Engineering of formate dehydrogenase for the acceptance of a biomimetic nicotinamide-based cofactor in the reduction of CO_2

Vanesa Teijeiro Seijas

University Master's Degree in Bioinformatics and biostatistics from Open University of Catalonia and University of Barcelona

Area 3, subarea 4 (Structural Bioinformatics and Design of Bioinformatic Applications)

Elisabeth Ortega Carrasco

David Merino Arranz

January 2020

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Engineering of formate dehydrogenase for the acceptance of a biomimetic nicotinamide-based cofactor in the reduction of CO_2*
Nombre del autor:	*Vanesa Teijeiro Seijas*
Nombre del consultor/a:	*Elisabeth Ortega Carrasco*
Nombre del PRA:	David Merino Arranz
Fecha de entrega (mm/aaaa):	01/2020
Titulación:	*Máster en Bioinformática y Bioestadística UOC-UB*
Área del Trabajo Final:	*Area 3, subarea 4 (Structural Bioinformatics and design of bioinformatic applications)*
Idioma del trabajo:	*Inglés*
Key words:	*Protein engineering, Nicotinamide biomimetic cofactor, Formate dehydrogenase*

Resumen del Trabajo:

La enzima formato deshidrogenase (FDH) es ampliamente utilizada para la regeneración del cofactor NAD(P)H. Debido al alto precio del cofactor natural, su regeneración es crítica cuando se trabaja con enzimas que lo requieren. Otro uso de FDH es la reducción de CO_2 a ácido fórmico. Los cofactores biomiméticos se crearon como una alternativa al cofactor natural. Sin embargo, no hay estudios que los integren con la enzima FDH. En este estudio, utilizando el acoplamiento molecular, el cofactor biomimético P3NAH fue seleccionado como el que mejor se acomoda en el centro activo de FDH de *Pseudomonas* (PsFDH) y puede reducir el CO_2. Con el objetivo de crear una mutante de PsFDH que acomode la parte de nicotinamida de P3NAH en una posición similar a la nicotinamida de NADH, se modificaron residuos en el túnel de acceso del cofactor y en el centro activo de la enzima. La estructura de las mutantes se generó con modelado de homología. Al cambiar el residuo Ser 380 por Gly, la nicotinamida de P3NAH se sitúa en una posición más cercana a la posición de NADH. La enzima sin modificar sitúa la molécula P3NAH con una distancia promedio entre átomos de P3NAH y residuos relevantes de 1.2 Å mientras que la mutante 2NADa_S380G lo acomoda con una distancia promedio de 0.25 Å. Además de los resultados de acoplamiento aceptables para P3NAH, la estabilidad de la proteína y el transporte de ligando de 2NADa_S380G es aceptable, por lo que esta mutante es más adecuada para la reacción con P3NAH que la enzima sin modificar.

Abstract:

Formate dehydrogenase (FDH) is an enzyme widely used in the industry for the regeneration of the cofactor NAD(P)H. Due to the expensive price of the natural cofactor, its regeneration is critical when working with enzymes that require it. Another use of formate dehydrogenase is the reduction of CO_2 to formic acid. Biomimetic cofactors were created as an alternative to the expensive natural cofactor NADH. However, there are no studies that integrate them with the enzyme FDH. In this study, using molecular docking, the biomimetic cofactor P3NAH was selected as the biomimetic cofactor that accommodates best in the active site of the enzyme FDH from *Pseudomonas* (PsFDH) and can reduce CO_2. With the objective of creating a mutant of PsFDH that accommodates the nicotinamide part of P3NAH in a position similar to the nicotinamide of NADH, residues were switched in the cofactor access tunnel and active site of the enzyme and the structure of the mutants was generated with homology modelling. When changing the residue Ser 380 for Gly, the nicotinamide part of P3NAH locates the closest when compared to the position of NADH in the WT enzyme. The WT enzyme locates P3NAH with an average distance between relevant atoms of P3NAH and relevant residues of 1.2 Å while the mutant 2NADa_S380G accommodates it with an average distance of 0.25 Å. Apart from acceptable docking results for P3NAH, the protein stability and the ligand transport of 2NADa_S380G is acceptable so this mutant is more suitable for reaction with P3NAH than the WT enzyme.

Dedication

For my amazing husband who supported me and helped me in all things great and small.

Acknowledgments

Elisabeth, muchas gracias por aceptar ser mi directora y por tus comentarios a lo largo de este trabajo. La verdad creo que debería haberte hecho muchas más preguntas y así podría aprender mucho más de ti. Espero sigamos en contacto en el futuro pues me pareces una gran profesional.

Moitas grazas familia! Grazas polo apoio incondicional en calquera aventura na que me embarco aínda que asemelle ben estrafalaria.

한국에서의 삶에 너무 좋은 경험과 관점을 선사해주어서, 우리 한국 가족들에게 너무 고마워요.

Eli, moitas grazas por esas charlas nas que o teléfono se queda sen batería! Encántame falar contigo por cómo me fas analizar calquera tipo de situación. Digo tamén por aquí que estou súper orgullosa da miña amiga actriz, creativa e traballadora como pouca outra xente que coñezo.

María, moitas grazas pola paciencia e mensaxes de voz con todo tipo de detalles. Resistencia e para adiante con esa opo! Acabarala sacando. Estou segura!

Ana, Anxos, Azu, Cris, Gonzalo, Mirta, Montse e Ro, a vida sería moi aburrida sen vós! Admírovos a todas vós!

Thank you Dr. Baig. When discussing with you my research I can see my weak parts and options for improvement. You are a good research ally and I think we will work together soon.

Thank to all my friends from Korea University. Among them, thank you to Abhra, Areli, Ayushi, Beesha, Carlos, Carolina, Depika, Dere, Dongsop, Eunjee, Hamid, Hansok, Hyejin, Irfan, Isra, Millika, Pau, Quyen, Peri, Robel, Sajjad, Santiago, Sunam and Theo. We are all from different fields but share the commitment with our work. I have no doubts you are good professionals and will be much better. Thank you for listening, sharing ideas and advise.

Thank prof. Ahn for your talks. You open my eyes and make me think in ways I could never imagine.

Thank you, Prof. Kim. In your laboratory I started working with enzymes and I fall in love with them. Thank you Dr. Adhikari, IS, Dr. Kumar, HS, JJ, NJH and YH for the research discussions. Specially thanks to Dr. Adhikari for assisting me during the production and purification of enzymes.

Erratum

The enzyme formate dehydrogenase from *Thiobacillus sp. KNK65MA* (TsFDH) should have been selected for conducting this study instead of the enzyme from *Pseudomonas* (PsFDH). The enzyme PsFDH was not reported in the bibliography for the reduction of CO_2. The right approach for the selection of the enzyme is:

1) Check all the enzymes reported for CO_2 reduction and that use a cofactor.

2) Among the reported CO_2 reduction enzymes, select the most thermostable enzyme with an available 3D structure in PDB.

Following this approach, the enzyme formate dehydrogenase from *Thiobacillus sp. KNK65MA* (TsFDH) (Kim et al., 2013) should have been selected.

List of Tables

List of Figures

Table of Contents

1. Introduction

1.1. Introduction to the Study

The topic of this study is to determine at least one beneficial mutation for doing protein engineering to the enzyme formate dehydrogenase to allow the use of a nicotinamide biomimetic cofactor instead of the natural cofactor NADH. The natural cofactor is expensive and its use limits the use of enzymes that depend on it in commercial processes.

Formate dehydrogenase is an enzyme widely used in the industry for the regeneration of cofactors and it was reported that some enzymes can convert CO_2 to formic acid. A reduction in the cost of the reaction for CO_2 conversion could contribute to create economical feasible industrial processes for CO_2 utilization, instead of releasing it to the atmosphere.

So far, there are no studies to determine the required mutations on the enzyme formate dehydrogenase to allow the use of a biomimetic cofactor for the transfer of an electron and a hydrogen atom to the substrate of the enzyme.

1.2. Background of the Study

1.2.1. FDH Applications

Formate dehydrogenase (FDH, EC 1.17.1.9) catalyses the reversible interconversion of formate and carbon dioxide. The enzyme can be used by itself to convert CO_2 to formate or in a cascade of enzymes (together with formaldehyde dehydrogenase and alcohol dehydrogenase) to convert CO_2 to methanol(Amao, 2018; Nielsen et al., 2019).

$$HCOO^- \rightleftharpoons CO_2 + H^+ + 2e^-$$

$$CO_2 + H^+ + 2e^- \rightleftharpoons HCOO^-$$

Formate dehydrogenase is widely used for the regeneration of the cofactor NAD(P)H in industrial applications.

1.2.2. Types of formate dehydrogenase

There is a great diversity of FDHs in terms of their structure, subunit composition and metabolic function. Some are soluble while others are bounded to membranes. There is a general classification of enzymes depending on the presence or absence of metal in their active site (Maia et al., 2017).

1.2.2.1. Metal-independent FDHs

The metal-independent FDHs are a uniform group of enzymes. They belong to the family D-isomer specific 2-hydroxyacid dehydrogenase(Vinals et al., 1993). These are globular proteins with one or two identical active sites.

The reaction takes place when the enzyme allows the cofactor and the substrate to reach close positions to transfer electrons between the two molecules(Castillo et al., 2008). In general, the oxidation of formate is favoured over the reduction of CO_2.

However, in the presence of high concentrations of cofactor, it is shifted. The cofactors used in nature for electron transfer are NADH or NAD(P)H.

1.2.2.2. Metal-dependent FDHs

The metal-dependent FDHs have molybdenum or tungsten in the active site and belong to the DMSO reductase family(Hille et al., 2014) or also classified as Mo/W-bisPGD enzymes(Grimaldi et al., 2013). The structures of these enzymes share a main 'alpha unit' that contains the Mo/W-bisPGD active site, and some have a 'beta subunit' (20-35 kDa) and a 'gamma-subunit' (12-18 kDa). Recently, a study classified the metal-dependent FDHs in six groups(Nielsen et al., 2019), as indicated in Figure 1.

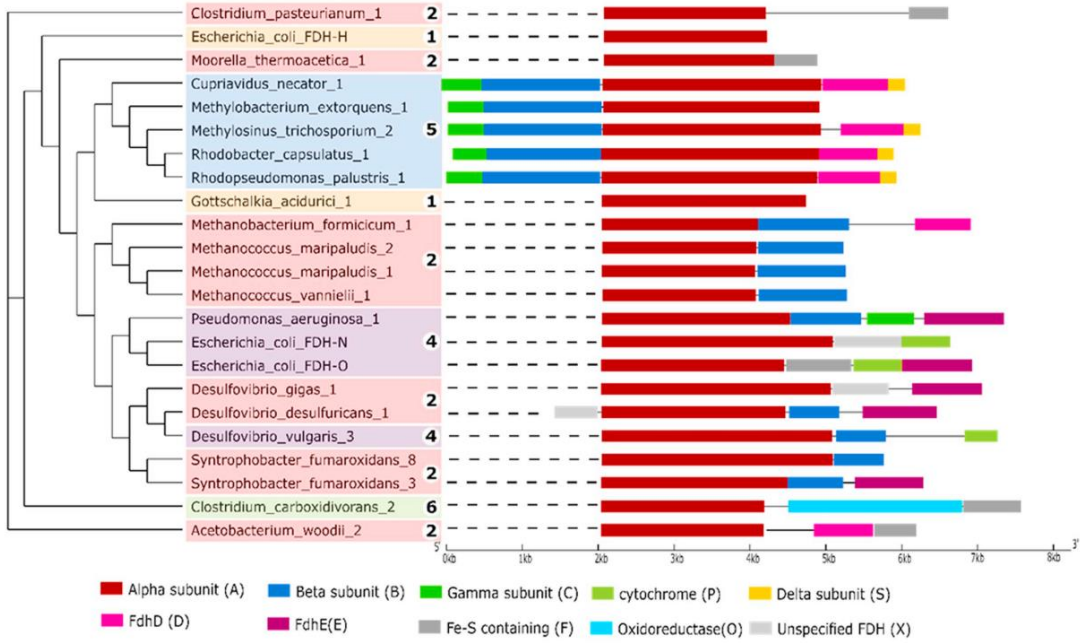


Figure 1. Classification of the metal-dependent FDHs(Nielsen et al., 2019)

The molybdenum or tungsten of the Mo/W-bisPGD in the active site, presents a terminal sulfur residue. The enzyme directs electrons through a ‘wire’ of FeS groups in order to reduce the heavy metal-sulfo group in the active site, and this reduced metal-sulfo group acts as a hydride donor for the reduction of CO_2 to formate. In comparison to the metal-independent FDHs, a wider variety of electron donors appear in nature, as cytochromes, ferredoxins, coenzyme F_{420}, and membrane quinols(Nielsen et al., 2019).

Various metal-dependent FDHs are oxygen sensitive. The mechanism is still unknown. The metal-dependent FDHs described in the bibliography that can tolerate oxygen are MeFDH (*Methylobacterium extorquens AM1*), RcFDH (*Rhodobacter capsulatus*) and CnFDH (*Cupriavidus necator*).

1.2.3. Biomimetic nicotinamide cofactors instead of NADH

The metal-independent formate dehydrogenase enzymes require a cofactor (NADH) for the supply of 2 electrons and one hydride to reduce CO_2 to formate. For each molecule of CO_2 reduced, one molecule of cofactor is required. The reduced natural cofactor NADH is expensive and not stable. One approach to decrease the cost of the reaction and make it economically feasible is to regenerate the cofactor. Studies can be

found in the bibliography for the regeneration of the cofactor using other enzymes, electrochemistry and photocatalysts. Further development is required in this area. Another option to decrease the cost of the reaction is to use biomimetic cofactors. These cofactors could also be regenerated to further decrease in the cost.

Several biomimetic cofactors were reported in the bibliography and recent reviews on the topic are worth to read(Guarneri et al., 2019; Knaus et al., 2016; Zachos et al., 2019). The names and structures are indicated in Table 1, Table 2 and Figure 2.

Table 1. Totally synthetic NCBs

Abbreviated name	**Full name**
BNAH	1-benzylnicotinamide
HPNA	p-hydroxyphenylnicotinamide
PNAH	1-phenylnicotinamide
P2NAH	1-phenylethylnicotinamide
P3NAH	1-(3-phenylpropyl)nicotinamide
BNNH	1-benzylnicotinonitrile
BNAcH	1-benzyl nicotinic acid
BAPH	1-benzyl 3-acetylpyridine
MNAH	1-methylnicotinamide
BuNAH	1-butylnicotinamide
BuOHNAH	1-(4-hydroxybutyl)nicotinamide
HEH	Hantzsch ester

Table 2. Semi synthetic NCBs

Abbreviated name	**Full name**
NRH	nicotinamide riboside
NMNH	nicotinamide mononucleotide
NFCH	nicotinamide flucytosine dinucleotide
NCDH	nicotinamide cytosine dinucleotide
PEG-NADH	polyethylene glycosylated-NADH
carba-NADH	carba-NADH

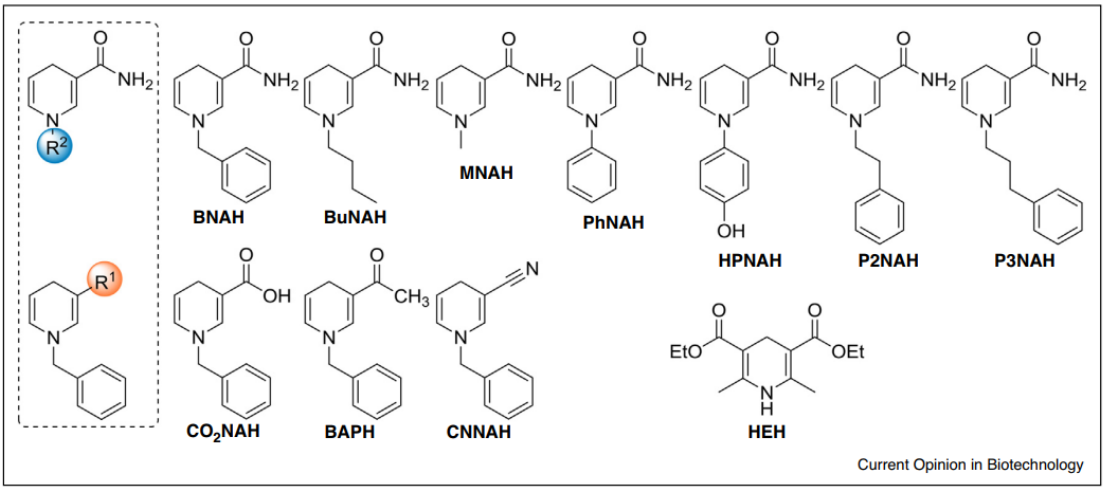


Figure 2. Structures of NCBs (reduced form) used for NAD-dependent oxidoreductases-catalysed reactions, and the Harntzsch ester HEH(Guarneri et al., 2019; Zachos et al., 2019)

The reduction potential of the biomimetic cofactors is different than the reduction potential for NADH as indicated in Figure 3. In addition, due to the pH-dependency of the CO_2 reduction potential, the reaction must take place around pH 5 in order to use biomimetic cofactors (refer to Figure 4).

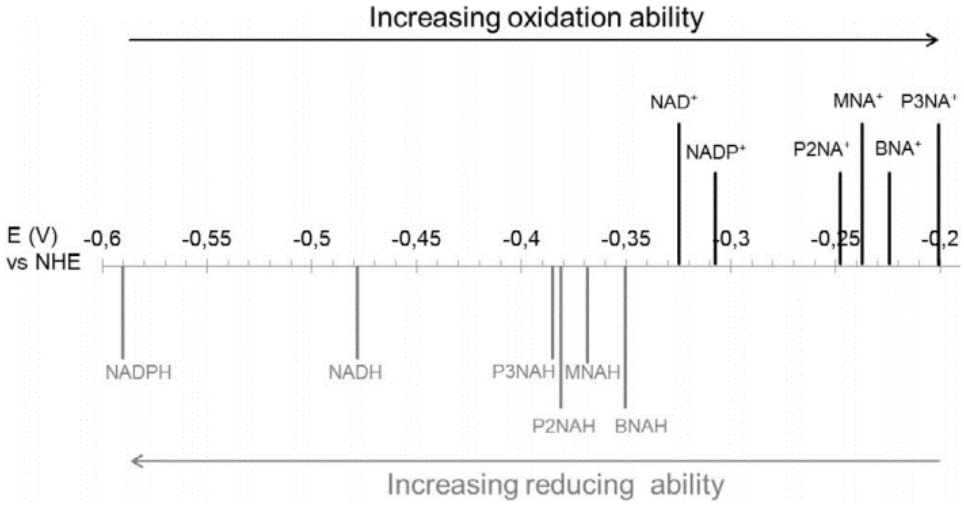


Figure 3. Ranking order of cathodic and anodic potentials in increasing order of cofactor reduction and oxidation abilities for NCBs(Nowak, Pick, Csepei, et al., 2017)

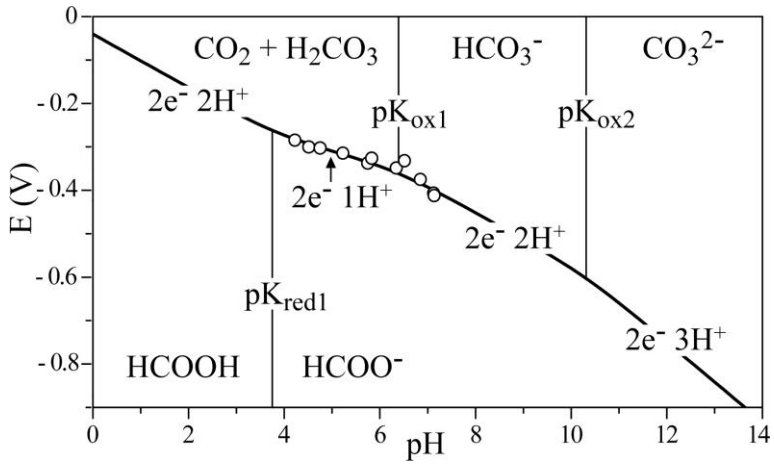


Figure 4. The pH-dependent reduction potential of CO_2 (Reda et al., 2008)

1.2.4. Engineering the access tunnel of the cofactor to the active site

Since a different cofactor will be used, the access tunnel of the wild enzyme might have steric restriction to the biomimetic cofactor. The target of this project is to modify residues located in the access tunnel of the cofactor or active site.

1.2.5. Previous protein modification for FDH

A review of the bibliography that mentions studies where mutations were done to FDH could be beneficial to target certain amino acids in the active site and avoid others. For example, FDH from *Candida boidinii* has been engineered to accept NADP(H) instead of NAD(H) (Andreadeli et al., 2008) (Wu et al., 2009). Other modifications done to FDH had the objective of improving its thermal stability, chemical stability and kinetic parameters(Tishkov & Popov, 2006).

1.2.6. Previous protein modifications of enzymes to accept biomimetic cofactors

Several authors have engineered oxidoreductase enzymes to promote the reactions with biomimetic cofactors(Knaus et al., 2016; Nowak, Pick, Lommes, et al., 2017; You et al., 2017). Learning from their studies which residues are important for the interaction with the biomimetic cofactor could help for targeting specific amino acids.

1.2.7. Enzyme engineering

Enzyme directed evolution has become the strategy of choice for tailoring the properties of enzymes. It relies on an interactive two-step protocol. Initially it generates molecular diversity by random mutagenesis and in vitro recombination. Then, it identifies library members with improvements in desired phenotype by high-throughput screening or selection(Lutz, 2010). This method generates a huge library of enzymes.

Semi-rational, smart or knowledge-based design utilizes information on protein sequence, structure and function, as well as computational predictive algorithms to preselect promising target sites and limited amino acid diversity for protein engineering. The focus on specific amino acid positions translates into dramatically reduced library sizes(Lutz, 2010).

1.3. Problem Statement

Formate dehydrogenase is an enzyme with a great potential due to its use for cofactor regeneration and also for its application in CO_2 reduction. So far, there are no studies about which residues should be modified to promote the reaction of the enzyme with any biomimetic cofactor.

1.4. Purpose of the Study

The objective of this study is to find at least one mutation that is beneficial for the reaction of formate dehydrogenase with a biomimetic cofactor in the reduction of CO_2. For this, a semi-rational design is used in order to create a small library of mutants with different residues in the active site of the enzyme and cofactor entrance tunnel.

The model enzyme is formate dehydrogenase, targeting the reduction of CO_2 to formate. This enzyme belongs to the family of oxidoreductases, which are widely used in the pharmaceutical industry because of their ability to generate specific isomers.

The protocol created as a result of this study could be extended to other enzymes that use NADH as a cofactor in order to change it for a cheaper compound and create new economical viable processes using green catalysts.

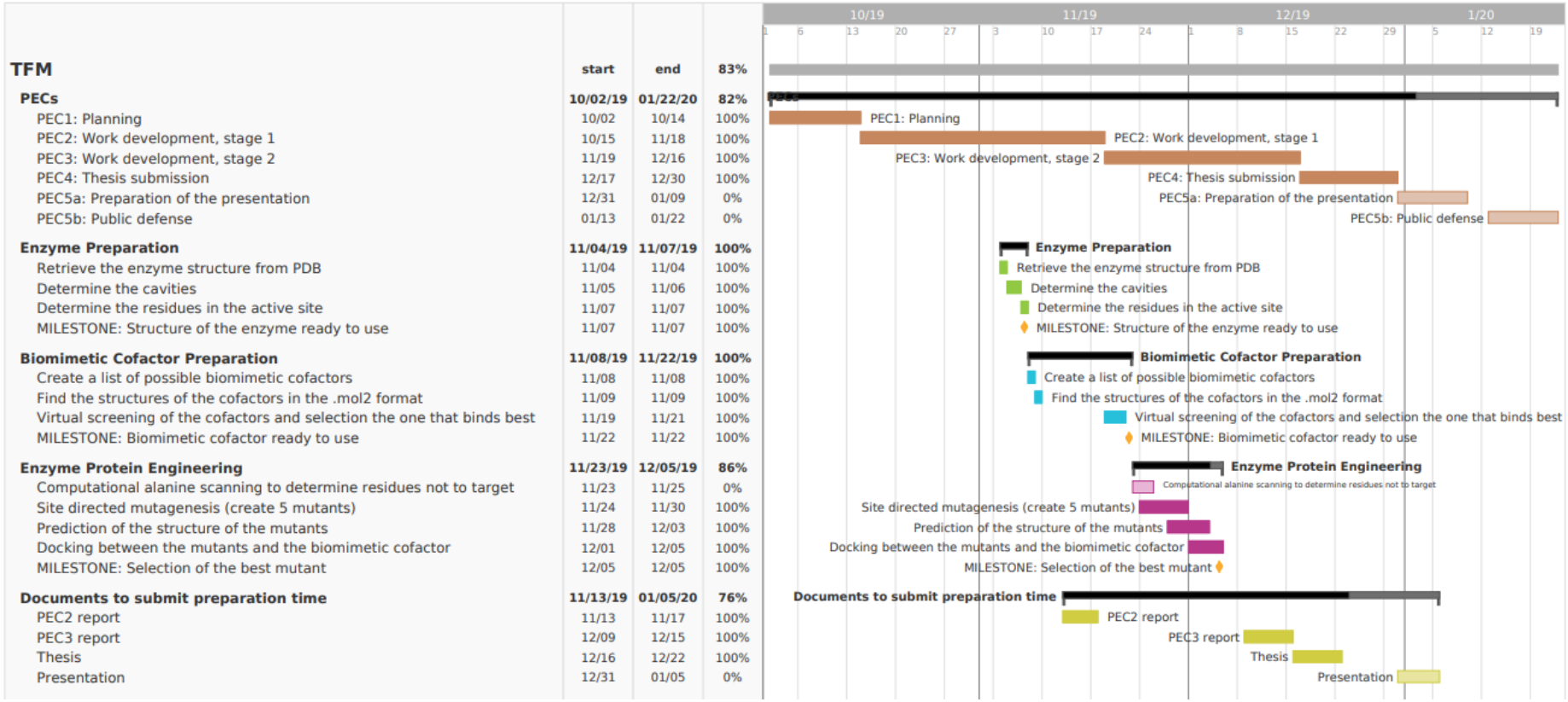
1.5. Planning, Timing and Tasks

1.5.1. Structure of the Project

This project is divided in 3 main sections. The tasks under each section are:

- SECTION 1: Enzyme preparation
  - Enzyme selection
  - Retrieve the structure from PDB
  - Determine the cavities of the enzyme
  - Determine the residues that can be targeted during mutations
- SECTION 2: Biomimetic cofactor preparation
  - Create a list of the possible cofactors to be used
  - Find the structures of the cofactors in .mol2 format
  - Virtual screening of several nicotinamide cofactor biomimetics for the selection of the most suitable for the enzyme
- SECTION 3: FDH protein engineering (modification of the access tunnel of the cofactor and active site)
  - Semi-rational protein engineering to create 8 mutants
  - Prediction of the structure of the protein after mutations
  - Docking between the final structure and the biomimetic cofactor
  - Selection of the best mutant
  - Evaluation of the viability of the mutant (protein stability and ligand transport towards the active site)

The Gantt diagram in Figure 5 indicates the time distribution of the tasks.



TFM	start	end	83%
PECs	**10/02/19**	**01/22/20**	**82%**
PEC1: Planning	10/02	10/14	100%
PEC2: Work development, stage 1	10/15	11/18	100%
PEC3: Work development, stage 2	11/19	12/16	100%
PEC4: Thesis submission	12/17	12/30	100%
PEC5a: Preparation of the presentation	12/31	01/09	0%
PEC5b: Public defense	01/13	01/22	0%
Enzyme Preparation	**11/04/19**	**11/07/19**	**100%**
Retrieve the enzyme structure from PDB	11/04	11/04	100%
Determine the cavities	11/05	11/06	100%
Determine the residues in the active site	11/07	11/07	100%
MILESTONE: Structure of the enzyme ready to use	11/07	11/07	100%
Biomimetic Cofactor Preparation	**11/08/19**	**11/22/19**	**100%**
Create a list of possible biomimetic cofactors	11/08	11/08	100%
Find the structures of the cofactors in the .mol2 format	11/09	11/09	100%
Virtual screening of the cofactors and selection the one that binds best	11/19	11/21	100%
MILESTONE: Biomimetic cofactor ready to use	11/22	11/22	100%
Enzyme Protein Engineering	**11/23/19**	**12/05/19**	**86%**
Computational alanine scanning to determine residues not to target	11/23	11/25	0%
Site directed mutagenesis (create 5 mutants)	11/24	11/30	100%
Prediction of the structure of the mutants	11/28	12/03	100%
Docking between the mutants and the biomimetic cofactor	12/01	12/05	100%
MILESTONE: Selection of the best mutant	12/05	12/05	100%
Documents to submit preparation time	**11/13/19**	**01/05/20**	**76%**
PEC2 report	11/13	11/17	100%
PEC3 report	12/09	12/15	100%
Thesis	12/16	12/22	100%
Presentation	12/31	01/05	0%

Figure 5. Project tracking

1.5.2. Changes in Comparison with the Initial Plan

All the planned milestones were achieved. In fact, 8 mutants were generated instead of the 5 initially planned and the selected mutant was evaluated in terms of protein stability and ligand transport.

Alanine scanning was substituted for a study of the residues that cannot be modified because are relevant for interaction with the substrate and with the other monomer. Because the structure of the enzyme and the role of certain residues in the active site of the enzyme are known, a semi-rational protein engineering design was done.

There are several options for molecular docking. In this study SwissDock was selected because it does not require computational power. However, waiting for the results slowed the evolution of the project (refer to Figure 6) and the results for the docking of the mutants with the cofactor took several days to be released and AutoDock Vina was used instead.

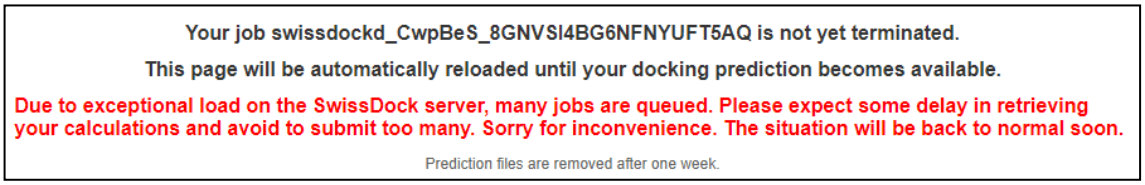


Figure 6. Announcement of SwissDock server

1.6. Description of Contents

This dissertation is divided in 6 sections and 3 appendices.

- Section 1(1. Introduction) explains the context for the research and the time planning of this project.
- Section 2 (2. Research Method) focuses on the methods used to carried out the research.
- Section 3 (3. Results and Discussion) summarizes and explains the results for each stage of the project.
- Section 4 (4. Conclusions and Recommendations) states the conclusions and recommendations for future work.
- Section 5 (5. Glossary) is a glossary with relevant concepts.
- Section 6 (6. References) collects relevant references.
- The appendices A, B and C collect all the raw data.

1.7. Short summary of the Final Result

Among the protein mutants generated, the mutant 2NADa_S380G (change the Ser 380 residue for a Gly) allows the nicotinamide group of the biomimetic cofactor P3NAH to locate in a close position to relevant residues for the reaction with CO_2 (or formic acid) when compared to the position in the WT enzyme.

2. Research Method

2.1. Research Design and Rationale

The objective of this project is to create a mutant of the enzyme formate dehydrogenase from PsFDH that can use a biomimetic cofactor instead of the natural cofactor NADH.

The first step is to select the biomimetic cofactor that shows the best fitting in the active site of the enzyme among the biomimetic cofactors with a suitable reduction potential previously selected. In order to predict the binding location of the biomimetic cofactor, molecular docking is done among the wild type enzyme and the different biomimetic cofactors.

Once the biomimetic cofactor is selected, several protein mutants are created. A rational design of the protein is done. Based on the predicted binding location among the wild type enzyme and the cofactor, residues in the active site are modified in order to accommodate the biomimetic cofactor. The structure of the mutants is determined using homology modelling. With the new structure, molecular docking is done to evaluate the binding affinity with the biomimetic cofactor.

A mutant that shows improvement for binding the biomimetic cofactor is selected. The viability of this mutant is determined by studying the protein stability and the ligand transportation within the enzyme tunnel.

A flow diagram of the work pipeline is indicated in Figure 7.

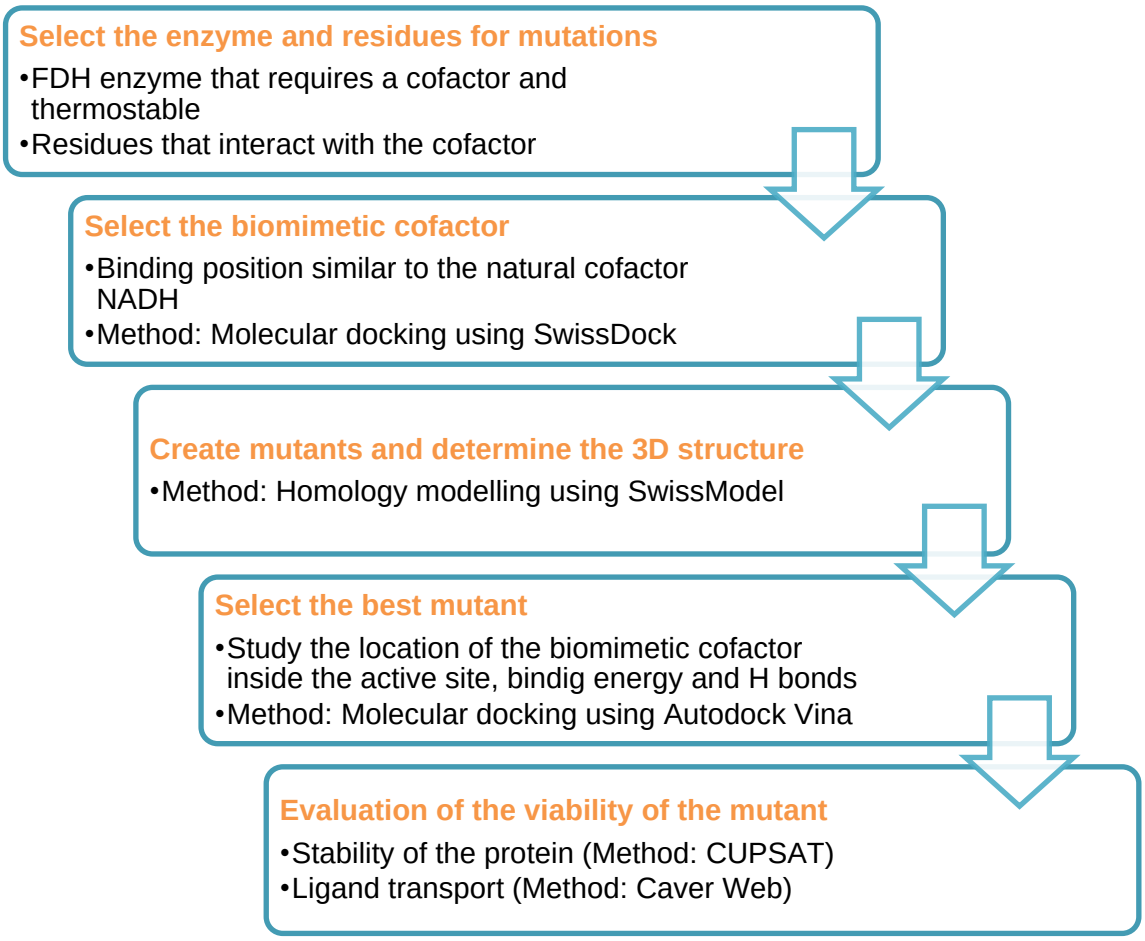


Figure 7. Flow diagram of the work pipeline.

2.2. Methodology and resources

2.2.1. Enzyme preparation

2.2.1.1. Enzyme selection

First, a list of the structures of the enzyme FDH available in PDB was created. Since this study is highly dependent on the structure of the enzyme, only enzymes with a 3D structure reported were considered. Among them, the enzyme that uses a cofactor and shows the best thermostability is selected.

2.2.1.2. Determine the cavities of the enzyme

The cavities of the enzyme and residues within a distance of 5A of the cofactor and substrate were studied using Chimera.

Other options to determine the cavities are:

- The web server CaverDock (Filipovic et al., 2019): https://loschmidt.chemi.muni.cz/caverdock/
- The web server fPocket (Le Guilloux et al., 2009):

https://github.com/Discngine/fpocket

2.2.1.3. Determine the residues in the active site and among enzyme subunits

The residues of the active side that are involved in the reaction with the enzyme substrate are reported in the bibliography. These residues are excluded from mutations to preserve the enzyme activity.

Residues involved in the interaction of the enzyme subunits are also reported in the bibliography and are excluded from mutations to preserve the 3D structure of the enzyme.

2.2.2. Biomimetic cofactor preparation

2.2.2.1. Create a list of the possible cofactors to be used

Only cofactors with a suitable reduction potential for the reduction of CO_2 were selected.

2.2.2.2. Find the structures of the cofactors in .mol2 format

The SMILES code of each cofactor was retrieved from PubChem and the .mol2 structure was created using Chimera.

2.2.2.3. Selection of the biomimetic cofactor

Molecular docking among each biomimetic cofactor and the structure of one subunit of the enzyme PsFDH was done using SwissDock (Grosdidier et al., 2011): http://www.swissdock.ch/

All the subunits of the enzyme PsFDH share the same structure so in order to speed up the simulations, docking was done in only one subunit. The steps are:

1. Preparation of the protein as a .pdb file.
    a. The experimental structure 2NAD appears in PDB.
    b. Only the monomer A is selected for docking studies.
    c. All the groups (azide, PEG, water and NADH) are removed using Chimera.
    d. Before docking hydrogens are added to the protein using Dock Prep in Chimera.
    e. The monomer of the protein is saved as a .pdb file.
2. Preparation of the ligand as a .mol2 file.
    a. The SMILES code is generated using PubChem.
    b. With the SMILES code, the 3D structure is created in Chimera.
    c. Hydrogens are added using Dock Prep in Chimera
    d. The ligand is saved as a .mol2 file.
3. Results are submitted in SwissDock specifying the parameters as indicated in Figure 8:
    a. Docking type: Accurate
    b. Region of interest: X center 48, Y center 39, Z center -6, X size 25, Y size 30 and Z size 30.
    c. Flexibility for side chains within 5 Å for all the biomimetic cofactors. In

the case of docking NADH, a flexibility of 3 Å was considered because SwissDock did not generate results for 5 Å.

4. Results were analyzed using Chimera.

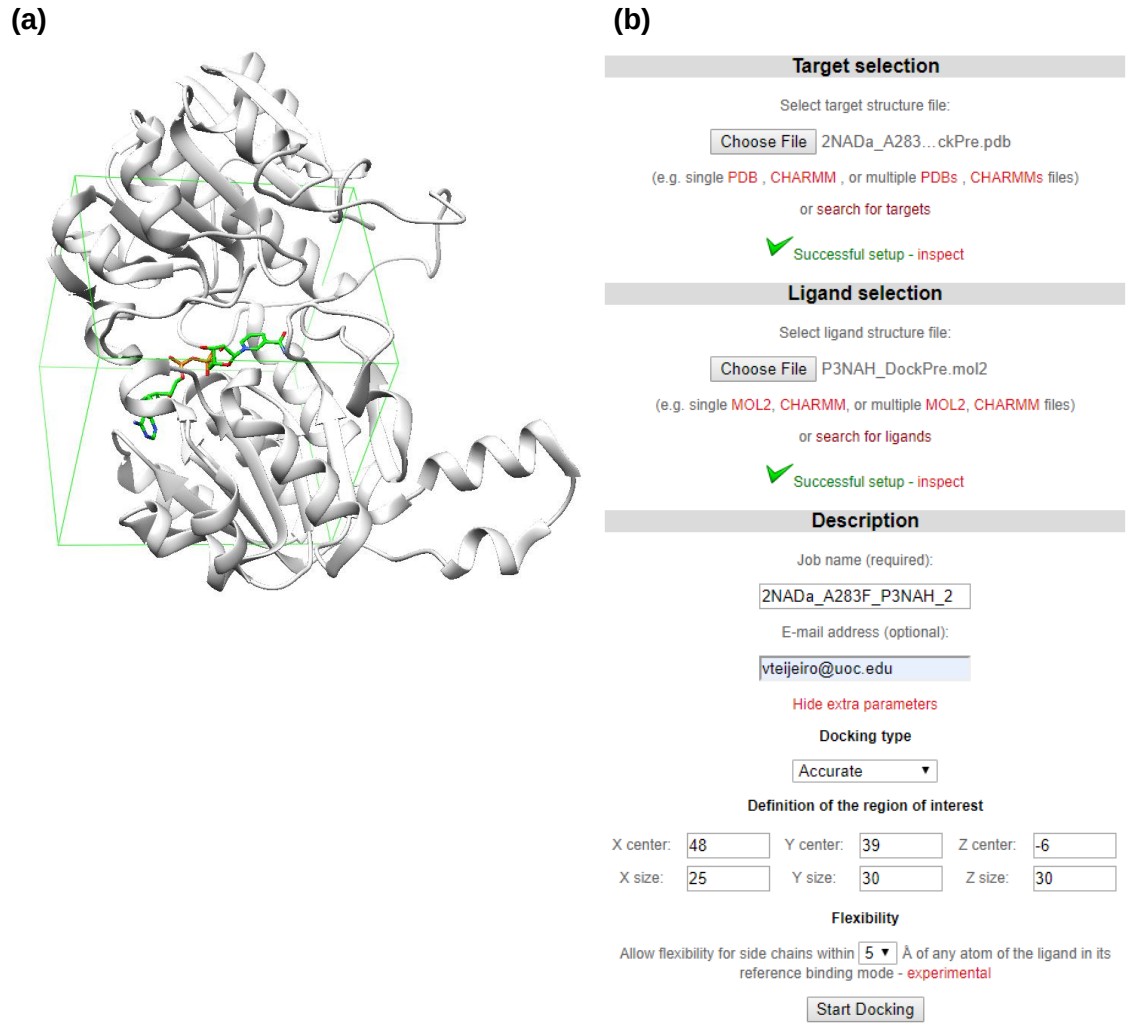


Figure 8. Docking specifications (a) Structure of the protein with the selected target volume, (b) Parameters in SwissDock.

Alternatives to SwissDock are:

- MTiOpenScreen(Labbé et al., 2015): http://bioserv.rpbs.univ-paris-diderot.fr/services/MTiOpenScreen/
- AutoDock Vina

The biomimetic cofactor that suits best the active is selected for further studies.

2.2.3. Protein engineering of FDH

2.2.3.1. Strategy for rational protein design

A total number of eight mutants were created following two different approaches.

- In the first approach, residues were modified to promote π–π stacking interactions among the protein and the aromatic group of the cofactor nicotinamide. This approach is summarized in Figure 9
- In the second approach, big residues in the entrance tunnel of the cofactor were changed to glycine. This approach is summarized in Figure 10.

Based on the docking results between the wild type enzyme and the cofactor P3NAH, residues near the location of the aromatic group of the cofactor (the nicotinamide group) are modified in order to promote π–π stacking interactions hoping for stabilization of the cofactor in that area and proper orientation for reaction with the substrate of the enzyme. The new residues introduced are phenylalanine and tyrosine. Tyrosine has a OH group that could create further stabilization if participating in a H bond. Tryptophan is another residue that has an aromatic group but because of its bigger size it can create steric effects keeping the cofactor from reaching the desired location so it was not selected for this experiment. These mutants are:

- Mutant 2NADa_A283F: Ala 283 changes to Phe
- Mutant 2NADa_A283Y: Ala 283 changes to Tyr
- Mutant 2NADa_G123F: Gly 123 changes to Phe
- Mutant 2NADa_G123Y: Gly 123 changes to Tyr

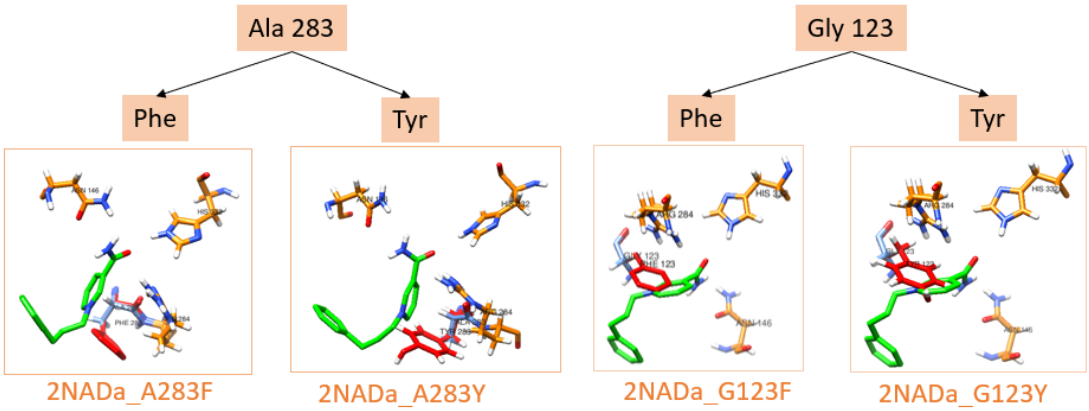


Figure 9. Mutants for promoting π–π stacking interactions.

The other set of mutations were orientated to create a wider cofactor binding groove. Since the biomimetic cofactor P3NAH is shorter than the natural cofactor NADH residues that participate in the stabilization of the phosphate and adenine group of NADH are not required. In fact, several of the residues in the tunnel can avoid the cofactor for reaching a lower conformation energy. Due to its small size, glycine was selected as the mutation to introduce. These mutants are:

- Mutant 2NADa_T376G: Thr 376 changes to Gly
- Mutant 2NADa_S380G: Ser 380 changes to Gly
- Mutant 2NADa_Y381G: Tyr 381 changes to Gly
- Mutant 2NADa_R222G: Arg 222 changes to Gly

All the residues selected for mutations are among the ones selected as “viable for mutations” in the previous study (refer to the section Prediction of beneficial mutations).

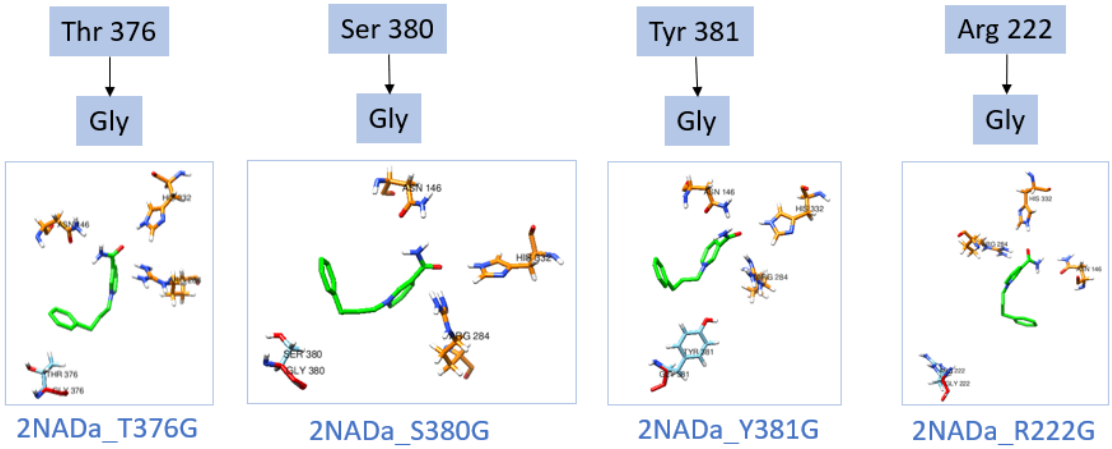


Figure 10. Mutants with a wider cofactor binding groove.

2.2.3.2. Creating mutants

Considering the best docking conformation of the biomimetic cofactor P3NAH in the wild type enzyme, residues around the cofactor area were selected for mutations. The FASTA sequence of the wild type enzyme was retrieved using Chimera and a certain amino acid was modified. The new FASTA sequence was submitted to the web server SWISS-MODEL(Waterhouse et al., 2018) for homology modelling for getting the 3D structure of the mutant enzyme.

Other options are available for the prediction of the structure of the protein after mutations:

- The web server DALI (Holm, 2019): http://ekhidna2.biocenter.helsinki.fi/dali/
- The web server I-TASSER (Yang & Zhang, 2015): https://zhanglab.ccmb.med.umich.edu/I-TASSER/

2.2.3.3. Analysis of binding between P3NAH and the mutants

After the 3D structure of each mutant was retrieved, molecular docking with the cofactor P3NAH was done using AutoDock Vina(Trott & Olson, 2010) in Chimera. Among the cofactor conformations provided by Chimera, the one with the closer similarity in the location of the nicotinamide group with the position of NADH in the wild enzyme was selected.

The successful mutant was selected considering the location of 3 carbons of the nicotinamide group of the cofactor (C1, C4 and C6) with respect to the atoms of relevant residues for the reaction of the cofactor with the substrate of the enzyme. Specifically, the atom ND2 of Asn 146, CZ of Arg 284 and NE2 His 332. The distance was compared with the atoms of the cofactor NADH when docking to the wild type enzyme. The number of H bonds between the cofactor and the enzyme and the binding score was also taking into consideration. Figure 11 shows the labelling of the atoms for the cofactor and the 3 residues considered.

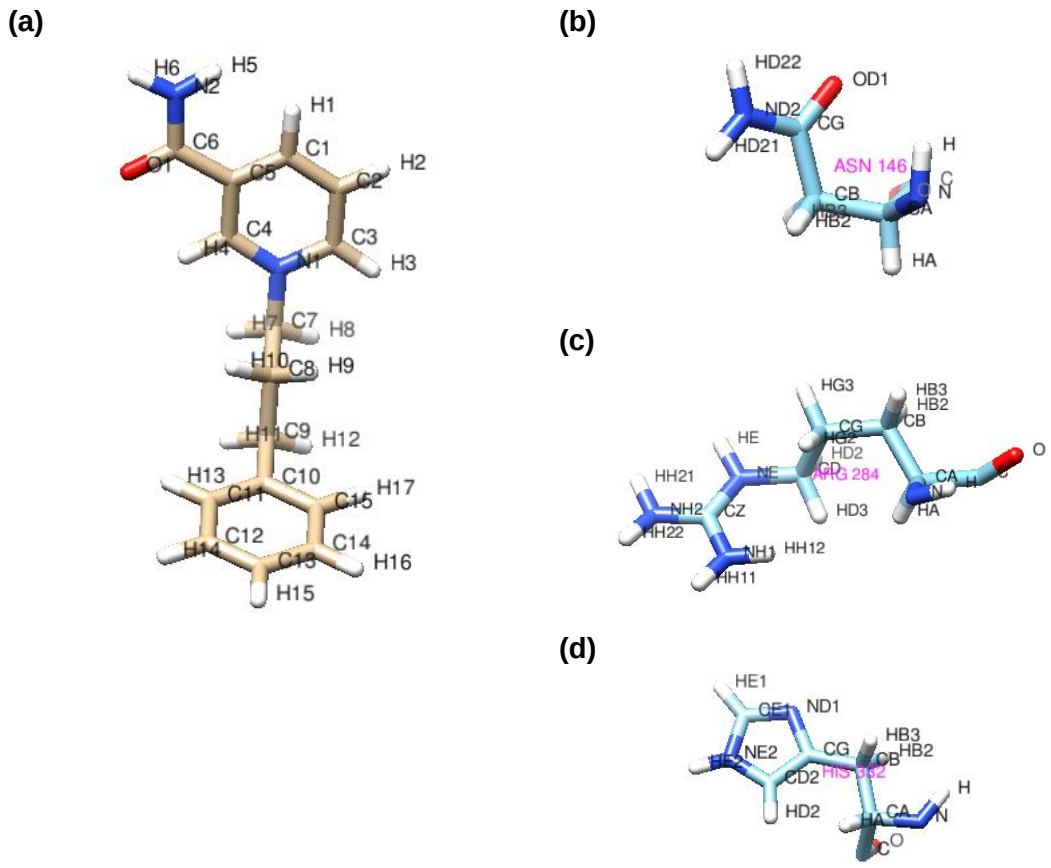


Figure 11. Atom labelling for (a) P3NAH, (b) Asn 146, (c) Arg 284 and (d) His 332.

2.2.4. Evaluation of the viability of the selected mutant

2.2.4.1. Protein stability

The stability of the protein after mutation was evaluated using the web server CUPSAT(Parthiban et al., 2006), accessible in http://cupsat.tu-bs.de/index.jsp.

CUPSAT is a tool to predict changes in protein stability upon point mutations. The prediction model uses amino acid-atom potentials and torsion angle distribution to assess the amino acid environment of the mutation site. Additionally, the prediction model can distinguish the amino acid environment using its solvent accessibility and secondary structure specificity.

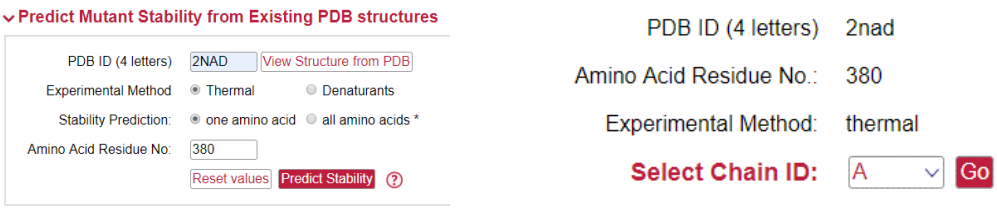


Figure 12. Data provided to CUPSAT for the evaluation of the stability of the mutant 2NADa_S380G.

2.2.4.2. Analysis of ligand transport

The study of the transport of the biomimetic cofactor within the tunnel of the mutant 2NADa_S380G was done using Cave Web(Stourac et al., 2019), accessible as a web server at https://loschmidt.chemi.muni.cz/caverweb/.

In Caver Web the starting point to calculate the tunnels must be indicated. It was found using Chimera. The starting point is the deepest part of the binding site. It changes for the wild and mutant enzyme because of changes in residues.

An alternative is CaverDock(Filipovic et al., 2019).

2.2.5. Additional resources

2.2.5.1. Visualization of the protein

Visualization of the protein was performed using UCSF Chimera(Pettersen et al., 2004): https://www.cgl.ucsf.edu/chimera/index.html

2.2.5.2. Gantt diagram

The Gantt diagrams were created with teamgantt (https://www.teamgantt.com/).

3. Results and Discussion

3.1. Selection of the FDH enzyme for this study

The PsFDH (formate dehydrogenase from *Pseudomonas sp. 101*) was selected as the enzyme to use in this project. This is an enzyme that requires a cofactor (NADH is the natural cofactor) and it shows the highest tolerance to high temperatures among all the structures reported in PDB.

3.1.1. Available crystal structures in PDB

The crystal structure of the protein is required to study how the mutations affect the docking with the substrate and cofactor. It could be possible to use a software to predict the structure based on the primary sequence that can be obtained from the gene. However, predictions of the tertiary structure of proteins based on the primary structure are not 100% reliable. Starting the modifications in a tertiary structure reported is more reliable.

3.1.2. Use of NADH cofactor

Only FDHs that work in the presence of NADH cofactor were considered for this study. These enzymes are metal-independent and stable under O_2. Stability under oxygen conditions is desirable specially when targeting CO_2 reduction as an application.

3.1.3. Thermostability

Thermostable enzymes are preferable when doing protein engineering because they can tolerate more changes in residues before losing the activity. PDB does not have the crystal structure of any thermostable FDH. Considering the experimental data previous published for T_m (temperature which provides with 50% inactivation in 20 min), and limiting to the crystal structures available in PDB, PsFDH is selected.

Table 3. Selection of FDH for this study.

Enzyme	PDB ID	Metal independent	Experimental T_m (°C)
ArFDH	3JTM	Yes	No information
CbFDH	5DNA	Yes	56.8 (Tishkov & Popov, 2006)
PsFDH	2GO1	Yes	63(Tishkov & Popov, 2006)
GrFDH	4XYG	Yes	55(Fogal et al., 2015)
TsFDH	3WR5	Yes	52.5(Tishkov & Popov, 2006)
MorFDH	2GSD	Yes	58(Tishkov & Popov, 2006)

3.1.4. Kinetic studies for CO_2 reduction

A recent review compared the kinetics of previous studies for the CO_2 reduction. For the class of non-metal FDHs, the enzyme from the organism *Pseudomonas oxalaticus* (PoFDH) provided the most favorable k_{cat}. The selected enzyme for this study, seems to be also a good candidate for CO_2 conversion.

Table 4. Overview of FDHs described in the literature to reduce CO_2 (Nielsen et al., 2019).

ID	Organism	Electron donor	Donor conc (mM)	Exp. pH	Class	Type	k_{cat} (red) (s^{-1})	K_m CO_2 (mM)	k_{cat}/K_m	Reference
SfFDH	*Syntrophobacter fumaroxidans*	MV^{2+}	1		Metal	2	282	–	–	(De Bok et al., 2003)
DdFDH	*Desulfovibrio desulfuricans*	MV^{2+}	0.2	7	Metal	2	47	0.02	2968	(Maia et al., 2016)
EcFDH (FDH-H)	*Escherichia coli*	MV^{2+}	0.1	7	Metal	1	1	8.3	0.1	(Bassegoda et al., 2014)
DvFDH	*Desulfovibrio vulgaris*	MV^{2+}	1		Metal	4	3	–	–	(da Silva et al., 2011)
AwFDH (FDH2)	*Acetobacterium woodii*	MV^{2+}	5	7	Metal	2	372	3.8	97.9	(Schuchmann and Müller, 2013)
SfFDH	*Syntrophobacter fumaroxidans*	Electrode	–		Metal	2	112	–	–	(Reda et al., 2008)
EcFDH (FDH-H)	*Escherichia coli*	Electrode	–		Metal	1	112	8.3	13.5	(Bassegoda et al., 2014)
EcFDH (FDH-H)	*Echerichia coli*	Electrode	–		Metal	1	–	2.5		(Yuan et al., 2018)
DdFDH	*Desulfovibrio desulfofuricans*	Electrode	–		Metal	2	–	–	–	(Cordas et al., 2019)
DvFDH	*Desulfovibrio vulgaris*	Electrode	–		Metal	3	11	–	–	(Miller et al., 2019)
AwFDH	*Acetobacter woodii*	Hydrogen	–		Metal	2	28	3.8	7.4	(Schuchmann and Müller, 2013)
TkFDH	*Thermoanaerobacter kuvui*	Hydrogen	–		Metal	2	2654	–	–	(Schwarz et al., 2018)
CbFDH	*Candida boidinii*	NADH	1	7.5	Non-metal	–	0.1	–	–	(Altas et al., 2017)
CbFDH	*Candida boidinii*	NADH	0.15	7.1	Non-metal	–	0.02	2.6	0.01	(Choe et al., 2014)
TsFDH	*Thiobacillus sp. KNK65MA*	NADH	0.15	7	Non-metal	–	0.3	0.95	0.34	(Choe et al., 2014)
CcFDH	*Clostridium carboxidivorans*	NADH	0.2		Metal	6	0.08	–	–	(Alissandratos et al., 2013)
RcFDH	*Rhodobacter capsulatus*	NADH	0.2		Metal	5	1.48	–	–	(Hartmann and Leimkühler, 2013)
PoFDH	*Pseudomonas oxalaticus*	NADH	100		Non-metal	–	3	40	0.08	(Ruschig et al., 1976)
CnFDH	*Cupriavidus necator*	NADH	0.2	7	Metal	5	11	2.7	4.07	(Yu et al., 2017)
MtFDH	*Myceliophthora thermophila*	NADH	1	6	Non-metal	–	0.1	0.44	0.23	(Altas et al., 2017)
CtFDH	*Chaetomium thermophilum*	NADH	1	5	Non-metal	–	0.02	3.29	0.01	(Altas et al., 2017)
CmFDH	*Candida methylica*	NADH	1	8	Non-metal	–	0.01	0.01	1.00	(Altas et al., 2017)

3.2. Characterization of the enzyme PsFDH

3.2.1. Mechanism for CO_2 reduction

Knowing the mechanism of reaction of the wild enzyme with the natural cofactor is important to determine which residues to modify.

In the case of non-metal FDH, the mechanism for formate oxidation and CO_2 reduction is shown in Figure 13. This is a reversible reaction and the steps in the case of formic acid oxidation are(Amao, 2018):

- First the cofactor joins the cofactor-binding site. Two hydrogen and carbonyl oxygen of the amide group of NAD^+ are trapped by hydrogen bond among these groups and amino acid residues (Thr, Asp and His), respectively.
- Next, formic acid joins the active site. The negatively charged formic acid is electrostatically trapped by positively charged arginine of amino acid residues and is trapped by hydrogen bond between oxygen of formic acid and His amino acid residue.
- At the active site, formic acid oxidation to CO_2 and NAD^+ reduction to NADH proceeds catalytically.
- Finally, CO_2 and NADH are released from active-site and active-site returns to its original state.

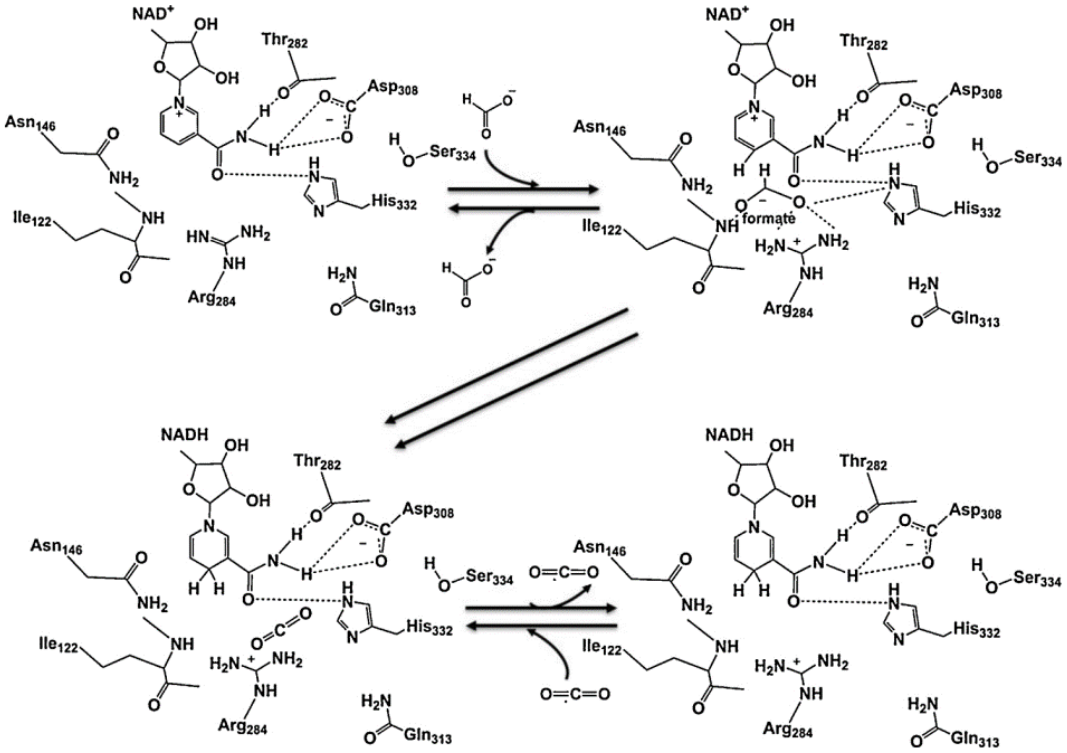


Figure 13. Mechanism of formic acid oxidation to CO_2 and CO_2 reduction to formic acid with non-metal FDH(Amao, 2018).

The NAD^+- dependent FDHs follow a hydride (H) transfer mechanism (see Figure 14). Particularly, the direct hydride transfer, which is the rate-determining step (RDS) of the overall catalytic cycle, occurs from the C-atom of the formate to the C4-atom of pyridine ring of NAD+, leading to NADH and CO_2. Although the reaction is reversible, the reaction rate for CO_2 reduction to formate is low. Mechanistically, the reaction involves a charge neutralization process of NAD+ and is therefore supposed to be governed by the electrostatics. In line with this reasoning, the neighbouring residues like Thr, Asp, Ser and Gly interact with the positively charged NAD+ ring and stabilize the positively charged ring, thereby promoting the hydride transfer from formate to C4(Mondal et al., 2015).

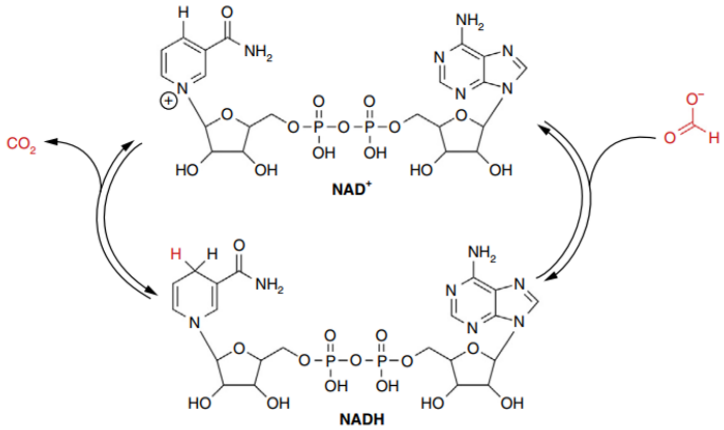
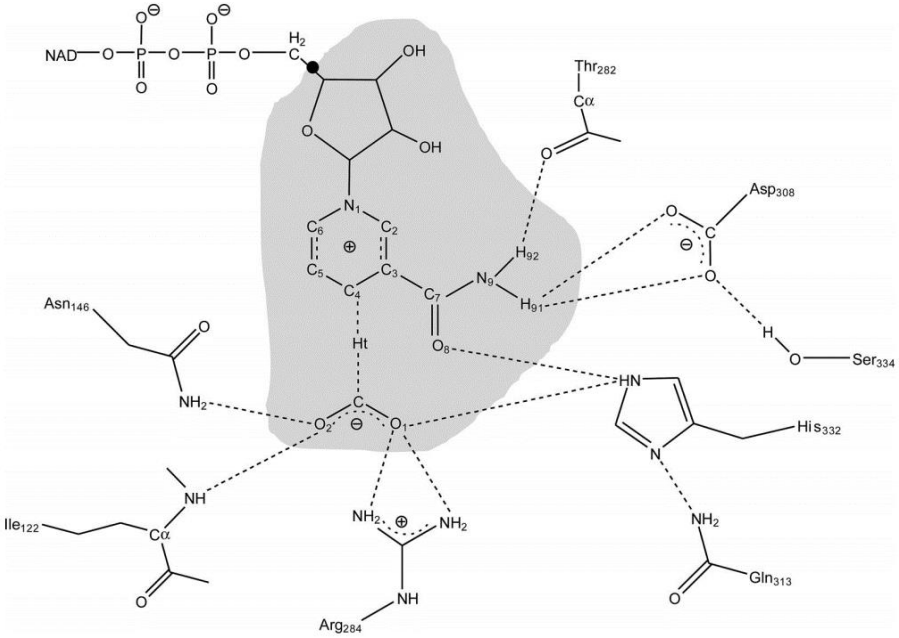


Figure 14. Hydride transfer mechanism(Mondal et al., 2015).

The enzyme compresses the substrate and the cofactor into a conformation close to the transition structure by means of favorable interactions with the amino acid residues of the active site, thus facilitating the relative orientation of donor and acceptor atoms to favour the hydride transfer as indicated in Figure 15. Moreover, a permanent field created by the protein reduces the work required to reach the transition state (TS) with a concomitant polarization of the cofactor that would favour the hydride transfer. In contrast, in water the TS is destabilized with respect to the reactant species because the polarity of the solute diminishes as the reaction proceeds, and consequently the reaction field, which is created as a response to the change in the solute polarity, is also decreased. Therefore protein structure is responsible of both effects; substrate preorganization and TS stabilization thus diminishing the activation barrier(Castillo et al., 2008).

The residues that cannot be modified because they are involve in the reaction are indicated in Figure 15 and their location relative to the cofactor and substrate appears in Figure 16 and Figure 17, respectively.

(a)



(b)

ILE 122
ASN 146
THR 282
ARG 284
ASP 308
GLN 313
HIS 332
SER 334

Figure 15. (a) details of the FDH active site(Castillo et al., 2008), (b) Relevant residues for the reaction mechanism.

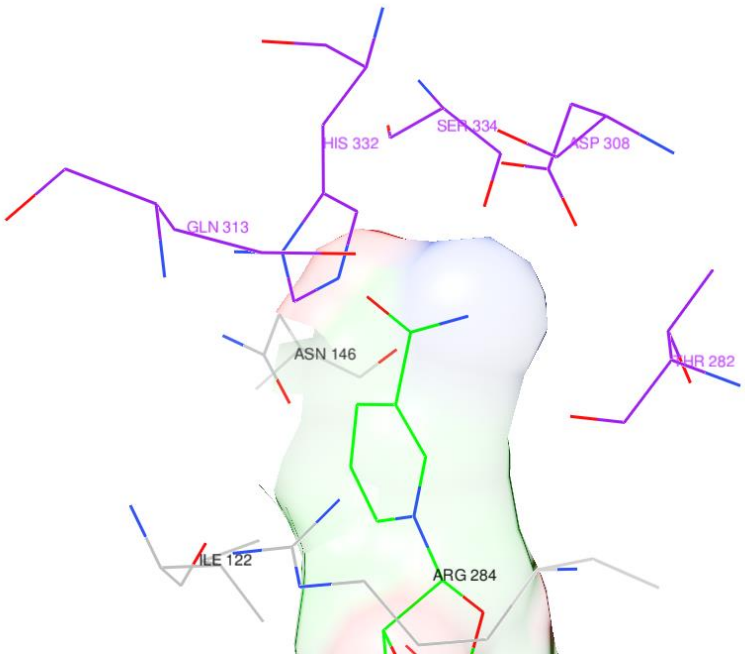


Figure 16. Relevant residues for the reaction mechanisms that interact with the nicotinamide group of the cofactor

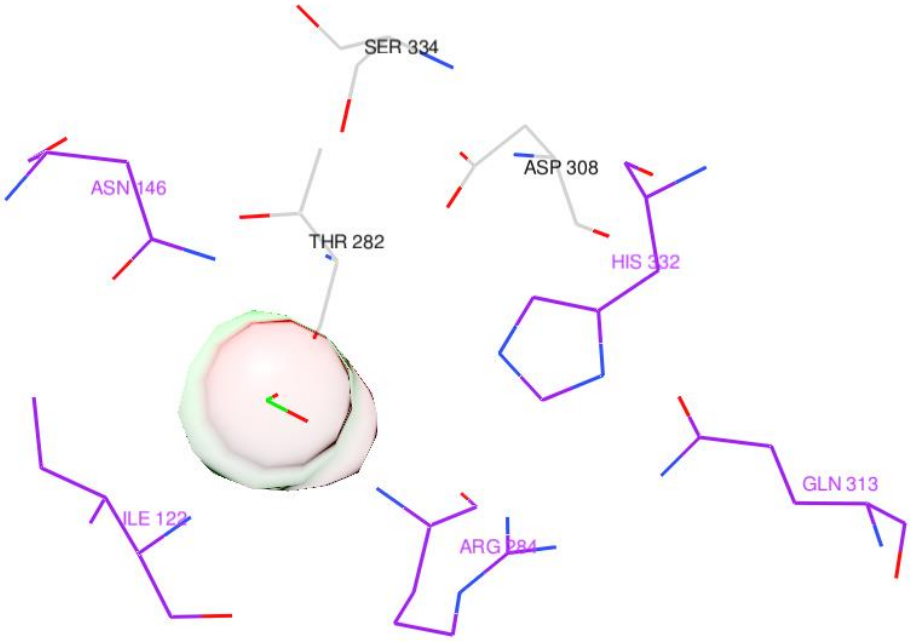


Figure 17. Relevant residues for the reaction mechanisms that interact with formic acid

3.2.2. Study of the enzyme structure

3.2.2.1. General concepts of PsFDH structure

The enzyme formate dehydrogenase from *Pseudomonas sp.* 101 belongs to the family of *D*-specific dehydrogenases of 2-hydroxy acids. It is from the class of α/β proteins. In the cell, this enzyme appears as a biologically active homodimer. Four dimers pack together (Filippova et al., 2005), as indicated in Figure 18.

Each subunit of the dimer consists of a globular two-domain structure, with the coenzyme-binding domain and the catalytic domains. Both domains have a similar structure, consisting of a parallel left-twisted β-sheet surrounded by α-helices.

The structure analysis in this project is based on the structures retrieved from PDB. The structures available in PDB of the enzyme formate dehydrogenase from *Pseudomonas sp*. 101 are:

- 2GO1: 1 subunit of PsFDH in complex with a sulfate ion (apoenzyme)
- 2GUG: 2 subunits PsFDH in complex with formate (holoenzyme)
- 2NAC: 4 subunits of PsFDH in complex with a sulfate ion (apoenzyme)
- 2NAD: 4 subunits of PsFDH in complex with NADH, azide ion and sulfate ion (holoenzyme)

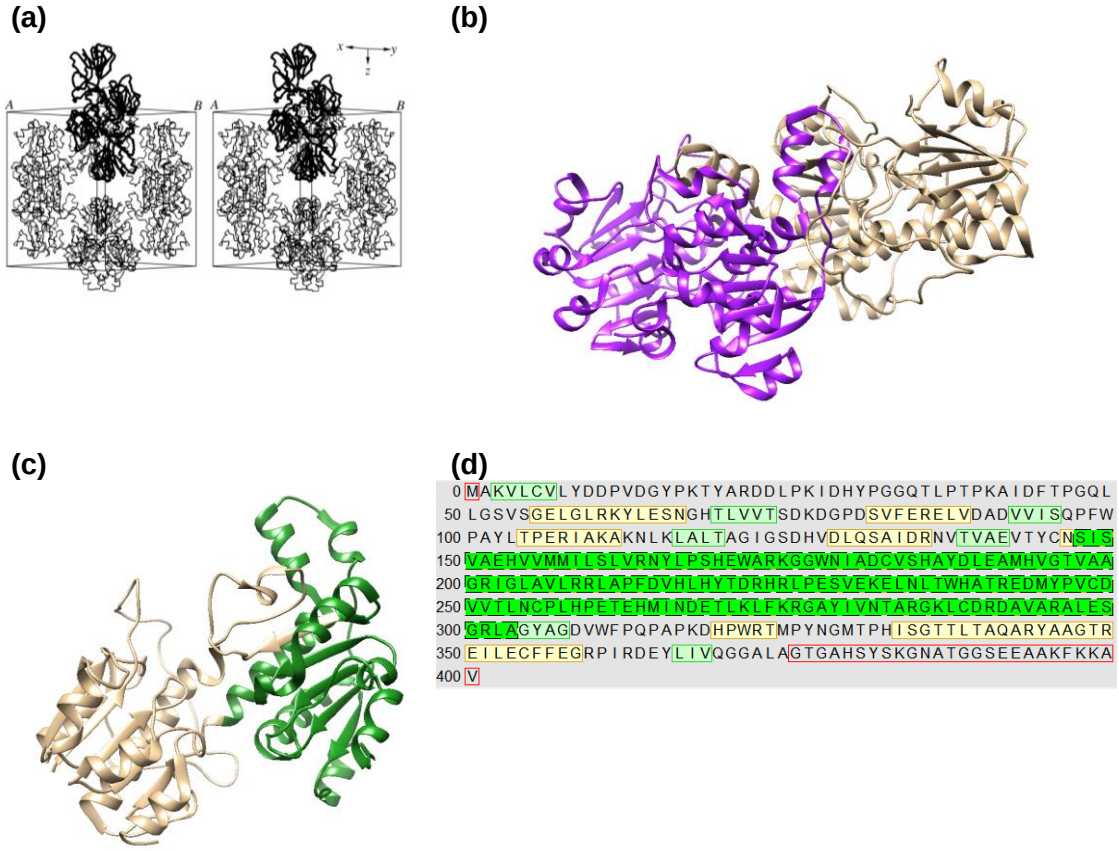


Figure 18. (a) Stereo diagram of the packing of FDH molecules in the unit cell (Filippova et al., 2005), (b) Structure of the dimer, (c) Structure of each enzyme with the catalytic domain in tan and the cofactor-binding domain in green, (d) Sequence of the enzyme with the cofactor-binding domain highlighted in green.

3.2.2.2. Coenzyme-binding domain

The coenzyme-binding domain is composed by the residues 147-333, indicated in Figure 19. The secondary structures in this domain consists of (Filippova et al., 2005):

- a β-sheet of seven parallel β-chains (colored in orange in the next figure)
- α-helices that link the β-sheet (colored in dark blue in the next figure)

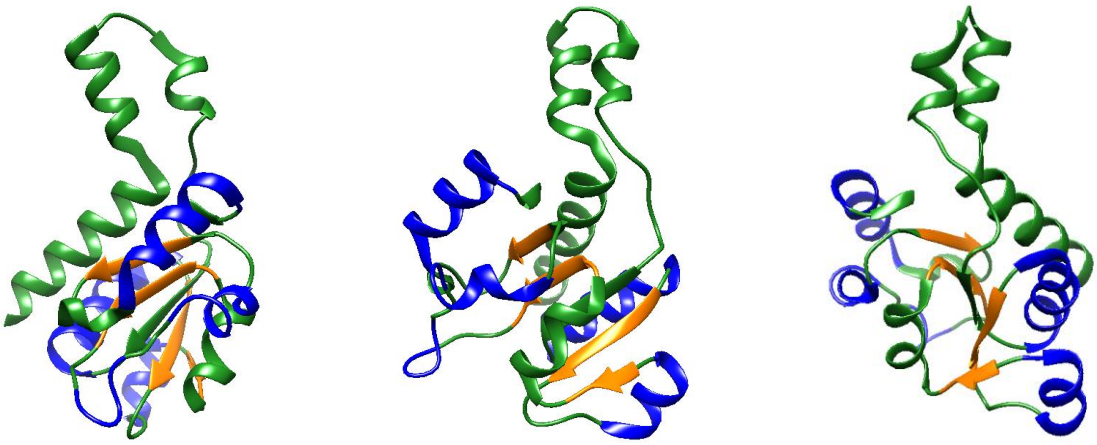


Figure 19. Structural analysis of the coenzyme-binding domain of PsFDH from 3 different perspectives.

In the holoenzyme, after the cofactor binds the enzyme, other residues cover the pocket to protect the reaction from the external solvent. Figure 20 shows a dimer with the residues of the coenzyme-binding domain highlighted. In the case of the holoenzyme, the cofactor is included in the structure.

When doing docking, the structure of the apoenzyme will be used to determine if the biomimetic cofactor can access through the channel. However, the structure of the holoenzyme will be used to determine if the cofactor can assemble properly in the active site facilitating contact with the substrate.

(a) Apoenzyme structure **(b) Holoenzyme structure**

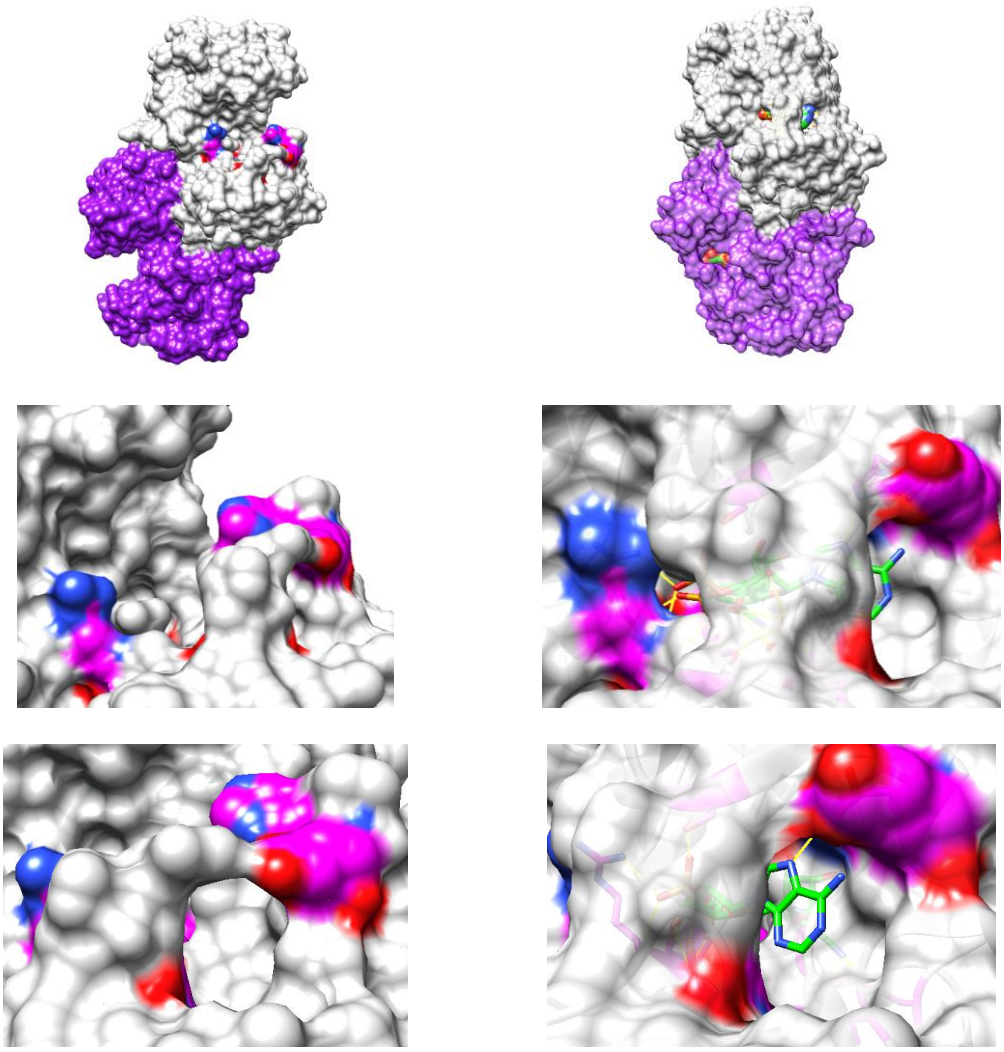
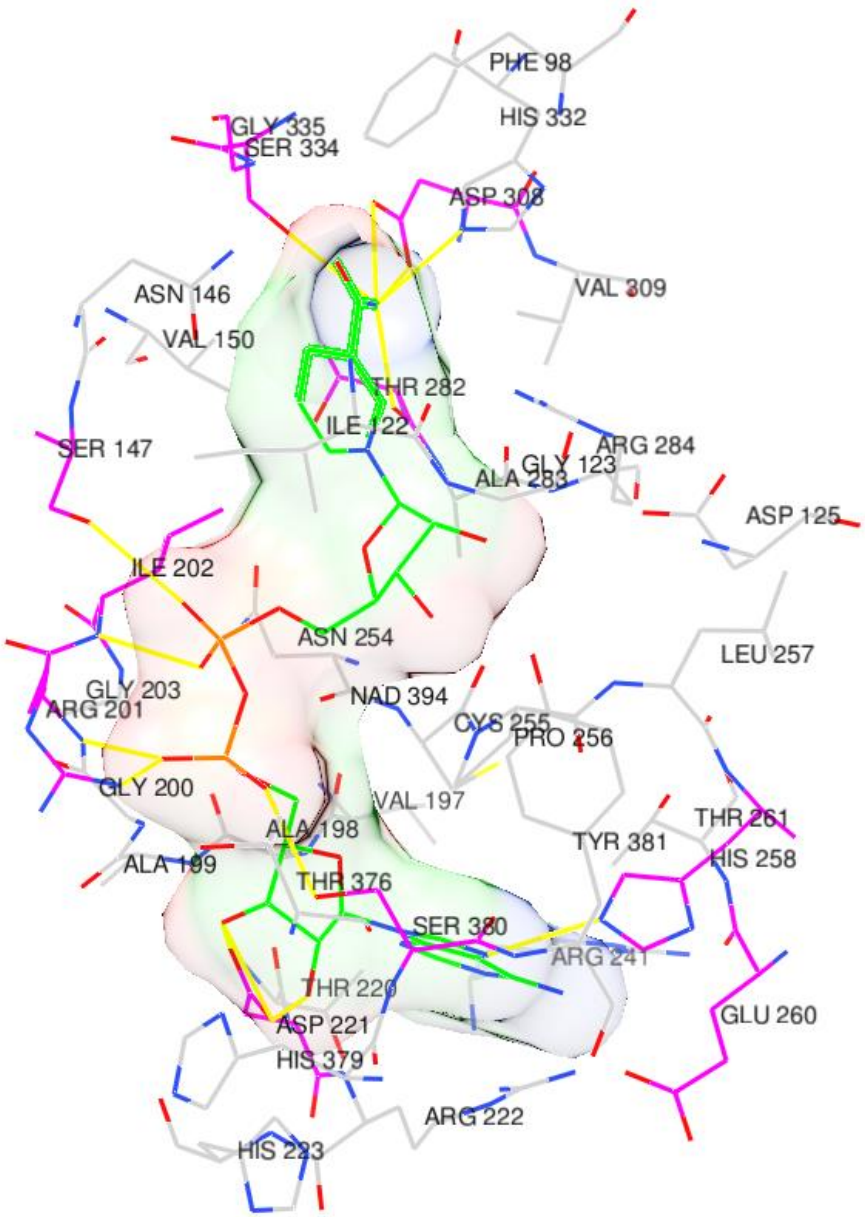


Figure 20. External views of the pocket where the cofactor binds the enzyme for the (a) apoenzyme and (b) holoenzyme. The highlighted residues in magenta were reported to be relevant for the reaction.

In the pocket where the coenzyme binds the enzyme, there are 38 residues within a distance of 5Å from the cofactor. The next figure indicates the residues around the cofactor. Hydrogen bonds are indicated with a yellow line and the residues reported as relevant in the bibliography are highlighted in magenta.

From this list of residues, the ones important for the catalytic reaction with CO_2 (or formate) and the residues involved in the dimer interactions will be excluded when doing site-directed mutagenesis.

(a)



(b)

PHE 98
ILE 122
GLY 123
ASP 125
ASN 146
SER 147
VAL 150
VAL 197
ALA 198
ALA 199
GLY 200
ARG 201
ILE 202
GLY 203
THR 220
ASP 221
ARG 222
HIS 223
ARG 241
ASN 254
CYS 255
PRO 256
LEU 257
HIS 258
GLU 260
THR 261
THR 282
ALA 283
ARG 284
ASP 308
VAL 309
HIS 332
SER 334
GLY 335
THR 376
HIS 379
SER 380
TYR 381

Figure 21. Residues within 5Å from the cofactor (a) conformation inside the pocket, (b) list.

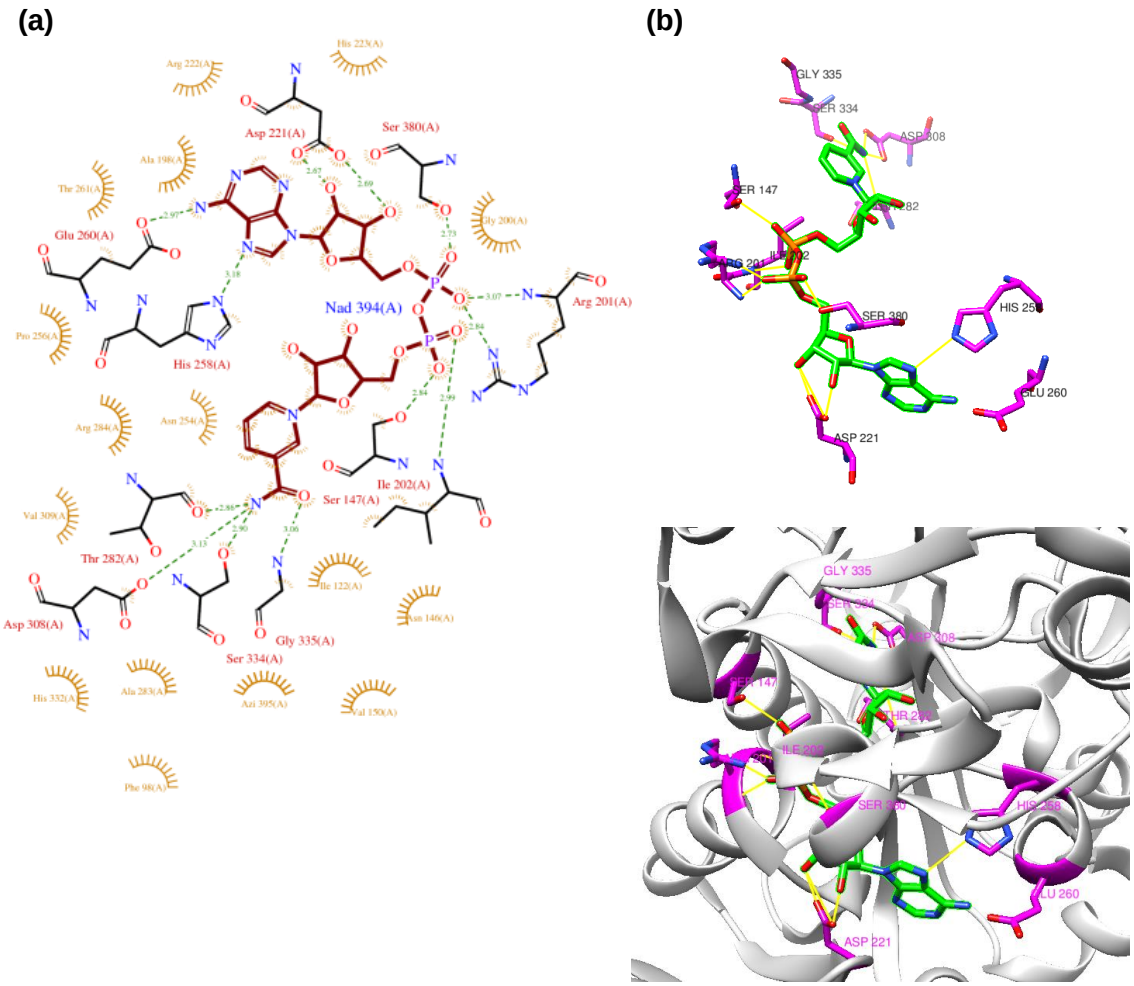


Figure 22. (a) Ligplot of interactions of PsFDH (2NAD) with the cofactor NAD. Retrieved from PDBsum, (b) visualization of the conformation the relevant residues take inside the structure of the enzyme.

3.2.2.3. Catalytic domain

The catalytic domain of the protein is composed of the residues 1-147 and 333-374(Filippova et al., 2005). The secondary structures that integrate the catalytic domain are indicated in Figure 23. The catalytic domain is composed of:

- a β-sheet consisting of five parallel β-chains (orange color)
- four α-helices that surround the β-sheet (dark blue color)
- a disordered loop (magenta color)
- a α-helix that forms part of the C-terminal part of the domain (cyan color)
- a β-hairpin that also forms part of the C-terminal part (red color)

The residues of the catalytic domain will not be modified in this project. It is relevant to know them because some residues from the catalytic domain also form part of the cofactor-binding domain. The residues that are present in both domains are not considered for mutations.

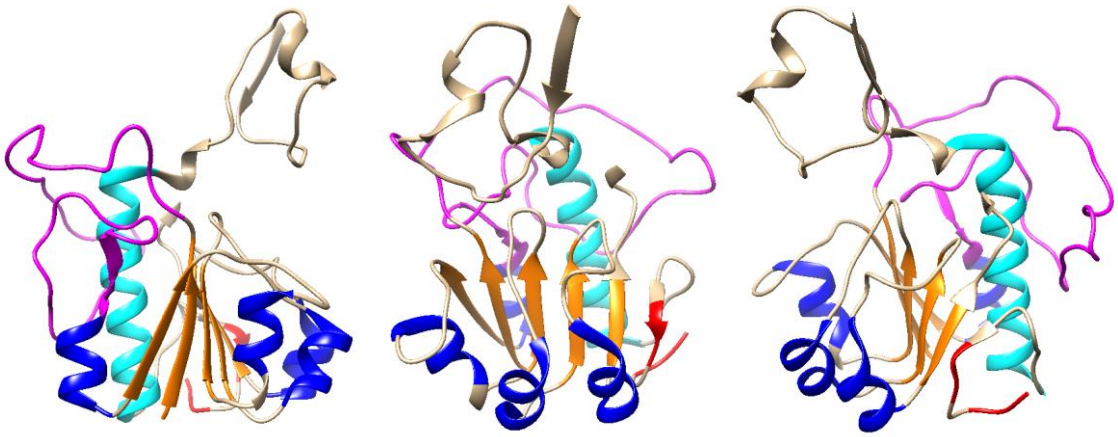
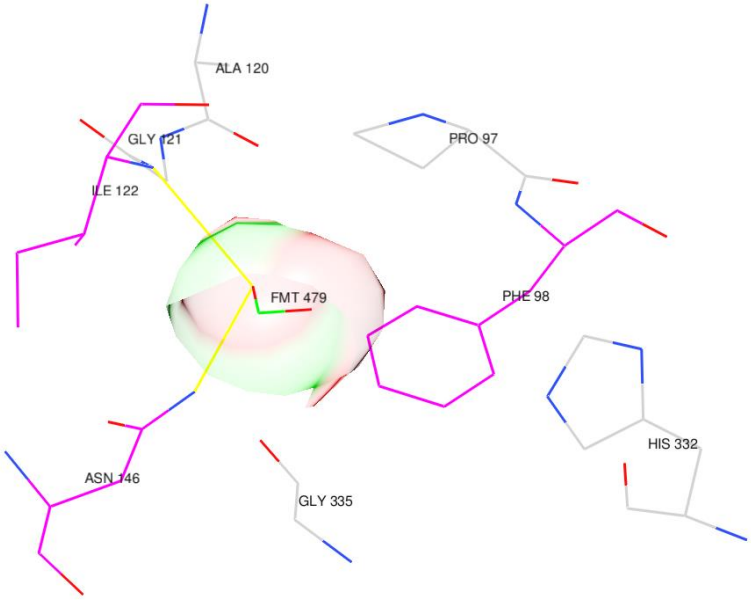


Figure 23. Structural analysis of the catalytic domain of PsFDH from 3 different perspectives.

There are 8 residues within 5Å from the substrate in the catalytic domain. From those, 3 are also part of the cofactor-binding domain. They are indicated in Figure 24.

(a) Residues around the substrate



(c) Residues within 5Å from the substrate:

PRO 97
PHE 98
ALA 120
GLY 121
ILE 122
ASN 146
HIS 332
GLY 335

(d) Residues that belong also to the cofactor-binding domain:

ILE 122
ASN 146
PHE 98

(b) Relevant residues during reaction

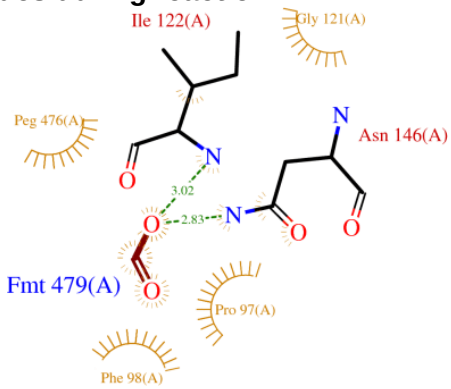


Figure 24. (a) Residues around the substrate, (b) Ligplot of interactions of PsFDH (2GUG) with formic acid, (c) Residues around the substrate, (d) Residues that belong to the formate and cofactor-binding domain. Retrieved from PDBsum.

The catalytic domain is located next to the cofactor-binding domain, as indicated in Figure 25. For this reason, there are residues that belong to both domains. In the figure, the residues that belong to the cofactor-binding domain appear in magenta and the ones that belong to the catalytic domain, in orange.

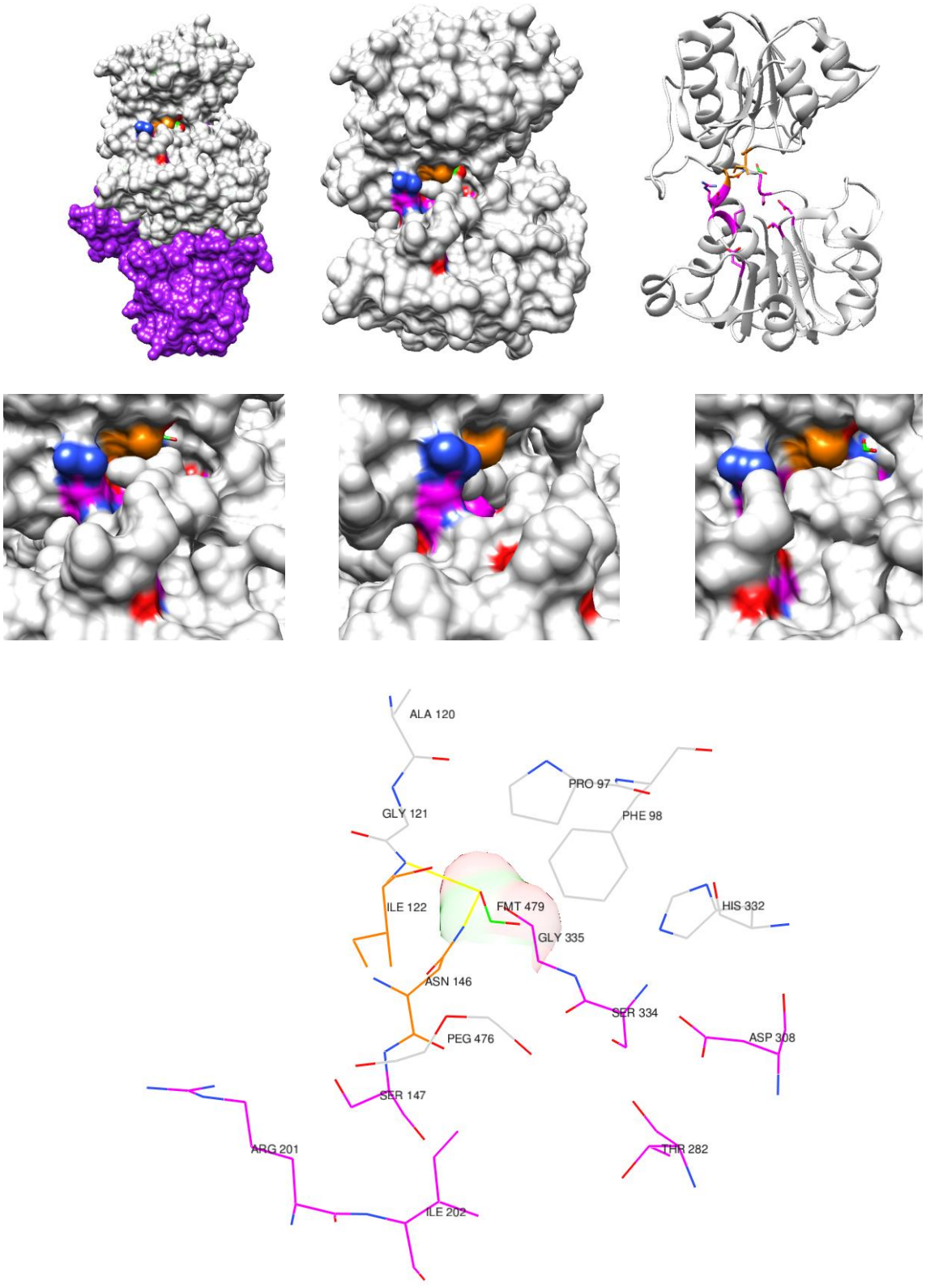


Figure 25. Location of the catalytic domain in relation to the cofactor-binding domain. Residues in magenta belong to the cofactor-binding domain and the ones in orange to the catalytic domain

3.2.2.4. Interactions between the subunits

The enzyme PsFDH appears as a biologically active dimer. The contacts between the molecules within the dimer are substantially stronger than the other contacts. For this pair of molecules, the area of the contact surface is equal to 3891 $Å^2$ and the molecules in a dimer are linked by 27 hydrogen bonds. The interactions between the molecules in the dimer are predominantly due to the residues belonging to the coenzyme-binding domain(Filippova et al., 2005). The residues responsible of these interactions were not modified. They are indicated in Figure 26 and Figure 27.

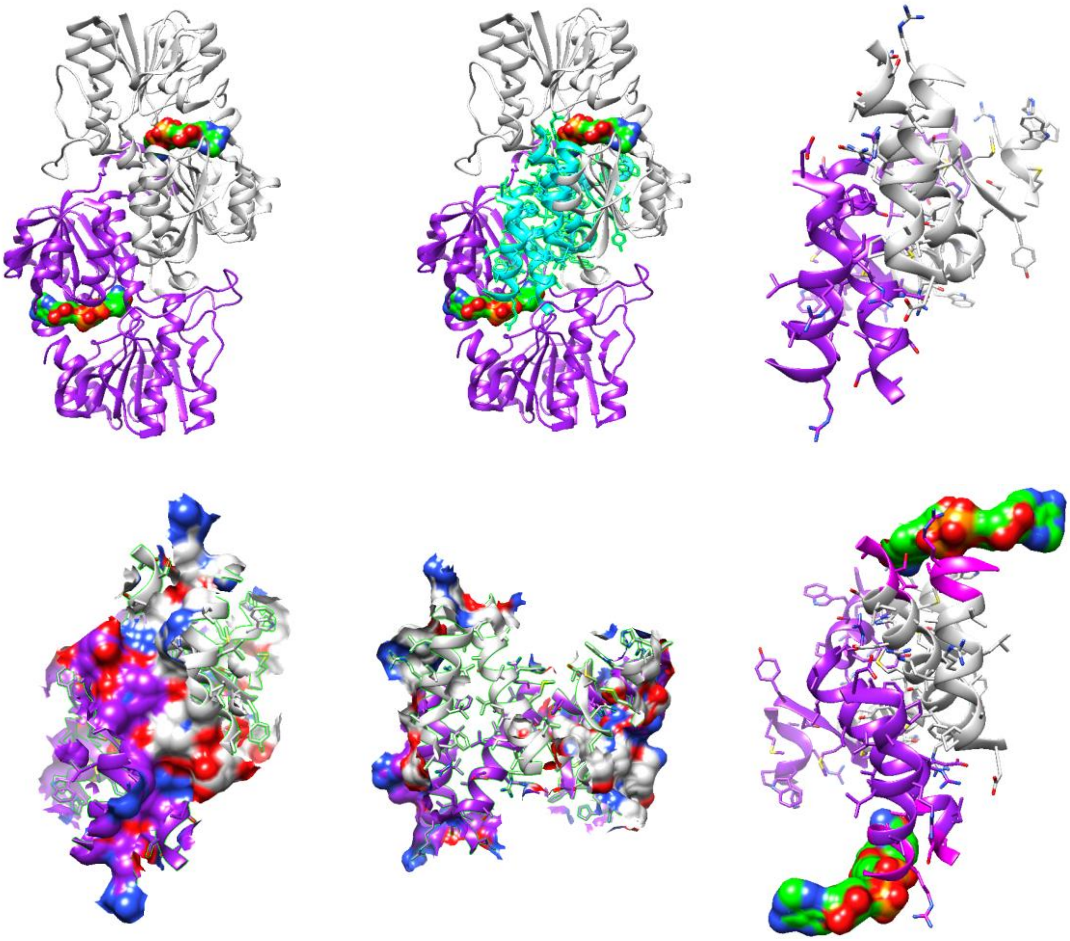
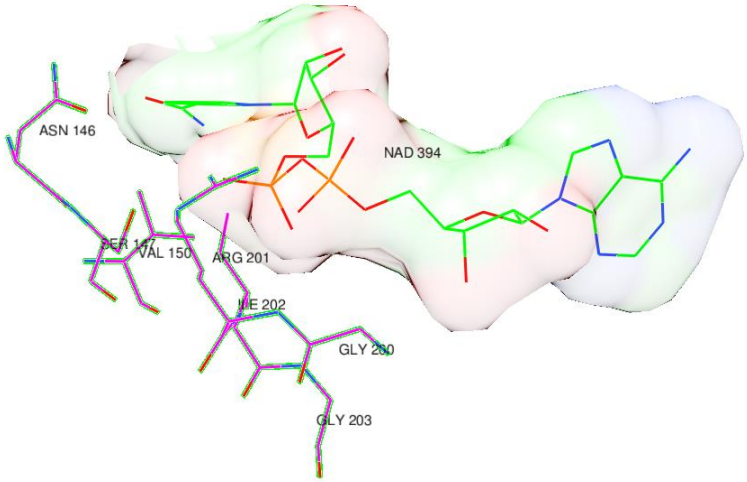


Figure 26. Residues that interact between 2 subunits in a PsFDH dimer

(a)



(b)

ASN 146
SER 147
VAL 150
GLY 200
ARG 201
ILE 202
GLY 203

Figure 27. Relevant residues for interaction among two subunits of a dimer, (a) distribution around the cofactor, (b) list of cofactors that belong to the interregional area between the 2 subunits and form part of the cofactor-binding site.

3.2.3. Target residues for mutations

There are 24 residues that can be targeted for mutations, indicated in Figure 28 and Figure 29. To determine these residues, first all the residues within a distance of 5Å from the cofactor were selected. Among them, the following residues were excluded:

- relevant residues for the reaction mechanism
- residues that influence the stability of the protein quaternary structure
- residues that belong to the catalytic active site (even not reported as relevant for the reaction mechanism)

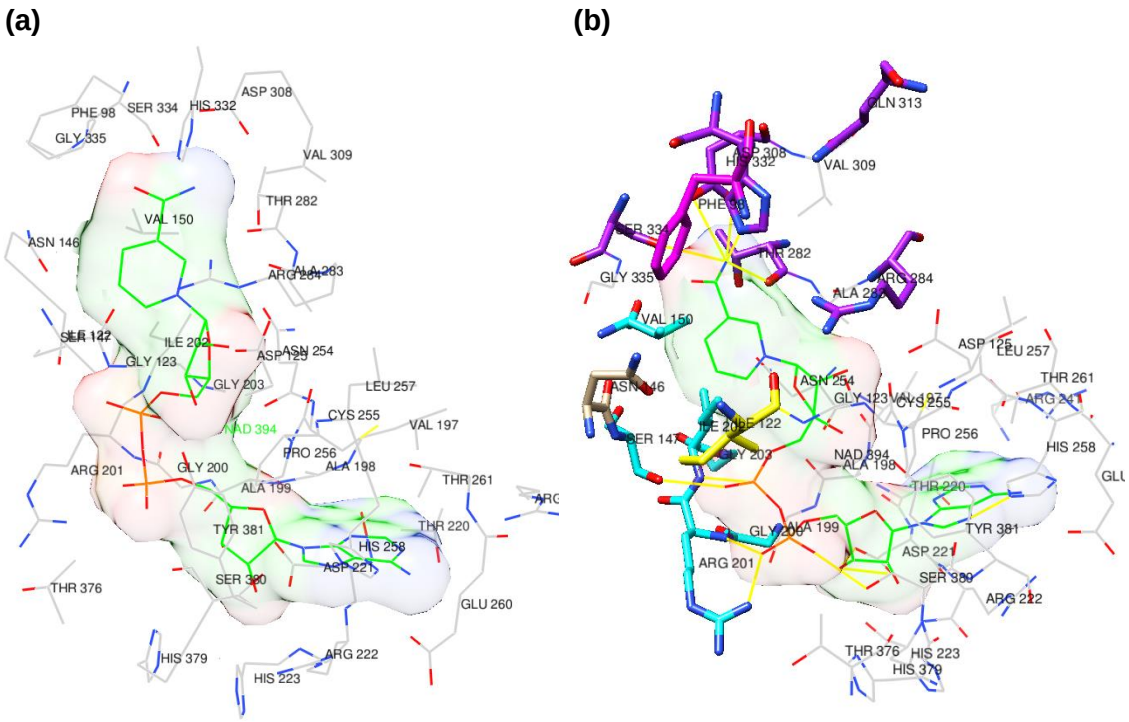
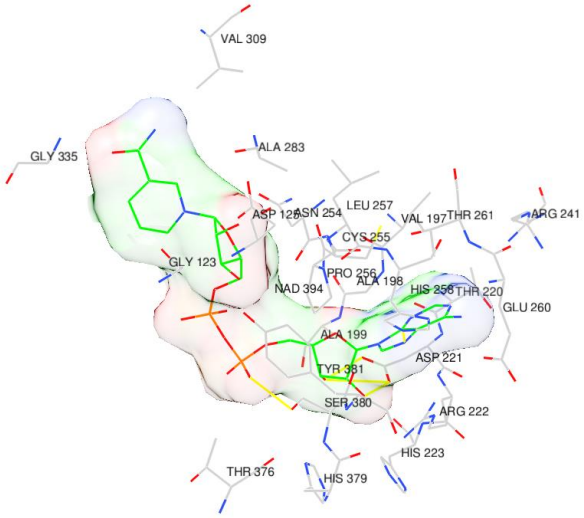
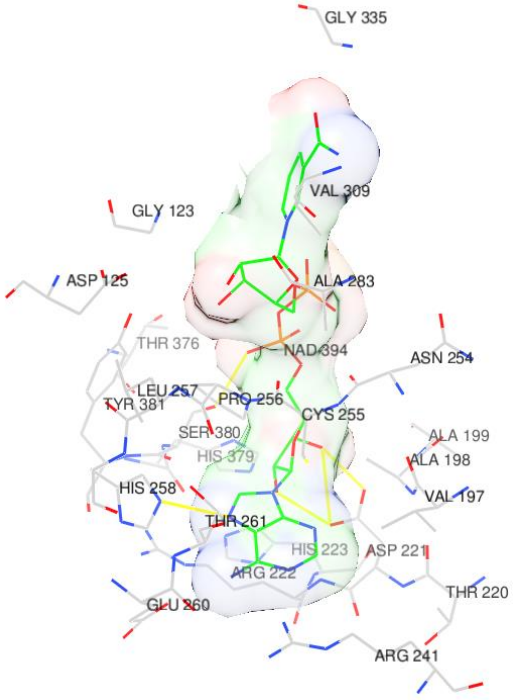


Figure 28. Residues in the cofactor-binding domain (a) all within 5Å, (b) discarded residues for mutations in magenta because they belong to the catalytic pocket, in cyan because they are important for the stability of the quaternary structure, in purple because they are relevant for the reaction mechanism, in tan because of all the previous reasons and in yellow for being part of the catalytic pocket and also relevant for the reaction mechanism.

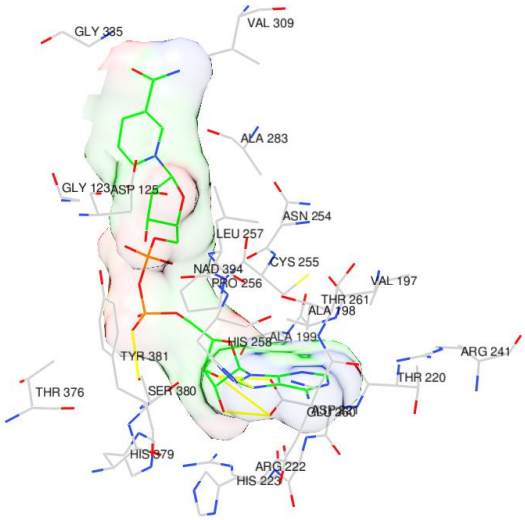
(a)



(b)



(c)



(d)

GLY 123
ASP 125
VAL 197
ALA 198
ALA 199
THR 220
ASP 221
ARG 222
HIS 223
ARG 241
ASN 254
CYS 255
PRO 256
LEU 257
HIS 258
GLU 260
THR 261
ALA 283
VAL 309
GLY 335
THR 376
HIS 379
SER 380
TYR 381

Figure 29. Electable residues for mutations. (a), (b) and (c) are different perspectives of the residues around the cofactor, (d) list of the residues.

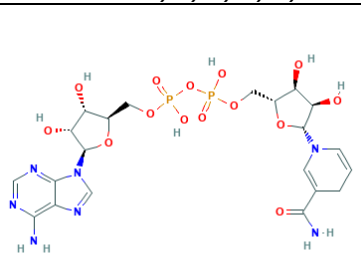
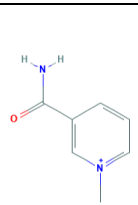
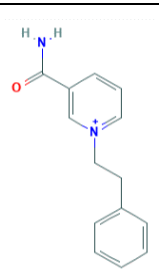
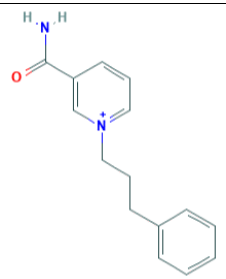
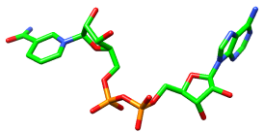
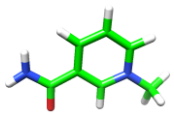
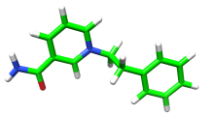
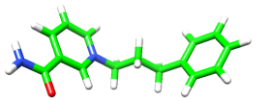
3.3. Selection of the biomimetic cofactor

3.3.1. Structures of several biomimetic cofactors according to redox potential

The objective of this project is to determine valuable mutations in the enzyme PsFDH to allow the use of a biomimetic cofactor instead of the natural cofactor NADH. Based on the redox potential, 3 biomimetic cofactors were studied (MNAH, P2NAH and P3NAH). Using docking, one of the cofactors was selected and mutations were done in the enzyme to promote the use of this cofactor.

Details of all the cofactors are indicated in the next tables. The 3D structure of NADH was isolated from the PDB structure 2NAD. The 3D structures of the other cofactors were generated using Chimera. First the SMILE code was generated with PubChem

Table 5. Cofactors that will be used in docking.

Abbreviated name	**NADH**	**MNAH**	**P2NAH**	**P3NAH**
IUPAC Name	[[(2R,3S,4R,5R)-5-(6-aminopurin-9-yl)-3,4-dihydroxyoxolan-2-yl]methoxy-hydroxyphosphoryl] [(2R,3S,4R,5R)-5-(3-carbamoyl-4H-pyridin-1-yl)-3,4-dihydroxyoxolan-2-yl]methyl hydrogen phosphate	1-methylpyridin-1-ium-3-carboxamide	1-(2-phenylethyl)pyridin-1-ium-3-carboxamide	1-(3-phenylpropyl)pyridin-1-ium-3-carboxamide
SMILES code	C1C=CN(C=C1C(=O)N)C2C(C(C(O2)COP(=O)(O)OP(=O)(O)OCC3C(C(C(O3)N4C=NC5=C(N=CN=C54)N)O)O)O)O	C1=CC=[N](C=C1C(=O)N)C	C1=CC=[N](C=C1C(=O)N)CCC2=CC=CC=C2	C1=CC=[N](C=C1C(=O)N)CCCC2=CC=CC=C2
2D Structure				
3D Structure				
PubChem CID	439153	457	410172	10947189

3.3.2. Docking for the selection of the biomimetic cofactor

Among the 3 tested biomimetic cofactors (MNAH, P2NAH and P3NAH), P3NAH appears in a more favourable position for reaction. Its nicotinamide group overlaps the nicotinamide of the natural cofactor. P3NAH is a bigger molecule than the other biomimetic cofactors so it can be stabilized inside the active site and binding to other pockets of the protein is more limited. P3NAH was selected for further studies.

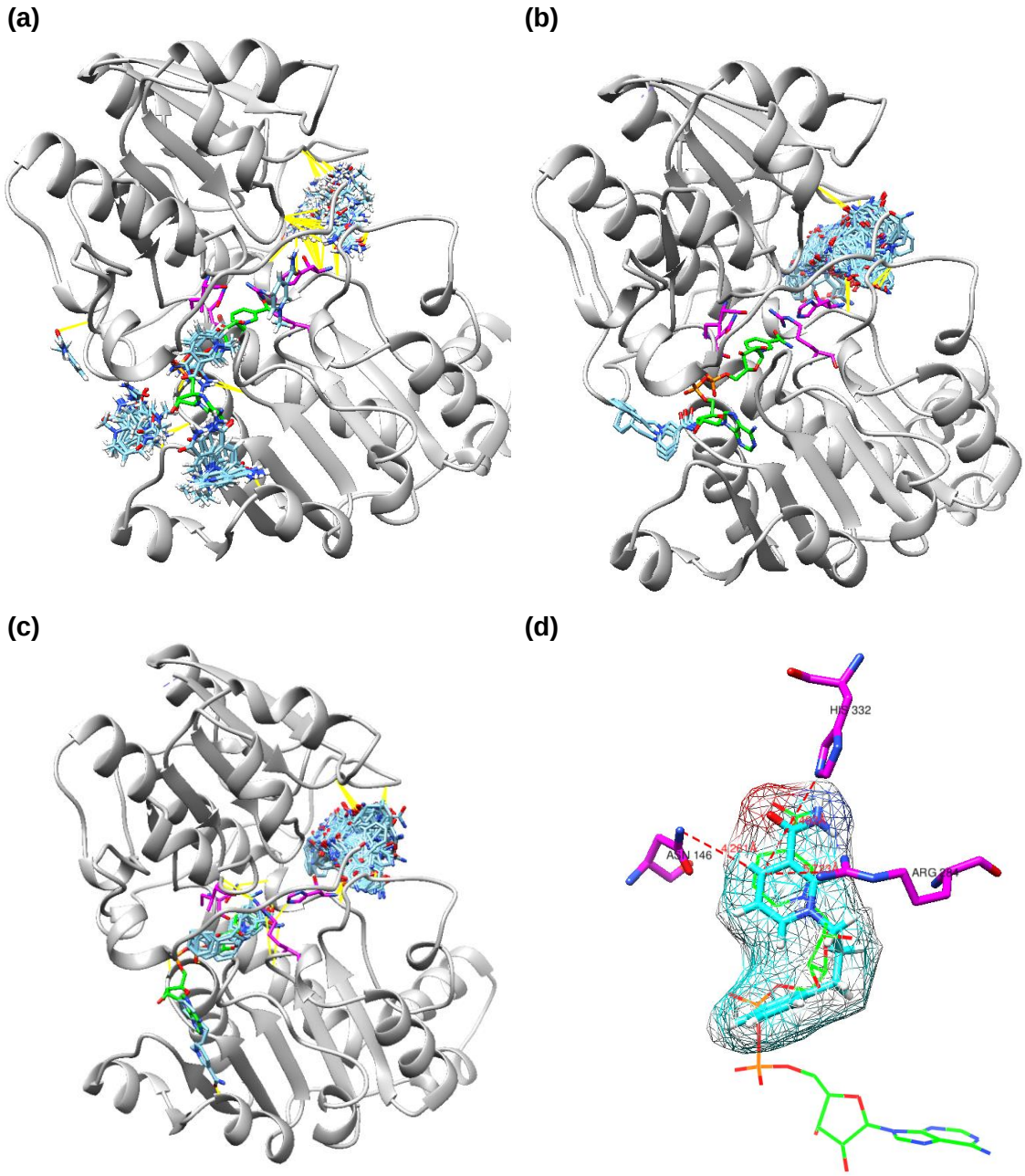


Figure 30. Clusters predicted by SwissDock for the cofactors (a) MNAH, (b) P2NAH and (c) P3NAH. (d) Superposition of the natural cofactor NADH in green and P3NAH in cyan.

3.4. Enzymes with mutations

3.4.1. Evaluation of the predicted structures for the mutants

The 3D structure of each mutant was created using homology modelling. The QMEAN score of the predicted structures for the mutants is indicated in Table 6. According to the QMEAN score, all the 3D structures generated are acceptable. This score provides an estimate of the "degree of nativeness" of the structural features observed in the model on a global scale. It indicates whether the QMEAN score of the model is comparable to what one would expect from experimental structures of similar size. QMEAN scores around zero indicate good agreement between the model structure and experimental structures of similar size. Scores of -4.0 or below are an indication of models with low quality.

Table 6. QMEAN scores of the predicted structures for the mutants.

Mutant	QMEAN
2NADa_A283F	0.27
2NADa_A283Y	0.29
2NADa_G123F	0.09
2NADa_G123Y	0.11
2NADa_T376G	0.32
2NADa_S380G	0.31
2NADa_Y381G	0.30
2NADa_R222G	0.33

3.4.2. Evaluation of binding between P3NAH and the mutants

Among all the mutants generated (refer to Table 7), the mutant 2NADa_S380G locates the cofactor closer to the 3 reference residues when compared with the wild type enzyme. The average distance of the nicotinamide group of P3NAH in 2NADa_S380G to the reference residues is 0.245 Å and for the WT enzyme is 1.2 Å. Figure 31 shows the location of the cofactor in the WT and 2NADa_S380G enzyme. The mutant 2NADa_S380G has the same number of H bonds between the cofactor and the residues in the enzyme pocket as the wild type enzyme with P3NAH (2 H bonds). However, the binding score is worse than the wild type enzyme, -7.2 in comparison with -7.6.

When aromatic residues were added to promote π–π stacking interactions (change of Ala 283 and Gly 123 for Phe and Tyr), P3NAH locates further from the reference residues. Both Phe and Tyr blocked the entrance of the cofactor.

In the case of changing big residues in the groove of the cofactor for Gly (Thr 376, Ser 380, Tyr 381 or Arg 222), there was only a relevant improvement for the mutant that has Gly instead of Ser 380.

Table 7. Evaluation of the mutants based on P3NAH docking

Enzyme	Score	H bonds	Difference with the atoms in NADH (Å)
WT	-7.6	2	1.200
2NADa_A283F	-6.7	1	2.424
2NADa_A283Y	-7.4	2	5.635
2NADa_G123F	-7.2	1	5.594
2NADa_G123Y	-7.3	1	5.573
2NADa_R222G	-7.5	2	1.348
2NADa_S380G	-7.2	2	0.245
2NADa_T376G	-7.2	2	1.409
2NADa_Y381G	-7.1	2	2.473

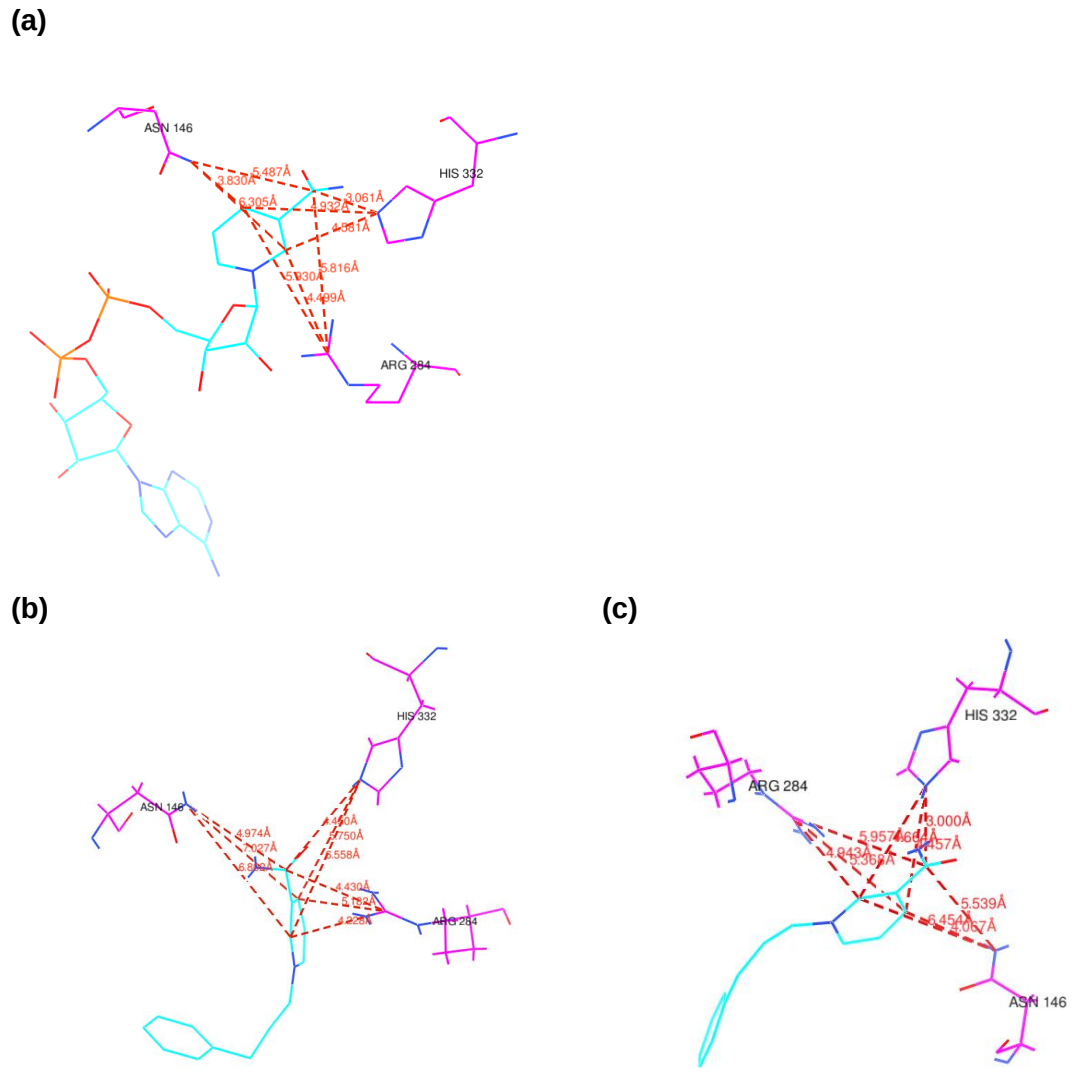


Figure 31. (a) Location of NADH in the WT enzyme, (b) P3NAH in the WT enzyme and (c) P3NAH in the mutant 2NADa_S380G.

3.5. Mutant viability evaluation

When modifying a protein, it is important to determine if the mutations have influence on its stability. In addition, in the case of enzymes, the access tunnel to the active site should allow the entrance and exit of the cofactor and/or substrate.

3.5.1. Protein stability evaluation

The mutant 2NADa_S380G has a higher stability than the wild type enzyme. The change in the residue serine to glycine provides a better stability to the protein.

(a)

Comprehensive Prediction Results

Mutation Site		Structural Features	
Protein:	2NAD	SS element:	Helix
Chain:	A	Solvent accessibility:	28.16%
Wild type AA:	SER	Torsion angles:	-103.4°, -11.3°
Residue ID:	380		

(b)

Amino Acid Mutations

Amino acid	Overall Stability	Torsion*	Predicted ΔΔG (kcal/mol
GLY	Stabilising	Unfavourable	1.8
ALA	Stabilising	Unfavourable	3.1
VAL	Stabilising	Unfavourable	2.9
LEU	Stabilising	Favourable	1.7
ILE	Stabilising	Unfavourable	3.3
MET	Stabilising	Unfavourable	2.5
PRO	Stabilising	Unfavourable	0.0
TRP	Stabilising	Favourable	2.5
THR	Stabilising	Favourable	0.3
PHE	Stabilising	Favourable	2.7
GLN	Stabilising	Favourable	1.2
LYS	Stabilising	Favourable	0.8
TYR	Stabilising	Favourable	2.0
ASN	Stabilising	Favourable	1.3
CYS	Stabilising	Favourable	3.2
GLU	Stabilising	Unfavourable	2.0
ASP	Stabilising	Favourable	2.1
ARG	Stabilising	Unfavourable	1.2
HIS	Stabilising	Favourable	2.1

Figure 32. Results from CUPSAT for the evaluation of the stability of the mutant 2NADa_S380G.

3.5.2. Analysis of ligand transport

The modification of the residues Ser 380 did not have a huge influence in the size of the tunnel, in fact, Caver Web indicates that the bottleneck for the mutant is 1.9 Å and for the wild type enzyme 2 Å. Serine 380 was recognize as one of the residues in the bottleneck, but changing it to glycine (a smaller residue) did not contribute to create a wider tunnel. The energy required for the cofactor to go in and out of the active site is similar in both proteins.

The modification of the residue did not contribute to an easier transport of the cofactor.

4. Conclusions and Recommendations

4.1. Conclusions

Among the 8 modification of the wild type enzyme for promoting π–π stacking interactions and create a wider groove, one viable mutant was found (2NADa_S380G). In this mutant a serine residue in the access tunnel of the cofactor was changed to glycine. The mutant can allocate the biomimetic cofactor in a more suitable position for reaction with the substrate of the enzyme than the wild type enzyme. P3NADH is a shorter molecule than NADH and does not bind as much as NADH, so the lack of the serine residue avoids steric effects. The mutations did not damage the stability of the protein and did not influence the transport of ligand toward the active site.

4.2. Limitations of the Study and Recommendations

This study was exclusively carried out with tools available online and with the software Chimera, which is available for free. With more computational power, molecular dynamics studies could be done to fully understand the influence of the mutations in the interaction between the enzyme and the cofactor.

Disadvantages of the work pipeline generated in this project are:

- The different steps are not automatically connected. For instance, after the 3D structure of one mutant is generated by homology modelling with SWISS-MODEL, the structure must be manually loaded for molecular docking in AutoDock Vina (or SwissDock).
- Dependence of tools based on web servers is not reliable because sometimes webs servers are too busy and take long time to send the results. For this reason, in the last part of this project molecular docking was carried out with AutoDock Vina in Chimera instead of with SwissDock.

With more computational power and knowledge, a work pipeline could be generated where the steps connect automatically, and more mutations are created and evaluated.

5. Glossary

Biomimetic nicotinamide-based cofactor: Synthetic molecule that contains a nicotinamide group and can be used in an enzymatic reaction as the cofactor

CO_2: Carbon dioxide

HM: Homology modelling. It consists on predicting the 3D structure of a protein based on similar sequences.

Molecular docking: Method for the prediction of the preferred orientation of a molecule when interaction with another in the active site.

Mutant: enzyme with at least one residue different in comparison to the wild type sequence

NADH: Nicotinamide adenine dinucleotide

P3NAH: 1-(3-phenylpropyl)nicotinamide

PDB: Protein Data Bank

Protein engineering: Conception and production of unnatural polypeptides, often through modification of amino acid sequences that are found in nature.

WT: Wild type enzyme (structure without any modifications)

6. References

Amao, Y. (2018). Formate dehydrogenase for CO2 utilization and its application. *Journal of Co2 Utilization*, *26*, 623–641. https://doi.org/10.1016/j.jcou.2018.06.022

Andreadeli, A., Platis, D., Tishkov, V., Popov, V., & Labrou, N. E. (2008). Structure-guided alteration of coenzyme specificity of formate dehydrogenase by saturation mutagenesis to enable efficient utilization of NADP+. *The FEBS Journal*, *275*(15), 3859–3869. https://doi.org/10.1111/j.1742-4658.2008.06533.x

Castillo, R., Oliva, M., Martí, S., & Moliner, V. (2008). A Theoretical Study of the Catalytic Mechanism of Formate Dehydrogenase. *The Journal of Physical Chemistry B*, *112*(32), 10012–10022. https://doi.org/10.1021/jp8025896

Filipovic, J., Vávra, O., Plhák, J., Bednar, D., Marques, S. M., Brezovsky, J., Matyska, L., & Damborsky, J. (2019). CaverDock: A Novel Method for the Fast Analysis of Ligand Transport. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1–1. https://doi.org/10.1109/TCBB.2019.2907492

Filippova, E. V., Polyakov, K. M., Tikhonova, T. V., Stekhanova, T. N., Boiko, K. M., & Popov, V. O. (2005). Structure of a new crystal modification of the bacterial NAD-dependent formate dehydrogenase with a resolution of 2.1 Å. *Crystallography Reports*, *50*(5), 796–800. https://doi.org/10.1134/1.2049398

Fogal, S., Beneventi, E., Cendron, L., & Bergantino, E. (2015). Structural basis for double cofactor specificity in a new formate dehydrogenase from the acidobacterium <Emphasis Type="Italic">Granulicella mallensis</Emphasis> MP5ACTX8. *Applied Microbiology and Biotechnology*, *99*(22), 9541–9554. https://doi.org/10.1007/s00253-015-6695-x

Grimaldi, S., Schoepp-Cothenet, B., Ceccaldi, P., Guigliarelli, B., & Magalon, A. (2013). The prokaryotic Mo/W-bisPGD enzymes family: A catalytic workhorse in bioenergetic. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1827*(8), 1048–1085. https://doi.org/10.1016/j.bbabio.2013.01.011

Grosdidier, A., Zoete, V., & Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, *39*(Web Server issue), W270–W277. https://doi.org/10.1093/nar/gkr366

Guarneri, A., van Berkel, W. J., & Paul, C. E. (2019). Alternative coenzymes for biocatalysis. *Current Opinion in Biotechnology*, *60*, 63–71. https://doi.org/10.1016/j.copbio.2019.01.001

Hille, R., Hall, J., & Basu, P. (2014). The mononuclear molybdenum enzymes. *Chemical Reviews*, *114*(7), 3963–4038. https://doi.org/10.1021/cr400443z

Holm, L. (2019). Benchmarking fold detection by DaliLite v.5. *Bioinformatics*, *btz536*. https://doi.org/10.1093/bioinformatics/btz536

Knaus, T., Paul, C. E., Levy, C. W., de Vries, S., Mutti, F. G., Hollmann, F., & Scrutton, N. S. (2016). Better than Nature: Nicotinamide Biomimetics That Outperform

Natural Coenzymes. *Journal of the American Chemical Society*, *138*(3), 1033–1039. https://doi.org/10.1021/jacs.5b12252

Labbé, C. M., Rey, J., Lagorce, D., Vavruša, M., Becot, J., Sperandio, O., Villoutreix, B. O., Tufféry, P., & Miteva, M. A. (2015). MTiOpenScreen: A web server for structure-based virtual screening. *Nucleic Acids Research*, *43*(Web Server issue), W448–W454. https://doi.org/10.1093/nar/gkv306

Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009). Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, *10*(1), 168. https://doi.org/10.1186/1471-2105-10-168

Lutz, S. (2010). Beyond directed evolution—Semi-rational protein engineering and design. *Current Opinion in Biotechnology*, *21*(6), 734–743. https://doi.org/10.1016/j.copbio.2010.08.011

Maia, L. B., Moura, I., & Moura, J. J. G. (2017). Molybdenum and tungsten-containing formate dehydrogenases: Aiming to inspire a catalyst for carbon dioxide utilization. *Inorganica Chimica Acta*, *455*, 350–363. https://doi.org/10.1016/j.ica.2016.07.010

Mondal, B., Song, J., Neese, F., & Ye, S. (2015). Bio-inspired mechanistic insights into CO<inf>2</inf> reduction. *Current Opinion in Chemical Biology*, *25*, 103–109. https://doi.org/10.1016/j.cbpa.2014.12.022

Nielsen, C. F., Lange, L., & Meyer, A. S. (2019). Classification and enzyme kinetics of formate dehydrogenases for biomanufacturing via CO2 utilization. *Biotechnology Advances*. https://doi.org/10.1016/j.biotechadv.2019.06.007

Nowak, C., Pick, A., Csepei, L.-I., & Sieber, V. (2017). Characterization of Biomimetic Cofactors According to Stability, Redox Potentials, and Enzymatic Conversion by NADH Oxidase from Lactobacillus pentosus. *Chembiochem: A European Journal of Chemical Biology*, *18*(19), 1944–1949. https://doi.org/10.1002/cbic.201700258

Nowak, C., Pick, A., Lommes, P., & Sieber, V. (2017). Enzymatic Reduction of Nicotinamide Biomimetic Cofactors Using an Engineered Glucose Dehydrogenase: Providing a Regeneration System for Artificial Cofactors. *ACS Catalysis*, *7*(8), 5202–5208. https://doi.org/10.1021/acscatal.7b00721

Parthiban, V., Gromiha, M. M., & Schomburg, D. (2006). CUPSAT: Prediction of protein stability upon point mutations. *Nucleic Acids Research*, *34*(Web Server issue), W239-242. https://doi.org/10.1093/nar/gkl190

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*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Reda, T., Plugge, C. M., Abram, N. J., & Hirst, J. (2008). Reversible interconversion of carbon dioxide and formate by an electroactive enzyme. *Proceedings of the National Academy of Sciences*, *105*(31), 10654–10658. https://doi.org/10.1073/pnas.0801290105

Stourac, J., Vavra, O., Kokkonen, P., Filipovic, J., Pinto, G., Brezovsky, J., Damborsky, J., & Bednar, D. (2019). Caver Web 1.0: Identification of tunnels and channels in proteins and analysis of ligand transport. *Nucleic Acids Research*, *47*(W1), W414–W422. https://doi.org/10.1093/nar/gkz378

Tishkov, V. I., & Popov, V. O. (2006). Protein engineering of formate dehydrogenase. *Biomolecular Engineering*, *23*(2), 89–110. https://doi.org/10.1016/j.bioeng.2006.02.003

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*(2), 455–461. https://doi.org/10.1002/jcc.21334

Vinals, C., Depiereux, E., & Feytmans, E. (1993). Prediction of structurally conserved regions of D-specific hydroxy acid dehydrogenases by multiple alignment with formate dehydrogenase. *Biochemical and Biophysical Research Communications*, *192*(1), 182–188. https://doi.org/10.1006/bbrc.1993.1398

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Wu, W., Zhu, D., & Hua, L. (2009). Site-saturation mutagenesis of formate dehydrogenase from Candida bodinii creating effective NADP+-dependent FDH enzymes. *Journal of Molecular Catalysis B: Enzymatic*, *61*(3–4), 157–161. https://doi.org/10.1016/j.molcatb.2009.06.005

Yang, J., & Zhang, Y. (2015). I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*, *43*(W1), W174–W181. PubMed. https://doi.org/10.1093/nar/gkv342

You, C., Huang, R., Wei, X., Zhu, Z., & Zhang, Y.-H. P. (2017). Protein engineering of oxidoreductases utilizing nicotinamide-based coenzymes, with applications in synthetic biology. *Synthetic and Systems Biotechnology*, *2*(3), 208–218. https://doi.org/10.1016/j.synbio.2017.09.002

Zachos, I., Nowak, C., & Sieber, V. (2019). Biomimetic cofactors and methods for their recycling. *Current Opinion in Chemical Biology*, *49*, 59–66. https://doi.org/10.1016/j.cbpa.2018.10.003

Appendix A: Structure of the mutants obtained by homology modelling

A.1. Wild type enzyme FASTA sequence

>2NAD A WT

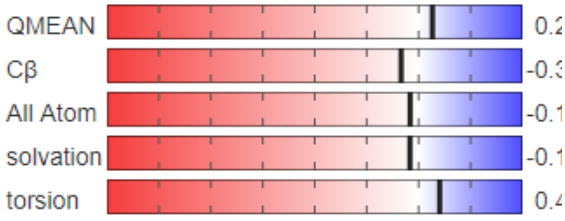
AKVLCVLYDDPVDGYPKTYARDDLPKIDHYPGGQTLPTPKAIDFTPGQLLGSVSGELG
LRKYLESNGHTLVVTSDKDGPDSVFERELVDADVVISQPFWPAYLTPERIAKAKNLKL
ALTAGIGSDHVDLQSAIDRNVTVAEVTYCNSISVAEHVVMMILSLVRNYLPSHEWARK
GGWNIADCVSHAYDLEAMHVGTVAAGRIGLAVLRRLAPFDVHLHYTDRHRLPESVEK
ELNLTWHATREDMYPVCDVVTLNCPLHPETEHMINDETLKLFKRGAYIVNTARGKLCD
RDAVARALESGRLAGYAGDVWFPQPAPKDHPWRTMPYNGMTPHISGTTLTAQARYA
AGTREILECFFEGRPIRDEYLIVQGGALAGTGAHSYSKGNATGGSE

A.2. Mutant 2NADa_A283F

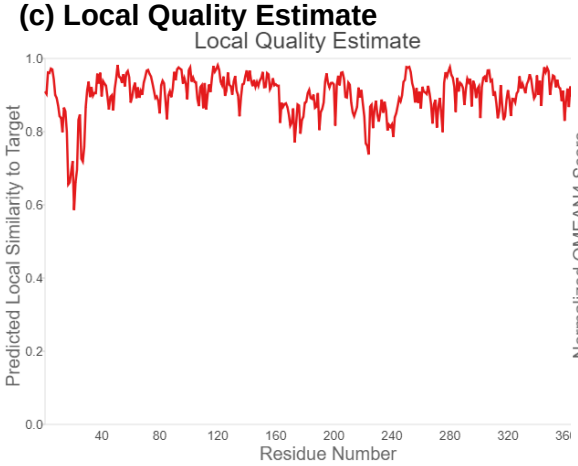
(a) Homology Modelling details

Mutant name	2NADa_A283F
Template	2NAD chain A
Change in residues	Ala 283 to Phe
QMEAN	0.27
Seq Identity (%)	99.74
Seq Similarity	0.62

(b) Global Quality Estimate

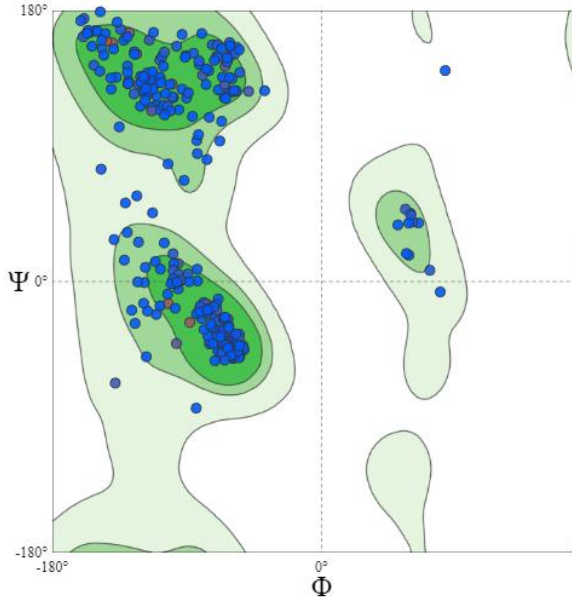


(c) Local Quality Estimate

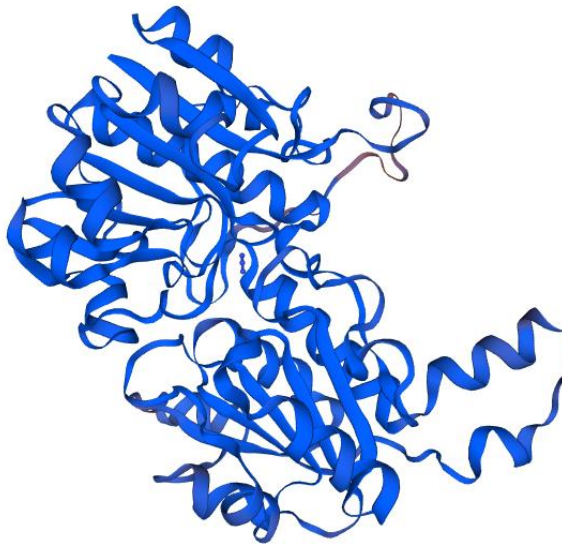


(d) Comparison

(e) Ramachandran Plot



(f) Structure



(g) FASTA sequence

>2NAD A Mutant Ala 283 to Phe

AKVLCVLYDDPVDGYPKTYARDDLPKIDHYPGGQTLPTPKAIDFTPGQLLGSVSGEL
GLRKYLESNGHTLVVTSDKDGPDSVFERELVDADVVISQPFWPAYLTPERIAKAKNL
KLALTAGIGSDHVDLQSAIDRNVTVAEVTYCNSISVAEHVVMMILSLVRNYLPSHEW
ARKGGWNIADCVSHAYDLEAMHVGTVAAGRIGLAVLRRLAPFDVHLHYTDRHRLP
ESVEKELNLTWHATREDMYPVCDVVTLNCPLHPETEHMINDETLKLFKRGAYIVNTF
RGKLCDRDAVARALESGRLAGYAGDVWFPQPAPKDHPWRTMPYNGMTPHISGTT
LTAQARYAAGTREILECFFEGRPIRDEYLIVQGGALAGTGAHSYSKGNATGGSE

(h) Residue Quality

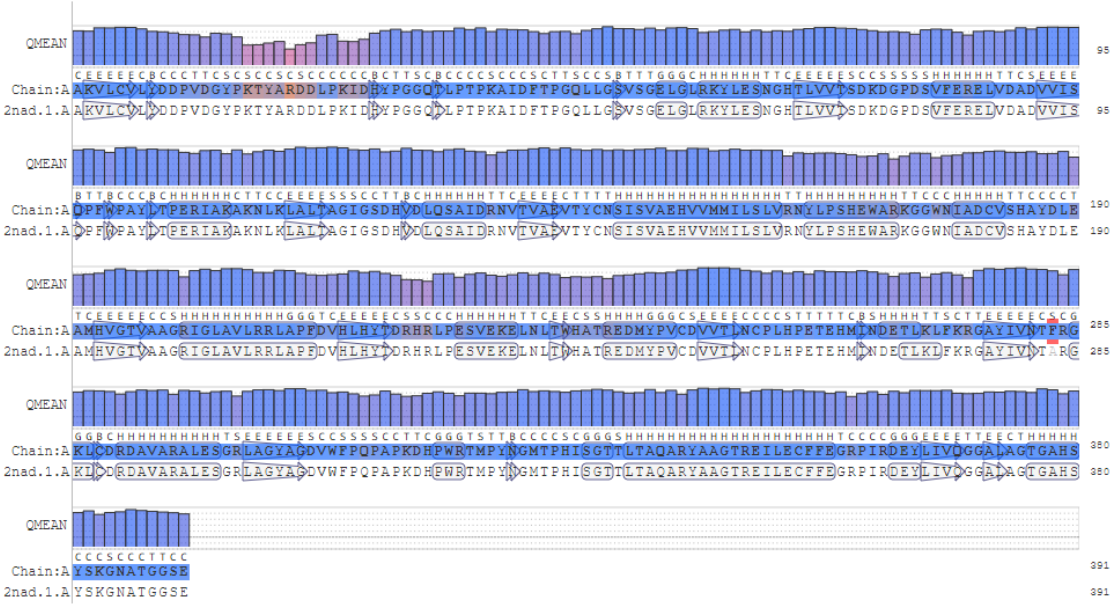


Figure 33. Homology modelling result for the mutant 2NADa_A283F.

(a) **(b)**

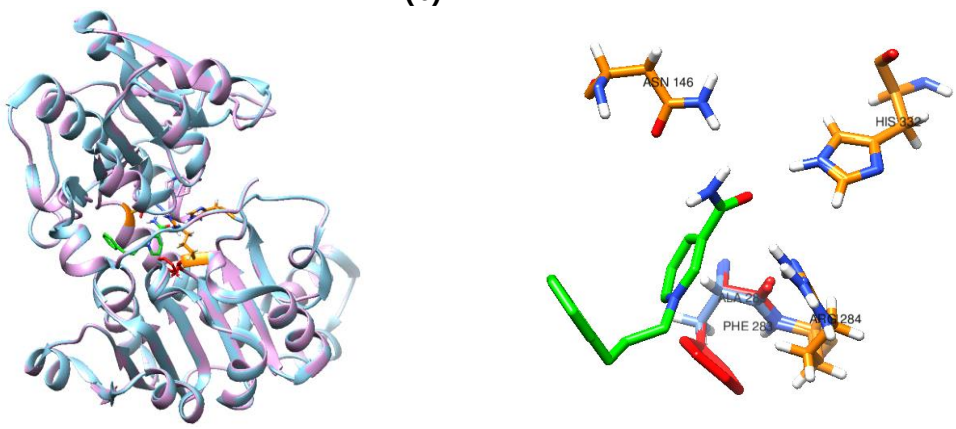


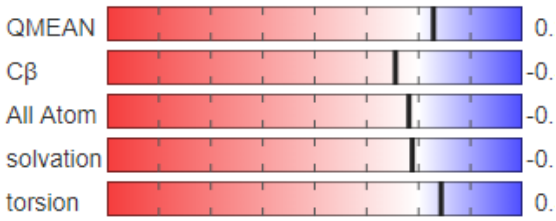
Figure 34. Mutant 2NADa_A283F (a) Comparison with the wild protein (in blue), (b) modified residue indicated in red and the biomimetic cofactor P3NAH in green with the best predicted position for the WT enzyme.

A.3. Mutant 2NADa_A283Y

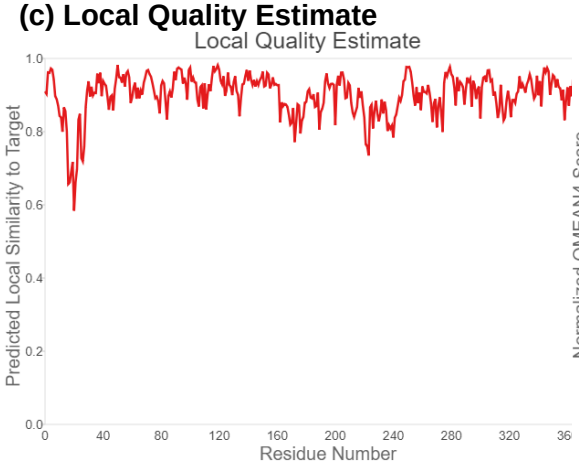
(a) Homology Modelling details

Mutant name	2NADa_A283Y
Template	2NAD chain A
Change in residues	Ala 283 to Tyr
QMEAN	0.29
Seq Identity (%)	99.74
Seq Similarity	0.62

(b) Global Quality Estimate

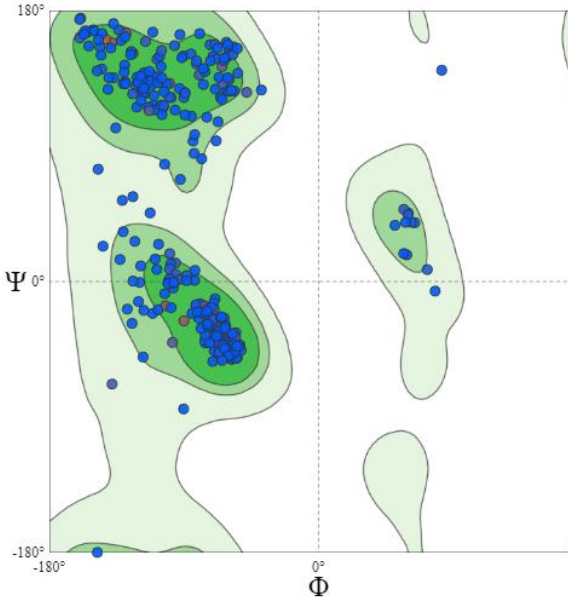


(c) Local Quality Estimate



(d) Comparison

(e) Ramachandran Plot



(f) Structure

(g) FASTA sequence

>2NAD A Mutant Ala 283 to Tyr
AKVLCVLYDDPVDGYPKTYARDDLPKIDHYPGGQTLPTPKAIDFTPGQLLGSVSGEL
GLRKYLESNGHTLVVTSDKDGPDSVFERELVDADVVISQPFWPAYLTPERIAKAKNL
KLALTAGIGSDHVDLQSAIDRNVTVAEVTYCNSISVAEHVVMMILSLVRNYLPSHEW
ARKGGWNIADCVSHAYDLEAMHVGTVAAGRIGLAVLRRLAPFDVHLHYTDRHRLP
ESVEKELNLTWHATREDMYPVCDVVTLNCPLHPETEHMINDETLKLFKRGAYIVNT
YRGKLCDRDAVARALESGRLAGYAGDVWFPQPAPKDHPWRTMPYNGMTPHISGT
TLTAQARYAAGTREILECFFEGRPIRDEYLIVQGGALAGTGAHSYSKGNATGGSE

(h) Residue Quality

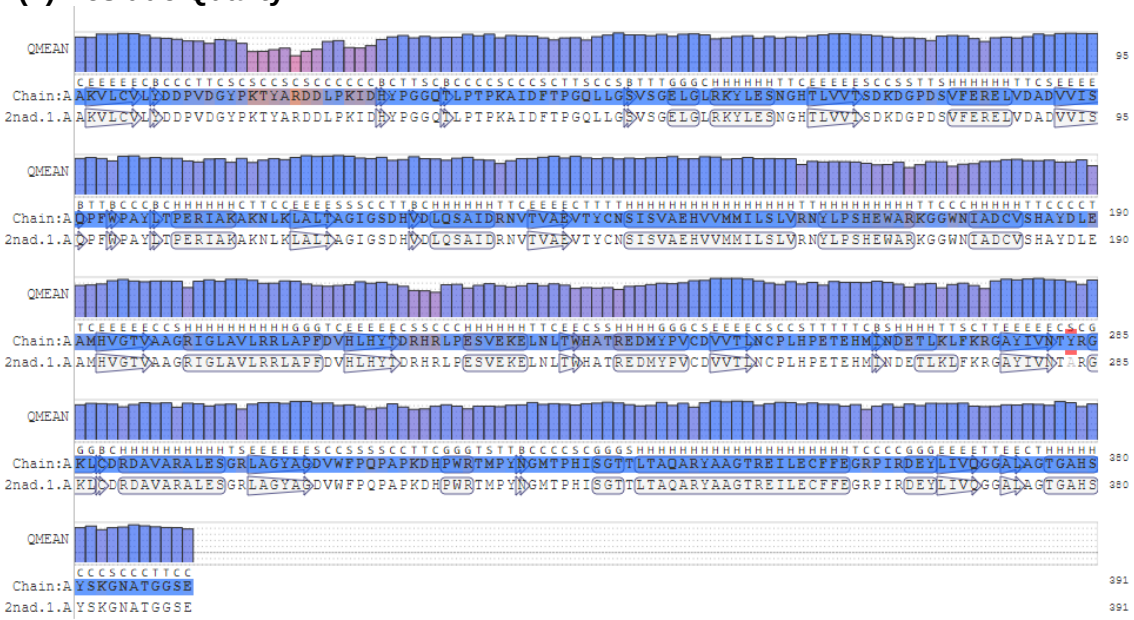


Figure 35. Homology modelling result for the mutant 2NADa_A283Y.

(a) **(b)**

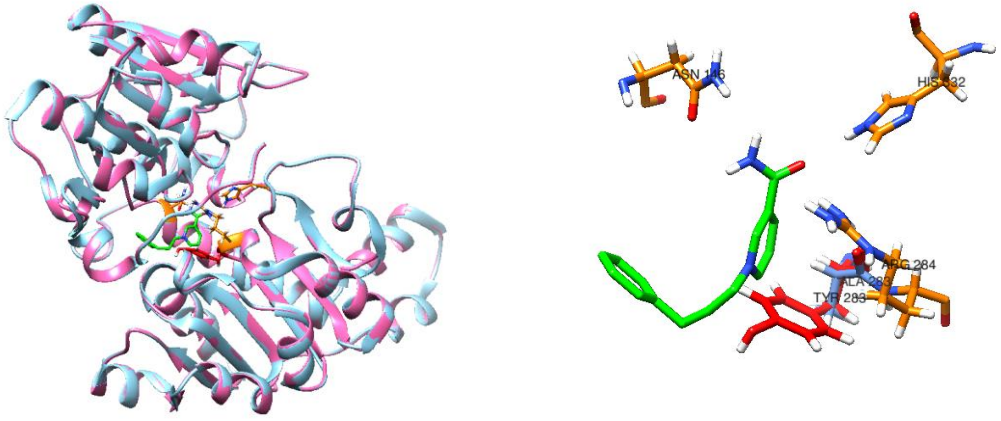


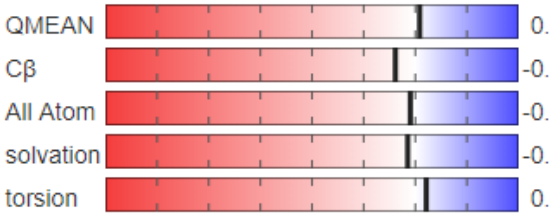
Figure 36. Mutant 2NADa_A283Y (a) Comparison with the wild protein (in blue), (b) modified residue indicated in red and the biomimetic cofactor P3NAH in green with the best predicted position for the WT enzyme.

A.4. Mutant 2NADa_G123F

(a) Homology Modelling details

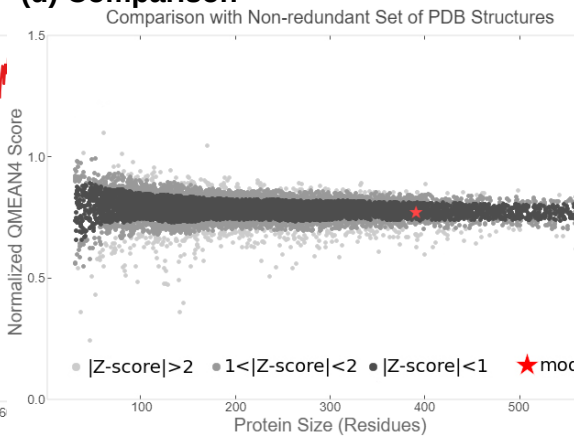
Mutant name	2NADa_G123F
Template	2NAD chain A
Change in residues	Gly 123 to Phe
QMEAN	0.09
Seq Identity (%)	99.74
Seq Similarity	0.61

(b) Global Quality Estimate

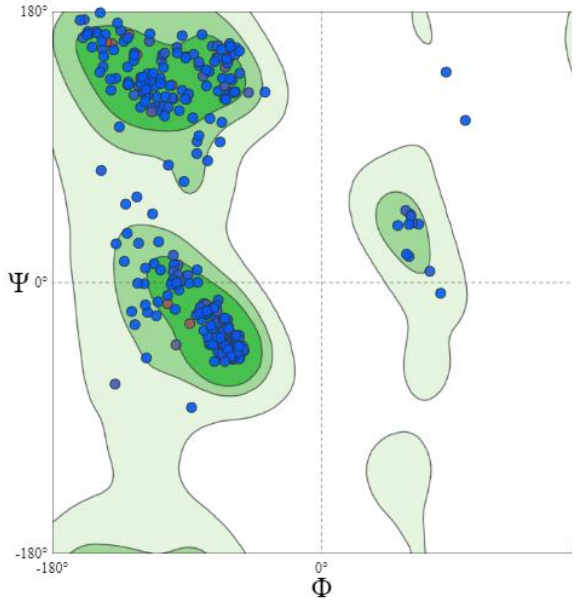


(c) Local Quality Estimate

(d) Comparison



(e) Ramachandran Plot



(f) Structure

(g) FASTA sequence

>2NAD A Mutant Gly 123 to Phe

AKVLCVLYDDPVDGYPKTYARDDLPKIDHYPGGQTLPTPKAIDFTPGQLLGSVSGEL
GLRKYLESNGHTLVVTSDKDGPDSVFERELVDADVVISQPFWPAYLTPERIAKAKNL
KLALTAGIFSDHVDLQSAIDRNVTVAEVTYCNSISVAEHVVMMILSLVRNYLPSHEW
ARKGGWNIADCVSHAYDLEAMHVGTVAAGRIGLAVLRRLAPFDVHLHYTDRHRLP
ESVEKELNLTWHATREDMYPVCDVVTLNCPLHPETEHMINDETLKLFKRGAYIVNT
ARGKLCDRDAVARALESGRLAGYAGDVWFPQPAPKDHPWRTMPYNGMTPHISGT
TLTAQARYAAGTREILECFFEGRPIRDEYLIVQGGALAGTGAHSYSKGNATGGSE

(h) Residue Quality

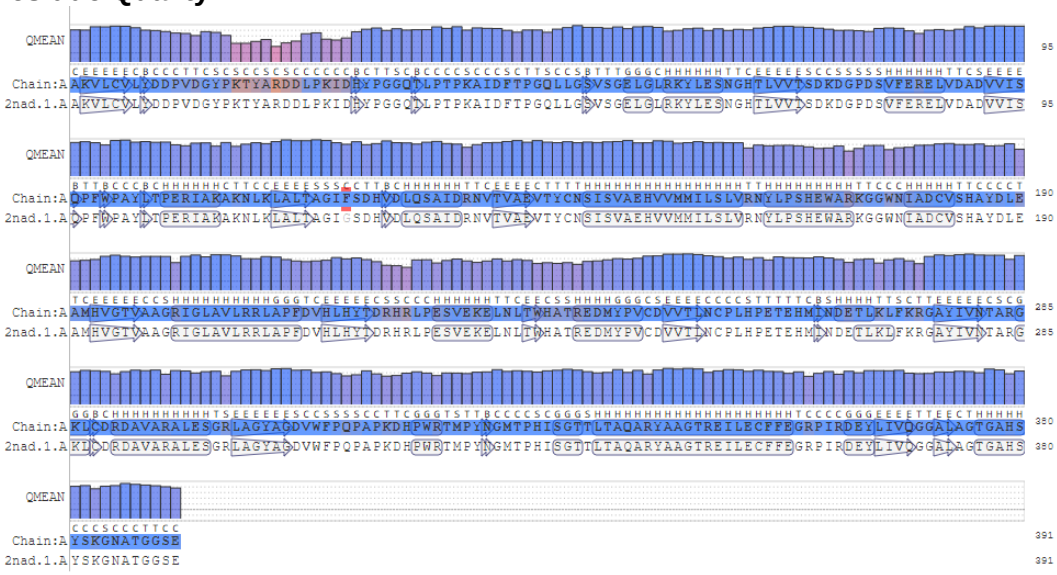


Figure 37. Homology modelling result for the mutant 2NADa_G123F.

(a) **(b)**

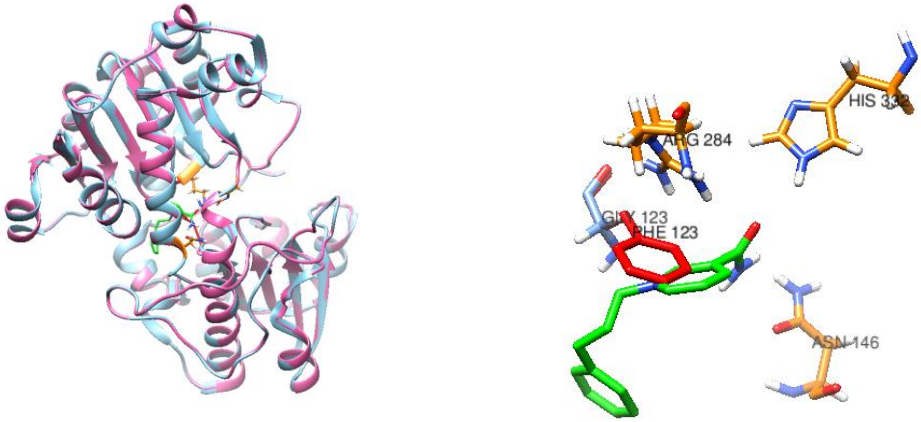


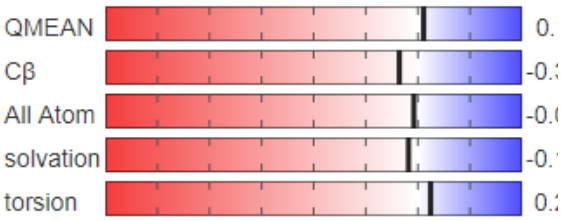
Figure 38. Mutant 2NADa_G123F (a) Comparison with the wild protein (in blue), (b) modified residue indicated in red and the biomimetic cofactor P3NAH in green with the best predicted position for the WT enzyme.

A.5. Mutant 2NADa_G123Y

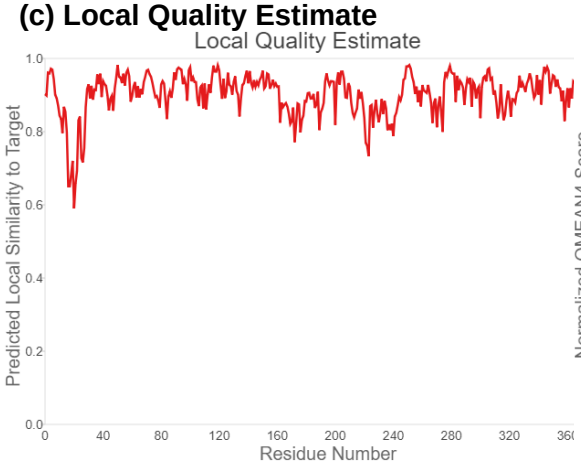
(a) Homology Modelling details

Mutant name	2NADa_G123Y
Template	2NAD chain A
Change in residues	G123 to Tyr
QMEAN	0.11
Seq Identity (%)	99.74
Seq Similarity	0.61

(b) Global Quality Estimate

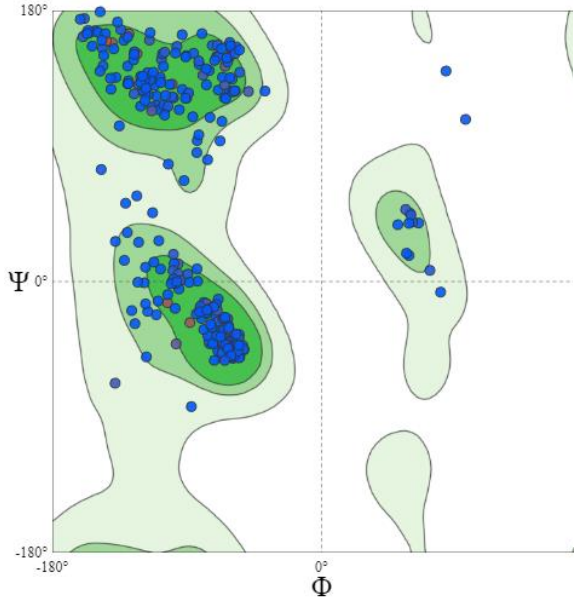


(c) Local Quality Estimate

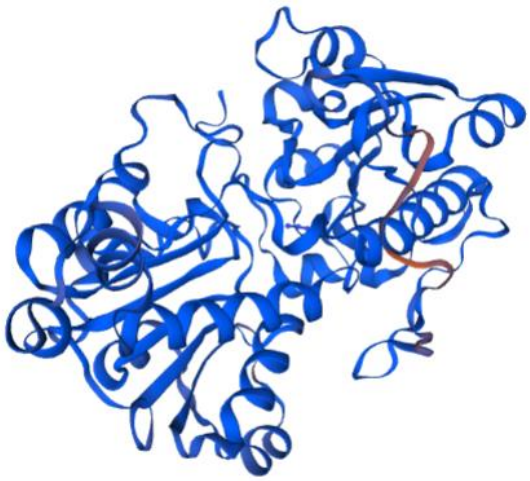


(d) Comparison

(e) Ramachandran Plot



(f) Structure



(g) FASTA sequence

>2NAD A Mutant Gly 123 to Tyr

AKVLCVLYDDPVDGYPKTYARDDLPKIDHYPGGQTLPTPKAIDFTPGQLLGSVSGEL
GLRKYLESNGHTLVVTSDKDGPDSVFERELVDADVVISQPFWPAYLTPERIAKAKNL
KLALTAGIYSDHVDLQSAIDRNVTVAEVTYCNSISVAEHVVMMILSLVRNYLPSHEW
ARKGGWNIADCVSHAYDLEAMHVGTVAAGRIGLAVLRRLAPFDVHLHYTDRHRLP
ESVEKELNLTWHATREDMYPVCDVVTLNCPLHPETEHMINDETLKLFKRGAYIVNT
ARGKLCDRDAVARALESGRLAGYAGDVWFPQPAPKDHPWRTMPYNGMTPHISGT
TLTAQARYAAGTREILECFFEGRPIRDEYLIVQGGALAGTGAHSYSKGNATGGSE

(h) Residue Quality

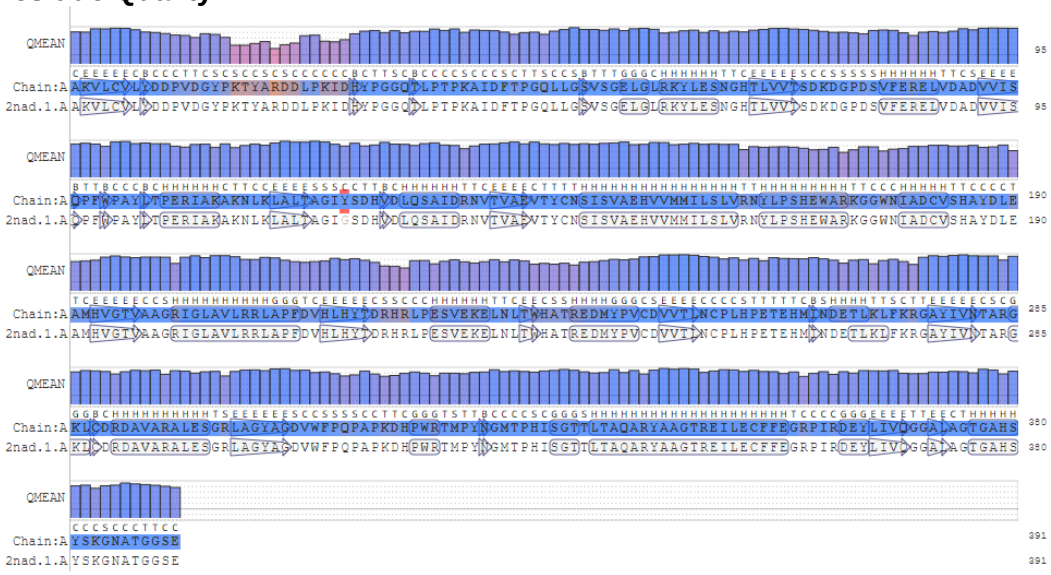


Figure 39. Homology modelling result for the mutant 2NADa_G123Y.

(a) **(b)**

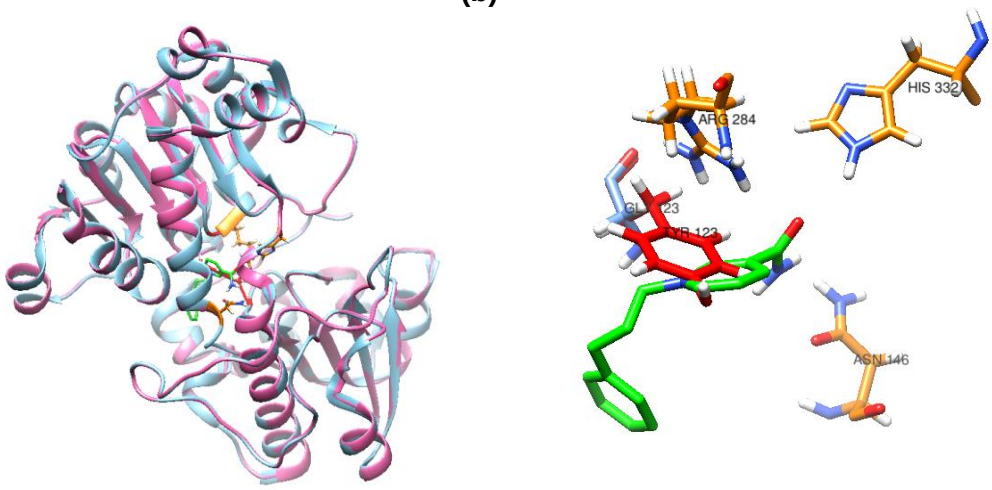


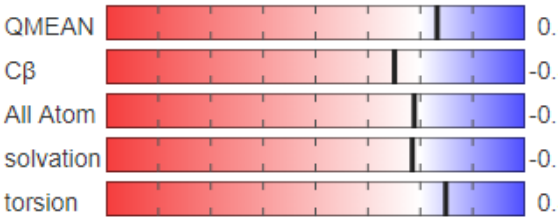
Figure 40. Mutant 2NADa_G123F (a) Comparison with the wild protein (in blue), (b) modified residue indicated in red and the biomimetic cofactor P3NAH in green with the best predicted position for the WT enzyme.

A.6. Mutant 2NADa_T376G

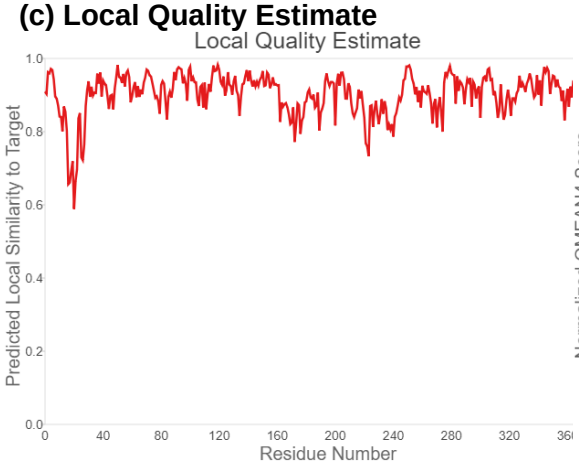
(a) Homology Modelling details

Mutant name	2NADa_T376G
Template	2NAD chain A
Change in residues	Thr 376 to Gly
QMEAN	0.32
Seq Identity (%)	99.74
Seq Similarity	0.61

(b) Global Quality Estimate

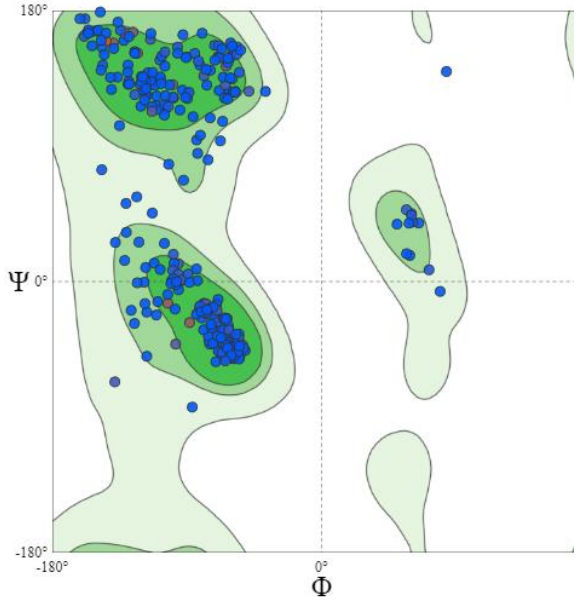


(c) Local Quality Estimate



(d) Comparison

(e) Ramachandran Plot



(f) Structure

(g) FASTA sequence

>2NAD A Mutant Thr 376 to Gly

AKVLCVLYDDPVDGYPKTYARDDLPKIDHYPGGQTLPTPKAIDFTPGQLLGSVSGEL
GLRKYLESNGHTLVVTSDKDGPDSVFERELVDADVVISQPFWPAYLTPERIAKAKNL
KLALTAGIGSDHVDLQSAIDRNVTVAEVTYCNSISVAEHVVMMILSLVRNYLPSHEW
ARKGGWNIADCVSHAYDLEAMHVGTVAAGRIGLAVLRRLAPFDVHLHYTDRHRLP
ESVEKELNLTWHATREDMYPVCDVVTLNCPLHPETEHMINDETLKLFKRGAYIVNT
ARGKLCDRDAVARALESGRLAGYAGDVWFPQPAPKDHPWRTMPYNGMTPHISGT
TLTAQARYAAGTREILECFFEGRPIRDEYLIVQGGALAGGGAHSYSKGNATGGSE

(h) Residue Quality

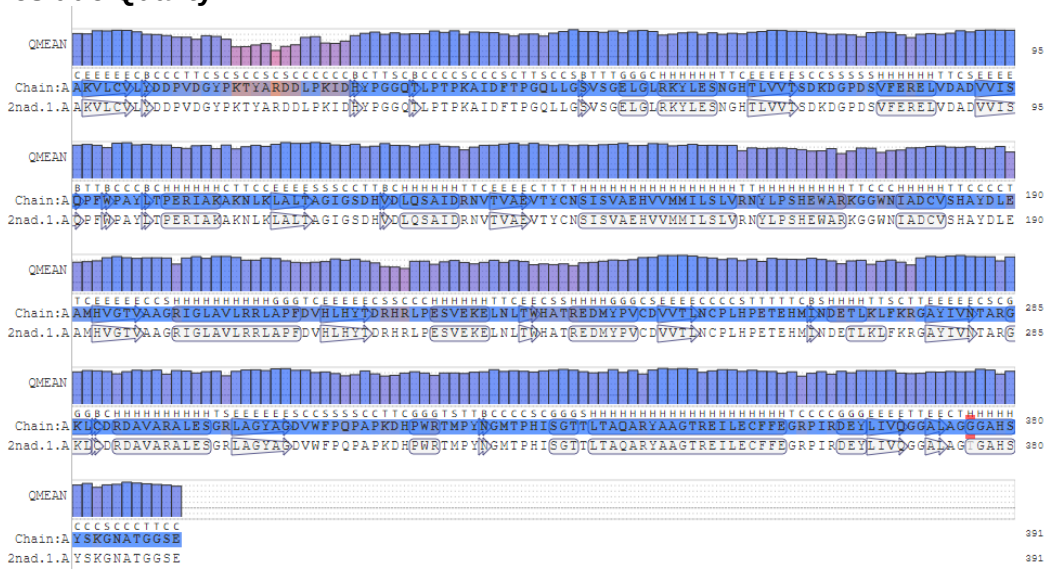


Figure 41. Homology modelling result for the mutant 2NADa_T376G.

(a) **(b)**

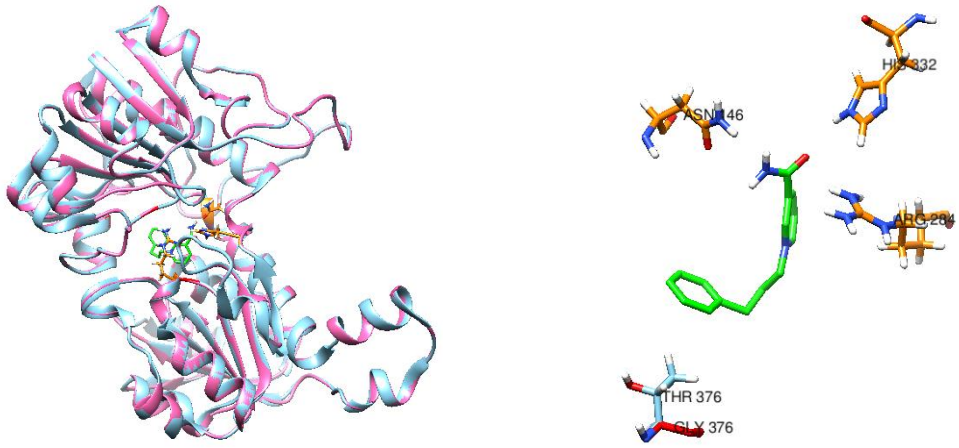


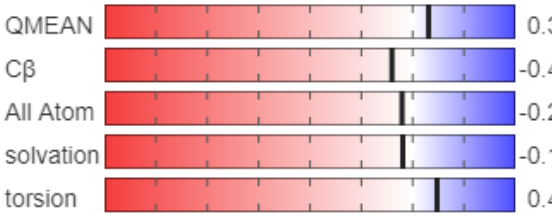
Figure 42. Mutant 2NADa_T376G (a) Comparison with the wild protein (in blue), (b) modified residue indicated in red and the biomimetic cofactor P3NAH in green with the best predicted position for the WT enzyme.

A.7. Mutant 2NADa_S380G

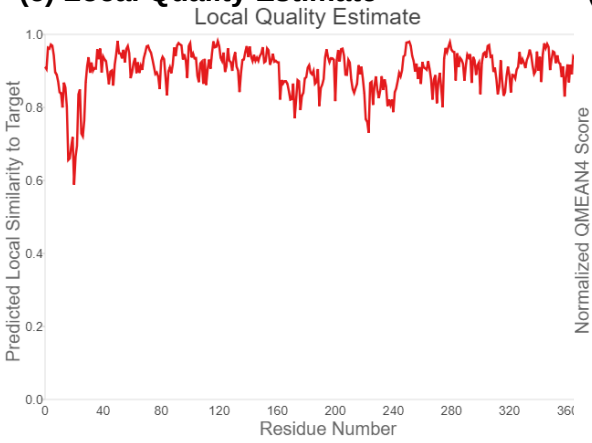
(a) Homology Modelling details

Mutant name	2NADa_S380G
Template	2NAD chain A
Change in residues	Ser 380 to Gly
QMEAN	0.31
Seq Identity (%)	99.74
Seq Similarity	0.62

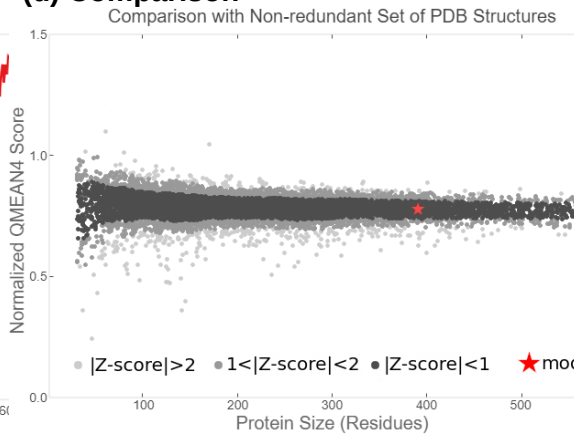
(b) Global Quality Estimate



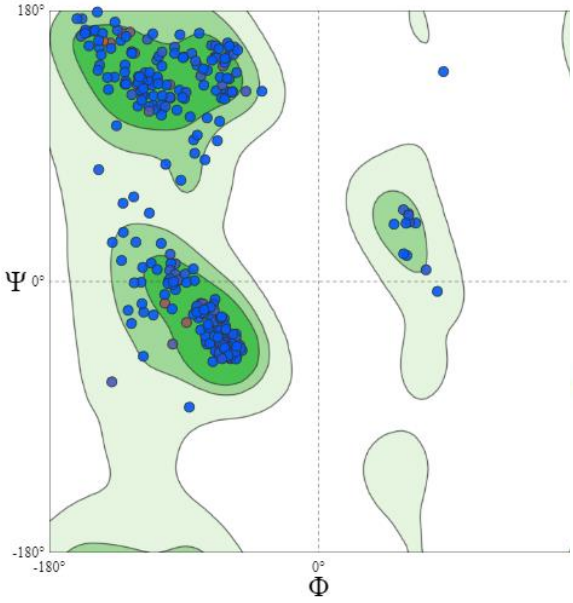
(c) Local Quality Estimate



(d) Comparison



(e) Ramachandran Plot



(f) Structure

(g) FASTA sequence

>2NAD A Mutant Ser 380 to Gly

AKVLCVLYDDPVDGYPKTYARDDLPKIDHYPGGQTLPTPKAIDFTPGQLLGSVSGEL
GLRKYLESNGHTLVVTSDKDGPDSVFERELVDADVVISQPFWPAYLTPERIAKAKNL
KLALTAGIGSDHVDLQSAIDRNVTVAEVTYCNSISVAEHVVMMILSLVRNYLPSHEW
ARKGGWNIADCVSHAYDLEAMHVGTVAAGRIGLAVLRRLAPFDVHLHYTDRHRLP
ESVEKELNLTWHATREDMYPVCDVVTLNCPLHPETEHMINDETLKLFKRGAYIVNT
ARGKLCDRDAVARALESGRLAGYAGDVWFPQPAPKDHPWRTMPYNGMTPHISGT
TLTAQARYAAGTREILECFFEGRPIRDEYLIVQGGALAGTGAHGYSKGNATGGSE

(h) Residue Quality

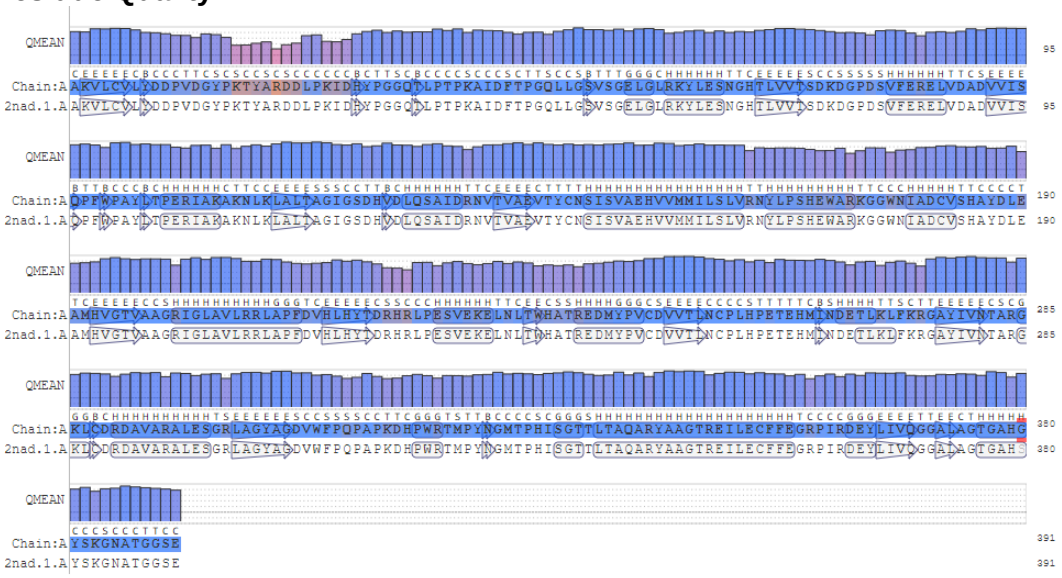


Figure 43. Homology modelling result for the mutant 2NADa_S380G.

(a) **(b)**

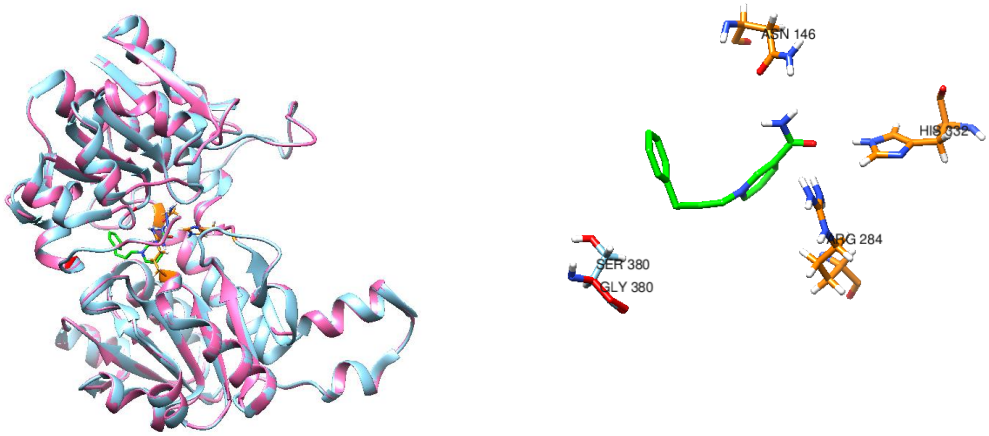


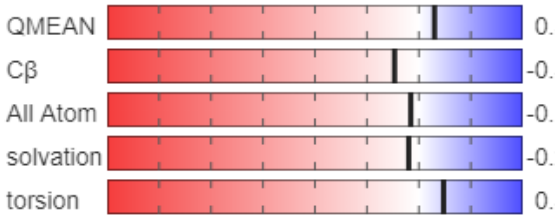
Figure 44. Mutant 2NADa_S380G (a) Comparison with the wild protein (in blue), (b) modified residue indicated in red and the biomimetic cofactor P3NAH in green with the best predicted position for the WT enzyme.

A.8. Mutant 2NADa_Y381G

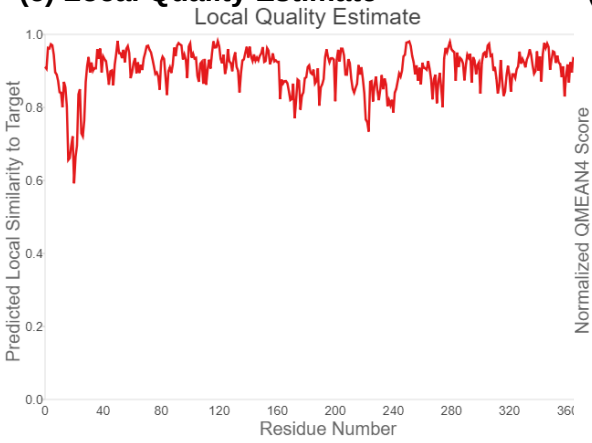
(a) Homology Modelling details

Mutant name	2NADa_Y381G
Template	2NAD chain A
Change in residues	Tyr 381 to Gly
QMEAN	0.30
Seq Identity (%)	99.74
Seq Similarity	0.61

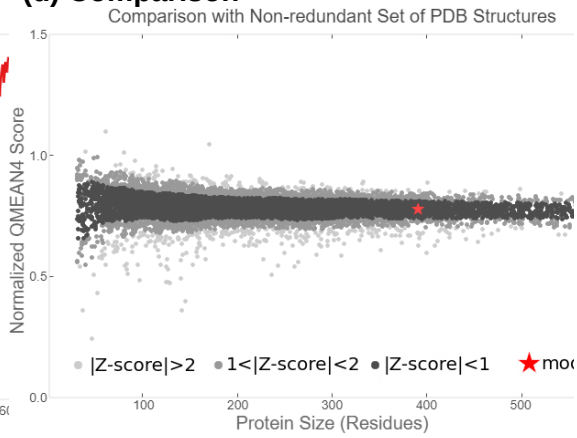
(b) Global Quality Estimate



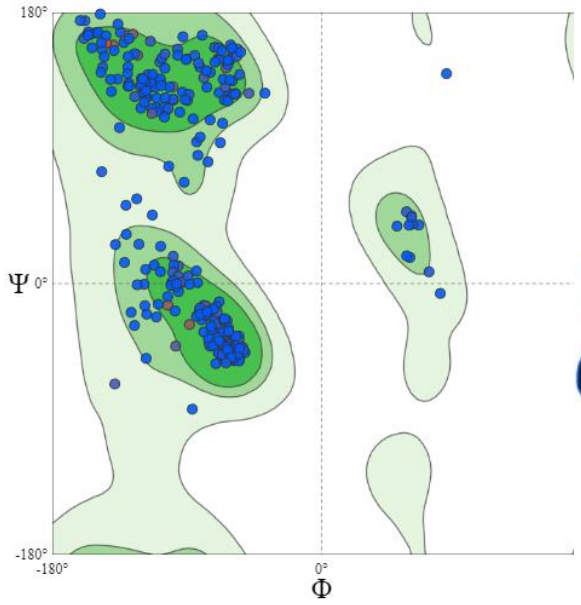
(c) Local Quality Estimate



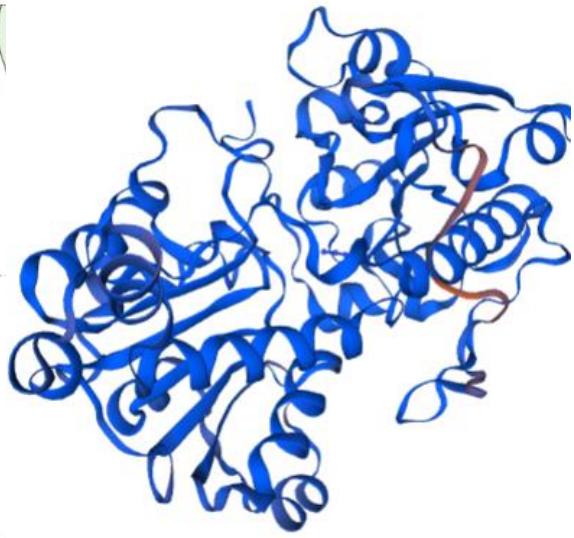
(d) Comparison



(e) Ramachandran Plot



(f) Structure



(g) FASTA sequence

>2NAD A Mutant Tyr 381 to Gly

AKVLCVLYDDPVDGYPKTYARDDLPKIDHYPGGQTLPTPKAIDFTPGQLLGSVSGEL
GLRKYLESNGHTLVVTSDKDGPDSVFERELVDADVVISQPFWPAYLTPERIAKAKNL
KLALTAGIGSDHVDLQSAIDRNVTVAEVTYCNSISVAEHVVMMILSLVRNYLPSHEW
ARKGGWNIADCVSHAYDLEAMHVGTVAAGRIGLAVLRRLAPFDVHLHYTDRHRLP
ESVEKELNLTWHATREDMYPVCDVVTLNCPLHPETEHMINDETLKLFKRGAYIVNT
ARGKLCDRDAVARALESGRLAGYAGDVWFPQPAPKDHPWRTMPYNGMTPHISGT
TLTAQARYAAGTREILECFFEGRPIRDEYLIVQGGALAGTGAHSGSKGNATGGSE

(h) Residue Quality

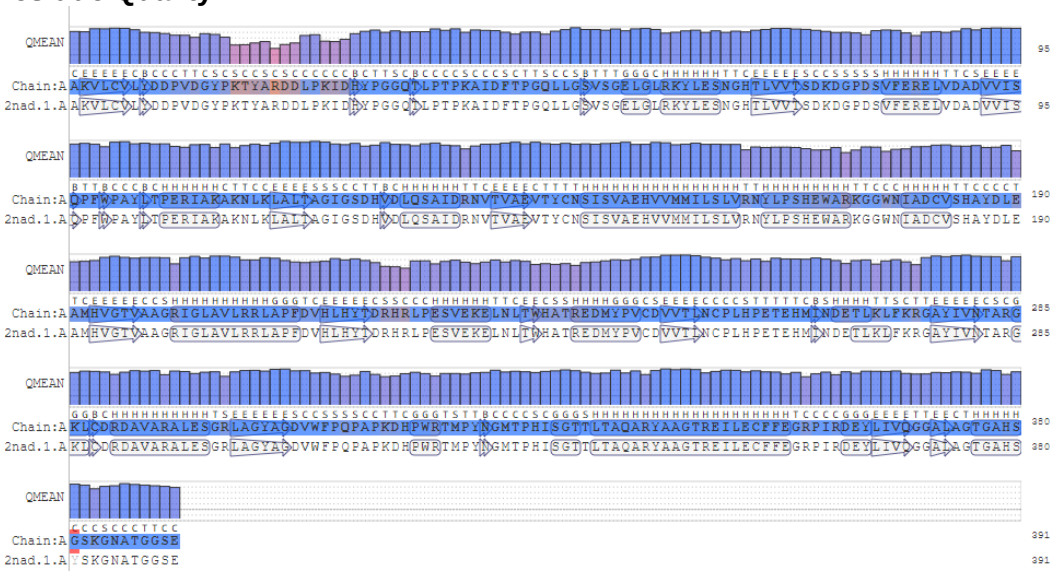


Figure 45. Homology modelling result for the mutant 2NADa_Y381G.

(a) **(b)**

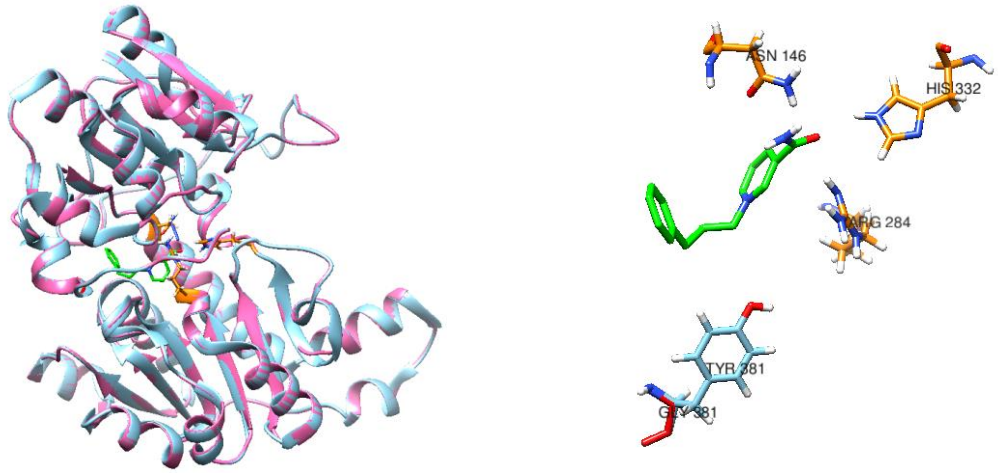


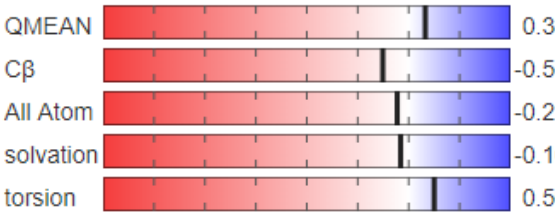
Figure 46. Mutant 2NADa_A283F (a) Comparison with the wild protein (in blue), (b) modified residue indicated in red and the biomimetic cofactor P3NAH in green with the best predicted position for the WT enzyme.

A.9. Mutant 2NADa_R222G

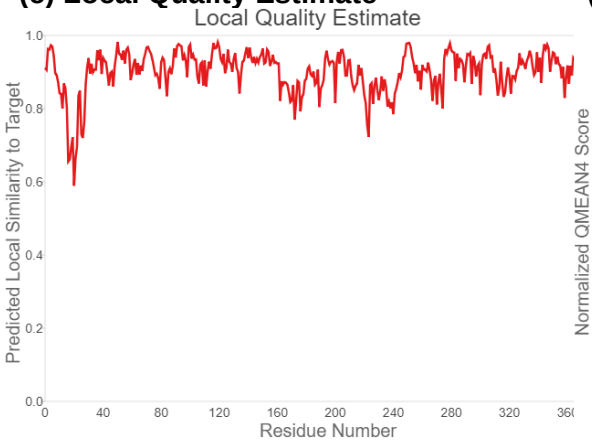
(a) Homology Modelling details

Mutant name	2NADa_R222G
Template	2NAD chain A
Change in residues	Arg 222 to Gly
QMEAN	0.33
Seq Identity (%)	99.74
Seq Similarity	0.61

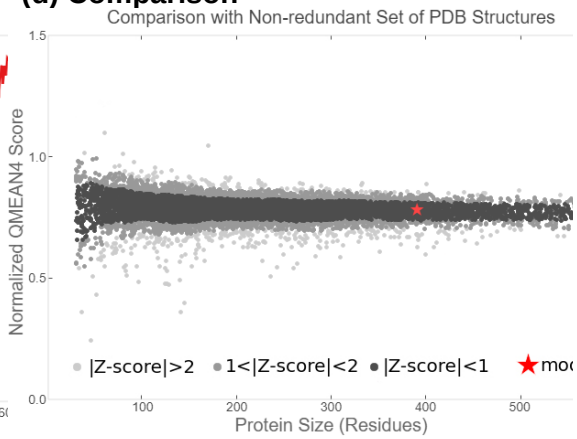
(b) Global Quality Estimate



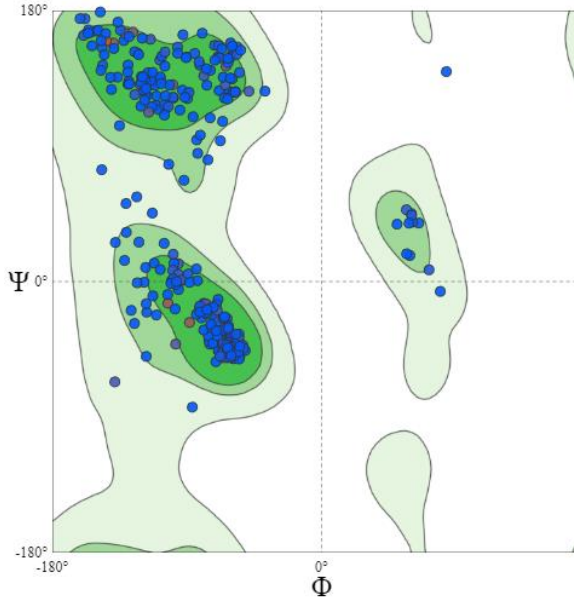
(c) Local Quality Estimate



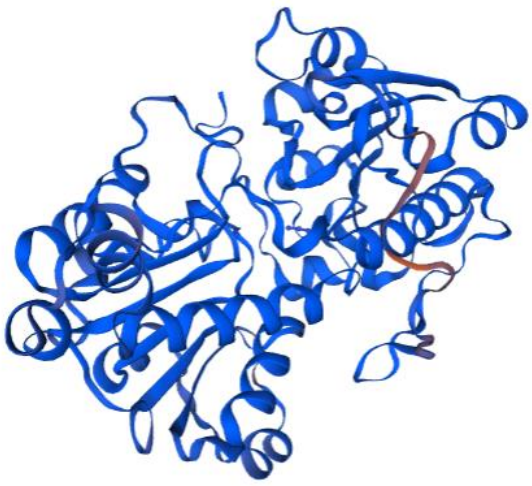
(d) Comparison



(e) Ramachandran Plot



(f) Structure



(g) FASTA sequence

>2NAD A Mutant Arg 222 to Gly
AKVLCVLYDDPVDGYPKTYARDDLPKIDHYPGGQTLPTPKAIDFTPGQLLGSVSGELGLRKYLESNGHTLVVTSDKDGPDSVFERELVDADVVISQPFWPAYLTPERIAKAKNL
KLALTAGIGSDHVDLQSAIDRNVTVAEVTYCNSISVAEHVVMMILSLVRNYLPSHEW
ARKGGWNIADCVSHAYDLEAMHVGTVAAGRIGLAVLRRLAPFDVHLHYTDGHRLP
ESVEKELNLTWHATREDMYPVCDVVTLNCPLHPETEHMINDETLKLFKRGAYIVNT
ARGKLCDRDAVARALESGRLAGYAGDVWFPQPAPKDHPWRTMPYNGMTPHISGT
TLTAQARYAAGTREILECFFEGRPIRDEYLIVQGGALAGTGAHSYSKGNATGGSE

(h) Residue Quality

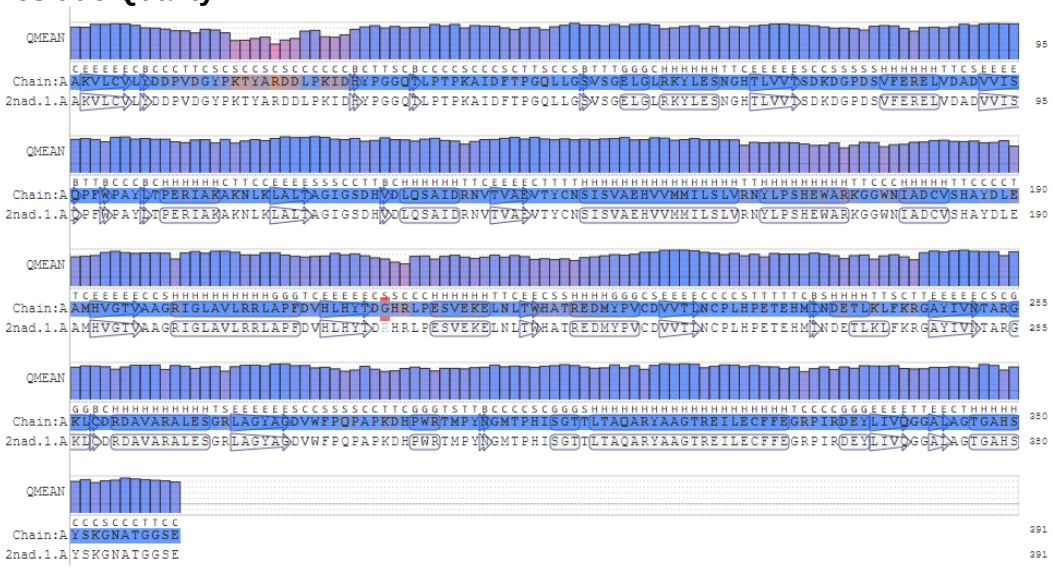


Figure 47. Homology modelling result for the mutant 2NADa_R222G.

(a) **(b)**

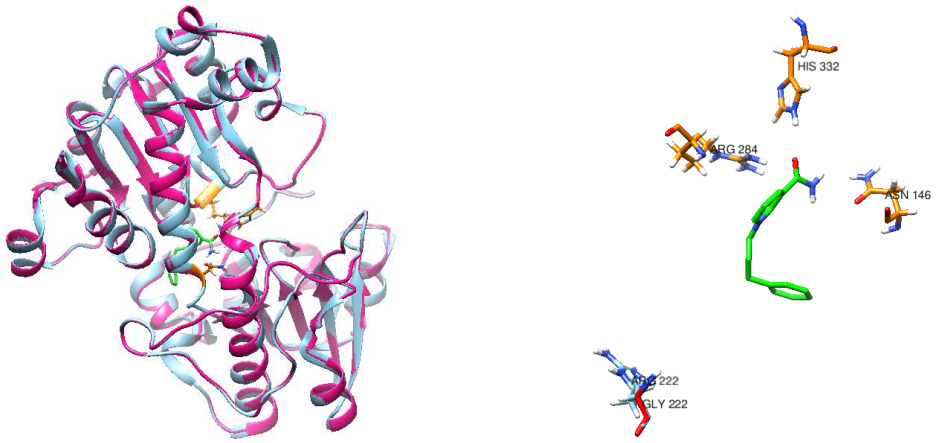


Figure 48. Mutant 2NADa_R222G (a) Comparison with the wild protein (in blue), (b) modified residue indicated in red and the biomimetic cofactor P3NAH in green with the best predicted position for the WT enzyme.

Appendix B: Docking of P3NAH in each enzyme

B.1. NADH distance from the relevant residues in the wild type enzyme

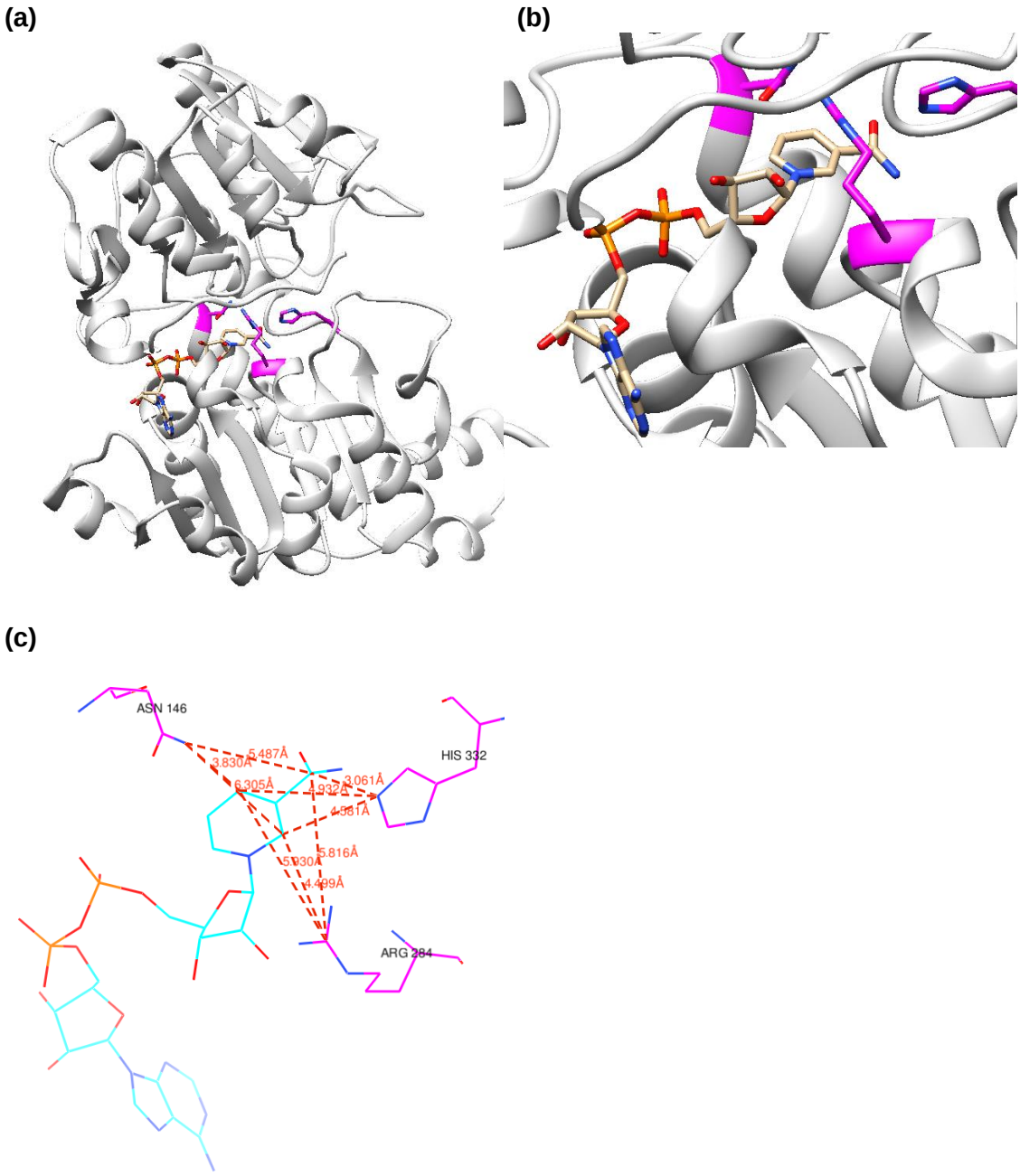


Figure 49. (a), (b), (c) Visual assessment of the position NADH in the wild type enzyme.

B.2. Docking between P3NAH and the different enzymes

B.2.1. Wild type enzyme

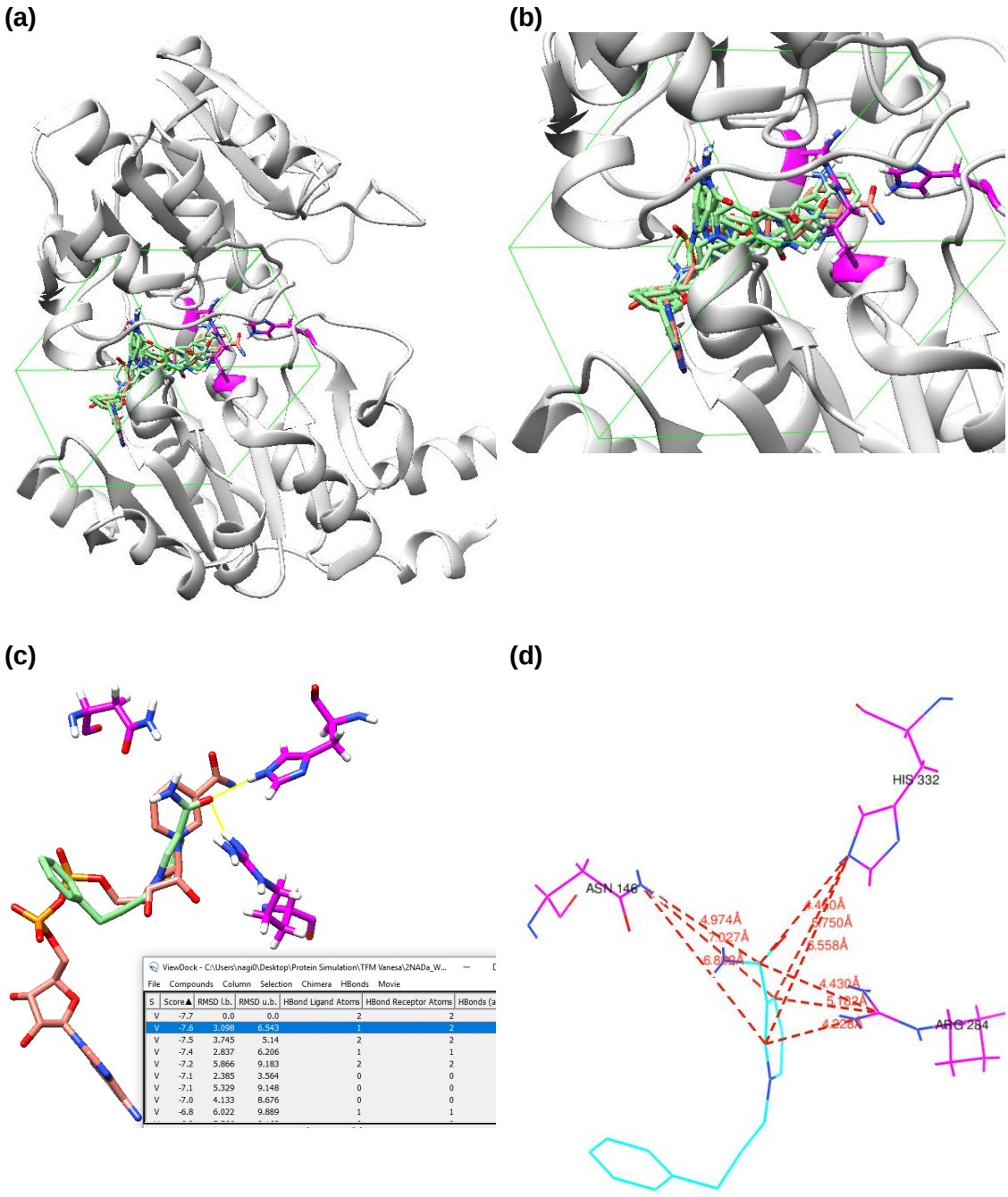


Figure 50. (a), (b), (c) Visual assessment of the docking results for the wild type enzyme with the cofactor NADH indicated in orange, and the several positions for the biomimetic cofactor P3NAH in green, (d) distance to relevant residues.

B.2.2. Mutant 2NADa_A283F

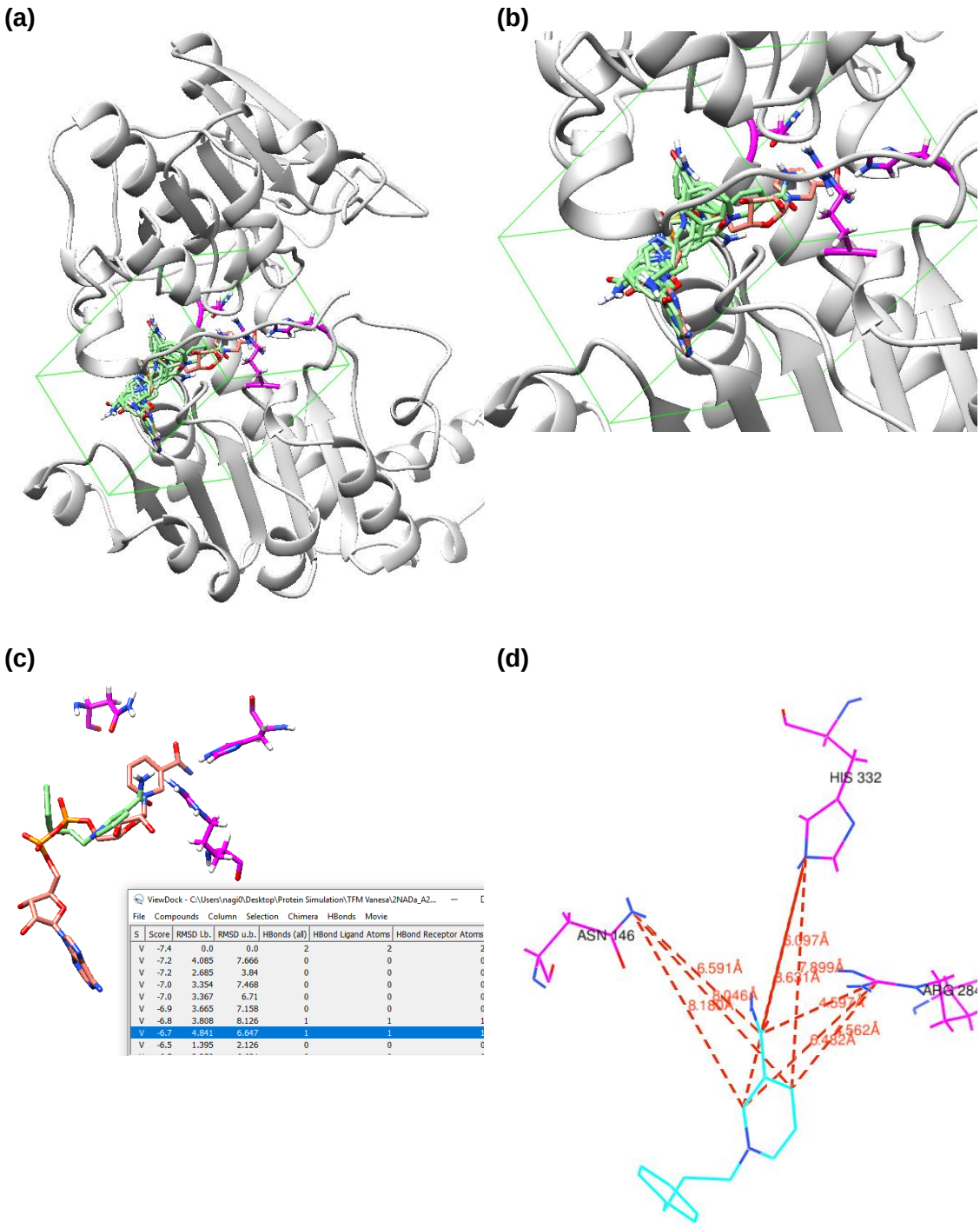


Figure 51. (a), (b), (c) Visual assessment of the docking results for the mutant 2NADa_A283F with the cofactor NADH indicated in orange, and the several positions for the biomimetic cofactor P3NAH in green, (d) distance to relevant residues.

B.2.3. Mutant 2NADa_A283Y

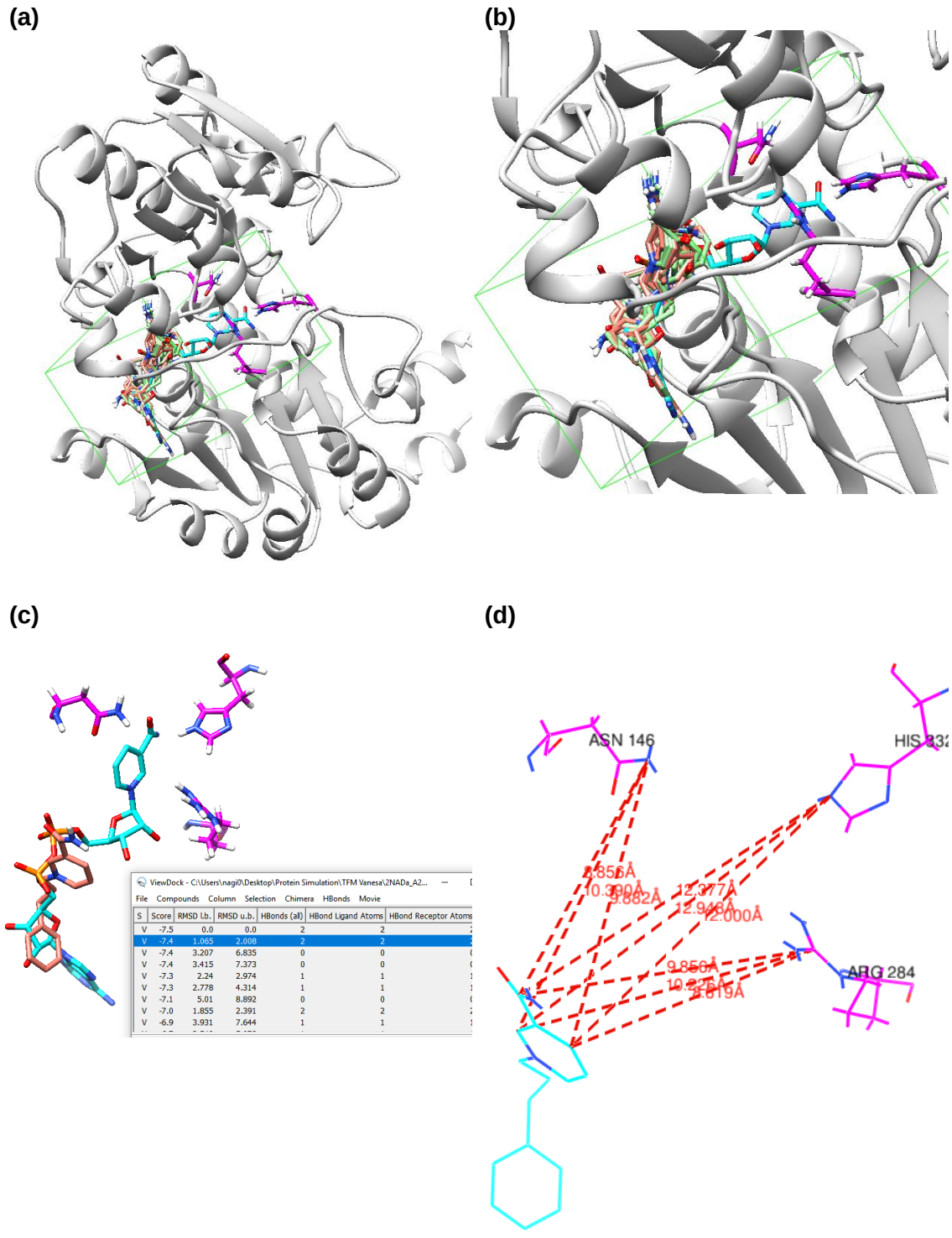


Figure 52. (a), (b), (c) Visual assessment of the docking results for the mutant 2NADa_A283Y with the cofactor NADH indicated in cyan, and the several positions for the biomimetic cofactor P3NAH in green and orange, (d) distance to relevant residues.

B.2.4. Mutant 2NADa_G123F

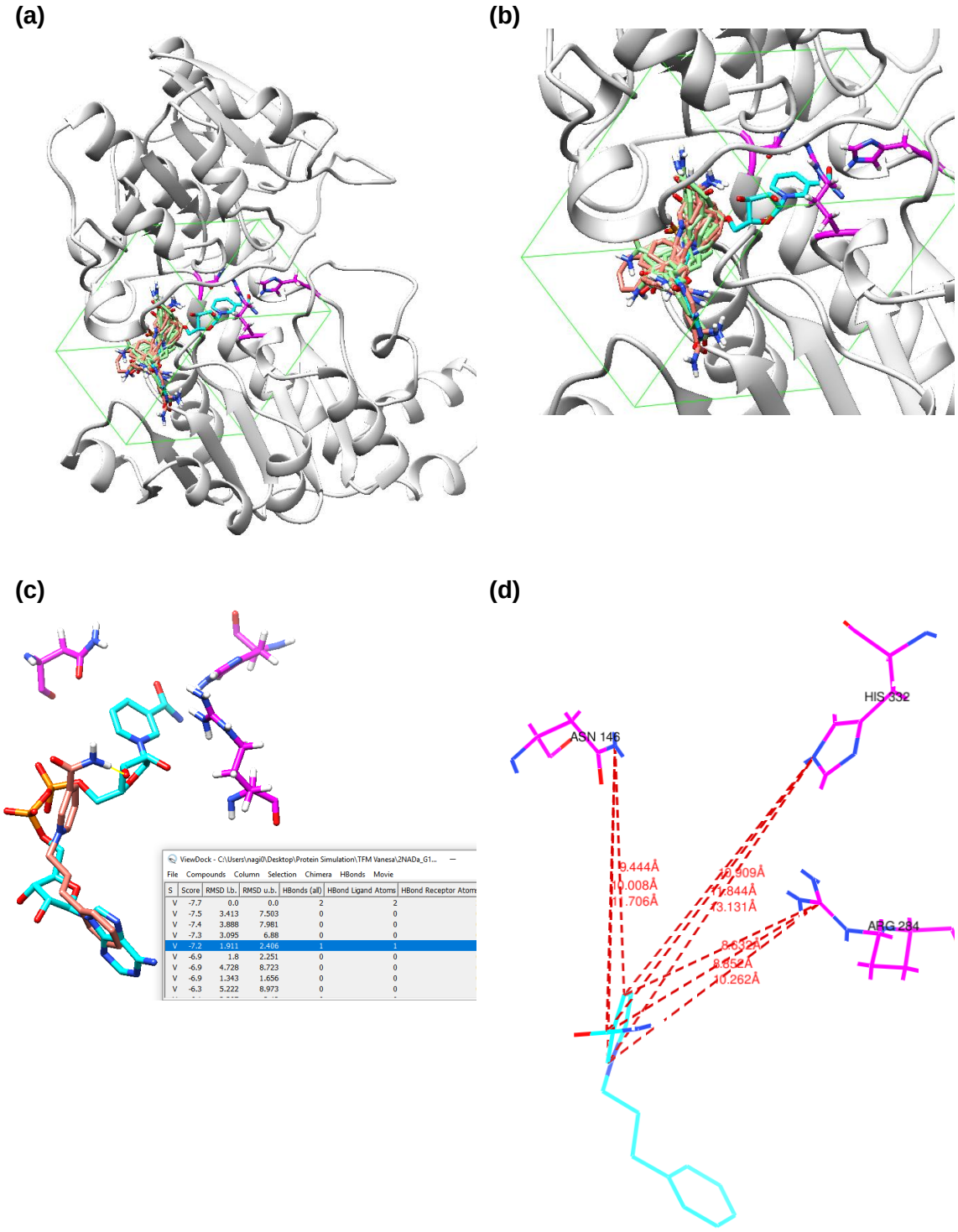


Figure 53. (a), (b), (c) Visual assessment of the docking results for the mutant 2NADa_A123F with the cofactor NADH indicated in cyan, and the several positions for the biomimetic cofactor P3NAH in green and orange, (d) distance to relevant residues.

B.2.5. Mutant 2NADa_G123Y

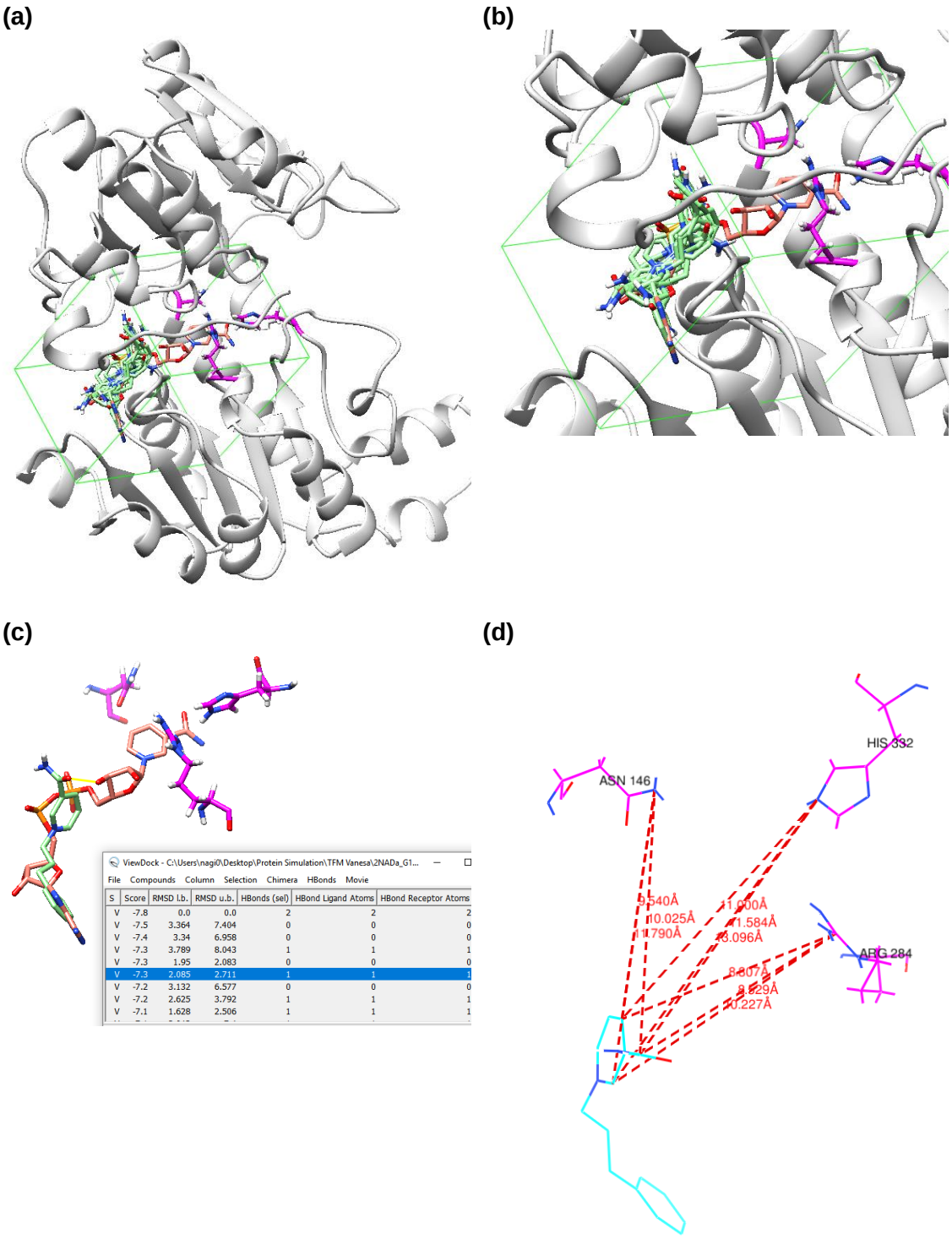


Figure 54. (a), (b), (c) Visual assessment of the docking results for the mutant 2NADa_G123Y with the cofactor NADH indicated in orange, and the several positions for the biomimetic cofactor P3NAH in green, (d) distance to relevant residues.

B.2.6. Mutant 2NADa_R222G

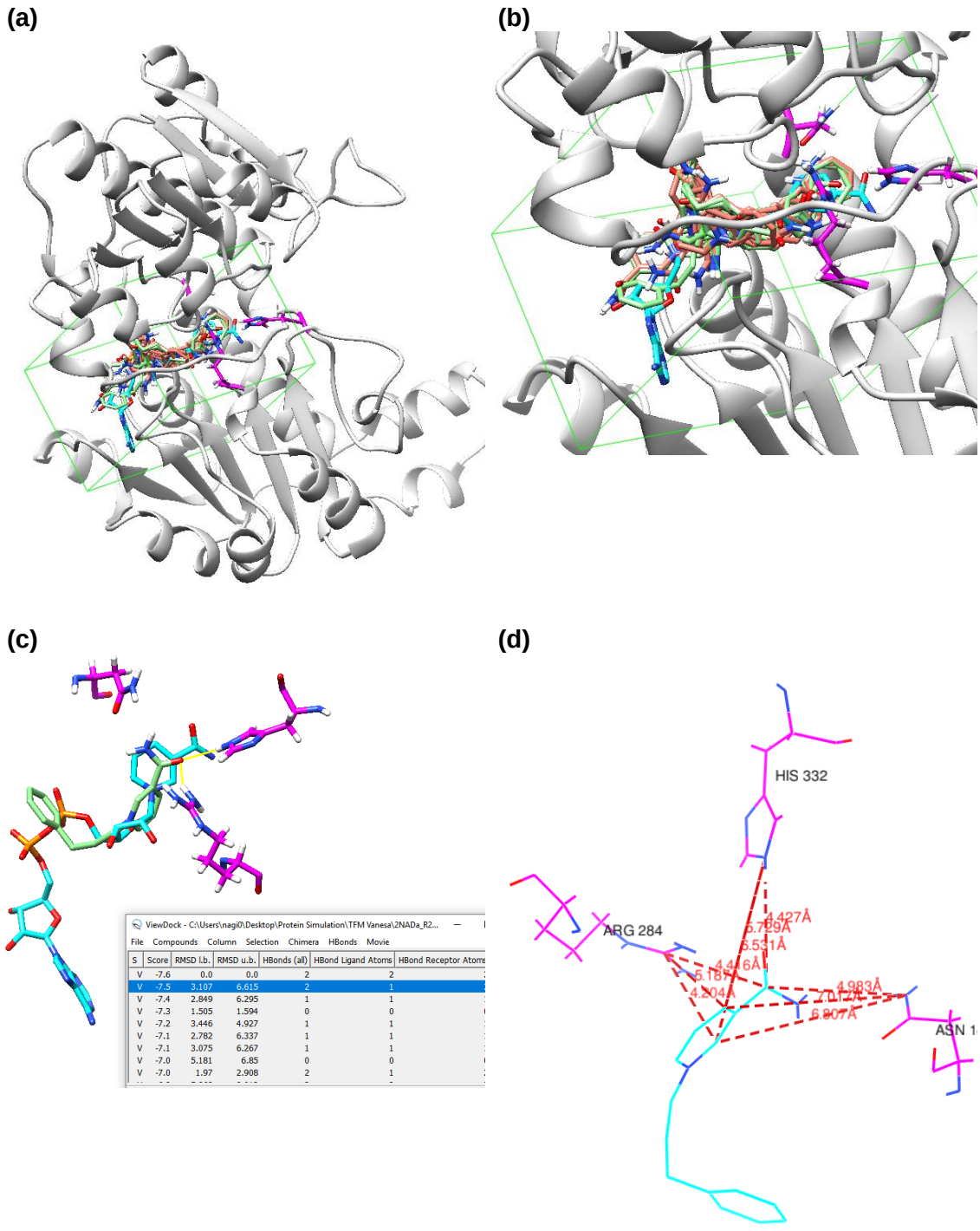


Figure 55. (a), (b), (c) Visual assessment of the docking results for the mutant 2NADa_R222G with the cofactor NADH indicated in cyan, and the several positions for the biomimetic cofactor P3NAH in green and orange, (d) distance to relevant residues.

B.2.7. Mutant 2NADa_S380G

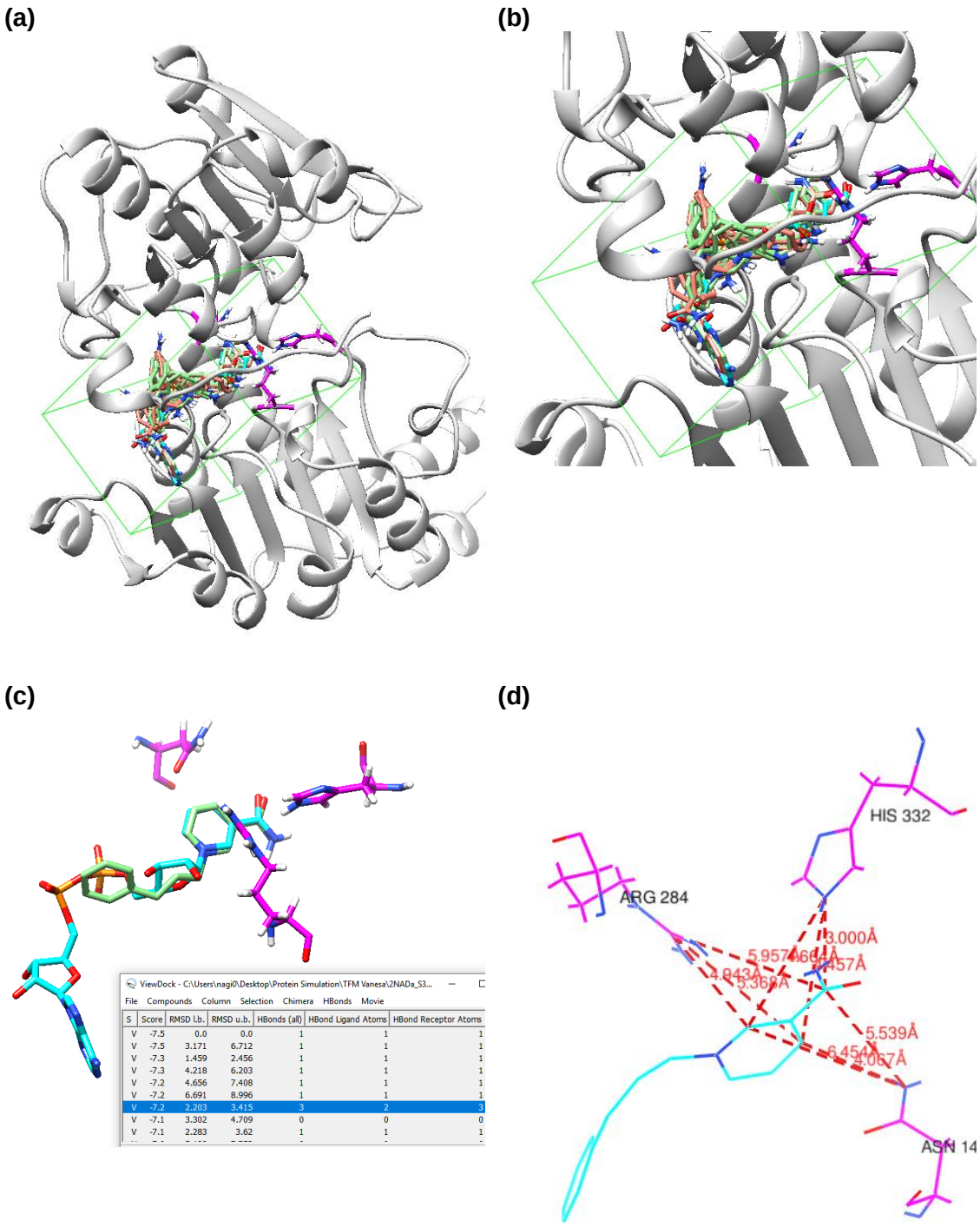


Figure 56. (a), (b), (c) Visual assessment of the docking results for the mutant 2NADa_S380G with the cofactor NADH indicated in cyan, and the several positions for the biomimetic cofactor P3NAH in green and orange, (d) distance to relevant residues.

B.2.8. Mutant 2NADa_T376G

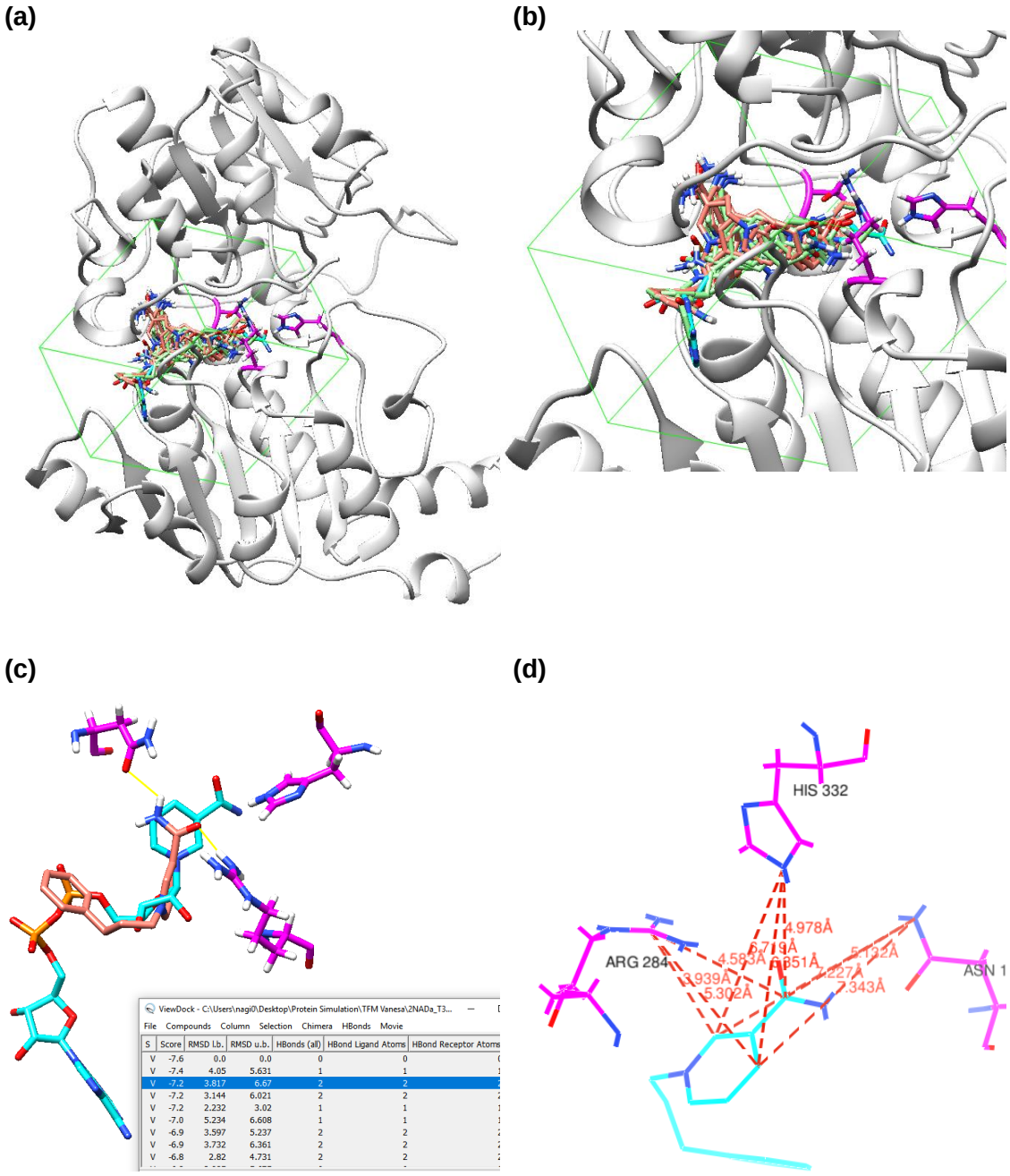


Figure 57. (a), (b), (c) Visual assessment of the docking results for the mutant 2NADa_T376G with the cofactor NADH indicated in cyan, and the several positions for the biomimetic cofactor P3NAH in green and orange, (d) distance to relevant residues.

B.2.9. Mutant 2NADa_Y381G

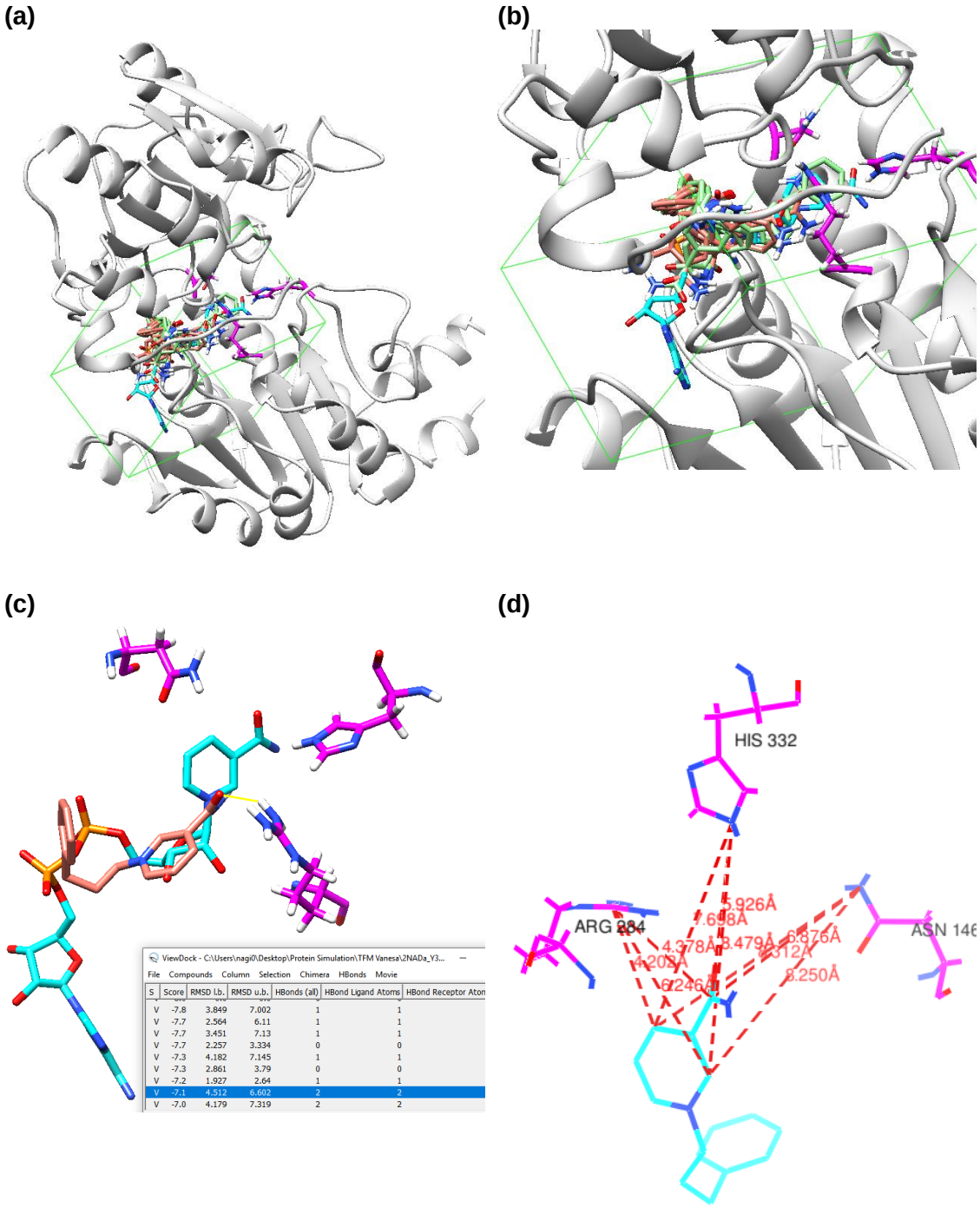


Figure 58. (a), (b), (c) Visual assessment of the docking results for the mutant 2NADa_Y381G with the cofactor NADH indicated in cyan, and the several positions for the biomimetic cofactor P3NAH in green and orange, (d) distance to relevant residues.

B.3. Numerical results

Table 8. Evaluation of the mutants docking with P3NAH.

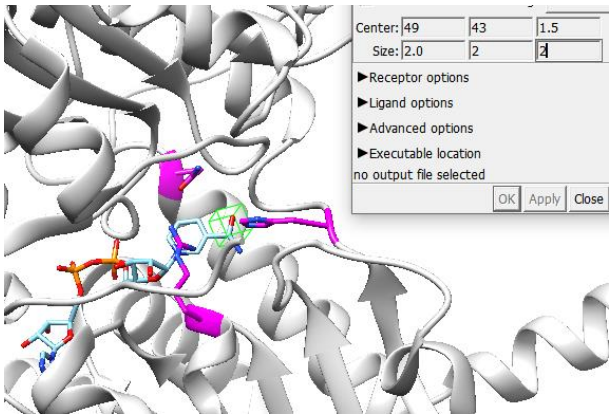
Enzyme	Binding evaluation		Distance between cofactor and residues (Å)				Difference with the atoms in NADH (Å)		
	Score	H bonds	Cofactor atoms	ARG 284 CZ	ASN 146 ND2	HIS 332 NE2			
NADH	NA	NA	C1	5.930	3.830	4.932			
			C4	4.499	6.305	4.581			
			C6	5.816	5.487	3.061			
WT	-7.6	2	C2	5.182	7.027	5.750	0.748	3.197	0.818
			C4	4.228	6.802	6.558	0.271	0.497	1.977
			C6	4.430	4.974	4.450	1.386	0.513	1.389
									1.200
2NADa_ A283F	-6.7	1	C1	4.562	8.046	7.899	1.368	4.216	2.967
			C4	6.482	8.180	8.631	1.983	1.875	4.050
			C6	4.597	6.591	6.097	1.219	1.104	3.036
									2.424
2NADa_ A283Y	-7.4	2	C1	8.619	9.882	12.000	2.689	6.052	7.068
			C4	10.226	10.390	12.948	5.727	4.085	8.367
			C6	9.856	8.856	12.377	4.040	3.369	9.316
									5.635
2NADa_ G123F	-7.2	1	C1	8.632	9.444	10.909	2.702	5.614	5.977
			C4	10.262	11.706	13.131	5.763	5.401	8.550
			C6	8.852	10.008	11.844	3.036	4.521	8.783
									5.594
2NADa_ G123Y	-7.3	1	C1	8.807	9.540	11.000	2.877	5.710	6.068
			C4	10.227	11.790	13.096	5.728	5.485	8.515
			C6	8.529	10.025	11.584	2.713	4.538	8.523
									5.573
2NADa_ R222G	-7.5	2	C1	5.187	5.729	7.017	0.743	1.899	2.085
			C4	6.807	6.531	6.204	2.308	0.226	1.623
			C6	4.427	4.983	4.416	1.389	0.504	1.355
									1.348
2NADa_ S380G	-7.2	2	C1	5.368	4.067	4.457	0.562	0.237	0.475
			C4	4.943	6.454	4.664	0.444	0.149	0.083
			C6	5.957	5.539	3.000	0.141	0.052	0.061
									0.245
2NADa_ T376G	-7.2	2	C1	5.302	7.343	6.351	0.628	3.513	1.419
			C4	3.939	7.227	6.719	0.560	0.922	2.138
			C6	4.583	5.132	4.978	1.233	0.355	1.917
									1.409

Table 8. Evaluation of the mutants docking with P3NAH (continued).

Enzyme	Binding evaluation		Distance between cofactor and residues (Å)				Difference with the atoms in NADH (Å)		
	Score	H bonds	Cofactor atoms	ARG 284 CZ	ASN 146 ND2	HIS 332 NE2			
2NADa_ Y381G	-7.1	2	C1	4.202	8.312	7.698	1.728	4.482	2.766
			C4	6.246	8.250	8.479	1.747	1.945	3.898
			C6	4.378	6.876	5.926	1.438	1.389	2.865
									2.473

Appendix C: Ligand Transport Analysis

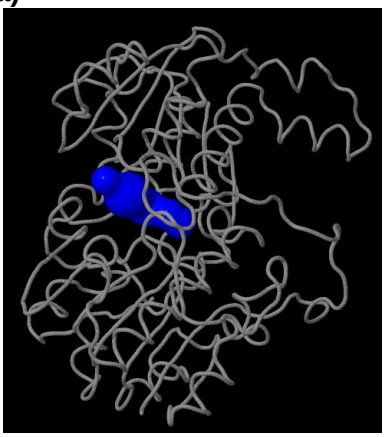
C.1. Wild type enzyme

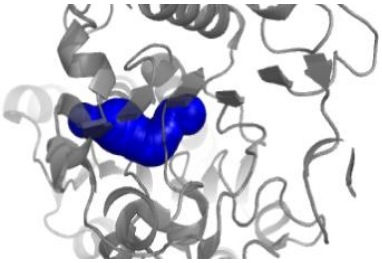


desired_radius 5
exclude_residue_names null
shell_radius 3
shell_depth 4
clustering_threshold 3.5
probe_radius 0.9
max_distance 3
compute_tunnel_residues yes
visualization_tunnel_sampling_step 0.5
seed 1

Figure 59. Configuration of Caver Web for the WT enzyme.

(a)





bottleneck radius [Å]:	2.0
length [Å]:	15.7
distance to surface [Å]:	13.5
curvature:	1.2
throughput:	0.75
number of residues:	40
number of bottlenecks:	1

(b)

radius [Å]:	2.0
distance from starting point [Å]:	14.5
coordinates [Å, Å, Å]:	[51.2, 34.3, -8.7]

Bottleneck residues (9):

amino acid	ID	chain
Arg	201	A
Ile	202	A
Asp	221	A
Pro	256	A
Thr	376	A
His	379	A
Ser	380	A
Ala	199	A
Gly	200	A

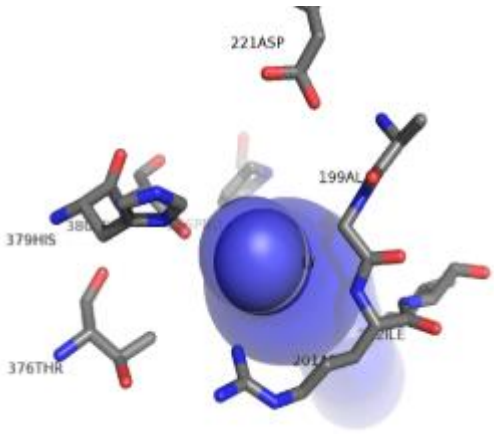


Figure 60. (a) Tunnel overview and (b) Bottleneck residues for the WT enzyme.

(a) Go inside **(b) Getting out**

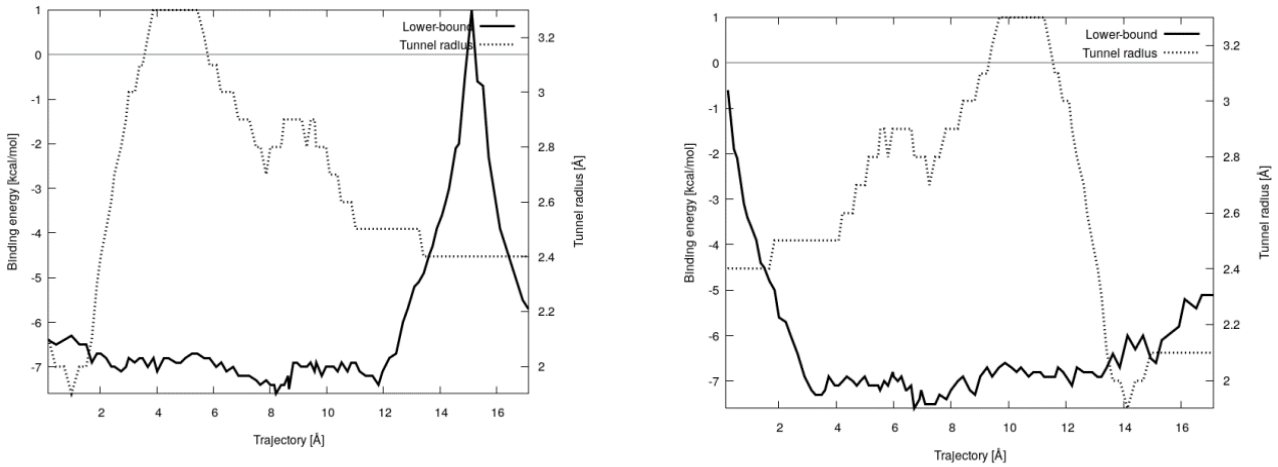
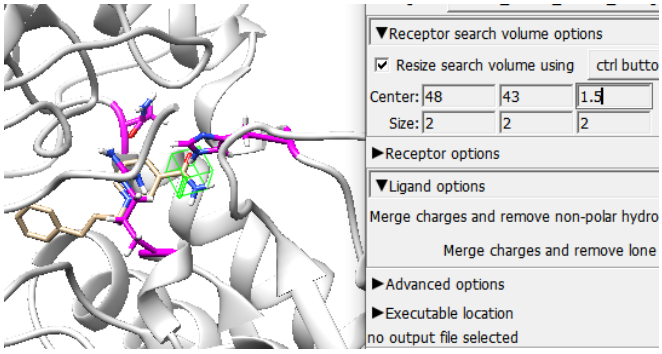


Figure 61. Energy profile of P3NAH for going inside (a) and outside (b) of the active site in the WT enzyme.

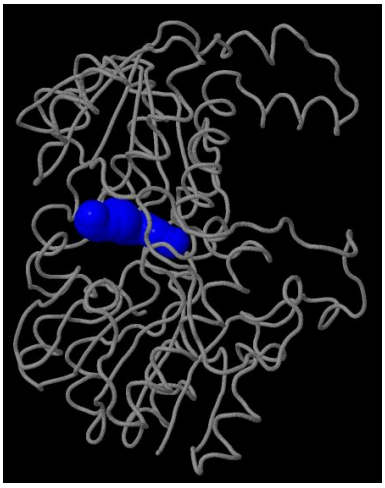
C.2. Mutant 2NADa_S380G

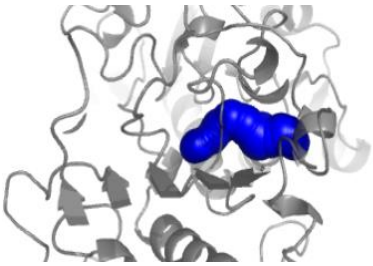


starting_point_coordinates 48.0 43.0 1.5
desired_radius 5
exclude_residue_names null
shell_radius 3
shell_depth 4
clustering_threshold 3.5
probe_radius 0.9
max_distance 3
compute_tunnel_residues yes
visualization_tunnel_sampling_step 0.5
seed 1

Figure 62. Configuration of Caver Web for 2NADa_S380G.

(a)





bottleneck radius [Å]:	1.9
length [Å]:	16.4
distance to surface [Å]:	13.2
curvature:	1.2
throughput:	0.72
number of residues:	38
number of bottlenecks:	1

(b)

radius [Å]:	1.9
distance from starting point [Å]:	0.0
coordinates [Å, Å, Å]:	[50.6, 42.5, 0.2]

Bottleneck residues (8):

amino acid	ID	chain
Gly	200	A
Arg	201	A
Asp	221	A
His	223	A
Thr	376	A
His	379	A
Gly	380	A
Ala	199	A

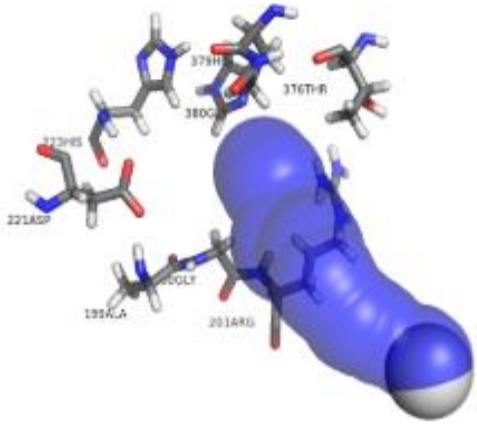


Figure 63. (a) Tunnel overview and (b) Bottleneck residues for 2NADa_S380G.

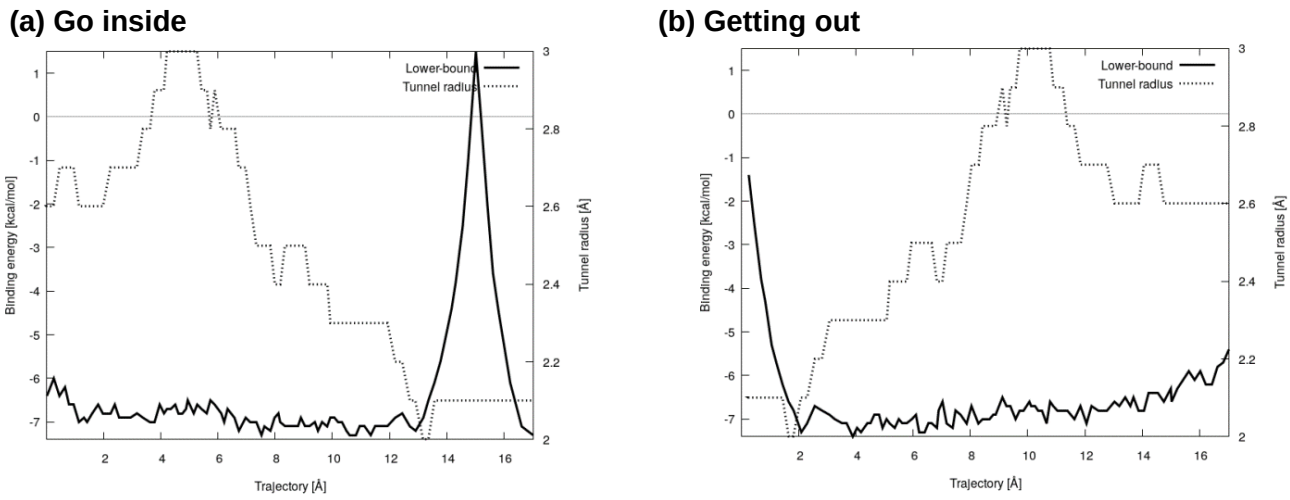


Figure 64. Energy profile of P3NAH for going inside (a) and outside (b) of the active site in 2NDAa_S380G.