Cro proteins have an all-alpha structure. Due to secondary structure switching, Cro has it's fourth and fifth helices replaced by a beta sheet (labelled fig1) [4].

The helix-turn-helix and beta sheets are labelled on the right.

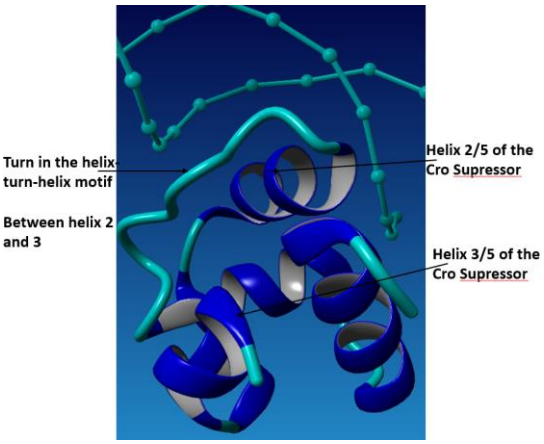


Figure 2: This labelled image shows the helix-turn-helix motif in the DNA CRO-Repressor. It occurs between helices 2 and 3.

Interactions

There are only two regions of hydrogen bonding between the Cro-Dimer and the alpha helix of DNA (**fig.3a**). The rest of interactions are hydrophobic. The DNA fragment contains the sequence of the **OR1 operator site**.

There are differences between Cro interaction with nonspecific and specific DNA.

CRO-Nonspecific DNA interaction is salt-sensitive. This means the interaction is mainly electrostatic (like in **fig.3b**): Cro-binds to DNA based on the salt concentration which can easily be controlled.

The central base pairs influence repressor affinity by affecting the relative ease with which the DNA backbone in this region can adopt the conformation imposed by the protein. The interaction is mainly facilitated by outer four base pairs of the 12 naturally occurring half-sites: 5'-ACAA. These are conserved operator positions that involve nearly all base-pair contacts. All these contacts are similar aside form a single major groove Van der Waals contact to base pair 5 and minor groove water-mediated hydrogen bonds to base pair 7 (**fig.3**). The central four base-pairs are not in contact with any atoms in the protein.

The hydrophobic interactions can be visualised in fig3b. Almost all interactions are simple hydrophobic interactions (red). However, there are also some magenta interactions which are ionic and hydrophobic. Ionic interactions explain the strong binding of the dimer to the helix. [4]

=

Figure 3a: Hydrogen Bonds
This figure shows the hydrogen bond interactions that occur between the CRO Repressor and the DNA-DNA alpha helix.

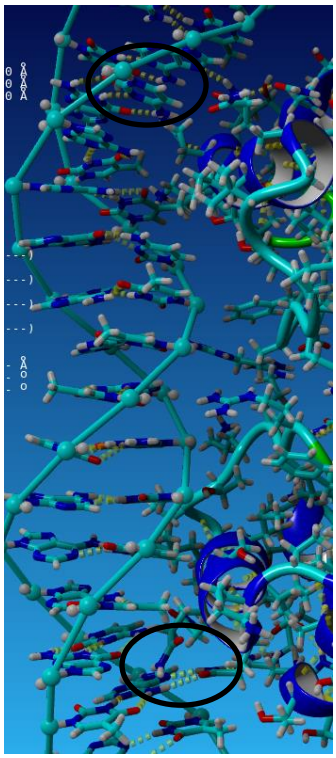
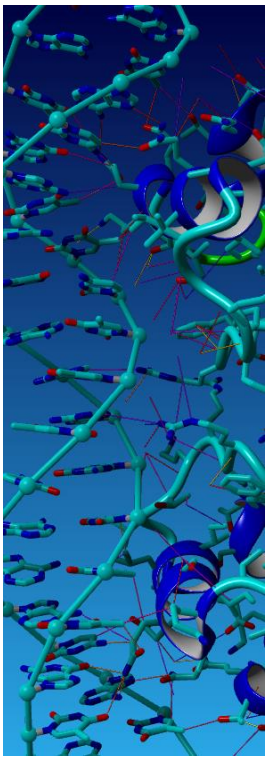


Figure 3b: All Interactions
This figure shows the hydrophobic and ionic interactions that occur between the DNA-CRO Repressor and the DNA alpha helix.



References:
[1] https://pubmed.ncbi.nlm.nih.gov/2038059/
[2] http://www.yasara.org/YASARAScienceManual.pdf
[3] https://pubmed.ncbi.nlm.nih.gov/9653036/
[4] https://pubmed.ncbi.nlm.nih.gov/3522575/

8. Protein ligand interactions

The enzyme dihydrofolate reductase (DHFR) catalyses the NADPH-linked reduction of dihydrofolate to tetrahydrofolate. The latter is vital to metabolism, as it is an intermediate in the biosynthesis of amino acids, thymidylate and purines. Inhibition of DHFR leads to depletion of thymidylate and purines and consequently to inhibition of DNA synthesis. DHFR is therefore an excellent target for drug design, and inhibitors of DHFR (also known as antifolates) that are in clinical use include the anti-bacterial trimethoprim, the anti-malarial pyrimethamine and the anti-neoplastic agent methotrexate. The structures of these inhibitor molecules and the folate substrates are shown in Figures 9-12 in Appendix 2.

The aim of this practical is to compare the structures of the three drugs trimethoprim (TMP), pyrimethamine (PYR) and methotrexate (MTX), and to investigate the binding of these molecules to the active site of DHFR using YASARA.

The DHFR-methotrexate complex

Using YASARA, open the file 3DFR.PDB. This file contains the crystallographically determined structure of the complex between DHFR (from *E. Coli*) and methotrexate (MTX) (*J.Biol.Chem. (1982)* ***257:*** *13650*). The molecule is displayed as CPK with the water molecules present in the crystal structure. To simplify viewing it, begin with clicking **VIEW**, **STYLE SCENE**, **RIBBON**. This displays the protein as a simple chain, with the ligands still displayed as CPK. There are two ligands, NADPH and methotrexate (MTX). As we are not studying the NADPH in this practical you can delete it by clicking on one of the atoms (it is the molecule with the yellow phosphorus atoms), right hand click, then select **DELETE**, **MOLECULE**.

Remember that it is possible to UNDO your last commands if you make an error (there are small forward and backward curved arrows on the menu bar).

The main feature of DHFR is the eight-stranded β-sheet, which is flanked by α-helices. You can explore some of the other **VIEW** commands to make this clearer. The MTX molecule binds in a deep cleft in the surface of the protein. To investigate the interactions between DHFR and MTX, display the protein as **TUBE**, then the following interacting residues can be displayed as **STICK** by **VIEW**, **SHOW ATOMS**, **RESIDUE**, then clicking on the following residue numbers (hold down ctrl to select multiple residues at once), followed by **OK**.

Active site residues:
Leu4, Trp5, Ala6, Leu19, Asp26, Leu27, His28, Phe30, Arg31, Ser48, Phe49, Pro50, Arg57, Ala97, Thr116

You can now look at what types of interactions MTX makes with DHFR. To look at the details, you might find it clearer to remove the backbone, which can be achieved by clicking **VIEW**, **HIDE SECONDARY STRUCTURE OF**, **ALL**.

Save this view by clicking **FILE**, **SAVE AS**, **YASARA SCENE**, as you may want to have another look at it later in the practical (a SCENE file saves the current view settings as well as the molecules, but can still be edited).

Question 16:
Locate the β-sheet and notice the twist between adjacent strands:
<u>Briefly</u> identify <u>some</u> of the protein residues that interact with the MTX molecule and note below the type of interaction, i.e. hydrogen bond, electrostatic or hydrophobic interactions. (You don't have to identify every interaction!)

Question 16 Answer:

Locate the β-sheet and notice the twist between adjacent strands.

Human DHFR is a 186 amino acid protein with a Mr of 20kDa. It consists of an adenosine-binding subdomain and a loop subdomain. The loop subdomain contains three loops: the F-G loop, the Met-20 loop and the G-H loop. The beta sheet encompasses 35% of the polypeptide sequence. The 8 strands are termed: G-H-F-A-E-B-C-D. The strands are all parallel apart from one strand: βH.

The beta sheet is the red region of the structure. It can be seen the most clearly in ribbon representation. This shows the structure as a large redarrows with a twist in between adjacent strands at their ends to connect them (**fig1**). There are five regions of beta strand which form a central 8 stranded beta sheet. These regions exist between amino acids 29-36, 39-46, 52-62, 66-70 and 74-76. These are connected by twists and loop regions between adjacent strands. The alpha helix portion exists between amino acids 71-73 and 43-50 in the E.coli dihydrofolate enzyme we are analysing. There is also cis peptide linking **Gly-97** and **Gly-98**. [1]

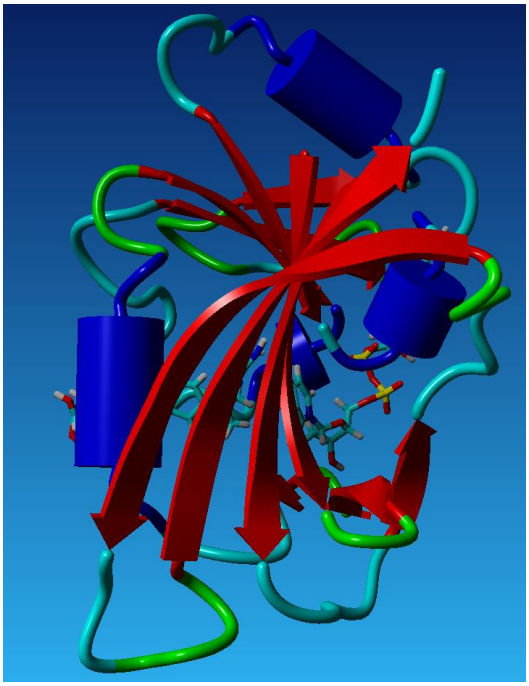


Figure 1: This the structure of DFHR-MTX. The DHFR is represented in ribbon form and MTX ligand as sticks. The twist can clearly be seen from this view. There is a clear cross between beta sheets that forms an X cross shape. This cross represents the twist between the 8 beta strands in the molecule.

<u>Briefly</u> identify <u>some</u> of the protein residues that interact with the MTX molecule and note below the type of interaction, i.e. hydrogen bond, electrostatic or hydrophobic interactions. (You don't have to identify every interaction!)

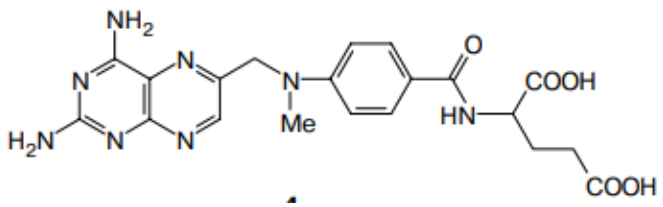
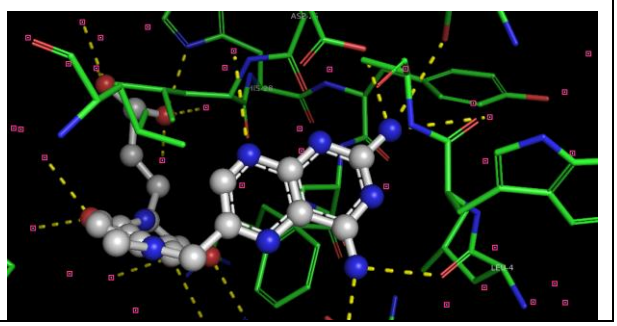


Figure 2: MTX interactions found



<u>The top of MTX: Methotrexate is bound in a 15-angstrom-deep cavity with the pteridine ring buried in a primarily hydrophobic pocket</u>

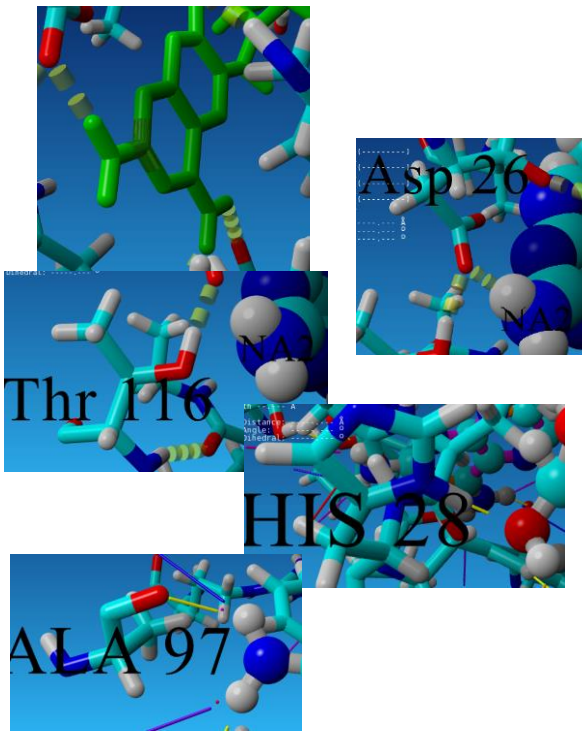
The bottom of MTX includes the heterocyclic aromatic ring. This ring contains nitrogen and carbons. The nitrogens in this ring act as hydrogen donors and form **4 hydrogen bonds** with the active site of MTX. Hydrogen bonds are formed with Asp-26, Thr-116, Leu-4 and Ala-97

Thr116 and Tryp 21 bind and stabilise water molecules in the active site. These water molecules form 2 additional hydrogen bonds with MTX's pteridine ring.

Thus there are 6 hydrogen bonds overall between MTX's heterocyclic ring and the active site environment of DHFR

There is also a positively charged Nitrogen in the ring. This positively charged nitrogen will form strong electrostatic interactions with the resonance-stabilised carboxyl group of Asp26.

Finally, there is some pi-pi stacking interaction between the aromatic ring in Tryp21 and the aromatic, heterocyclic ring of MTX

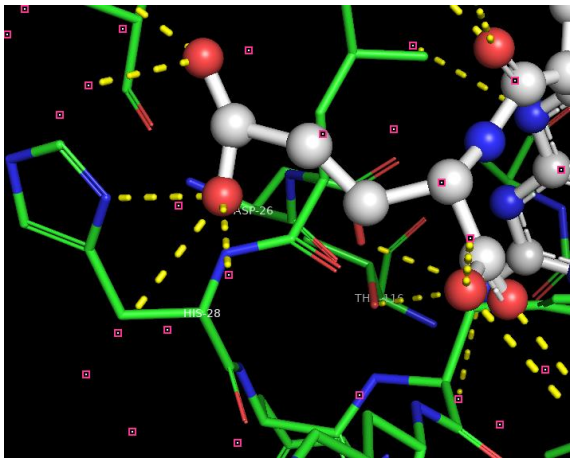


<u>The bottom of MTX: Methotrexate is bound in a 15-angstrom-deep cavity with the pteridine ring buried in a primarily hydrophobic pocket</u>

On the other side of the MTX molecule there are a lot of carboxyl groups. There form around **3 hydrogen bonds** with the active site of MTX.

The 3 hydrogen bonds occur specifically with basic residues like His-28 and Arg-57. For example, the carboxyl group a the very end (c-terminal) can form a hydrogen bond with basic **His28** within the active site of DHFR. The other carboxyl group (in the middle) can form 2 hydrogen bonds with **Arg57**.

In addition, the carboxyl groups can interact with water molecules within the active site and **form 7 more hydrogen bonds.**



<u>References</u>

[1] *BMRB featured system: Dihydrofolate reductase* (no date) *BMRB*. Available at: https://bmrb.io/featuredSys/dhfr/dhfr1.shtml#:~:text=Human%20DHFR%20is%20a%20186,helices%20with%20connecting%20loop%20regions

Conformational analysis of the drug trimethoprim.

Trimethoprim (TMP) is an anti-bacterial agent that also acts by binding with high affinity to, and inhibiting, the enzyme DHFR. You can build this using the following SMILES string:

c1nc(N)nc(N)c1Cc2cc(OC)c(OC)c(OC)c2

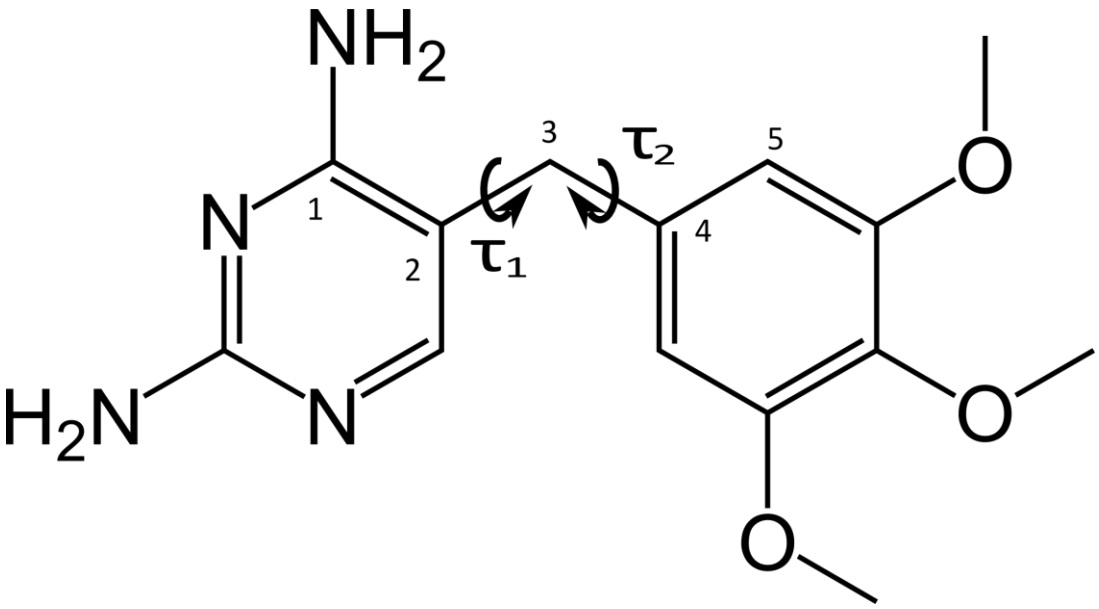


Figure 5: Trimethoprim

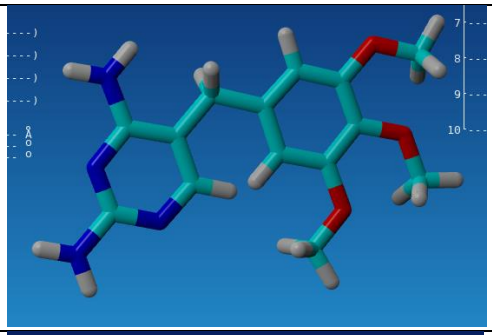
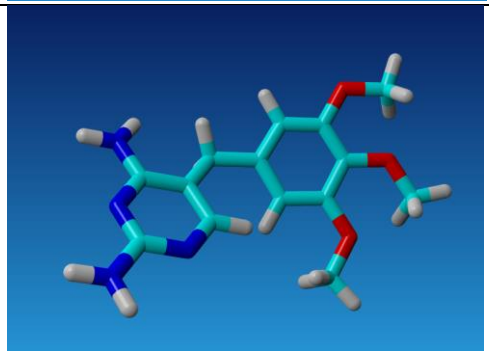
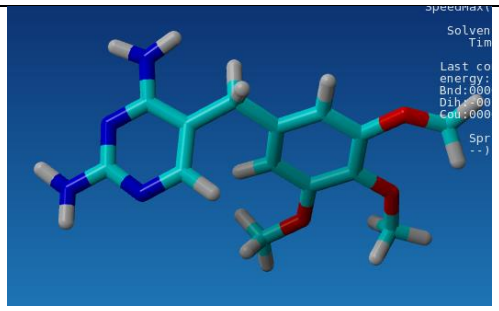
Set the bond angle at the central carbon atom between the two rings to be 118° which was established from an X-ray crystallographic structure determination of the molecule. You can do this by clicking the three atoms defining the angle, then right hand click on the white flashing atom and select **GEOMETRY**, **BOND ANGLES**. You can use the slider to change the angle. If you need to centre the molecule, click on one of the atoms, right hand click **CENTRE**, **ATOM**. This centres the molecule on the atom you selected

You are now going to find the lowest energy conformation of TMP. There are two dihedral (torsion) angles, defined as τ1 and τ2, shown in Figure 9 above. The aim of this part of the practical is to vary these angles and explore the energy landscape shown below (Figure 6). This is similar to what you did to explore the Ramachandran plot in the first practical. However, in this case, first set τ1 and τ2 to a high energy value (the black parts of the energy landscape) and measure the energy as you did before for the α-helix and ethane (don't forget to define the simulation cell and forcefield Amber99, as in sections 3 and 6, Graphics Practical part I). Then to minimise the structure, select **STEEPEST DESCENT**, turn the **SIMULATOR** on, and allow the structure to reach an energy minimum. Because the simulator is not the same as that used to generate the energy landscape in Figure 6, the actual energy values are not directly comparable, but the areas of high and low energy are a property of the shape of the molecule and will therefore be the same.

Question 17:
Check to see what effect the cycles of energy minimisation have had on your structure. Re-measure your dihedral (torsion) angles, your central bond angle and your new energy value. How have they been changed?

Question 17 answer:

Overview: The central carbon atom is between two other carbon atoms. The angle between the central carbon atom and the atom on the left is T1 and the angle on the right is T2.

The angle	The image	Energy Values
The molecule with default torsion angles in yasara. Angle between central carbons = 118. Angle determined by X-ray crystallographic structure of the molecule		Energy = 341103.02 kJ/mol
T1 and T2 **high energy configuration** using the black part of the energy landscape Angle T1 High Energy: 164 Angle T2 High Energy: -141		Energy = 398021.61 kJ/mol Very high and endothermic: would never form spontaneously
Low energy configuration generated by the simulation using steepest descent Angle T1 Low energy: -138 Angle T2 Low energy: -161		Energy = -1354.90 kJ/mol Much lower and more viable. Negative so is exothermic and will form spontaneously

Conclusion/Analysis: The energy value for the low energy configuration is far lower. This means that there is far more energy released in the formation of the molecule generated by the simulator, and thus it is a far more stable structure. This is expected as its dihedral angles are far more favourable in the Ramachandran plot: the dihedral angle has decreased significantly for T1 into the negative.

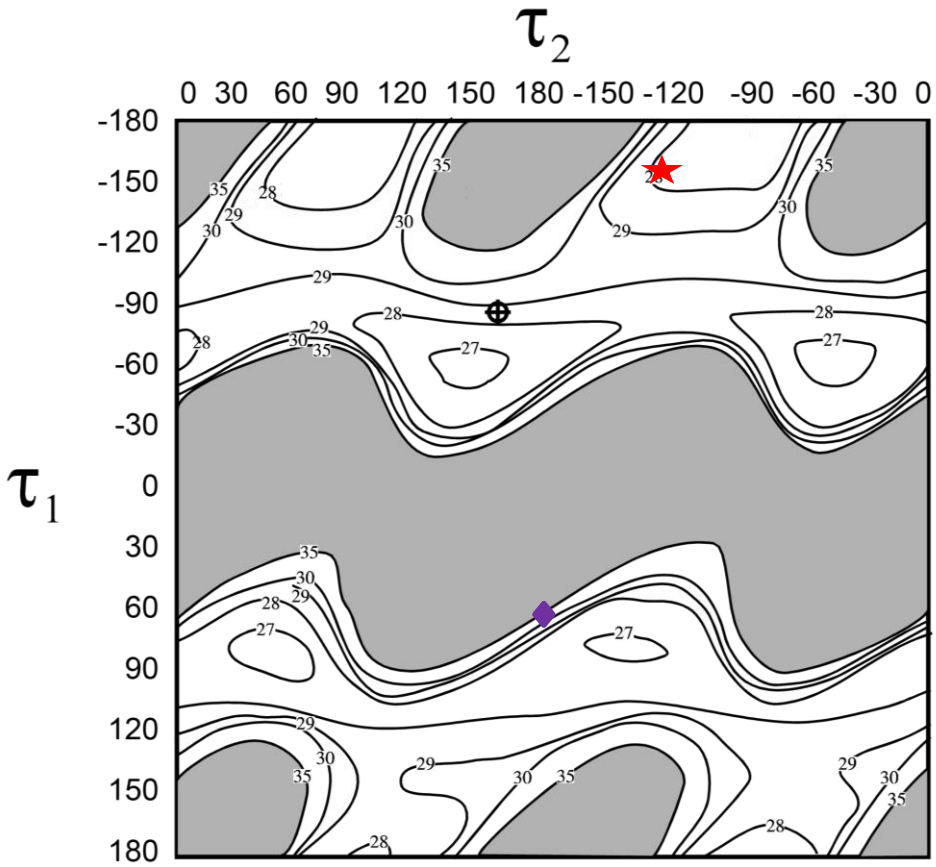


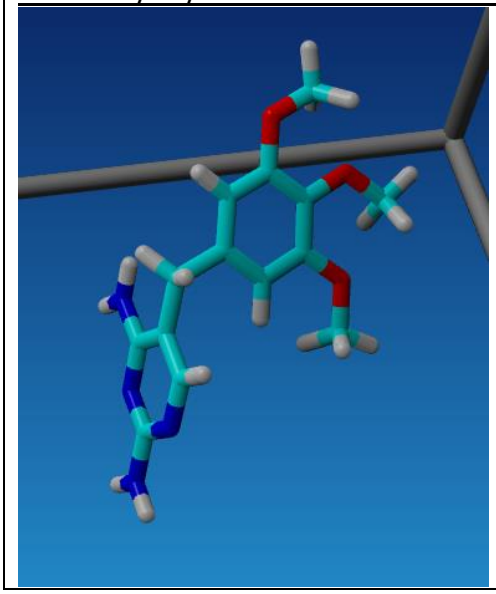
Figure 6: Conformational potential energy for trimethoprim as a function of rotations about the exocyclic C-C bonds, τ1 and τ2. The contour levels are in units of kcal/mol. The conformation observed in the crystal structure of TMP is marked with a cross. *(Koetzle and Williams, 1976). Copyright 1976 American Chemical Society.*

Note: -90 degrees is the same as +270 degrees; -120 is the same +240 degrees, *etc.*

Question 18:

Compare your minimised structure with the complete energy surface of the molecule, and either mark the approximate position on Figure 6 above or note the coordinates below. Have you found the global energy minimum of the structure?

Question 18 answer

The X-Ray Crystal Structure in Yasara:	Angles and Energy of X-Ray crystal structure of TMP
	T1: -80 degrees T2: 160 degrees Energy: -1055.98 KJ/mol

The minimised structure calculated by Yasara. This is marked on the Ramachandran plot using a **red star:** ★	Angles and Energy of minimised structure of TMP Angle T1 Low energy: -131 Angle T2 Low energy: -161 Energy: -1354.90 kJ/mol

The Global Energy Minimum of the structure:

The complete energy surface of the molecule has contours broadly spread out in the Ramachandran plot. The global anergy minimum of the structure corresponds to the regions marked '27' on the Ramachandran plot.

The minimised structure is not within this region, so I have not found the global energy minimum of the structure. The minimised structure is instead within contour 28.

Question 19:

Compare your structure with that of the X-ray crystal structure of TMP marked ⊕ on the energy landscape (Figure 6) above. Is the X-ray structure at the global energy minimum?

Question 19 answer

As mentioned above, the star is the simulation structure's region on the plot:

The bond angles of the marked, x-ray crystal, structure are T1 (-80) and T2 (160), as opposed to the bond angles of 176 and -128 in the energy minimum, respectively. This x-ray crystal structure of trimethoprim has an energy of **-1055.98 kJ/mol**. This is a low energy favourable structure; however it is not as low as the energy minimum of **-1366.62 kJ/mol** created by the simulation. This increase in energy of ~300kj/mol has taken placed the structure (marked +) into contour 29 region. This region is higher energy than the energy minimised structure (contour 28) and significantly higher than the global energy minimum (contour 27).

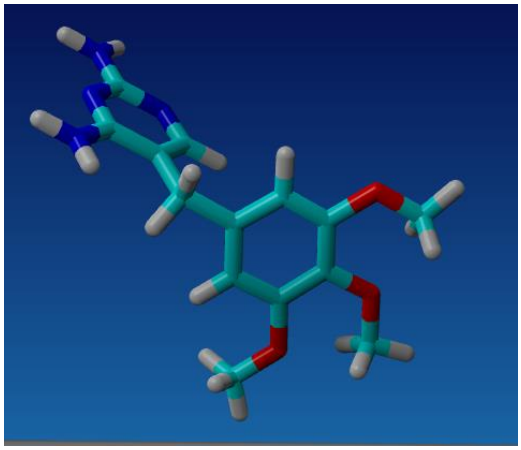
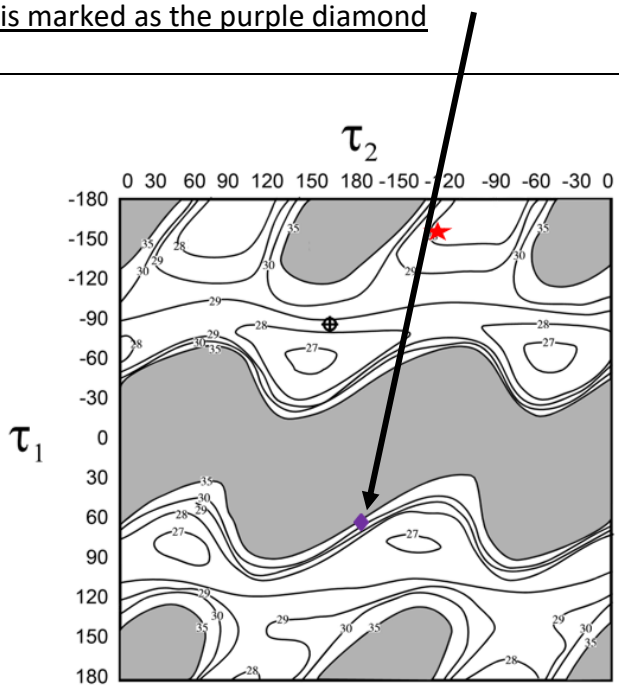
However, the crystal structure is still exothermic and it this favourable (it releases energy). A positive value of energy would mean the structure is endothermic and thus requires energy input to form.

Trimethoprim and its interaction with DHFR

In the preceding section you built the structure of **trimethoprim**, established the lowest energy configuration of the molecule, and found that this differed slightly from that found in the **crystal structure determination of TMP**. The crystal structure of the complex between TMP and DHFR has also been determined, and the conformation of the enzyme-bound TMP is given by the torsion angle values **τ_1 = 177° and τ_2= 76°**. Set these torsion angles, and also set the bond angle at the central C atom between the two rings to 118°, then measure the energy, and indicate where this conformation lies in relation to the others on the energy surface for TMP.

Question 20:
Is it a conformation likely to be adopted by the molecule free in solution?

Question 20 answer

TMP-DHFR complex conformation of TMP	Angles and Energy
	Bond dihedral T1: 177 Bond dihedral T2: 76 C-C-C bond angle: 118 Energy: The energy of the trimethoprim crystal structure when that binds the DHFR active site **is -990.36 kJ/mol.** This configuration is slightly more endothermic than the usual X-Ray crystal structure form. This is made possible by the active site which forms hydrogen bonds around the configuration that energetically stabilise it.
Ramachandran Plot Location: The new structure is marked as the purple diamond	Indicate where this conformation lies in relation to the others on the energy surface for TMP.
	The energy of -990.36 kJ/mol makes this a favourable exothermic structure so it will form spontaneously. Its energy level is very similar to the X-ray crystal structure of trimethoprim on its own and is only slightly more endothermic (the X-ray crystal structure is -1055.98 KJ/mol and is thus slightly lower energy) due to the addition of the TMP ligand in the DHFR active site. It is located within the highest energy region of all the structures previously looked at. This region is within contours 30-35 and the highest energy regions. As expected, like the TMP crystal structure, this is also more endothermic than the theoretical 'energy minimum' conformation that has an energy of -1366.62/mol; an extremely low energy value that is hard to achieve in practice.

Is it a conformation likely to be adopted by the molecule free in solution?

The most common way to determine if a conformation is likely to be adopted by a molecule free in solution is to compare its molecular energy to other conformations. The lower it is than other conformations, the more likely it is to be adopted by the molecule.

Hydration energy is important for determining if a molecule comformation can be adopted as it is the energy released when a molecule interacts with water molecules. This can affect the conformation adopted by the molecule due to an increased stability of certain conformations in the prescence of water.

Although still highly exothermic (~-990kj/mol), this conformation is the highest energy conformation compared to the X-Ray structure (-1055.98kj/mol) and the minimised structure (-1366.62kj/mol). This means it is less exothermic and thus less favourable than the other conformations and thus they will take precedence to form in solution.

The only redeeming factor could be the ability of the conformation to interact with water as bonds with water will provide additional stability and lower the overall energy of the system. Hydrophilic molecules tend to adopt conformations that maximise the interactions with water.

Since all these molecules are equally hydrophilic, then it simply comes down to their overall energy. The TMP-DHFR comformation is in the highest energy region of the Ramachandran plot. As a result, it is not likely to be adopted by the molecule free in solution compared to the other, lower energy conformations like the X-Ray structure and the minimised structure.

Save this conformation of TMP as a PDB file, as you will need to read it in again shortly.

Now you will open a file containing a simplified version of the active site of DHFR and will dock TMP into this site. Click **FILE**, **LOAD**, **YASARA SCENE**, **colisite.sce**. This has residues from the active site shown in STICK, and a "framework" (magenta). The framework will act as a guide to help with the orientation when you carry out the docking procedure. Inspect this model and note the relative disposition of the residues and their different chemical nature. The following residues are represented:

1 Aspartic Acid
1 Phenylalanine
3 Isoleucines, 2 with their adjacent peptide units
1 Leucine

Load the TMP molecule with the correct angles that you saved previously.

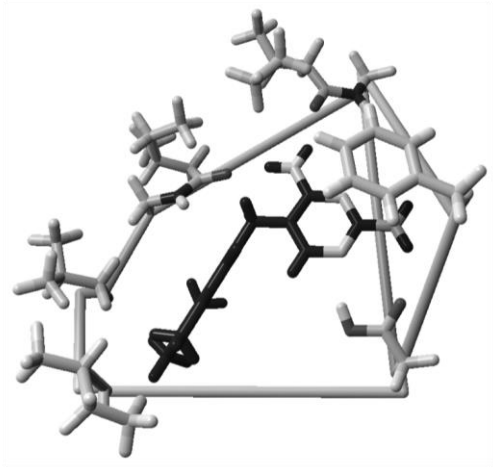
Compare the structures of the active site and the TMP molecule, and by rotating and translating the TMP relative to the site, dock TMP into the active site of DHFR. When you have finished your docking, you can save it as a **YASARA SCENE** file (this will save it looking exactly like what you see on the screen), or as a screenshot.

You may find it useful to use some of the following keys to help with the movement of the molecules – it can be easier to control molecules than using the mouse:

SHIFT: toggles between object 1, object 2 and object all (see bottom right of screen). All the move commands then relate to whichever object is selected. 1 is DHFR, 2 is TMP, all is both together
A and S: X rotation
D and F: Y rotation
G and H: Z rotation
Q and W: X translation
E and R: Y translation
T and Y: Z translation

Look for possibilities of H-bonding, electrostatic or hydrophobic interactions in both the active site residues and the TMP molecule. You can identify and measure the H-bonds and hydrophobic interactions – remember hold down ctrl to select more than one atom, and it will report the distance under ATOM PROPERTIES on the left-hand side of the screen, at the bottom.

N.B. You may not find this easy! To help you, the following figure shows the location of TMP relative to the active site residues.



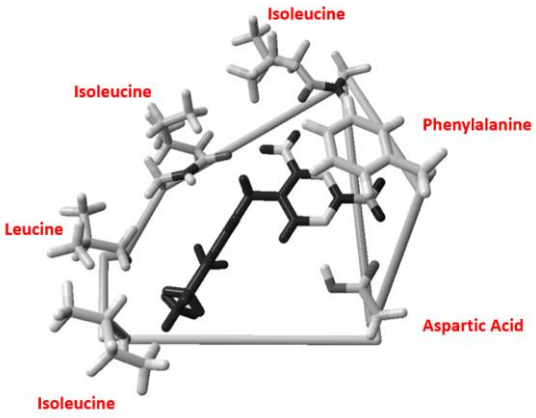
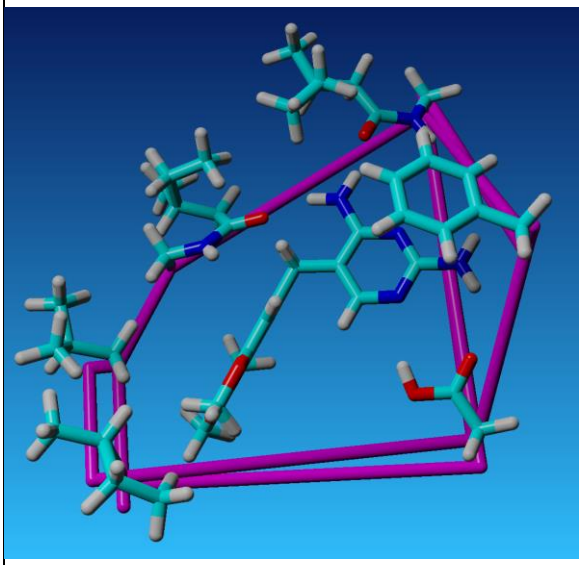
Question 21:

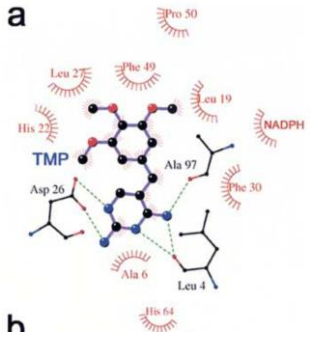
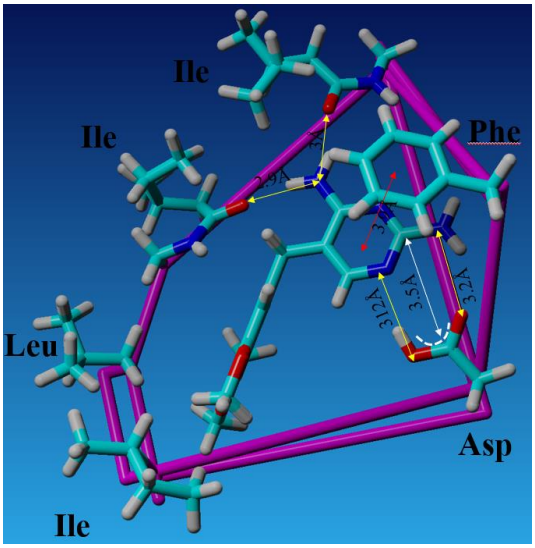
When you have established a model of the complex, note the hydrogen bond lengths and other non-covalent interactions. You can either annotate Figure 7 above, insert an annotated image, or list the interactions below.

How do the lengths of the H-bonds and any close van der Waals interactions compare with the ideal values for these distances (see Appendix 1)?

Question 21 Answer

When you have established a model of the complex, note the hydrogen bond lengths and other non-covalent interactions. You can either annotate Figure 7 above, insert an annotated image, or list the interactions below.

Labelled Image of figure 7 amino acids	**Image of the structure obtained in Yasara**
	

Hydrogen Bonds	**Other Interactions (apply to both)**
Theoretical Potential of 5 hydrogen bonds in a real DHFR active site The bonds will form at a maximum length of 3.2 angstroms. This is longer than the bond lengths in the appendix which states that hydrogen bonds can only range from 2.9+/- one angstrom  This theoretical number is from a research paper: https://www.researchgate.net/figure/Interactions-between-DHFR-and-the-ligands-a-TMP-and-b-NADPH-H-bonds-distances-below_fig2_226484688 This has an additional hydrogen bond for Leucine 4. In our active site this is replaced by alanine and the distance is too far to form such interactions: only possible at 3.5 angstroms which is too far for hydrogen bonds. **Actual** formation of 4 hydrogen bonds The simplified DHFR active site is slightly different and is ILE residues instead of Leucine and Alanine alongside Aspartate in the active site. The constant residue is Aspartate which will **form 2 hydrogen bonds** with the nitrogen in the ring and electrostatic interactions with the amino group. There will then be **2 more hydrogen bonds** facilitated by both isoleucine residues interacting with hydrogens on the protruding amino group in TMP.	Van der Waals interactions As always, Van der Waals interactions can form between non-polar regions of the molecules and the active sites. In the active site, non-polar regions include the bulky alkyl groups on isoleucine residues and leucine residues. In the TMP the non-polar regions include the carbon chain in the centre of the molecule and with the carbon-oxygen ring at the end. There are also pi-pi stacking non-covelant interactions between the aromatic ring in phenylalanine and the Electrostatic interactions Due to the resonance stabilised carboxylate group in Aspartate, this can create a negative charge. This will be attracted to the positively charged amine group of TMP. The amine group in TMP is positively charged because the hydrogens induce electrons onto themselves from the Nitrogen, giving it a strong partial positive charge.
4 hydrogen bonds formation in the simplified active site of DHFR To achieve 4 hydrogen bonds (yellow), the lengths of the bonds must reach 3.2 angstroms and have a minimum of 2.8 angstroms. This is opposed to the appendix which a range of +/- 0.1 angstroms around 2.9. The hydrophobic nonpolar interactions must also occur at a length of 3-4 angstroms between phenylalanine and the heterocyclic ring in TMP (red). The electrostatic interactions (white) between aspartate and the nitrogen in TMP's heterocyclic ring will occur at 3.5 angstroms, which is higher than the 3.2 angstrom mentioned in the appendix. The non-polar pi-pi stacking interactions between phenylalanine and the ring in TMP will occur at a length of 4 angstroms.	

Explaining the image

This means that hydrogen bonds in the active site of the protein vary much more than within peptide bonds and G/C base pairs in DNA.

In the active site, the hydrogen bonds vary from 2.8 angstroms all the way up to 3.2 angstroms

This also goes for hydrophobic interactions which are usually 3.20 angstroms according to the appendix. However, in the active site, the hydrophobic interactions can go up to 4 angstroms.

This is the same for electrostatic interaction which also go up to 3.5 angstroms within the protein active site. The **electrostatic interaction stabilises and accounts for the long hydrogen bonds between aspartate and the nitrogens in the heterocylic ring of TMP.**

Question 22:

Do you consider that the interactions you have found are sufficient to account for the high affinity binding of TMP to DHFR?

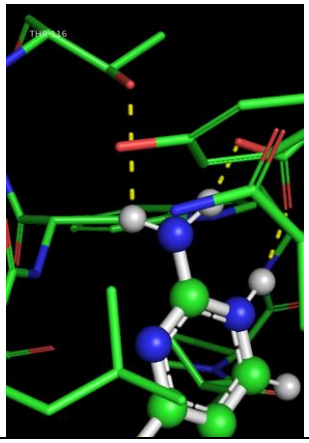
Question 22 Answer

The energy of the TMP on its own is -1037.94 kj/mol and the energy of the TMP in complex configuration is slightly more endothermic at -990.36 kJ/mol KJ/mol. This means there only needs to be a small amount of energy input in order to form the enzyme-substrate complex: ~130kj/mol.

Long hydrogen bonds and van der waals forces are generally quite low energy. There is also Pi-Pi stacking between the phenyalanine and the heterocyclic aromatic ring in the TMP molecule. In addition, the electrostatic interactions between aspartic acid and the amine group account for a huge amount of energy which will certainly cover this energy change the conformation of TMP. Thus these interactions are more than sufficient to account for the small increase in energy of the TMP structure in its less favourable confirmation within the active

There are also other interactions involved to consider that are not present in the active site and further stabilise the TMP. For example, an external threonine residue (threonine 116) also interacts with the TMP and can form another hydrogen bond.

This threonine residue is in the top left of the image on the right.



Question 23:

When the aspartic acid residue in the active site of DHFR was replaced by asparagine in a site-directed mutagenesis experiment, the affinity for TMP was reduced by a factor of 30. Explain how your model of the complex is consistent with this experimental observation.

Question 23 Answer

The aspartic residue is essential for binding of TMP in the active site of DHFR. This is because, as can be seen in question 20 aspartic acid forms 2 essential hydrogen bonds with the Nitrogens in the heterocyclic aromatic ring of the TMP substrate. This is key to position the substrate in the active site on one side.

Furthermore, the aspartic acid residue has a negative resonant structure. This negative resonance strongly attracts the partially positively charged amino group on the heterocyclic ring in TMP. Electrostatic interactions are very strong and thus the removal of this interaction will have a detrimental effect on affinity.

Specifically replacing aspartic acid with asparagine would also introduce an NH2 R group that cannot form hydrogen bonds and would even repel the nitrogens within the substrate's heterocylic aromatic ring due to the partial positive charge on both nitrogens. This would not only lack the strong affinity that exists between asparagine and the ring, but it would also generate repulsion that would significantly reduce affinity for TMP in the DHFR active site.

The structure of pyrimethamine

You will now look at the interaction of other inhibitors with DHFR. Delete both molecules by clicking **FILE**, **NEW**. Now you can build pyrimethamine using the smiles string below:

c1(CC)nc(N)nc(N)c1c2ccc(Cl)cc2

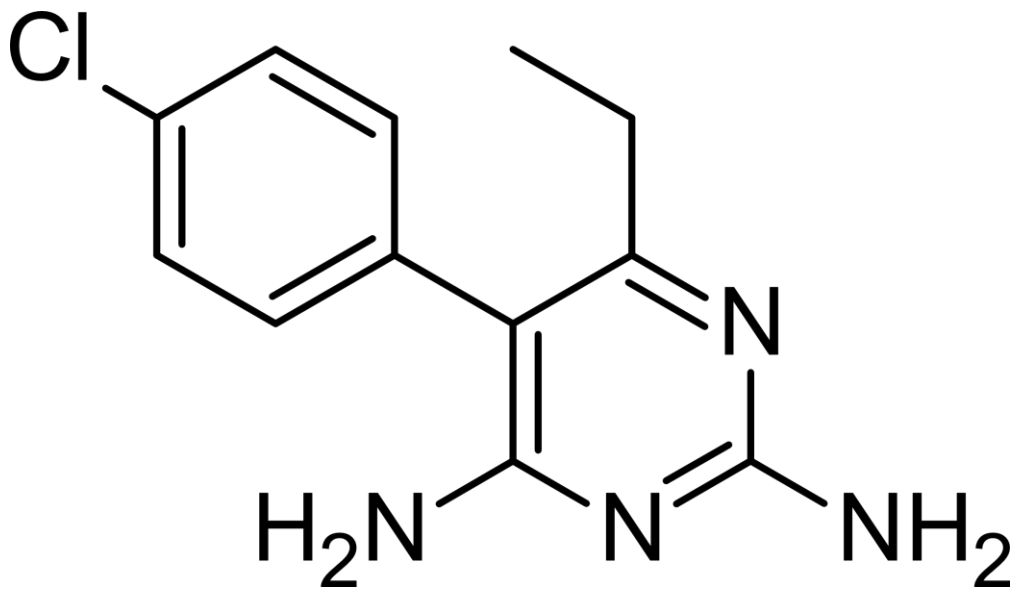


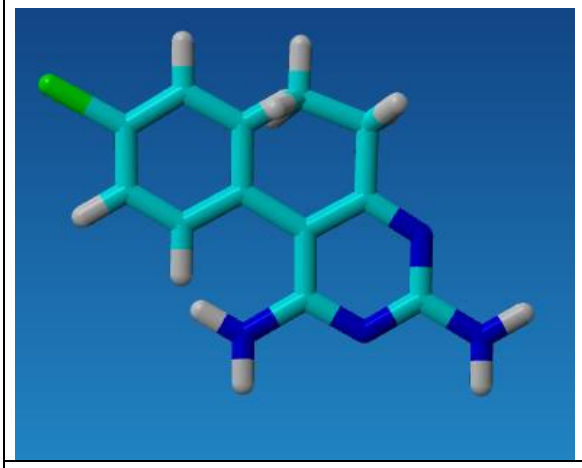
Figure 8: Pyrimethamine

Run the minimiser to find the lowest energy structure, and then save it.

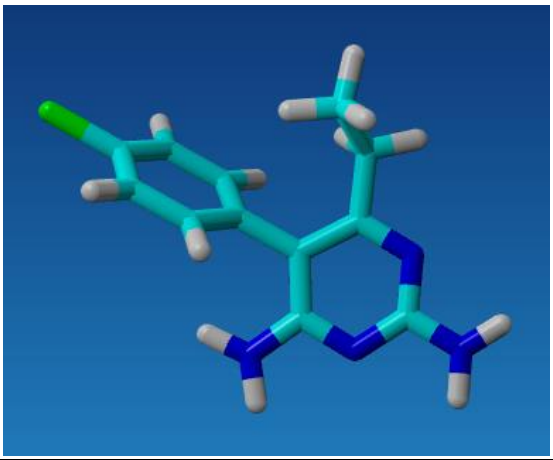
Question 24:
How does your minimised structure differ from the starting structure?
Compare the structures and conformational freedom of TMP and PYR.

Question 24 Answer

Starting Structure
Energy: 3069780252453.30 kJ/mol



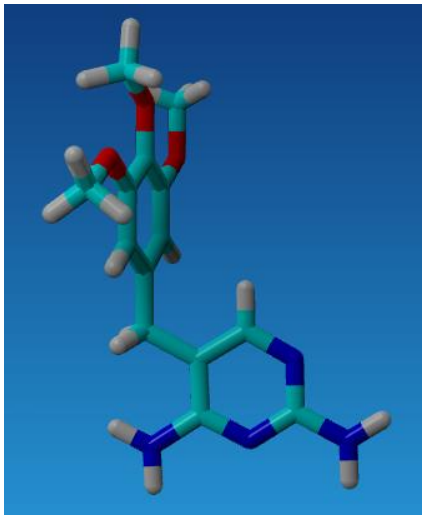
Minimised Structure
Energy: -1413.90 kJ/mol



How does your minimised structure differ from the starting structure?

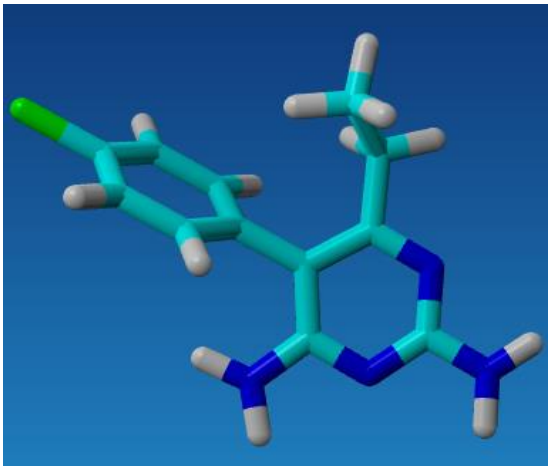
The minimised structure is very different from the starting structure. The aromatic ring top left is rotated to avoid any contact between its hydrogen atoms and hydrogens on the Nitrogens, aswell as the ethyl group attached to the heterocyclic ring bottom left of the molecule. This removes all incidence of steric clash which causes a massive decrease in energy to a highly exothermic state that is very favourable to form. The alkyl group is also rotated around 90 degrees to point towards the screen. The chlorine remains facing away from the heterocyclic nitrogen ring. This somewhat of a dipole in the molecule as a whole: the chlorine side is negative and the amino groups protruding on the ring side are partially positive.

TMP Structure and Conformational Freedom



-1058.91 kJ/mol

PYR structure and conformational freedom



-1413.90 kJ/mol

Compare the structures and conformational freedom of TMP and PYR.

The structures are very similar: they both have 2 rings connected via carbon-carbon bonds. TMP is slightly different mainly because there is an additional carbon in the bridge connecting the two carbon rings and certain groups replaced.

The energies between the two comformations above is not comparabale as the TMP structure is its conformation in the DHFR active site, not its minimised structure like PYR. The minimised structure of TMP has an energy of **-1354.90 kJ/mol** (question 18). This is very similar to the energy of the PYR minimised structure: -**1413.90 kJ/mol**. Thus the molecules have a very similar energy and are both equally exothermic.

The additional carbon in TMP means there is less steric clash between hydrogens protruding from the two rings. The bulky alkyl group on PYR restricts the movement of the aromatic ring to a comformation as free as TMP. This bridge in TMP allows more rotation as it is a central carbon 'pivot' that increases the conformational potential and rotation between the two aromatic groups in the molecule.

Another key difference is that in TMP one ring has oxygen and another ring has Nitrogen: this means the rings are electrostatically attracted to eachother and they try and folder onto eachother to form hydrogen bonds, however the rotation around the long carbon bridge between them restricts this and prevents it from happening.

PYR also has a heterocyclic Nitrogen carbon ring but the other ring is a carbon ring with chlorine attached at a position the furthest away from the heterocyclic nitrogen ring. Chlorine is less electronegative than the oxygens in TMP and it faces away from the heterocyclic Nitrogen carbon ring. The CL ring's position in space is almost adjacent to the heterocyclic nitrogen ring.

PYR's additional ethyl group attached to its heterocyclic Nitrogen ring is repelled by the hydrogen groups on the CL carbon ring and thus moves away from that ring in the minimised structure. On the other hand, TMP has a hydroxyl group in the same place which can be invaluable for forming hydrogen bonds.

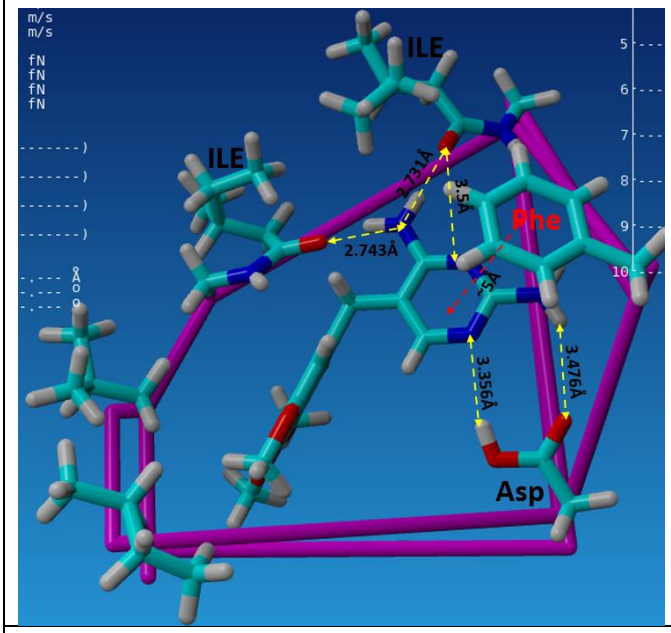
Overall TMP and PYR minimised structures have very similar energies, owed to their very similar structure. However, the conformational freedom of TMP is slightly better due to a carbon 'bridge' between the two rings that can act as a pivot, in addition to no bulky alkyl groups that can cause steric clash in certain positions. The presence of more hydroxyl groups and carboxyl groups also means TMP can be stabilised more easily by polar interactions and thus can exist in higher energy states than PYR.

You can now dock pyrimethamine into the active site of DHFR as you did for TMP. (Read in the colisite.sce YASARA scene file, followed by your saved minimised pyrimethamine).

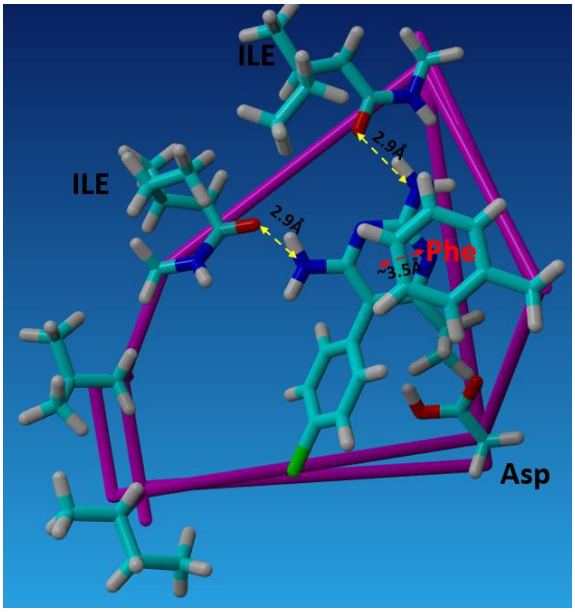
Question 25:
Can you dock pyrimethamine in the same way as you did TMP?

Question 25 Answer

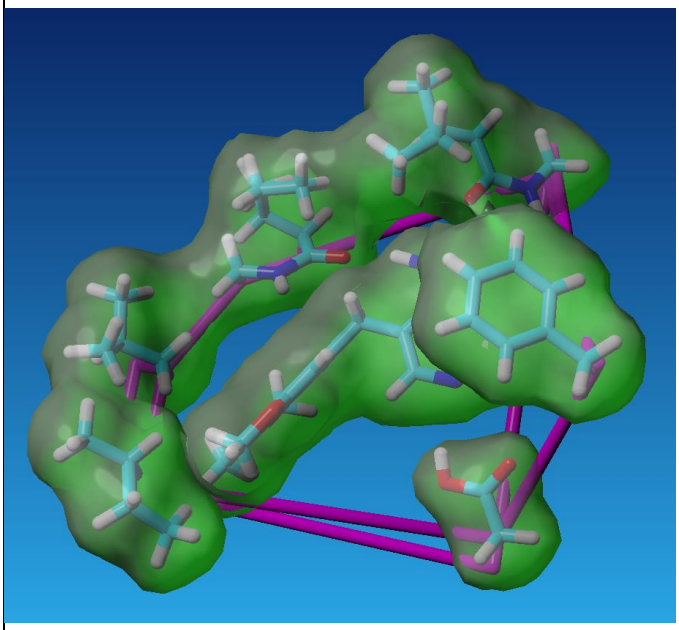
TMP docked in DHFR



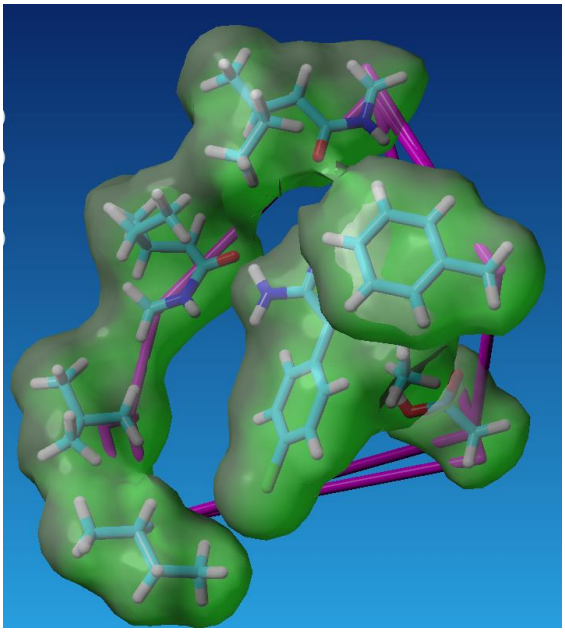
PYR docked in DHFR like TMP



Surface Representation TMP



Surface Representation PYR



Can you dock pyrimethamine in the same way as you did TMP?

No

Since the bottom groups are the same, most of the interactions are retained from TMP to PYR. This includes 2 of the hydrgeon bonds between the heterocyclic nitrogen carbon ring and the active site of DHFR and the pi-pi stacking interactions between the rings. As a result, this region of the substrate PYR is wedged quite into the DHFR active site.

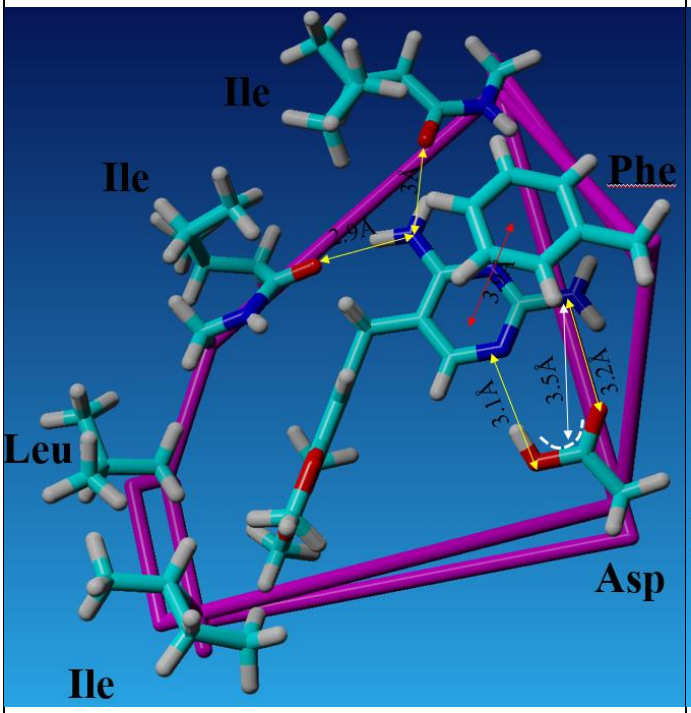
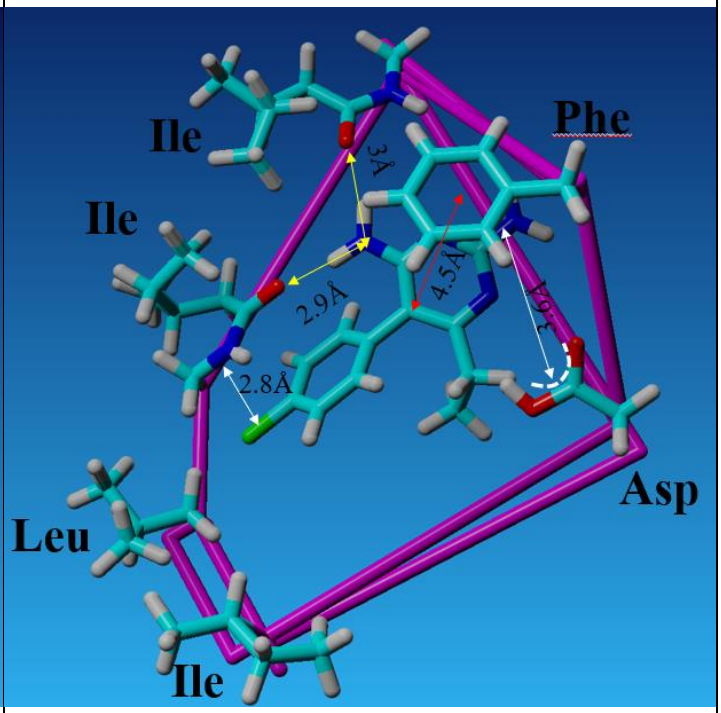
However, in this specific 'minimised' comformation of PYR it is impossible to dock the exact same way as TMP. The conformational freedom of the PYR is slightly less that TMP and thus the rings are not perpendicular around the C-C bond, as they are in TMP. This means that the molecule is more slanted in the active site and some regions cannot form hydrogen bonds. This means, in order to form hydrogen bonds with ILE residues the PYR molecule is positioned in a way where it **alkly group clashes with the aspartate residue on the bottom right** of the active site image. This can be seen in the surface representation where the surfaces overlap between the active site and the PYR molecule.

In order for PYR to dock and form similar interactions to TMP, the molecule has to be rotated **clockwise** to expose its amino group to the same interactions. This also removes the steric clash between the alkyl group and the aspartate residue on the bottom right of the active site.

Question 26:
What do you think are the relative affinities (qualitatively) of PYR and TMP for DHFR?

Question 26 Answer

Images of the PYR and TMP in the active sites for comparison
White: Electrostatic
Yellow: Hydrogen bonds
Red: Pi-Pi

TMP	PYR
	

What do you think are the relative affinities (qualitatively) of PYR and TMP for DHFR?

The main difference that affects affinity in PYR is the **chlorine group protruding from the carbon ring instead of the oxygen groups within the carbon ring in TMP**. Since the oxygen groups in TMP heterocyclic ring are ether groups and wedged between two carbon groups they don't form any hydrogen bonds in TMP. Thus they cannot form hydrogen bonds with the hydrogens protruding from isoleucine and leucine in the active site. However, PYR has a Cl group instead which extends from the carbon ring. Thus, this highly electronegative group is exposed and free to form hydrogen bonds with the hydrogens via its strong partial negative charge. It also has the potential to form strong electrostatic interactions with any basic residues in this region of the active site. In particular, it can form electrostatic interactions with the nitrogen in the ILE residue nearby it as this nitrogen is bound to an oxygen which will give it a partial positive charge. This means **both sides of the molecule can interact strongly** with the active site.

However, PYR cannot form hydrogen bonds between its amino group and nitrogen in the ring with the aspartate residue in the active site as the distance is too long. This means it loses 2 hydrogen bonds relative to TMP. The loss of these hydrogen bonds will decrease some interactions, however the distance of 3.6 angstroms between nitrogen and aspartate is still enough to facilitate strong electrostatic interactions which are more important.

The phenylalanine residue in PYR is slightly further away which will decrease the strength of pi pi stacking between the two rings but there will still be some interaction as these can reach 5 angstroms.

Overall, PYR forms less hydrogen bonds than TMP but it has stronger interactions and interacts with the active site strongly on both sides of the molecule, when in a conformation that facilitates binding. The most important point is the addition of the Cl group which allows anchoring of PYR on **both sides** of the molecule. This means PYR is very similar to TMP in terms of affinity but it is slightly better.

References

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7102388/

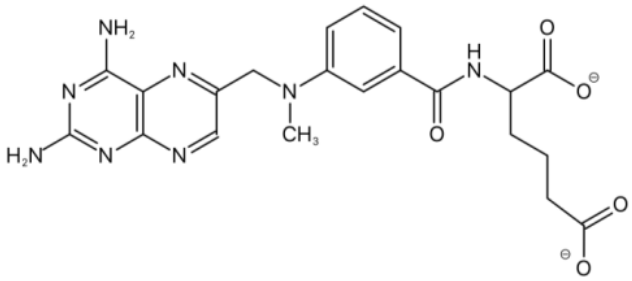
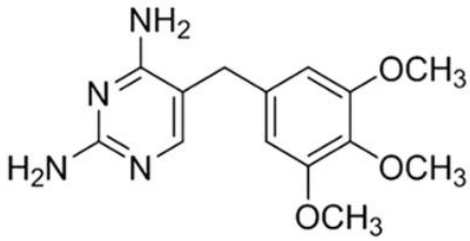
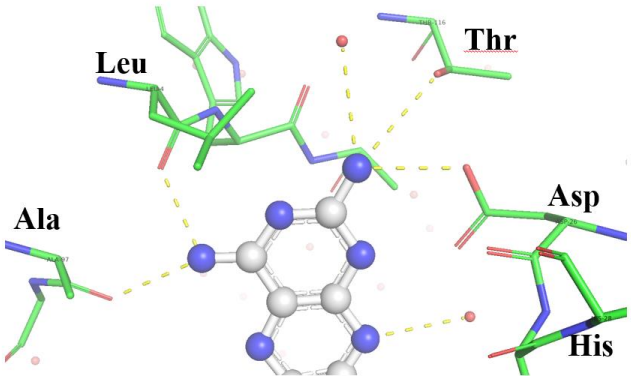
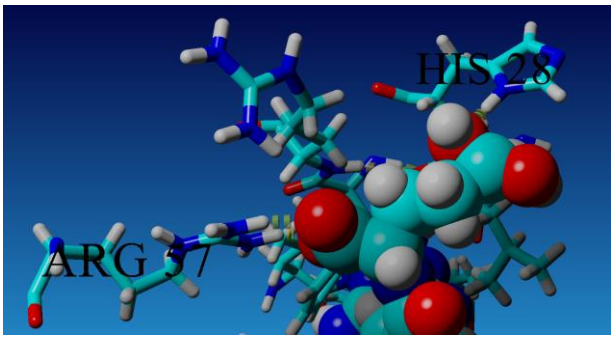
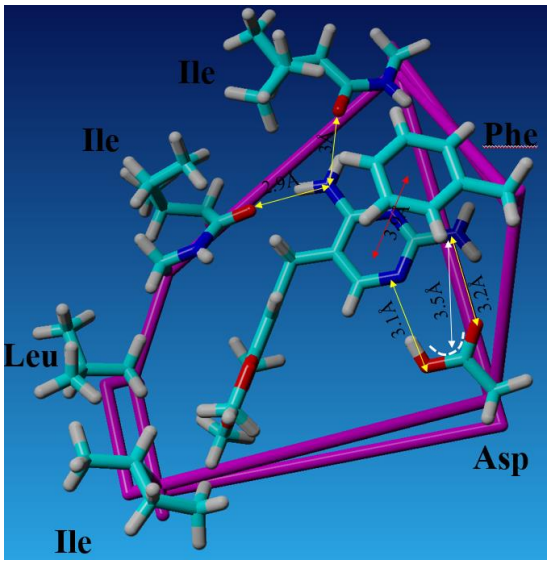
Methotrexate-DHFR interaction

Methotrexate (MTX) is a direct analogue of the natural folate substrate (see Appendix 2) and binds to DHFR $2x10^3$ times more tightly than folate. (The binding constants for MTX and folate are $2x10^8$ M^{-1} and $1x10^5$ M^{-1} respectively). Note that MTX is a very flexible molecule. Compare your docked TMP structure with that of the DHFR-MTX complex.

Question 27:

Does MTX bind to DHFR in a similar mode to that of TMP? (Look back at the structure of MTX bound to DHFR that you viewed for Q.16)

Question 27 Answer

Methotrexate Structure	**TMP Structure**
Figure 12: Methotrexate 	 **Trimethoprim**
Methotrexate bound to DHFR	**TMP bound to DFHR**
Heterocyclic ring:  Glutamic acid tail: 	 In the TMP simplified active site, Leu and alanine are replaced by isoleucine.
Does MTX bind to DHFR in a similar mode to that of TMP? (Look back at the structure of MTX bound to DHFR that you viewed for Q.16) *What interactions does this acidic tail of the molecule make with DHFR?*	

Is it possible for the site to accommodate the glutamic acid tail of MTX?
Overall, MTX certainly binds similarly but much more strongly than TMP and has a higher affinity for the DHFR active site. This is largely owed to its larger molecule size that can form more interactions.

MTX is a classical antifolate whereas TMP is a nonclassical antifolate. Methotrexate is a very similar substrate to TMP: It also has the heterocyclic nitrogen carbon ring which forms hydrogen bonds with the active site of DHFR. However, TMP is a smaller molecule and has one less ring in the heterocyclic nitrogen ring system. In addition, on the other side of the molecule, methotrexate has exposed carboxyl groups whereas TMP has ether groups which are far less electronegative and are not involved in hydrogen bonding. A peptide bond in the middle of methotrexate also increases the interaction potential. Overall, methotrexate is a much bigger molecule that can form many more interactions than TMP.

The first discrepancy is hydrogen bond count. Methotrexate can form many hydrogen bonds within the active site of DHFR, totalling 8 hydrogen bonds: 6 from the heterocyclic nitrogen ring and 2 between the conjugated carboxylate groups on the other end of the molecule (its glutamic acidic tail) with basic residues like Histidine 28 and Arginine 57. On the other hand, TMP can form a maximum of 4 hydrogen bonds: all facilitated by the nitrogen atoms in the heterocyclic nitrogen carbon ring. Both molecules have electrostatic interactions (aswell as hydrogen bonds) between the amine group and the aspartate residue of the active site.

Furthermore, the carboxylate groups in the glutamic acid tail of MTX exhibit resonance and are negatively charged in the ring. This means they can form strong electrostatic interactions with the amine groups in the Histidine and Arginine residues aswell. This creates a very strong anchorage of the acidic tail inside DHFR.

The second discrepancy are non-polar interactions. MTX has 3 aromatic rings in its structure whereas TMP has only two. Furthermore, two of the rings are joined in MTX and create a bigger conjugated electron system. This means that MTX forms far stronger pi-pi stacking and hydrophobic interactions that TMP.

The hydrogen bonds on both sides of the methotrexate molecule means it is anchored strongly into the active site at two positions (on both ends of the molecule). TMP is only strongly anchored on the side with the heterocyclic nitrogen ring: the other side is held in the active site by weak induced dipole interactions between the ether groups and the active site's non-polar residues.

The additional nitrogen in the middle of the MTX compared to a carbon in the middle of the TMP is quite significant. This nitrogen in MTX has a methyl group attached which is electron inducing. This means that the nitrogen is partially positively charged and can form electrostatic interactions with residues near it, like the conjugated, electron dense ring of phenylalanine 49. This methyl group can also form further induced dipole forces.

Finally the peptide bond in the middle of MTX facilitates two extra hydrogen bonds with water in the active site via the exposed carbonyl group. This further anchors MTX into the active site by the middle of the molecule, unlike TMP which is only anchored in the middle by induced dipole interactions between its carbon chain and non-polar active site residues.

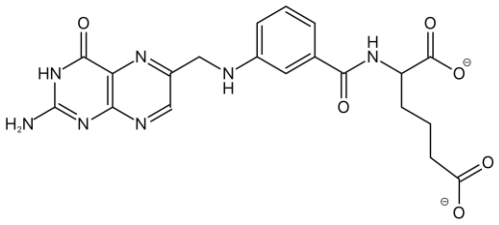
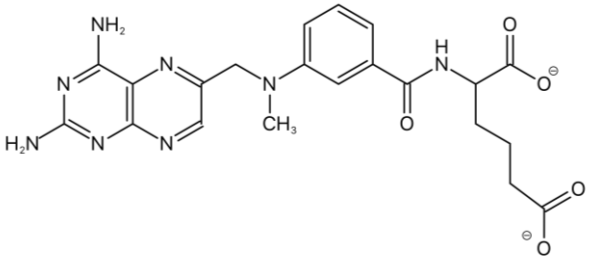
References

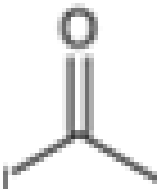
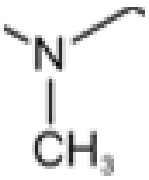
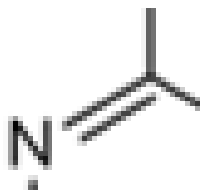
[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7102388/

Question 28:

Inspect and compare the structures of folate and MTX in Appendix 2 (Figures 9 and 12). If we assume that MTX and folate bind in the same way to DHFR, what is the difference, in terms of the number of H-bonds, van der Waals interactions etc. between the binding of folate to DHFR and the binding of MTX?

Question 28 Answer

Folate	**Methotrexate (MTX)**
	

Difference 1	**Difference 2**	**Difference 3**
Folate 	Folate 	Folate 
Mtx 	Mtx 	Mtx 

Inspect and compare the structures of folate and MTX in Appendix 2 (Figures 9 and 12). If we assume that MTX and folate bind in the same way to DHFR, what is the difference, in terms of the number of H-bonds, van der Waals interactions etc. between the binding of folate to DHFR and the binding of MTX?

Context

Folate and methotrexate have very similar structures. They both have the same organisation of structure: a double phenyl ring on the left and a long carbon chain on the right, connected by a nitrogen bridge. Their differences lie in how they fit into the active site: folate is planar and does not fit into the DHFR active site as well as the 3 dimensional MTX conformation. Dihydroflate, not folate, is the substrate that MTX competes with for DHFR.

Differences between folate and MTX

They **first key difference** is a hydrogen on the nitrogen bridge in the middle of folate as opposed to a methyl group on the nitrogen bridge in the middle of methotrexate. This may make a small increase in polarity in the middle of the molecule for Folate as it's secondary nitrogen group with the hydrogen can form hydrogen bonding whereas the tertiary nitrogen in MTX will be completely non polar and blocked by carbon atoms. This creates 1 additional hydrogen bonding opportunity for folate that MTX did not have, whether it be with a water molecule or with a residue in the active site. However, in the **active site** this removes the dipole-dipole interaction between the CH3 on the nitrogen bridge and the Leucine 21 (due to the substitution of CH3 for hydrogen at this position).

The **second key difference** is in the double phenyl ring on the left. Folate has one less nitrogen-carbon double bond in the ring. This will affect the electron delocalised system in the left side of the ring that could change van der waals interactions; however, there is the addition of a hydrogen on the Nitrogen in the ring which can form hydrogen bonds. This creates an 1 additional hydrogen bond in this location that MTX did not have. However, this may slightly affect the induced dipole interaction between leucine 55 of **the active site** and the ring delocalised system.

The **third key difference** is the carbonyl group in the Folate ring instead of an amine group in Methotrexate. The amine group here in MTX was able to form 3 hydrogen bonds with 2 different residues (Asp and Thr) and a water molecule. The replacement of this group with a carbonyl group in Folate can only facilitate 1 hydrogen bond with a water molecule. This removes 2 hydrogen bonds. This would have a negative impact on affinity to the DHFR **active site**. Furthermore, this amine group is electrostatically positive in MTX due to electon induction from the carbon double bond in the phenyl ring and interacts with the negative aspartate residue in the active site with strong electrostatic interactions. Thus, the removal of this amine group in folate also removes strong electrostatic interactions.

As previously established, MTX can form **8 hydrogen bonds** overall (6 from the heterocyclic ring and two from the glutamic acid tail). Despite the differences, folate can also form 8 hydrogen bonds (5 from the heterocyclic ring, one from the tertiary amine in the centre and two from the glutamic acid tail). However, the rearrangement of hydrogen bonds in folate has negated additional non covalent interactions that were facilitated by MTX: strong electrostatic interactions between aspartate and the amino group, induced dipole interactions between the methyl group on the nitrogen bridge and leucine in the active site and finally weaker pi pi stacking interactions due to the removal of a double bond in the heterocyclic nitrogen carbon aromatic ring.

The loss of DHF affinity and catalytic efficiency is generally smaller than the loss of MTX affinity when amino acids of the active site are substituted. This is often attributed to formation of different contacts with either ligand due to the 180° inversion of the pterin ring of bound DHF relative to MTX.

References

[1] https://www.jbc.org/article/S0021-9258(17)49292-1/fulltext
[2] *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2738603/*

Question 29:

The difference in K_a between folate and MTX (a factor of $2x10^3$) can be converted into a difference in the free energy of binding by means of the relation

$$\Delta G = -RT \ln K_a \qquad (R = 8\ J\ K^{-1}\ mol^{-1} \text{ or } 2\ cal\ K^{-1}\ mol^{-1})$$

(If you want to convert from ln (log to base e) to $\log_{10}$ (log to base 10) then:

$$\ln K_a = 2.3 \times \log_{10} K_a).$$

Calculate the difference between the ΔG values for the binding of folate and the binding of MTX. (Use a value of T = 310K). Do you consider that this difference in the free energy of binding can be accounted for entirely by the different non-covalent interactions that you have identified? If not, what other factors might affect the relative affinities?

Answer:

As mentioned earlier, the binding constants for MTX and folate are $2x10^8\ M^{-1}$ and $1x10^5\ M^{-1}$ respectively This means the MTX binds to DHFR 2x103 times more tightly than folate.

MTX-DHFR = $-RT\ln K_{a\ MTX\text{-}DHFR} = -8 \times 310 \times \ln(2x10^8) = -47{,}402$ J/mol = -47.402 kJ/mol

Folate-DHFR = $-RT\ln K_{a\ Folate\text{-}DHFR} = -8 \times 310 \times \ln(2x10^5) = -28{,}552$ J/mol = -28.552 kJ/mol

Free energy for binding folate to DHFR is almost half as MTX representing a significant difference in affinity. This cannot be explained by hydrogen bond count, as they are the same, but mainly due to lack of electrostatic interactions and van der waals forces in folate compared to MTX. However, such a big difference in energy cannot just be explained by interactions and the folding of the molecule in the active site is also very important: folate is a planar molecule and MTX is not which means MTX has a more snug fit into the active site of the DHFR.

https://www.nature.com/articles/srep15328

Question 30:

Now build on your experience of analysing protein structures and write a report of no more than 2000 words describing a protein structure or a protein/protein, protein/substrate or protein/nucleic acid complex of your choice. In addition to a description of the structure, the focus of the report should be on what the structure tells you about its function. Keep these points in mind as you prepare the report:

1. You can freely select your protein structure of choice from the PDB under two conditions:
    a. You must pick an *experimental* structure determined by X-ray crystallography, NMR spectroscopy or cryo electron microscopy. If there are several identical structures of the same protein use your judgement as to which one to use. Note that the same protein in complex with different ligands may not necessarily have the same structure.
    b. There should be an accompanying article in the PDB entry. If there is no publication associated with a structure it possibly means it is not that important.
2. Context to the structure analysis in the report should be taken from a total of about twelve articles, including the one coming with the PDB entry. They should be properly cited and referenced using any conventional but consistent style.
3. Illustrate the report with at least 4 figures that you prepare in Yasara (or PyMol if you prefer). Ensure that you choose carefully the way in which you depict the protein to support the point you want to highlight. For example: cartoon view of whole protein is useful for overview of structure but useless to show details of protein ligand interaction. The latter would be illustrated by zooming in on a small portion of the protein with sidechains and important contacts shown. It may be helpful to add other figures to support the text, either taken from the literature (if properly referenced and acknowledged) or home-made.
4. This must be entirely your own work.
5. **IMPORTANT: This question carries 60% of the marks for the practical report.**

Human Serum Albumin: *A Paradox of Structural Simplicity and Functional Complexity*

What is Human Serum Albumin?

Human serum albumin (HSA) is globular protein of 585 amino acids and 66,438Da. It is encoded by a single copy gene ALB gene on the long arm of homo sapiens chromosome 4. Due to the addition of an N-terminus pro-peptide (6 residues) and signal peptide (18 residues) the translated protein is 609 amino acids, until processing in the hepatocyte cytoplasm [1]. HSA is a member of a protein family of homologous proteins with an extraordinary ligand binding capacity and distinctive structural features [2]. This facilitates the protein as a depot and carrier for many endogenous and exogenous compounds [1]. HSA's primary function is as the main determinant of plasma oncotic pressure and modular of fluid distribution between body compartments. Its secondary function is to bind ligands like water, zinc, calcium, sodium, potassium, fatty acids, hormones, bilirubin and drugs; facilitated by its unique and intricate structure. It is the most abundant protein in human blood plasma, contributing to around half of serum protein [3].

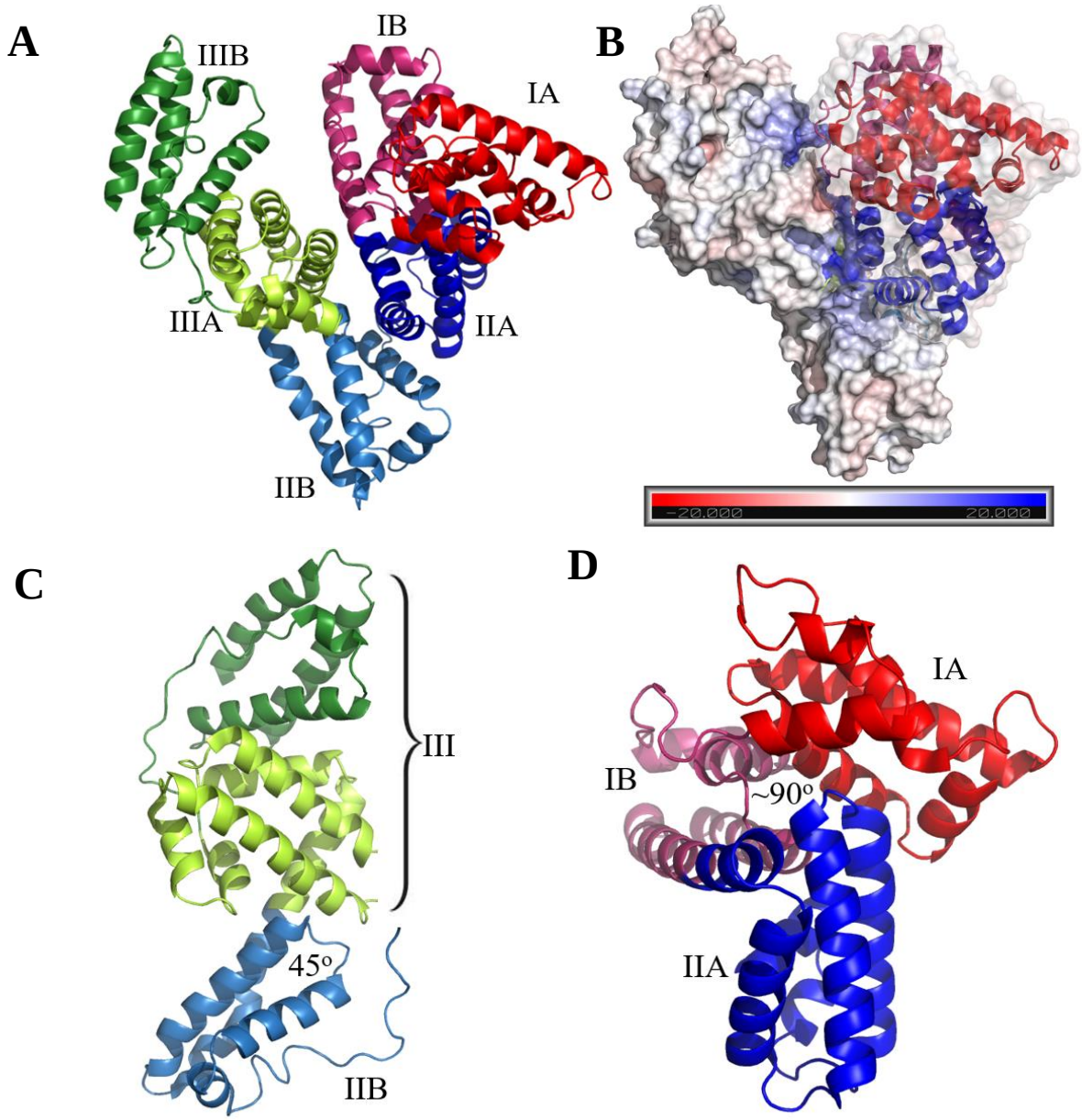


Figure 1: Cartoon structure of HSA made in Pymol [4] A) This image depicts the 6 subdomains within the 3 domains of HSA. The first domain is coloured in shades of red, second in shades of blue and third in shades of green. **B)** This image was created using an **APBS plugin** for pymol which depicts the electrostatics of HSA's surface, colour coded according to the legend below. **C)** This is the Y junction between III and IIB discussed on the next page. The angle is shown as ~45 degrees **D)** This is the 'T' configuration discussed on the next page between I and IIB. The angle is 90 degrees (perpendicular).

Human Serum Albumin General Structure

The secondary structure of HSA is 68% alpha helices and has no B sheet element. It contains three homologous domains indicated as I (1-195), II (196-383) and III (384-585) (**fig.1a**). However, the structure is not so simple as each domain contains two separate subdomains (a and b) of ten helices: six (h1-h6) and four (h7-h10), respectively, connected by a long-extended loop (fig1b). The total number of helices is, however 28 (not 30) because the last helices of domains I and II are fused to the first helices of the next domain [4]. The structure is held together by 17 pairs of intramolecular disulphide bridges but there is one free cysteine residue, Cys34, which is the only thiol group in the protein that does not take part [2]. The cysteine bonds are essential to curve the subdomains and hold the structure together to cluster aliphatic and aromatic side chains in the inner surface to form each subdomain's hydrophobic core. This topological feature is seen in every subdomain except IA-helix 3: the lone Cys34 resides between IA-helix 2 and IA-helix 3 [4]. This cysteine can form intermolecular disulphide linkages and it is oxidised by cysteine/glutathione in 30-40% of blood HSA molecules [1].

The primary factor which facilitates different active site affinities is the positioning of the subdomains relative to each other: despite structural similarity, the domains interact with neighbour domains in different ways which creates a highly asymmetric environment with a variety of binding sites. A perpendicular T-Shaped arrangement is formed between subdomain IIA and the interface region between subdomains IA and IB, facilitated by hydrophobic interactions and hydrogen bonds (**fig.1d**). A unique Y-Shaped assembly between domains II and III is formed via a 45^{o} protrusion of domain III from subdomain IIB (**fig.1c**). The relative motions of these domain structures imply flexibility of HSA in different conditions [4].

Human serum albumin has a long lifetime of 28-36 days that can introduce structural modifications that affect ligand binding and anti-oxidant properties. In addition to a higher proportion of acidic to basic residues, HSA's primary structure is predominantly composed of ionized residues which give it a net charge of around -15 and an acidic isoelectric point (pI) of 4.7 at physiological PH7 (**fig.1b**). This is essential in its function and regulation in the blood [5]. This high number of acidic and basic residues allows HSA structure to undergo reversible transitions at different pH values: an extended (E) conformation below 2.7, a fast-migrating (F) form between 2.7 and 4.3, the neutral (N) form between pH 4.3 and 8.0 and finally a basic (B) form at a greater pH, characterised by the loss of alpha helix and increased in affinity for some ligands [6].

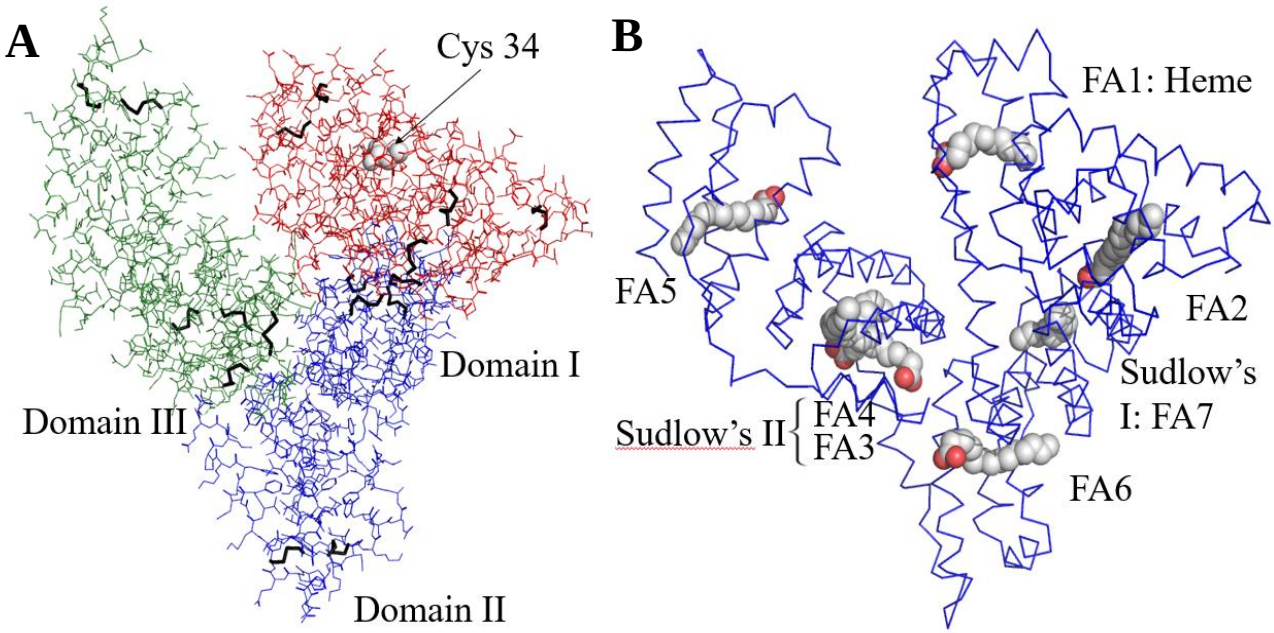


Figure 2: Ribbon structure of HSA made in Pymol [4][8]. A) This image depicts all 17 disulphide bond pairs as black lines within the HSA structure. There are, however, 35 cysteine residues in HSA and the free Cys34 position is depicted as 'ball and stick; and labelled. [A] **B)** This image depicts all 7 fatty acid binding sites in HSA. Sites 3 and 4 are grouped as 'Sudlow's Site I' and site 7 is 'Sudlows Site II'. The palmitic fatty acid moieties are depicted as ball and stick, coloured by element [B].

Function of Human Serum Albumin

Aliphatic Fatty Acids – Eggs, Nuts, Fish, Milk

An essential function of Human Serum Albumin is to carry substrates in the blood and its structure is specifically designed for essential fatty acids ligands (**fig.3**). Palmitic acid is a hexadecenoic acid with a 16-carbon chain; it is the most common saturated fatty acid found in animals, plants and microorganisms and makes up 44% of total fats. The palmitate anion is the observed from of palmitic acid at physiological pH (7.4) [7]. There are 7 binding sites on the molecule occupied in different ways by fatty acids, facilitated by folding and arrangement of the subdomains to create environments of high affinity (**fig.2b**). Sites 1-5 have well placed clusters of amino acids which co-ordinate the binding of the fatty acid and anchor its carboxylate moiety by electrostatic interactions to basic and polar side-chains. A notable difference in these 5 sites is how the fatty acid is folded in the site: the longer the methylene tail of the fatty acid, the stronger electrostatic and hydrophobic interactions with the pocket. Palmitic acid is a 'long fatty acid' (>16C) and binds strongly. Sites 6 and 7 are 'lower affinity' [8].

The interaction homology of the first five sites can be seen by the binding of palmitic acid in the most open and accessible site, site I within subdomain IB (**fig.3a**). Fatty acids in these sites bind in the same orientation with the carboxylate group hydrogen-bonded to a water molecule. In site 1, this is coordinated by the side chain hydroxyl of Tyr161 and carboxyl oxygen of Leu182 which a water molecule in place that interacts and anchors the fatty acid carboxyl group [8]. Arg117 also contributes to hydrogen bonding interactions with the palmitic acid carboxylate moiety [4]. For longer chain fatty acids, like palmitic acid, the hydrophobic tail curls around the inside surface of the cavity toward His146 at the lower end of the cavity opening [8]. The distance of 11.3Å can be reduced quickly to the ~5Å maximum distance of Van der Waals interactions by the folding of the fatty acid chain which is 43.7Å in length (**fig.3b**) [9]. Site 2 is between IIA and IIB and similar but it is one of the most enclosed fatty acid binding sites on HSA and long fatty acid chains extend further into the pocket. Site 3 is also similar but broader running from subdomain IIIA to IIB: long fatty acid chains fold into a U-bend configuration. Site 4 is narrow and long and longer chains just extend further into the hydrophobic tunnel of subdomain IIIA; site 5 is the exact same, running through sub-domain IIIB [8].

The unique structure of site 6 is located in a shallow trench on the surface of the protein, between IIA and IIB (**fig.3c**). It lacks a basic/polar amino acid clusters to anchor the carboxylate moiety as in sites 1-5. The ligand density is stronger on the IIB side, including side chains Arg209, Lys351 and Ser480, which transiently interact with the carboxyl group in a variety of positions, with no conserved electrostatic interaction. However, the middle portion of the methylene tail is well anchored via a said chain 'strap' holding it in position, formed between Arg209 to both Asp324 and Glu354 (**fig.3c**)[8]. Site 7 is significantly smaller and has been hypothesised as a binding site specific for shorter chain fatty acids: this is supported by evidence which tried to covalently modify lys220 in site 7 subdomain IIa as this was blocked by small chain fatty acids but not long chain fatty acids (>16C), confirming binding [10].

Phenolic Acids – Fruit, Vegetables, Cereals

Many plant-based foods introduce phenolic acids into the diet; these are not aliphatic and have a phenolic ring in their structure. Gallic acid (trihydroxybenzoic acid) has antioxidant effects and low toxicity; it is a hydrolysed tannin which primarily enters the body from tea and fruits like strawberries, grapes and bananas [11]. Gallic acid binds to serum albumin in the blood after intestinal absorption and which introduces the importance of the unusual single tryptophan residue, Trp214, within the structure of HSA (most proteins have more) (**fig.4a**) [1]. This tryptophan residue helps position the gallic acid moeity in subdomain IIA of HSA to a position ~10.4Å away where it interacts with hydrophobic residues concentrated in the pocket at a distance of ~3Å (**fig4**). This is because the nitrogen of the tryptophan is attracted to the hydroxyl group of gallic acid. Within the pocket, gallic acid forms many hydrogen bonds between its protruding hydroxyl groups and nitrogen of basic amino acids which shield the gallic acid such as Arg257 and Arg222. The tyrosine 150 in the pocket is particularly important as it has an aromatic component which can form pi-pi stacking hydrophobic interactions with the aromatic ring in the gallic acid moiety. These hydrophobic interactions can also extend from gallic acid to Ala261 and Leu219/238 via induced Van der Waals forces with their methyl groups. The shielding of gallic acid by the pocket decreases its absorbance spectra [11].

Figure 3: Active Site representation of HSA made in *PyMol* [8]. A) Active site of fatty acid binding site 1, within four helices at the centre of subdomain IB. An additional Arg117 is located in a crevice formed by the helix cluster of subdomain IB and the extended loop between subdomains IA and IB. The water molecule is pink and polar interactions are shown by dashed yellow lines. **B)** The 11.3A distance depicted between the bottom His146 residue and the PLM tail. Histidine is often important in catalytic triads and His146 is the gate to the cavity. **C) Circled:** The salt 'strap' formed by carboxyl groups on Asp324 and Glu341 with the nitrogen from Arg209 is circled. This holds the palmitic acid in place.

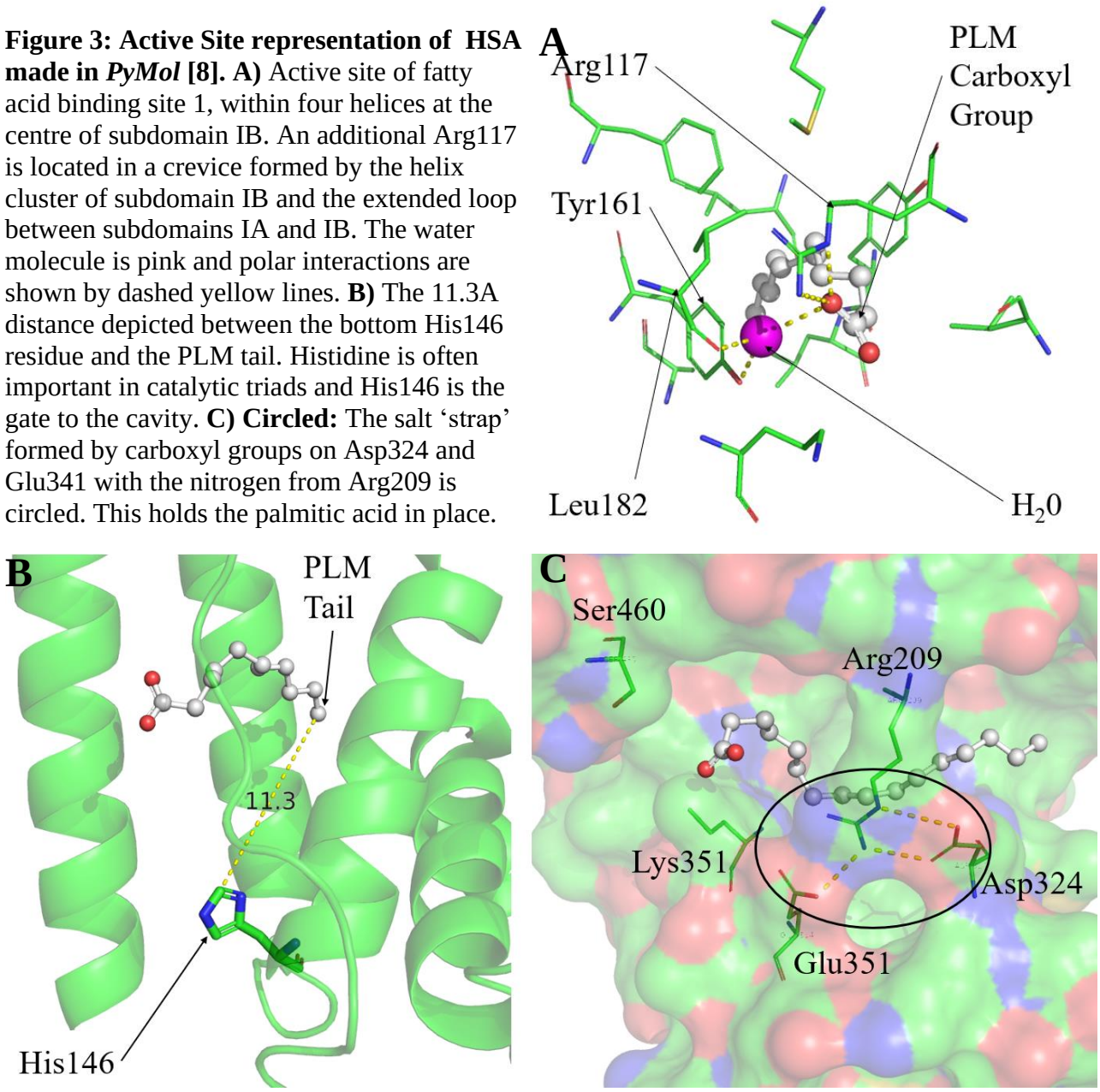
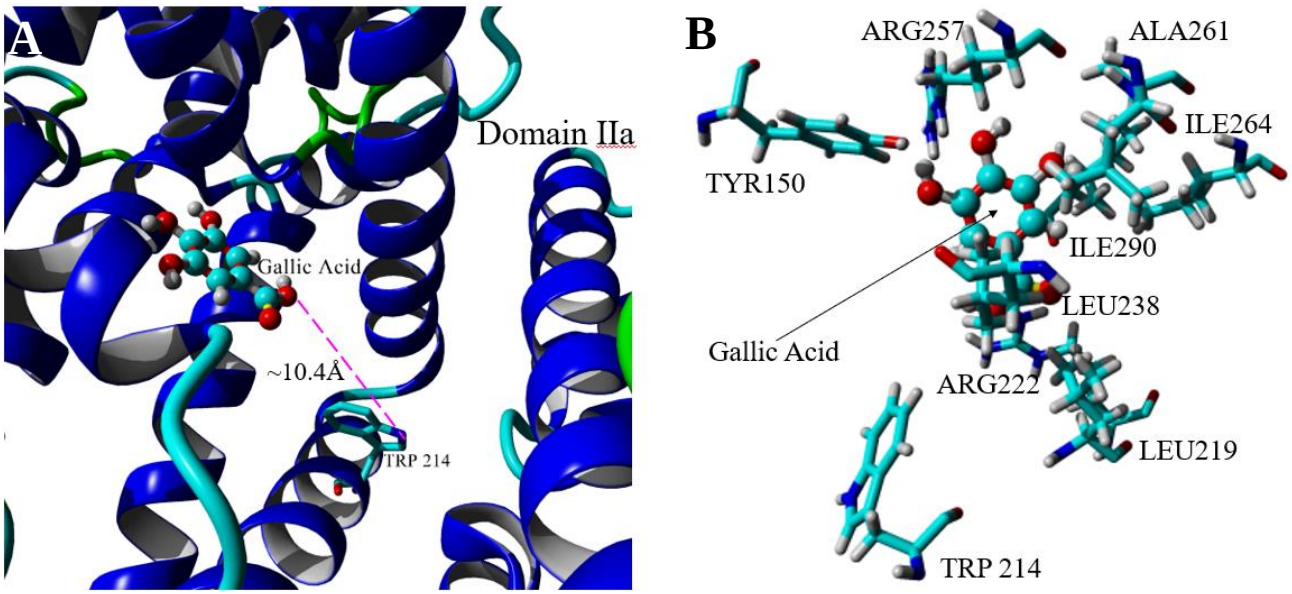


Figure 4: Active site representation of HSA made in *Yasara* [4][11]. A) A combined structure docking a ligand free crystallised pdb of HSA with a gallic acid molecule built from a smile string calculated using its structure as C1=C(C=C(C(=C1O)O)O)C(=O)O [11]. The unique single TRP 214 residue is labelled and its distance between its nitrogen and the hydroxyl of gallic acid is marked in pink. **B)** The active site of HSA (FA6) within subdomain IIA. Relevant amino acids are labelled.



Anion Binding of Human Serum Albumin

HSA is an essential transition metal ion transporter in plasma: it binds approximately 45% of circulating calcium, 45% of circulating magnesium and 80% of circulating zinc. It has the highest affinity to zinc, followed by calcium and then magnesium. The binding is facilitated by potentially more than 2 calcium binding sites; there is a shared binding site between zinc and calcium at Asp273 which also suggests a crosstalk between zinc and calcium transport in the blood [3]. HSA was long described as a 'sponge' which binds all and everything non-specifically. However, in the last decades, along with regulatory roles of the protein, specific binding sites with specific metal preferences were inferred from spectroscopic techniques and limited analysis of the structural amino acid environment. There are four metal binding sites in most mammalian albumins including HSA and BSA: the N-terminal site (NTS), Cys34 residue (and surrounding environment), Site A (MBS – multi metal binding site) and Site B (unknown location) (**fig.5a**) [12].

The **NTS** site is the hallmark of albumin metal binding. It is composed of the first three Asp-Ala-His residues in the albumin sequence (hence N-terminal). These make up four donor atoms involved in coordination which prefer metal ions capable of a square-planar formation like Cu(II) and Ni(II): NTS bound to HSA is the second largest pool of copper in human blood serum (after ceruloplasmin). The donor atoms include the His-3 and free N-terminal amine group which facilitate the formation of peptide bonds in the ion coordination [12]. The **Cys34 site** involves the lone cysteine mentioned earlier, that is not involved in HSA's 17 disulphide bridges [2]. In reduced HSA, it contains a free thiol group with a very low pKa of 5. This Cys34 is located in a cleft between helices 2 and 3 and subdomain IA. This site is specific for metal ions that bind effectively to HSA via a single, high enthalpy, metal-sulphur bond. **Site A** is an extensively studied site. It is the 'multi binding site (MBS)' because it is the preferential site for Zn(II), one of the two thermodynamically equivalent sites for Cd(II) and a secondary weaker binding site for both Cu(II) and Ni(II). This site consists of imidazole nitrogen atoms (from His247 and His67) located in the axial positions of a distorted trigonal bipyramidal arrangement around the metal ion. Asn99 and Asp249 provide carboxyl ligands and the final, 5th, ligand is water (**fig.5b**). The His67 and Asn99 are from domain I whereas His247 and Asp249 are from domain II. There is also a 6th, more distant ligand which is provided by the His247 amid oxygen [12]. **Site B** is non-localised and is specific for cadmium (Cd(II)) with the same affinity as site A. Chemical shift shows that there is not more than one nitrogen ligand involved in binding. It is a primary site for Mn(II) ions and can also bind Zn(II), albeit with much lower affinity.

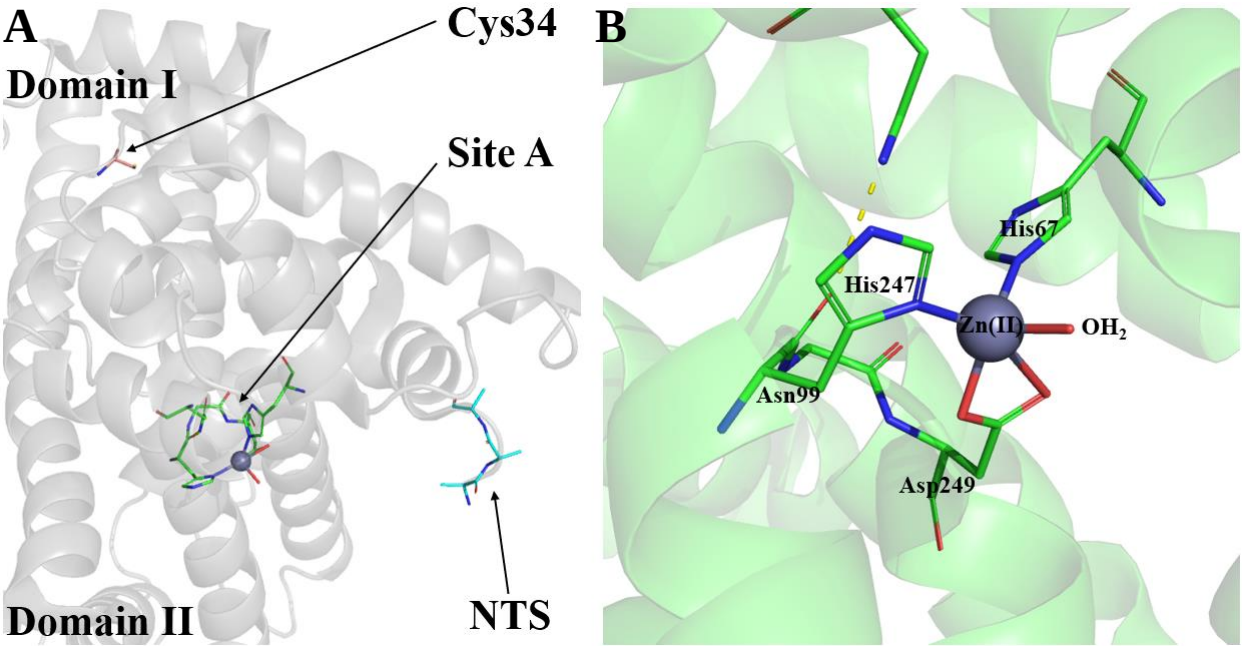


Figure 5: Ion Binding Sites made in *PyMol* [12][13][14]. A) Discovered locations of 3 of the 4 sites (Cys34, Site A and NTS) all within Domain II. The location of Site B has not yet been determined. **B)** The Site A active site with a Zn(II) ion bound. The key residues and water are labelled, forming a bipyramidal arrangement around the metal ion.

Drugs and Research

The future and utilisation of our knowledge lies in albumin-based drug delivery. The two major drug binding sites are named Sudlow's site 1 (binds drugs like warfarin) and Sudlow site II (binds drugs like ibuprofen) (**fig.2b**). Furthermore, drug metabolites (furosemide and salicylic acid) and Nonsteroidal Anti-Inflammatory Drugs (ibuprofen) can react covalently with HSA: the free thiol group of cysteine 34 is an attractive feature for forming such bonds in drug delivery [14]. Albumin-associated drugs are being developed to enhance drug biodistribution and bioavailability due to the reversibility of their binding. These drugs incorporate components, like fatty acids above, to potentiate albumin association and increase drug efficacy. Ligand binding capacity of HSA (as defined by crystallographic studies) are depicted in table 1: the sites correspond to sites visualised in figure 2b [15]. Ibuprofen can bind to sites FA6 and FA3/4 and has a fully crystallised structure (**fig6**).

FA1	FA3/4: Sudlow Site II	FA5	FA6	FA7: Sudlow Site I	CLEFT
Hemin 2º: Azapopazone 2º: Indomethacin 2º: TIB	Thyroxine 4 Diflunisal Diazepam Halothane Ibuprofen Indoxyl Sulphate Propofol 2º: CMPF	2º: Oxyphe-nbutzaone 2º: Propofol	2º: Diflunisal 2º: Halothane 2º: Ibuprofen	Azapropazone CMPF DIS Indomethacin Oxyphenobutazone Phenylbutazone TIB Warfarin 2º: Indoxyl-Sulphate 3º: Difflunisal	Thyroxine 5 2º: Indipamide

Table 1: Ligand Binding Capacity of HSA: Sites correspond to those visualised in figure 2b [19]**.** The 'cleft' is an additional site, mainly for thyroxines that is located in the interface between subdomains IIA and IIB. Ibuprofen is highlighted as it is depicted in figure 5.

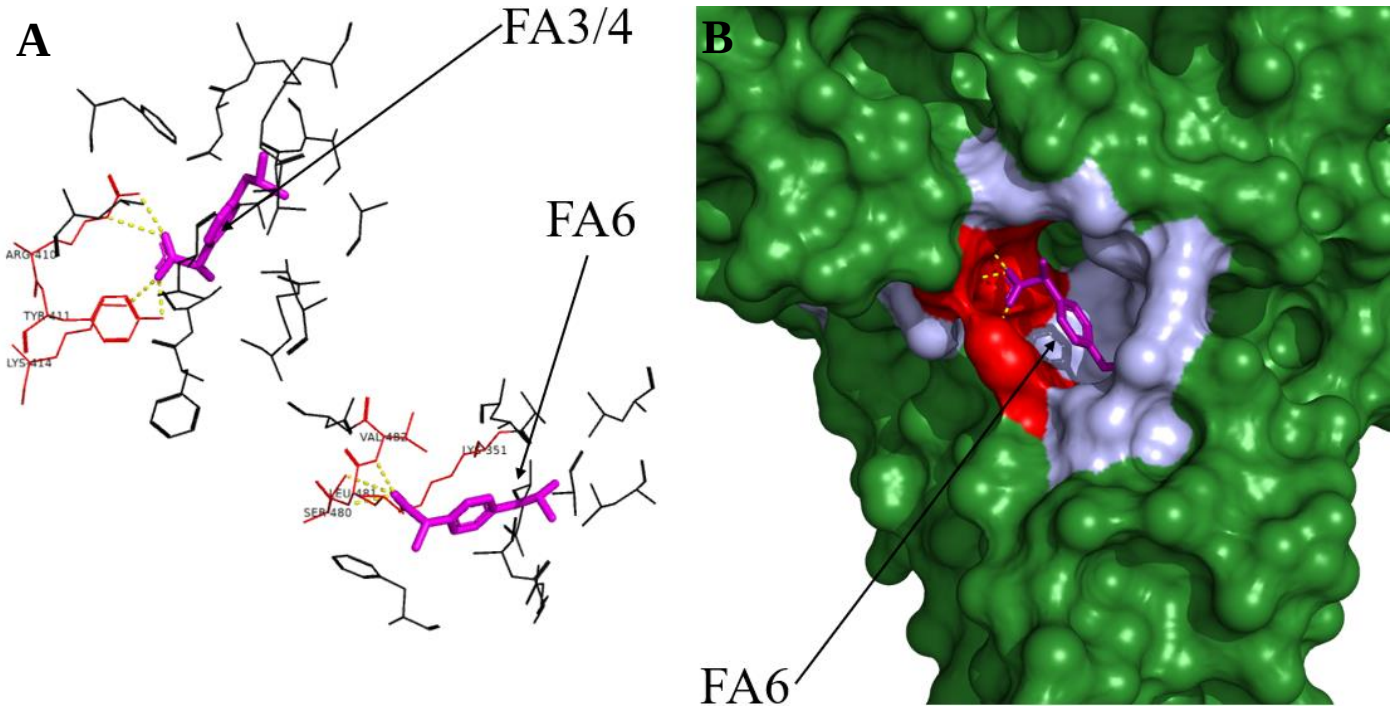


Figure 6: Ibuprofen in the active sites of HAS made in *PyMol* [15]: A) This depicts ibuprofen in FA3/4 (Sudlow's Site II) on the left and FA6 on the right. The ibuprofen molecule is purple, polar residues forming hydrogen bonds are red and the non-polar residues are gray. **B)** This is a surface representation, depicting the ibuprofen ligand in FA6. The molecule surface is green, polar surface in red, non-polar surface in grey and the ibuprofen ligand is purple.

Conclusion

Overall, serum albumin is the most abundant soluble protein in the body of all vertebrates and one of the longest known and most studied of all proteins. However, the ability of this protein to have such a myriad of affinities despite only two major binding sites is a paradox that still puzzles scientists [1]. There is great research into fatty acid binding and ion binding discussed in this report. However, the details of thermodynamics and kinetics of binding between metal ions and HSA is limited since there is a lack of X-ray structures of HSA in metal bound states; Site B's localisation is also yet to be discovered [12]. Furthermore, there is even possibility of an interplay between ions like Zn(II) and fatty acid binding that can 'translate certain aspects of the organismal energy state into global zinc signals'. This is mediated by interdependent binding between Zn(II) and fatty acid binding at the FA2 site: fatty acids may serve as an allosteric switch for Zn(II) binding to plasma [16]. Thus, ligand binding is not so simple as one compound. These factors all make understanding serum albumin ever more complex and a true paradox of functional complexity despite structural simplicity .

References:

[1] Theodore Peters, 2 - The Albumin Molecule: Its Structure and Chemical Properties, Editor(s): Theodore Peters, All About Albumin, Academic Press, 1995, Pages 9-II, ISBN 9780125521109

[2] Gabriella Fanali, Alessandra di Masi, Viviana Trezza, Maria Marino, Mauro Fasano, Paolo Ascenzi, Human serum albumin: From bench to bedside, Molecular Aspects of Medicine, Volume 33, Issue 3, 2012, Pages 209-290

[3] Albumin_Human (2022) UniProt. Available at: https://www.uniprot.org/uniprotkb/P02768/entry

[4] S. Sugio, A. Kashima, S. Mochizuki, M. Noda, K. Kobayashi, Crystal structure of human serum albumin at 2.5 Å resolution, *Protein Engineering, Design and Selection*, Volume 12, Issue 6, June 1999, Pages 439–446

[5] Theodore Peters, 4 - Genetics: The Albumin Gene, Editor(s): Theodore Peters, All About Albumin, Academic Press, 1995, Pages 133-187

[6] Paolo Ascenzi, Mauro Fasano, Allostery in a monomeric protein: The case of human serum albumin, Biophysical Chemistry, Volume 148, Issues 1–3, 2010, Pages 16-22, ISSN 0301-4622,

[7] Carta G, Murru E, Banni S, Manca C. Palmitic Acid: Physiological Role, Metabolism and Nutritional Implications. Front Physiol. 2017 Nov 8;8:902

[8] Ananyo A Bhattacharya, Tim Grüne, Stephen Curry, Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin11Edited by R. Huber, Journal of Molecular Biology, Volume 303, Issue 5, 2000, Pages 721-732.

[9] Ruettinger, A., Kiselev, M. A., Hauss, T., Dante, S., Balagurov, A. M., & Neubert, R. H. (2008). Fatty acid interdigitation in stratum corneum model membranes: a neutron diffraction study. *European biophysics journal : EBJ*, *37*(6), 759–771.

[10] J.R Brown, P Shockley. **Serum albumin: structure and characterisation of its ligand binding sites**
P Jost, O.H Griffith (Eds.), Lipid-Protein Interactions, vol. 1, Wiley & Sons, New York (1982), pp. 26-68

[11] López-Yerena A, Perez M, Vallverdú-Queralt A, Escribano-Ferrer E. Insights into the Binding of Dietary Phenolic Compounds to Human Serum Albumin and Food-Drug Interactions. Pharmaceutics. 2020 Nov 21;12(11):1123

[12] Wojciech Bal, Magdalena Sokołowska, Ewa Kurowska, Peter Faller, Binding of transition metal ions to albumin: Sites, affinities and rates, Biochimica et Biophysica Acta (BBA) - General Subjects, Volume 1830, Issue 12, 2013, Pages 5444-5455,

[13] Handing, K. B., Shabalin, I. G., Kassaar, O., Khazaipoul, S., Blindauer, C. A., Stewart, A. J., Chruszcz, M., & Minor, W. (2016). Circulatory zinc transport is controlled by distinct interdomain sites on mammalian albumins. Chemical science, 7(11), 6635–6648

[14] Larsen MT, Kuhlmann M, Hvam ML, Howard KA. Albumin-based drug delivery: harnessing nature to cure disease. Mol Cell Ther. 2016 Feb 27;4:3

[15] Jamie Ghuman, Patricia A. Zunszain, Isabelle Petitpas, Ananyo A. Bhattacharya, Masaki Otagiri, Stephen Curry, Structural Basis of the Drug-binding Specificity of Human Serum Albumin, Journal of Molecular Biology, Volume 353, Issue 1, 2005, Pages 38-52

[16] Barnett, J. P., Blindauer, C. A., Kassaar, O., Khazaipoul, S., Martin, E. M., Sadler, P. J., & Stewart, A. J. (2013). Allosteric modulation of zinc speciation by fatty acids. *Biochimica et biophysica acta*, *1830*(12), 5456–5464

Pdb Files used:

[4] 1AO6
[8] 1E7H
[13] 5IJF
[15] 2BXG

Words: 2,347