bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A computational study about the mechanism of action of metformin on hepatic gluconeogenesis, focused on its ability to create stable pseudo-aromatic copper complexes

Fabio Rebecchi[1*]

[1]Department of Bioinformatics, Calyx Informatics, Rome, Italy
[*]Corresponding Author
fabio@calyx-science.it

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Metformin is the best therapeutic choice for treating type 2 Diabetes.

Despite this, and the fact it has been prescribed worldwide for decades, its mechanism of gluconeogenesis inhibition is still unknown.

In the following work a novel mechanism of inhibition is suggested: that metformin performs its action on the target enzyme not as a pure molecule but, after sequestering endogenous cellular copper, as a copper complex.

This result was obtained using chemoinformatics methods including homology modeling for the creation of the target enzyme's tridimensional virtual structure, molecular docking for both the determination of the movement of the prosthetic group inside its cavity and for the identification of the best ligand poses for the metformin copper complexes, and eventually pharmacophore modeling and virtual screening to find alternative virtual leads that could achieve similar effects.

The simulations show the complex binding as a non competitive inhibitor to the large exit of the mitochondrial glycerophosphate dehydrogenase enzyme's FAD cavity, preventing FAD movement inside the cavity and/or quinone interaction and therefore its electron transfer function.

The proposed mechanism seems to be successful at explaining a wide range of existing experimental results, both regarding measurements of metformin non-competitive inhibition of GPD2 and the role of copper and pH in its action.

The virtual screening outcome of at least two similarly active purchasable molecule hints to an easy way to experimentally test the proposed mechanisms.

In fact, the virtual leads are very similar to the copper complex but quite different from metformin alone, and a laboratory confirmation of their activity should plausibly imply that metformin acts in synergy with copper, giving us the ability to design new antidiabetic drugs in a novel and more rational fashion, with significant savings in research costs and efforts.

Keywords: Metformin, Diabetes, Gluconeogenesis, GPD2, Copper Complex, Molecular Docking

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Section 1 Introduction

1.1 Metformin: a brief story

Metformin is a molecule that belongs to the biguanidine class of drugs (Figure 1.1).
Its hypoglycemic effect in animals has been well known since 1929, and the same effect on humans was discovered in 1962. (Gottlieb & Auld, 1962).
It is the first therapeutic resource for Type 2 Diabetes Mellitus, being able to inhibit hepatic gluconeogenesis, and preventing hyperglycemia, without causing increase in insulin secretion, hypoglycemia or weight gain. (Bennet et al., 2010)
It has been prescribed for over forty years now (Rojas, Chávez-castillo & Torres, 2014).
Furthermore, metformin showed the ability to decrease the incidence of several kind of cancers in diabetic patients who underwent metformin therapy (Karabulut-Bulan, 2016; Rojas, Chávez-castillo & Torres, 2014), and it recently attracted the attention of anti-ageing researchers because it has been noted to extend lifespan in nematodes and mice (Onken & Driscoll, 2010).

1.2 Metformin and gluoconeogenesis: a review

In spite of the long and widespread use of metformin, its mechanism of action is unknown.
Through the years, many investigations addressed the problem, often with contradictory results.
The earliest studies focused on respiratory Complex I, and (Owen & Halestrap, 2000) found that millimolar concentrations of metformin mildly inhibit respiratory chain's Complex I in isolated mitochondria and this effect seemed to be directly linked to the gluconeogenesis inhibition.
Then, researchers shifted their attention to the LKB1/AMPK pathway, with (Zhou et al, 2001) reporting that metformin needs AMPK to exert its inhibitory effect on gluconeogenesis, taking place through AMPK inhibition of SREBP-1 gene expression, and hypothesized that metformin directly activates AMPK. On the road traced by these results a subsequent study (Shaw et al, 2005), pointed out that metformin needs LBK1, the AMPK upstream activator, to decrease blood glucose level.
However, (Savage et al, 2006) and (Fullerton et al, 2013) showed that the inhibition of gluconeogenic transcription could be achieved by a parallel mechanism also, i.e. it could be due to the sensitization of transcription to insulin where metformin activity depends on ACC and induces an AMPK/ACC mediated decrease of hepatic lipid content.
Furthermore, the AMPK signalling inhibition hypothesis is contradicted by the results of (Foretz et al, 2010), that found that metformin depended solely on the ability to decrease hepatic energy charge to exert its anti-hyperglicemic action, performed even in mice lacking AMPK or LKB1, and by the results of (Miller et al, 2013), showing that metformin increases cAMP concentration and cAMP is

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

capable of inhibiting adenylyl cyclase preventing its activation by glucagon and therefore gluconeogenesis as well.

But the reduction of hepatic energy charge's hypothesis has its opponent too.

(Hawley, Olsen & Hardie, 2002) in fact, showed not only that metformin does not affect regulation of AMPK by upstream kinases and phosphatases, but also that it activates AMPK without altering the AMP/ADP and the ADP/ATP ratios.

In this intricate and contradictory scenario, a recent study (Madiraju et al, 2014) took into consideration all the previous results listed above and suggested, backed by original and exhaustive experimental data, that metformin actually reduces gluconeogenesis by non-competitive inhibition of the mitochondrial glycerophosphate dehydrogenase (GPD2).

Supplementary insight on the issue is given by four other studies that target the interaction of biguanides with copper and the metal role in their action.

(Repiščák et al., 2014) shows that, among biguanides, metformin is not only the best at inhibiting gluconeogenesis, but it's the most effective at forming stable pseudo-aromatic copper complexes, suggesting a correlation between the two facts. Moreover, the complex stability increases with deprotonation, an event that is more likely to happen in the high pH environment of mitochondria.

(Logie et al., 2012) found that glucose production in liver cells is no longer decreased by metformin when cellular copper is sequestered, suggesting that copper is essential to metformin action.

(Annapurna & Rao, 2007) found that, in alloxan induced diabetic rabbits, Cu^{2+} and Ni^{2+} metal complexes of metformin showed a significant increase in hypoglicemic activity when compared to the pure drug. This pattern seems common in biguanides, since the antiproteolytic action of metformin and phenformin shows a significant increase for the metal complexes when compared to pure drugs (Sweeney et al, 2003).

1.3 Project objectives and scope

(Madiraju et al., 2014) proposes a mechanism of action based on the hypothesis that FAD, the GPD2's prosthetic group, needs to perform an oscillatory sliding-in/sliding-out movement in order to first oxidise glycerol-3-phosphate and then transfer the electrons through the transport chain, and that metformin blocks this movement by docking with the smaller of the cavity's two entrances.

But a pilot simulation carried out as part of this work, performed using molecular docking software on 2RGH, the bacterial alpha-glycerophosphate oxidase that is a template for GPD2, showed three weaknesses in this hypothesis.

The first was that the free binding energy of metformin docked in the cavity's small entrance is too little to account for the inhibition constant (KI) found in (Madiraju et al., 2014).

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The second was that metformin seems to bind to this target only when glycerol-3-phosphate, its substrate, is absent, therefore lacking one of the required characteristics of a non-competitive inhibitor.

The third was that this mechanism isn't able to explain why the absence of copper halts the inhibitory action.

The same simulation showed that the weaknesses above are overcome when a square planar copper-complex of metformin (Figure 1.2) is used as a ligand instead, this time with a preferential docking area in the cavity's large entrance.

This preliminary result seemed to suggest that metformin performs its inhibitory action by collecting endogenous cellular copper and by docking on GPD2 as a copper-complex, and this mechanism could explain not only the inhibitory effects measured in (Madiraju et al., 2014) but account at the same time both for the role of copper in the process as highlighted in (Logie et al., 2012) and for the increased potency of the metal complex shown in (Annapurna & Rao, 2007).

This prompted to carry out a systematic series of molecular docking simulations aimed at understanding in depth the multiple interactions between GPD2, FAD, glycerol-3-phosphate, metformin alone, monometformin copper complex, dimetformin copper complex and coenzyme Q10, the final electron acceptor. Simulation runs were performed with phenformin too, a similar antihyperglycemic biguanide.

Objectives of these simulations were to identify the most favourable FAD positions, both alone and bond with glycerol-3-phosphate, and the most favourable poses for metformin and its complex to achieve inhibition by blocking the FAD's electron transport function.

The best docking poses, achieved by the copper complex, were then exploited to perform a virtual screening on databases of commercial molecules with the aim of identifying lead compounds able to exert a similar action.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Section 2: Methods

2.1 Protein structures retrieval

PDB structure of 2RGH was retrieved from Protein Data Bank (Berman et al., 2000).
GPD2 amino acid sequence in FASTA format was retrieved from UniProt (The UniProt Consortium, 2014).

2.2 Ligand structures retrieval

PDB structures of FAD, G3P and co-enzyme Q10 were retrieved from Protein Data Bank (Berman et al., 2000).
PDB structures of metformin and phenformin were retrieved from Drug Bank (Wishart et al., 2006).
The lead candidates structures were obtained in MOL2 file format from Zinc (Irwin et al., 2012) and from ChemSpider (Royal Society of Chemistry, 2016), and subsequently converted into PDB by Open Babel (O'Boyle et al., 2011), and eventually refined in Discovery Studio (Dassault Systèmes, 2016).

2.3 Protein structures modeling

Since GPD2's crystallographic structure was unavailable, the GPD2 FAD cavity was modeled by substituting the 2RGH cavity residues with the corresponding residues of the GPD2 amino acid sequence.
First, in order to achieve the original protein structure, as it was prior to the processing needed for the crystallography procedure, the seleno-methionine residues of the 2RGH structure were substituted with methionine residues using the Macromolecules/replace function of Discovery Studio (Dassault Systèmes, 2016).
The residues forming the FAD cavity in 2RGH were identified using Receptor-Ligand Interaction Explorer of Discovery Studio (Dassault Systèmes, 2016).
The other cavity residues, corresponding to the small and the large cavity entrance, were manually identified in Autodock Tools (Morris et al., 2009).

The correspondence between the GPD2 and the 2RGH residues was determined using the target-template alignment function present on I-Tasser (Yang et al., 2015) that returned 2RGH as one of the top threading templates for the GPD2's FASTA sequence in input (Figure 2.1).

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The substitution of the 2RGH cavity residues with the corresponding GPD2 residues was performed with Discovery Studio (Dassault Systèmes, 2016) using the Macromolecules/replace function.

2.4 Ligand structures modeling

The dimetformin (Figure 2.2) square planar copper complex was created in Discovery Studio (Dassault Systèmes, 2016) using the bond distances reported in (Repiščák et al., 2014).
The protonation states of residues and ligands at various pHs (Figure 2.3). were determined by MarvinBeans (ChemAxon, 2016).

2.5 Molecular docking

A first series of molecular dockings were performed for FAD into the GPD2's cavity, in order to assess the plausibility and the extent of a FAD movement inside it.
Then glyceraldehyde 3-phosphate (G3P) was docked with various positions of FAD to assess the most suited FAD positions for docking G3P, and coenzyme Q10 was docked as well to understand its target site for electrons retrieval.
Eventually, the following ligands were docked to the protein with FAD in different positions, to assess their ability to act as non competitive inhibitors: pure metformin, pure phenformin, dimetformin copper complex, , dimetformin copper complex, monophenformin copper complex, and the two alternative lead candidates described in the Results section.
Molecular docking of ligands on the GPD2 cavity was performed using Autodock 4 (Morris et al., 2009) and Autodock Vina (Trott & Olson, 2010).
On Autodock 4, a parameter file with atomic data for copper was added, since copper wasn't supported in the standard version.
PDBQT format files for ligands and for the protein structure, necessary for Autodock input, were prepared by adding hydrogens, computing Gasteiger charges (Gasteiger & Marsili, 1980) and merging non polar hydrogens on Autodock Tools (Morris et al., 2009). Where necessary, bonds to be considered flexible in ligands and residues were set up on PDBQT files using Autodock Tools (Morris et al., 2009). The total number of flexible bonds in a simulation ranged from zero to 25.
The coordinates Grid Boxes (Figure 2.4), necessary to delimit the simulation docking area, were prepared using Autodock Tools (Morris et al., 2009). Grids' volume was between 1000 and 2700 cubic angstroms.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The searches were set up with a higher level of accuracy than standard, at the expense of calculation speed, with 25 millions of maximum evaluations set on Autodock 4 (Morris et al., 2009) and the exhaustiveness parameter set on 25 on Autodock Vina. (Figure 2.5)

The number of poses to be returned was set on 50 inside an energy range of maximum 50 kcal/mol.

The inhibition constant Ki was calculated as nanomoles applying the following formula:

$$K_i = 10^9 \times \exp\left(\frac{1000 \times E_b}{T \times R}\right),$$

where E_b is the binding energy expressed in kcal/mol, T is the temperature (298K) expressed in Kelvins and R is the gas constant approximated as 1.987 cal K^{-1} mol^{-1} .

2.6 Analysis of ligand interaction and binding cavities

A preliminary control of results was performed using Autodock Tools (Morris et al., 2009) and Chimera (Pettersen et al., 2004).

The analysis of the docked poses and of the interaction of the ligands with the cavity residues was performed using the Receptor-Ligand Interaction Explorer of Discovery Studio (Dassault Systèmes, 2016).

2.7 Pharmacophore and virtual screening

A pharmacophore descriptor based on the dimetformin copper complex was created using the manual pharmacophore creation feature (Figure 2.6) of ZincPharmer (Koes & Camacho, 2012).

The coordinates and radius of each functional group were derived from distance measurements taken in Discovery Studio (Dassault Systèmes, 2016) on the copper complex PDB model.

A virtual screening was performed on the "Purchasable Compounds" of Zinc (Irwin et al., 2012).

The parameters used for the search were the following:

- Molecular Weight: less than 250
- Max hits 100: 1
- Max hits per conf: 1
- Max hits per mol: 1

Two new compounds were designed from scratch too, using MarvinSketch (ChemAxon, 2016), following simple similarity considerations with the monometformin and the dimetformin copper complexes. The compounds are described in the Results section.

In order to obtain other lead candidates, a similarity search on these new compounds was performed on ChemSpider (Royal Society of Chemistry, 2016),

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.8 Hardware used

The following hardware was employed for Molecular Docking:

A MacBook Pro with an 8-core Intel i7 2.8 GHZ CPU and 16 GB 1600 MHz DDR3 RAM.

The following hardware was employed for all other activities:

An Alienware PC with an Intel Core i7-6820HK (Quad-Core, 8MB Cache, 4.1GHz) CPU and 32 GB RAM.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Section 3 Results

3.1 Homology modeling

The analysis of the FAD binding cavity of 2RGH showed it is made up of the 66 residues reported in Table 3.1.

This result includes the edges of both the small and the large entrance.

Confronting the above residues with the corresponding ones in the GPD2 sequence gave the result highlighted in Figure 3.1, with 24 residues not matching one another, corresponding to the 36% of the total.

After substituting into the 2RGH tridimensional structure the non matching residues from GPD2, the binding cavity 66 residues became the ones reported in Table 3.2.

Figure 3.2 shows the cavity and the position of the small and of the large entrance in the newly modeled cavity.

3.2 Molecular docking

3.2.1 FAD sliding movement

The redocking of FAD (Figure 3.3) in the new cavity gave a best pose with a RMSD of atomic positions less than 0.2 angstroms from the original 2RGH crystal structure pose (Colussi et al., 2008), as shown in Figure 3.4, confirming a good reliability of the docking software. The estimated free binding energy of the original FAD from 2RGH crystallography was $-14.4\frac{kcal}{mol}$, while the energy of the best pose for the FAD docked in the new cavity was $-14.7\frac{kcal}{mol}$.

This position, from now on labeled as the "fully slid in", was deemed the most stable by the docking calculations.

Three other reasonably stable FAD poses were found by the software, and are summarised, together with their labels and the fully slid in pose, in Table 3.3.

The distance from the crystallographic position is reported to quantify the extent of the FAD's movement.

In Figures 3.5, 3.6, 3.7 and 3.8 the tridimensional representation of the four poses is shown.

3.2.2 G3P docking

The results of the docking of G3P molecule (Figure 3.9) with the receptor containing the four different FAD poses are summarised in Table 3.4.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The distance between the G3P's hydrogen and the FAD's nitrogen in position 5 on the isoalloxine ring (Figure 3.10) is reported because G3P's oxidation reaction is believed to initiate from these two atoms (Yeh, Chinte & Du, 2008).
In Figures 3.11-3.13 the cavity interaction of the three G3P docking poses is shown, with highlights on the bond types.

Overall, G3P showed a decreasing affinity for the binding site with the increasing of the FAD sliding out movement. This agrees with the fact (Fraaije & Mattevi, 2000) that in flavoproteins the docking position of the substrate is buried well below the protein surface to protect the reaction site from the solvent.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.2.3 FADH2 interactions and redox reactions

In order to assess the most favourable positions for FAD's reduction and for subsequent $FADH_2$ oxidation and electron transfer to ubiquinone, the cavity interactions of the nitrogens in position 1 and 5 on the isoalloxine ring of FAD (Figure 3.14) were analysed for all the four FAD positions.
According to (Fraaije & Mattevi, 2000) and to (Ghanem & Gadda, 2008), in fact, the FAD in flavoproteins needs a hydrogen bond donor close to position 5 to initiate the reduction reaction, and a positive residue close to position 1 to stabilise the negative charge developed in the intermediate step of the reduction process.
The results are summarised in Table 3.5.

3.2.4 Coenzyme Q10 docking on GPD2

In order to assess the possibility of a direct electron transfer from $FADH_2$ to the final receptor, Coenzyme Q10 (Figure 3.15), the ubiquinone was docked with GPD2 large entrance and its immediate surroundings.
The results of the docking of coenzyme Q10 with the receptor's large entrance, in correspondance to the four different FAD poses, are summarised in Table 3.6.
Coenzyme Q10 shows a good binding ability only when FAD is fully slid in or when it's slightly slid out. The latter position is slightly preferential from the binding energy point of view, but the former is more specific because it's characterised by bonds that, on average, are more directional in nature. Finally, the oxidising oxygens of coenzyme Q10 are closer, in the slightly slid out FAD pose, to the target nitrogen in position 5 of the isoalloxine ring. (see Figures 3.21-3.24)
A mechanism where first the substrate (G3P) docks to the flavine to reduce it and then, after substrate leaves, an ubiquinone docks, in turn, to get reduced, is an example of a so called ping-pong mechanism, something not uncommon in flavoproteins as shown by (Maenpuen et al, 2015).
In Figures 3.16 and 3.17 the 2D interaction diagrams are shown.
In Figures 3.18 and 3.19 the tridimensional representation of the two binding poses is shown.

3.2.5 Metformin at the small entrance

Pure metformin was docked to the small entrance of the FAD cavity in order to assess its ability to block the FAD sliding in movement.
The results of the docking of the metformin molecule with the receptor's small entrance, in correspondance to the four different FAD poses, are summarised in Table 3.7.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Metformin doesn't bind to the small entrance when FAD is fully slid in, and its binding becomes stronger along with FAD's sliding out movement.

In Figures 3.20, 3.21 and 3.22 the tridimensional representation of the three poses is shown.

3.2.6 Metformin at the large entrance

Pure metformin was docked to the large entrance of the FAD cavity in order to assess its ability to block the FAD sliding out movement.

The results of the docking of metformin molecule with the receptor's large entrance, in correspondance to the four different FAD poses, are summarised in Table 3.8.

All metformin binding poses at the large entrance resulted either far from FAD or with free binding energies not better than -4.0 kcal/mol.

3.2.7 Dimetformin copper complex at the large entrance

In order to assess their ability to block the FAD sliding movement or to block the coenzyme Q10 docking, both the cationic and the neutral metformin copper complexes (Figure 3.23) were docked to the small and to the large entrance of the FAD cavity.

The planar pseudo-aromatic structures of the complexes were realised using the (Repiščák et al., 2014) hypothesis of a metal coordination through a σ-donation and a π-backdonation with a delocalised π ring system being formed (Figure 3.24).

They both showed affinity with the large entrance of the binding cavity, while they were unable to dock to the small entrance.

The results of the docking of the copper complexes with the receptor's large entrance, in correspondance to the full slid in FAD pose, are summarised in Table 3.9.

The detailed interaction of the neutral complex with the cavity residues is summarised in Figure 3.25.

It's worth noticing two facts about the above figure:

1. the interactions of the cationic complex are very similar yet less effective, with the Pi-anion repulsion substituted by a charge-charge attraction with the oxygen of ASP300, and two missing hydrogen bonds caused by the fact that the two hydrogen bond acceptors of the neutral complex are substituted by the protonated nitrogens of the cationic form
2. the Pi-anion repulsion characterising the neutral complex pose has an indirect stabilising effect on the pose, because it "pushes away" the negative aromatic cloud from the ring allowing it to be involved more strongly in the three Pi-cation interactions

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The tridimensional representation of the neutral copper complex position relative to the interacting residues is shown in Figure 3.26.

When the pose of the dimetformin copper complex docked with the fully slid in FAD and the pose of the slightly slid out FAD are superimposed a collision is shown, demonstrating that the copper complex is able to block the FAD sliding out movement. This is shown in Figure 3.27.

The more protruding FAD positions are blocked as well.

Finally, a docking simulation was performed, docking G3P to the cavity's large entrance when the dimetformin copper complex was bound too, showing that G3P docks to GPD2, and close to the FAD's head, with the same poses and binding energies no matter if the copper complex is present or not, and therefore the complex satisfies the requisites of a non-competitive inhibitor.

In Figures 3.28 and 3.29 this "cohabitation" between G3P and the complex is shown.

3.2.8 Phenformin docking at the small and at the large entrance

Pure phenformin (Figure 3.30) was docked to both the small and the large entrance of the FAD cavity, in analogy of what was done with metformin in paragraphs 3.2.5 and 3.2.6 above.

The results are similar to the ones of metformin, with phenformin showing no significant binding to the large exit and with a binding to the small exit becoming stronger along with FAD's sliding out movement, with better binding energies than metformin, as summarised in Table 3.9.

In Figure 3.31 phenformin binding to the cavity's small entrance when FAD is slightly slid out is shown.

3.2.9 Monophenformin copper complex at the large entrance

The monophenformin copper complex (Figure 3.32) was docked to both the small and the large entrance of the FAD cavity.

The complex is imagined to be furtherly coordinated, in solution, with two molecules of water, that are displaced by the cavity residues at the moment of docking.

The results of the docking of the monophenformin copper complex with the cavity's large entrance, in correspondance to the full slid in FAD pose, are summarised in Table 3.10. No significant interaction was shown with the cavity's small entrance.

The interactions of the complex with the cavity residues are shown in Figure 3.33.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

When the pose of the monophenformin copper complex docked with the fully slid in FAD and the pose of the slightly slid out FAD are superimposed a collision is shown, demonstrating that the complex is able to block the FAD sliding out movement. This is shown in Figure 3.34.
The more protruding FAD positions are blocked as well.

3.2.10 Monometformin copper complex at the large entrance

In analogy with the monophenformin simulation, a mono-deprotonated monometformin copper complex (Figure 3.35) was docked to both the small and the large entrance of the FAD cavity.
The complex is imagined to be furtherly coordinated, in solution, with two molecules of water, that are displaced by the cavity residues at the moment of docking.
The results of the docking of the monometformin copper complex with the cavity's large entrance, in correspondance to the full slid in FAD pose, are summarised in Table 3.11. No significant interaction was shown with the cavity's small entrance.
The interactions of the complex with the cavity residues is shown in Figure 3.36.
When the pose of the monometformin copper complex docked with the fully slid in FAD and the pose of the slightly slid out FAD are superimposed a collision is shown, demonstrating that the complex is able to block the FAD sliding out movement. This is shown in Figure 3.37.
The more protruding FAD positions are blocked as well.
A final simulation showed the same non-interference with G3P docking, as already observed for $[Cu(Met\text{-}H)_2]$. In Figure 3.38 G3P can be seen docked when $[Cu(Met\text{-}H)]^{1+}$ is present.

3.3 Virtual screening

3.3.1 Leads identification

The pharmacophore modeled as a result of the analysis of the binding cavity interactions of $[Cu(Metf\text{-}H)_2]$ is represented in Figure 3.39, where the colors used are the standards for H-bond donor, H-bond acceptor, hydrophopic domain and aromatic ring functions.
The first virtual screening query returned 100 results, the only one fullfilling all the pharmacophore requirements was [4-amino-6-(isopropylthio)-2-methyl-pteridin-7-yl]amine, shown in Figure 3.40.
A second search was performed too, this time on ChemSpider (Royal Society of Chemistry, 2016) based on similarity considerations with the monometformin and the monophenformin complexes.
In more detail, the compound sought after had the following characteristics:

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. the ability to coordinate in a square planar fashion with Cu^{2+} or Ni^{2+}, possibly with stronger ligands than the amine groups, after consulting the NIST database (Smith & Marshall, 2004) the choice fell upon a dithioic acid functional group
2. the ability to perform a hydrogen bond donor function close to the coordinating group
3. the presence of a hydrophobic tail at least four carbons long

The search result was butyl carbamodithioic acid, represented in Figure 3.41.

3.3.2 Docking of the virtual screening leads

The results of the docking of the two leads to the cavity's large entrance are summarised in Table 3.12.

[4-amino-6-(isopropylthio)-2-methyl-pteridin-7-yl]amine, thanks to its high number (4) of hydrogen bond acceptors, formed a netword of bonding that was even better than the one formed by the original copper complex, with a final binding energy result slightly worse than $[Cu(Metf\text{-}H)_2]$ probably because of the torsional energies involved in the adjustment of the cavity residues around the ligand. $[Ni(BCA\text{-}H)]^{1+}$, instead, was slightly penalised with respect to the monometformin copper complex probably because its slimmer shape caused the interaction with some residues to be weaker, for instance the hydrogen bond formed with ASP300 resulted 0.1 angstroms longer (2.6 against 2.5).

The detailed interaction diagram of the ligands [4-amino-6-(isopropylthio)-2-methyl-pteridin-7-yl]amine and of $[Ni(BCA\text{-}H)]^{1+}$ with each residue are shown in Figures 3.42-3.43.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Section 4 Discussion

The 2RGH cavity presents a good similarity with the GPD2 sequence with 64% residues in common (Figure 3.1), and the docking of FAD in the modeled cavity shows a RMSD between the original and the new FAD positions of just 0.2 angstroms (Figure 3.4). The modeled cavity can therefore be considered a good approximation of the real one.

FAD's fully slid in position must be the environment where G3P's oxidation takes place, because of the following facts:

1. it corresponds to FAD's lowest binding energy to GPD2 (Table 3.3), therefore is the stablest and the most suited to accept a substrate;
2. it corresponds to G3P's lowest binding energy to GPD2 (Table 3.4), therefore it's where G3P spends most of its time on GPD2;
3. it's the only position characterised by the presence of a hydrogen bond donor close to position 5 nitrogen (Table 3.5), a characteristic needed to initiate the reaction (Ghanem & Gadda, 2008);
4. In that position G3P is oxidised well below the protein surface, exactly what flavoproteins do to protect reactants from solvent (Fraaije & Mattevi, 2000).

The above facts rule out (Madiraju et al., 2014)'s hypothesis that metformin can exert its action at the cavity's small entrance, because metformin doesn't bind to the small entrance when FAD is in the fully slid in position (Table 3.7), and this is exactly the moment when G3P binds best (1.-4. above): but a non-competitive inhibitor must bind the receptor equally well both when the substrate is and is not docked, and not only when the substrate (G3P) is away (Strelow et al., 2012).

Both $[Cu(Met\text{-}H)]^{1+}$ and $[Cu(Met\text{-}H)_2]$, instead, through binding at the cavity's large entrance show the ability to interfere with the post G3P oxidation steps, and therefore with the subsequent electron transfer, in one or both of the following ways:

I. preventing the sliding out movement of $FADH^-$, needed to reach the slightly slid out position where the negative charge is stabilised by HIS259 (Table 3.5; Fraaije & Mattevi, 2000; Ghanem & Gadda, 2008), a step needed before a proton is taken from the environment and complete reduction to $FADH_2$ is finalised and subsequent electron transfer takes place;
II. preventing coenzyme Q10 docking in front of $FADH_2$ and subsequent $FADH_2$'s oxidation and electron transfer (Tables 3.8 and 3.11).

Furthermore, both complexes have inhibition constants inside (Table 3.8) or close (Table 3.11) to (Madiraju et al., 2014)'s Ki measured range of $(3.8 - 5.5)\times 10^4 nM$, they dock effectively only when

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in deprotonated forms (Tables 3.9 & 3.12) so at high pH values, and they don't interfere with G3P binding (Figures 3.28 & 3.38), satisfying the requisites of a non-competitive inhibitor.

So in both cases I. and II. above, the suggested mechanisms are able to fit with both with the experimental data of (Madiraju et al., 2014) and with the ideas that mitochondrial high pH (Repiščák et al., 2014), copper (Logie et al., 2012; Annapurna & Rao, 2007) and pseudo-aromaticity in metformin complexes (Repiščák et al., 2014), must be involved.

The model is able to explain in a similar way (Table 3.10 & Figure 3.31) also the antihyperglycemic action of phenformin, a fellow biguanide and predecessor of metformin as an antidiabetic.

In conclusion, this computational study suggests that metformin performs its inhibitory action on gluconeogenesis by sequestering endogenous cellular copper and docking as a metal complex at mitochondrial glycerophosphate dehydrogenase in correspondence of the FAD cavity's large entrance, this way blocking the electron transfer mechanism taking place after G3P oxidation.

The proposed alternative drugs resulting from the virtual screening represents a possibility to shed light on the issue directly in the laboratory, since evidence of an inhibitory action of [4-amino-6-(isopropylthio)-2-methyl-pteridin-7-yl]amine would confirm the mechanism suggested for the dimetformin complex, while an inhibitory action of $[Ni(BCA\text{-}H)]^{1+}$ would confirm the mechanism suggested for the monometformin complex.

Needless to say, a better understanding of metformin mechanism of action would open the possibility to design new antidiabetic drugs taylormade to the characteristics of the target cavity.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

List of abbreviations

ACC = Acetyl-CoA carboxylase;

ADP = Adenosine diphosphate;

AMP = Adenosine monophosphate;

AMPK = 5' adenosine monophosphate-activated protein kinase;

ATP = Adenosine triphosphate;

BCA = Butyl carbamodithioic acid;

cAMP = Cyclic adenosine monophosphate;

FAD = Flavin adenine dinucleotide;

FASTA = FAST-All, is a text-based format for representing either nucleotide sequences or peptide sequences ("All" means that it works with any alphabet), in which nucleotides or amino acids are represented using single-letter codes;

G3P = Glyceraldehyde 3-phosphate;

GPD2 = Glycerophosphate dehydrogenase 2, the mitochondrial glycerophosphate dehydrogenase;

KI = Inhibition constant;

LKB1 = Liver Kinase B1;

LOMETS = Local Meta-Threading-Server, is an on-line web service for protein structure prediction;

Metf = Metformin;

Metf-H = Deprotonated metformin, where the deprotonation occurs at the secondary amine in position 2 of the parent chain;

MOL2 = Molecular file format 2;

NIST = National Institute of Standards and Technology;

PDB = Protein Data Bank file format;

PDBQT = Protein Data Bank, Partial Charge (Q), & Atom Type (T)) format;

Phen = Phenformin;

Phen-H = Deprotonated phenformin, where the deprotonation occurs at the secondary amine in position 2 of the parent chain.

SDF = Structure data format file;

SREBP-1 = Steroid regulatory element binding protein 1.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

- Bennett, W. L., Maruthur, N. M., Singh, S., Segal, J. B. and Wilson, L. M. (2011) 'Annals of Internal Medicine Comparative Effectiveness and Safety of Medications for Type 2 Diabetes : An Update Including New Drugs and 2-Drug Combinations', *Annals of Internal Medicine*, year 2011, vol. 154, pp.602-613, [Online]. DOI: 10.7326/0003-4819-154-9-201105030-00336 (Accessed 1 September 2016)
- ChemAxon. (2016) Marvin Suite software, release 16. [Online]. Available from: <https://www.chemaxon.com/download/marvin-suite/> (Accessed 1 September 2016)
- Colussi, T., Parsonage, D., Boles, W., Matsuoka, T., Mallett, T. C., Karplus, P. A. and Claiborne, A. (2008) 'Structure of alpha-glycerophosphate oxidase from Streptococcus sp.: A template for the mitochondrial alpha-glycerophosphate dehydrogenase', *Biochemistry*, 47(3), pp. 965–977, [Online]. DOI: 10.1021/bi701685u (Accessed 1 September 2016)
- Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2016, San Diego: Dassault Systèmes, 2016. [Online]. Available from: <http://accelrys.com/products/collaborative-science/biovia-discovery-studio/> (Accessed 1 September 2016)
- Foretz M.H.S., Leclerc J., Zarrinpashneh E., Soty M., Mithieux G., Sakamoto K., Andreelli F., Viollet B. (2010) 'Metformin inhibits hepatic gluconeogenesis in mice independently of the LKB1/AMPK pathway via a decrease in hepatic energy state.' Journal of Clinical Investigation, 120, pp. 2355–2369, [Online]. DOI: 10.1172/JCI40671 (Accessed 1 September 2016)
- Fraaije, M. W. and Mattevi, A. (2000) 'Flavoenzymes: Diverse catalysts with recurrent features', *Trends in Biochemical Sciences*, 25(3), pp. 126–132, [Online]. DOI: 10.1016/S0968-0004(99)01533-9 (Accessed 1 September 2016)
- Gasteiger, J; Marsili, M. (1980) 'Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges.' *Tetrahedron*, 1980, 36, 3219–3228. [Online]. DOI: 10.1016/0040-4020(80)80168-2 (Accessed 1 September 2016)
- Ghanem, M. and Gadda, G. (2006) 'Effects of reversing the protein positive charge in the proximity of the flavin N(1) locus of choline oxidase', *Biochemistry*, 45(10), pp. 3437–3447, [Online]. DOI: 10.1021/bi052514m (Accessed 1 September 2016)
- Gottlieb, B. and Auld, W. H. (1962) 'Metformin in treatment of diabetes mellitus.', *British Medical Journal*, 1(5279), pp. 680–2, [Online]. DOI: 10.1136/bmj.1.5293.1692-b (Accessed 1 September 2016)
- Hawley S.A., Olsen G.S., Hardie D.G.. (2002) 'The antidiabetic drug metformin activates the AMP-activated protein kinase cascade via an adenine nucleotide-independent mechanism.', *Diabetes*, 51, pp.2420–2425, [Online]. DOI: 10.2337/diabetes.51.8.2420 (Accessed 1 September 2016)
- H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne. (2000) 'The Protein Data Bank,', *Nucleic Acids Research*, Vol. 28, No. 1, 2000, pp. 235-242, [Online]. DOI: 10.1093/nar/28.1.235 (Accessed 1 September 2016)
- Irwin, Sterling, Mysinger, Bolstad and Coleman. (2012) 'ZINC: A Free Tool to Discover Chemistry for Biology.', *J. Chem. Inf. Model.*, 2012, 52 (7), pp 1757–1768, [Online]. DOI: 10.1021/ci3001277 (Accessed 1 September 2016)
- Karabulut-Bulan, O. (2016) 'New Insight Into Metformin Action: Diabetes, Prostate Cancer, Ion Channels', *Marmara Pharmaceutcal Journal*, 20(2), p. 216, [Online]. DOI: 10.12991/mpj.20162014802 (Accessed 1 September 2016)
- Koes D.R., Camacho C. J. (2012) 'ZINCPharmer: pharmacophore search of the ZINC database', *Nucleic Acids Research*, (2012), 40 (W1):W409-W414. [Online]. DOI: 10.1093/nar/gks378 (Accessed 1 September 2016)
- Logie, L., Harthill, J., Patel, K., Bacon, S., Hamilton, D. L., Macrae, K., McDougall, G., Wang, H. H., Xue, L., Jiang, H., Sakamoto, K., Prescott, A. R. and Rena, G. (2012) 'Cellular responses to the metal-binding properties of metformin', *Diabetes*, 61(6), pp. 1423–1433, [Online]. DOI: 10.2337/db11-0961 (Accessed 1 September 2016)
- Maenpuen, S., Watthaisong, P., Supon, P., Sucharitakul, J., Parsonage, D., Karplus, P. A., Claiborne, A. and Chaiyen, P. (2015) 'Kinetic mechanism of l-α-glycerophosphate oxidase from

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mycoplasma pneumoniae', *FEBS Journal*, 282(16), pp. 3043–3059, [Online]. DOI: 10.1111/febs.13247 (Accessed 1 September 2016)

- Madiraju, A. K., Erion, D. M., Rahimi, Y., Zhang, X.-M., Braddock, D. T., Albright, R. a, Prigaro, B. J., Wood, J. L., Bhanot, S., MacDonald, M. J., Jurczak, M. J., Camporez, J.-P., Lee, H.-Y., Cline, G. W., Samuel, V. T., Kibbey, R. G. and Shulman, G. I. (2014) 'Metformin suppresses gluconeogenesis by inhibiting mitochondrial glycerophosphate dehydrogenase', *Nature*, 510(7506), pp. 542–546, [Online]. DOI:10.1038/nature13270 (Accessed 1 September 2016)
- Morris G., Huey R., Lindstrom W., Sanner M. F., Belew R. K., Goodsell D., Olson A. J. (2009) 'AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility', *Journal of Computational Chemistry, 2009 Dec; 30(16): 2785–2791.* [Online]. DOI: 10.1002/jcc.21256 (Accessed 1 September 2016)
- O'Boyle N.M., Banck M., James C.A., Morley C., Vandermeersch T., Hutchinson G. R. (2011) 'Open Babel: An open chemical toolbox', *Journal of Cheminformatics*, year 2011, 3:33, [Online]. DOI: 10.1186/1758-2946-3-33 (Accessed 1 September 2016)
- Onken, B. and Driscoll, M. (2010) 'Metformin Induces a Dietary Restriction–Like State and the Oxidative Stress Response to Extend C. elegans Healthspan via AMPK, LKB1, and SKN-1', *PLoS ONE*, 5(1), p. 8758, [Online]. DOI: 10.1371/journal.pone.0008758 (Accessed 1 September 2016)
- Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E. (2004) 'UCSF Chimera--a visualization system for exploratory research and analysis.', *Journal of Computational Chemistry*, 2004 Oct, 25(13), 1605-12. [Online]. DOI: 10.1002/jcc.20084 (Accessed 1 September 2016)
- Repiščák, P., Erhardt, S., Rena, G. and Paterson, M. J. (2014) 'Biomolecular mode of action of metformin in relation to its copper binding properties', *Biochemistry*, 53, pp. 787–795, [Online]. DOI:10.1021/bi401444n (Accessed 1 September 2016)
- Rojas, J., Chávez-castillo, M. and Torres, W. (2014) 'Metformin in Cancer: Chemical Pathways for Tumoral Control Independent of Amp- Dependent Kinase', *Journal of Endocrinology, Diabetes and Obesity*, 2(2): 1036, [Online]. ISSN: 2333-6692, Available from: <https://www.jscimedcentral.com/Endocrinology/endocrinology-spid-antidiabetic-drug-metformin-1036.pdf> (Accessed 1 September 2016)
- Royal Society of Chemistry. (2016) ChemSpider. [Online]. Available from: <http://www.chemspider.com/> (Accessed 1 September 2016)
- Shaw R.J., Vasquez D., Koo S.H., Bardeesy N., Depinho R.A., Montminy M., Cantley L.C. (2005) 'The kinase LKB1 mediates glucose homeostasis in liver and therapeutic effects of metformin', *Science*, 310, pp. 1642–1646, [Online] DOI: 10.1126/science.1120781 (Accessed 1 September 2016)
- Smith R. M. and Martell A. E. (2004) 'NIST Critically Selected Stability Constants of Metal Complexes Database', *National Institute of Standards and Technology*, [Online]. Available from: <http://www.nist.gov/srd/nist46.cfm> (Accessed 1 September 2016)
- Strelow J., Dewe W., Iversen P.W., Brooks H.B., Radding J.A., McGee J., Weidner J. (2012), 'Mechanism of Action Assays for Enzymes'. *Assay Guidance Manual*, [Internet]. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004-.Available from: http://www.ncbi.nlm.nih.gov/books/NBK92001/
- The UniProt Consortium. (2015) 'UniProt: a hub for protein information', *Nucleic Acids Research*, Vol. 43, No. 1, 2015, D204-D212, [Online]. DOI: 10.1093/nar/gku989 (Accessed 1 September 2016)
- Trott O., Olson A. J. (2010) 'AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading.', *Journal of Computational Chemistry*, 31, pp. 455-461, [Online]. DOI: 10.1002/jcc.21334 (Accessed 1 September 2016)
- Wishart D.S., Knox C., Guo A.C., Shrivastava S., Hassanali M., Stothard P., Chang Z., Woolsey J. (2006) 'DrugBank: a comprehensive resource for in silico drug discovery and exploration.', *Nucleic Acids Research*, vol. 34 No. 1, 2006, D668-72 [Online]. DOI: 10.1093/nar/gkj067 (Accessed 1 September 2016)
- Yang J., Yan R., Roy A., Xu D., Poisson D., Zhang Y. (2015) 'The I-TASSER Suite: Protein structure and function prediction.', Nature Methods, 12: 7-8 (2015). [Online]. DOI: 10.1038/nmeth.3213 (Accessed 1 September 2016)
- Yeh, J. I., Chinte, U. and Du, S. (2008) 'Structure of glycerol-3-phosphate dehydrogenase, an essential monotopic membrane enzyme involved in respiration and metabolism.', *Proceedings of the National Academy of Sciences of the United States of America*, 105(9), pp. 3280–5, [Online]. DOI: 10.1073/pnas.0712331105 (Accessed 1 September 2016)

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

- Zhou G., Li Y., Chen Y., Shen X., Fenyk-Melody J., Wu M., Ventre J., Doebber T., Fujii N., Musi N., Hirshman M.F., Goodyear L.J., Moller D.E. (2001) 'Role of AMP-activated protein kinase in mechanism of metformin action. ', *Journal of Clinical Investigation*, 108, pp. 1167–1174, [Online]. DOI: 10.1172/JCI13505 (Accessed 1 September 2016)

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

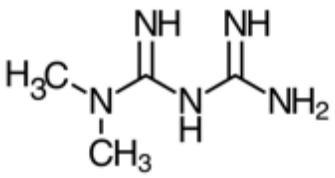


Figure 1.1 – Structural formula of metformin

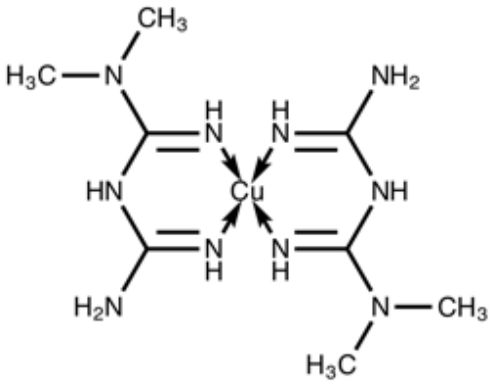
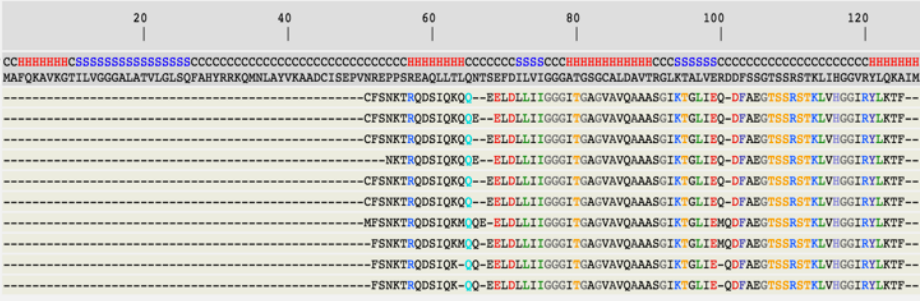


Figure 1.2 – Structural formula of a metformin Cu^{2+} copper complex

Top 10 threading templates used by I-TASSER

(I-TASSER modeling starts from the structure templates identified by LOMETS from the PDB library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of template alignments. I-TASSER only uses the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. The templates in this section are the 10 best templates selected from the LOMETS threading programs. Usually, one template of the highest Z-score is selected from each threading program, where the threading programs are sorted by the average performance in the large-scale benchmark test experiments.)

Rank	PDB Hit	Iden1	Iden2	Cov	Norm. Z-score	Download Align.
1	user	0.30	0.25	0.71	10.00	Download
2	2rghA	0.30	0.25	0.71	2.55	Download
3	2rghA	0.29	0.25	0.72	2.73	Download
4	2rghA	0.29	0.25	0.71	4.12	Download
5	2rghA	0.30	0.25	0.71	5.23	Download
6	2rghA	0.28	0.25	0.72	5.83	Download
7	2rgoA	0.30	0.25	0.73	4.22	Download
8	2rgoA	0.30	0.25	0.73	6.00	Download
9	2rghA	0.30	0.25	0.71	3.65	Download
10	2rghA	0.30	0.25	0.71	4.08	Download



(a) All the residues are colored in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color. Coloring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are colored in dark shade. (more about the colors used)

(b) Rank of templates represents the top ten threading templates used by I-TASSER.

(c) Ident1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.

(d) Ident2 is the percentage sequence identity of the whole template chains with query sequence.

(e) Cov represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.

(f) Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.

(g) Download Align. provides the 3D structure of the aligned regions of the threading templates.

(h) The top 10 alignments reported above (in order of their ranking) are from the following threading programs:
1: User 2: MUSTER 3: wdPPAS 4: Neff-PPAS 5: SP3 6: FFAS03 7: SPARKS-X 8: FFAS03 9: HHSEARCH I 10: HHSEARCH

Figure 2.1 – The correspondence between the GPD2 residues and the residues of 2RGH, the best threading template in I-Tasser

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

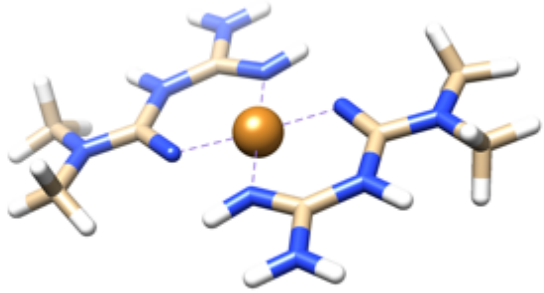


Figure 2.2 – A ligand modeling example: the dimetformin neutral copper complex visualised in Chimera

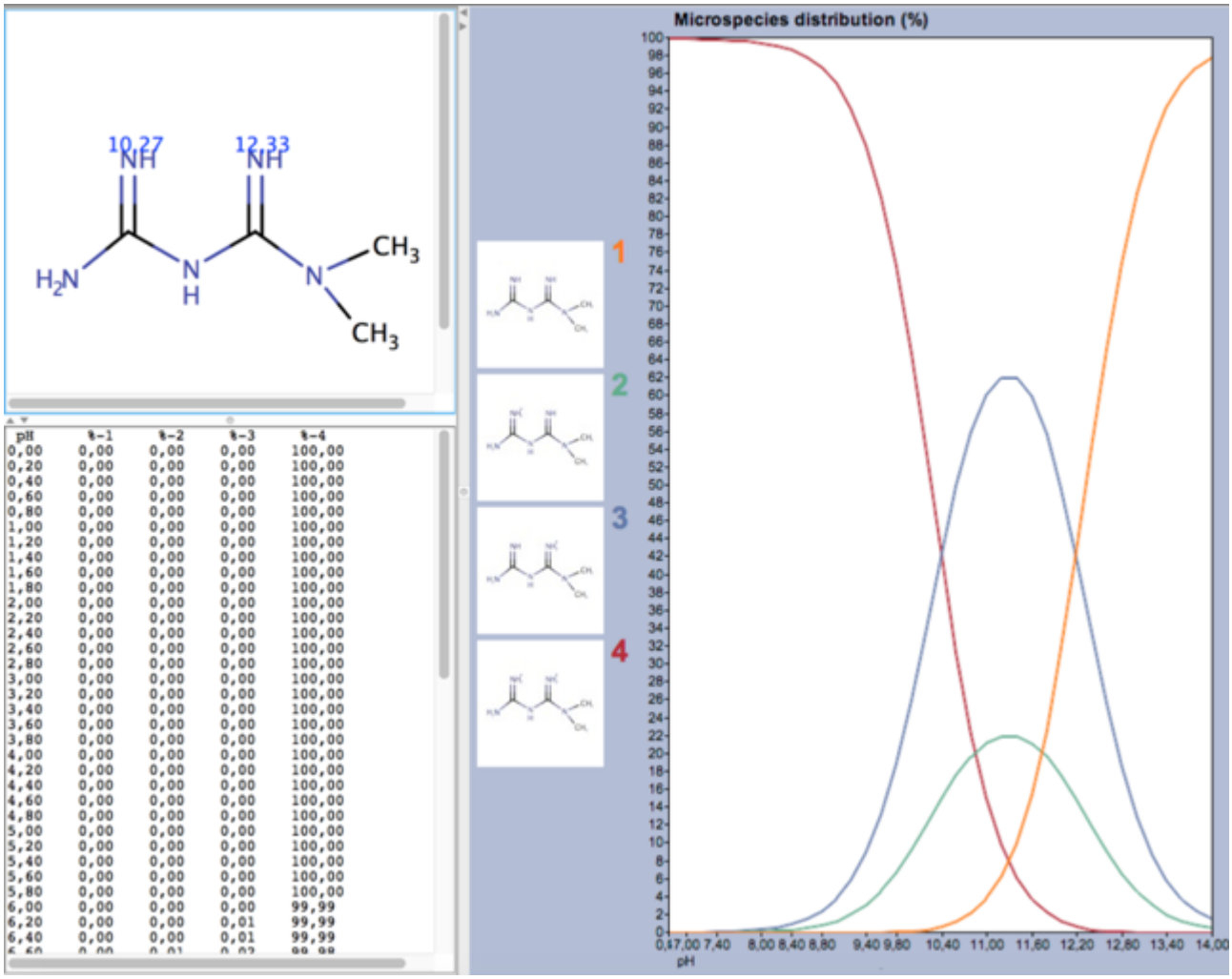


Figure 2.3 – An example of speciation curves obtained by MarvinBeans, in this case for metformin; species 1 is non-protonated, species 2 and 3 are mono-protonated, species 4 is di-protonated.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

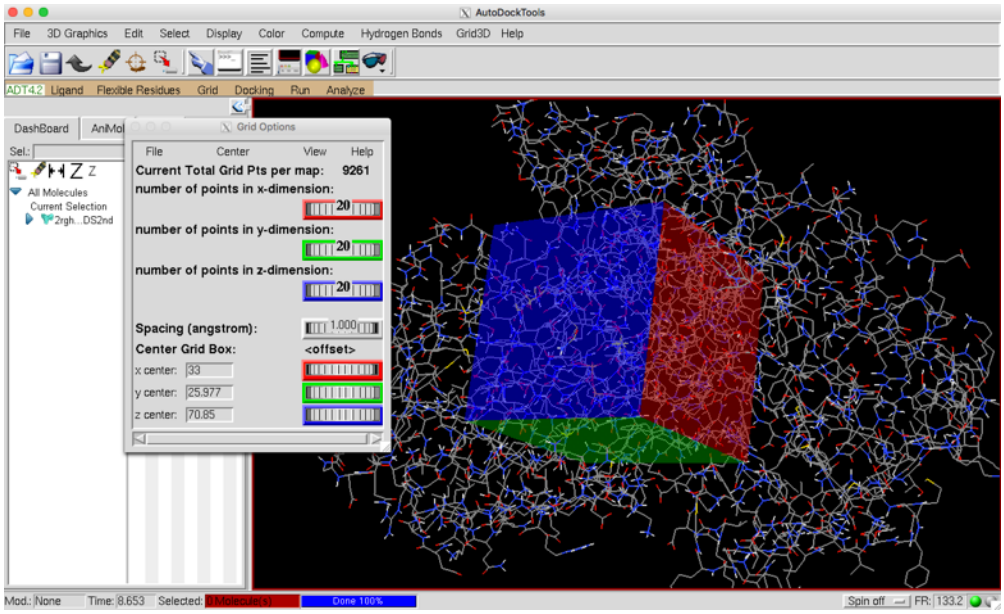


Figure 2.4 – The set up of a coordinates grid (Grid Box) on Autodock Tools to delimit the docking area

```
MGLTools-1.5.6 — -bash — 70×44
MBP-di-Fabio:MGLTools-1.5.6 fabiorebecchi$ ./vina --receptor 2rghFADG3
PFMSEsubGDP22ndIN118_1.pdbqt --ligand phenformincopperdouble.pdbqt --c
enter_x 33.4 --center_y 25.8 --center_z 74.7 --size_x 20 --size_y 20 -
-size_z 20 --out docking158.pdbqt
#################################################################
# If you used AutoDock Vina in your work, please cite:          #
#                                                               #
# O. Trott, A. J. Olson,                                        #
# AutoDock Vina: improving the speed and accuracy of docking    #
# with a new scoring function, efficient optimization and       #
# multithreading, Journal of Computational Chemistry 31 (2010)  #
# 455-461                                                       #
#                                                               #
# DOI 10.1002/jcc.21334                                         #
#                                                               #
# Please see http://vina.scripps.edu for more information.      #
#################################################################

Detected 8 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 370544877
Performing search ...
0%   10   20   30   40   50   60   70   80   90   100%
|----|----|----|----|----|----|----|----|----|----|
***************************************************
done.
Refining results ... done.

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1        -10.6      0.000      0.000
   2        -10.1      0.625      9.906
   3         -8.9      3.826      5.178
   4         -8.9      3.825     10.488
   5         -8.9      4.180      5.805
   6         -8.7      4.073      6.627
   7         -8.4      2.336     10.337
   8         -8.4      2.145     10.126
   9         -8.1      3.991     10.212
Writing output ... done.
```

Figure 2.5 – An example of input/output screen for Autodock Vina inside the command line shell

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

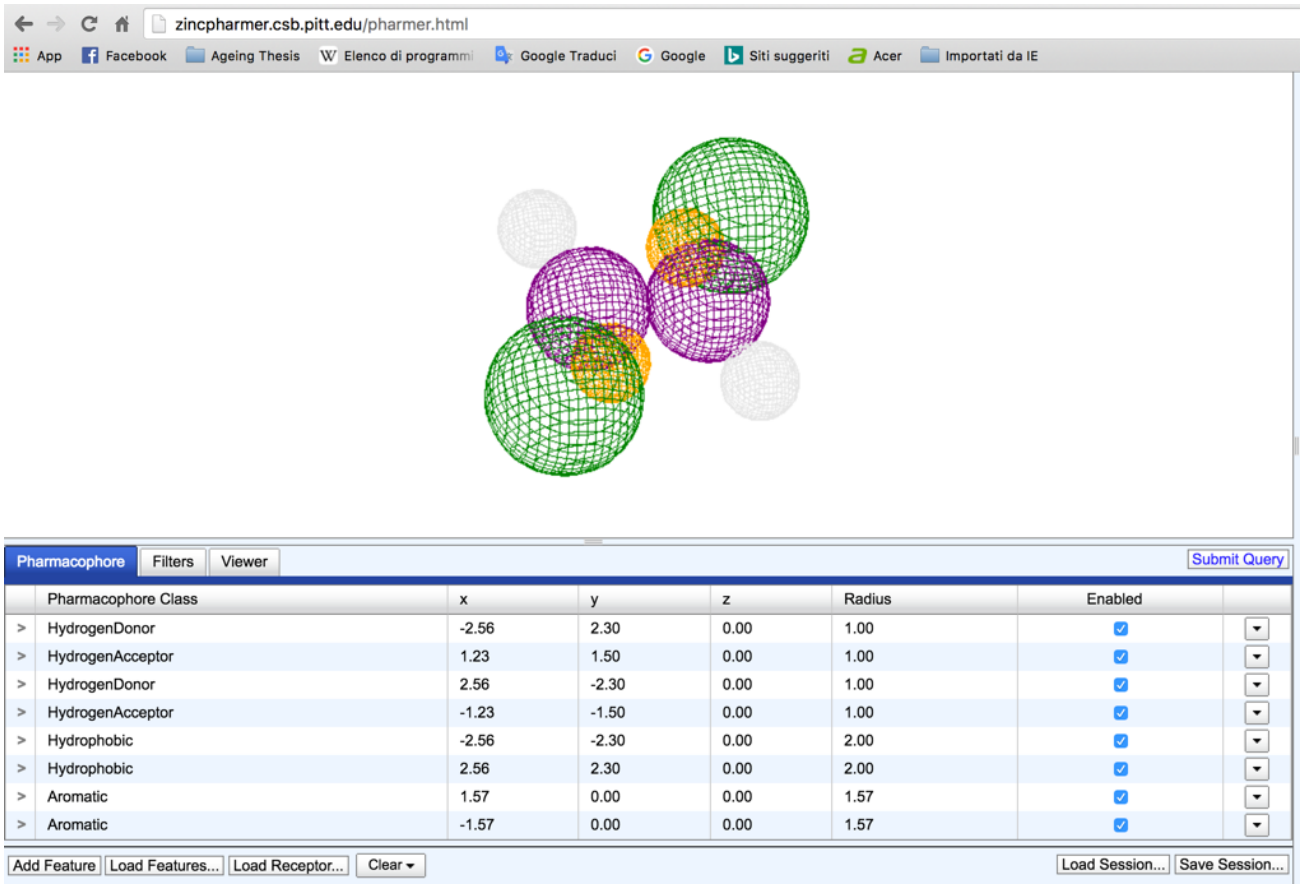


Figure 2.6 – Pharmacophore modeling interface of the ZincPharmer website

Residue's number	24	25	26	27	28	29	47	48	49	50	55	56	57	59	60	61	63	64	65	68	69	70	78	79	195	196	197	198	231	232	233	234	235
2RGH	I	G	G	G	I	T	I	E	M	Q	G	T	S	R	S	T	L	V	H	I	R	Y	V	V	M	K	A	V	T	S	G	P	W
GDP2	.	.	.	.	A	.	V	.	.	R	.	.	.	.	.	.	.	I	.	V	.	.	M	.	.	M	E	.	A	T	.	.	F

Residue's number	236	237	238	239	240	241	242	255	256	257	258	259	274	275	276	277	282	283	284	285	298	300	301	344	345	346	347	348	349	428	429	430	431
2RGH	V	D	K	V	R	N	L	T	K	G	I	H	Y	F	D	T	G	R	M	V	T	D	T	G	L	R	P	L	L	G	K	I	T
GDP2	.	.	S	.	.	.	M	S	A	.	V	.	L	D	.	A	R	V	.	I	.	.	.	.	I	.	.	.	V	.	.	.	.

	Inner cavity	19/29 = 66% residues in common
	Small entrance	10/16 = 63% residues in common
	Large entrance	13/21 = 62% residues in common
*	Total	42/66 = 64% residues in common

Figure 3.1 – Cavity's residues comparision between 2RGH and GPD2

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

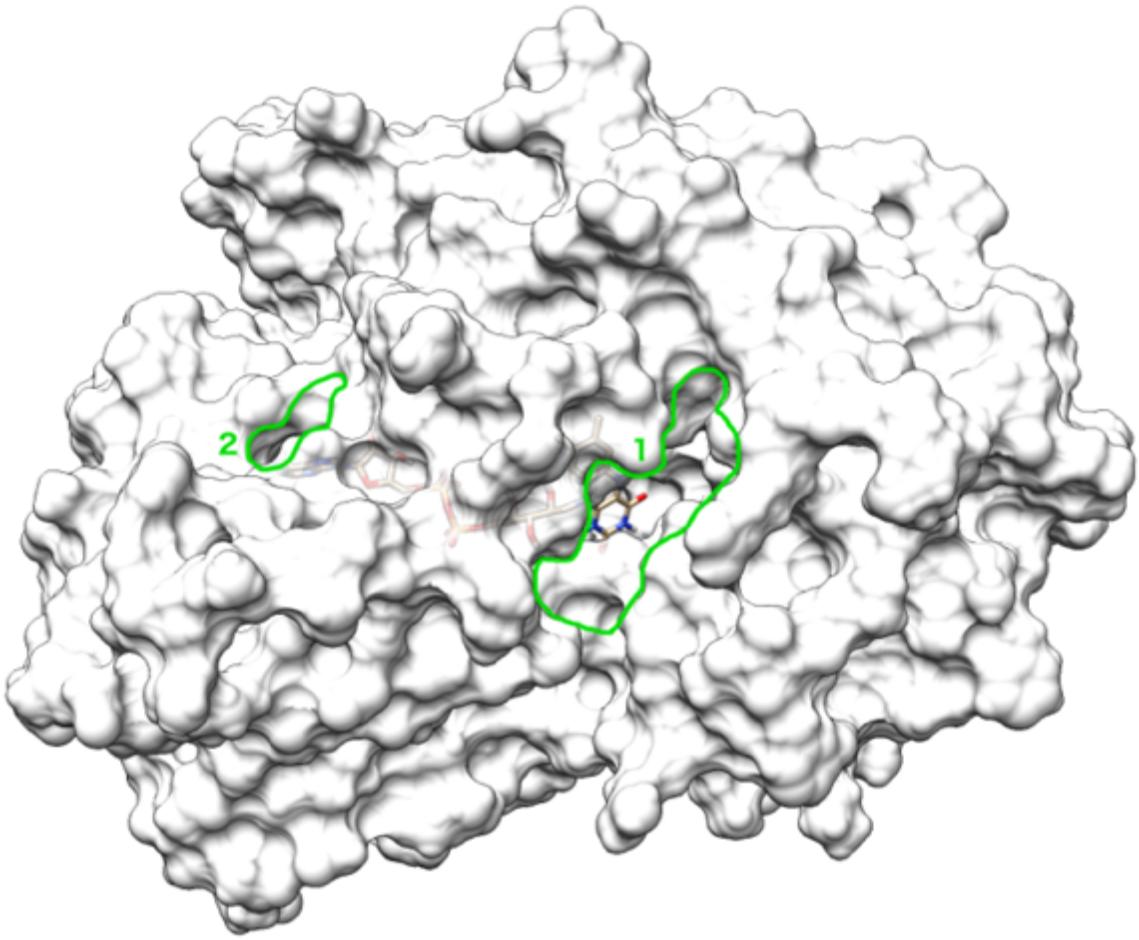


Figure 3.2 – The newly modeled cavity structure with docked FAD, the edges of the small (2) and large (1) entrances are highlighted in green. The white hydrophobicity surface has 50% transparency, showing the FAD body inside.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

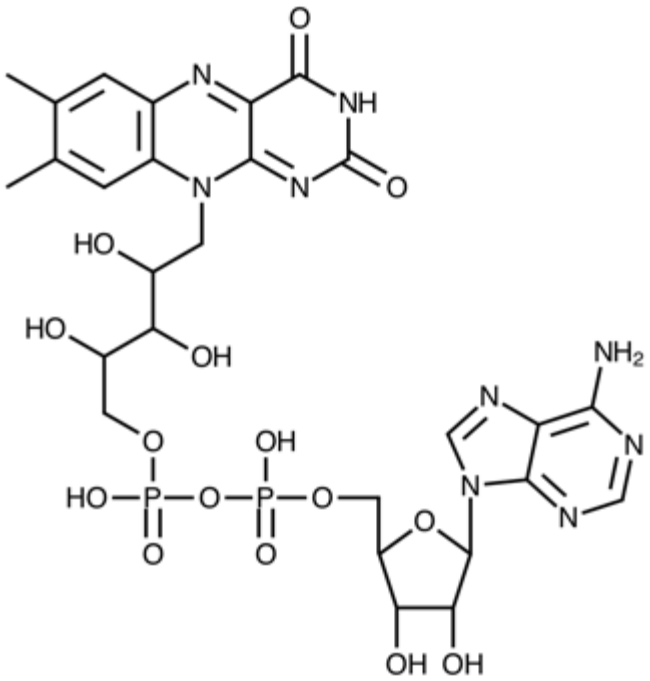


Figure 1.3 – The structural formula of flavin adenine dinucleotide (FAD)

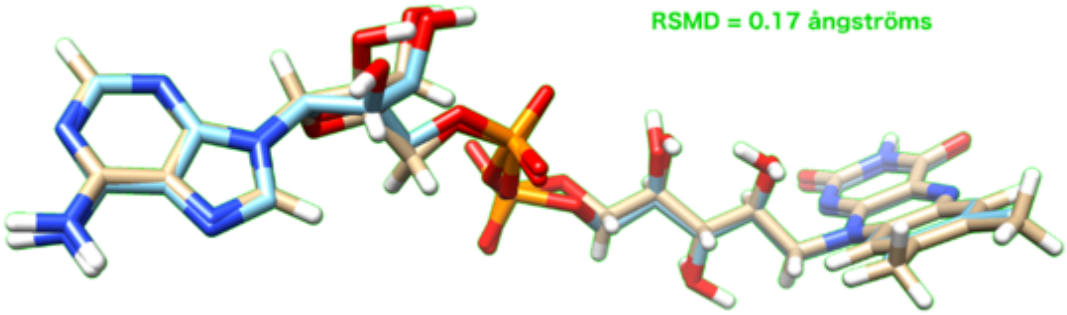


Figure 3.4 – Root mean square deviation of atomic positions between original 2RGH crystal structure FAD and docked FAD in the newly modeled GPD2 cavity.

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

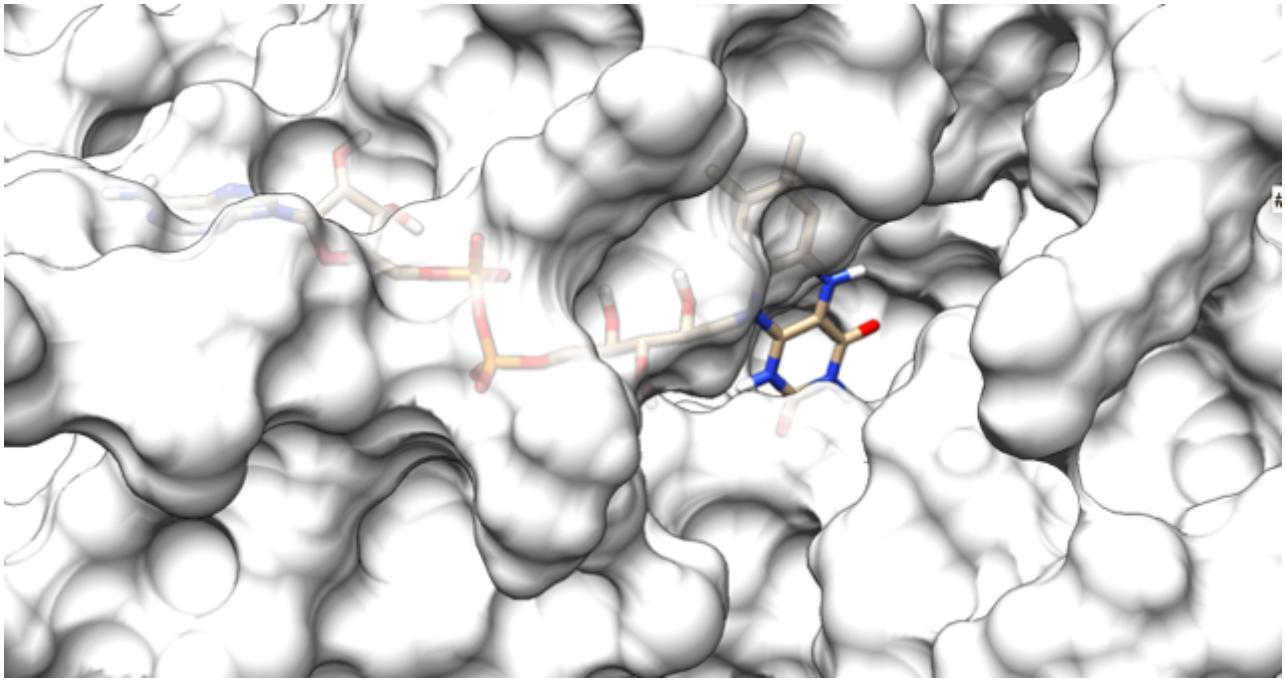


Figure 3.5 – The fully slid in FAD's pose in the cavity, corresponding to a free binding energy of -14.9 kcal/mol

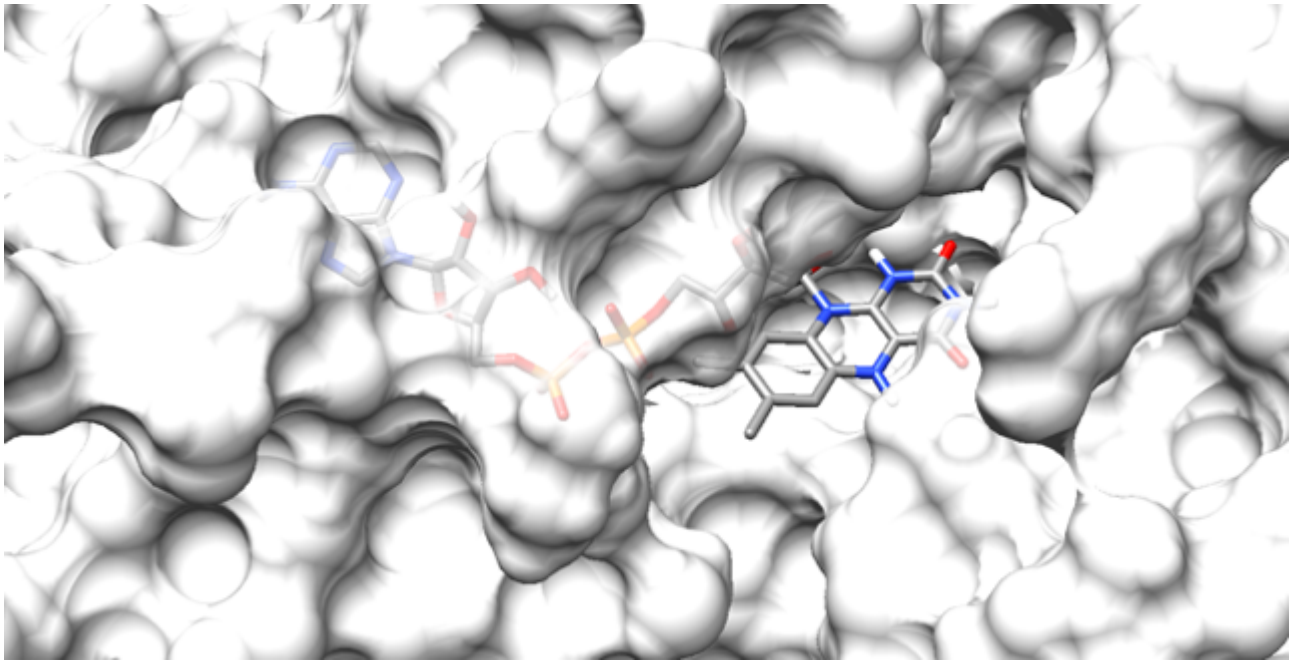


Figure 3.6 – The slightly slid out FAD's pose in the cavity, corresponding to a free binding energy of -12.6 kcal/mol

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

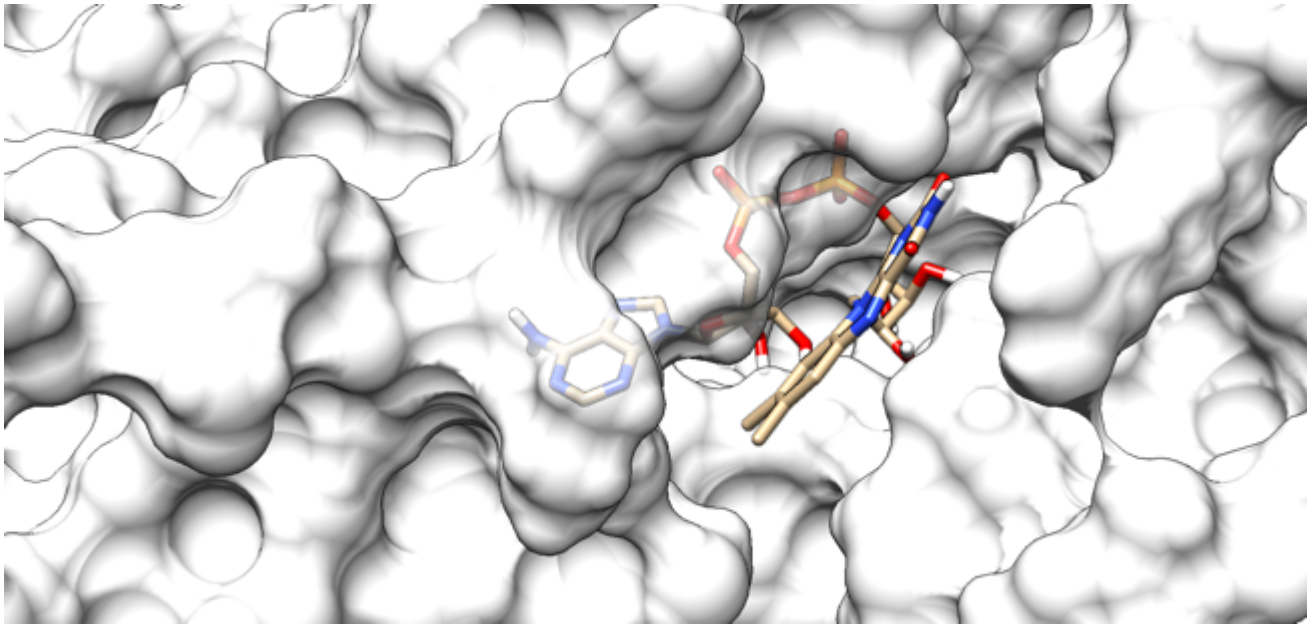


Figure 3.7 – The significantly slid out FAD's pose, corresponding to a free binding energy of -11.9 kcal/mol

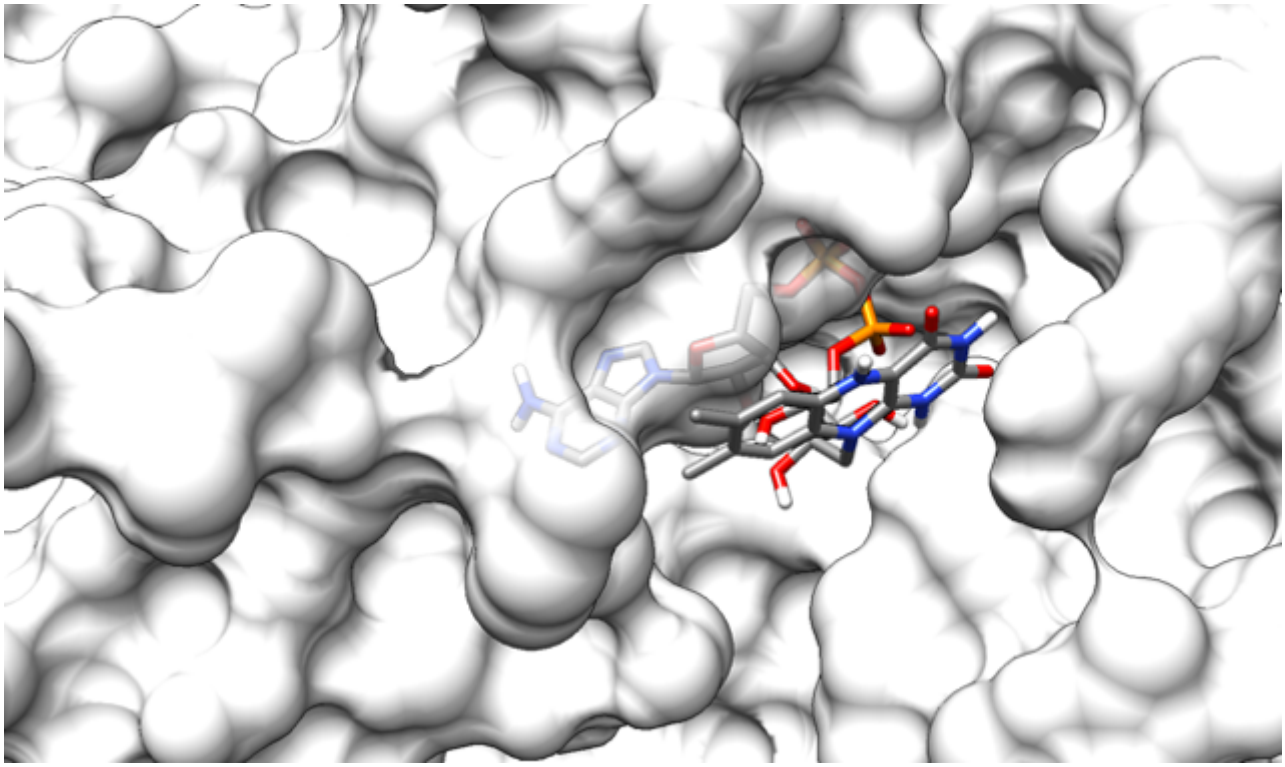


Figure 3.8 – The extremely slid out FAD's pose in the cavity, corresponding to a free binding energy of -10.6 kcal/mol

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

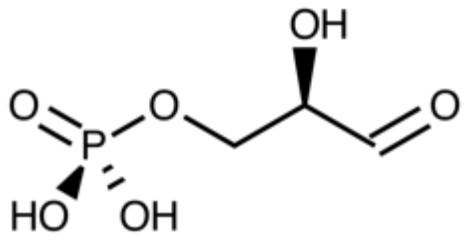


Figure 3.9 – Structural formula of glyceraldehyde 3-phosphate (G3P)

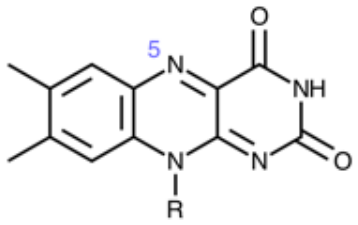


Figure 3.10 – The isoallixine ring of FAD, showing the nitrogen in position 5

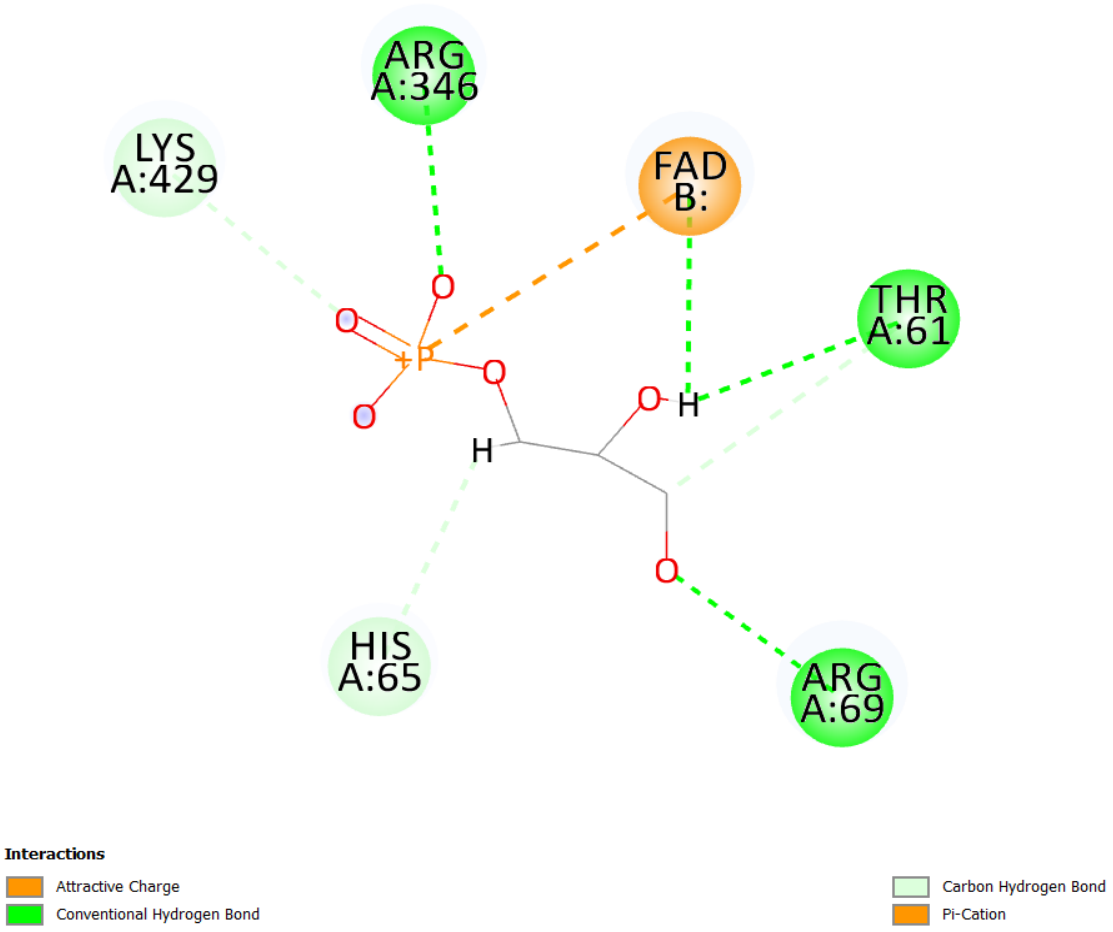


Figure 3.11 – Interaction diagram describes the bonds formed by G3P in the cavity with the fully slid in FAD

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

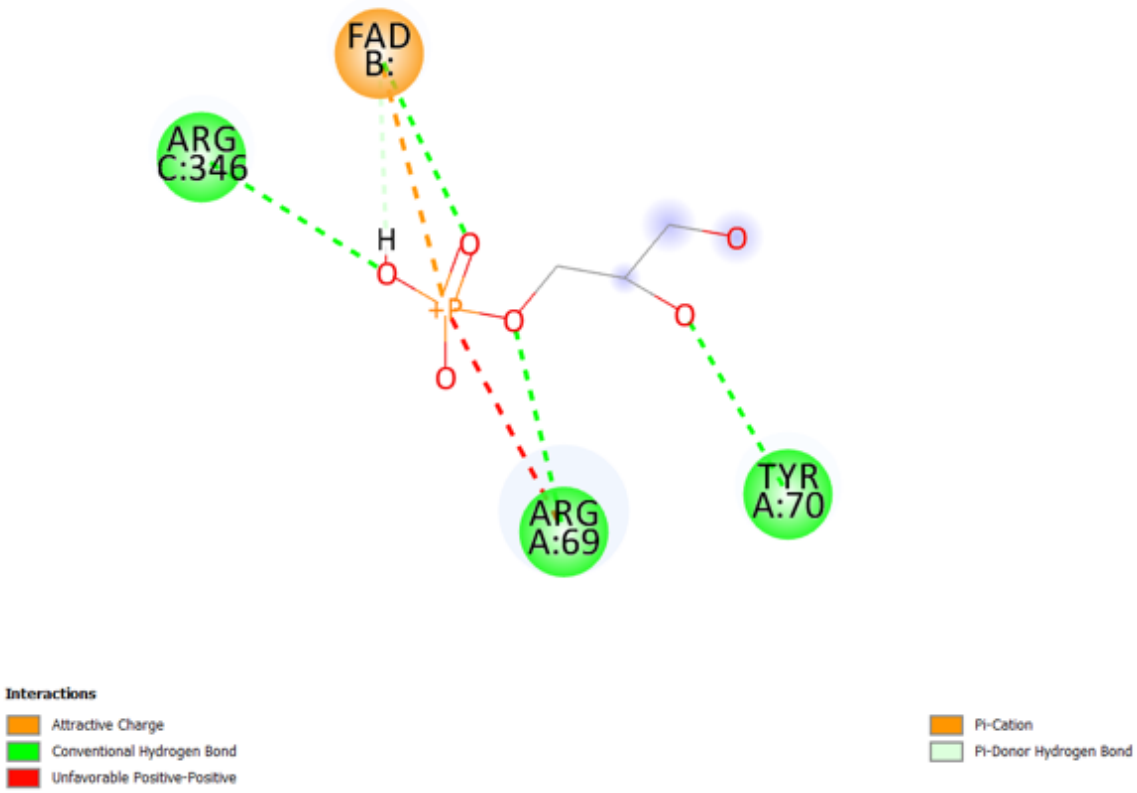


Figure 3.12 –Interaction diagram describes the bonds formed by G3P in the cavity with the slightly slid out FAD

bioRxiv preprint doi: https://doi.org/10.1101/130211; this version posted April 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

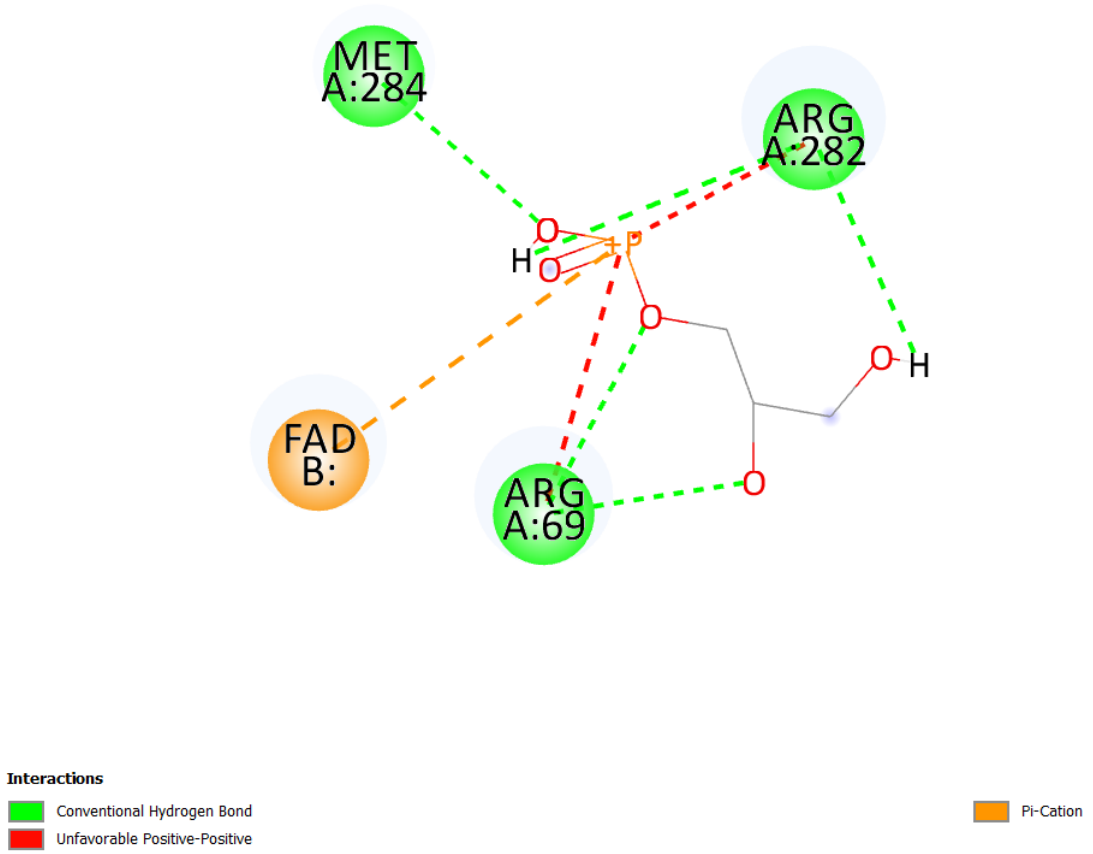


Figure 3.13 –Interaction diagram describes the bonds formed by G3P in the cavity with the significantly slid out FAD

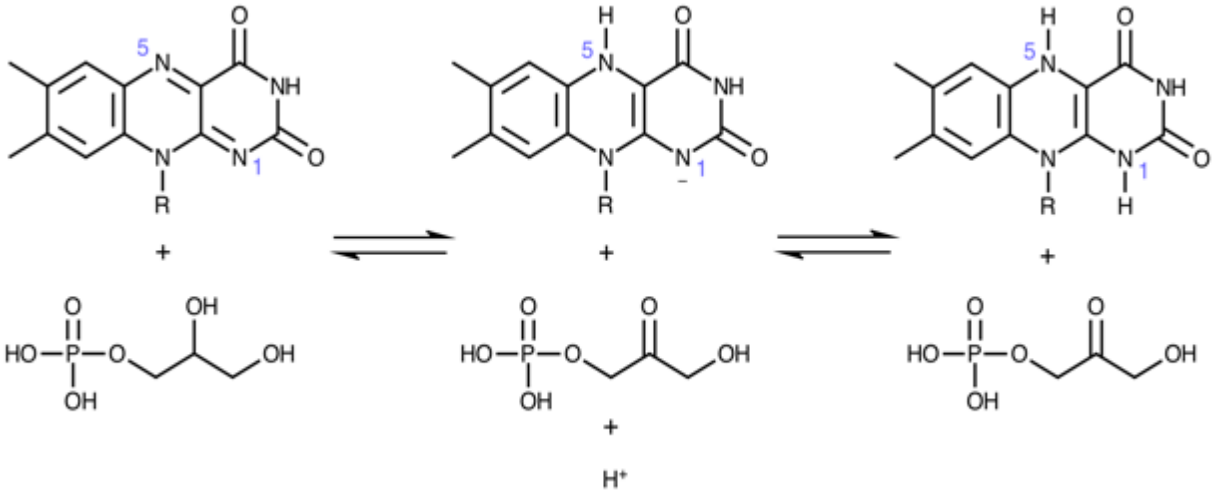


Figure 3.14 – The two steps of the FAD reduction process by glyceraldehyde 3-phosphate (G3P)

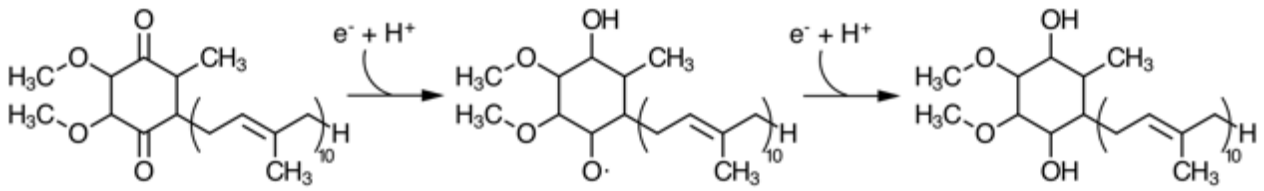


Figure 3.15 – Coenzyme Q10 and its reduction mechanism from left to right

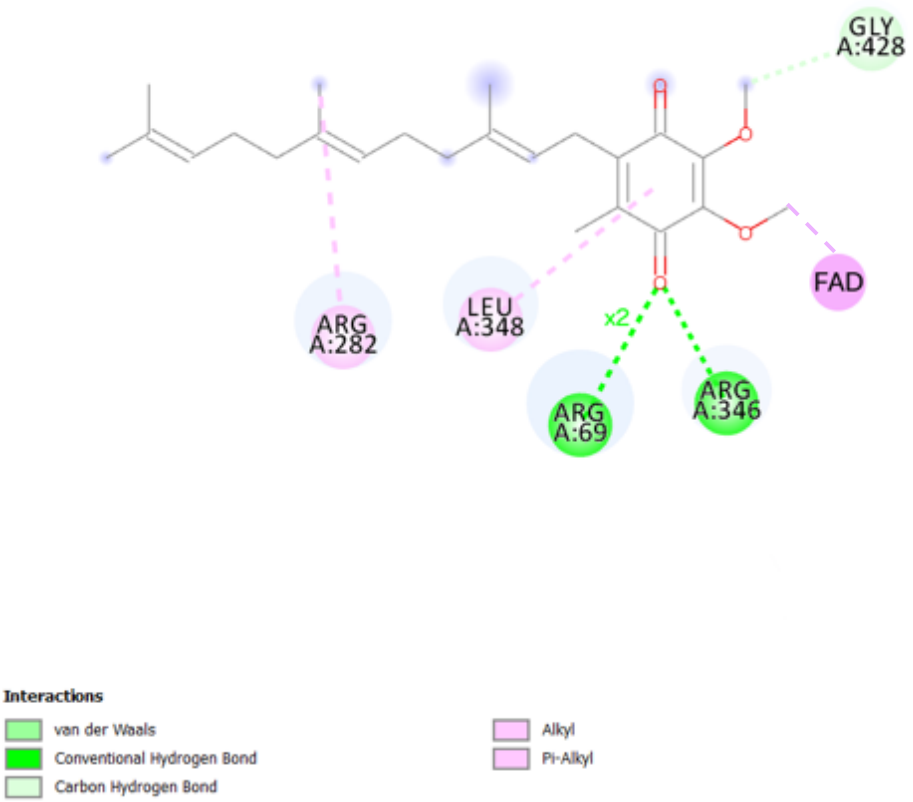


Figure 3.16 – Coenzyme Q10 and its cavity interactions in a 2D representation when FAD is fully slid in

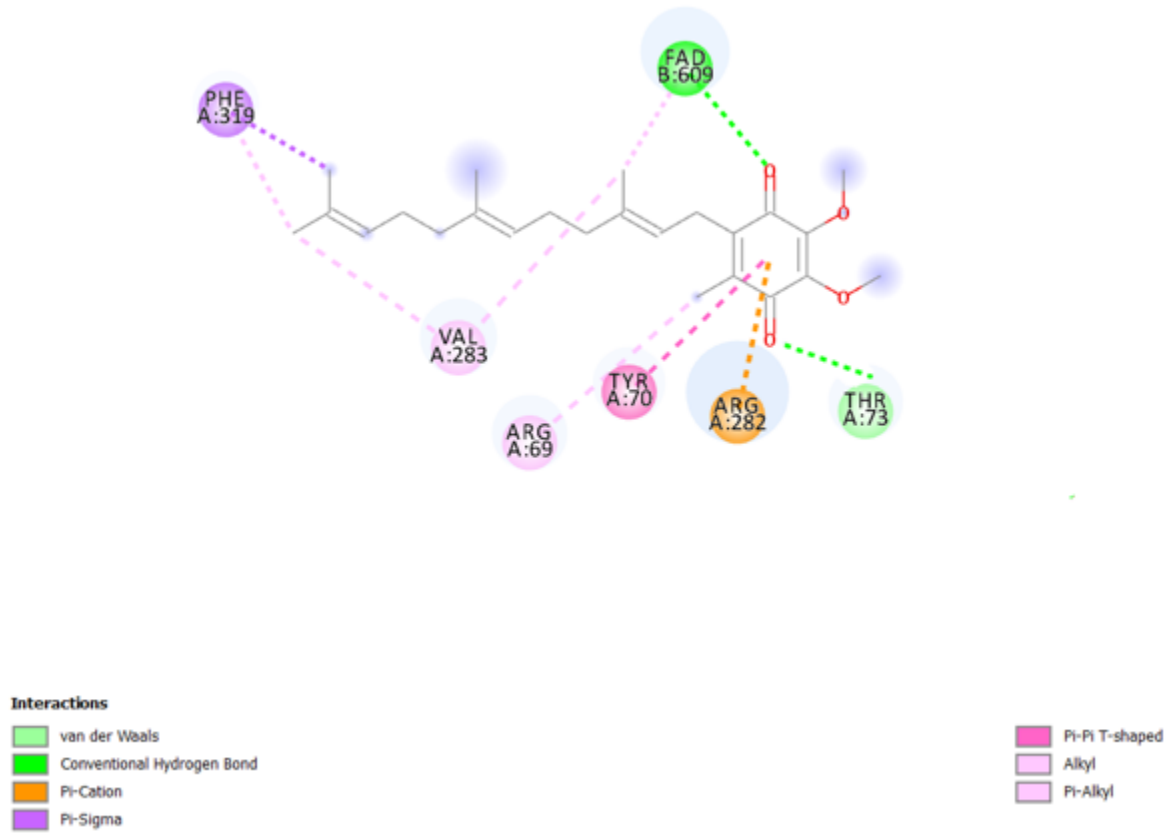


Figure 3.17 – Coenzyme Q10 and its cavity interactions in a 2D representation when FAD is slightly slid out

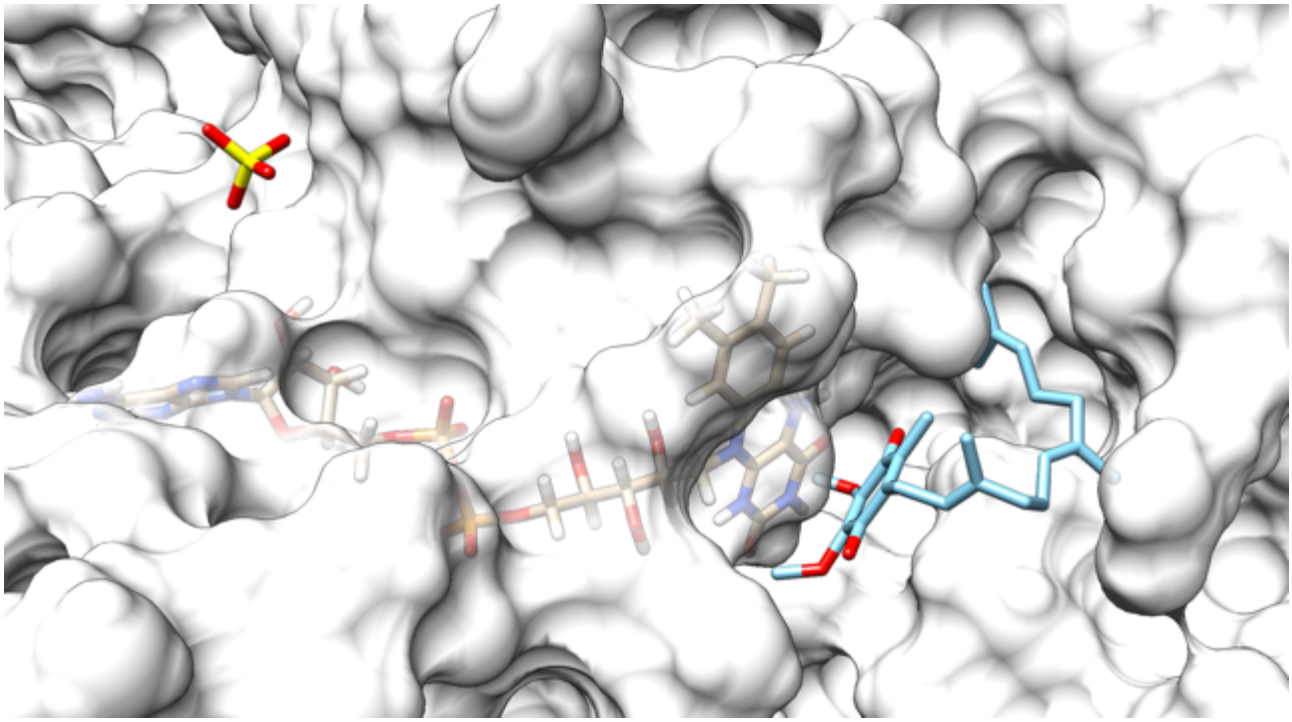


Figure 3.18 – Coenzyme Q10 pose while FAD is fully slid in, having a free binding energy of -6.1 kcal/mol

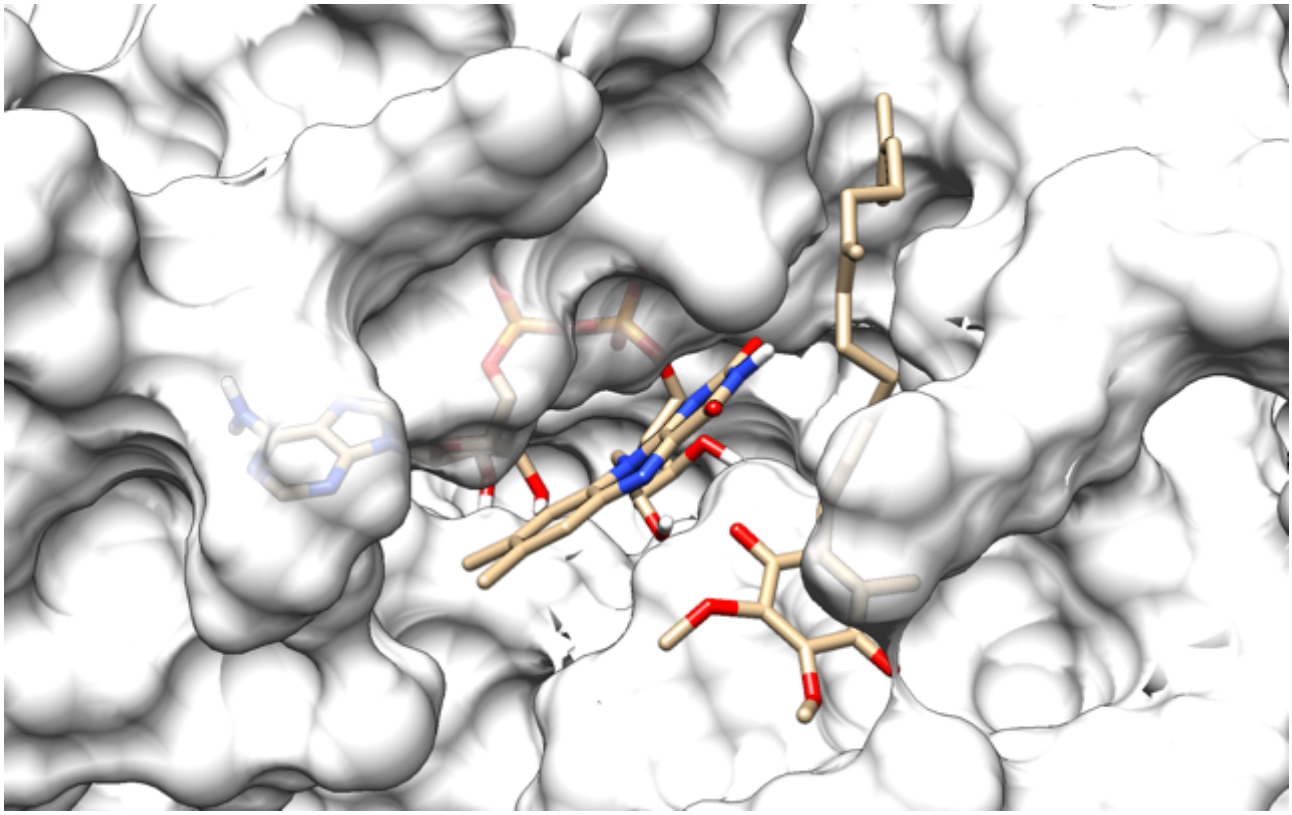


Figure 3.19 – Coenzyme Q10 pose while FAD is slightly slid out, having a free binding energy of -6.6 kcal/mol

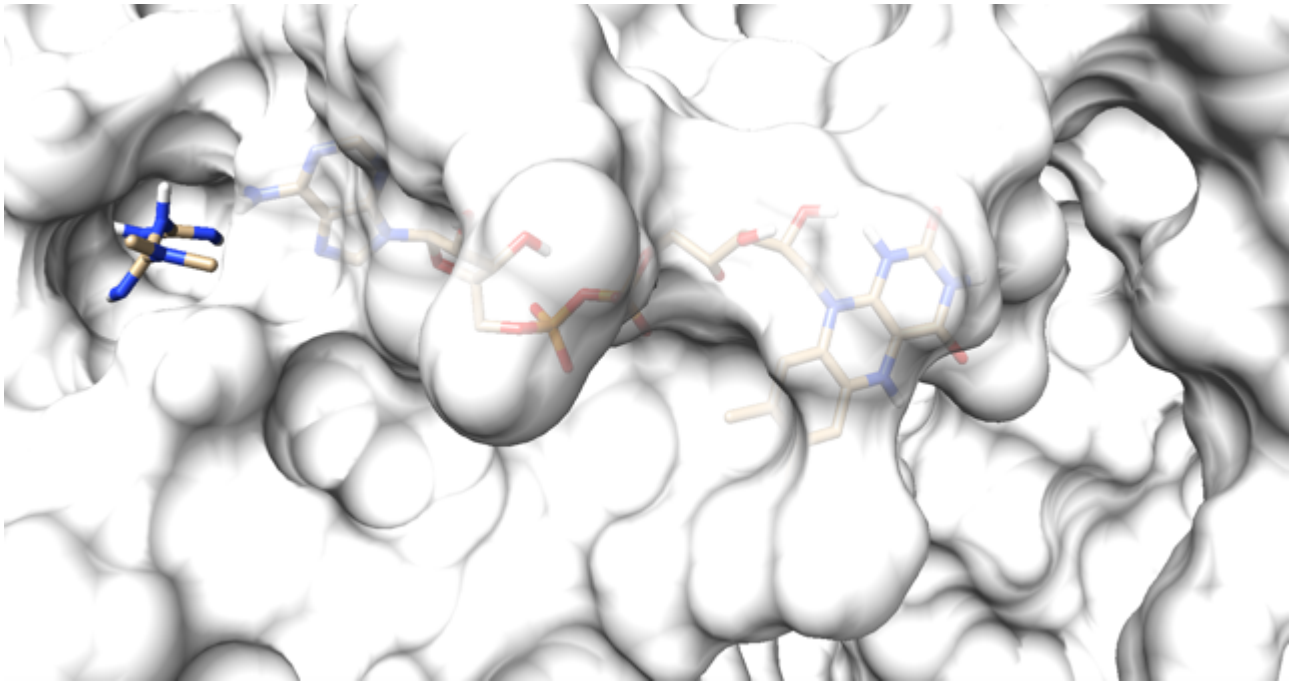


Figure 3.20 – Metformin pose while FAD is slightly slid out, having a free binding energy of -5.0 kcal/mol

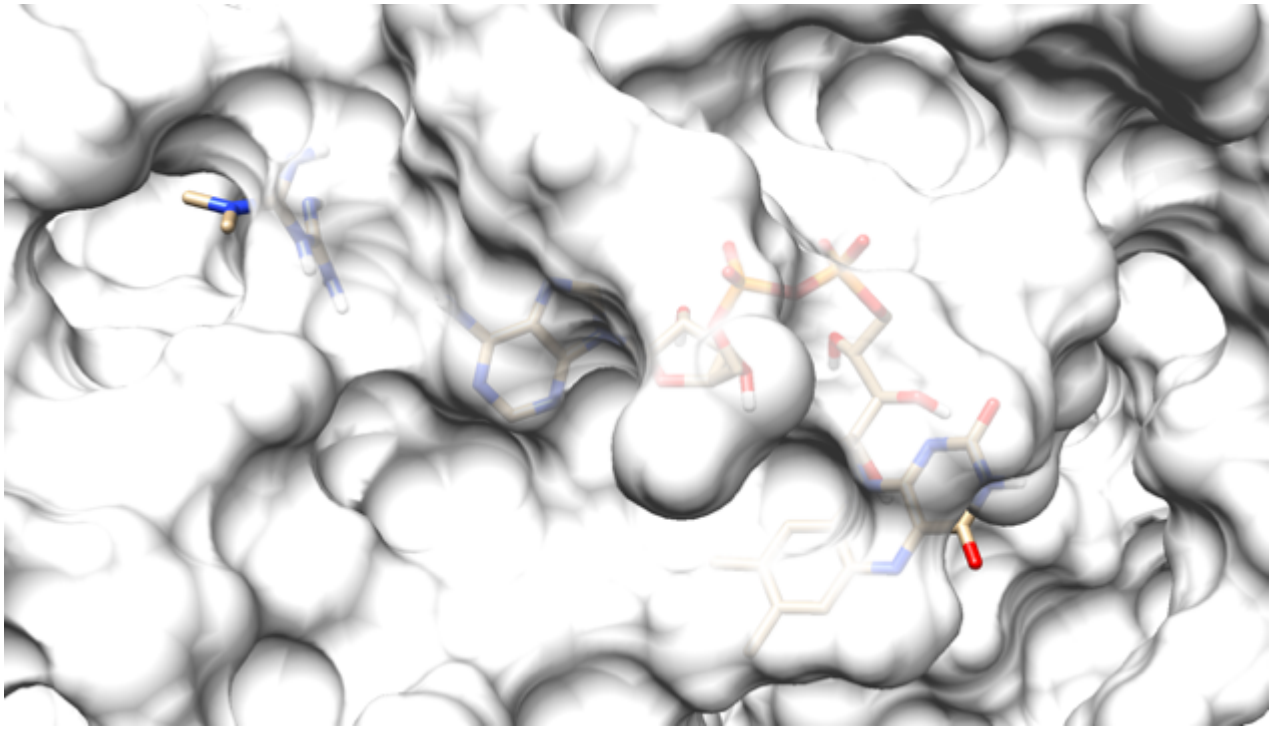


Figure 3.21 – Metformin pose while FAD is significantly slid out, having a free binding energy of -5.3 kcal/mol

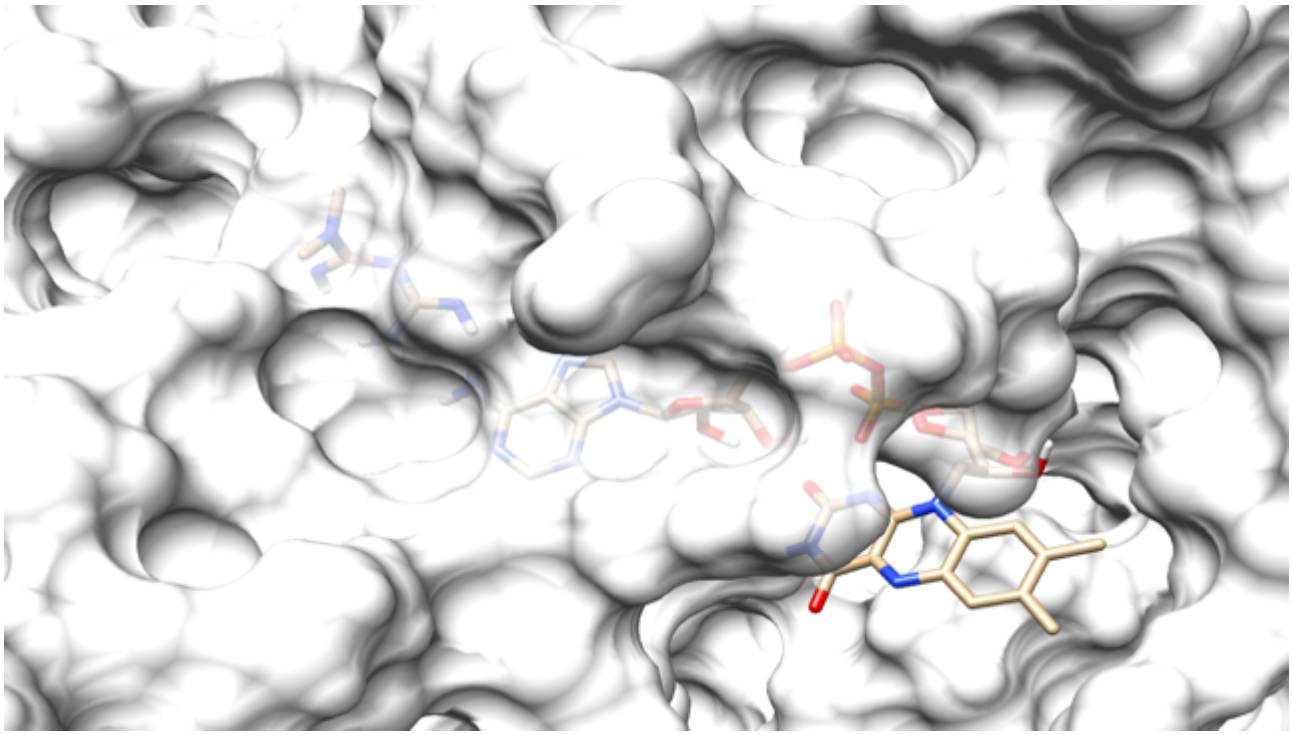


Figure 3.22 – Metformin pose while FAD is extremely slid out, having a free binding energy of -5.8 kcal/mol

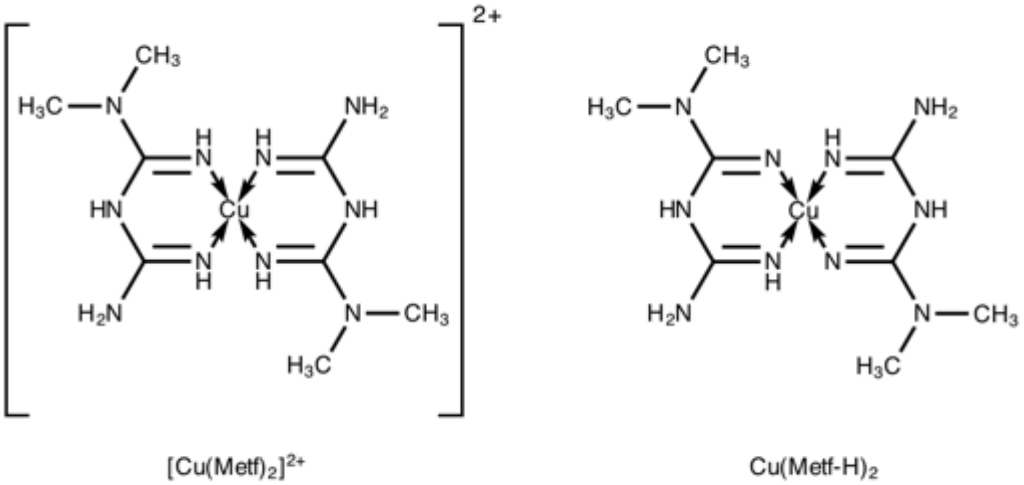


Figure 3.23 – The cationic (left) and neutral (right) dimetformin copper complexes

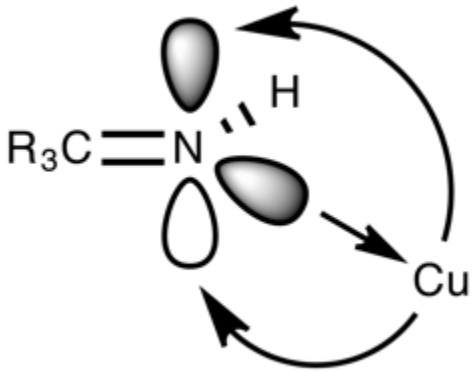


Figure 3.24 – Representation the of the copper coordination through a σ-donation and a π-backdonation

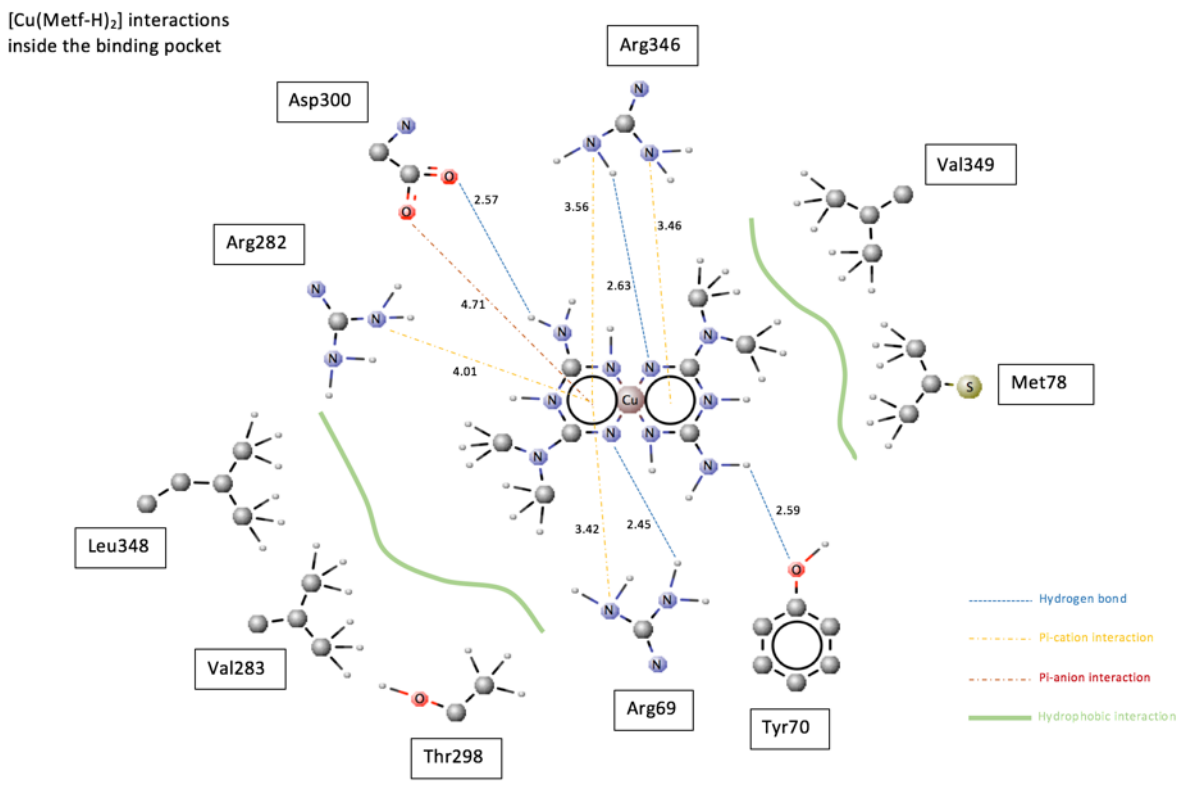


Figure 3.25 – [Cu(Metf-H)$_2$] interactions with cavity residues, bond lengths are expressed in angstroms

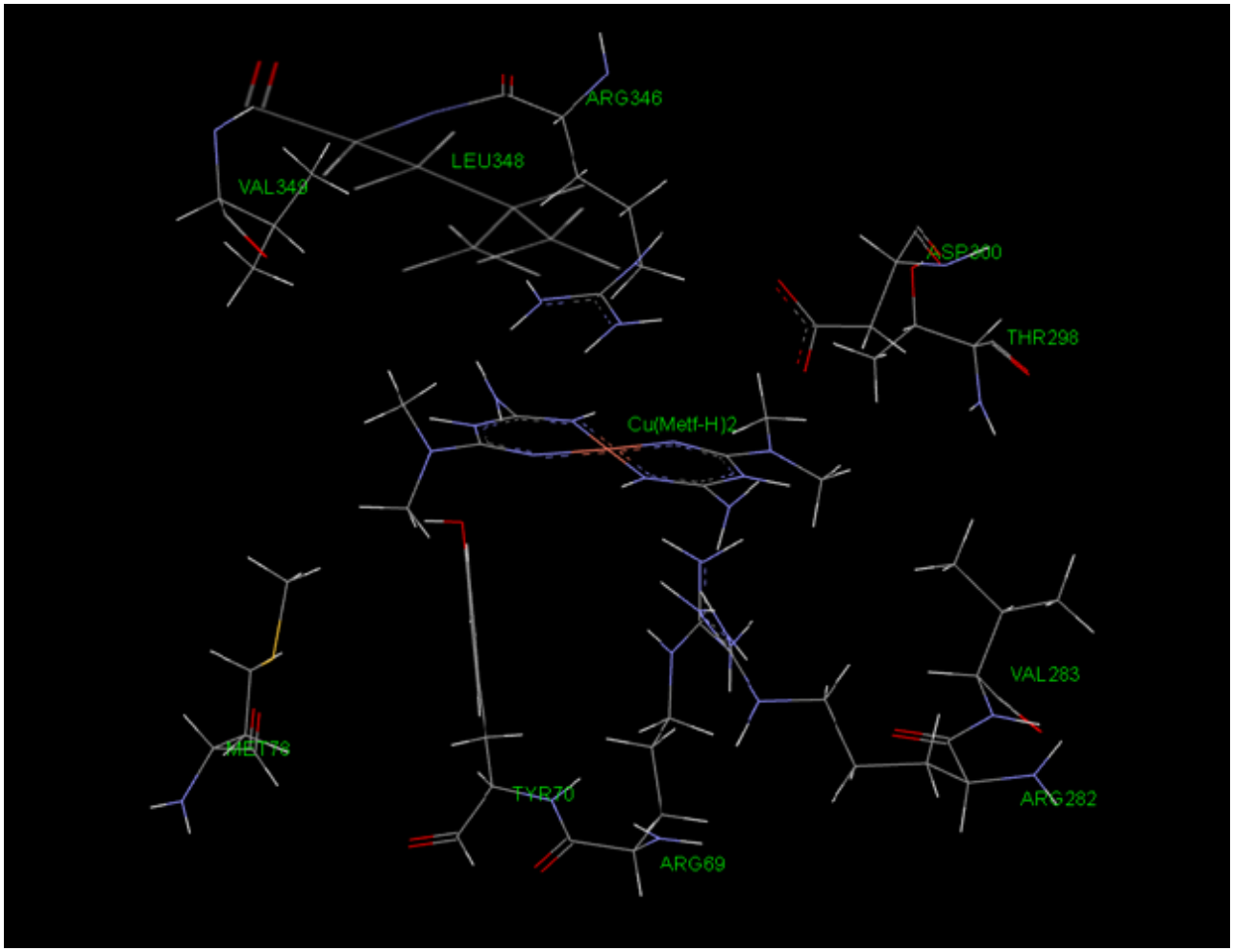


Figure 3.26 – [Cu(Metf-H)$_2$] position relative to cavity residues

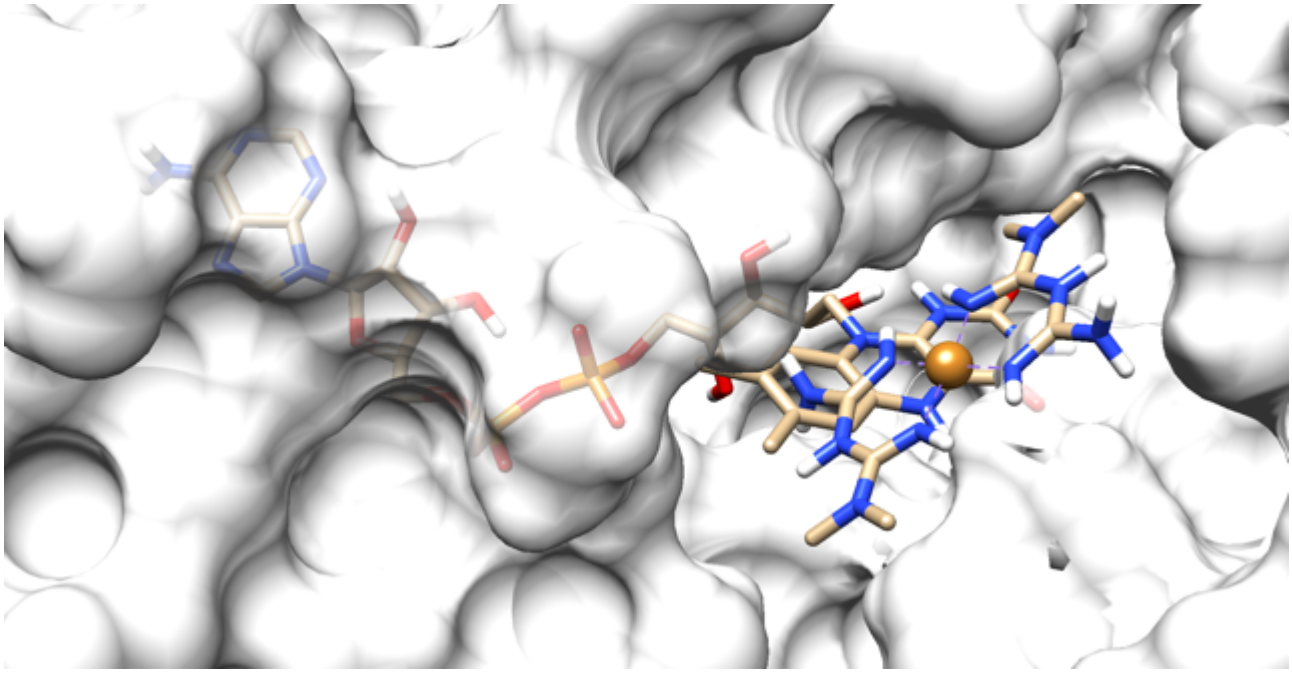


Figure 3.27 – [Cu(Metf-H)$_2$] collision with slightly slid out FAD, GPD2 is shown as a 40% transparent HF surface

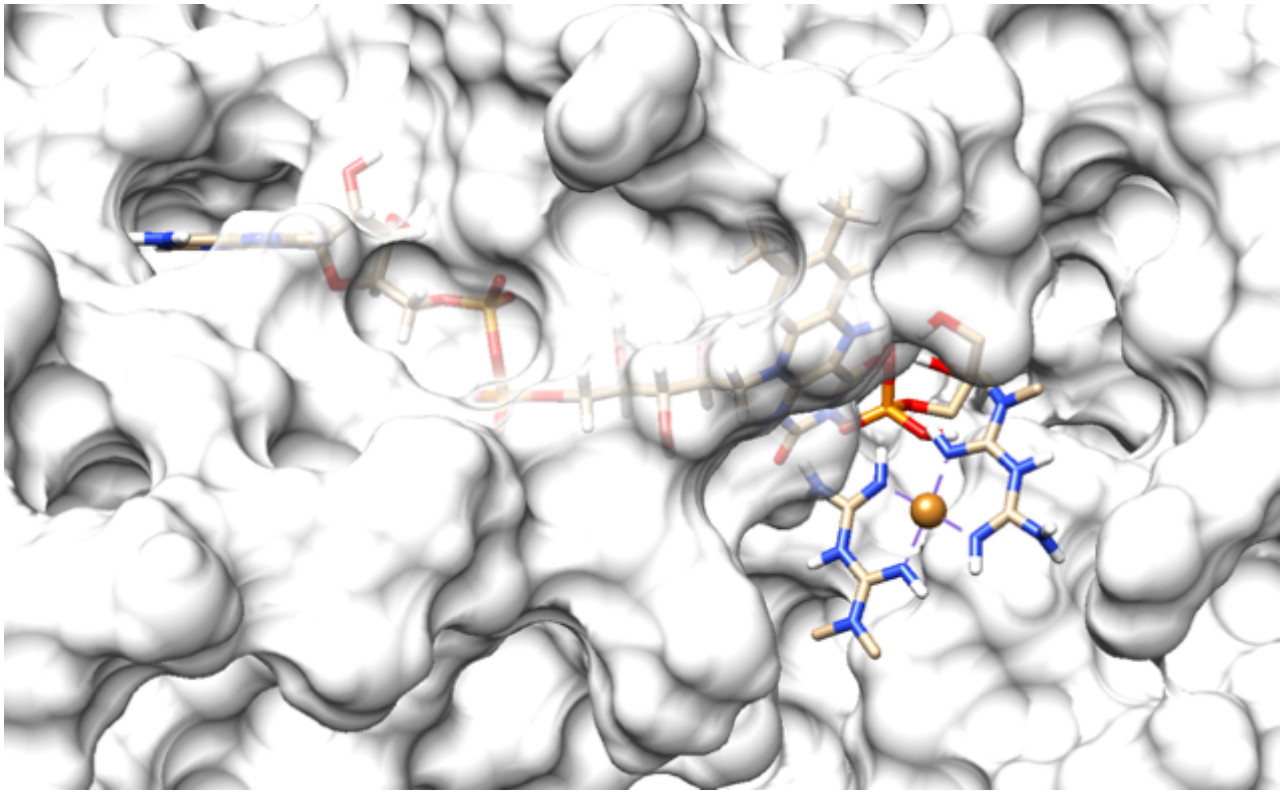


Figure 3.28 – G3P docked to GPD2 when FAD is slid in and [Cu(Metf-H)$_2$] is present, frontal view

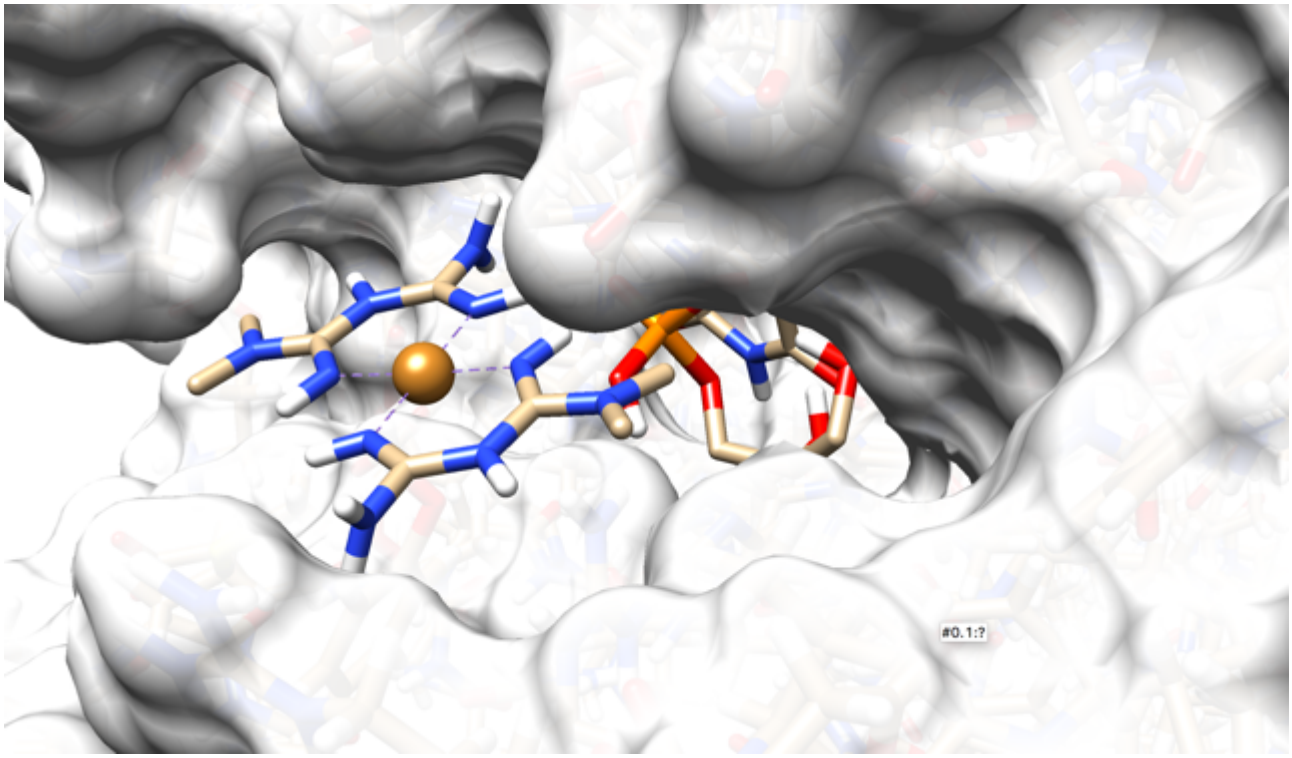


Figure 3.29 –G3P docked to GPD2 when FAD is slid in and [Cu(Metf-H)$_2$] is present, lateral view

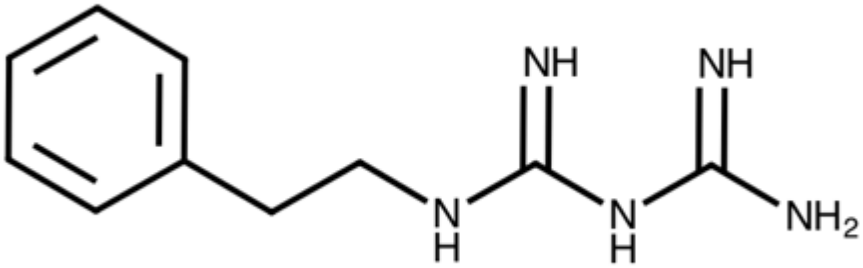


Figure 3.30 – The structural formula of phenformin

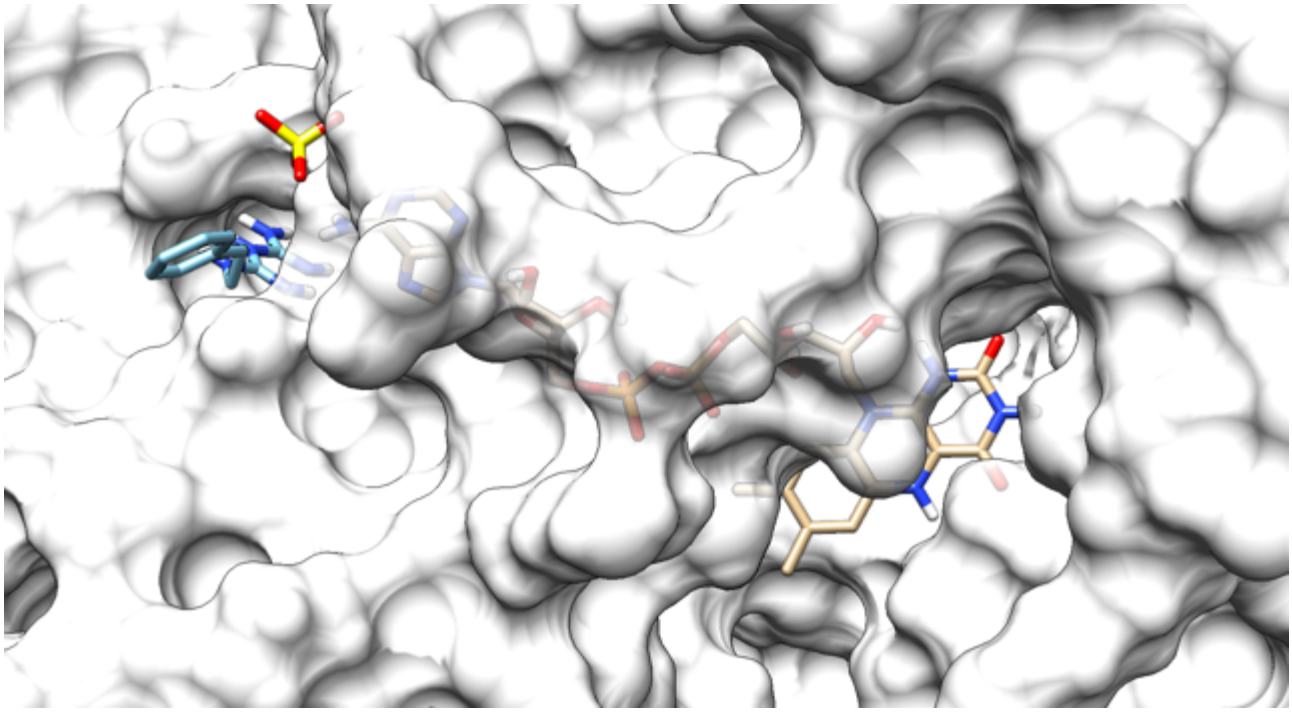


Figure 3.31 – Phenformin binding to the cavity's small entrance when the FAD is slightly slid out, corresponding to a binding energy of -5.7 kcal/mol

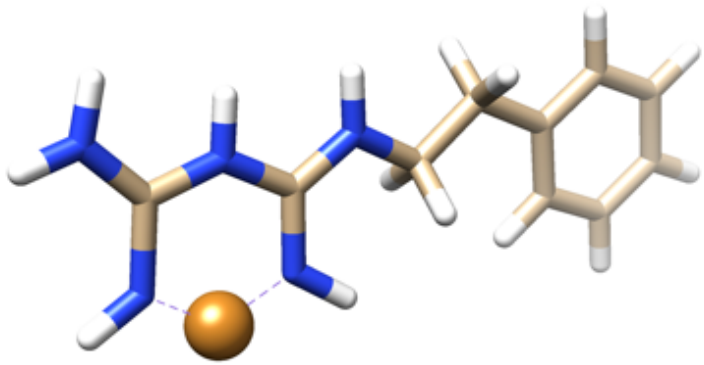


Figure 3.32 – Monophenformin copper complex tridimensional representation

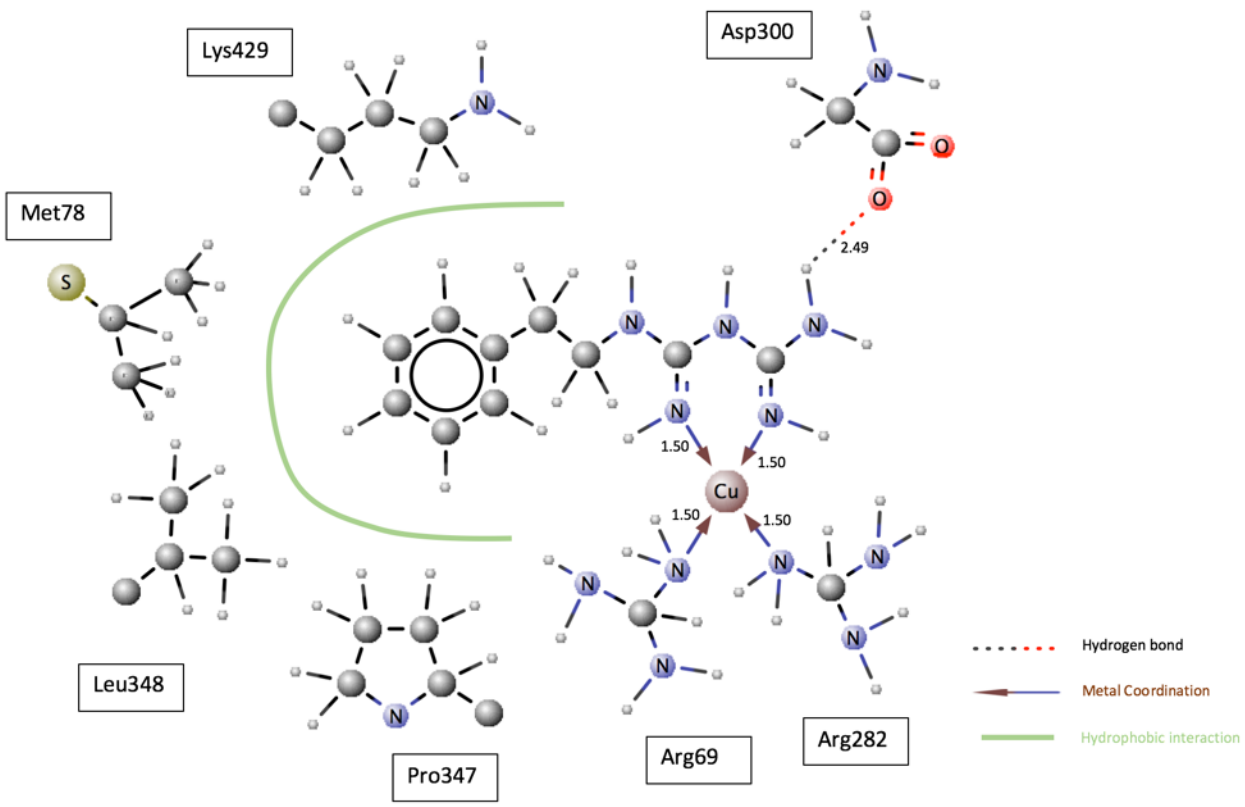


Figure 3.33 – Monophenformin copper complex interactions with cavity residues, bond lengths are expressed in angstroms

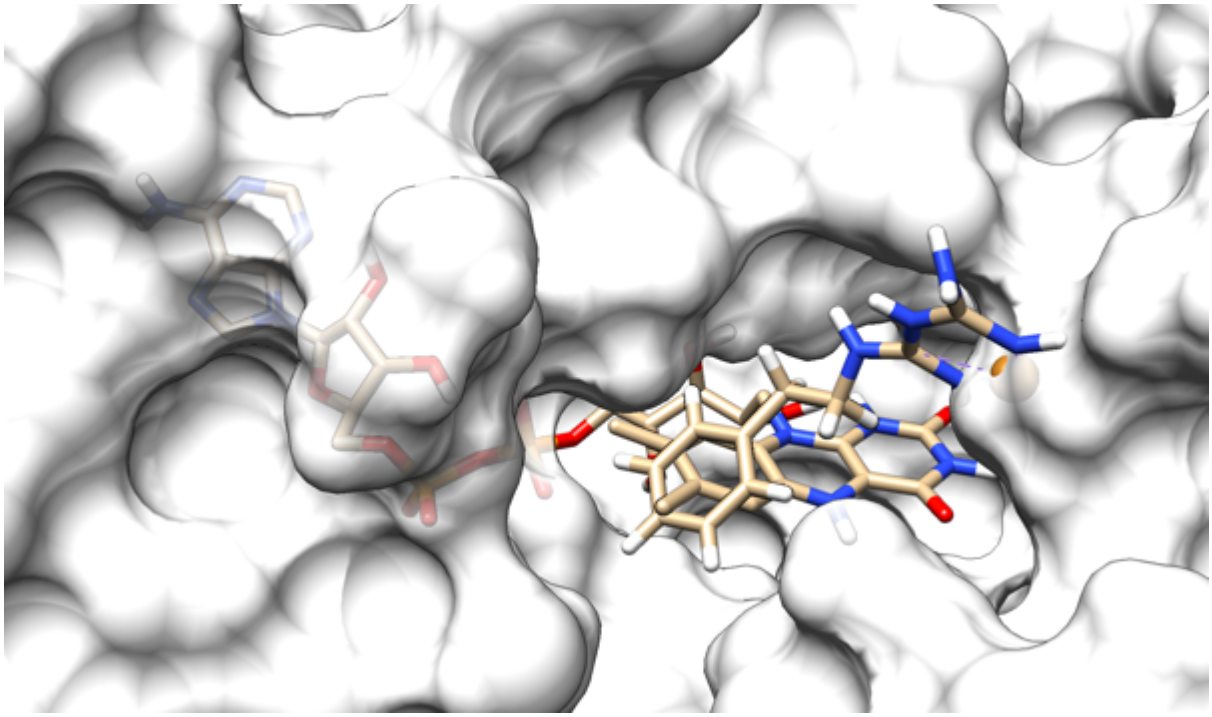


Figure 3.34 – Monophenformin complex collision with slightly slid out FAD

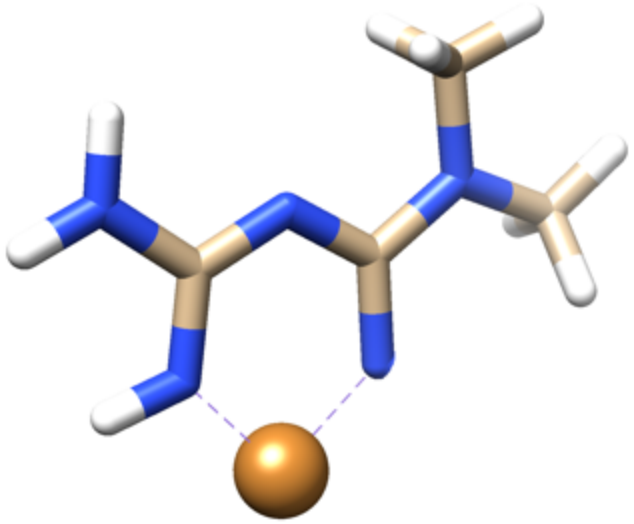


Figure 3.35 – Monometformin copper complex tridimensional representation

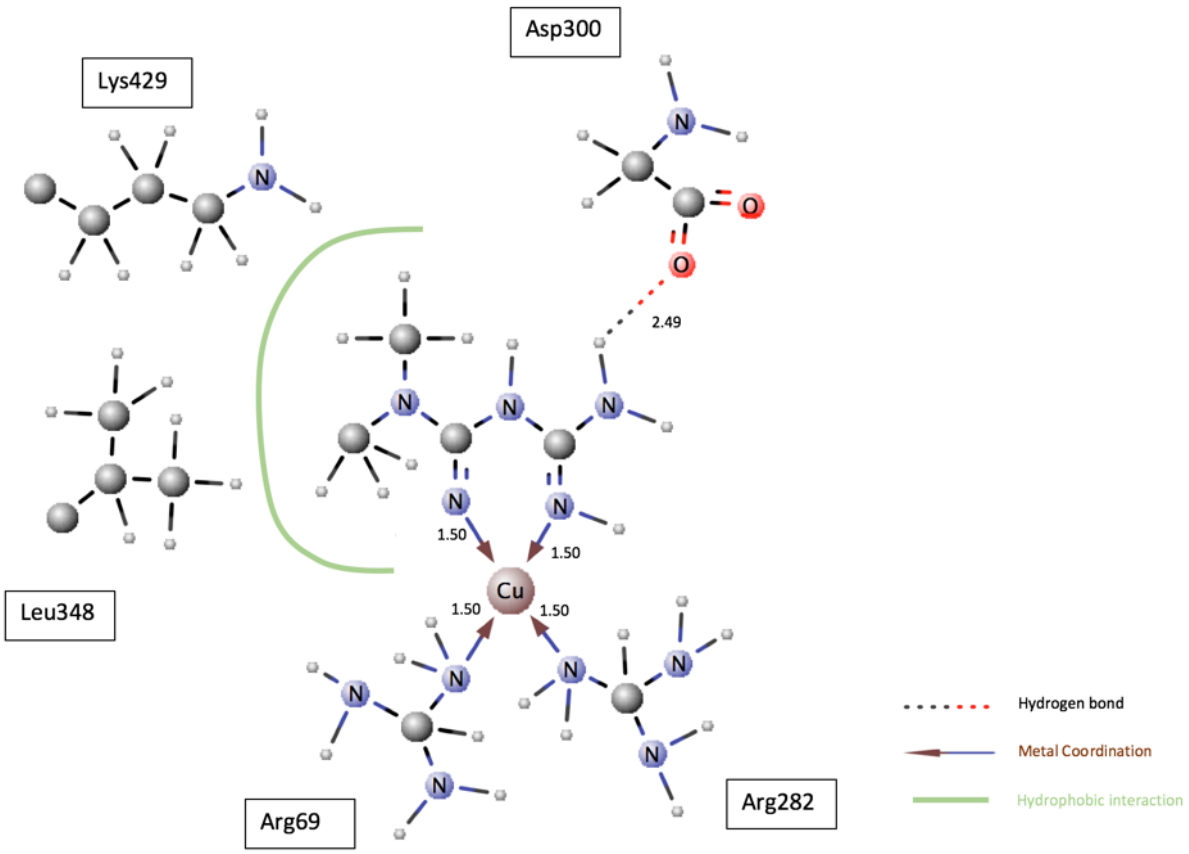


Figure 3.36 – Monometformin copper complex interactions with cavity residues, bond lengths are expressed in angstroms

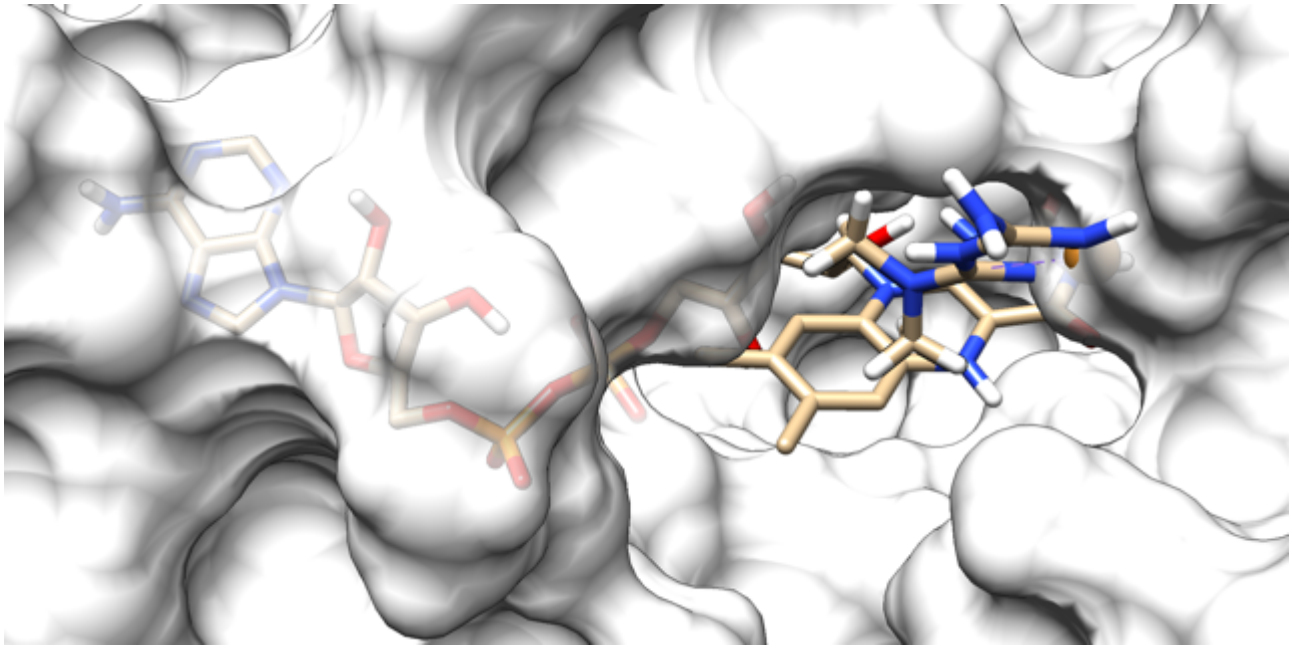


Figure 3.37 – Monometformin complex collision with slightly slid out FAD

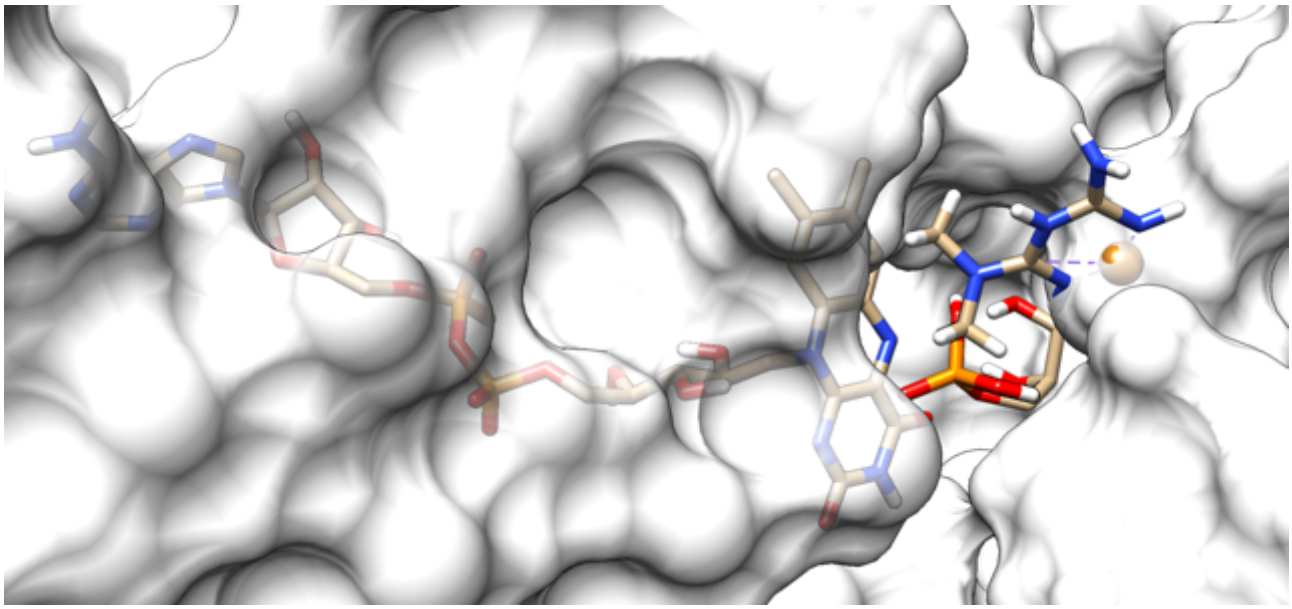
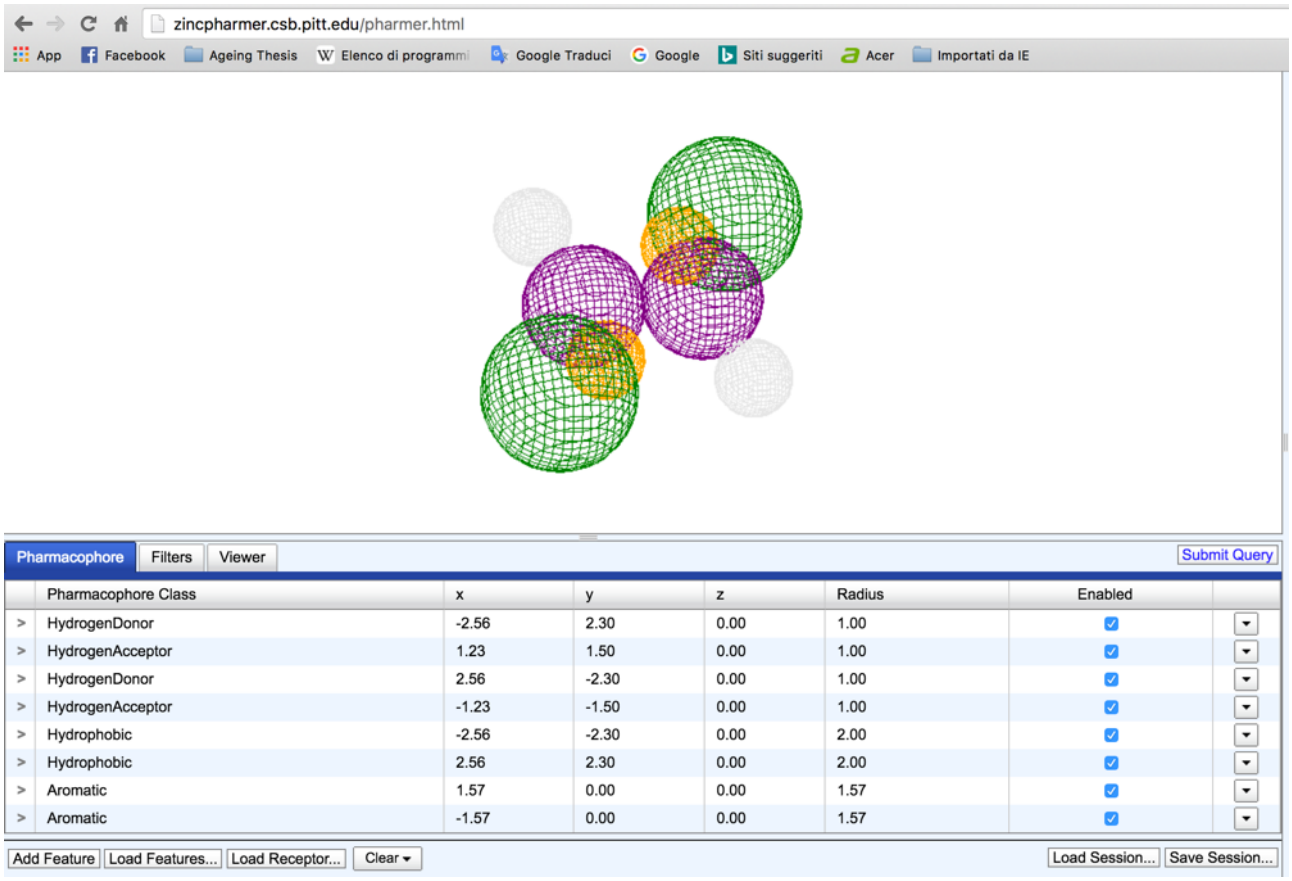


Figure 3.38 – G3P docked to FAD in fully slid in position when $[Cu(Met\text{-}H)]^{1+}$ is present



	Pharmacophore Class	x	y	z	Radius	Enabled
>	HydrogenDonor	-2.56	2.30	0.00	1.00	✓
>	HydrogenAcceptor	1.23	1.50	0.00	1.00	✓
>	HydrogenDonor	2.56	-2.30	0.00	1.00	✓
>	HydrogenAcceptor	-1.23	-1.50	0.00	1.00	✓
>	Hydrophobic	-2.56	-2.30	0.00	2.00	✓
>	Hydrophobic	2.56	2.30	0.00	2.00	✓
>	Aromatic	1.57	0.00	0.00	1.57	✓
>	Aromatic	-1.57	0.00	0.00	1.57	✓

Figure 3.39 – [Cu(Metf-H)$_2$] pharmacophore shape and data as modeled on Zinc Pharmer

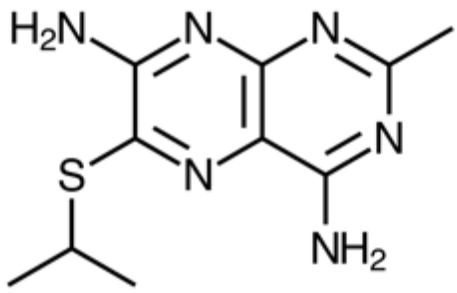


Figure 3.40 – Structure of [4-amino-6-(isopropylthio)-2-methyl-pteridin-7-yl]amine

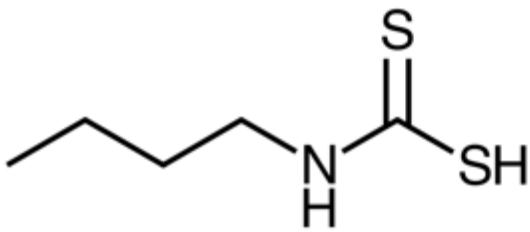


Figure 3.41 – Structure of butyl carbamodithioic acid

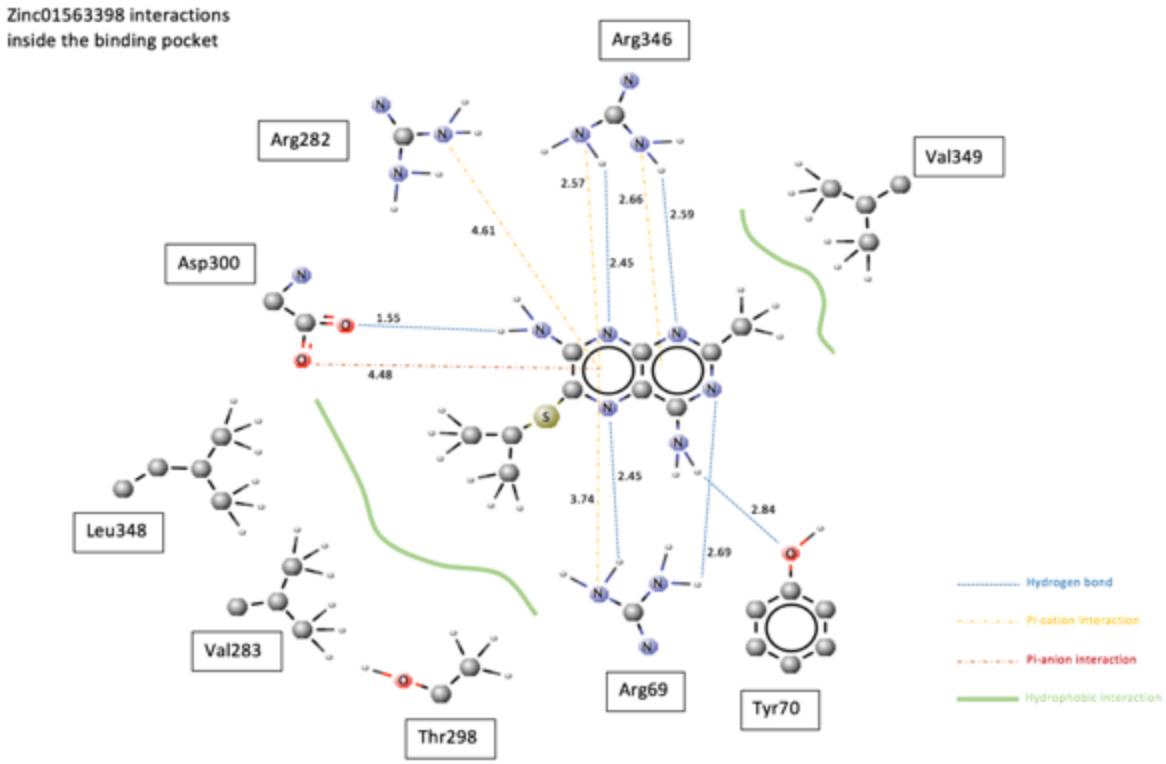


Figure 3.42 – [4-amino-6-(isopropylthio)-2-methyl-pteridin-7-yl]amine's 2D interaction diagram, bond lengths are in angstroms

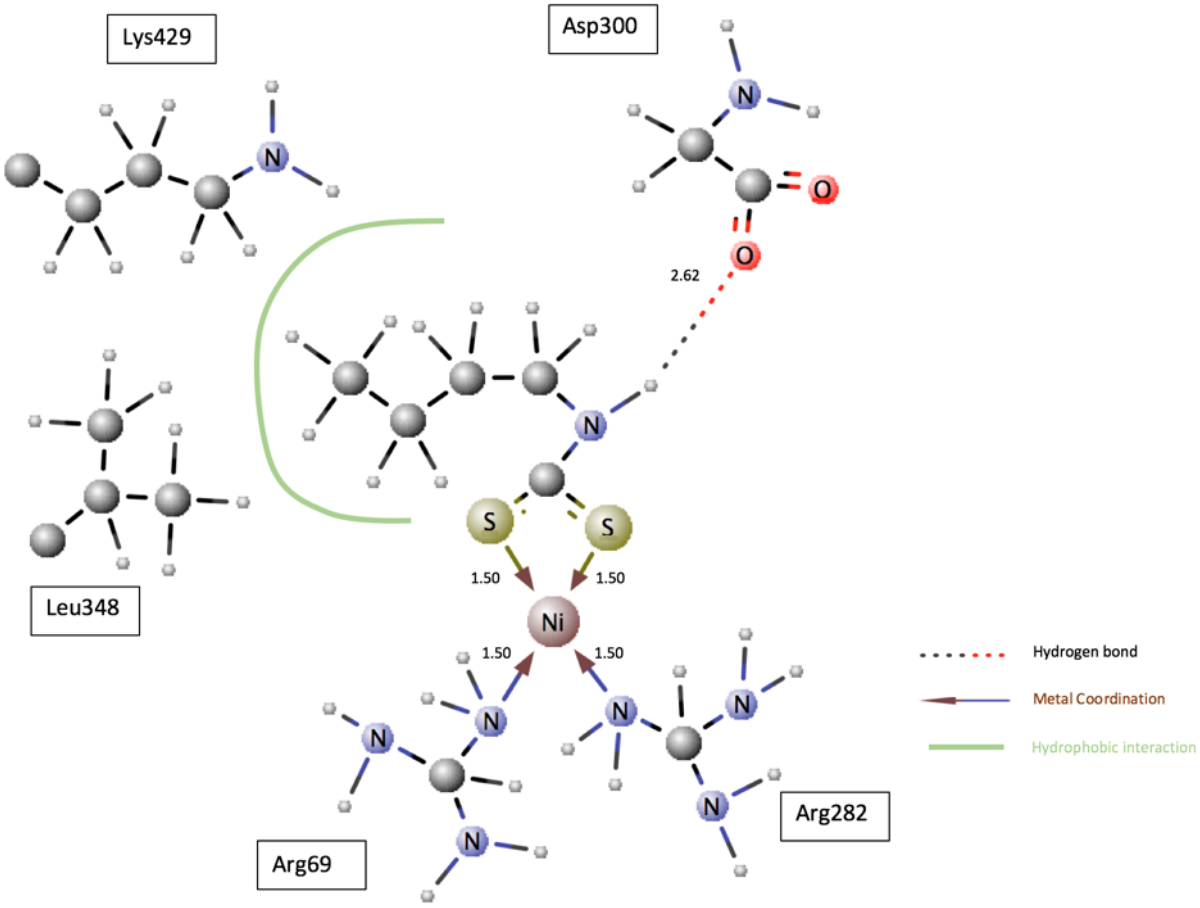


Figure 3.43 – butyl carbamodithioic acid's 2D interaction diagram, bond lengths are in angstroms

Tables

List of 2RGH's residues
GLY25, GLY26, GLY27, ILE28, THR29, ILE47, GLU48, MET49, GLN50, GLY55, THR56, SER57, ARG59, SER60, THR61, LEU63, VAL64, HIS65, ILE68, ARG69, TYR70, MET78, VAL79, MET195, LYS196, ALA197, VAL198, THR231, SER232, GLY233, PRO234, TRP235, VAL236, ASP237, LYS238, VAL239, ARG240, ASN241, LEU242, THR255, LYS256, GLY257, ILE258, HIS259, TYR274, PHE275, ASP276, THR277, GLY282, ARG283, MET284, VAL285, THR298, ASP300, THR301, GLY344, LEU345, ARG346, PRO347, LEU348, VAL349, GLY428, LYS429, ILE430, THR431

Table 3.1 – List of the residues forming the original 2RGH cavity

List of GPD2's residues
ILE24, GLY25, GLY26, GLY27, ALA28, THR29, VAL47, GLU48, MET49, ARG50, GLY55, THR56, SER57, ARG59, SER60, THR61, LEU63, ILE64, HIS65, VAL68, ARG69, TYR70, MET78, VAL79, MET195, MET196, GLU197, VAL198, ALA231, THR232, GLY233, PRO234, PHE235, VAL236, ASP237, SER238, VAL239, ARG240, ASN241, MET242, SER255, ALA256, GLY257, VAL258, HIS259, LEU274, ASP275, ASP276, ALA277, ARG282, VAL283, MET284, ILE285, THR298, ASP300, THR301, GLY344, ILE345, ARG346, PRO347, LEU348, VAL349, GLY428, LYS429, ILE430, THR431

Table 3.1 – List of the residues forming the modeled GPD2 cavity

FAD position	FAD binding energy/kcal mol^{-1}	Distance from the crystallographic position/Å
"Fully slid in"	-14.7	0
"Slightly slid out"	-12.8	3
"Significantly slid out"	-11.9	8
"Extremely slid out"	-10.7	10

Table 3.3 – Summary of the reasonably stable FAD poses inside the cavity, as computed by the molecular docking

FAD position	G3P binding energy/ kcal mol^{-1}	Distance between the G3P hydrogen and the FAD nitrogen/Å	Hydrogen bonds	Other bonds (electrostatic, Pi carbon-hydrogen)	Repulsive interactions
Fully slid in	-6.0	2.4	4	5	0
Slightly slid out	-5.1	3.4	4	2	1
Significantly slid out	-4.5	3.4	4	1	2
Extremely slid out	NA	G3P doesn't bind to FAD	-	-	-

Table 3.4 – Results of the G3P interaction with different FAD poses, as computed by the molecular docking

FAD position	Nitrogen at position 1	Nitrogen at position 5
Fully slid in	No interactions	2.5 Å hydrogen bond with THR61
Slightly slid out	3.2 Å proximity with HIS259	No interactions
Significantly slid out	No interactions	No interactions
Extremely slid out	No interactions	No interactions

Table 3.5 – Cavity interactions of the nitrogens in position 1 and 5 on the isoalloxine ring of $FADH_2$

FAD position	Coenzyme Q10 binding energy/kcal mol^{-1}	Distance from position 5 nitrogen/ angstroms	Hydrogen bonds	Other bonds
Fully slid in	-6.1	4.7	3	4
Slightly slid out	-6.6	3	2	8
Significantly slid out	NA - no significative binding	NA - no significative binding	-	-
Extremely slid out	NA - no significative binding	NA - no significative binding	-	-

Table 3.6 – Results of the coenzyme Q10 molecular docking with the large entrance with different FAD poses

FAD position	Metformin binding energy/kcal mol^{-1}	Metformin inhibition constant(KI)/nM	Hydrogen bonds	Other bonds
Fully slid in	NA – The small cavity is occupied by FAD	NA – The small cavity is occupied by FAD	-	-
Slightly slid out	-5.0	2.15×10^5	2	1
Significantly slid out	-5.3	1.30×10^5	1	4
Extremely slid out	-5.8	5.57×10^4	3	4

Table 3.7 – Results of the metformin molecular docking with the small entrance with different FAD poses

FAD position	Metformin binding energy/cal mol^{-1}	Metformin inhibition constant(KI)/nM
Fully slid in	-4.0	1.16×10^6
Slightly slid out	No significant binding occurred	No significant binding occurred
Significantly slid out	No significant binding occurred	No significant binding occurred
Extremely slid out	No significant binding occurred	No significant binding occurred

Table 3.8 – Results of the metformin molecular docking with the large entrance with different FAD poses

Copper complex	Binding energy /kcal mol^{-1}	Inhibition constant(KI) /nM	Hydrogen bonds	Other bonds	Repulsive interactions	Number of residues involved in hydrophobic interactions	Ability to prevent docking of Q10 in the cavity
$[Cu(Metf\text{-}H)_2]$	-5.9	4.71×10^4	4	4	1	5	yes
$[Cu(Metf)_2]^{2+}$	-5.4	1.09×10^5	2	5	0	5	yes

Table 3.9 – Results of the dimetformin copper complex molecular docking with the large entrance with different FAD poses

FAD position	Phenformin binding energy/kcal mol^{-1}	Phenformin inhibition constant(KI)/nM	Hydrogen bonds	Other bonds
Fully slid in	NA – The small cavity is occupied by FAD	NA – The small cavity is occupied by FAD	-	-
Slightly slid out	-5.7	6.60×10^4	2	3
Significantly slid out	-6.0	3.97×10^4	1	7
Extremely slid out	-6.3	2.39×10^4	3	7

Table 3.10 – Results of the phenformin molecular docking with the small entrance with different FAD poses

Copper complex	Binding energy/kcal mol^{-1}	Inhibition constant(KI)/nM	Hydrogen bonds	Coordination bonds	Hydrophobic interactions	Ability to prevent docking of Q10 in the cavity
$[Cu(Phen)]^{2+}$	-5.8	4.71×10^4	1	2	4	yes

Table 3.11 – Results of the monophenformin copper complex molecular docking with the large entrance with fully slid in FAD

Copper complex	Binding energy/kcal mol^{-1}	Inhibition constant(KI)/nM	Hydrogen bonds	Coordination bonds	Hydrophobic interactions	Ability to prevent docking of Q10 in the cavity
$[Cu(Met\text{-}H)]^{1+}$	-5.7	6.60×10^4	1	2	2	yes

Table 3.12 – Results of the monometformin copper complex molecular docking with the large entrance with fully slid in FAD

Virtual lead	Binding energy/ kcal mol^{-1}	Inhibition constant(KI)/nM	Hydrogen bonds	Other bonds	Repulsive interactions or steric bumps	Number of residues involved in hydrophobic interactions
[4-amino-6-(isopropylthio)-2-methyl-pteridin-7-yl]amine	-5.8	$5.57x10^4$	6	4	1	4
[Ni(BCA-H)]$^{1+}$	-5.6	7.81×10^4	1	2	0	2

Table 3.13 – Results of docking of the virtual screening leads in the GPD2 cavity with FAD in fully slid in position