DOCTOR OF SCIENCES

Characterization of the oligomeric interface of Lactate Dehydrogenase-B (LDH-B) through crystallographic and molecular dynamics approaches

Van Gysel, Megane

Award date:
2023

Awarding institution:
University of Namur

UNIVERSITÉ DE NAMUR

FACULTÉ DES SCIENCES

Characterization of the oligomeric interface of Lactate Dehydrogenase-B (LDH-B) through crystallographic and molecular dynamics approaches

A thesis submitted by **Mégane Van Gysel**
for the fulfilment of the requirements
for the degree of Doctor in Sciences

Thesis Committee:

Prof. Jérôme de Ruyck (Université de Lille)

Prof. Raphaël Frédérick (Université Catholique de Louvain)

Prof. Yoann Olivier (Université de Namur, president of the jury)

Dr. Catherine Michaux (Université de Namur)

Prof. Johan Wouters (Université de Namur, thesis advisor)

February 2023

Université de Namur
FACULTE DES SCIENCES
Rue de Bruxelles, 61 – 5000 Namur, Belgique

Characterization of the oligomeric interface of Lactate Dehydrogenase-B (LDH-B) through crystallographic and molecular dynamics approaches

by Mégane Van Gysel

Abstract

Cancer cells acquired specific characteristics, called Hallmarks, such as cell metabolism dysregulation. It is in this context that we study Lactate Dehydrogenase (LDH), an active tetrameric enzyme involved in lactate metabolism. LDH catalyzes the interconversion of pyruvate into L-lactate, in the presence of the NAD cofactor, at the end of the glycolytic pathway. Lactate, previously considered as a waste product, has been shown to be essential for the anabolic growth and proliferation of cancer cells. Thus, LDH is involved in several processes, such as metabolic cooperation, autophagy, angiogenesis, invasion, and metastasis. Based on these considerations, LDH has been identified as a promising target for cancer treatment. Despite the development of many inhibitors targeting its active site, LDH remains a challenging therapeutic target. A second strategy targeting the self-association of LDH homo-tetramer can be considered. It aims at destabilizing the protein complex with disruptors interacting at the tetrameric interface of the protein. Therefore, the main objective of this work consists in a better understanding of the tetramerization process for one of the two main isoforms of LDH, LDH-B. For this purpose, a multi-step strategy has been established.

The first part of this work is dedicated to the LDH-B tetrameric interface, via the study of two tetramerization sites. For this purpose, different forms of LDH-B were used: a truncated form (LDH-Btr), as well as single-point mutants. The protein systems were produced, purified, and studied by crystallography to determine their three-dimensional structures. The second tetramerization site was also characterized by molecular dynamics simulations, using the Steered Molecular Dynamics method. This consists in studying the tetramer self-disassembly of LDH-B. This approach has allowed the identification of five hotspots, corresponding to residues essential for the tetramerization of the protein. The second part of this work was dedicated to the characterization of the binding mode of molecules interacting at the tetrameric interface of LDH-B. The LDH-B-ligand complexes were studied by crystallography. The structural data has allowed to highlight two distinct binding sites.

Ph.D. thesis in Sciences
Namur, February 2023

Université de Namur
FACULTE DES SCIENCES
Rue de Bruxelles, 61 – 5000 Namur, Belgique

Caractérisation de l'interface oligomérique de la Lactate Déshydrogénase B (LDH-B) via des approches cristallographique et de dynamique moléculaire

par Mégane Van Gysel

Résumé

Les cellules cancéreuses sont dotées de caractéristiques spécifiques, appelées *Hallmarks*. L'une d'entre elle consiste en la dérégulation du métabolisme cellulaire. C'est dans ce contexte que nous étudions la Lactate Déshydrogénase (LDH), une enzyme active sous forme tétramérique impliquée dans le métabolisme du lactate. Elle catalyse la réaction d'interconversion du pyruvate en L-lactate en présence du co-facteur NAD à la fin de la voie glycolytique. Le lactate, longtemps considéré comme un produit résiduel, a été mis en évidence comme essentiel pour la croissance anabolique, ainsi que la prolifération des cellules cancéreuses. Ainsi, la LDH est impliquée dans plusieurs processus, tels que la coopération métabolique, l'autophagie, l'angiogenèse ou encore l'invasion et la métastatisation. Basée sur ces évidences, la LDH a été définie comme une cible prometteuse dans le traitement du cancer. Malgré la conception de nombreux inhibiteurs ciblant son site actif, la LDH reste un défi de taille en tant que cible thérapeutique. Une seconde stratégie visant à cibler l'auto-association de la LDH (homo-tétramère) peut être envisagée. Elle a pour but de déstabiliser le complexe protéique à l'aide de disrupteurs interagissant à l'interface tétramérique de la protéine. Dès lors, l'objectif principal de ce travail consiste en une meilleure compréhension du processus de tétramérisation d'un des deux principaux isoformes de la LDH, la LDH-B. Pour ce faire, une stratégie en plusieurs étapes a été établie.

La première partie de ce travail est consacrée à la caractérisation de l'interface tétramérique de la LDH-B, via l'étude de deux sites de tétramérisation. Pour ce faire, différentes formes de la LDH-B ont été utilisées : une forme tronquée (LDH-Btr), ainsi que plusieurs mutants ponctuels. Les différents systèmes protéiques ont été produits, purifiés et étudiés par cristallographie en vue de déterminer leurs structures tridimensionnelles. Le second site de tétramérisation a également été caractérisé par simulations de dynamique moléculaire, via la méthode de *Steered Molecular Dynamics* qui consiste à étudier le processus de dissociation des différentes sous-unités de la LDH-B. Cette étude a permis de mettre en évidence cinq *« hotspots »*, correspondant à des résidus essentiels à la formation d'un tétramère stable de LDH-B. La seconde partie de ce travail a été consacrée à la caractérisation du mode de liaison de molécules se liant à l'interface tétramérique de la LDH-B. Les complexes LDH-B-ligand ont été étudiés par cristallographie. Les données structurales ont permis de mettre en évidence deux sites distincts de liaison des molécules.

Dissertation doctorale en Sciences
Namur, Février 2023

Au terme de cette thèse, j'aimerais remercier toutes les personnes qui ont contribué de près ou de loin à son élaboration.

Tout d'abord, j'aimerais remercier mon promoteur, le professeur Johan Wouters, de m'avoir accueillie au sein de son laboratoire, et ce dès le mémoire. Merci de m'avoir donné goût à la chimie biologique structurale, mais aussi de m'avoir guidée et donné de précieux conseils tout au long de ces cinq années.

J'aimerais aussi exprimer toute ma gratitude envers le Dr. Laurence Leherte. Merci pour toutes les connaissances que vous m'avez transmises en dynamique moléculaire, pour le temps passé à discuter et à m'aider sur mon projet de thèse.

J'aimerais également remercier les membres de mon comité d'accompagnement et de mon jury de thèse (les professeurs Jérôme de Ruyck, Raphaël Frédérick, Yoann Olivier et le Dr. Catherine Michaux), d'avoir accepté d'évaluer ce travail.

Ensuite, je voudrais remercier toutes les équipes avec qui j'ai pu échanger et collaborer sur le projet PDR Télévie, les équipes des professeurs Pierre Sonveaux et Marianne Fillet, et notamment l'équipe du professeur Raphaël Frédérick. Merci beaucoup aux Dr. Léopold Thabault et Maxime Liberelle, ainsi qu'à Chiara Brustenga et Perrine Savoyen avec qui j'ai eu l'occasion de travailler sur le projet LDH.

Je tenais à remercier le Télévie qui a financé ces recherches et qui m'a permis de travailler sur un sujet passionnant. Merci aussi à Aurélie Pirlot pour sa bonne humeur lors des différentes activités Télévie auxquelles j'ai eu la chance de participer.

Cette thèse n'aurait pas été possible sans l'aide scientifique de nombreuses personnes et l'accès à différentes infrastructures. Merci aux différents laboratoires dans lesquels j'ai été accueillie pour réaliser différentes expériences, le département de pharmacie, l'URBC, mais aussi l'URBM. Merci aux membres de ces différentes équipes pour leur accueil, ainsi que leur aide. J'ai également une pensée pour les « local contacts » du synchrotron Soleil. Merci aux Dr. William Shepard, Pierre Montaville, Serena Sirigu et Martin Savko des lignes PX1 et PX2-A pour leur support, leurs conseils et leur disponibilité lors des différents runs synchrotron. Ce travail n'aurait pas non plus été possible sans l'accès aux ressources computationnelles : merci au CECI. J'ai aussi une pensée pour Frédéric Wautelet et Laurent Demelenne que j'aimerais remercier pour leur support informatique et leur aide tout au long de cette thèse.

Bien évidemment, une mention spéciale pour mes collègues (anciens et actuels) du CBS. J'aimerais d'abord remercier les « anciens », Quentin, Sébastien, Thomas et Andréa, avec qui j'ai partagé le début de ma thèse. Mention particulière pour Quentin et Sébastien, mes collègues et co-équipiers du Télévie. Merci pour toutes les activités Télévie qu'on aura eu l'occasion de

partager ensemble ! Merci aussi à Nikolay et toutes ses anecdotes racontées sur les temps de midi. J'aimerais aussi remercier tous mes autres collègues avec qui j'ai eu l'occasion de partager quelques mois ou même l'entièreté de ma thèse. Une chose dont je suis certaine, c'est que vous retiendrez mes super goûts musicaux et ma culture musicale très étendue pour la variété française ! Ce que je retiendrais moi, ce sont nos fameux runs synchrotron, tous les pokés, les sushis, les gateaux,…qu'on aura partagés, les moments de rédaction que j'aurai eu l'occasion de partager avec certaines d'entre vous, tous ces moments de thèse… qui n'auraient pas été pareil sans vous ! Alors merci à vous les filles, Laurie, Marie, Kalina, Elise, Manon et Jeanne. Je ne vous oublie pas non plus, Gilles et Tanguy. Merci pour votre bonne humeur et votre humour à toute épreuve, pour ces nombreuses galettes des rois partagées…la fin de thèse n'aurait pas non plus été pareil sans vous ! Je tiens également à remercier tous les étudiants avec qui j'ai eu l'occasion de travailler : mes deux mémorants, Jordane et Jonathan, mais aussi mes stagiaires, Ugo, Arnaud et Sophie !

Une pensée toute particulière pour Sébastien Marx, Seb. Merci d'avoir été un parrain d'unif si cool et de m'avoir épaulée lors de mes premières années d'étude. Merci aussi de m'avoir transmis ta passion pour la recherche lorsque tu m'encadrais en mémoire. Ça nous a mené à devenir collègues par la suite. Merci pour tous ces moments de thèse (et d'après thèse), et notamment toutes les activités Télévie qu'on a eu l'occasion de partager. Mais ça ne s'arrête pas qu'à ça, merci d'être un ami si important !

J'ai aussi une pensée toute particulière pour ma famille et mes amis. Merci à vous tous pour votre soutien durant toutes ces années. Sans vous, je ne serai pas qui je suis aujourd'hui ! J'ai une pensée toute particulière pour mes parents. Merci de toujours m'avoir épaulée dans mes projets, et ce depuis le début. Il y a 11 ans, nous allions visiter pour la première fois l'université de Namur, me voilà 11 plus tard toujours aussi bien accompagnée !

Finalement, je voudrais terminer en remerciant Thomas Dal Maso, Tom. Merci pour ton aide précieuse, ton soutien sans faille et tes encouragements dans chaque moment important, même les plus difficiles. Pour le reste, tu sais déjà tout !

List of abbreviations

3D	Three-dimensional
6xHis-tag	6 x Histidine tag
Acetyl-CoA	Acetyl Coenzyme A
ADME	Absorption, Distribution, Metabolism and Excretion
AFM	Atomic Force Microscopy
ASU	Asymmetric Unit
ATP	Adenosine Triphosphate
BAX	Bcl-2-associated X
BLOSUM	Blocks Substitution Matrix
CAM	Chloramphenicol
CC-value	Correlation Coefficient value
C_f	Final concentration
CHARMM	Chemistry Harvard Molecular Mechanics
COM	Center of Mass
Cryo-EM	Cryogenic Electron Microscopy
DSF	Differential Scanning Fluorimetry
E. coli	*Escherichia coli*
EGFR	Epidermal Growth Factor Receptor
ELISA	Enzyme-Linked Immunosorbent Assay
ETC	Electron Transport Chain
$FADH_2$	Flavine Adenine Dinucleotide (reduced form)
FBS	Fragment-Based Screening
FT	Flow Through
FRET	Föster Resonance Energy Transfer
FX	Fluoxetine
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GLUT-1	Glucose Transporter 1
H2B	Histone 2B
HIF-1	Hypoxia-Inducible Factor-1
hLDH-B	Human Lactate Dehydrogenase B
HTS	High Throughput Screen
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
ITC	Isothermal Calorimetry
ITF	Intrinsic Tryptophan Fluorescence

KAN	Kanamycin
K_d	Dissociation constant
K_M	Michaelis-Menten constant
LB	Lysogeny Broth
LDH	Lactate Dehydrogenase
logP	Logarithm of Partition coefficient
Mb	Megabody
MBP	Maltose-Binding Protein
MCT	Monocarboxylate Transporter
MD	Molecular Dynamics
MDM2	Mouse Double Minute 2
MDMX	Mouse Double Minute X
MP	Mass Photometry
MS	Mass Spectrometry
MST	Microscale Thermophoresis
MW	Molecular Weight
NAD^+	β-Nicotinamide Adenine Dinucleotide (oxidized form)
NADH	β-Nicotinamide Adenine Dinucleotide (reduced form)
nanoDSF	nanoscale Differential Scanning Fluorimetry
Nb	Nanobody
NDSB-195	Non-Detergent Sulfobetaine 195
NMR	Nuclear Magnetic Resonance
NMR TOCSY	Nuclear Magnetic Resonance Total Correlation Spectroscopy
NMR WaterLOGSY	Nuclear Magnetic Resonance Water-Ligand Observed via Gradient Spectroscopy
OD	Optical Density
o/n	Overnight
OXM	Oxamate
OXPHOS	Oxidative Phosphorylation
PAGE	Polyacrylamide Gel Electrophoresis
PDB	Protein Data Bank
PDH	Pyruvate Dehydrogenase
PD-1	Programmed cell Death-1
PD-L1	Programmed cell Death-Ligand 1
PEG	Polyethylene Glycol
PHD-2	Prolyl Hydroxylase-2
pH_e	Extracellular pH

pI	Isoelectric point
PPI	Protein-Protein Interaction
PSI	Protein Structure Initiative
R	Resolution
RMSD	Root-Mean-Square Deviation
RMSF	Root-Mean-Square Fluctuation
RPM	Rotation Per Minute
RSA	Relative Surface Accessibility
RT	Room Temperature
SAR	Structure-Activity Relationship
SD	Standard Deviation
SDS-PAGE	Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SMD	Steered Molecular Dynamics
STD NMR	Saturation Transfer Difference Nuclear Magnetic Resonance
T	Temperature
TCA cycle	Tricarboxylic Acid cycle
TEV	*"Tobacco etch virus"*
T_m	Melting temperature
TMAO	Trimethylamine N-Oxide
TRI	Triprolidine
V-ATPase	Vacuolar-type ATPase
VEGF	Vascular Endothelium Growth Factor
V_f	Final volume
v_{pull}	Pulling velocity
XRD	X-Ray Diffraction

List of amino acids

Amino acid	3-letter code	1-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophane	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

List of LDH-B forms used in the work

Acronym	Name	Description
LDH-B	Lactate Dehydrogenase B	Full-length form
LDH-Btr	Truncated Lactate Dehydrogenase B	19 N-terminal amino acids are deleted
LDH-Btr$_{cleaved}$	Cleaved truncated Lactate Dehydrogenase B	6xHis-tag cleaved
6xHis-LDH-Btr	Truncated Lactate Dehydrogenase	N-terminal 6xHis-tag
LDH-Btr-6xHis	Truncated Lactate Dehydrogenase	C-terminal 6xHis-tag
LDH-B^{E61A}	Mutated Lactate Dehydrogenase B	Glutamate 61 mutated to alanine
LDH-B^{D64A}	Mutated Lactate Dehydrogenase B	Aspartate 64 mutated to alanine
LDH-B^{H67A}	Mutated Lactate Dehydrogenase B	Histidine 67 mutated to alanine
LDH-B^{L70A}	Mutated Lactate Dehydrogenase B	Leucine 70 mutated to alanine
LDH-B^{F71A}	Mutated Lactate Dehydrogenase B	Phenylalanine 71 mutated to alanine

Table of Contents

Preface

This work is funded by Télévie through a "*Projet De Recherche (PDR)"* that involves the collaboration between teams from different universities (UCLouvain, ULiège and UNamur). The PDR research project aims to study the LDH protein family to design new therapeutic drugs via a multidisciplinary approach. This PhD thesis focuses on the structural characterization of the LDH-B system, corresponding to one of the two main isoforms, using both X-ray crystallography and Molecular Dynamics.

Therefore, this work is organized into different chapters. The first chapter corresponds to the introduction and addresses several topics useful for the understanding of this work. The introduction first discusses the protein-protein interaction topic as a challenge and relevant therapeutic target in drug design, and particularly for cancer treatment. Therefore, the energy metabolism of normal and cancer cells (metabolic dysregulation by cancer cells) is discussed. This part focuses on the biological function of LDH which is a protein identified as an interesting therapeutic target in cancer treatment. LDH is then described from a structural point of view before describing the two strategies proposed to target the LDH protein in cancers. The second strategy, targeting the self-association of LDH homo-tetramer via its tetrameric interface, is further developed. Finally, the introduction discusses the project context.

The second chapter corresponds to the presentation of the results. This chapter is divided into three parts (parts I, II and III). Parts I and II are dedicated to the LDH-B tetrameric interface, via the study of the tetramerization sites 1 and 2, respectively. For this purpose, different forms of LDH-B were used: a truncated form (for the study of the tetramerization site 1), as well as single-point mutants (for the study of the tetramerization site 2). These protein systems are produced, purified, and studied by crystallography to determine their three-dimensional structures. The second tetramerization site is also characterized by molecular dynamics simulations, using Steered Molecular Dynamics method. This approach consists in studying the tetramer self-disassembly of LDH-B via a computational approach. Part III is dedicated to the characterization of the binding mode of molecules interacting at the tetrameric interface of LDH-B. For this purpose, crystallography is used to study LDH-B-ligand complexes.

Finally, the third chapter describes the conclusions for each part of the work and presents several perspectives that could be envisaged for further characterizing the LDH-B system.

The reference LDH-B structure that has been used in this work is the PDB structure "1I0Z" (from the Protein Data Bank). All results presented within this work, follow the same residue numbering as the reference LDH-B structure.

Introduction

1. The human proteome

The human proteome contains a multitude of proteins, ~20 000 non-modified proteins if we consider that one gene codes for one single protein[1,2]. Nevertheless, if we consider phenomena such as alternative splicing, single amino acid polymorphism and post-translational modifications, the number of different human proteins can reach a few billion[1]. Proteins are polypeptide entities organized in four different structural levels (Fig. I.1)[3–5].

- The primary structure refers to the amino acid sequence of the protein. This one determines the three-dimensional (3D) structure of the protein and consequently its function.
- The secondary structure corresponds to the local spatial arrangement of a polypeptide backbone, such as α-helix or β-sheet (sidechains are not considered).
- The tertiary structure is defined as the global 3D arrangement of full protein atoms, formed by the different elements of secondary structure.
- The quaternary structure corresponds to a higher-structural level wherein several subunits (polypeptide chains) interact to form the protein complex. These subunits, or monomers, can be identical (homo-oligomer) or different (hetero-oligomer) (amino acid composition). The final oligomeric state governs the function of the protein.

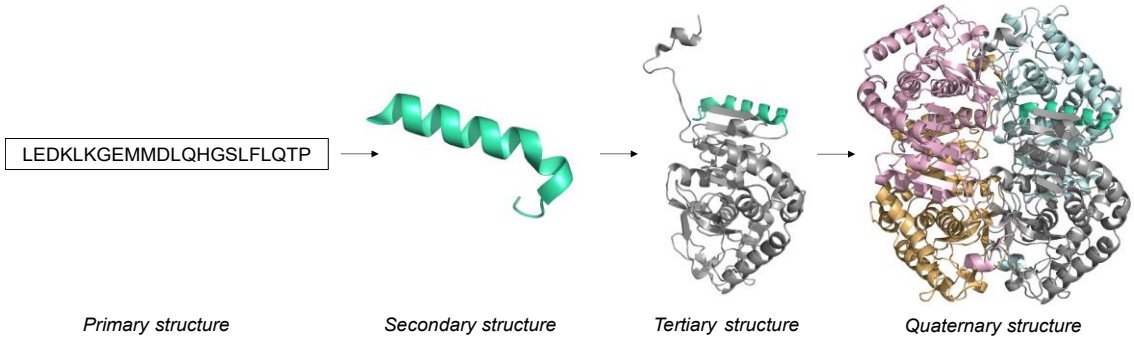


Figure I.1: Representation of the four structural levels of a protein: primary, secondary, tertiary, and quaternary structures

2. The human interactome: protein-protein interactions as a target in drug design

Because of their function, proteins are at the core of the cellular machinery. Through protein-protein interactions (PPIs), proteins have a specific biological function allowing them to regulate the cellular response/mechanisms. All these interactions define the

human interactome. The specific biological activity of proteins can be dysregulated in many diseases. Therefore, targeting the PPIs is an interesting strategy in drug design[6–8].

2.1. Targeting PPIs as a challenge?

Targeting PPIs may seem challenging compared to targeting the active site of a protein for several reasons, such as[9–13]:

- The nature of the protein-protein interface: the size of this interface is larger (between 1 000 $Å^2$ and 10 000 $Å^2$) compared to the size of the active site (between 100 $Å^2$ and 600 $Å^2$) that is usually targeted for drug development. Then, the surface tends to show some flatness, making it more difficult to highlight potential binding sites (cavities). In addition, these binding sites composed by residues involved in the protein-protein interaction are non-contiguous. Finally, the interface generally presents some lipophilic characteristics.
- The size of the available chemical space: due to the nature of the protein-protein interface, the molecules used to target this interface may not respect Lipinski's rule of five (molecular weight (MW) ≥ 400 Da and logP > 4 of a PPI modulator compared to MW between 200 and 500 Da and LogP < 5 of a druggable molecule). Therefore, this makes it more difficult to identify PPI modulators from the available chemical libraries (mainly composed of druggable molecules).

Nevertheless, to overcome these constraints, the development of computational tools allows the highlighting of binding pockets and the assessment of the hotspots. Hotspots are defined as residues involved in the interactions of the protein-protein interface, and thus in the stabilization of the protein complex. These hotspots can be identified and evaluated from an energetic point of view by Ala-scan. This constitutes a good starting point in the design of PPI modulators[10,12,14]. In addition, different types of modulators exist to target PPIs as described in the next section.

2.2. PPI modulators

Three classes of modulators can be used to target PPIs, such as small molecules, protein modulators, and peptides/peptidomimetics[12,15]. Some examples (non-exhaustive list) of PPI modulators (in clinical trials or launched on the market) developed as cancer treatment are presented for each class of modulators.

Small molecules can be used to target PPIs. Nevertheless, their use is best suited for smaller and more compact protein-protein interfaces, where hotspots are close to each

other. The molecules selected as modulators have generally some different characteristics (MW ≥ 400 Da, LogP > 4) than the usual druggable molecules used to target a protein active site. Moreover, small molecule modulators can be designed by fragment-based screening (FBS)[9,11,12,14]. Affinity and selectivity can be potentially increased by designing targeted covalent inhibitors[16–19]. An example of a small molecule modulator is Venetoclax (Venclexta® and Venclyxto®) inhibiting the binding of prosurvival protein Bcl-2 to the proapoptotic Bcl-2-Associated X protein (Bcl-2/BAX complex) to restore the proapoptotic activity of the protein in cancer cells[20–22]. A second example is Paclitaxel (Taxol®, used in cancer chemotherapy) which acts as a microtubule-stabilizing drug, by binding and stabilizing the β-tubulin filament[7,23].

When protein-protein interfaces are larger, peptide/peptidomimetic modulators can serve as an alternative to small molecules. Indeed, they present a high selectivity due to their large surface of interaction. In addition, peptides present other advantages such as lower toxicity and ease of synthesis[12,24,25]. However, their lack of permeability to cell membranes, as well as their sensitivity to proteases *in vivo* could be problematic. Nevertheless, some chemical modifications on side chains, e.g. the insertion of D-amino acids in peptides or the use of cyclized and stapled peptides allow overcoming these problems[12]. Peptides can also be conjugated with cell-penetrating peptides[26,27]. Moreover, they can be used as starting materials for the design of peptidomimetic and small molecule modulators[12]. For example, the stapled α-helix peptide (ALRN-6924) is able to target the protein-protein interface of the protein complex MDM2/p53 or MDMX/p53 by inhibition of the Mouse Double Minute 2 or X (MDM2 or MDMX). Therefore, the p53-dependent tumor suppressor activity is recovered[20,28,29].

The last approach includes different types of proteins, such as monoclonal antibodies, nanobodies (found in camelids) or synthetic (artificial) antibodies. They can be used to target a PPI by competing with the protein binder or by binding to an allosteric site of the protein. In general, antibodies are used for targeting extracellular proteins (e.g. receptors) due to their poor cell barrier permeability[12,15]. An example of a protein modulator used as a drug is the monoclonal antibody Cetuximab (Erbitux®, used in combination with cancer chemotherapy) directed against the Epidermal Growth Factor Receptor (EGFR), i.e., a tyrosine kinase receptor involved in the growth of cancer cells. Cetuximab interacts with the soluble extracellular region of EGFR avoiding receptor dimerization (mandatory for proper cell signal transduction) and preventing the binding of ligands (such as

Epidermal Growth Factor (EGF))[7,30,31]. Another example is the monoclonal antibody Durvalumab (Imfinzi®, used in cancer treatment such as non-small-cell lung cancer) directed against Programmed cell Death-Ligand 1 (PD-L1). The binding of PD-L1 to Programmed cell Death-1 (PD-1) receptor avoids the T-cell function[20,32,33].

2.3. Mode of action of modulators and focus on mechanisms of homomeric disruption

To target the PPIs, different types of modulators exist and are classified according to their mode of action: the PPI disruptors and stabilizers that induce the destabilization or the stabilization of the PPI, respectively. Moreover, each of these modulator groups can be divided into orthosteric and allosteric PPI modulators (Fig. I.2)[12,14,20,34].

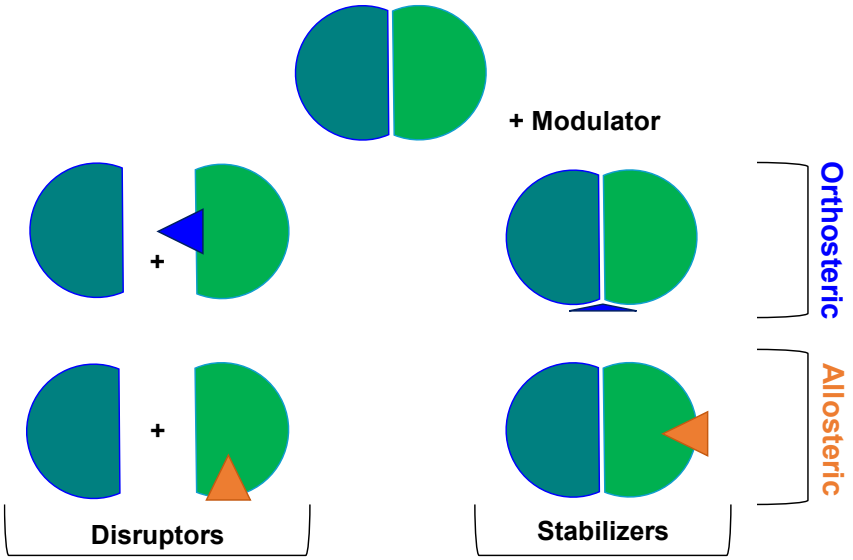


Figure I.2: PPI modulators – Orthosteric (in blue) or allosteric (in orange) PPI modulators leading to the PPI stabilization or destabilization. Figure adapted from Fischer *et al.*[34]

The formation of the protein complex via a PPI can be destabilized in presence of PPI disruptors. The allosteric disruptors bind to an allosteric site of the protein, i.e., they do not interfere with the protein-protein interface itself. This binding causes a change of affinity (via a conformational change of the protein), preventing the proteins to form a complex via establishing the protein-protein interactions. Moreover, orthosteric disruptors, have the same function as allosteric disruptors, but follow a different mechanism of action. These modulators bind at the protein-protein interface, competing directly with the protein binder. These mechanisms can be also used for stabilizing the PPIs with allosteric and orthosteric stabilizers[12,14,20,34].

PPIs can be established between two identical proteins (identical primary structure) to form homo-oligomers, but also between two different proteins (different primary structure), forming hetero-oligomers[8]. In this work, we will focus on PPIs in homo-oligomers allowing the protein homo-association. The mechanisms of homomeric disruption by modulators are presented in the next section[7]. More details and examples of hetero-oligomer systems can be found in the scientific literature or in the review of Lu *et al.*[20].

Different mechanisms of homomeric disruption are presented in Figure I.3 and described below[7]:

- *Direct homo-oligomeric disruption* (A): the mechanism corresponds to the intercalation of a modulator at the protein-protein interface of the complex. This leads to the dissociation of the homo-oligomer into several subunits.
- *Subunit intercalation* (B): as for the previous mechanism, it consists in the intercalation of a modulator at the protein-protein interface of the complex. Nevertheless, this does not lead to the dissociation of the homo-oligomer, but to a conformational change of the protein, resulting in a loss of function of the complex.
- *Capture of an inactive monomeric conformation* (C): this corresponds to the interaction of the modulator with a monomer. This interaction blocks the monomer into an inactive conformation, preventing the protein homo-association. The mechanism is possible when an important concentration of monomer is available in the solution.
- *Promotion of inactive multimeric complexes* (D): it consists in the interaction of a modulator with an allosteric site of the protein, leading to the stabilization of the protein into an inactive multimeric conformation.

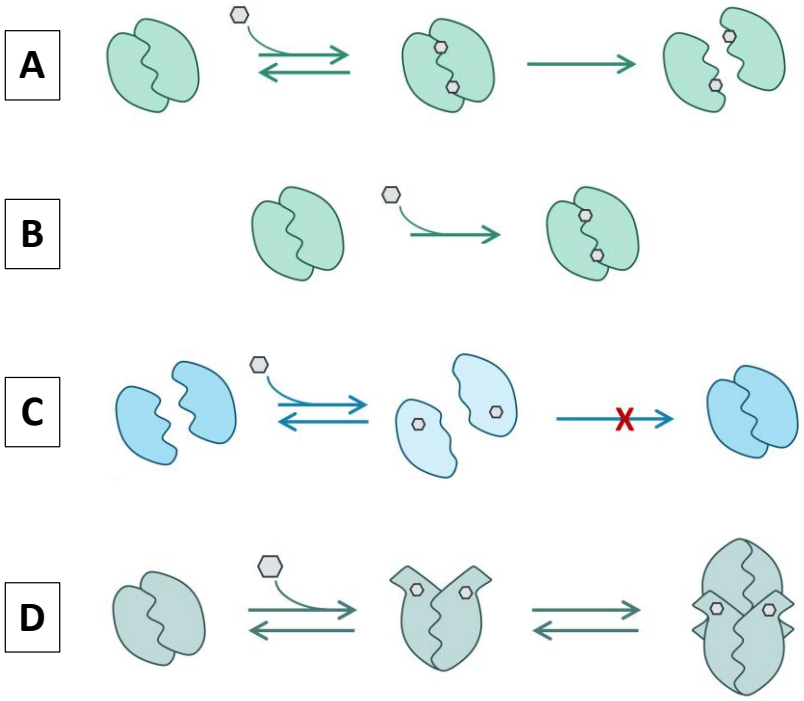


Figure I.3: Mechanisms of homomeric disruption – (A) Representation of the direct homo-oligomeric disruption mechanism – (B) Representation of the subunit intercalation mechanism – (C) Representation of the capture of an inactive monomeric conformation mechanism – (D) Representation of the promotion of inactive multimeric complexes mechanism – Figure adapted from Thabault *et al.*[7]

2.4. Techniques and focus on X-ray crystallography

Unlike inhibitors targeting the active site of proteins, the homomeric disruptors cannot be identified by conventional High Throughput Screens (HTS), such as via catalytic activity assays. It is necessary to adapt the techniques used to monitor the oligomerization states, and in particular in this work for identifying modulators targeting protein self-assembly[7,12]. The adapted methods include[7,11,12]: analytical centrifugation, thermal shift assay, Föster Resonance Energy Transfer (FRET), Enzyme-Linked ImmunoSorbent Assay (ELISA), Intrinsic Tryptophan Fluorescence (ITF), native Mass Spectrometry (MS), PolyAcrylamide Gel Electrophoresis (PAGE) and Size Exclusion Chromatography (SEC). Some of these methods may be subject to false positives, due to protein aggregation or allosteric binding. Therefore, other techniques, such as IsoThermal Calorimetry (ITC), Nuclear Magnetic Resonance (NMR) and X-ray crystallography, can be used to validate and characterize the modulators[12]. X-ray crystallography is one of the main techniques used in this work to provide structural data about protein-disruptor complexes.

The use of crystallography provides a three-dimensional structure of the studied protein-ligand complex[7,11,12]. Different methodologies can be used to obtain the complexes, such as co-crystallization and soaking technique. Co-crystallization consists in incubating the ligand with the purified protein, usually at 4°C (the incubation time depends on the system). Crystallization assays are then performed using this protein-ligand solution. The soaking uses a protein crystal (without studied ligand) transferred in a crystallization solution containing the desired ligand to let it diffuse in the crystal lattice[35,36].

When many ligands must be tested, the soaking technique is more adapted to obtain the complex structure. However, for ligands causing a large conformational change of the protein, such as disruptors, co-crystallization is more appropriate[36]. A large conformational change could lead to a weakening of the crystal while soaking. In addition, soaking requires an accessible binding site for the ligand, and this last must not be too large to diffuse within the crystal[35,36].

3. Cell metabolism and biological function of LDH

As mentioned before, targeting PPIs is an interesting approach in drug design. This approach constitutes an emerging and promising strategy in the development of new cancer therapies. Indeed, many PPIs are involved in the dysregulation of the oncogenic pathways, which can promote tumorigenesis, tumor progression, invasion, or metastasis[9,37,38]. It is in this context that we study L-Lactate Dehydrogenase, a tetrameric enzyme involved in the energy metabolism of normal and cancer cells and identified as a promising therapeutic target in cancer treatment. In the next sections, the metabolism of normal and cancer cells, as well as the biological function of LDH are presented.

3.1. Normal cells

In aerobic conditions, normal cells metabolize glucose via the cell respiration pathway (Fig. I.4). Glucose enters the cell via the GLUT-1 transporter and then is converted into pyruvate through the glycolysis pathway. Pyruvate is decarboxylated into acetyl-CoA by the Pyruvate Dehydrogenase (PDH) in mitochondria. Then, the acetyl-CoA is metabolized through the Krebs cycle, also called the tricarboxylic acid cycle (TCA cycle). The oxidation of reduced NADH and $FADH_2$ coenzymes, generated during the TCA cycle, by the electron transport chain (ETC) allows the phosphorylation of ADP into ATP molecules via the oxidative phosphorylation (OXPHOS)[5,39,40]. In the end, 38 molecules of ATP, six molecules of carbon dioxide (CO_2) and six molecules of water (H_2O) are produced per glucose molecule[41].

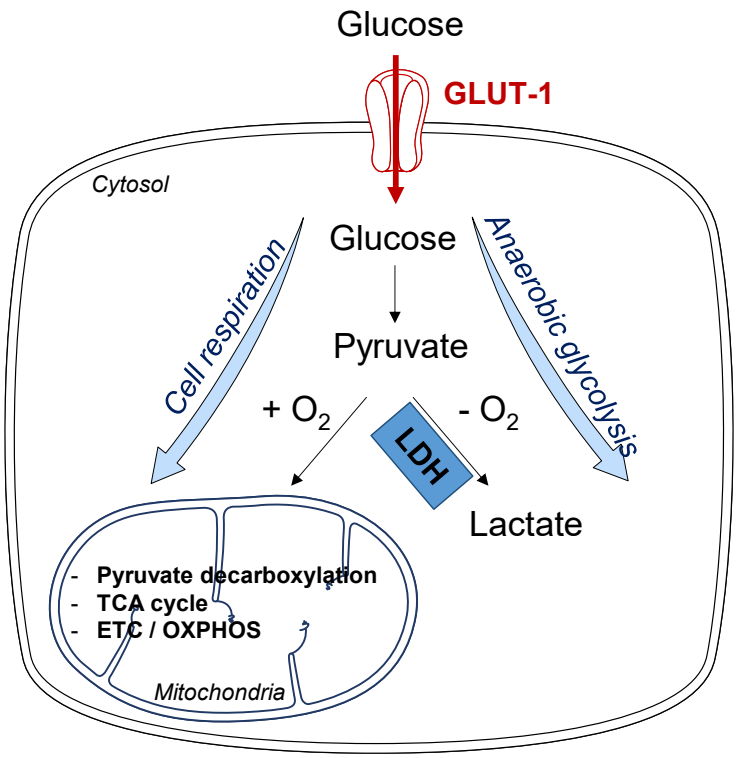


Figure I.4: Representation of glucose metabolism in normal cells in aerobic and anaerobic conditions. Figure adapted from Vander Heiden *et al.*[39]

Nevertheless, under anaerobic conditions, pyruvate produced through the glycolysis pathway is reduced to L-lactate by the lactate dehydrogenase in presence of NADH cofactor (Fig. I.4)[5,39,40]. This pathway is called anaerobic glycolysis and two molecules of ATP, two molecules of lactate and two molecules of H_2O are produced per glucose molecule[41]. The main role of the LDH enzyme is to reduce pyruvate into lactate. This reaction will be accompanied by the reoxidation of the cofactor NADH into NAD^+ (see section 4.2 for details), essential to maintain the glycolytic flux. Moreover, the lactate produced through the anaerobic glycolysis can be transported out of the cell (via monocarboxylate transporters (MCT)) and used as a source of energy for parenchymal cells or as a precursor for the gluconeogenesis[42].

The cytosolic LDH enzyme is involved in glucose metabolism through the anaerobic glycolysis pathway as previously described. Nevertheless, the enzyme can also have other biological functions while located in the peroxisome, the mitochondria or the nuclei of the cell[42]. One of the functions of the peroxisome is to degrade the long chains of fatty acids (carbon atoms > 22) via the β-oxidation reaction. This reaction will be accompanied by the reduction of NAD^+ into NADH. Therefore, to maintain the redox balance in the peroxisome, NADH is oxidized again to NAD^+ with LDH enzyme[42–44]. Regarding LDH

function in mitochondria, the lactate produced through the anaerobic glycolysis can be reoxidized into pyruvate to be used as a source of energy or as a precursor for gluconeogenesis[42]. Finally, it has been shown that LDH located in the nucleus plays a role in the regulation of gene transcription, such as the histone 2B (H2B). Indeed, LDH is in complex with glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and allows to maintain an optimal ratio NAD^+/NADH for the gene transcription[42,45].

3.2. Cancer cells

Cancer cells are cells that have undergone several mutations and have acquired specific characteristics, called Hallmarks, such as unregulated proliferation[46,47]. Reprogramming energy metabolism (dysregulation of glucose metabolism) is a well-known Hallmark of cancer. Cancer cells show an increase in glucose uptake compared to normal cells (Fig. I.5)[39,48,49]. Indeed, the high glycolytic flux reflects their adaptation to a hypoxic environment via the anaerobic glycolysis (as observed for normal cells – Fig. I.4), or to their unregulated proliferation via the aerobic glycolysis[50].

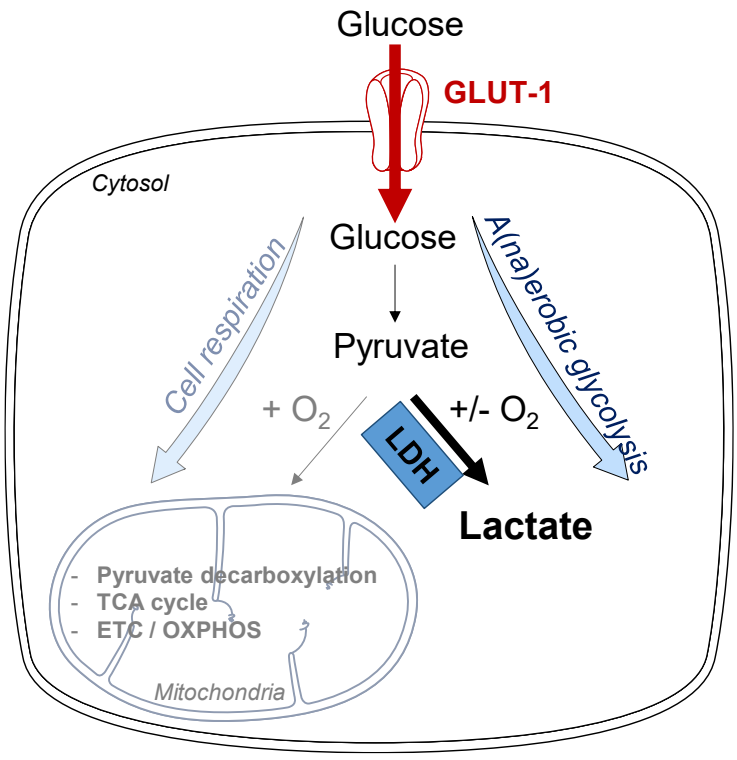


Figure I.5: Representation of glucose metabolism in cancer cells in aerobic and anaerobic conditions. Figure adapted from Vander Heiden *et al.*[39]

Thus, cancer cells adapt their energy metabolism via the reduction reaction of pyruvate to lactate, regardless of the oxygen supply in the cell (the mitochondria remain functional and a small percentage (~5%) of pyruvate is metabolized through cell respiration). This

phenomenon is called the Warburg effect[39,48,49]. The energy yield is two ATP molecules formed per molecule of glucose consumed[41]. Therefore, the high glycolytic flux (via the increase of glucose uptake in the cell) compensates for the low ATP yield (compared to the respiration pathway). This provides the energy and the biomass necessary to maintain the anabolic growth of cancer cells[39,48–50]. Moreover, the lactate product of the glycolysis pathway obtained through the reduction of pyruvate with LDH is considered as an essential product for the anabolic growth and proliferation of cancer cells[50]. This promotes several phenomena such as metabolic cooperation[51–53], autophagy[52,54], sustaining of intracellular redox equilibrium[42,55], angiogenesis[56,57], invasiveness and metastasis[53,58].

a) Metabolic cooperation and autophagy

Within the same tumor, a fluctuation in the availability of resources (O_2 and glucose) exists. Therefore, cancer cells show adaptability to substrates depending on the available resources. The cells close to the blood vessels, called oxidative cancer cells, are in normoxic conditions and have more access to glucose compared to the cancer cells in hypoxic conditions (inside the tumor and far away from blood vessels), called glycolytic cancer cells. To compensate for the fluctuation of available resources, cancer cells can establish metabolic cooperation (Fig. I.6).

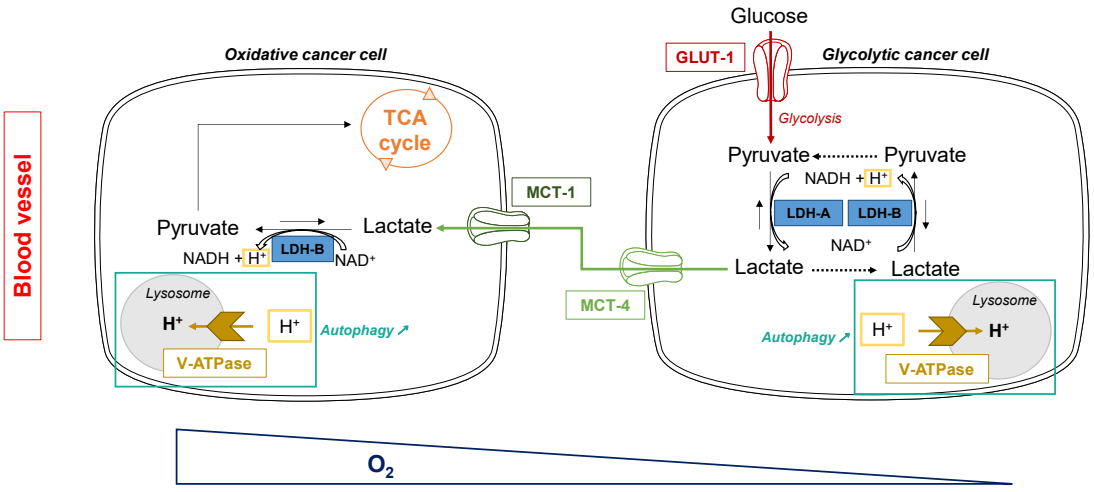


Figure I.6: Representation of the metabolic cooperation between oxidative and glycolytic cancer cells in the tumor and the V-ATPase-dependent acidification of lysosome-promoting autophagy. Figure adapted from Brisson *et al.*[52]

Although the oxidative cancer cells can use glucose to generate ATP via cell respiration, those cells will not promote this metabolic pathway. Therefore, the available glucose diffuses within the tumor and is taken up by glycolytic cancer cells via the GLUT-1

transporter. Glucose is converted to pyruvate which is reduced to lactate via anaerobic glycolysis. This reaction is catalyzed by the LDH-A enzyme in presence of the cofactor NADH. Lactate is then transported via monocarboxylate transporters (MCT-4 and MCT-1) into the oxidative cancer cells and oxidized in pyruvate. This reaction is catalyzed by the LDH-B enzyme in presence of the cofactor NAD^+. Then, the pyruvate will be used in the TCA cycle to generate ATP. In short, the oxidative cancer cells recycle the lactate generated by the glycolytic cancer cells. The uptake of lactate into oxidative cells is possible via the MCTs transporters. Finally, the oxidation of lactate into pyruvate is catalyzed by the presence of LDH-B enzyme[51,52].

Another strategy used by cancer cells to cope with the fluctuation of available resources is to promote autophagy (Fig. I.6). This phenomenon aims at degrading various damaged proteins and organelles, in order to provide the biosynthetic precursors necessary for the anabolic growth of cancer cells and thus promote cancer cell growth[52]. Indeed, the oxidation of lactate in pyruvate in glycolytic and oxidative cancer cells is accompanied by the reduction of NAD^+ cofactor in NADH + H^+. The protons generated by the reduction of the cofactor allow the acidification of lysosomes via the V-ATPase (a proton pump of lysosomes), promoting autophagy[52,54].

b) Sustaining of intracellular redox equilibrium

Different signaling pathways are NAD^+ dependent that can regulate, among others, the cancer cell metabolism. Therefore, due to the constant degradation of NAD^+ within the different signaling reactions, it is necessary to resynthesize NAD^+ to maintain the NAD^+/NADH ratio within the cancer cell and allow its proliferation. The increased lactate production through the anaerobic glycolysis pathway induces the regeneration of NAD^+ via the reduction of pyruvate in lactate accompanied by the reoxidation of NADH in NAD^+. It allows sustaining the intracellular redox equilibrium necessary for the functioning of the cell[42,55,59,60].

c) Angiogenesis

Again, to cope with the heterogeneous distribution of nutrients within the tumor, cancer cells promote angiogenesis (Fig. I.7). This corresponds to a phenomenon that aims to expand the vascular network in order to increase the supply of oxygen and nutrients to hypoxic tumor cells[56].

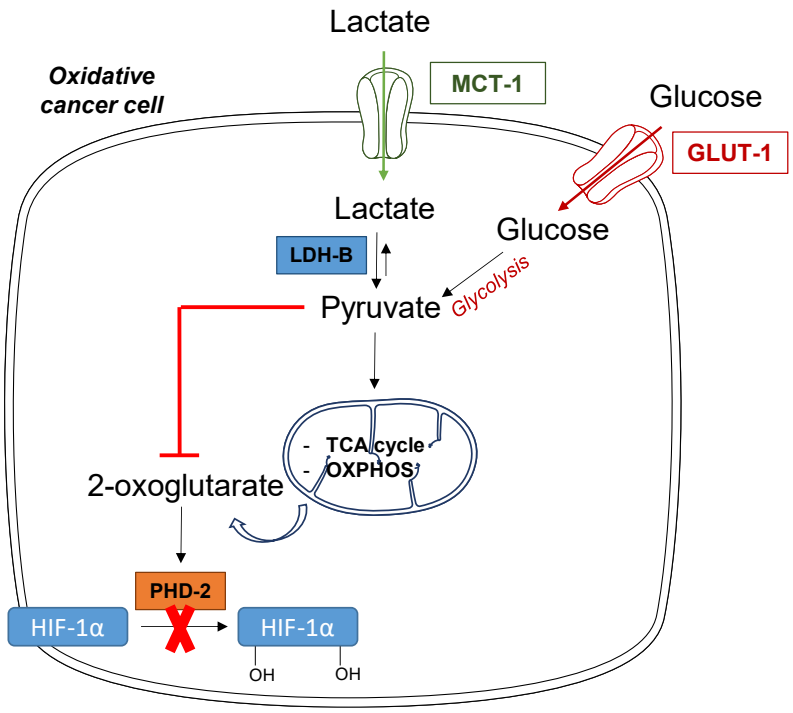


Figure I.7: Representation of angiogenesis phenomenon in oxidative cancer cells. Figure adapted from Sonveaux *et al.*[56]

The main regulator of angiogenesis is the Hypoxia-Inducible Factor-1 (HIF-1) corresponding to a heterodimer constituted by two subunits HIF-1α and HIF-1β (constitutive nuclear protein)[56,61]. In normoxic endothelial cells, HIF-1α is hydroxylated by the Prolyl Hydroxylase-2 (PHD-2) in presence of the 2-oxoglutarate substrate. Hydroxylated HIF-1α is inactive and will be degraded by the proteasome.

HIF-1 is a transcription factor inducible by hypoxia. Nevertheless, in oxidative cancer cells (normoxic conditions), its activity can be stimulated by exogenous lactate (produced through anaerobic glycolysis of glycolytic cancer cells – see Fig. I.6). Lactate is oxidized in pyruvate through the reaction catalyzed by LDH-B. Therefore, the concentration of pyruvate in oxidative cancer cells increases and the pyruvate competes with the 2-oxoglutarate (substrate of PHD-2). The hydroxylation reaction of HIF-1α is inhibited and HIF-1α remains in its active form. At this stage, HIF-1α binds to HIF-1β and other cofactors in the nucleus. The formed multiprotein complex initiates the transcription of different genes, such as the Vascular Endothelium Growth Factor (VEGF), promoting angiogenesis[56,57].

d) Invasiveness and metastasis

One of the characteristics of malignant tumors is their propensity for invasion and metastasis. This is promoted by the acidosis of the primary tumor, inducing in particular the degradation of the extracellular matrix (ECM)[53,58]. The extracellular pH (pH_e) of cancer cells is between 6.5 and 6.9, while pH_e of normal cells is between 7.2 and 7.4[58]. Therefore, the acidification of the tumor environment may be correlated with the abnormal metabolic activity of cancer cells (increased glycolytic flux), resulting in the lactic acid release via the MCT transporters[62,63] and proton extrusion[53].

4. Structural characterization of Lactate Dehydrogenase

4.1. Primary structure of LDH

Three main genes encode for proteins belonging to the L-lactate dehydrogenase (LDH) family in human: ldh-a, ldh-b, and ldh-c. Each of these genes encodes for one of the three different subunits: LDH-M, LDH-H and LDH-X subunit, respectively[42,64,65]. Their primary structure is composed of 332 amino acids (for LDH-M (Uniprot entry P00338) and LDH-X (Uniprot entry P07864)) and 334 amino acids (for LDH-H (Uniprot entry P07195)). Moreover, the three LDH forms have a high percentage of sequence identity and similarity. The values were calculated using Emboss Needle[66] and presented in Table I.1.

Table I.1: Percentage of sequence identity and similarity between the three LDH forms (LDH-M, LDH-H and LDH-X). Values are calculated with Emboss Needle[66] (BLOSUM62 similarity matrix) from the LDH subunit amino acid sequences. The lower value corresponds to the percentage of sequence identity and the value in brackets to the percentage of sequence similarity

	LDH-M	LDH-H	LDH-X
LDH-M	/	75.1% (88.6%)	75.3% (88.9%)
LDH-H	/	/	69.2% (83.2%)

The ldh-c is a testis-specific gene and encodes for LDH-X subunit that is expressed in testis[64,65]. We will mostly focus on the two main LDH isoforms (LDH-A and LDH-B) constituted by LDH-M and LDH-H subunits, respectively (see section 4.3).

4.2. Secondary and tertiary structures of LDH

The primary structure of the different LDH forms determines their tertiary structure, i.e., the 3D arrangement of the full protein containing different secondary structure elements[5]. LDH is a protein that has been studied for over 50 years. While the first complete

structure of LDH dates from 1970[67], the first crystallographic structures of both main human isoforms were solved in 2001 (Protein Data Bank (PDB) entries: 1I01 and 1I0Z)[65]. As of today, 56 crystallographic structures of both human LDH isoforms have been deposited in the PDB (52 and four structures for LDH-A and LDH-B, respectively). Table I.2 describes the crystallographic structures of the human LDH-B isoform which is the subject of this work.

Table I.2: Description of hLDH-B crystallographic structures available in the Protein Data Bank

PDB entry	Description	Resolution (Å)
1I0Z[65] (2001)	Structure of hLDH-B in complex with NADH and oxamate	2.1
1T2F[68] (2004)	Structure of hLDH-B in complex with NAD^+ and 4-hydroxy-1,2,5-oxadiazole-3-carboxylic acid	3.0
7DBJ[69] (2021)	Structure of hLDH-B in complex with NADH, oxamate and AXKO-0046	1.5
7DBK[69] (2021)	Structure of hLDH-B in complex with NADH	1.8

All the data obtained from the crystallographic structures of LDH allowed to further characterize the protein from a structural point of view. LDH subunits are organized in two domains: the first domain is characterized by a "Rossmann"-type fold and the second is the mixed α/β "substrate binding" domain (Fig. I.8A). The "Rossmann"-type fold is a common structural motif found in a large proportion of protein (~20% in the protein structures from PDB). This first domain is composed of a 6-stranded parallel β-sheet flanked by 3 helices on each site (Fig. I.8A-B in green). One of its main structural features is the cavity formed at the junction of the β-strands 3 and 4, allowing the binding of nucleotides such as NAD(P) or FAD cofactor[70,71].

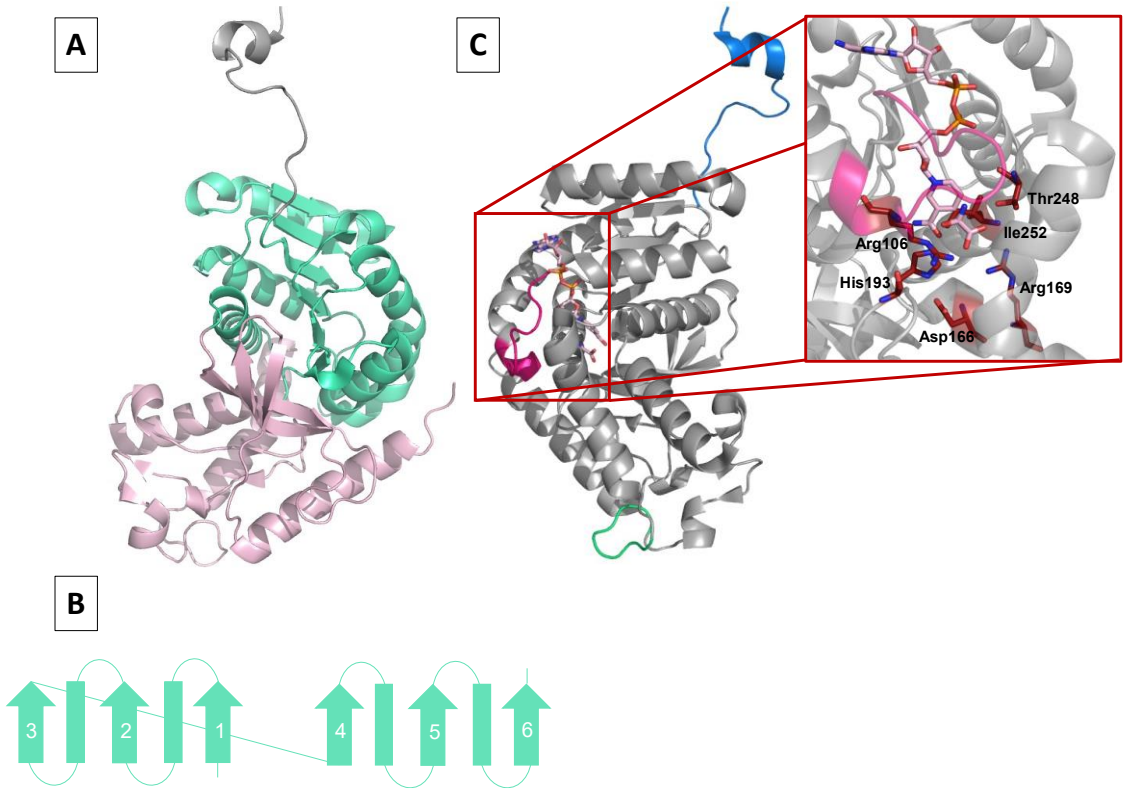


Figure I.8: Representation of LDH tertiary structure. The model used for the representation is the monomer A (reference monomer) from LDH-B structure (PDB entry: 1I0Z) – (A) Representation of the two domains composing LDH subunit: the "Rossmann"-type fold domain in green and the mixed α/β "substrate binding" domain in light pink – (B) Schematic representation of the "Rossmann"-type fold domain (C) Highlighting of different secondary structure elements in LDH subunit: tetramerization arm in blue, antigenic loop in green, active site loop in pink. Zoom on active site (red box): the substrate analog (OXM) and cofactor (NADH) in the active site are represented in light pink sticks and the essential residues constituting the active site are in red sticks

The second domain corresponds to 4 β-strands and 3 α-helices (Fig. I.8A in light pink)[65]. The interface between this domain and the "Rossmann"-type fold domain provides a binding site for the substrate. These two binding sites constitute the active site of LDH and its open/closed conformation is mediated by the active site loop (defined as a flexible part of the protein) (Fig. I.8C in pink)[65]. L-lactate dehydrogenase belongs to the 2-hydroxy acid oxidoreductase family and catalyzes the interconversion of pyruvate to L-lactate in presence of the reduced NADH cofactor at the end of the glycolytic pathway (Fig. I.9)

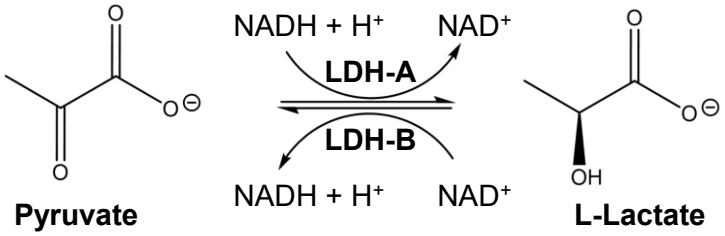


Figure I.9: Interconversion reaction of pyruvate to L-lactate in presence of NAD cofactor catalyzed by LDH enzyme

The active site is defined by essential residues including Asp166, Arg169, His193, Thr248 and Ile252 (Fig. I.8C – red box). The reaction occurs within the catalytic site of the protein following an ordered mechanism: the cofactor binds first the end of a central β-sheet of the "Rossmann"-type fold domain. Then, the substrate can bind to the substrate binding site located at the interface of the two protein domains. Finally, once the substrate is bound, it can interact via a hydrogen bond with Arg106 (part of the active site loop). This interaction leads to the closing of the active site loop. Therefore, in absence of the cofactor and/or the substrate, the active site loop will adopt an open conformation[42,65,69,72].

Another flexible part is the antigenic loop (Fig. I.8C in green). The loop has been identified as a region with large variability in the amino acid composition between the different forms of LDH[65,73]. Moreover, the 19 N-terminal amino acids are defined as the tetramerization arm due to their involvement in the assembly of the LDH quaternary structure (Fig. I.8C in blue).

4.3. Quaternary structure

The LDH protein is catalytically active when it assembles as a tetramer[64]. Indeed, the quaternary structure of LDH has a significant impact on the structure of the active site, by maintaining its geometry and its hydrogen bonds network[72,74]. The tetrameric oligomerization state prevents water from entering the active site (hydrogen bond network and electrostatic environment are not affected) and also provides rigidity to an adjacent α-helix[72].

The LDH protein is assembled into a tetramer thus constituted by four subunits. Using M and H subunits as building blocks, different oligomers can be formed: three hetero-oligomers and two homo-oligomers can be obtained via the combination of M and H subunits (Fig. I.10). The three hetero-oligomers MH_3, M_2H_2 and M_3H are called LDH-2, LDH-3 and LDH-4, respectively. The two homo-oligomers H_4 and M_4 are called LDH-1 or LDH-B and LDH-5 or LDH-A, respectively. The LDH-B isoform is found mainly in

aerobic tissues (e.g. cardiac muscle), while the LDH-A isoform is found mainly in anaerobic tissues (e.g. skeletal muscle, liver)[64,65].

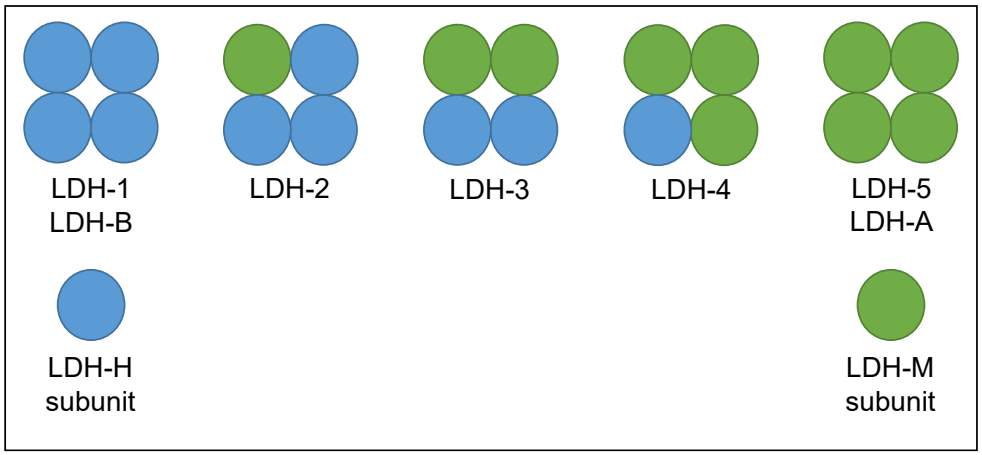


Figure I.10: Representation of the LDH tetrameric oligomers: three hetero-oligomers (LDH-2, LDH-3 and LDH-4) and two homo-oligomers (LDH-1 and LDH-5) are obtained via the combination of LDH-H and LDH-M subunits. Figure adapted from Valvona *et al.*[64]

The two main isoforms of LDH, LDH-A and LDH-B, assemble into homo-tetramers. These ones correspond to the association of two dimers into a tetramer (dimer of dimer)[75–77]. The works of Thabault *et al.*[50] and Jafary *et al.*[78] suggest that the tetramer is formed by the association of the dimers AC and BD (see Fig. I.12A). The assembly of LDH proteins as tetramers is mediated by the presence of the tetramerization arm, corresponding to the 19 N-terminal amino acids. Indeed, the tetramerization arm of each monomer interacts and wraps around the two adjacent monomers[59,65]. In addition, there is a 22-amino acid α-helix ended by a short loop identified as essential for the tetramer stabilization of LDH proteins[79]. These two structure elements will be detailed in section 5.2. Therefore, the tetrameric assembly of the LDH protein has a molecular weight of about 140 kDa.

5. LDH as a therapeutic target for cancer

LDH is an enzyme at the core of lactate metabolism, essential for the anabolic growth and the proliferation of cancer cells. Based on this evidence, LDH was highlighted as an interesting therapeutic target for cancer treatment. Therefore, different strategies can be used to target LDH protein, such as targeting its active site or its oligomeric interfaces.

5.1. Strategies to target LDH protein

a) Active site

LDH has been identified as a key enzyme in the glycolytic pathway with the conversion of pyruvate into lactate, preferentially catalyzed by LDH-A. Due to the implication of this isoform in the metabolic switch promoting cancer cell survival and tumor growth, studies were initially interested in further targeting LDH-A[80–84]. To inhibit its catalytic activity, many selective inhibitors of LDH-A targeting its active site were designed, such as the GNE140 compound (Genentech company)[83,85,94,95,86–93]. Nevertheless, the involvement of LDH-B has also been demonstrated in cancer pathogenesis as described before. LDH-B promotes lactate-fueled oxidative cancer cell respiration via the metabolic cooperation between glycolytic and oxidative cancer cells[51–53]. In addition, LDH-B controls lysosome activity and autophagy in cancer[52,54]. Therefore, the design of inhibitors targeting the two isoforms of LDH seems interesting[96]. Several non-specific inhibitors to LDH-A, such as Galloflavin, are also described in the literature[80,97–99].

Nevertheless, the nature of the active site of LDH leads to some drawbacks[42,50]. First, there exists a loss of selectivity for the LDH enzyme with other NAD-dependant enzymes. Indeed, as described in section 4.2, LDH protein is composed of two domains: the mixed α/β "substrate binding" domain and the "Rossmann"-type fold domain where the cofactor links[65]. Because of the close proximity of the two binding sites, inhibitors competing with the active site (via the cofactor and/or substrate binding sites) may decrease their specificity toward the LDH enzyme[42,50]. Secondly, LDH active site has a high polar amino acid content that can be explained by the large solvent exposure. Therefore, inhibitors designed to target the active site have challenging ADME (Absorption, Distribution, Metabolization and Excretion) properties[42,50]. Finally, the concentration of LDH in cancer cells is high (in the micromolar range), preventing the complete inhibition of the protein, even in the presence of the most potent inhibitors[50]. To overcome these drawbacks, another strategy can be considered. It consists of targeting allosteric sites, such as the oligomeric interfaces of LDH.

b) Oligomeric interface

As mentioned before, LDH-A and LDH-B are homo-tetramers formed by the assembly of two dimers (dimer of dimer) (Fig. I.12A)[75–77]. Therefore, due to the difficulties that can be encountered when designing selective and potent inhibitors targeting the active site of the protein, another strategy can be used[50,59]. This consists in targeting an allosteric

site, such as the protein-protein interactions (PPIs) at the dimer-dimer interface (defined as the tetrameric interface) of the LDH enzyme, with the aim of disrupting its active tetrameric state. The PPIs at the monomer-monomer interface (forming a dimer) are defined as the dimeric interface.

Currently, several reports have highlighted new allosteric inhibitors of LDH protein. The team of Shibata *et al.* has identified a small molecule (AXKO-0046) as a selective inhibitor of LDH-B (EC_{50} = 42 nM) by targeting the dimeric interface (Fig. I.11A). They have suggested an uncompetitive inhibitory mechanism[69]. Moreover, three other works described new allosteric inhibitors of LDH proteins, by targeting PPIs at the tetrameric interface. The team of Friberg *et al.* has identified selective inhibitors of LDH-A: phthalimide (IC_{50} = 308.0 nM) and dibenzofuran (IC_{50} = 757.0 nM) derivatives (Fig. I.11B)[100]. The team of Jafary *et al.* worked on the design of LDH-A linear peptide inhibitors that mimic the tetramerization arm, highlighting the ability of peptides to destabilize the LDH-A interacting subunit in solution[78]. In addition, the team of Nadal-Bufi *et al.* has also worked on the design of a LDH-A β-hairpin peptide inhibitor (cGmC9 with an IC_{50} = 2.5 µM) targeting a β-sheet region (Fig. I.11C – orange box) involved in PPIs at the tetrameric interface of LDH-A[24]. This peptide is designed from the gomesin antimicrobial peptide. It has a β-hairpin structure (two antiparallel β-strands joined by a β-turn that looks like a hairpin) with two disulfide bonds and is cyclized (Fig. I.11C)[24,101].

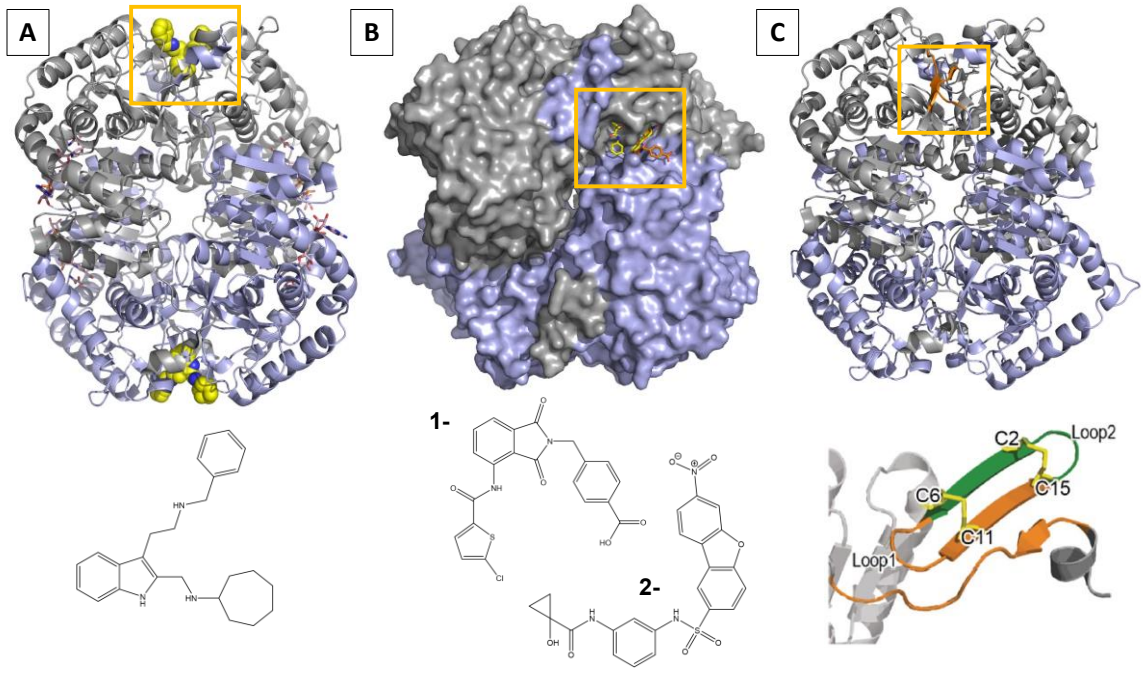


Figure I.11: (A) Crystallographic structure of LDH-B in complex with AXKO-0046 inhibitor (PDB entry: 7DBJ) and representation of the AXKO-0046 structure[42] – (B) Superposition of crystallographic structures of LDH-A in complex with phthalimide (PDB entry: 6SBU) and dibenzofuran (PDB entry: 6SBV) derivatives and representation of the phthalimide (1) and dibenzofuran (2) derivatives structures[89] – (C) Representation of the crystallographic structure of LDH-A (PDB entry: 1I10[65]) with the suggested binding site of the β-hairpin peptide (orange box) and representation of the cGmC9 β-hairpin peptide close to the tetramerization arm[24]

5.2. Project context

The project in which this work is included uses the same strategy of targeting PPIs at the tetrameric interface of LDH. For this purpose, two sites were already highlighted, by the UCLouvain collaborators, to be essential for the tetramerization of the protein[50,59]. The description of tetramerization sites will focus on the LDH-B isoform, which is the subject of this work (Fig. I.12).

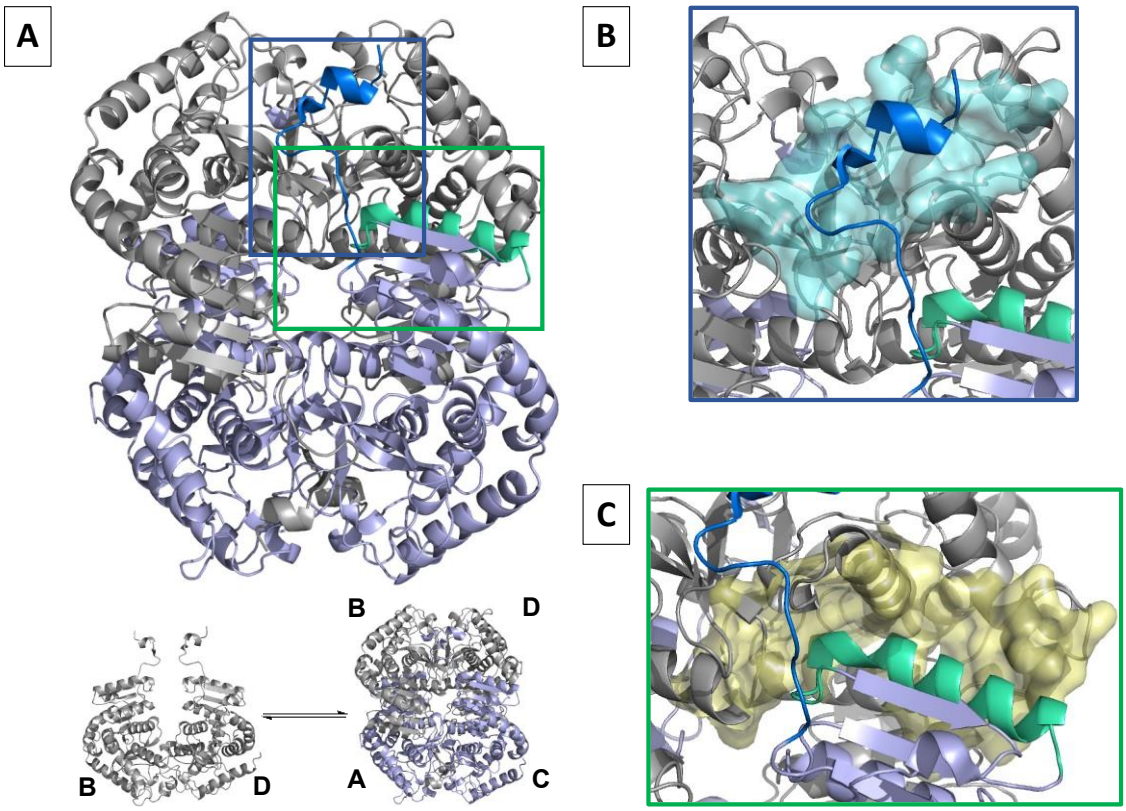


Figure I.12: (A) Representation of the two dimers (AC dimer in light purple and BD dimer in grey) composing the LDH-B tetramer (PDB entry: 1I0Z) and highlighting of tetramerization site 1 (blue box) and 2 (green box) – (B) Zoom on the tetramerization arm (in dark blue) interacting with the tetramerization site 1 (in light blue surface) – (C) Zoom on the 22-amino acid α-helix (in green) interacting with the tetramerization site 2 (in yellow surface)

- Tetramerization site 1

LDH-B tetramer assembly is mediated by the tetramerization arm, corresponding to the 19 N-terminal amino acids (residue Ala1 to residue Pro19 – ATLKEKLIAPVAEEEATVP). The N-terminal arm is formed by a short α-helix, followed by a β-strand and ends up with a loop. The tetramerization arm of each monomer interacts and wraps around the two adjacent monomers. It binds two different pockets on the two adjacent monomers, via (non)polar interactions[59,65]. This interacting site with the tetramerization arm is defined as the tetramerization site 1 in this work and is described in the work of Thabault *et al.*[59] (Fig. I.12B).

The work of Thabault *et al.*[59] aims to design stapled peptides mimicking the tetramerization arm to disrupt the oligomeric state of LDH. For this purpose, a truncated form of LDH-B (LDH-Btr), i.e., by deleting the 19 N-terminal amino acids corresponding to the tetramerization arm, has been produced and purified. This form was further characterized (Table I.3).

Table I.3: Characterization of the two LDH-B forms (LDH-Btr and LDH-B) via different properties - The Molecular Weight (MW) (determined by Size Exclusion Chromatography (SEC); the first value corresponds to the experimental MW, and the value in brackets is the theoretical MW), the affinity for NADH (by Microscale Thermophoresis (MST)), the catalytic properties (following the formation of NADH by fluorescence) and the thermal stability (melting temperature (T_m) determined by nano Differential Scanning Fluorimetry (nanoDSF))[59]

	MW (kDa)	K_{d-NADH} (µM)	Catalytic properties	T_m (°C)
	SEC	***MST***	***NADH fluorescence***	***nanoDSF***
LDH-Btr	78 (73)	28.8 ($CI_{95\%}$: 25.1 to 33.1)	K_M (NAD^+) = 0.587 ± 0.057 mM K_M (Lactate) = 33.77 ± 6.17 mM V_{max} = 0.0122 ± 0.003 µM.s^{-1}	58
LDH-B	130 (146)	23.2 ($CI_{95\%}$: 19.5 to 27.6)	K_M (NAD^+) = 0.208 ± 0.005 mM K_M (Lactate) = 1.686 ± 0.284 mM V_{max} = 0.190 ± 0.012 µM.s^{-1}	75

In contrast to the full-length form of LDH-B (LDH-B) which is a stable and active tetramer in solution, LDH-Btr is a well-folded dimer in solution with low catalytic activity and lower thermal stability in comparison with LDH-B (Table I.3). Indeed, the LDH-Btr protein was characterized by SEC and was in a native dimeric state (MW_{exp} = 78 kDa, MW_{th} = 73 kDa). The affinity of the truncated LDH-B for NADH was assessed by MST, highlighting a similar affinity of the two forms of LDH-B (LDH-B and LDH-Btr) for the cofactor. This suggests that the "Rossman"-type fold domain is well folded. In addition, the catalytic properties of LDH-Btr highlighted a low activity when compared with LDH-B protein (the Michaelis-Menten constant (K_M) is 3-fold and 20-fold larger for the NAD^+ cofactor and the substrate lactate, respectively and a large decrease of the maximal velocity (V_{max}) is also observed). Finally, the thermal stability of LDH-Btr was investigated by nanoDSF and showed a destabilization of the truncated form, with a ΔT_m of 17°C (corresponding to the difference between the T_m of LDH-B and LDH-Btr). The truncation of the tetramerization arm in LDH-Btr has allowed gaining accessibility to the tetramerization site 1 of the protein.

Based on the LDH-B crystallographic structure, the pharmacological properties of a first linear peptide (LB19), corresponding to the 19 N-terminal amino acids, were evaluated

on the LDH-Btr protein. The interaction between LB19 and LDH-Btr was characterized by MST and Nuclear Magnetic Resonance Water-Ligand Observed via Gradient Spectroscopy (NMR WaterLOGSY). These two biophysical methods highlighted that LB19 interacts with the LDH-Btr protein (K_d = 1.78 mM [1.14 – 2.77 mM]). Data from both spectra of NMR WaterLOGSY and two-dimensional Nuclear Magnetic Resonance Total Correlation Spectroscopy (NMR TOCSY) suggested that LB19 N-terminal residues are more involved in the interaction of the peptide with the LDH-Btr protein.

Starting from the LB19 linear peptide, other peptides have been designed, such as LB8, by trimming the residues of the C-terminal part while keeping those (from the N-terminal part) mainly involved in the interaction with the LDH-Btr protein. Therefore, the LB8 peptide is composed of the eight N-terminal residues of LB19 (ATLKEKLI), corresponding exactly to the α-helix of the tetramerization arm of LDH-B protein. LB8 also interacts with LDH-Btr (K_d = 1.05 mM [0.55 – 2.01 mM]). Thereafter, the structure-activity relationship (SAR) of the eight residues (LB8) interacting with the tetramerization site 1 was assessed by testing 12 structural analogs of LB8 (single amino acid substitutions). Furthermore, these data were correlated with the Ala-scan performed on the eight N-terminal residues of LDH-B protein. This highlighted that the lipophilic residues, i.e., Leu3, Leu7 and Ile8, appear to be key in the interaction between the LB8 peptide and the tetramerization site 1 via their aliphatic side chains which are oriented towards the hydrophobic cavities of the tetramerization site 1 (Fig. I.13).

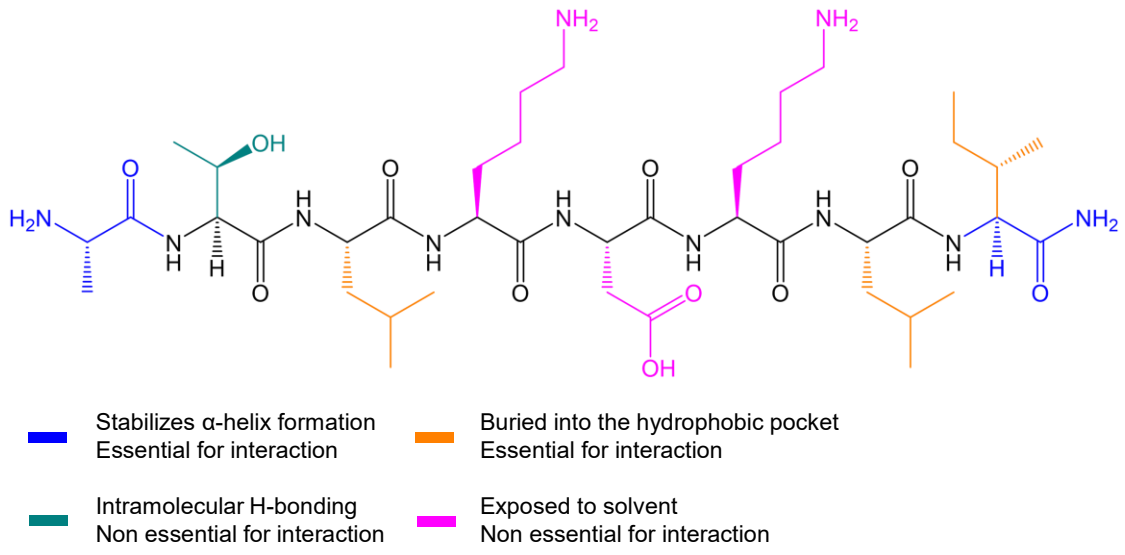


Figure I.13: Structure-activity relationship of LB8 interacting with the tetramerization site 1. Figure adapted from Thabault *et al.*[59]

By circular dichroism, LB8 presents a poor helical propensity which could be the origin of the low affinity of the peptide for LDH-Btr (K_d = 1.05 mM [0.55 – 2.01 mM]). Therefore, to stabilize its active helical conformation, the adopted strategy consisted of constraining the peptide via a macrocyclization approach using cysteine stapling with an α-helix-promoting agent. After optimization, the best cyclic peptide was macrocycle 7 constrained in positions i and i+4 with p-tetrafluorophenyl (Fig. I.14).

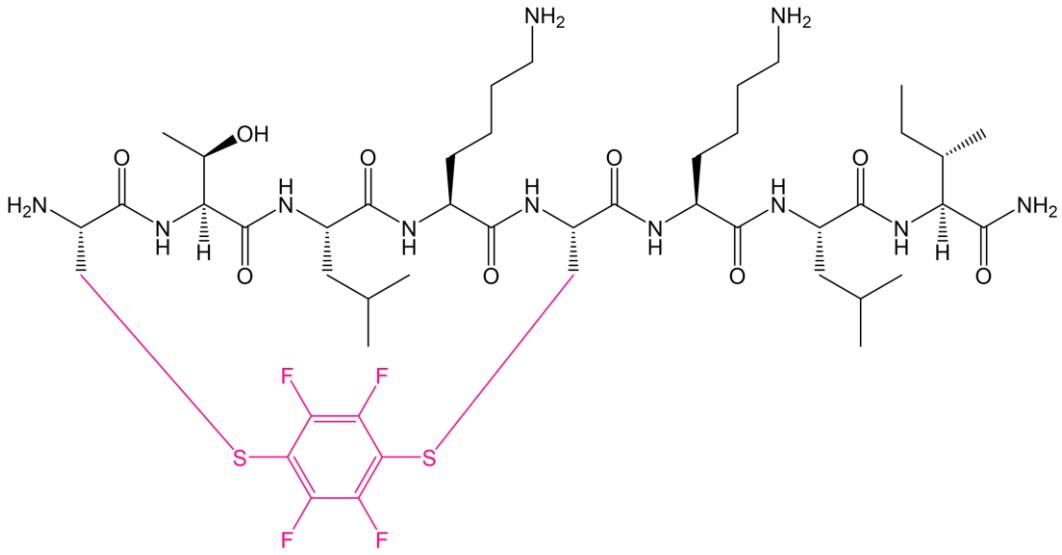


Figure I.14: Structure of macrocycle 7[59]

The designed macrocycle 7 can compete with the tetramerization arm of the two isoforms of LDH (LDH-A and LDH-B), resulting in the destabilization and the disruption of the protein complex in solution. Indeed, macrocycle 7 has a higher affinity for the LDH-Btr protein (K_d = 11 µM [$CI_{95\%}$ = 9 – 14 µM]). Furthermore, it also interacts with LDH-A and LDH-B at higher concentrations with a K_d = 117 µM [$CI_{95\%}$ = 94-144 µM] and K_d = 380 µM [$CI_{95\%}$ = 315 – 457 µM], respectively. The interaction of macrocycle 7 with both isoforms suggests a shift of the tetramerization arm through macrocycle 7 to reach the tetramerization site 1.

- Tetramerization site 2

The work of Thabault *et al.*[50] has revealed the existence of a second allosteric site that is defined as the tetramerization site 2 in this work. This site was found via the study of the LDH-Btr protein which is present as a tetramer at high concentration in solution (self-interaction of the LDH-Btr dimer protein: K_d = 1.25 µM [0.96 – 1.62 µM]), despite the truncation of the tetramerization arm previously described. Indeed, it was highlighted there exists an equilibrium between the dimeric and tetrameric forms thanks to the second tetramerization site.

Therefore, the stability of the tetrameric form of LDH-B is also mediated by the presence of the 22-amino acid α-helix ended by a short loop (residue Leu54 to residue Pro75 - LEDKLKGEMMDLQHGSLFLQTP) interacting with the second tetramerization site (Fig. I.12C)[50]. The tetramerization site 2 is defined by the following residues: N164, A168, R169, R171-L173, E176, H181-S183, S237, E240, V241, K243-K245, Y247-A251, D258, E261, K265-L267, and R269. Therefore, this second site will be divided into three parts in the following work:

- Tetramerization site 2 (part-1): N164, A168, R169, R171-L173, E176, H181-S183
- Tetramerization site 2 (part-2): S237, E240, V241, K243-K245, Y247-A251, D258
- Tetramerization site 2 (part-3): E261, K265-L267, R269

The work of Thabault *et al.*[50] aimed to further characterize the tetrameric interface of LDH protein through this new allosteric site (tetramerization site 2) using epitope mapping techniques and peptides. For this purpose, the pharmacological properties of the LP22 peptide, corresponding to the 22-amino acid α-helix interacting with the tetramerization site 2, were evaluated. The interaction between LP22 and LDH-Btr/LDH-B was characterized by MST and NMR WaterLOGSY. These two biophysical methods highlighted that LP22 does not interact with LDH-B, while it interacts with the LDH-Btr protein (K_d = 156 µM). Furthermore, the effect of LP22 on the thermal stability of LDH was assessed. In presence of LP22, a stabilization of dimeric LDH-Btr was observed (ΔT_m = 2.8°C at 500 µM of LP22), while the tetrameric LDH was destabilized in a concentration-dependent manner (EC_{50} = 47 µM [32 – 68 µM]). All these data allowed the team of R. Frédérick to highlight that LP22 interacts at the tetrameric interface of LDH and destabilizes the tetrameric protein form.

Then, they identified the essential part of LP22 responsible for the binding: the peptide GP16, corresponding to the 16 C-terminal amino acids (GEMMDLQHGSLFLQTP) when the six N-terminal of LP22 are trimmed (LEDKLK). The properties of GP16 were also evaluated: the peptide interacts in a similar way to LP22 (similar WaterLOGSY spectrum and K_d = 240 µM). Moreover, GP16 was shown to destabilize the tetrameric protein in a concentration-dependent manner (EC_{50} = 262 µM [142 – 383 µM]).

Finally, they assessed the contribution of each of the 16 residues in the LDH-B protein (GEMMDLQHGSLFQTP) to identify hotspots, i.e., the residues contributing to the stability of the LDH-B oligomeric state. For this purpose, they have performed an alanine scanning of LDH-B for each of these 16 residues and evaluated the properties of these mutants (Table I.4).

Table I.4: Characterization of LDH-B mutants[50]

Protein	MW (kDa)	EC_{50} (M) (Gdn.HCl)	T_m (°C)	Ratio 350/330 nm
LDH-B	155 ± 17	0.953 ± 0.012	74.5 ± 0.1	1.04
LDH-Btr	88 ± 13	< 0.1	57.5 ± 0.1	0.87
G60A	142 ± 10	0.735 ± 0.018	69.8 ± 0.1	1.05
E61A	77 ± 18	< 0.1	58.9 ± 0.1	0.95
M62A	141 ± 14	0.891 ± 0.012	71.6 ± 0.1	1.04
M63A	149 ± 18	0.845 ± 0.010	68.7 ± 0.1	1.04
D64A	143 ± 12	0.521 ± 0.012	56.4 ± 0.1	1.03
L65A	137 ± 23	0.630 ± 0.011	61.1 ± 0.1	1.03
Q66A	134 ± 16	0.893 ± 0.010	73.1 ± 0.1	1.04
H67A	154 ± 15	0.619 ± 0.016	67.2 ± 0.1	1.05
G68A	153 ± 18	0.722 ± 0.015	73.5 ± 0.1	1.05
S69A	144 ± 13	0.580 ± 0.009	66.0 ± 0.1	1.04
L70A	148 ± 17	< 0.1	67.1 ± 0.1	0.90
F71A	70 ± 17	< 0.1	53.9 ± 0.1	0.94
L72A	137 ± 10	0.348 ± 0.008	62.0 ± 0.1	1.03
Q73A	143 ± 10	0.636 ± 0.020	71.3 ± 0.1	1.05
T74A	145 ± 17	0.439 ± 0.012	64.9 ± 0.1	1.04
P75A	142 ± 17	0.752 ± 0.011	72.2 ± 0.1	1.05

Table I.4 includes different values allowing the characterization of the LDH-B mutants. The molecular weight of particles in solution was calculated by Mass Photometry (MP) and allows to determine the oligomerization state(s) of LDH-B protein in solution. The chemical stability was assessed by the dissociation of the homotetrameric form of LDH-B upon the addition of guanidinium hydrochloride (Gdn.HCl). The EC_{50} value can be determined by following the tryptophan fluorescence intensity (λ_{exc} = 286 nm; λ_{em} = 350 nm) to check the tetrameric integrity of LDH. Furthermore, the protein thermal stability was investigated by nanoDSF via determining the melting temperature and the 350/330 nm ratio was used to monitor the unfolding events of the mutants (tryptophan fluorescence intensity ratio (λ_{exc} = 286 nm; λ_{em} = 350/330 nm)). For mutants having denaturation patterns similar to the one of LDH-Btr, the 350/330 nm initial ratio is lower and followed by a red shift, while for mutants with a similar pattern to the tetrameric LDH-B, the initial ratio is higher and followed by a blue shift.

Based on this information, authors have identified four residues (Glu61, Asp64, Leu70 and Phe71) essential for the stability of tetrameric LDH-B (Table I.4 in green). Indeed, when these residues are mutated into alanine, the resulting mutant proteins present features similar to the LDH-Btr protein (dimeric form, chemical and thermal stabilities).

Objectives and strategy

Aim of the study and strategy

Because LDH is an enzyme at the core of lactate metabolism, essential for the anabolic growth and the proliferation of cancer cells, the protein was highlighted as an interesting therapeutic target for cancer treatment. Despite the design of many active site inhibitors, LDH remains a challenging therapeutic target. Hence, it remains necessary to consider other mechanisms of inhibition, such as targeting the active oligomeric state of the protein.

Based on the information presented in the introduction, the main goal of this work consists in a better characterization of the LDH-B tetramerization process. For this purpose, a general strategy was proposed (Fig. O.1).

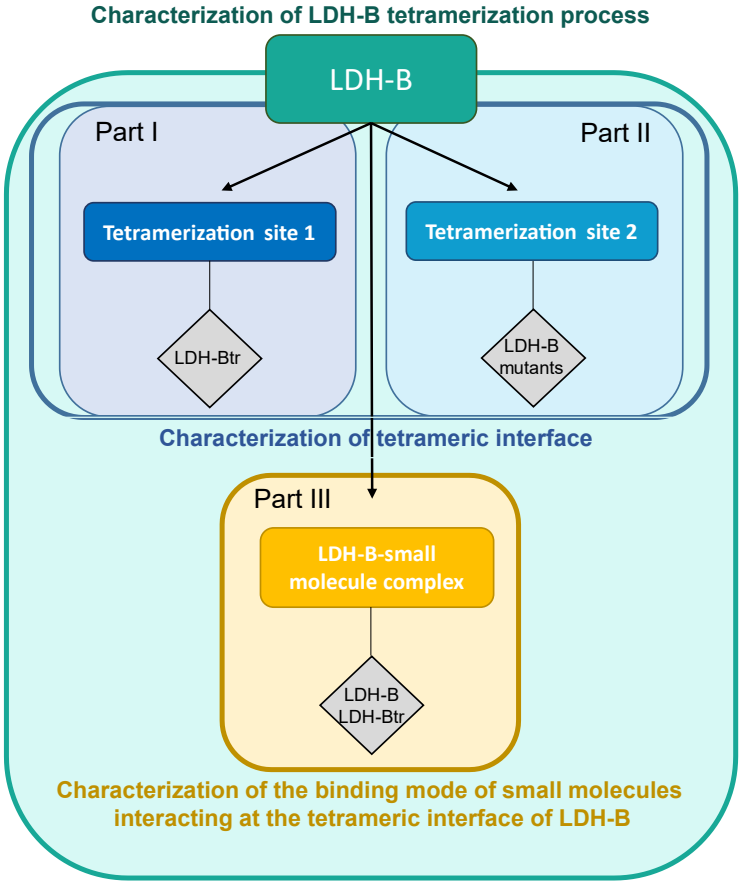


Figure O.1: Representation of the general strategy of the project

Using the LDH-B enzyme, the two first parts of the work consist in further characterizing the two tetramerization sites previously identified[50,59]. Part I will characterize the tetramerization site 1 by the structural study of a truncated form of LDH-B (LDH-Btr), where the 19 N-terminal amino acids were deleted. First, this chapter will present the methodology used to obtain protein crystals from which the structure of LDH-Btr will be

solved by X-ray crystallography. Then, the chapter will describe the LDH-Btr structure in a draft form for scientific publication.

Part II will study the tetramerization site 2 via the characterization of LDH-B mutants. The mutants will be studied by crystallography, and by molecular dynamics using the steered molecular dynamics methodology. Parts I and II will allow a better characterization of the tetrameric interface of LDH-B.

The third part of the work consists in characterizing the binding mode of small molecules previously identified, interacting at the tetrameric interface[79]. LDH-B-ligand complexes will be studied by X-ray crystallography using both full-length and truncated forms of the protein.

Results

Part I

Structural study of truncated LDH-B (LDH-Btr) - Characterization of the tetramerization site 1

1. Macromolecule crystallization: definition

The crystallization of a macromolecule can be defined as the self-assembly of molecules into a regular periodic 3D network[102,103]. The crystallization process requires the fulfilment of several conditions that can be classified into three categories (Fig. 1.1).

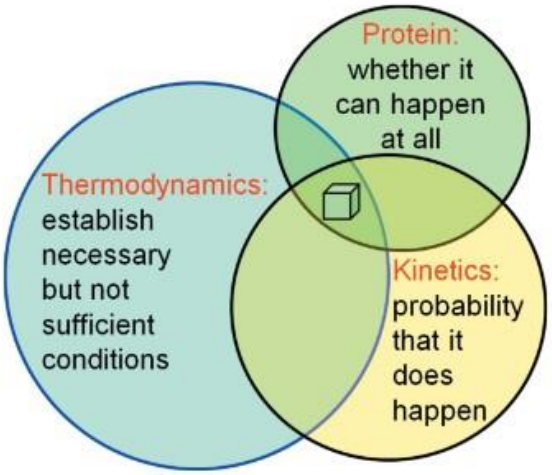


Figure 1.1: Conditions required to achieve protein crystallization – figure from Rupp *et al.*[102]

- Macroscopic processes being governed by thermodynamics, the formation of a crystal must be thermodynamically possible, i.e., by the existence of an equilibrium between a protein-rich phase (where crystal forms) and its growth solution. These macroscopic conditions are necessary but not sufficient[102].
- Microscopic processes, such as nucleation and crystal growth, are controlled by kinetics. These conditions must be favorable to allow the macroscopic processes[102].
- A final requirement to consider is the crystallizability of the protein (specific to each system), i.e., the tendency for the protein molecules to self-assemble into a periodic network resulting in a protein crystal. If any of the three requirements are not met, no protein crystal can be obtained[102,103].

Regarding the crystallizability of the protein, several factors influencing the stability of the protein may be responsible for the success of growing protein crystals. Stability can be divided into two types: compositional and conformational stability[103]. The **compositional stability** refers to the homogeneity within the protein solution used to perform the protein crystallization assays. The second factor, **conformational stability**, is linked to the protein flexibility or disordered regions that may be less likely to form protein crystals.

Many parameters can be optimized to improve the compositional and conformational stabilities of a given protein, i.e., its crystallizability. The first part of this chapter describes the strategy implemented to obtain a structure of the truncated form of LDH-B (LDH-Btr) and the results obtained by acting on different parameters (homogeneity, tag, stability and plasmid design). The second part of the chapter is dedicated to the structural characterization of the LDH-Btr protein from the structure that has been obtained. This part is included as a scientific paper draft.

2. Plasmid description

The amino acid sequence of LDH-Btr is based on the sequence of human LDH-B (Uniprot entry P07195), where the N-terminus part was truncated (residues 2 to 20) leading to a 315 residues protein.

6xHis-LDH-Btr – pET-28a(+) plasmid construction: this plasmid was provided by Prof. R. Frédérick from UCLouvain (Fig. 1.2A). The DNA sequence coding for LDH-Btr was cloned into a bacterial overexpression vector pET-28a(+). This vector allows the expression of N-terminal 6xHis-tag linked to a thrombin cleavage site. The cloning was performed using NdeI and Bpu1102I restriction sites. The sequence of 6xHis-LDH-Btr is composed of 337 amino acids.

LDH-Btr-6xHis – pET-24a(+) plasmid construction: this second plasmid was designed and ordered from Genscript© (Fig. 1.2B). Codon optimization was performed to improve the production of the protein in bacteria. The sequence was cloned into a bacterial overexpression vector pET-24a(+). This vector allows the expression of C-terminal TEV cleavage site followed by a 6xHis-tag. The restriction sites used for the cloning are NdeI and XhoI. The full sequence of LDH-Btr-6xHis is composed of 330 amino acids.

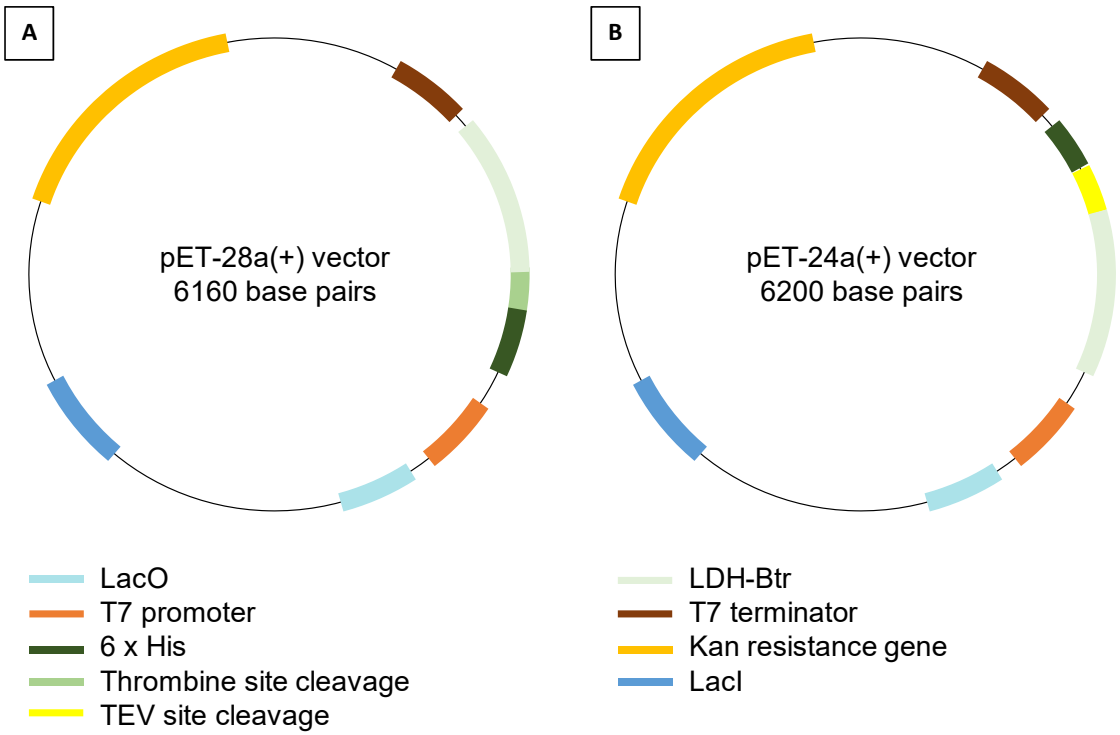


Figure 1.2: Representation of the two plasmid constructions coding for LDH-Btr – (A) pET-28a(+) plasmid construction – (B) pET-24a(+) plasmid construction

3. Crystallization assays of 6xHis-LDH-Btr protein

3.1. First crystallization assays

Results in this section were obtained using the 6xHis-LDH-Btr protein from the pET-28a(+) plasmid construction. First crystallization trials of LDH-Btr were performed using the protein purified by Immobilized Metal Affinity Chromatography (IMAC), allowing the separation of impurities from the protein of interest. On the elution profile (Fig. 1.3A), the first peak, named FT, is the flow through (FT), corresponding to all impurities that do not bind to the column. Once the FT is washed, an imidazole gradient is applied to elute the protein of interest from the column. At low imidazole concentration, a "shoulder" appears first on the chromatogram, corresponding to 75 kDa impurities with low affinity for the column. Then at higher imidazole concentration, the peak corresponding to the 6xHis-LDH-Btr protein is observed on the chromatogram. The LDH-Btr monomer has a MW of ~36 kDa. Each fraction reported on the chromatogram (a to c) corresponds to the fractions analyzed by SDS-PAGE (Fig. 1.3B).

After this first purification step, the protein presents high purity: fraction c was kept for the first crystallization assays.

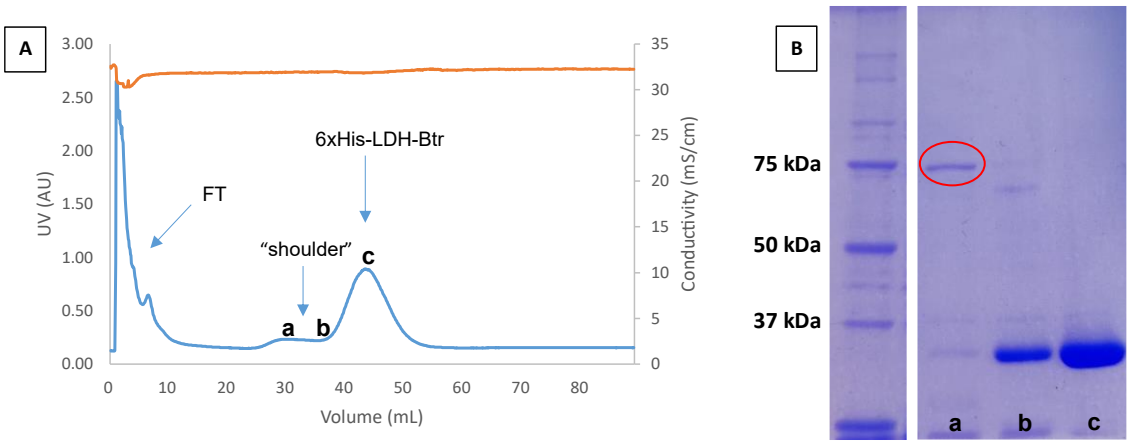


Figure 1.3: (A) Chromatogram of 6xHis-LDH-Btr purification using IMAC; elution gradient (30 to 250 mM imidazole – 45 min – 1 mL.min^{-1}); UV 280 nm (AU) in blue; conductivity (mS.cm^{-1}) in orange – (B) SDS-PAGE (10% acrylamide); impurities at 75 kDa are circled in red. The complete gel is reported in Appendix 1.A

Thereafter, two crystallization condition screens were performed. Data are shown in Table 1.1.

Table 1.1: Crystallization assays of LDH-Btr – screening of crystallization conditions

Screening	Protein conc. (mg.mL^{-1})	Kit	Drop*	T (°C)	Protein buffer	Ligands	Method	Results
1	8.5	HR2-110 / HR2-112	1/1	RT	50 mM phosphate buffer pH 7.5, 0.1 M NaCl, 10% (v/v) glycerol, 1 mM DTT	/	Sitting drop	No crystal
2	5.2						Sitting & hanging drop	Salt crystals, "fibrous" pattern, no diffraction

* Mother liquor/protein ratios

Table 1.1 shows the conditions used for the two screenings with the crystallization screening kits HR2-110 and HR2-112 (from Hampton Research*). The first one was performed with a protein concentration of 8.5 mg.mL^{-1} and no crystals were obtained. At this protein concentration, a precipitate was observed in most of the wells from the crystallization plate, which can be explained by a too-high concentration of protein. A second screening was performed using a protein concentration of 5.2 mg.mL^{-1}. At lower protein concentration, crystals were obtained in several wells of the screening plate. These conditions were then repeated and optimized with the hanging drop method to

*https://hamptonresearch.com/

refine the crystallization conditions. Nevertheless, the analysis of these crystals by X-ray diffraction (Soleil Synchrotron, Paris) has shown that they do not correspond to protein crystals.

At this stage, the addition of purification steps for 6xHis-LDH-Btr protein would increase the purity and homogeneity of the protein solution used for the next crystallization assays, improving the compositional stability parameter. The influence of protein purity and homogeneity is tested in the next section.

3.2. Influence of the purity and homogeneity of the 6xHis-LDH-Btr protein

Two more purification steps were added to the first IMAC. The first purification step is an anion exchange chromatography (Fig. 1.4A). During purification, a NaCl gradient was applied, and several protein populations elute at different elution times depending on the salt concentration. Those populations are represented by the different overlapping peaks on the chromatogram. Peaks resolution could not be improved despite changing salt elution gradient conditions. To obtain the most homogeneous protein mixture, only the main peak was kept (green box). Although the protein already exhibits high purity (Fig. 1.4C (b)), a gel filtration step was added afterward (Fig. 1.4B). All fractions (a to e) collected during the two purification steps were analyzed by SDS-PAGE (Fig. 1.4C). At the end of the purification steps, a yield of about 30% is obtained.

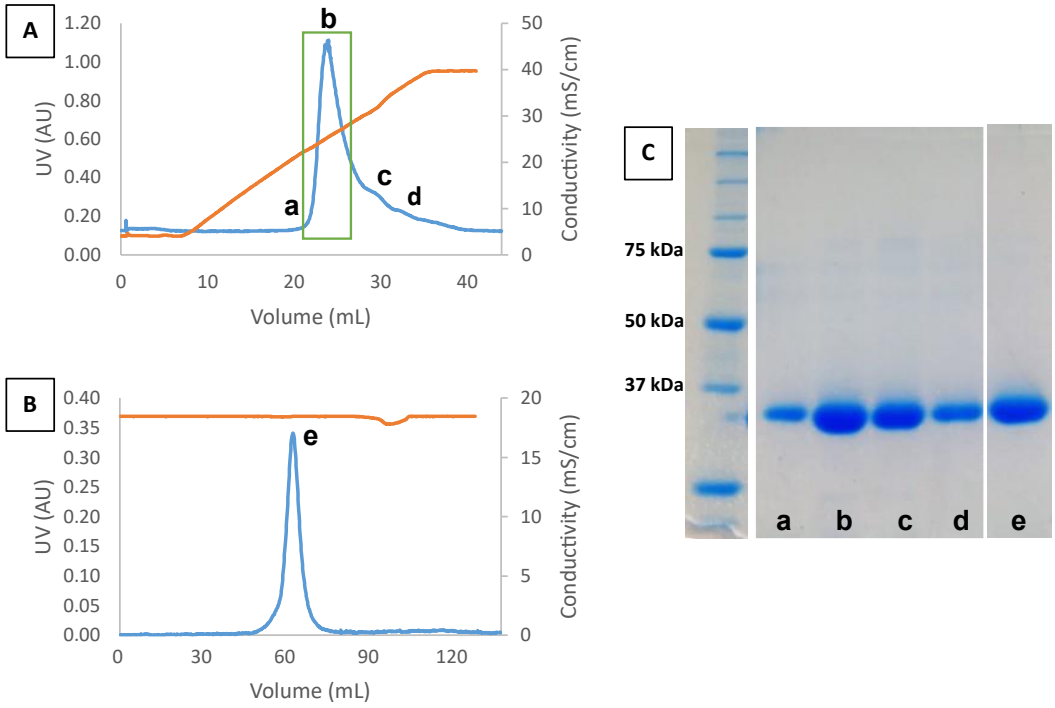


Figure 1.4: (A) Chromatogram of 6xHis-LDH-Btr purified by anion exchange chromatography; elution gradient (10 to 350 mM NaCl – 30 min – 1 mL.min^{-1}); UV 280 nm (AU) in blue; conductivity (mS.cm^{-1}) in orange – (B) Chromatogram of 6xHis-LDH-Btr purified by size exclusion chromatography; UV 280 nm (AU) in blue; conductivity (mS.cm^{-1}) in orange – (C) SDS-PAGE (10% acrylamide); colored with Gelcode Blue Stain Reagent. The complete gel is reported in Appendix 1.A

Three crystallization condition screenings were then performed using protein obtained after size exclusion chromatography. Conditions are listed in Table 1.2:

Table 1.2: Crystallization assays of 6xHis-LDH-Btr – screening of crystallization conditions

Screening	Protein conc. (mg.mL^{-1})	Kit	Drop*	T (°C)	Protein buffer	Ligand	Method	Results
3	5.4	HR2-110 / HR2-112	1/1	RT	50 mM phosphate buffer pH 7.5, 0.1 M NaCl, 10% (v/v) glycerol, 1 mM DTT	/	Sitting drop	Salt crystals
4	5.1	MD1-01 / MD1-02	1/2 – 1/1 – 2/1	RT				Salt crystals
5	5.1	Wizard I / Wizard II	1/1 – 2/1	4°C				Salt crystals

* Mother liquor/protein ratios

The protein concentration used is in the same range as the concentration determined during the screening 2 (Table 1.1). Several crystallization condition commercial kits were used to diversify the type of crystallization conditions to be screened (HR2-110 and

HR2-112 from Hampton Research[†], MD1-01, MD1-02, Wizard I and Wizard II from Molecular Dimensions[‡]). In addition, the diffusion rate between the protein-containing drop and the crystallization solution well can be modified by changing the protein/mother liquor ratios in the drop, or by incubating the crystallization plate at 4°C (screening 5). Crystals were obtained in all screenings under several crystallization conditions. Nevertheless, the analysis of these crystals by X-ray diffraction shows they are not protein crystals.

3.3. Influence of the presence of 6xHis-tag

The presence of a tag must be assessed to check if it can influence the crystallizability of the protein. As mentioned in Figure 1.2, pET-28a(+) construct is expressed with a 6xHis-tag linked to a thrombin cleavage site at the N-terminus of the protein. For 6xHis-LDH-Btr, this part, formed by 20 amino acids (cleavage site + 6xHis-tag), tends to be flexible and thus could decrease the conformational stability of the protein by interfering with crystal packing. Therefore, the tag may be unfavorable to the crystallization of this specific protein[103,104]. In the case of LDH-Btr, previous crystallization assays (with tagged protein) did not result in protein crystal formation. For that reason, it seems interesting to perform the cleavage of the tag via the thrombin cleavage site. This part was performed with LDH-Btr protein purified by IMAC. An optimization of thrombin cleavage was performed by testing several conditions: different ratios between thrombin units/mg protein and different incubation times. Each sample was analyzed by SDS-PAGE (see Appendix 1.B). Based on this optimization, the condition corresponding to five units of thrombin per mg of protein and 20 hours of incubation was selected for thrombin cleavage.

Using the optimized conditions, thrombin cleavage of 6xHis-LDH-Btr could be achieved (Fig. 1.5).

[†]https://hamptonresearch.com/
[‡]https://www.moleculardimensions.com/

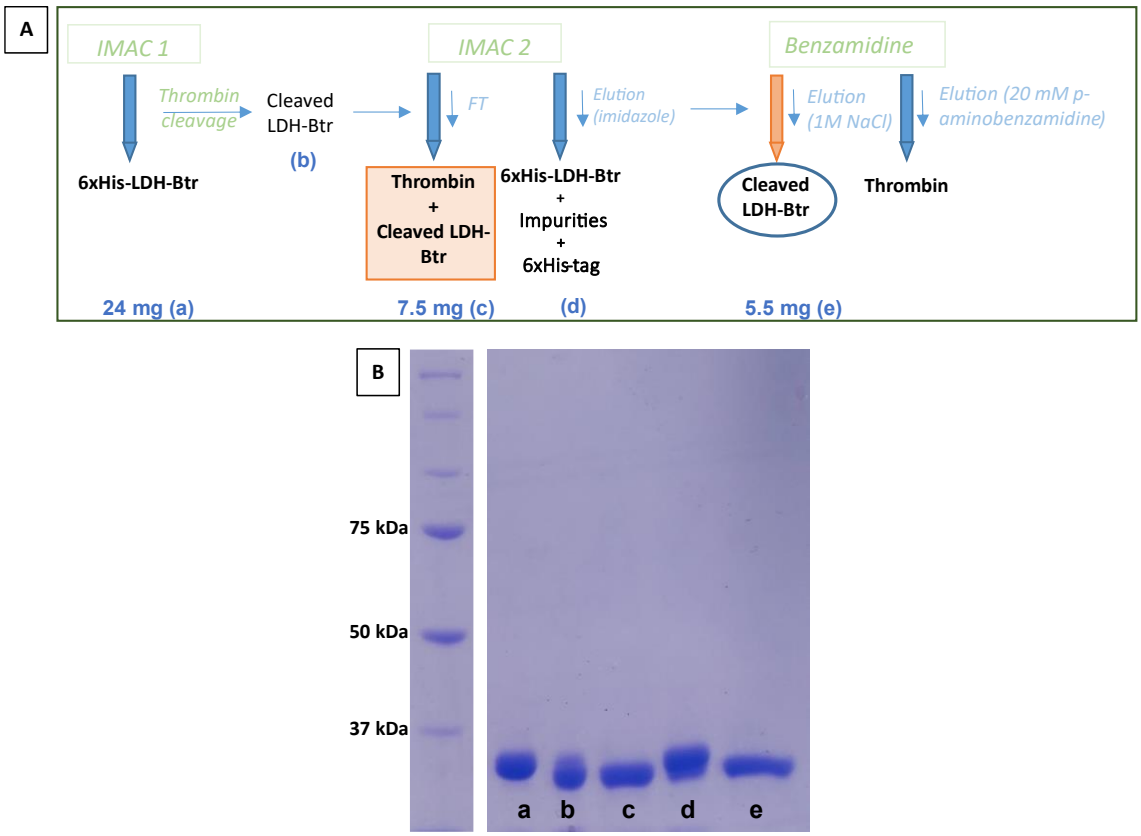


Figure 1.5: (A) Purification steps for thrombin cleavage; sample used with benzamidine column (orange box) – (B) SDS-PAGE (10% acrylamide); analysis of the different steps of purification (a to e; Fig. 1.5A). The complete gel is reported in Appendix 1.A

From IMAC-purified LDH-Btr (a), cleavage was performed using optimized conditions (b). Two purification steps (c to e) are required to separate the cleaved protein from the remaining uncleaved protein and thrombin. At the end of step e, a yield of about 20% is obtained. Before performing further crystallization assays using LDH-Btr$_{cleaved}$, the thermal stability of the protein was studied.

3.4. Thermal stability investigation of LDH-B by Differential Scanning Fluorimetry (DSF)

In addition to already studied parameters aiming to improve the crystallizability of the protein, thermal stability was also investigated. Consequently, this parameter can be evaluated by the DSF method[105,106]. DSF consists in applying a temperature scan to the protein incubated in the presence of a dye (SYPRO^{TM} Orange in our case). This one will interact with the exposed hydrophobic core of the protein as it denatures over temperature increase. The interaction between the dye and the protein can be followed by fluorescence. A sigmoid curve is obtained, and the melting temperature (T_m) can be

evaluated. It is defined as the temperature where 50% of the protein is unfolded and corresponds to the inflection point of the sigmoidal curve (Fig. 1.6).

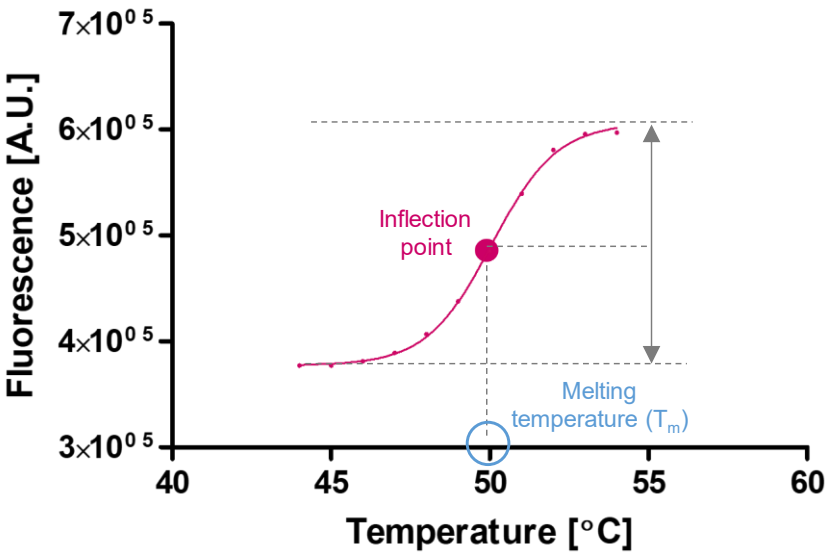
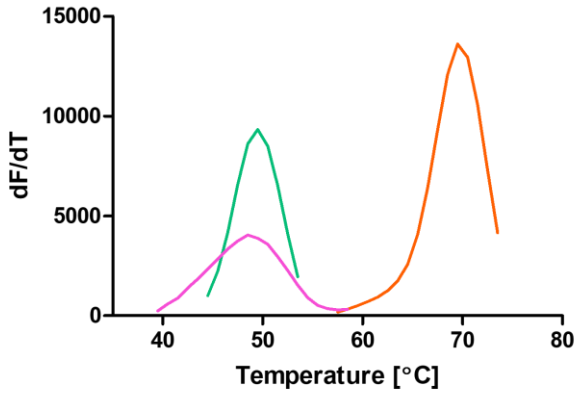


Figure 1.6: Representation of a DSF sigmoidal curve. The inflection point is represented in pink and the melting temperature (T_m) corresponds to the intersection of the inflection point with the x-axis

Therefore, the stability of both truncated forms of LDH-B (cleaved and uncleaved) could be compared with the full-length form of LDH-B (LDH-B). The first derivative has been calculated from the sigmoidal curves and the maxima correspond to the T_m (Fig. 1.7).



	$T_{melting}$ (°C)
LDH-B	69.6 ± 0.2
6xHis-LDH-Btr	49.5 ± 0.5
LDH-Btr$_{cleaved}$	48.4 ± 0.4

Figure 1.7: DSF profiles for different forms of LDH-B (LDH-B full length (LDH-B) in orange; truncated LDH-B (6xHis-LDH-Btr) in green and cleaved truncated LDH-B (LDH-Btr$_{cleaved}$) in pink) (n=4). Representation of the first derivative of fluorescence as a function of temperature (°C) – melting temperatures are determined from the maximum of the first derivatives of fluorescence

A difference of about 20°C is observed between truncated forms of LDH-B and the full-length form. This destabilization comes from the truncation of the tetramerization arm, known as essential in the tetramerization process of the protein[59,65]. However,

considering both melting temperature values for LDH-Btr and LDH-Btr$_{cleaved}$, cleavage of the N-terminal 6xHis-tag does not significantly affect the thermal stability of the truncated protein.

This loss of stability by the truncation of the tetramerization arm could explain the difficulty to crystallize the LDH-Btr protein. Therefore, thermal stability looks interesting as a parameter to consider to influence the protein crystallization[103,107]. Increasing the thermal stability of the truncated form by optimizing the buffer conditions (variable pH, buffer composition), as well as the addition of additives for crystallization assays can prevent protein unfolding or aggregation[105].

a) Buffer condition optimization by DSF

Buffer conditions were optimized by testing 30 different buffers (varying buffering agent, pH and salt concentration)[108]. Figure 1.8 represents the difference in melting temperature (ΔT_m) for each tested condition. This corresponds to the difference between the T_m when the protein is incubated in the tested buffer conditions and the control T_m, i.e., when the protein is incubated in the buffer after purification (50 mM phosphate buffer pH 7.5, 150 mM NaCl).

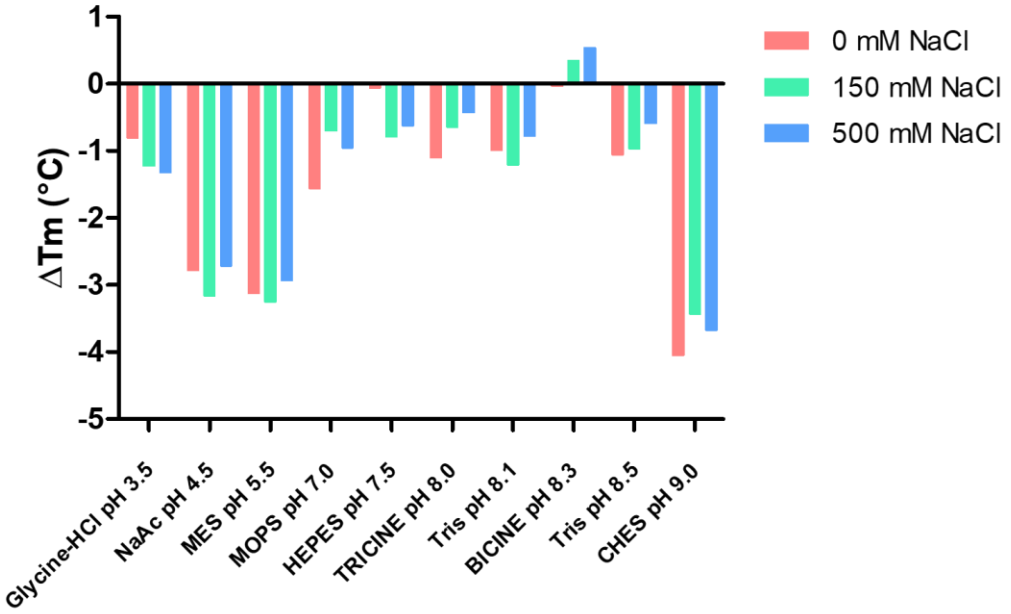
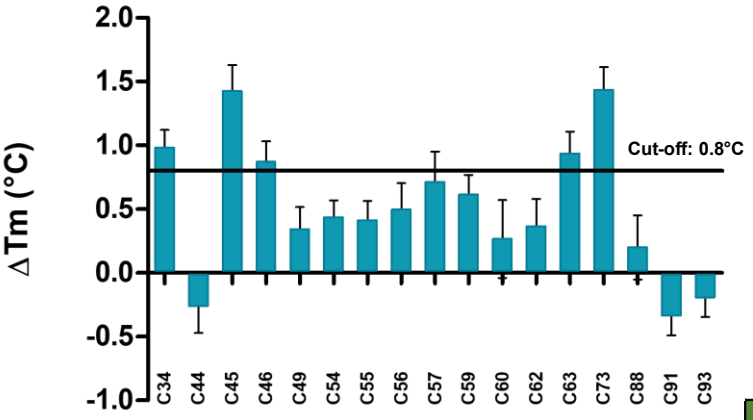


Figure 1.8: Buffer screening by DSF on LDH-Btr$_{cleaved}$ – representation of ΔT_m. The tested buffers (composition and variable pH) in presence of variable salt concentration (no salt in pink; 150 mM NaCl in green and 500 mM NaCl in blue) are presented

Only one buffer condition (100 mM Bicine pH 8.3, 500 mM NaCl) stabilizes the protein ($\Delta T_m > 0$) compared to the initial phosphate buffer.

b) Additives screening by DSF

The screening of additives was carried out using the protein in the buffer condition described in the previous section (Fig. 1.9).



	ΔT_m (°C)
C34 - Glycine (0.1 M)	1.0 ± 0.1
C45 - B-Nicotinamide adenine dinucleotide hydrate (0.01 M)	1.4 ± 0.2
C46 - Adenosine-5'-triphosphate disodium salt hydrate (0.01 M)	0.9 ± 0.2
C63 – NDSB-195 (0.3 M)	0.9 ± 0.2
C73 – Trimethylamine N-oxide dihydrate (3% (w/v))	1.4 ± 0.2

Figure 1.9: Additive screening (Hampton research kit (HR2-428)) using DSF on the LDH-$Btr_{cleaved}$ (n=4) – representation of ΔT_m

96 additives were tested by DSF on LDH-$Btr_{cleaved}$ to identify compounds that could stabilize the protein and potentially favor protein crystal formation. The additives that stabilized the protein in the first screening were tested a second time (see results Fig. 1.9). Figure 1.9 represents the ΔT_m obtained for the 17 selected additives (the list of additives is available in Appendix 1.C), corresponding to the difference between the T_m of the protein in the presence or in absence of tested additive. Those having a stabilization effect of at least 0.8°C were selected for the next crystallization tests. Indeed, binding of several ligands to a protein improves the conformational stability and thus the propensity of the protein to crystallize[103]. The five selected additives include the enzyme cofactor (NADH), NDSB-195 (corresponding to a non-detergent sulfobetaine preventing the aggregation of proteins and promoting their folding, renaturation, and crystallization[109–

[112]) and, finally, trimethylamine N-oxide dihydrate (TMAO) known to stabilize proteins by acting as a surfactant at the interface of folded protein surfaces[113].

Based on these optimized buffer conditions and the selected additives, new crystallization assays were performed with LDH-Btr$_{cleaved}$ (Table 1.3).

Table 1.3: Crystallization assays of LDH-Btr – screening of crystallization conditions with LDH-Btr$_{cleaved}$ in optimized buffer conditions in presence of selected additives

Screening	**Protein conc. (mg.mL^{-1})**	**Kit**	**Drop***	**T (°C)**	**Protein buffer**	**Additives**	**Method**	**Results**
6	4.1	HR2-110 / HR2-112	1/1	RT	100 mM Bicine pH 8.3, 500 mM NaCl	C34 / C45 / C46/ C63 / C73	Sitting drop	Salt crystals

* Mother liquor/protein ratios

For the last screening, the LDH-Btr$_{cleaved}$ concentration (~4 mg.mL^{-1}) used was lower than previously (~5 mg.mL^{-1}), because the protein precipitated at higher concentrations. In this screening, crystals were obtained under several crystallization conditions. Nevertheless, the analysis of these crystals by X-ray diffraction showed they are not protein crystals.

Cleavage of the His-tag, as well as optimization of buffer conditions and addition of additives, did not result in protein crystals for LDH-Btr. By comparing the thermal stability for the LDH-Btr$_{cleaved}$ used for crystallization assays (See Table 1.3, screening 6) and the 6xHis-LDH-Btr (Fig. 1.7), it can be observed that the peak for the cleaved protein is broader than for the 6xHis-LDH-Btr. This indicates that the initial sigmoidal curve is less sharp and thus the transition between the folded and unfolded state of the LDH-Btr$_{cleaved}$ protein is slower and covers a wider range of temperature[106,108]. As a result, the conformational homogeneity could decrease and thus decreases the chances of crystallizing the LDH-Btr[107].

3.5. Summary table of crystallization assays

Table 1.4: Summary table of 6xHis-LDH-Btr crystallization assays

Screening	Protein concentration ($mg.mL^{-1}$)	Kit	Drop*	Temperature (°C)	Protein buffer	Ligands	Method	Results	
1	8.5	HR2-110 / HR2-112	1/1	RT	50 mM phosphate buffer pH 7.5, 0.1M NaCl, 10% (v/v) glycerol, 1 mM DTT	/	Sitting drop	No crystal	IMAC purification (see §3.1)
2	5.2	HR2-110 / HR2-112	1/1	RT		/	Sitting drop / hanging drop	Salt crystals, « fibrous » pattern, no diffraction	
3	5.4	HR2-110 / HR2-112	1/1	RT	50 mM phosphate buffer pH 7.5, 0.1M NaCl, 10% (v/v) glycerol, 1 mM DTT	/	Sitting drop	Salt crystals	Addition of purification steps (see §3.2)
4	5.1	MD1-01 / MD1-02	1/2 – 1/1 – 2/1	RT		/	Sitting drop	Salt crystals	
5	5.1	Wizard 1 / Wizard 2	1/1 – 2/1	4°C		/	Sitting drop	Salt crystals	
6	4.1	HR2-110 / HR2-112	1/1	RT	100 mM Bicine pH 8.3, 500 mM NaCl	C34 / C45 / C46 / C63 / C73	Sitting drop	Salt crystals	His-tag cleavage, optimized buffer conditions and selected additives (see §3.3 & 3.4)

* Mother liquor/protein ratios

The first part of this chapter was dedicated to optimizing the compositional and conformational stabilities of the protein to improve the crystallizability of the 6xHis-LDH-Btr.

Different screenings of crystallographic conditions were performed using the LDH-Btr (see Table 1.4). The first two screenings were performed using 6xHis-LDH-Btr purified by IMAC and varying the protein concentration. Those screenings didn't lead to protein crystals. Therefore, two purification steps were added to increase the purity of 6xHis-LDH-Btr, as well as its compositional stability by improving the homogeneity of the protein used for the crystallization assays. Moreover, these three screenings (screenings 3, 4 and 5) were performed with different commercial screening kits and the crystallization conditions were different (ratio, T). Nevertheless, these screenings did not result in obtaining protein crystals. Finally, the last screening was performed with the LDH-Btr$_{cleaved}$ protein. In order to improve the conformational stability of the protein, the His-tag was cleaved via the thrombin cleavage site and an optimization of the buffer conditions, as well as an additive screening, were performed. Again, this screening did not result in protein crystals.

4. Crystallization assays of LDH-Btr-6xHis

In order to increase the chance of getting LDH-Btr protein crystals, a second plasmid construct (pET-24a(+)) was designed by adding the 6xHis-tag to the C-terminus of the protein (LDH-Btr-6xHis) (see Fig. 1.2). These two constructs were then compared via an online protein crystallization prediction server.

4.1. Crystallization prediction: XtalPred-RF server

a) Method

XtalPred[114] is a server available online, providing an algorithm to predict the crystallizability of a protein, that is the "probability of yielding well-diffracting crystals that allow structure determination"[115]. The original version of XtalPred (Expert Pool method) allowed crystallization probabilities to be determined from eight different physicochemical features: length (protein size), isoelectric point, gravy index, the longest disorder region, instability index, percentage of coil structures, coiled coils and insertion score. Therefore, from these features, the proteins can be classified into five "crystallization classes". Each class represents a different crystallization success and failure rate observed from the Protein Structure Initiative (PSI) targetDB database[115,116].

An updated version of XtalPred (XtalPred-RF)[115], based on the Random Forest classifier to predict crystallizability class, is also available online. This version considers the eight features from the original version, but also three additional physicochemical surface features of the protein, by predicting the relative surface accessibility (RSA): surface ruggedness, surface hydrophobicity and surface entropy[115]. Moreover, an update of the training and testing sets has been done considering data until 2011-2012 (2005 in the initial version). Therefore, from these features, the proteins can be classified into 11 "crystallization classes". This version of XtalPred-RF will be used in the following section to predict the crystallizability of the LDH-Btr protein produced from the two plasmid constructs.

b) Results

Eleven features were considered to assign a crystallization class to 6xHis-LDH-Btr and LDH-Btr-6xHis with the 6xHis-tag in the N-terminal or C-terminal position, respectively (Fig. 1.10).

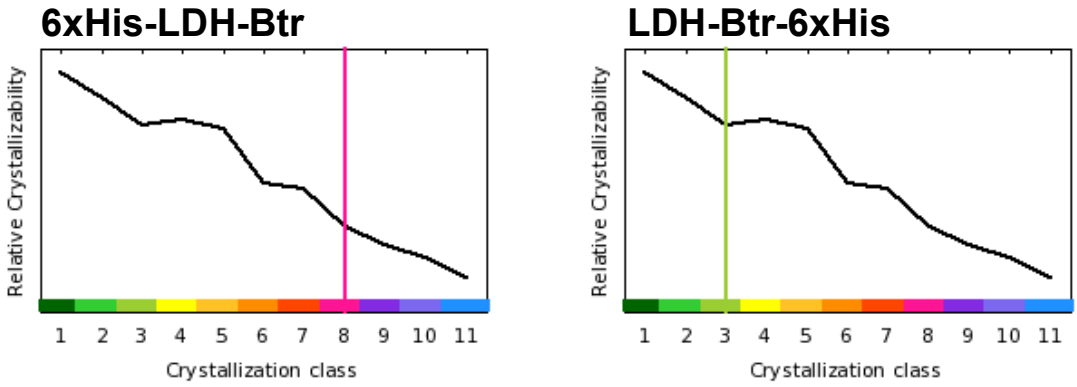


Figure 1.10: Crystallization class generated with XtalPred-RF[115] based on the sequence of the two plasmid constructs for LDH-Btr

A score of eight is obtained for the crystallization class of 6xHis-LDH-Btr, whereas a score of three is obtained for the second protein LDH-Btr-6xHis. Higher scores correspond to a lower probability to crystallize the protein. At this stage, XtalPredRF results highlight the real interest to change the purification tag position to favor the chances of crystallization of the LDH-Btr protein. The different features considered to establish the crystallization class are listed in Table 1.5.

Table 1.5: Features considered for the determination of the crystallization class of the two forms of LDH-Btr – more favorable characteristics for C-terminal 6xHis-tag construct are colored in orange

	6xHis-LDH-Btr	LDH-Btr-6xHis
Length	337	330
Isoelectric point (pI)	6.39	6.11
Gravy index	-0.04	-0.02
The longest disorder region	24	13
Instability index	25.74	23.27
Coil structures (%)	39	36
Coiled coils	0	21
Insertion score	0.03	0.03
Surface ruggedness	1.00	0.98
Surface entropy	-1.14	-1.14
Surface hydrophobicity	-0.39	-0.38

Highlighted features can give an explanation to better understand the difference in crystallization class for the 6xHis-LDH-Btr and the LDH-Btr-6xHis.

The length of the protein (number of residues composing the polypeptide chain of the protein) is directly related to the two protein constructs from which proteins were obtained. Indeed, for the LDH-Btr-6xHis, the part composed of the 6xHis-tag and the TEV-protease cleavage site is shorter (15 residues) than the one for the 6xHis-LDH-Btr (20 residues). This one is composed of the 6xHis-tag and the thrombin cleavage site. The chain length can, in our case, be directly related to another feature to consider for predicting protein crystallizability, namely **the longest disordered region** (Fig. 1.11A-B). It differs for each of the two LDH-Btr proteins directly resulting from the differences in plasmid construction.

Figure 1.11 (A-B) shows two plots with the number of proteins from the database (left y-axis) and the probability of crystallization (right y-axis) as a function of the longest disorder region (x-axis). Blue sticks represent the observed crystallization successes, while grey sticks are the observed crystallization failures. This feature can be predicted using the DISOPRED2 classifier[117] based on a training set of protein chains from PDB crystallographic structures (resolution (R) ≤ 2Å).

When considering the longest disorder region feature, 6xHis-LDH-Btr (Fig. 1.11A) has a crystallization probability lower than 0.3, whereas it is higher than 0.4 for LDH-Btr-6xHis

(Fig. 1.11B). This observation highlights the importance of reducing the length of the protein as much as possible by modifying the chosen plasmid construction.

A second feature that may explain a higher crystallizability of LDH-Btr-6xHis compared with 6xHis-LDH-Btr is **the instability index**[118] (Fig. 1.11C-D). From reference groups of stable and unstable proteins, the occurrence of dipeptides is studied, allowing the assignment of a weighted value of instability to each dipeptide. Based on these values, as well as on the length of the sequence of the protein of interest, an instability index can be calculated. Lower values correspond to higher crystallization probability. The 6xHis-LDH-Btr has an instability index of 25.74 (Fig. 1.11C), corresponding to a crystallization probability of about 0.45. The LDH-Btr-6xHis has an instability index of 23.27 (Fig. 1.11D), increasing the theoretical probability of crystallization around 0.5.

Finally, the updated version of XtalPred (XtalPred-RF), based on the Random Forest classifier calculates some features of the predicted protein surface using the NetSurfP algorithm[115,119]. One of these characteristics is **the surface ruggedness** (Fig. 1.11E–F). The latter is defined as "the ratio between surface area calculated as the sum of absolute solvent accessibilities of individual residues (predicted by NetSurfP) to the total accessible area expected for a protein of a given molecular mass"[115]. Indeed, the number of protrusions and cavities on the surface of the protein can have an impact on its crystallizability[115]. The 6xHis-LDH-Btr has a value of 1.00 for surface ruggedness while the LDH-Btr-6xHis has a value of 0.98. By having a slightly lower surface ruggedness value, the probability of crystallization increases from about 0.55 (Fig. 1.11E) to 0.60 (Fig. 1.11F). Moreover, it can be noticed that for values lower than 1, the number of observed successes is higher than the number of observed failures. This is not the case when the surface ruggedness value is higher than 1.

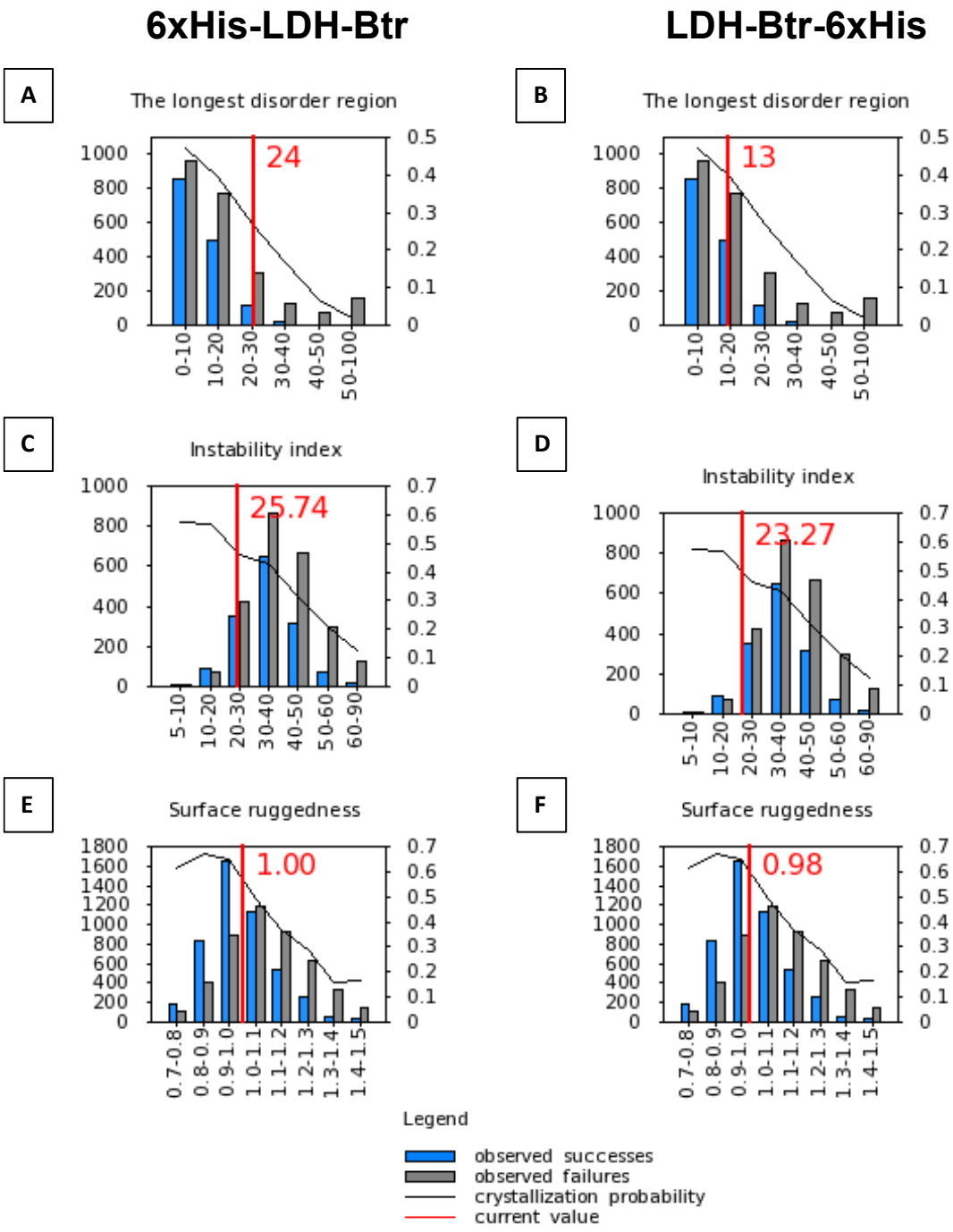


Figure 1.11: Features predicted by XtalPred server for 6xHis-LDH-Btr (A-C-E) and LDH-Btr-6xHis (B-D-F) – Plot of the number of proteins from the database (left y-axis) and the probability of crystallization (right y-axis) as a function of the studied features (x-axis)

The crystallizability of a protein depends on many characteristics such as those highlighted by the XtalPred prediction server. Since LDH-Btr-6xHis (new plasmid) gives better theoretical scores when using XtalPred, that construct will be used to perform the next crystallization assays.

c) Evaluation of the cleaved 6xHis-LDH-Btr crystallizability

The crystallizability of the cleaved form of 6xHis-LDH-Btr (LDH-Btr$_{cleaved}$) was also assessed via the XtalPred server. The cleavage of the 6xHis-tag allows to improve the crystallizability class: a score of three is obtained for the cleaved form while a score of eight was obtained for the uncleaved 6xHis-LDH-Btr. Among the characteristics considered, it is mainly the longest disorder region parameter that is responsible for the score improvement (Fig. 1.12). In the case of the cleaved form, it corresponds to seven residues and 24 for the 6xHis-LDH-Btr.

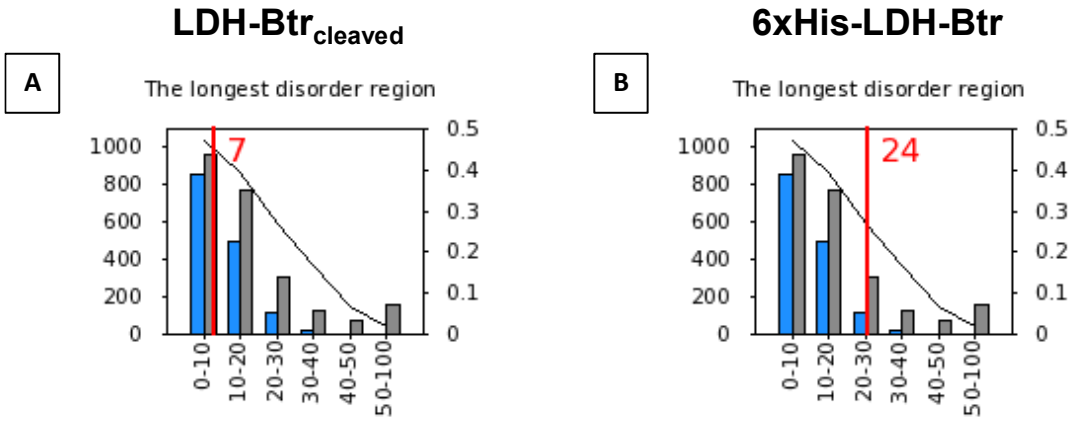


Figure 1.12: The longest disorder region feature predicted by XtalPred server for cleaved (A) and uncleaved (B) 6xHis-LDH-Btr. Plot of the number of proteins from the database (left y-axis) and the probability of crystallization (right y-axis) as a function of the studied features (x-axis)

Despite a predicted favorable score for LDH-Btr$_{cleaved}$ crystallization, no protein crystal was obtained. As described before, several factors can cause this. First, a low protein concentration is used for LDH-Btr$_{cleaved}$ crystallization assays (~4 mg.mL^{-1}). In addition, by thermal shift analysis, it seems that the conformational homogeneity has decreased following the cleavage of the 6xHis-tag. Therefore, the use of the new plasmid construct (pET-24a(+)) encoding for the LDH-Btr-6xHis was chosen and the produced and purified protein is used for the next LDH-Btr crystallization assays.

4.2. Crystallization assays of LDH-Btr-6xHis

The LDH-Btr-6xHis protein was produced from the new plasmid construct (Fig. 1.2B). It was then purified by IMAC under the same conditions as the 6xHis-LDH-Btr (see section 3.1). In addition, a second purification step by size exclusion chromatography was also performed for some crystallization assays.

Table 1.6: Crystallization assays of LDH-Btr-6xHis – screenings of crystallization conditions with LDH-Btr-6xHis purified by IMAC (1-2) and size exclusion chromatography (3-4). Composition of the protein buffers: Tris – 50 mM Tris-HCl pH 7.6, 150 mM NaCl, 5%(v/v) glycerol / Hepes - 25 mM Hepes pH 7.5, 150 mM NaCl, 5% (v/v) glycerol

Screening	Protein conc. ($mg.mL^{-1}$)	Kit	Drop*	T (°C)	Protein buffer	Ligands	Method	Results
1	7, 10, 13, 17	JCSG-plus™ (MD1-37)	1/1	RT	Hepes	/	Sitting drop	Salt crystals, no diffraction, protein crystals (too low resolution)
2	9, 13	JCSG-plus™ (MD1-37)	1/1	RT	Tris	OXM (3x) & NADH (3x**)	Sitting drop	Protein crystals
3	9, 13, 18, 26	JCSG-plus™ (MD1-37)	1/1	RT	Hepes	/	Sitting drop	Salt crystals, no diffraction
4	9, 13, 18, 26	JCSG-plus™ (MD1-37)	1/1	RT	Hepes	OXM (1.5x**) & NADH (1.5x**)	Sitting drop	Salt crystals, no diffraction

* Mother liquor/protein ratios

** Ligand concentration (ligand/protein ratios)

Several screenings of crystallization conditions were performed with the LDH-Btr-6xHis protein. A protein concentration ranging from 7 to 26 $mg.mL^{-1}$ was used. In addition, the different assays were performed with the JCSG-plus (MD1-37 from Molecular Dimensions[††]) commercial kit with a mother liquor/protein ratio of 1/1 in the drop at room temperature. Two different protein buffers were also used: Hepes (25 mM Hepes pH 7.5, 150 mM NaCl, 5%(v/v) glycerol) and Tris (50 mM Tris-HCl pH 7.6, 150 mM NaCl, 5%(v/v) glycerol). Finally, in screenings 2 and 4, the NADH cofactor and a substrate analog, oxamate (OXM), were incubated in the presence of the protein (1.5 and 3 times the protein concentration).

All the information concerning LDH-Btr-6xHis protein crystals, as well as the structure of the protein, are given in section 5, which is presented in draft form for scientific publication in the Acta Crystallographica journal (section F).

[††]https://www.moleculardimensions.com/

5. Structural characterization of truncated LDH-B (LDH-Btr) – Acta F publication

Crystallographic structure of a truncated form of hLDH-B in complex with NADH and oxamate

Abstract LDH proteins have been highlighted as interesting targets for cancer therapy due to their involvement in cancer cell metabolism. One strategy to inhibit these proteins is to target protein self-assembly. This work reports the structural analysis of a truncated form of LDH-B (LDH-Btr) for which the 19 N-terminal amino acids were deleted. This truncated part has been identified to be essential for the active tetrameric state of the protein. The protein structure was solved and refined in complex with its cofactor and a substrate analog at a resolution of 2.98 Å.

Keywords: ***human*** **LDH-B; truncated form; cancer**

1. Introduction

L-Lactate dehydrogenase (LDH) is active as a tetramer and belongs to the 2-hydroxy acid oxidoreductase family. The protein is involved in lactate metabolism. The function of the enzyme is to catalyze the interconversion of pyruvate + NADH + H^+ to L-lactate + NAD^+. Human LDH has two major subunits: the M form found mainly in anaerobic tissues (skeletal muscle, liver) and the H form found mostly in aerobic tissues (cardiac muscle). By assembly, these can form five different isoforms: three hetero-tetramers (M_3H, M_2H_2, MH_3) and two homo-tetramers, M_4 and H_4, also called LDH-5 or LDH-A and LDH-1 or LDH-B, respectively. The first homo-tetramer catalyzes preferentially the pyruvate reduction to lactate, while the second favors the lactate oxidation to pyruvate[59,65].

From a structural point of view, each monomer of LDH protein is composed of two domains: the first one is a "Rossmann"-type fold domain and the second one is a mixed α/β "substrate-binding" domain. In a first time, the cofactor binds to the Rossmann fold domain, and then the substrate can bind to the interface of the second domain[65]. Once the substrate binds, the active site switches to a closed conformation via a flexible loop. Therefore, it can interact with the C-terminal α-helix of the protein by vdW interactions

(Fig. 1). Nevertheless, this loop remains in an open conformation in the absence of substrate[69].

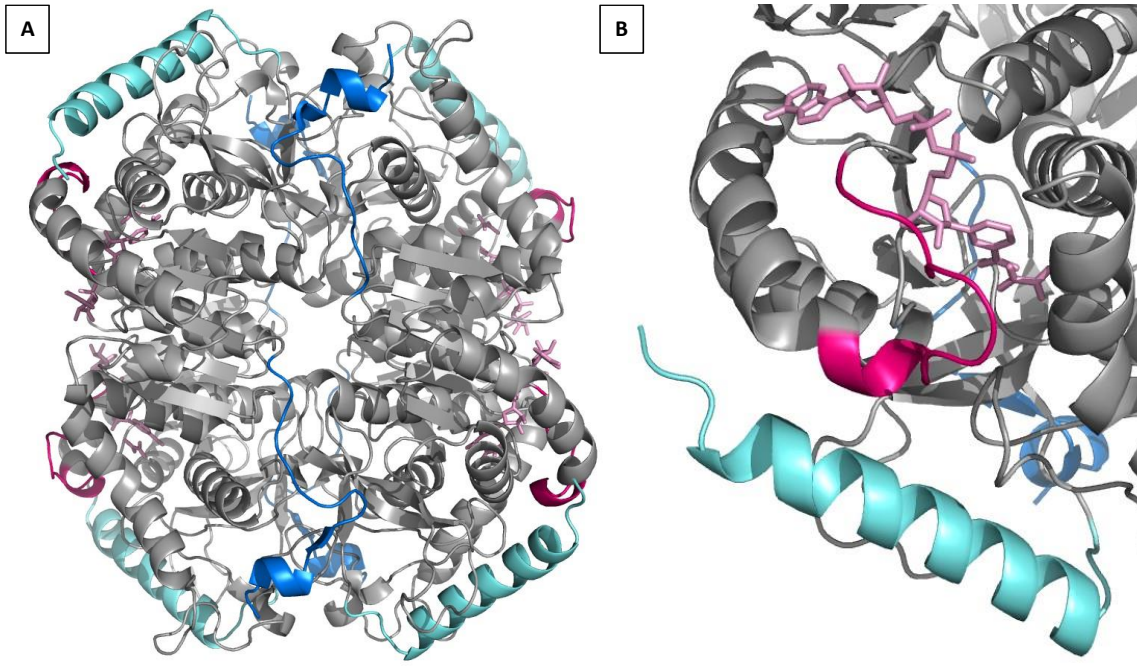


Figure 1 (A) Structural representation of human LDH-B (PDB entry: 1I0Z). Highlighting of different elements from the full-length hLDHB structure: tetramerization arm in blue, active site loop and cofactor/substrate analog in magenta and pink, respectively and the C-terminal α-helix in cyan – (B) Zoom on the active site. The loop, near the C-terminal α-helix, is in a closed conformation, due to the presence of cofactor and oxamate in the active site.

Moreover, the LDH protein assembles as a tetramer. The N-terminal tetramerization arm of each monomer, composed by residues 1-19 for LDH-B (in blue, Fig. 1), interacts and wraps around the two adjacent monomers allowing the assembly of the quaternary structure[65]. Currently, four structures of hLDH-B are available in the Protein Data Bank (PDB) (Table 1).

Table 1 Description of hLDH-B structures available in the PDB

1I0Z[65]	Structure of hLDH-B in complex with NADH and oxamate
1T2F[68]	Structure of hLDH-B in complex with NAD^+ and 4-hydroxy-1,2,5-oxadiazole-3-carboxylic acid
7DBJ[69]	Structure of hLDH-B in complex with NADH, oxamate and AXKO-0046
7DBK[69]	Structure of hLDH-B in complex with NADH

LDH enzymes have been identified as interesting targets for cancer therapy due to their implication in cancer cell metabolism. In fact, LDH promotes phenomena such as

angiogenesis[56,57], metabolic cooperation[51–53], and autophagy[52,54]. Different strategies can be used to target LDH proteins, such as the development of active site-directed inhibitors[90,91,97]. A second strategy consists in targeting the protein self-assembly. For this purpose, the tetramerization site can be studied, particularly by characterizing a truncated form of LDH-B (LDH-Btr) lacking 19 N-terminal amino acids (called tetramerization arm). Thabault and coworkers have shown that this truncation leads to protein dimerization in solution[59]. It is in this context that this truncated form of LDH-B is studied. In this paper, the structure of a truncated form of LDH-B (LDH-Btr) is presented.

2. Materials and methods

2.1. Macromolecule production and purification

LDH-Btr was expressed in *E. coli* BL21 (DE3) bacteria transformed with a pET-24a(+) plasmid coding for the protein of interest. The protein is expressed with a TEV cleavage site followed by a 6xHis-tag at the C-terminus. Bacteria were grown at 37°C until reaching an optical density (OD) of 0.6. The expression of LDH-Btr was induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) (C_f = 0.5 mM) at 20°C for 20 hours while shaking. After centrifugation, the pellet was resuspended in lysis buffer (50 mM Tris-HCl pH 8.5, 300 mM NaCl, 5 mM imidazole, 10 mM $MgCl_2$, 5% (v/v) glycerol, anti-proteases (cOmplete™, EDTA-free Protease Inhibitor Cocktail Tablets, Roche – one tablet/50 mL solution)) and bacteria were disrupted by sonication (10 x 30 seconds spaced with 30 seconds resting time).

The His-tagged LDH-Btr was purified by Immobilized Metal Affinity Chromatography (IMAC). After loading, the column is washed with buffer A (50 mM Tris pH 8.5, 300 mM NaCl, 30 mM imidazole, 10 mM $MgCl_2$, 5% (v/v) glycerol). The protein is then eluted using an imidazole gradient over 45 min (buffer B: 50 mM Tris pH 8.5, 300 mM NaCl, 250 mM imidazole, 10 mM $MgCl_2$, 5% (v/v) glycerol). The protein buffer was then exchanged using a desalting column with buffer C (50 mM Tris-HCl pH 7.6, 150 mM NaCl, 5% (v/v) glycerol). Finally, LDH-Btr was flash-freezed in liquid nitrogen and stored at -80°C.

2.2. Crystallization

The crystallization assays were performed using the sitting-drop vapor-diffusion method using a crystallization condition screening kit (JCSG-plus™ MD1-37) in 96-well plates

(Table 2). The protein was concentrated at 9 and 13 mg.mL^{-1} (in buffer C). The protein was previously incubated with NADH and oxamic acid (ratio protein/ligand: 1/3) for at least one hour on ice. Crystals were obtained after 10 days in 0.2 M lithium sulfate, 0.1 M sodium acetate pH 4.5 and 50% (v/v) PEG 400. Crystals were flash-freezed in liquid nitrogen.

Table 2 Crystallization conditions

Method	Sitting-drop
Plate type	96-well plates
Temperature (K)	293
Protein concentration (mg.mL^{-1})	9
Buffer composition of protein solution	50 mM Tris-HCl pH 7.6, 150 mM NaCl, 5% (v/v) glycerol
Composition of reservoir solution	0.2 M lithium sulfate, 0.1 M sodium acetate pH 4.5, 50% (v/v) PEG 400
Volume (µL) of drop and mother liquor/protein ratio	2, 1/1
Volume of reservoir (µL)	50

2.3. Data collection, processing, and refinement

Data collection was performed at Soleil synchrotron (Gif-sur-Yvette, France) on PROXIMA-1 beamline at 100 K, using a Dectris EIGER X 16M detector at a single wavelength of 0.97857Å. Data were processed with *XDS*[120] and the structure was solved by molecular replacement with Phaser[121] with a monomer of hLDH-B in absence of ligand (PDB entry: 1I0Z)[65]. Model building and refinement were performed with Phenix[122] and Coot[123]. Ligands in the structure were added to the structure using *eLBOW*[124]. Statistics for data collection and processing are presented in Table 3. Moreover, statistics for the final structure are summarized in Table 4.

Table 3 Data collection and processing for LDH-Btr

Values for the outer shell are given in parentheses.

Diffraction source	PROXIMA-1 beamline, Soleil
Wavelength (Å)	0.97857
Temperature (K)	100
Detector	Dectris EIGER X 16M
Space group	P 2_1 2_1 2_1
a, *b*, *c* (Å)	71.53, 85.66, 207.10
α, β, γ (°)	90.00, 90.00, 90.00
Resolution range (Å)	44.31 – 2.98 (3.09 – 2.98)
Total No. of reflections	360 566 (36 218)
No. of unique reflections	26 768 (2 626)
Completeness (%)	99.91 (100.00)
Multiplicity	13.5 (13.8)
$\langle I/\sigma(I)\rangle$	10.69 (1.76)
$R_{rmeas.}$	0.339 (1.697)
$CC_{1/2}$	0.995 (0.613)
Overall *B* factor from Wilson plot (Å^2)	63.72

Table 4 Structure solution and refinement for LDH-Btr

Values for the outer shell are given in parentheses.

Resolution range (Å)	44.31 – 2.98 (3.09 – 2.98)
Completeness (%)	99.91 (100.00)
No. of reflections, working set	26 758 (2 627)
No. of reflections, test set	1 337 (130)
Final R_{cryst}	0.2260 (0.2821)
Final R_{free}	0.3153 (0.3801)
No. of non-H atoms	
Protein	9 747
Ligand	360
Water	153
Total	10 193
R.m.s. deviations	
Bonds (Å)	0.004
Angles (°)	0.37
Average *B* factors (Å^2)	
Protein	60.55
Ligand	63.42
Water	48.80
Ramachandran plot	
Most favoured (%)	93.92
Allowed (%)	5.76

3. Results and discussion

3.1. Overall structure

The crystals of LDH-Btr were obtained in the orthorhombic space group $P2_12_12_1$. The asymmetric unit (ASU) is composed of four monomers (A, B, C and D) (Fig. 2A). Chain A of the protein was completely refined (from residue Asn20 to Leu333), while chains B, C and D were partially refined. Indeed, for these three chains, a variable part of the active site loop has not been refined (between residues Gln100 and Glu104 for chain B, Arg99 and Leu109 for chain C, and Val98 and Val110 for chain D). Moreover, the length of the chains varies due to the partial refinement of the C-terminal extremity (TEV cleavage site + 6xHis-tag): chain A was completed from residue Asn20 to Phe338, Asn 20 to Leu341 for chain B, Asn20 to Gly340 for chain C and Asn20 to His343 for chain D. Each monomer contains an NADH cofactor molecule and an oxamate substrate analog molecule in their active site. Furthermore, one glycerol and five Polyethylene Glycol (PEG) molecules are present in the structure of LDH-Btr.

The collected diffraction data has allowed solving the structure of LDH-Btr at a resolution of 2.98Å (R_{work}: 0.2260 and R_{free}: 0.3153). A general loss of secondary structure is observed for the LDH-Btr structure and is notably reflected in the Root-Mean-Square Deviation (RMSD) values presented in Table 4 (left column) when the four monomers of LDH-Btr are compared to monomer A (reference monomer) of full-length LDH-B (PDB entry: 1I0Z[65]). Moreover, the structure has an average B factor of 60.45Å^2. In general, three regions of the LDH-Btr protein are found to possess higher B factor values: the active site loop (residues 97-108) due to its flexibility to adopt an open/closed conformation, the antigenic loop known as a flexible region[65] and the C-terminal part of the protein.

3.2. Structural features of LDH-Btr protein

3.2.1. Substrate binding in the active site

The LDH-Btr protein was crystallized in the presence of the cofactor NADH and oxamate which is a substrate analog of the protein. One molecule of cofactor and one of oxamate are found in the active site of each of the four monomers. As mentioned in the introduction, their presence in the active site should lead to a switch of the active site loop (residues 97-108) to a closed conformation. The loop can then interact with the C-terminal α-helix, causing its displacement[69]. Nevertheless, only the active site loop from monomer A is in a closed conformation, while for the three other monomers, the

loop could not, or only partially, be refined (see section 3.1.). Indeed, no clear electron density was obtained at this region, probably because of the flexibility of this part of the protein, making it impossible to visualize it during the refinement of the structure of LDH-Btr protein. Moreover, the displacement of the C-terminal α-helix is not observed for any of the four monomers and is highlighted by high RMSD values between the C-terminal α-helix of the four monomers from LDH-Btr structure and the α-helix of the reference monomer A from full-length LDH-B structure (PDB entry: 1I0Z) (Table 5 – right column).

Table 5 Deviation between conformation of monomers as illustrated by the Root-Mean-Square Deviation (RMSD) values (nm). Left column: RMSD values between each of the four monomers from the LDH-Btr structure (monomer A to D) and monomer A (reference monomer) from the full-length hLDH-B structure (PDB: 1I0Z). Right column: RMSD values between the C-terminal α-helix (residues 307-332) from each of the four monomers from LDH-Btr structure (monomer A to D) and the C-terminal α-helix (residues 307-332) from the monomer A (reference monomer) from the full-length hLDH-B structure (PDB: 1I0Z). Each value was calculated with the "align" option of Pymol software[125] and considering only the backbone of the protein.

	Reference monomer (1I0Z)	**C-terminal α-helix (reference monomer)**
Monomer A	0.38	0.68
Monomer B	0.43	0.64
Monomer C	0.52	0.64
Monomer D	0.48	1.15

It has been shown that LDH-B protein associates as a dimer when its tetramerization arm is truncated (LDH-Btr). Moreover, LDH-Btr is less active compared to the full-length LDH-B. Indeed, the truncated form is ~three times less affine for the cofactor (K_M NAD^+: 0.587 ± 0.057 mM and 0.208 ± 0.005 mM for LDH-Btr and LDH-B, respectively) and ~20 times less for the substrate (K_M lactate: 33.77 ± 6.17 mM and 1.686 ± 0.284 mM for LDH-Btr and LDH-B, respectively)[59]. Nevertheless, it is not surprising to find the cofactor and a substrate analog in the active site of the LDH-Btr monomers, since it has an affinity for NADH similar to the one in the full-length LDH-B (K_d (LDH-Btr): 28.8 µM and K_d (LDH-B): 23.2 µM)[59].

3.2.2. Oligomerization state

Experimentally, it has been shown that tetramerization arm deletion led to a native dimeric state of the protein in solution. Indeed, the 19 N-terminal residues have been

identified to be essential for the tetramerization of the protein[59,65]. Nevertheless, from the structure of LDH-Btr presented in Figure 2A, a crystallographic tetramer is found in the asymmetric unit. Indeed, a second tetrameric interface has been discovered and highlighted to be involved in the tetramerization process. This interface is made with a 22-amino acid α-helix ended by a short loop (residues 54-75) interacting with the opposite dimer (Fig. 2B). At higher concentrations in solution, the dimeric LDH-Btr can self-assembly (K_d=1.25 µM), allowing the formation of a "weak" tetramer or also called "dimers of dimers" via this new tetrameric interface[50].

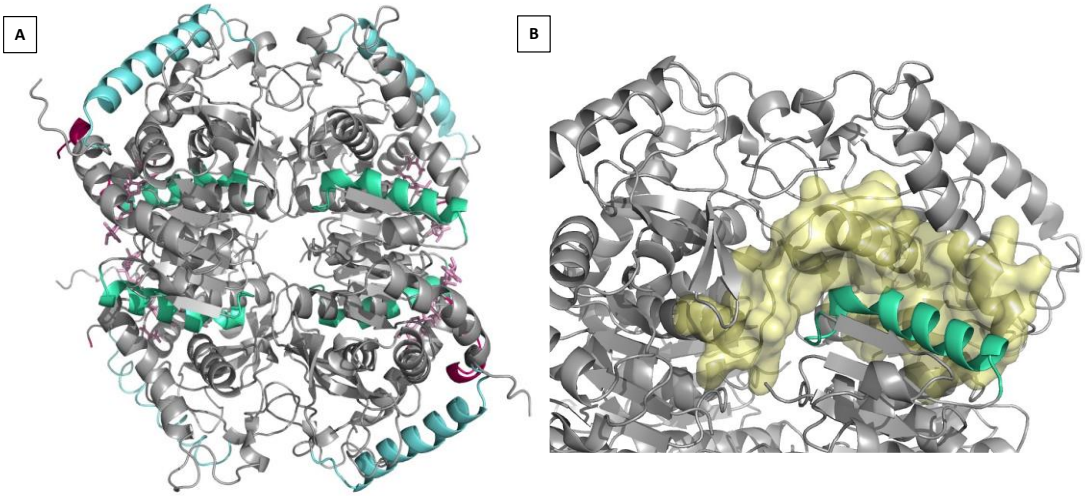


Figure 2 (A) Structural representation of truncated LDH-B (LDH-Btr). Highlighting of different elements from the LDH-Btr structure: active site loop and cofactor/substrate analog, in magenta and pink, respectively, C-terminal α-helix in cyan and 22-amino acid α-helix in green – (B) Zoom on the second tetramerization site: representation of this site in surface (yellow) with the interacting 22-amino acid α-helix (green)

Our crystal structure of LDH-Btr confirms that at high protein concentration, like in this crystal, this protein forms tetramers while dimers can be observed in solution at lower, more physiological, concentrations. This underlines one limit of using crystallography to study the oligomerization of proteins and calls for additional biophysical experimental (and theoretical) approaches.

Acknowledgements The authors would like to thank the SOLEIL synchrotron for the access to synchrotron facilities and PROXIMA-1 staff members for their help during the experiments.

6. Discussion and conclusions

The first part of this chapter described the strategy implemented to obtain a structure of LDH-Btr and proceeded in two sub-parts.

The first one is the part related to the pET-28a(+) plasmid construction coding for the 6xHis-LDH-Btr protein. At this stage, the strategy applied was to improve the crystallizability of the protein by acting on its stability, in particular on the compositional and conformational stabilities. The methods used to improve the stability are shown in Table 1.7[103].

Table 1.7: Summary of the methods used to improve compositional and conformational stabilities and their aims

	Adding purification steps	**Removal 6xHis-tag**	**Buffer screening**	**Additive screening**
Compositional stability	Yes	Yes	No	No
Conformational stability	No	Yes	Yes	Yes
Aim	Increase purity and homogeneity	Reduce disordered protein part	Optimize buffer conditions	Highlight protein stabilizing additives

Nevertheless, the use of this strategy did not lead to LDH-Btr protein crystals. Therefore, to have a chance to crystallize the truncated form of LDH-B, another strategy was used, i.e., gene construct optimization. The two major changes in the new plasmid construction are:

- The change in position of the 6xHis-tag (+ cleavage site) at the C-terminal extremity of the protein. This position is preferred to avoid the tag and the cleavage site interfering in the environment where the N-terminal tetramerization arm was truncated.
- The modification of cleavage site nature. A TEV site has been chosen for this new plasmid construction to reduce the number of purification steps when the tagged protein is cleaved. Moreover, the choice of this cleavage site allows to reduce the length of the disordered part at the C-terminal extremity, promoting the conformational stability of LDH-Btr.

In addition, another change is the addition of cofactor and substrate analog in crystallization assays of the LDH-Btrx6His.

In conclusion, the change in plasmid construction, as well as the addition of cofactor and substrate analog, led to protein crystals formation. These were analyzed by X-ray diffraction (Soleil Synchrotron, Paris) and the data allowed to solve the structure of LDH-Btr protein presented in the second part of this chapter. This part of the chapter also highlights that there are different approaches to promote obtaining crystallizable protein systems. Indeed, the diversity of the available experimental methods, but also the development of IT tools (such as XtalPred) allow to improve the chances of protein crystallization[126,127]. Nevertheless, protein crystallization remains an empirical phenomenon.

The second part of the chapter is dedicated to the structural characterization of the LDH-Btr structure. Based on the obtained structure, different information can be highlighted, such as about the active site, and about the oligomerization state of LDH-Btr.

- The unrefined part of the active site loop for three of the four monomers in the structure highlights the flexibility of this part. This could be typical of an open or intermediate conformation for the active site loop. At this stage, crystallographic data alone do not explain this observation.
- The ASU of LDH-Btr structure is composed of a tetramer, although the truncation of the tetramerization arm leads to a dimeric form of LDH-Btr in solution. The crystallographic tetramer can be explained because, at higher concentrations, it has been shown that the dimer can be self-assembled into a "weak tetramer" in solution[50]. There exists an equilibrium between the dimeric and tetrameric forms of LDH-Btr. Therefore, it seems unlikely to obtain a dimer by crystallography, since the protein concentrations used for crystallization assays are high (> 240 µM). The equilibrium is thus in favor of the tetrameric form. This observation underlines one limit of using crystallography to study the oligomerization of proteins and calls for additional biophysical experimental approaches.

Therefore, it seems interesting to characterize this second tetramerization site (tetramerization site 2). In this work, it will be characterized by crystallography and molecular dynamics simulations. These studies will be the subject of the results of part II.

Part II

Structural study of LDH-B mutants - Characterization of the tetramerization site 2

1. Structural analysis of LDH-B mutants

The second tetramerization site is investigated by a structural study of LDH-B mutants: LDH-B^{D64A}, LDH-B^{H67A}, and LDH-B^{F71A}. These mutants correspond to the mutation of aspartate 64 (Asp64), histidine 67 (His67) and phenylalanine 71 (Phe71) to an alanine (Ala) residue in the sequence of LDH-B protein. The three mutants have been chosen among the 16 mutants already studied in solution by Prof. R. Frédérick's group (UCLouvain)[50] (see Table I.4 in the introduction (section 5.2) for more details). Table 2.1 presents the characterization in solution of these three mutant proteins that are studied by X-ray crystallography in this section (Table 2.1).

Table 2.1: Characterization of different forms of LDH (LDH-B, LDH-Btr, LDH-B^{D64A}, LDH-B^{H67A}, LDH-B^{F71A}) by Mass Photometry (MP) and nanoscale Differential Scanning Fluorimetry (nanoDSF)[50]

Protein	MW (kDa)	EC_{50} (M) (Gdn.HCl)	T_m (°C)	Ratio 350/330 nm
LDH-B	155 ± 17	0.953 ± 0.012	74.5 ± 0.1	1.04
LDH-Btr	88 ± 13	< 0.1	57.5 ± 0.1	0.87
LDH-B^{D64A}	143 ± 12	0.521 ± 0.012	56.4 ± 0.1	1.03
LDH-B^{H67A}	154 ± 15	0.619 ± 0.016	67.2 ± 0.2	1.05
LDH-B^{F71A}	70 ± 17	< 0.1	53.9 ± 0.1	0.94

Table 2.1 includes different values allowing the characterization of the mutants. As a reminder, the molecular weight of particles in solution was calculated by Mass Photometry (MP) and allows to determine the oligomerization state(s) of LDH-B protein in solution. The chemical stability was assessed by the dissociation of the homotetrameric LDH-B upon the addition of guanidinium hydrochloride (Gdn.HCl). The EC_{50} value can be determined by following the tryptophan fluorescence intensity (λ_{exc} = 286 nm, λ_{em} = 350 nm) to check the tetrameric integrity of LDH-B. Finally, the protein thermal stability was investigated by nanoscale Differential Scanning Fluorimetry (nanoDSF) via determining the melting temperature (T_m) and the 350/330 nm ratio was used to monitor the unfolding events of the mutants[50].

Based on their observation, LDH-B^{D64A} and LDH-B^{H67A} form a tetramer, while LDH-B^{F71A} behaves mainly as a dimer in solution and presents a chemical stability similar to the one of LDH-Btr. Moreover, the mutation in LDH-B^{D64A} and LDH-B^{F71A} leads to a large thermal destabilization in comparison with LDH-B: ΔT_m ~ 18 and 20°C, respectively. Finally, LDH-B^{F71A} presents a denaturation pattern comparable to the dimeric LDH-Btr observed by a 350/330 nm ratio being lower than for the tetrameric proteins presented in Table 2.1[50]. Therefore, through their distinct profile (oligomerization states, chemical and thermal stability profiles), these three mutants seem to be perfect candidates for structural investigations.

1.1. Structural study: LDH-B^{H67A}

a) Expression and purification

A protocol for the expression and purification similar to the one for LDH-Btr-6xHis is used. From a pET-24a(+) plasmid construct coding for LDH-B^{H67A} mutant, the protein is expressed with a TEV cleavage site followed by a 6xHis-tag at the C-terminus. Then, the mutant was purified by Immobilized Metal Affinity Chromatography (IMAC) with the addition of a fixed concentration of imidazole.

b) Overall structure

The structure of LDH-B^{H67A} was solved and refined at a resolution of 2.38 Å (data completeness: 99.22 %) in the orthorhombic space group $P2_12_12_1$. The other structure statistics are reported in Appendix 2A. The asymmetric unit (ASU) is composed of four monomers (A, B, C, and D) forming a tetramer. Each chain composing the protein was completely refined (Ala1 to Leu333). Nevertheless, the length of observed chains varies due to the partial refinement of C-terminal extremity (TEV cleavage site + 6xHis-tag): Ala1 to Tyr337 (chain A); Ala1 to Asn335 (chains B and D) and Ala1 to Leu336 (chain C). Four molecules of glycerol are present in the structure. The mutant LDH-B^{H67A} has some structural features similar to LDH-B (see part I – section 5(1)). Furthermore, the protein was crystallized without substrate and cofactor, resulting in an open conformation of the active site loop as already described in the litterature[65,69].

Each monomer of the structure was superimposed with the monomer A (reference monomer) of the LDH-B structure (PDB entry: 1I0Z). The generated RMSD values are shown in Table 2.2.

Table 2.2: RMSD values (nm) between each of the four monomers from LDH-B^{H67A} structure (monomer A to D) and monomer A from the LDH-B structure (PDB entry: 1I0Z) (reference monomer). Each value was calculated with the "align" option of Pymol software[117] and considering only the C_α of the protein

	RMSD (nm) - Reference monomer
Monomer A	0.42
Monomer B	0.38
Monomer C	0.39
Monomer D	0.36

Similar RMSD values are obtained for all monomers from the structure when superimposed with the reference monomer. The open conformation of the active site loop of LDH-B^{H67A} protein can explain the RMSD values with the reference monomer. Indeed, the LDH-B protein (PDB entry: 1I0Z) was crystallized in presence of an analog of substrate and cofactor, resulting in a closed conformation of the active site loop and the shift of the C-terminal α-helix of the protein (Fig. 2.1). No other major changes are observed in the structure of the mutant compared to LDH-B structure.

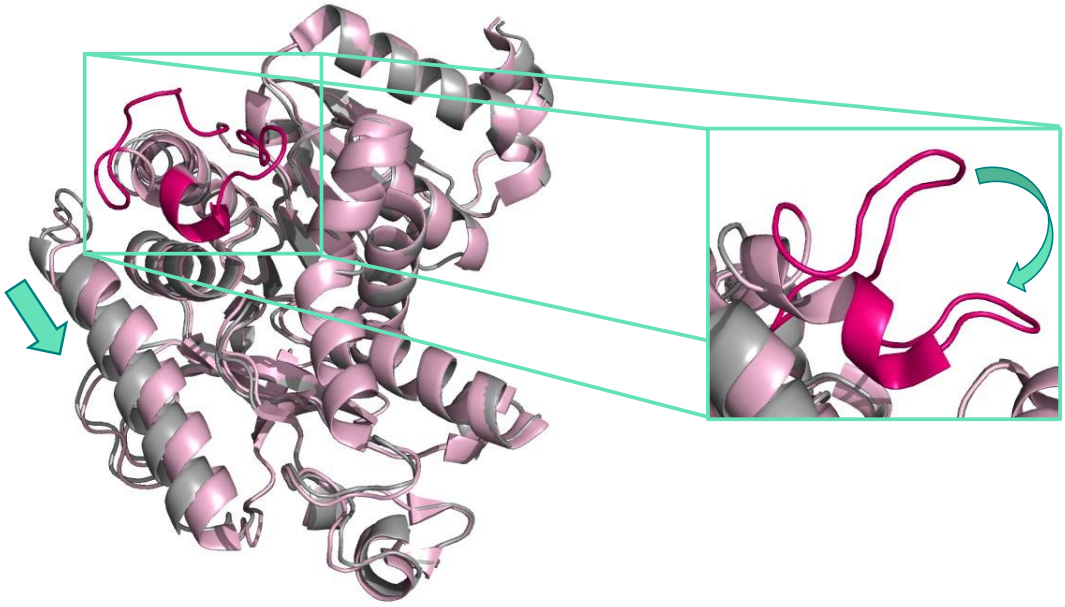


Figure 2.1: Superimposition of reference monomer (monomer A from LDH-B structure) (pink) with monomer B from LDH-B^{H67A} structure (grey). Zoom on the active site loop colored in dark pink. Arrows indicate the shift of structural elements between the two monomers (from open to closed conformation)

c) Interactions formed by His67 / Ala67

The environment of His67 residue was investigated when this residue is mutated to alanine (Fig. 2.2).

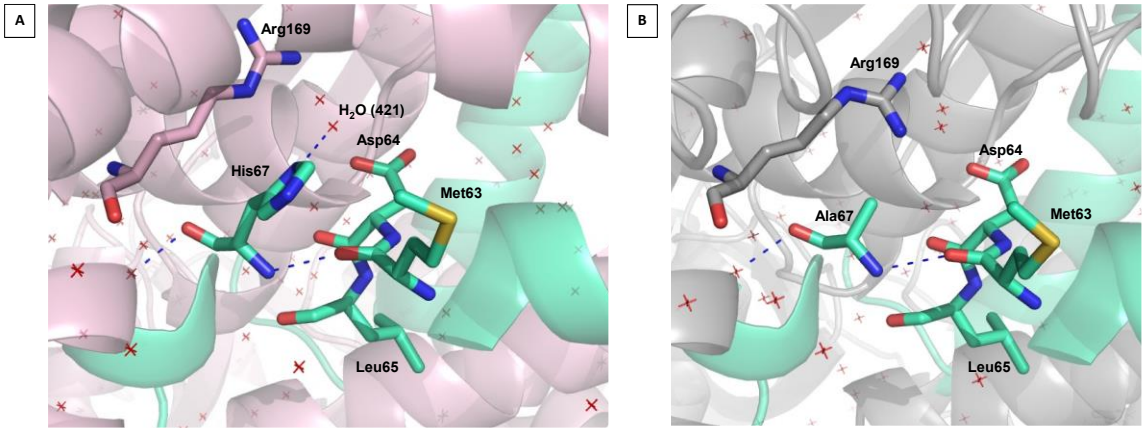


Figure 2.2: (A) Representation of residues interacting with His67 in wild-type LDH-B structure (PDB entry: 1I0Z (reference monomer)) – (B) Representation of residues interacting with Ala67 in LDH-B^{H67A} structure (monomer A) – Hydrogen bonds are represented by blue dotted lines and water molecules by red crosses

The two hydrogen bonds established with the backbone of His67 are preserved when mutated in an alanine residue. Moreover, a hydrogen bond between the histidine side chain and a water molecule (H_2O 421) stabilizes the 22-amino acid α-helix (Fig. 2.2A). When the histidine is mutated to an alanine, this hydrogen bond disappears (Fig. 2.2B). Nevertheless, it does not lead to a major conformational change in the environment of the mutated residue. Details regarding the interactions presented in Figure 2.2 are reported in Appendix 2B.

1.2. Structural study: LDH-B^{D64A}

a) Expression and purification

A protocol for the expression and purification similar to the one for LDH-Btr-6xHis is used. From a pET-24a(+) plasmid construct coding for LDH-B^{D64A} mutant, the protein is expressed with the TEV cleavage site followed by a 6xHis-tag at the C-terminus. Then, the mutant was purified by IMAC using an imidazole gradient for the elution of the protein of interest.

b) Overall structure

The structure of LDH-B^{D64A} was solved and refined at a resolution of 2.00 Å (data completeness: 72.90 %) in the monoclinic space group $P2_1$. Data collected from the crystals show high mosaicity and anisotropy. Therefore, some images of the dataset were removed and a correction for anisotropy was applied. The other statistics obtained for

the structure are reported in Appendix 2A. The ASU is composed of four monomers (A, B, C, and D) forming two different dimers. The tetramers of the protein are generated by symmetry. Moreover, protein chains were partially refined. Indeed, the part from Glu236 (chain B), Ser237 (chains A and C), or Ala238 (chain D) to Leu244 (chains A, B, and D), or Lys245 (chain C) is not refined (no density observed). Moreover, the length of the chains varies due to the partial refinement of the C-terminal extremity (TEV cleavage site + 6xHis-tag): Ala1 to Leu336 (chain A); Ala 1 to Asn335 (chain B); Ala 1 to Tyr337 (chain C) and Ala 1 to Leu336 (chain D). Five molecules of Polyethylene Glycol (PEG) are present in the structure. The LDH-B^{D64A} mutant has some structural features similar to LDH-B (see part I – section 5(1)). Furthermore, the protein was crystallized without substrate and cofactor, resulting in an open conformation of the active site loop.

c) Interactions formed by Asp64 / Ala64

As mentioned in the previous section, a part of an α-helix (named "D64-interacting α-helix") has not been refined for all monomers. The unrefined part corresponds to a part of the α-helix interacting with the aspartate 64 (Asp64) when the residue is not mutated (Fig. 2.3A). Each monomer of the LDH-B^{D64A} has been superimposed with the reference monomer of the LDH-B structure (PDB entry: 1I0Z) (Fig. 2.3B).

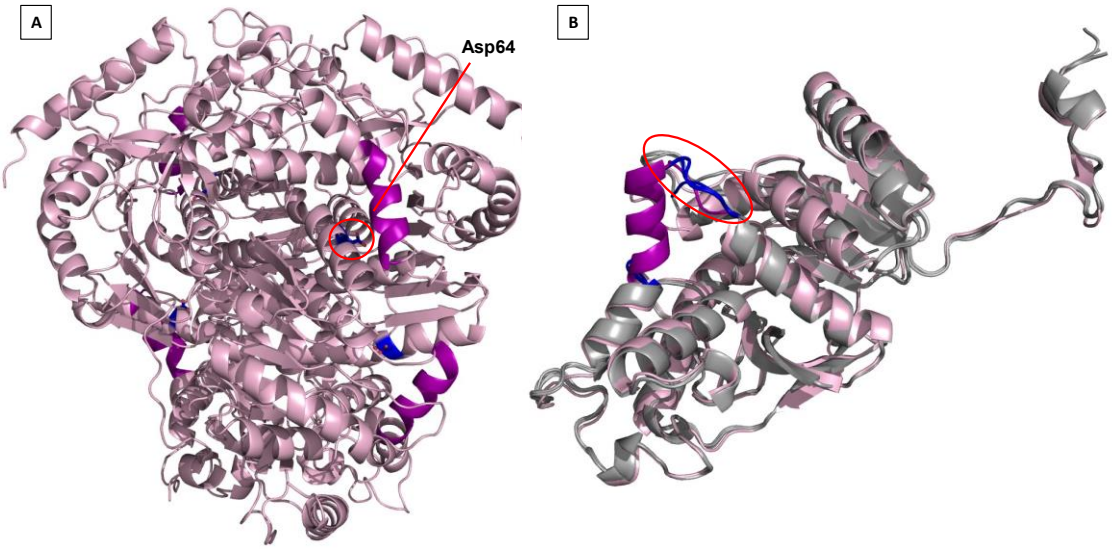


Figure 2.3: (A) Representation of LDH-B structure (PDB entry: 1I0Z). Non-refined part of D64-interacting α-helix in LDH-B^{D64A} is visualized in purple in LDH-B structure and Asp64 residue is represented in blue stick – (B) Superimposition of the four monomers (grey) of LDH-B^{D64A} structure with reference monomer (pink) of LDH-B structure (PDB entry: 1I0Z). The unrefined part of the D64-interacting α-helix is represented in purple (reference monomer) and the agitated residues surrounding the unrefined part are colored in blue for the four monomers of LDH-B^{D64A}

The residues surrounding the unrefined part of the D64-interacting α-helix present a higher B factor (between 25.0 and 50.0 $Å^2$) compared to the average B factor of the protein (22.3 $Å^2$), highlighting the destabilization of mutated residue environment caused by the mutation of Asp64 to Ala. Moreover, the interactions established with Asp64 or Ala64 were analyzed (Fig. 2.4).

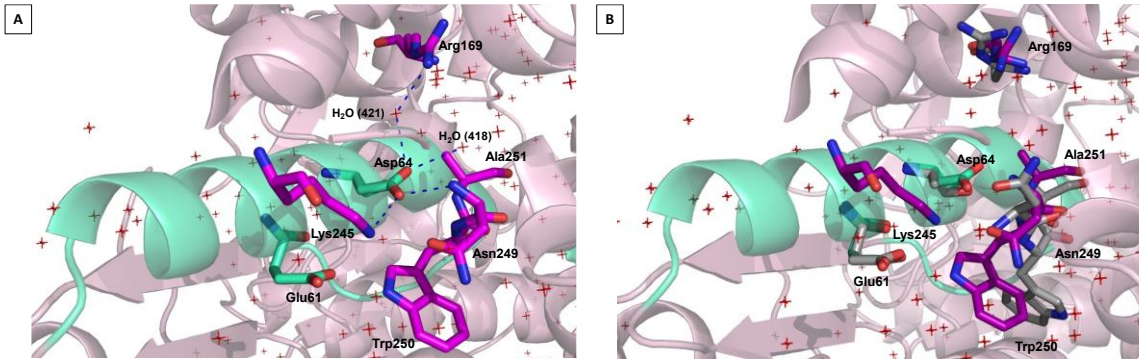


Figure 2.4: (A) Representation of residues interacting with Asp64 in LDH-B structure (PDB entry: 1I0Z (reference monomer)). Hydrogen bonds are represented by blue dotted lines and water molecules by red crosses – (B) Superimposition of residues (purple) interacting with Asp64 (green) from LDH-B structure and residues (grey) interacting with Ala64 (grey) from LDH-B^{D64A} structure

The Asp64 establishes a hydrogen bond network with its surrounding residues and two water molecules (H_2O 421 being the same water molecule involved in the hydrogen bond with His67 and H_2O 418) (Fig. 2.4A). When the residue is mutated to an alanine, the hydrogen bond network is disrupted which causes a rearrangement of several surrounding residues (Arg169, Asn249 to Ala251) (Fig. 2.4B). Upon Asp64 mutation, the number of interactions between residues decreases, making them more agitated and by consequence impossible to refine in the structure. Details of the interactions presented in Figure 2.4 are reported in Appendix 2B.

1.3. Structural study LDH-B^{F71A}

a) Expression and purification

A protocol for the expression and purification similar to the one for LDH-Btr-6xHis is used. From a pET-24a(+) plasmid construct coding for LDH-B^{F71A} mutant, the protein is expressed with a TEV cleavage site followed by a 6xHis-tag at the C-terminus. Then, the mutant was purified by IMAC using an imidazole gradient for the elution of the protein of interest.

b) Overall structure

The structure of LDH-B^{F71A} was solved and refined at a resolution of 3.22 Å (data completeness: 99.6 %) in the trigonal space group R32. Other statistics obtained for the structure can be found in Appendix 2A. The ASU is composed of six monomers (A, B, C, D, E, and F) forming a tetramer and a dimer. The second tetramer is generated by symmetry from the dimer of the ASU. The chains were completely refined (Ala1 to Leu333), except for chain F (Ala1 to Leu330). Nevertheless, the length of the other chains varies due to the partial refinement of the C-terminal extremity (TEV cleavage site + 6xHis-tag): Ala1 to Glu337 (chain A); Ala1 to Asn335 (chain B); Ala1 to Glu334 (chains C and D). Five molecules of glycerol and 10 molecules of PEG are present in the structure. The mutant LDH-B^{F71A} has some structural features similar to LDH-B (see part I – section 5(1)). Furthermore, the protein was crystallized without substrate and cofactor, resulting in an open conformation of the active site loop.

Each monomer of the structure was superimposed with the reference monomer from the LDH-B structure (PDB entry: 1I0Z). The generated RMSD values are shown in Table 2.3.

Table 2.3: RMSD values (nm) between each of the six monomers from LDH-B^{F71A} structure (monomer A to F) and the reference monomer from the LDH-B structure (PDB entry: 1I0Z). Each value was calculated with the "align" option of Pymol software[117] and considering only the C_α of the protein

	RMSD (nm) - Reference monomer
Monomer A	0.55
Monomer B	0.61
Monomer C	0.61
Monomer D	0.54
Monomer E	0.52
Monomer F	0.41

The RMSD values are similar for all monomers within the structure when superimposed on the reference monomer of the wild-type LDH-B structure, except for monomer F. As mentioned before, the open conformation of the active site loop of LDH-B^{F71A} protein can explain the RMSD values with the reference monomer of LDH-B. Moreover, RMSD values are higher than values calculated for the other mutants (see Table 2.2). That can be explained by a general loss of secondary structure, due to the lower quality of the crystal. Indeed, the mutant LDH-B^{F71A} is the least stable of the three studied mutants in

solution (see Table 2.1) and could therefore explain the difficulty to obtain good diffracting protein crystals. For monomer F, the RMSD value is lower because the active site loop for this monomer is in closed conformation despite the absence of substrate and cofactor in the active site. This highlights the high flexibility of this loop when it is not stabilized by interactions with the substrate and the C-terminal α-helix. Moreover, the environment of Phe71 was investigated when this residue is mutated to an alanine (Fig. 2.5).

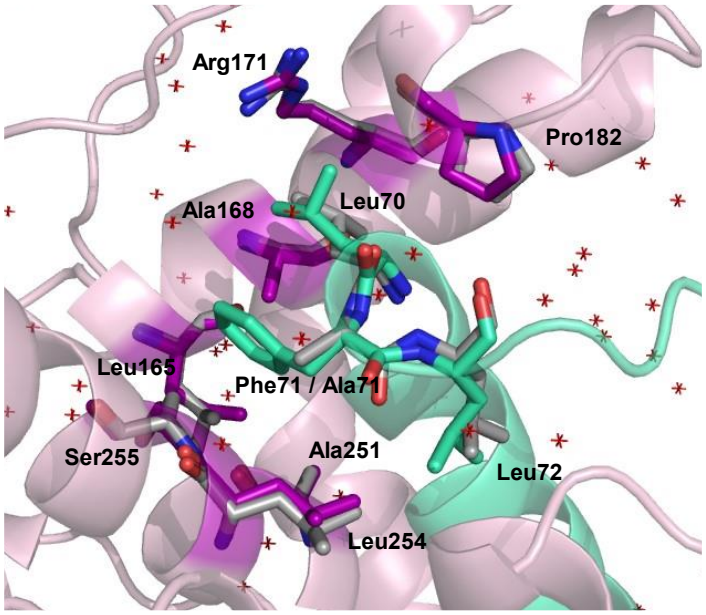


Figure 2.5: Superimposition of LDH-B (PDB entry: 1I0Z – reference monomer) and LDH-B^{F71A} (monomer A) structures. Representation of residues in purple (LDH-B) or grey (LDH-B^{F71A}) interacting with hydrophobic residues (Leu70, Phe71, and Leu72) in green (LDH-B) or grey (LDH-B^{F71A}).

The superimposition of the structures (LDH-B (PDB entry: 1I0Z) and LDH-B^{F71A}) does not show any significant structural changes. The mutation of Phe71 to Ala preserves the hydrophobic core (Leu-Phe/Ala-Leu) interacting with the surrounding residues, despite the loss of vdW interactions (amide π-stacked) between Leu254/Ser255 residues and the side chain of Phe71. Nevertheless, it does not lead to a conformational change in the environment of the mutated residue. Details of the interactions presented in Figure 2.5 are reported in Appendix 2B.

1.4. Discussion

Structural information obtained from the structures presented above can be linked to the results obtained in solution for these mutants (see Table 2.1).

- **LDH-B^{H67A}**: the ASU is composed of four monomers forming a crystallographic tetramer and the mutation of His67 to Ala does not lead to a conformational

change in the environment of the mutated residue. These data can be correlated with those in solution. Indeed, the mutant also forms a tetramer in solution and the mutation of His67 leads to a small destabilization (ΔT_m (LDH-B – LDH-B^{H67A}) ~ 7°C)[50].

- **LDH-B^{D64A}**: the ASU is composed of four monomers forming two dimers. Tetramers are generated by symmetry. Moreover, it can be observed in the structure that the hydrogen bond network disappears when Asp64 is mutated to Ala, leading to a rearrangement of the surrounding residues. Moreover, when the residue is mutated, a part of the D64-interacting α-helix is not refined in the mutant structure, highlighting this destabilization. These structural data can also be correlated with those in solution. Indeed, the mutant forms a tetramer in solution, as observed in the crystal structure and the mutation of Asp64 to Ala does lead to a large destabilization in solution (ΔT_m (LDH-B – LDH-B^{D64A}) ~ 18°C)[50].
- **LDH-B^{F71A}**: the ASU is composed of six monomers forming a tetramer and a dimer. The second tetramer is generated by symmetry from the dimer of the ASU. Despite the loss of interactions with the Phe71 sidechain upon mutation into an alanine residue, no major conformational change was observed in the structure of LDH-B^{F71A}. Nevertheless, the protein forms mostly a dimer in solution, although a small population of tetramers (19%) is also observed[50], which may explain the results obtained in crystallography. Indeed, a high protein concentration is used for crystallization assays, the tetramer has more chance to crystallize (because of the higher stability) than the dimeric form of LDH-B^{F71A}. Moreover, the mutation of Phe71 to Ala has been highlighted as one of the most destabilizing mutations (ΔT_m (LDH-B – LDH-B^{F71A}) ~ 20°C)[50]. However, no conformational changes were observed when comparing the structures LDH-B and LDH-B^{F71A}.

The analysis of the mutants has allowed to highlight structural characteristics that could be correlated with the results obtained in solution, e.g., for the LDH-$B^{H67A\ and\ D64A}$ mutants. Nevertheless, for the mutant LDH-B^{F71A}, the crystallographic data do not allow to understand the effect of the mutation on the behavior of the protein in solution. For this purpose, Molecular Dynamics simulations have been performed starting from the crystallographic coordinates of each mutant presented above, to complete the data obtained by structural analysis.

2. Characterization of LDH-B mutants by Molecular Dynamics simulations

2.1. Molecular Dynamics simulations based on solved crystallographic structures

Using the LDH-B mutant structures presented above, 300 ns-production Molecular Dynamics (MD) simulations have been performed at 300 K and 1 bar. These simulations aimed to highlight potential conformational changes or protein destabilization (using RMSD and Root-Mean-Square Fluctuation (RMSF) values) that occur as a result of residue mutations (Asp64, His67, and Phe71 mutated to Ala) within the different proteins. A RMSD analysis is used to compare two structures via the atom-positional root-mean-square deviation (in our case the comparison between the structure before and after 300 ns of MD simulations)[128]. A RMSF analysis is used to characterize the fluctuation within a structure, by identifying the more flexible parts of the protein[128]. The RMSF profiles for the crystallographic structures are generated from the experimental B-factors (see Materials and methods section for more details). The study of these potential conformational changes focuses on the 22-amino acid α-helix (containing the mutated residues) interacting with the tetramerization site 2.

The analysis of the RMSD and RMSF profiles of the mutants LDH-B^{H67A} and LDH-B^{F71A} did not reveal any conformational change along the MD trajectories (see Appendices 2C – 2D). Therefore, for LDH-B^{H67A}, this corroborates the data obtained in solution and by crystallography. The LDH-B^{F71A} mutant occurs predominantly as a dimer in solution, while a small tetramer population is also observed[50]. As for the results obtained in crystallography, the MD simulation does not explain the dimeric behavior of LDH-B^{F71A} in solution. One explanation could be that the simulation time is too short to observe any conformational change in LDH-B^{F71A} mutant. Moreover, a MD simulation using a dimer of the LDH-B^{F71A} mutant (not obtained in crystallography) could give more information about how the mutation affects the oligomeric behavior of the protein.

Since the RMSD profile for the 22-amino acid α-helix does not show significant deviation (see Appendix 2C), the RMSD profile for the tetramerization site 2 (interacting with the α-helix) during the 300 ns MD simulation is presented in Figure 2.6 (B-C). The site is defined as before (see Introduction part, section 5.2) and has been divided into three parts:

- Tetramerization site 2 (part-1): N164, A168, R169, R171-L173, E176, H181-S183

- Tetramerization site 2 (part-2): S237, E240, V241, K243-K245, Y247-A251, D258 (containing residues of D64-interacting α-helix)
- Tetramerization site 2 (part-3): E261, K265-L267, R269

It is highlighted that the presence of D64A mutation in LDH-B protein results in a larger amplitude of the RSMD profile for the tetramerization site 2 when compared with RMSD profile of LDH-B (Fig. 2.6B). The RMSD profiles for each part of the tetramerization site 2 were also generated (Fig. 2.6C). This highlights that the part-2 (including S237, E240, V241, K243-K245, Y247-A251 and D258 residues) has the highest contribution to the observed deviation.

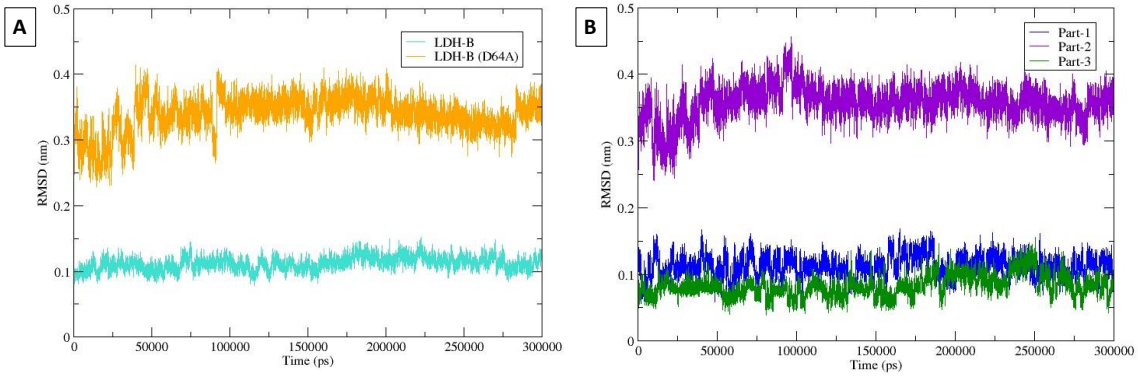


Figure 2.6: (A) RMSD profiles of the tetramerization site 2 of LDH-B (cyan) and LDH-B^{D64A} (orange) during the 300 ns MD simulation at 300 K and 1 bar – (B) RMSD profiles of the three parts of the LDH-B^{D64A} tetramerization site 2 during the 300 ns MD simulation at 300 K and 1 bar

These results can be complemented by a RMSF analysis of the LDH-B^{D64A} C_α atoms (Fig. 2.7). Data of LDH-B^{D64A} protein structure obtained by crystallography (LDH-B (D64A) crystallography) show that the entire protein is stable, except for the tetramerization arm which appears to be more agitated. The experimental and MD RMSF profiles of LDH-B^{D64A} can be compared. In this case, the agitation of several parts of the protein (tetramerization arm, active site loop, antigenic loop, and C-terminal extremity) is higher due to their high flexibility. Moreover, a fluctuation is observed for the residues 194 to 196, corresponding to a coil in the protein structure. This is also observed for some monomers of LDH-B protein. In addition, a large fluctuation of a part of the D64-interacting α-helix (some common residues with those constituting the tetramerization site 2 (part-2)) is observed. The fluctuation of this region is not visible on the experimental RMSF profile of LDH-B^{D64A} (Fig. 2.7 in green) because this part could not be refined. Also, no fluctuation has been observed for this region regarding the MD RMSF profile of LDH-B (Fig. 2.7 in cyan). Therefore, when Asp64 is mutated, the part-2 of the tetramerization site 2 is more flexible.

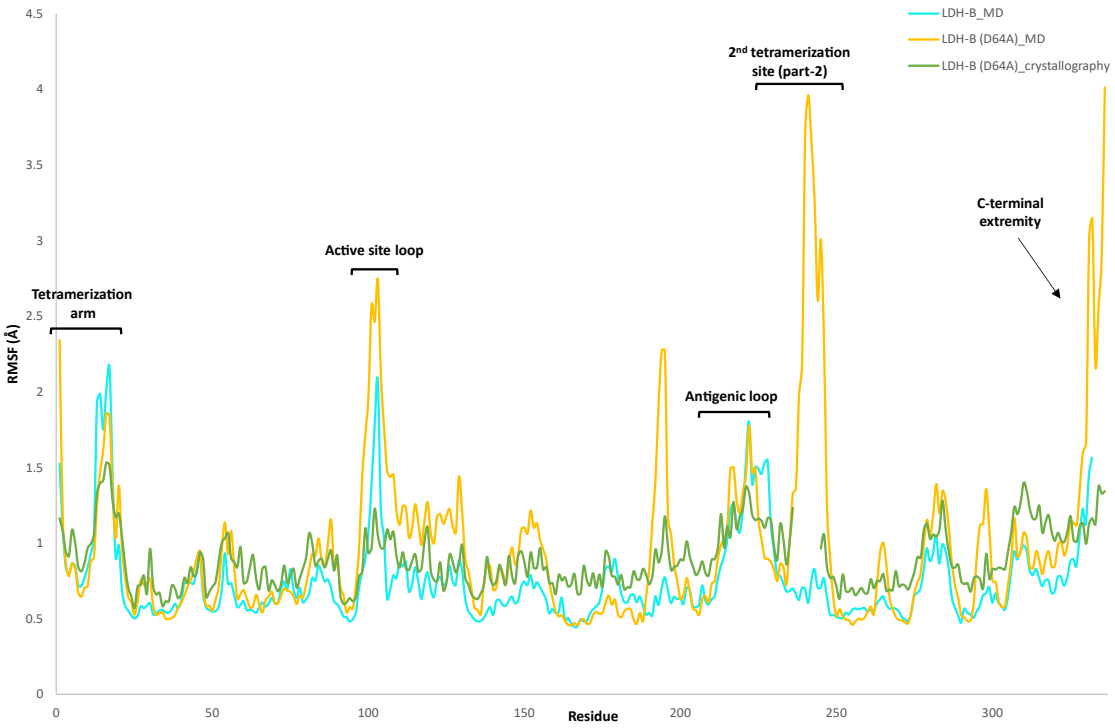


Figure 2.7: Experimental (from crystallographic structure – in green) and MD (in yellow) RMSF profiles of LDH-B^{D64} C_α and MD RMSF profile of LDH-B C_α (in cyan)

Regarding LDH-B^{D64A} protein, the mutation of Asp64 to Ala leads to the destabilization of a part of the D64-interacting α-helix composing the tetramerization site 2 (part-2). These data are in agreement with the results obtained by crystallography. Indeed, in the structure of LDH-B^{D64A}, a part of the D64-interacting α-helix could not be refined.

To complete these data, a second MD approach was used, i.e., Steered Molecular Dynamics (SMD), which allows for studying the protein self-disassembly of the LDH-B protein.

2.2. Steered Molecular Dynamics (SMD)

a) Methodology

Steered Molecular Dynamics (SMD) simulation is a computational approach inspired by experimental methods such as atomic force microscopy (AFM), optical tweezers, biomembrane force probe, or surface force apparatus[129,130]. These experimental approaches aim at the study of protein-ligand or protein complex interactions and provide a macroscopic view of the protein structure-function relationships. SMD simulations provide information from an atomic point of view and can therefore be complementary to the experimental approaches mentioned above[129].

The principle of the SMD approach is the application of a time-dependent external force in a given direction, through a damped virtual harmonic spring, on an (group of) atom(s) from a ligand or a protein[129–133]. The external force corresponds to[131]:

$$\vec{F}_{ext} = k(\vec{x}_0 + \vec{v}_{pull}t - \vec{x})$$

where k corresponds to the stiffness of the applied harmonic spring, $\vec{v}_{pull}$ is the velocity vector along which the restraint point moves (the pulling velocity), $\vec{x}_0$ is the initial position vector of the restraint point, and $\vec{x}$ corresponds to the reaction coordinate vector.

The SMD method can be used to study a variety of systems, such as a ligand-protein system or a protein-protein system[131,134]. In this work, the LDH-B tetramer protein self-disassembly is investigated via the application of an external force (along the y-axis) on the center of mass (COM) of one dimer (BD in blue), while the second dimer (AC in orange) is fixed (Fig. 2.8).

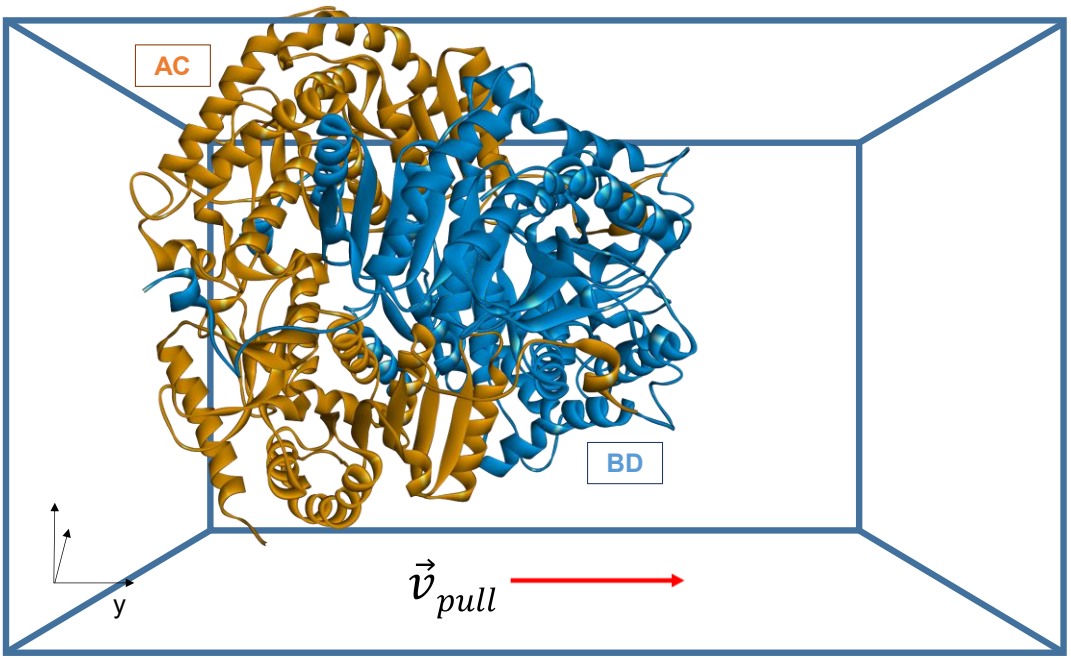


Figure 2.8: Application of a pulling velocity vector ($\vec{v}_{pull}$) along the y-axis on the center of mass of BD dimer (blue). The AC dimer (orange) is fixed in space. The two dimers form the tetramer of LDH-B protein. Figure generated from LDH-B structure (PDB entry: 1I0Z)

b) SMD optimization on LDH-B system: equilibration steps

The SMD results presented below are generated for the LDH-B structure (PDB entry: 1I0Z). Before performing a SMD simulation, a conventional Molecular Dynamics (MD) simulation was carried out at 300 K and 1 bar for 20 ns to reach a stable conformation of

LDH-B structure. It can be evaluated by calculation of the Root-Mean-Square Deviation (RMSD) (Fig. 2.9).

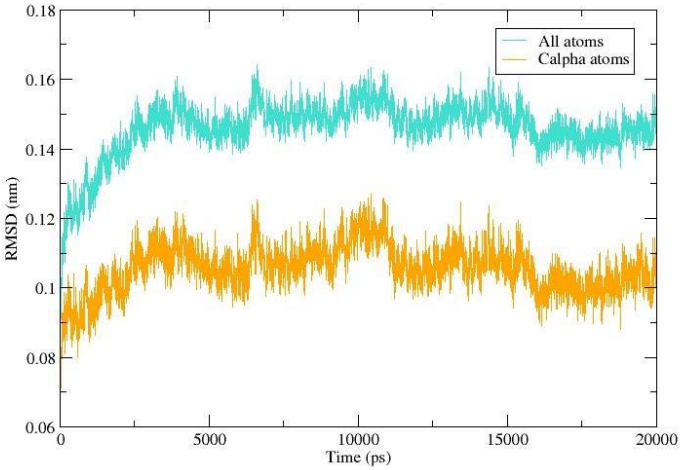


Figure 2.9: RMSD profiles of the LDH-B protein (PDB structure entry: 1I0Z) for C_α (orange) and all (cyan) atoms as a function of MD simulation time (20 ns)

The LDH-B protein reached an equilibrated state after ~4 ns of MD simulation. The RMSD values stabilize at ~0.15 nm for all atoms (cyan) and ~0.11 nm for C_α atoms (orange). The final LDH-B conformation obtained after this equilibration step was used as the starting point of the SMD simulation.

c) SMD optimization: selection of pulling velocity (v_{pull}) and spring constant (k) parameters

Different pulling velocities were tested (0.002, 0.005, and 0.01 $nm.ps^{-1}$). The selected velocity should be as small as possible to avoid generating disequilibrium and causing significant errors in simulation results[130,132]. Based on the separation trajectories, a pulling velocity of 0.005 $nm.ps^{-1}$ was chosen for the next SMD simulations.

On the other hand, the spring constant should be high enough to reach the breaking point corresponding to the separation of the two dimers, but not too high so that the measured force is not masked by the background noise[130,132,135]. In this work, different spring constants were tested (500, 1000, 2000, and 3000 $kJ.mol^{-1}.nm^{-1}$) while keeping the pulling velocity to the value of 0.005 $nm.ps^{-1}$ as selected previously (Fig. 2.10). The smaller the spring constant, the later the breaking point is reached and distinct. In this work, different position and angle constraints were applied on the system. Therefore, there is very little background noise observed on the graph. Based on the separation trajectories, a spring constant of 500 $kJ.mol^{-1}.nm^{-1}$ was chosen for the next SMD simulations.

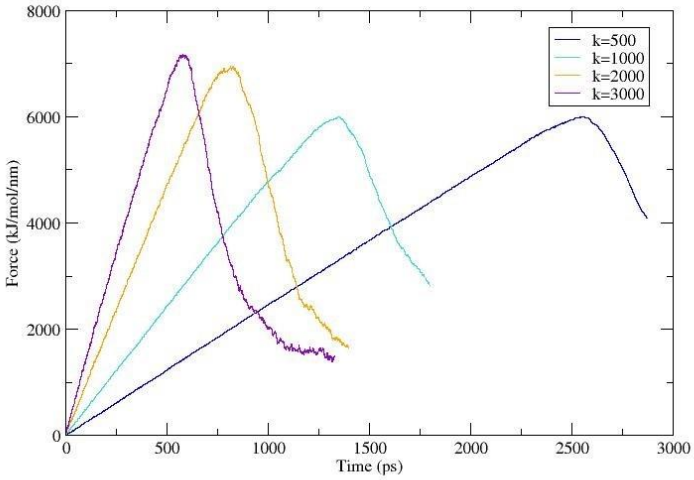


Figure 2.10: Influence of spring constant on rupture force ($kJ.mol^{-1}.nm^{-1}$) profiles as a function of time (ps)

2.3. Results

The SMD approach has been applied to the LDH-B structure (PDB entry: 1I0Z), with the following parameters: v_{pull} = 0.005 $nm.ps^{1}$ and k = 500 $kJ.mol^{-1}.nm^{-1}$.

a) Force and COM position profiles

As shown in Figure 2.11A, the force increases linearly with time until the breaking point is reached around 2670 ps. It corresponds to the separation of the two LDH-B dimers. This force profile can be correlated with the profile of the center of mass (COM) position as a function of time (Fig. 2.11B). The COM position of the pulled BD dimer of LDH-B vs the fixed AC dimer remains stable until ~2250 ps, where a slight variation is observed. From 2670 ps, the COM position strongly varies, corresponding to the separation of the two dimers. Figure 2.11D shows LDH-B states during the simulation: at t = 0 ps, the tetramer is in equilibrium. At t = 2500 ps, LDH-B is still assembled as a tetramer, just before the two dimers separate around t = 2670 ps. Finally, at t = 2970 ps, the two dimers are totally separated.

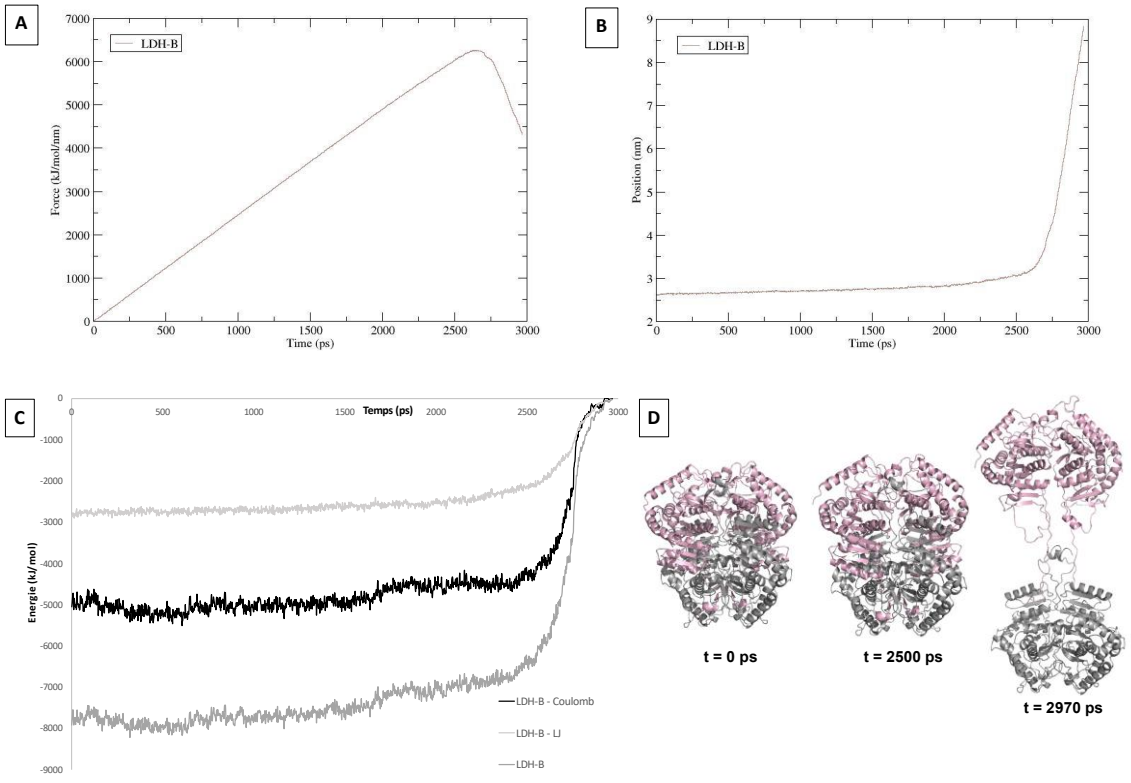


Figure 2.11: SMD simulation with LDH-B system from LDH-B structure (PDB entry: 1I0Z) – (A) Profile of the force (kJ.mol^{-1}.nm^{-1}) as a function of time (ps) – (B) Profile of the COM position (BD dimer) (nm) as a function of time (ps) – (C) Energy profile of BD dimer interacting with AC dimer during the SMD simulation. Total interaction energy (kJ.mol^{-1}) is presented in grey, short-range Lennard-Jones and short-range Coulomb energy contributions are in light grey and black, respectively – (D) Different states of LDH-B during the SMD simulation at t = 0 ps, t = 2500 ps and t = 2970 ps. The AC dimer (fixed) is represented in grey and BD dimer (mobile) in pink

b) Energy analysis of LDH-B during SMD simulation

The interaction energy profile between the dimers during the separation is presented in Figure 2.11 (C). At time t = 0 ps, the LDH-B system is in equilibrium with a low total interaction energy. Once SMD simulation starts, i.e., when the external force is applied on the COM of BD dimer to disassemble the tetramer, the system is no longer in equilibrium. Therefore, the energy fluctuates slightly at the beginning of the simulation. Thereafter, the fluctuation is more important between ~1700 ps and 2500 ps. The energy variation is due to the disappearance of the weakest interactions at the tetrameric interface of the protein, while the main interactions between the 22-amino acid α-helix and BD dimer allow to maintain the tetrameric state of LDH-B protein. Finally, the total interaction energy increases strongly during the separation of the two dimers, until reaching ~0 kJ.mol^{-1} when the tetramer is completely disassembled (Fig. 2.11D, t = 2970 ps). This highlights that despite the application of an external force on the LDH-B system, the presence of interactions maintains for a while the assembly of the two

dimers into a tetramer. The decomposition of the total interaction energy into short-range Lennard-Jones (Fig. 2.11C in light grey) and short-range Coulomb (Fig. 2.11C in black) energy contributions allows a better understanding of the nature of the interactions present at the tetrameric interface of LDH-B protein. The contribution of the Coulomb energy term (representative of the electrostatic interactions) is larger than the Lennard-Jones energy term (including the vdW interactions) to the total interaction energy. The interactions established at the tetrameric interface will be investigated later.

As mentioned before, two different tetramerization sites, sites 1 and 2, have been highlighted to be essential for the tetramerization of LDH-B. To complete the characterization of the tetramerization site 2, which is the object of this chapter, an energy analysis of the residues constituting the 22 amino-acid α-helix is presented in Table 2.4. The values correspond to the average and standard deviation (SD) of the interaction energy between each residue of monomer A (from the fixed AC dimer) and the mobile BD dimer during the SMD simulation. The energy values for residues of monomer C are reported in Appendix 2E-1 and present the same trend as for monomer A.

Table 2.4: Mean interaction energy of the residues from the 22-amino acid α-helix with BD dimer calculated over the first 2000 ps of the SMD trajectory. The different contribution to the total interaction energy ($kJ.mol^{-1}$) for monomer A is presented: short-range Coulomb energy contribution ($kJ.mol^{-1}$) and short-range Lennard-Jones energy contribution ($kJ.mol^{-1}$). The total interaction energy ($kJ.mol^{-1}$) corresponds to the sum of Coulomb and Lennard-Jones energy contributions

	Coulomb contribution ($kJ.mol^{-1}$)	**Lennard-Jones contribution ($kJ.mol^{-1}$)**	**Total interaction energy ($kJ.mol^{-1}$)**
Leu54	-0.11 ± 0.01	-0.56 ± 0.02	-0.67 ± 0.02
Glu55	-0.73 ± 0.02	-1.47 ± 0.03	-2.20 ± 0.03
Asp56	-4.79 ± 0.64	-12.93± 0.32	-17.72 ± 0.72
Lys57	4.33 ± 0.65	-14.75 ± 0.82	-10.42 ± 1.05
Leu58	-0.41 ± 0.03	-1.62 ± 0.04	-2.03 ± 0.05
Lys59	2.43 ± 0.18	-10.67 ± 0.09	-8.24 ± 0.20
Gly60	-5.03 ± 0.22	-19.29 ± 0.12	-24.32 ± 0.25
Glu61	-186.52 ± 4.6	-4.97 ± 0.61	-191.49 ± 4.60
Met62	-0.79 ± 0.02	-2.54 ± 0.05	-3.33 ± 0.05
Met63	0.66 ± 0.16	-21.07 ± 0.62	-20.41 ± 0.64
Asp64	-281.12 ± 0.99	-21.38 ± 0.54	-302.50 ± 1.13
Leu65	-0.75 ± 0.03	-16.45 ± 0.24	-17.20 ± 0.24
Gln66	-0.51± 0.87	-14.68 ± 0.52	-15.19 ± 1.01
His67	-51.24 ± 1.20	-43.24 ± 1.70	-94.48 ± 2.08
Gly68	0.16 ± 0.04	-13.07 ± 0.16	-12.91 ± 0.16
Ser69	-4.74 ± 0.63	-13.77 ± 0.64	-18.51 ± 0.89
Leu70	0.28 ± 0.44	-39.98 ± 0.44	-39.70 ± 0.62
Phe71	-8.31 ± 0.36	-45.93 ± 0.30	-54.24 ± 0.47
Leu72	-6.93 ± 0.39	-9.96 ± 0.15	-16.89 ± 0.42
Gln73	-8.63 ± 3.10	-22.15 ± 0.45	-30.78 ± 3.13
Thr74	-1.71 ± 0.11	-3.61 ± 0.25	-5.32 ± 0.27
Pro75	-4.02 ± 0.15	-17.93 ± 0.56	-21.95 ± 0.58

The total interaction energy values allow to highlight the most important residues of the 22-amino acid α-helix involved in the assembly of the two dimers (Table 2.4 in green). Indeed, Glu61, Asp64, His67, Leu70 and, Phe71 largely contribute to the interaction energy. The decomposition of this energy into Coulomb and Lennard-Jones terms reveals the variable contribution of these residues: the Coulomb energy contribution of the charged Glu61 and Asp64 to the total interaction energy is larger, while the Lennard-Jones term contributes predominantly to the total energy for the neutral and apolar Leu70 and Phe71 residues. Concerning the His67 residue, the Coulomb and Lennard-Jones terms contribute in a similar way to the total energy. Therefore, Glu61 and Asp64 residues are more involved in stabilizing electrostatic interactions, including salt bridge and hydrogen bond established with the opposite BD dimer. Leu70 et Phe71 residues are more involved in stabilizing vdW interactions with BD dimer. Finally, His67 is involved in stabilizing hydrogen bond and vdW interactions.

Based on crystallographic data presented previously, and the chemical environment of the five residues identified in Table 2.4, the list of interactions in which the residues are involved is reported in Table 2.5 and Figure 2.12. The energy profiles and mean interaction energy for these interactions involving Glu61, Asp64, His67, Leu70 and Phe71 of monomer C are reported in Appendix 2E-2 and 2E-3 and present the same trend than for the interactions presented in this section.

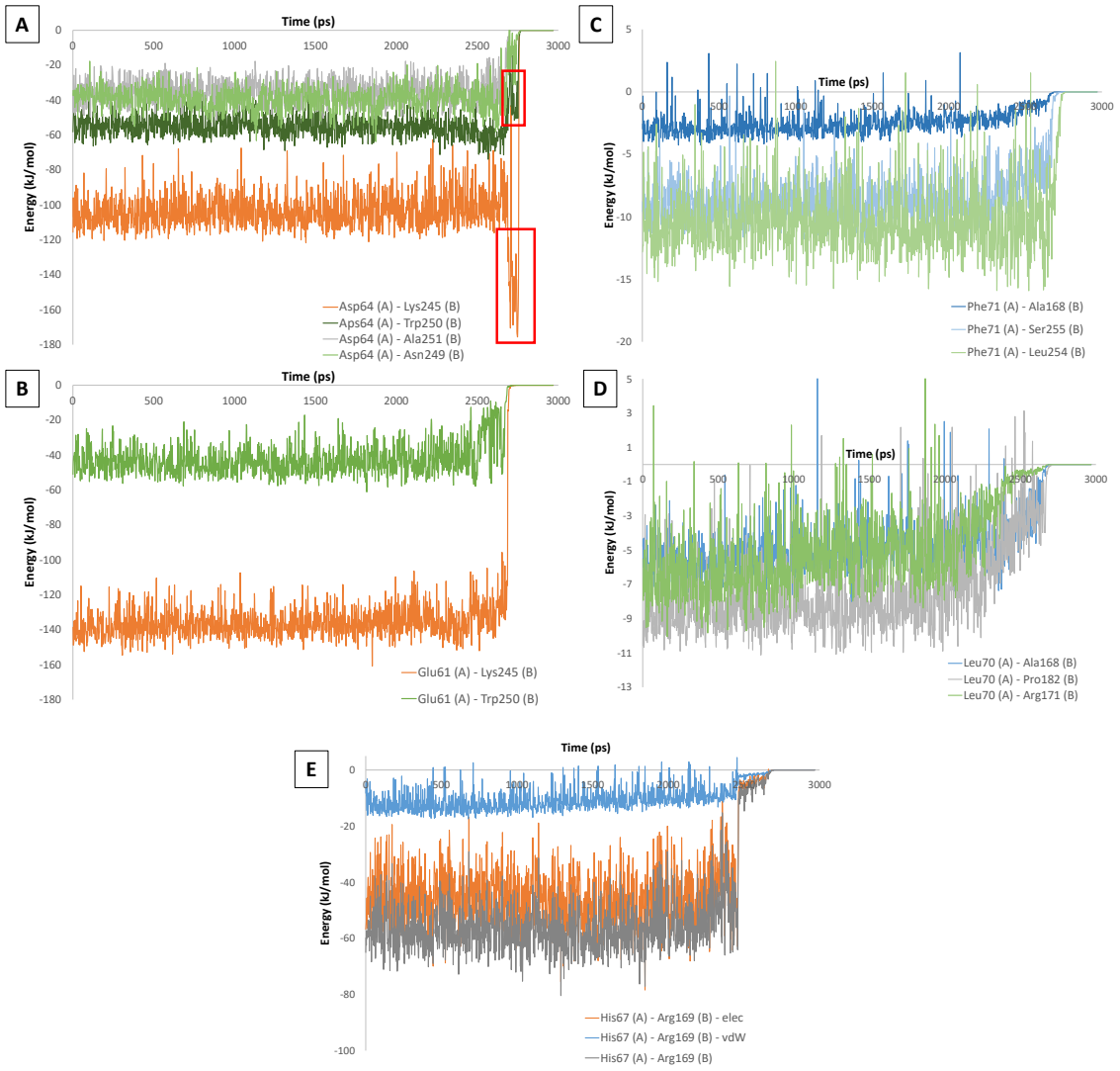


Figure 2.12: Energy profiles for interactions established between Glu61 (A), Asp64 (A), His67 (A), Leu70 (A), and Phe71 (A) residues and the BD dimer – (A) Energy profiles (total interaction energy ($kJ.mol^{-1}$)) for interactions between Asp64 (A) and the BD dimer – (B) Energy profiles (total interaction energy ($kJ.mol^{-1}$)) for interactions between Glu61 (A) and the BD dimer – (C) Energy profiles (total interaction energy ($kJ.mol^{-1}$)) for interactions between Phe71 (A) and the BD dimer – (D) Energy profiles (total interaction energy ($kJ.mol^{-1}$)) for interactions between Leu70 (A) and the BD dimer – (E) Energy profiles (total interaction energy ($kJ.mol^{-1}$) in grey, Coulomb energy contribution in orange ($kJ.mol^{-1}$) and Lennard-Jones energy contribution ($kJ.mol^{-1}$) in blue) for the interaction between His67 (A) and the BD dimer

Table 2.5: Mean interaction energy for different interactions between Glu61 (A), Asp64 (A), His67 (A), Leu70 (A), and Phe71 (A) from the 22-amino acid α-helix and residues of the BD dimer. The energy values were calculated over the first 2000 ps of SMD trajectory

	Coulomb contribution ($kJ.mol^{-1}$)	**Lennard-Jones contribution ($kJ.mol^{-1}$)**	**Total energy ($kJ.mol^{-1}$)**
Glu61 (A) – Lys245 (B)	-140.72 ± 0.73	3.51 ± 0.29	-137.21 ± 0.79
Glu61 (A) – Trp250 (B)	-41.67 ± 0.53	-2.49 ± 0.45	-44.16 ± 0.70
Asp64 (A) – Lys245 (B)	-109.97 ± 0.77	5.59 ± 0.27	-104.38 ± 0.82
Asp64 (A) – Asn249 (B)	-40.23 ± 0.78	0.80 ± 0.40	-39.43 ± 0.88
Asp64 (A) – Trp250 (B)	-47.61 ± 0.49	-6.90 ± 0.34	-54.51 ± 0.60
Asp64 (A) – Ala251 (B)	-30.99 ± 0.22	-4.52 ± 0.18	-35.51 ± 0.28
His67 (A) – Arg169 (B)	-45.97 ± 0.68	-11.46 ± 0.47	-57.43 ± 0.83
Leu70 (A) – Pro182 (B)	0.42 ± 0.13	-8.62 ± 0.03	-8.20 ± 0.13
Leu70 (A) – Arg171 (B)	4.14 ± 0.15	-9.83 ± 0.46	-5.69 ± 0.48
Leu70 (A) – Ala168 (B)	-0.17 ± 0.03	-5.05 ± 0.18	-5.22 ± 0.18
Phe71 (A) – Ala168 (B)	0.69 ± 0.01	-3.42 ± 0.11	-2.73 ± 0.11
Phe71 (A) – Leu254 (B)	1.06 ± 0.06	-11.53 ± 0.06	-10.47 ± 0.09
Phe71 (A) – Ser255 (B)	-0.74 ± 0.04	-8.14 ± 0.03	-8.88 ± 0.05

The total interaction energy profiles for the interactions involving Glu61 (A), Asp64 (A), Leu70 (A), and Phe71 (A) residues are shown in Figure 2.12 A-D. For the His67 (A)-Arg169 (B) interaction, the total interaction energy observed during the SMD simulation, as well as the contributions of the electrostatic and vdW interactions, is shown in Figure 2.12E. As mentioned before, Glu61 and Asp64 are mainly involved in stabilizing electrostatic interactions, such as hydrogen bonds or salt bridges. The energy for the two interactions with Glu61 residue increases from 2670 ps to reach ~0 $kJ.mol^{-1}$ at ~2760 ps. The profiles for the two interactions Asp64 (A)-Asn249 (B) and Asp64 (A)-Ala251 (B) are similar. The energy increases from ~2690 ps to reach ~0 $kJ.mol^{-1}$ at 2760 ps. The energy profiles for the two other interactions involving Asp64 (Asp64 (A)-Lys245 (B) and Asp64 (A)-Trp250 (B)) are also similar (the energy reaches a value of about 0 $kJ.mol^{-1}$ at ~2780 ps). Interestingly, there is a decrease in the interaction energy between 2690 and 2750 ps (Fig. 2.12A – red boxes). This decrease can be explained by the loss of interactions between Glu61 (A) and Lys245 (B) or Trp250 (B). Therefore, it is assumed that a rearrangement occurs at the level of the part-2 of tetramerization site 2 (residue Lys245 (B) to Trp250 (B)) which interacts with Asp64 (see Appendix 2F).

Regarding the stabilizing vdW interactions between Leu70 (A)/Phe71 (A) and the BD dimer, two interaction clusters have been identified according to their disappearance time during the SMD simulation (Table 2.6). The interactions of cluster 1 tend to disappear first during the SMD simulation. These interactions are involved in the maintenance of the LDH-B tetramer but do not induce its disassembly when they disappear (see

Fig. 2.11D, t=2500 ps), in comparison with the interactions of cluster 2. Indeed, the vdW interaction between Phe71 (A) and Leu254 (B)/Ser255 (B) tend to maintain the tetramer of LDH-B over larger times together with the interactions involving Glu61 and Asp64. The classification of the vdW interactions between Leu70 (C)/Phe71 (C) and the BD dimer in two clusters is reported in Appendix 2E-4.

Table 2.6: Characterization of vdW interactions between Leu70 (A)/Phe71 (A) and BD dimer

	Interactions	**$t_{\text{interaction disappears}}$ (ps)**	**$t_{\text{energy reached 0}}$ (ps)**
Cluster 1	Leu70 (A) – Pro182 (B)	2100	2730
	Leu70 (A) – Arg171 (B)	2086	2704
	Leu70 (A) – Ala168 (B)	2470	2720
	Phe71 (A) – Ala168 (B)	2446	2700
Cluster 2	Phe71 (A) – Leu254 (B)	2656	2778
	Phe71 (A) – Ser255 (B)	2652	2760

Moreover, when analyzing the energy profile of His67 (A) (Fig. 2.12E), involved in hydrogen bond and vdW interaction with Arg169 (B), the energy starts to increase from 2460 ps to reach ~0 $kJ.mol^{-1}$ at 2718 ps. As described previously, these interactions are involved in the maintenance of the LDH-B tetramer but do not induce its disassembly when they disappear.

c) Energy analysis of LDH-B mutants during SMD simulation

To further characterize the effect of the five residues (Glu61, Asp64, His67, Leu70, and Phe71) on the rupture force during the SMD simulation, the corresponding mutant proteins are studied. Starting from the LDH-B structure (PDB entry: 1I0Z), the mutation for each of the five residues was inserted using Pymol software[125]. After energy minimization and equilibration steps, the SMD simulations were performed with the mutants LDH-B^{E61A}, LDH-B^{D64A}, LDH-B^{H67A}, LDH-B^{L70A}, and LDH-B^{F71A}. SMD simulations were also performed on LDH-B (PDB entry: 1I0Z) and LDH-Btr (also obtained from LDH-B structure) as controls.

As shown in Figure 2.13 (A), the force increases linearly with time until the breaking point is reached. It corresponds to the separation of the two LDH-B dimers. Depending on the mutant studied, the rupture force varies ($F_{\text{LDH-Btr}} < F_{\text{LDH-B (D64A)}} < F_{\text{LDH-B (E61A)}} \sim F_{\text{LDH-B (F71A)}} < F_{\text{LDH-B (H67A)}} < F_{\text{LDH-B (L70A)}} < F_{\text{LDH-B}}$). This force profile can be correlated with the profile of COM position as a function of time (Fig. 2.13B). The energy profile between dimers during the separation trajectory is presented in Figure 2.13 (C). As for Figure 2.11 (C), a similar analysis of the energy profiles can be performed for the different

LDH-B mutants. Depending on the mutated residue, the interaction energy between the two dimers of the system differs from the energy of LDH-B system and reached ~0 kJ.mol^{-1} at different times during the simulation. All these data highlight the importance of residues that have been mutated and their involvement in maintaining the tetrameric state of the protein. Indeed, the mutations of Glu61, Asp64 and Phe71 will further facilitate the disassembly of the tetramer in comparison with the residue His67 and Leu70.

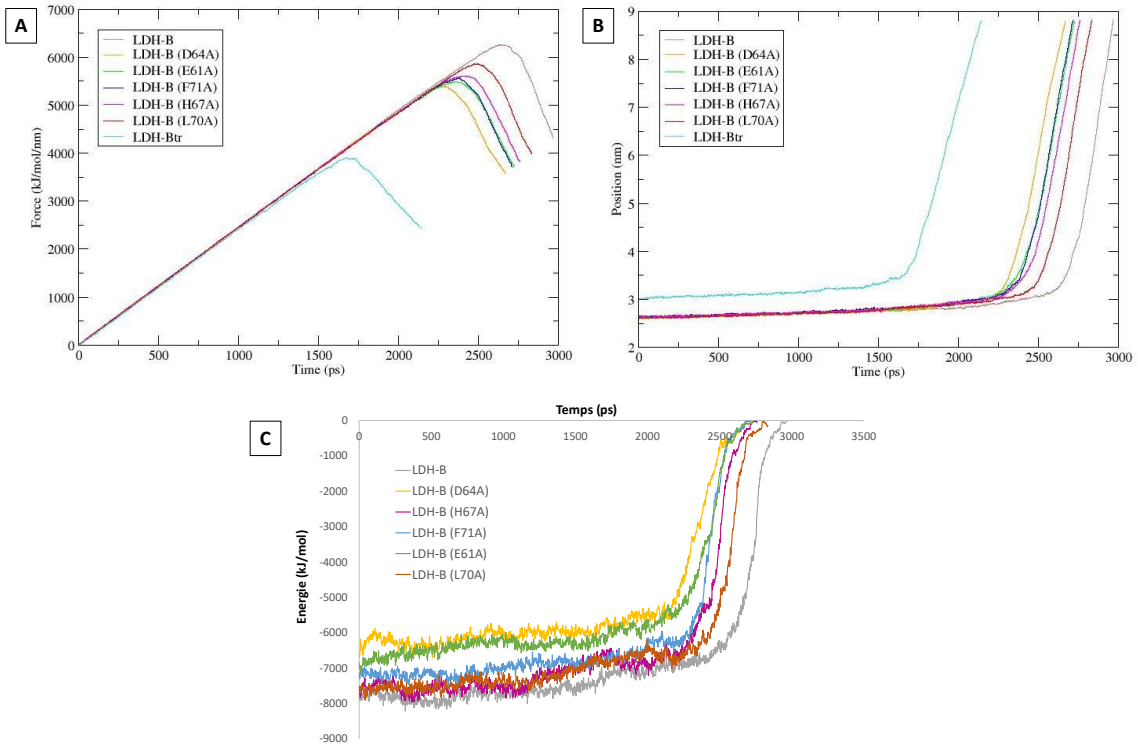


Figure 2.13: SMD simulation with LDH-B, LDH-B mutant and LDH-Btr systems (LDH-B, LDH-B^{E61A}, LDH-B^{D64A}, LDH-B^{H67A}, LDH-B^{L70A}, LDH-B^{F71A} and LDH-Btr) – (A) Profile of the measured force (kJ.mol^{-1}.nm^{-1}) as a function of time (ps) – (B) Profile of the COM position (BD dimer) (nm) as a function of time (ps) – (C) Energy profile of BD dimer interacting with AC dimer during the SMD simulation: the total interaction energy is presented (kJ.mol^{-1})

2.4. Conclusions – SMD

The Steered Molecular Dynamics method has allowed to better characterize the self-disassembly of LDH-B into dimers, and more particularly the 22-amino acid α-helix interacting with the tetramerization site 2. Based on these SMD simulations, different information is highlighted.

The contribution of the short-range Coulomb energy term to the total AC/BD interaction energy for the entire LDH-B system is larger than the short-range Lennard-Jones energy term. This highlights that the electrostatic interactions contribute more than vdW

interactions in maintaining the tetrameric state of LDH-B protein. This oligomerization state is allowed by the existence of the tetrameric interface composed of two tetramerization sites: the tetramerization site 1 with which the tetramerization arm interacts and the tetramerization site 2 interacting with the 22-amino acid α-helix.

The SMD simulations have been performed to better characterize the second tetramerization site interacting with the 22-amino acid α-helix. The analysis of interaction energy (between AC and BD dimers) allowed to identify five residues (Glu61, Asp64, His67, Leu70, and Phe71) being more important for the assembly of the two dimers into a tetramer. These results correlated well with the experimental ones in solution, i.e., Glu61, Asp64, Leu70, and Phe71 which have been identified as hotspots, defined as essential for the tetramerization of the protein[50].

Even if the electrostatic interactions are stronger and disappear later during the simulation, the Lennard-Jones energy contribution (including vdW interactions) must not be neglected, especially for the Phe71 residue. These interactions break down at the same time as the electrostatic interactions, highlighting the contribution of Phe71 in tetramer assembly. Moreover, from an energy point of view, the interactions established between Glu61 (A)/Asp64 (A) with Lys245 (B) seem essential for the tetramer assembly. This correlates with crystallographic data. Indeed, when Asp64 is mutated in Ala, a part of the D64-interacting α-helix is not stabilized.

Finally, the study of mutated LDH-B systems by SMD simulations allowed to confirm the importance of the mutated residue in the LDH-B tetramer stability. The results suggested that Glu61, Aps64 and Phe71 residues are more involved in the tetramer stability than the residues His67 and Leu70 (in agreement with experimental data). Indeed, when the residues Glu61, Aps64 and Phe71 are mutated to Ala, the mutant LDH-B protein is mainly in a dimeric form in solution (LDH-B^{E61A} and LDH-B^{F71A}) and very destabilized (for the three mutants). In addition, the mutation of His67 in LDH-B protein has a lower effect on the protein quaternary structure (lower chemical and thermal destabilization). Nevertheless, analysis of L70A mutation shows to have a higher effect in solution compared to what is observed in Molecular Dynamics simulation.

3. Conclusions

The first part of this chapter (section 1) consists in crystallographic studies of three mutants identified as promising candidates for characterizing the tetramerization site 2. This study allowed to confirm some data obtained in solution and presented in Table 2.1 (such as for LDH-B^{D64A} and LDH-B^{H67A} mutants)[50]. Nevertheless, this approach is not sufficient to understand the influence of mutations on the LDH-B system, such as conformation change and oligomerization state. Indeed, the crystallography requires working at high protein concentrations, which is not always representative of the native protein oligomerization state in solution.

Therefore, the crystallographic study can be complemented by a Molecular Dynamics approach. First, the LDH-B protein and LDH-B mutants (LDH-B^{D64A}, LDH-B^{H67A} and LDH-B^{F71A}) have been studied from a 300 ns-production. These data allowed to confirm some data shown by crystallography (such as the unstabilized part of D64-interacting α-helix for LDH-B^{D64A} and no major conformational change for LDH-B^{H67A}). Nevertheless, the effect of the F71A mutation on the behavior of LDH-B^{F71A} has not been elucidated. Finally, the study of the LDH-B system by SMD allowed a better understanding of the self-disassembly of tetrameric LDH-B. Five residues have been highlighted to be essential for the tetramer stability of LDH-B protein. Four of them correspond to hotspots that have been highlighted in solution[50].

Part III

Structural study of LDH-B - ligand complexes - Identification of molecules interacting at the tetrameric interface

1. Identification of small molecules interacting at the tetrameric interface of LDH proteins

This chapter is a collaborative work with Prof. R. Frédérick's team (UCLouvain) which aims at identifying small molecules interacting at the tetrameric interface of LDH enzymes. Therefore, this chapter will consist in the analysis of crystallographic structures of protein-small molecule complexes. Structural data will provide more information on the binding mode of identified compounds.

The strategy used for the identification of small molecules, from a chemical library and interacting at the tetrameric interface of LDH enzymes, is described in the publication of UCLouvain collaborators[79]. The screening strategy consists of three steps:

- Screening by nanoDSF: this first step determines the effect of compounds on the thermal stability profile on both dimeric truncated (LDH-Btr) and tetrameric (LDH-A and LDH-B) LDH proteins. Selected compounds destabilize the full-length proteins and have no effect or increase the thermal stability of the dimeric LDH-Btr.
- Hit validation by MicroScale Thermophoresis (MST) and Saturation-Transfer Difference Nuclear Magnetic Resonance (STD NMR): the second step consists of evaluating if compounds (at 1 mM) can bind the LDH-Btr.
- Hit characterization:
    1) Ligand affinity determination (K_d) for both tetrameric and dimeric LDH using MST. Ligands that would interact at the tetrameric interface should have a lower affinity for the tetrameric enzyme compared to the dimeric.
    2) Assessment of ligand that prevents LDH tetramer association using a protein renaturation experiment[24] (acidic conditions) of LDH-B.

3) Epitope mapping of ligands by STD-NMR using LDH-Btr. This experiment aims to identify important parts of ligand for interaction with protein.
4) Competition experiments by STD-NMR: identification of ligands that compete with macrocycle 7 and LP-22 peptides that have been previously identified to interact with first and second tetramerization sites, respectively (see Introduction section 5.2)[50,59]. This experiment aims to identify on which protein site the ligand interacts.

Each step of this screening allows to refine hit selection and finally to identify the best candidates.

Four promising ligands have been identified by the UCLouvain collaborators, i.e. maprotiline, imipramine, triprolidine and fluoxetine, using the previously described strategy and were used for crystallization assays (see details in Appendix 3A). The structure of compounds, for which a crystallographic structure with LDH-B has been obtained by co-crystallization, are presented in Figure 3.1.

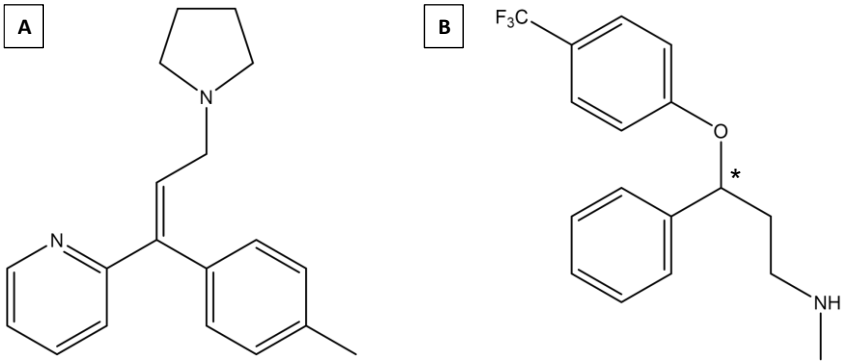


Figure 3.1: Structure of triprolidine (A) and fluoxetine (B) (* corresponds to the asymmetric carbon).

Triprolidine (TRI) showed a low affinity for LDH-B and LDH-A tetrameric proteins (K_d > 1 mM and 760 µM ± 357 µM, respectively), while a higher affinity is observed for LDH-Btr (K_d = 79 ± 37 µM). In addition, in presence of triprolidine, a destabilization of the tetrameric LDH-A and LDH-B is observed (ΔT_m = -3.3°C (LDH-A) and ΔT_m = -3.1°C (LDH-B)), while a stabilization of the dimeric LDH-Btr is observed (ΔT_m = 0.2°C)[79]. The ability of the compound to prevent the LDH tetramer association is also assessed via a renaturation assay as described in the work of Nadal-Bufi *et al.*[24]. In a first time, the tetrameric protein is partially denatured in acidic conditions (pH = 2.5), before recovering the physiological pH (pH = 7.4) in presence or absence of selected ligands (allowing the

subunits to refold and to reassociate into a tetramer). Subsequently, an IC_{50} value can be determined by evaluating the enzymatic activity of the tetrameric protein in the presence or absence of the compound (IC_{50} = 7.6 mM)[79]. Concerning the fluoxetine compound, experiments are ongoing, and data are not published. Nevertheless, this compound presents a similar profile (similar properties) to the triprolidine molecule.

2. Crystallization of LDH-B-small molecule complexes: structural analysis

2.1. LDH-B-triprolidine

a) Overall structure

The structure of LDH-B in complex with triprolidine (TRI) was solved and refined at a resolution of 2.68 Å (data completeness: 99.78 %) in the orthorhombic space group $P2_12_12_1$. Other statistics obtained for the structure can be found in Appendix 3B. The asymmetric unit (ASU) is composed of four monomers (A, B, C and D) forming a tetramer. Each chain composing the protein was completely refined (Ala1 to Leu333). Nevertheless, the length of the chains varies due to the partial refinement of the C-terminal extremity (TEV cleavage site + 6xHis-tag): Ala1 to Glu334 (chains A, B and C); Ala1 to Asn335 (chain D). LDH-B-triprolidine complex presents the same main structural features as LDH-B (see Part I – section 5(1)). Therefore, the apoprotein was crystallized (without substrate and cofactor), resulting in an open conformation for the active site loop. Moreover, one molecule of PEG and 13 molecules of glycerol from the crystallization solution are present in the structure, as well as two molecules of triprolidine.

b) LDH-B-triprolidine interactions

Two molecules of triprolidine are present in the newly solved structure. The first (TRI1) is located in the active site. This is probably due to the absence of substrate and cofactor. The second (TRI2) is interacting close to the tetramerization arm of monomer A (Figure 3.2). TRI2 has an occupancy of 77%, a B factor of 53.99 $Å^2$ (protein average B factor: 44.26 $Å^2$) and a CC-value (Correlation Coefficient) of 0.774. These parameter values support the presence of TRI2 in the structure. Indeed, the B factor value for a ligand should be as close as possible to the protein average B factor. In addition, the CC-value represents whether the ligand correlates with electron density. This value should be as high as possible: a minimum value of 0.6 is expected for a ligand.

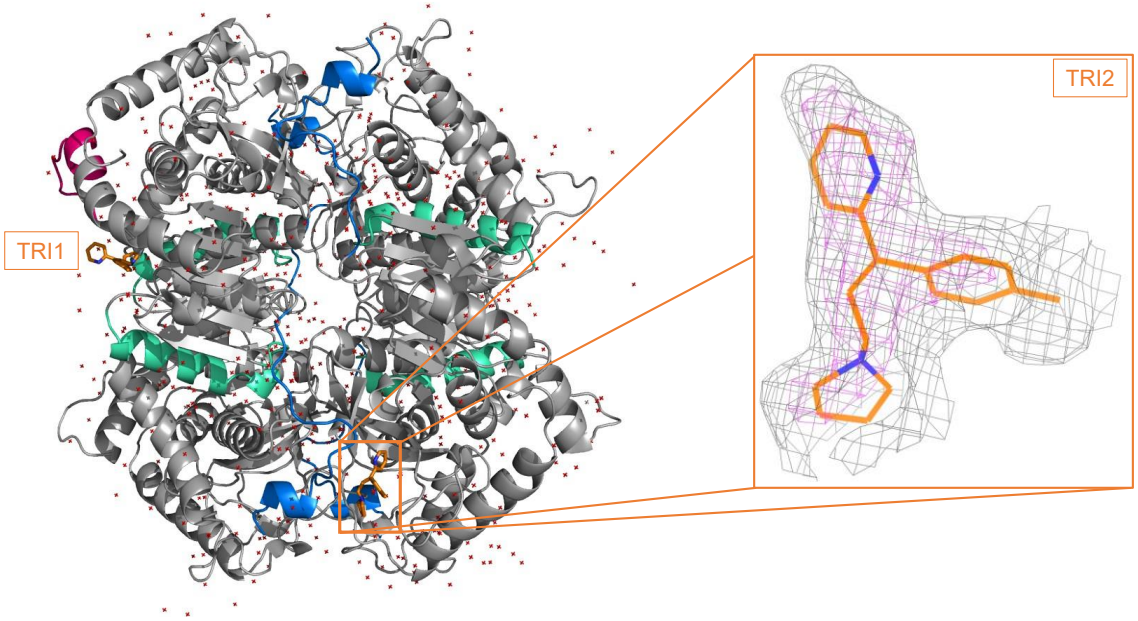


Figure 3.2: Structure of LDH-B in complex with triprolidine (TRI). Tetramerization arms are represented in blue, 22-amino acid α-helix in green, active site loop (monomer A) in magenta and triprolidine molecules are in orange sticks (TRI1 and TRI2) – zoom on TRI2 (orange box): superposition of the 2Fo-Fc map in magenta mesh (σ = 0.7 and carve = 1.6) with the Fo-Fc polder map (σ = 3.0 and carve = 2.3)

As mentioned before, TRI2 is located close to the tetramerization arm of monomer A (Fig. 3.3A). In addition to van der Waals (vdW) interactions established between TRI2 and the tetramerization arm, one hydrogen bond is observed between the pyridine group of ligand and Lys155 from monomer C. The ligand also interacts via a pi-cation interaction with the amine group of Lys155 (C) sidechain. The interaction site of TRI2 with LDH-B is shown in Figure 3.3(B) using the surface representation. Details of interactions presented in Figure 3.3(A) are shown in Appendix 3C.

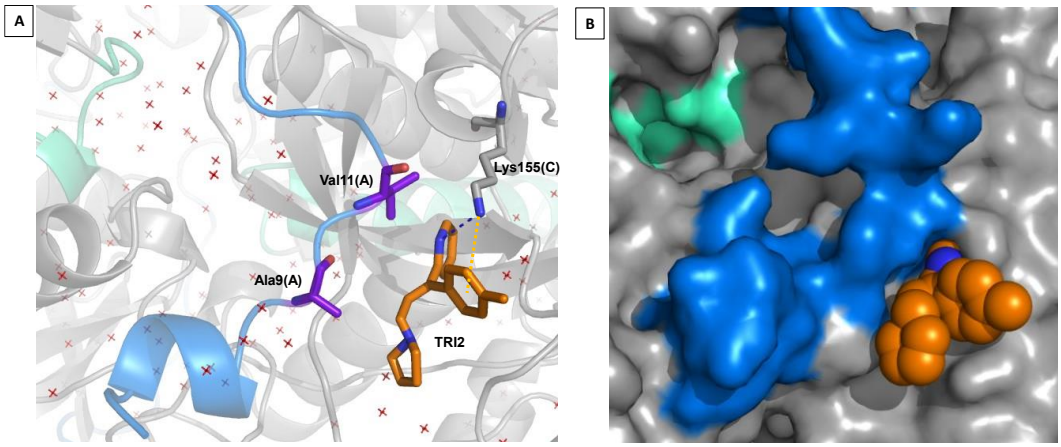


Figure 3.3: (A) Representation of LDH-B residues that interact with TRI2 in LDH-B-triprolidine complex structure. Hydrogen bond and pi-cation interaction between Lys155 (C) and pyridine group and phenyl ring are represented by blue and yellow dotted lines respectively, water molecules are in red and the tetramerization arm in blue – (B) Surface representation for LDH-B protein with TRI2 shown as spheres

2.2. LDH-B-fluoxetine

a) Overall structure

The structure of LDH-B in complex with fluoxetine (FX) was solved and refined at a resolution of 2.07 Å (data completeness: 98.39 %) in the monoclinic space group $P2_1$. Other statistics obtained for the structure can be found in Appendix 3B. The asymmetric unit (ASU) is composed of 12 monomers (A to L). Only one tetramer is found in the ASU and the others (four in total) are formed by symmetry (see Appendix 3D). Each chain composing the protein was completely refined (Ala1 to Leu333). Nevertheless, the length of the chains C, F, H, J and K varies due to the partial refinement of the C-terminal extremity (TEV cleavage site + 6xHis-tag): Ala1 to Glu334. The LDH-B-FX complex presents the same main structural features as LDH-B (see Part I – section 5(1)). Therefore, the protein was crystallized in presence of an analog of the substrate and with the cofactor, resulting in a closed conformation of the active site loop. Moreover, eight molecules of Bis-Tris, eight molecules of PEG and 21 molecules of glycerol from the crystallization solution are present in the structure, as well as eight molecules of fluoxetine. Compared to LDH-B-TRI complex crystallized in $P2_12_12_1$, the structure with FX was crystallized in the monoclinic $P2_1$ space group. The loss of crystal symmetry is probably related to the presence of FX in the crystal lattice.

b) LDH-B-fluoxetine interactions

LDH-B-FX is composed of 12 monomers, and we can observe one FX molecule in eight of them. The representation of a tetramer (A-B-G-H) allows to locate FX ligands (FX-A, FX-B, FX-C, FX-D, FX-F, FX-G, FX-I and FX-K) close to the tetramerization arm (Fig. 3.4). The presence of each of FX ligands could be assessed based on different parameters presented in Table 3.1. B factor values for each FX ligand (ranging from 45 to 60 $Å^2$) were compared with the average B factor of the structure (43.01 $Å^2$). Occupancy of FX ligands is between 60 et 70%. Moreover, fluoxetine has an asymmetric carbon, and the complex was crystallized in presence of a racemic mixture of fluoxetine. Therefore, during the refinement process and when adding FX ligands to the structure, both enantiomers were tested and the data for the two enantiomers are presented in Table 3.1. Nevertheless, the quality of the structure did not allow for discrimination against one enantiomer from the other.

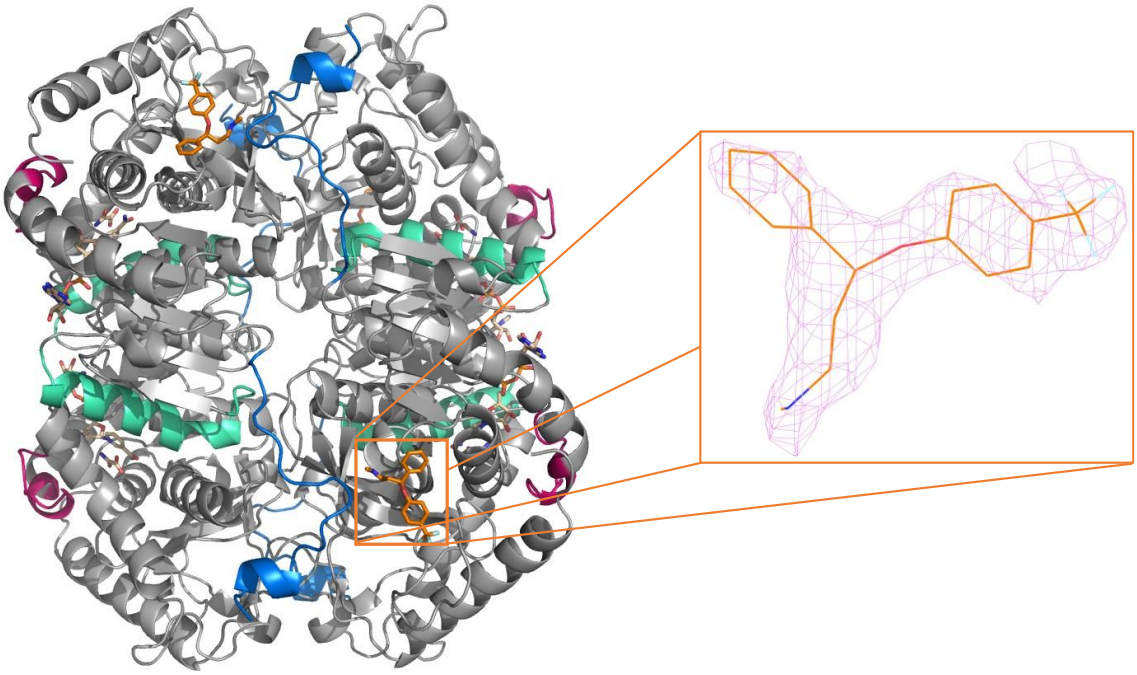


Figure 3.4: Structure of LDH-B in complex with fluoxetine – representation of tetramer composed by monomers A, B, G and H. Tetramerization arms are represented in blue, 22-amino acid α-helix in green, active site loop in magenta, substrate analog (OXM) and cofactor (NADH) in salmon and fluoxetine (FX) molecules are in orange sticks (FX-A, FX-B and FX-G) – zoom on FX-B (orange box): representation of the 2Fo-Fc map in magenta mesh (σ = 0.7 and carve = 1.6)

Table 3.1: Evaluation of the presence of ligands in the structure LDH-B-FX. A value for each parameter was assigned for each FX ligand. Fluoxetine is found in the following monomers: monomer A (FX-A), monomer B (FX-B), monomer C (FX-C), monomer D (FX-D), monomer F (FX-F), monomer G (FX-G), monomer I (FX-I) and monomer K (FX-K). Parameter values for enantiomers S and R are shown from structure refinement 97 and 78, respectively

	B factor (Å^2)	Occupancy	CC-value	Enantiomer
Structure	43.01	/	/	/
FX-A	51.53	0.64	0.707	S
	54.48	0.65	0.570	R
FX-B	46.55	0.65	0.788	S
	49.65	0.64	0.755	R
FX-C	52.84	0.68	0.713	S
	56.42 / 56.45	0.60 / 0.40	0.760 / 0.756	R
FX-D	54.19	0.62	0.723	S
	56.39	0.61	0.633	R
FX-F	52.12	0.69	0.734	S
	54.24	0.64	0.641	R
FX-G	54.44	0.66	0.710	S
	53.47	0.66	0.640	R
FX-I	53.56	0.68	0.723	S
	52.35	0.68	0.743	R
FX-K	59.63	0.67	0.735	S
	61.62	0.69	0.729	R

When superimposing monomers containing a FX molecule, we were able to observe two FX ligand clusters in the structure (Fig. 3.5(B-C)). Indeed, it has been shown that different orientations of FX ligands were possible, depending on the electron density observed. The first cluster (in orange in Fig. 3.5(A)) includes FX ligands oriented with their amine group pointing to Glu13 and the fluorinated ring directed to Glu286. The second one (in cyan in Fig. 3.5(A)) includes FX ligands that are oriented with their fluorinated ring towards Glu13.

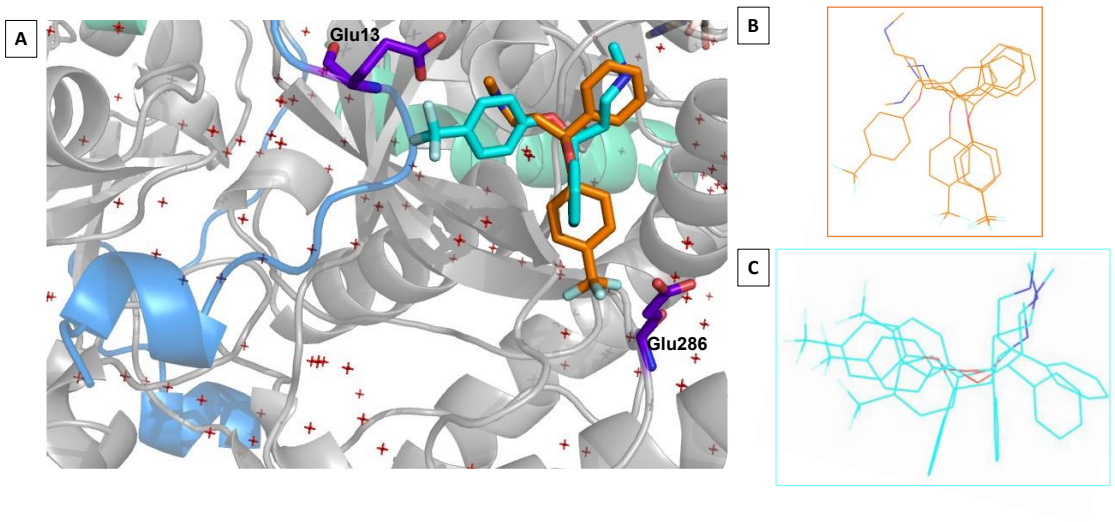


Figure 3.5: Representation of the two FX ligand clusters in LDH-B-FX structure – (A) The clusters correspond to different orientations adopted by FX ligands (FX (cluster 1) in orange sticks and FX (cluster 2) in cyan sticks) – (B) Superposition of FX ligands composing cluster 1 – (C) Superposition of FX ligands composing cluster 2

Therefore, after identifying the two major orientations that FX ligands can adopt, interactions for both FX ligand clusters (clusters 1 and 2) with LDH-B protein were highlighted in Figure 3.6. All the interactions are listed for each of the ligand clusters. Since the ligands within the same cluster show a similar but not identical orientation (Fig. 3.5(B-C)), different interactions are observed. More details about interactions are given in Appendix 3C.

For cluster 1 (Fig. 3.6(A1)), a halogen bond can occur between the fluor atom and Glu286 or Pro10. Moreover, other interactions take place between FX ligand and LDH-B protein, such as a salt bridge between the amine group of FX and Glu13, and a pi-cation interaction between the fluorinated ring of FX and Lys155. Finally, FX is also stabilized by a pi-pi stacking (T-shape) interaction between the fluorinated ring of FX and Trp148.

Regarding cluster 2 (Fig. 3.6(A2)), some of the interactions described for cluster 1 are also present, such as a pi-cation interaction between FX and Lys155 and a pi-pi stacking interaction between FX and the side chain of Trp148. Moreover, other interactions are observed between FX ligand and LDH-B protein: FX ligand is stabilized by different hydrogen bonds with Gly152, Lys155, His156 and water molecules and also by a salt bridge between the amine group of FX and Glu15.

The surface representation of FX in complex with LDH-B for cluster 1 (Fig. 3.6(A2)) and cluster 2 (Fig. 3.6(B2)) highlights the position in which the ligand is inserted and interacts with LDH-B.

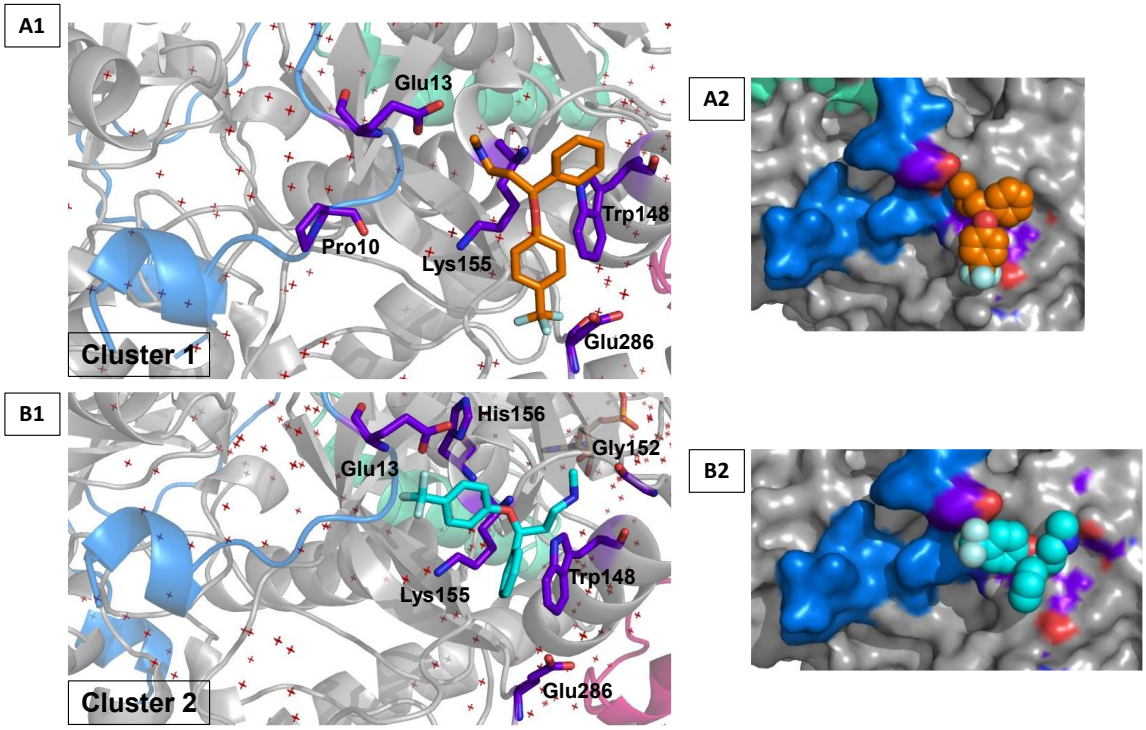


Figure 3.6: (A1-B1) Representation of residues interacting with FX molecules from clusters 1 (A1) and 2 (B1) in LDH-B-fluoxetine complex structure – (A2-B2) Surface representation of FX from clusters 1 in orange (A2) and 2 in cyan (B2) in complex with LDH-B

3. Characterization of triprolidine and fluoxetine binding sites

Figure 3.7 illustrates the location of triprolidine and fluoxetine ligands in the two LDH-B crystallographic structures that have been solved and described in the previous sections.

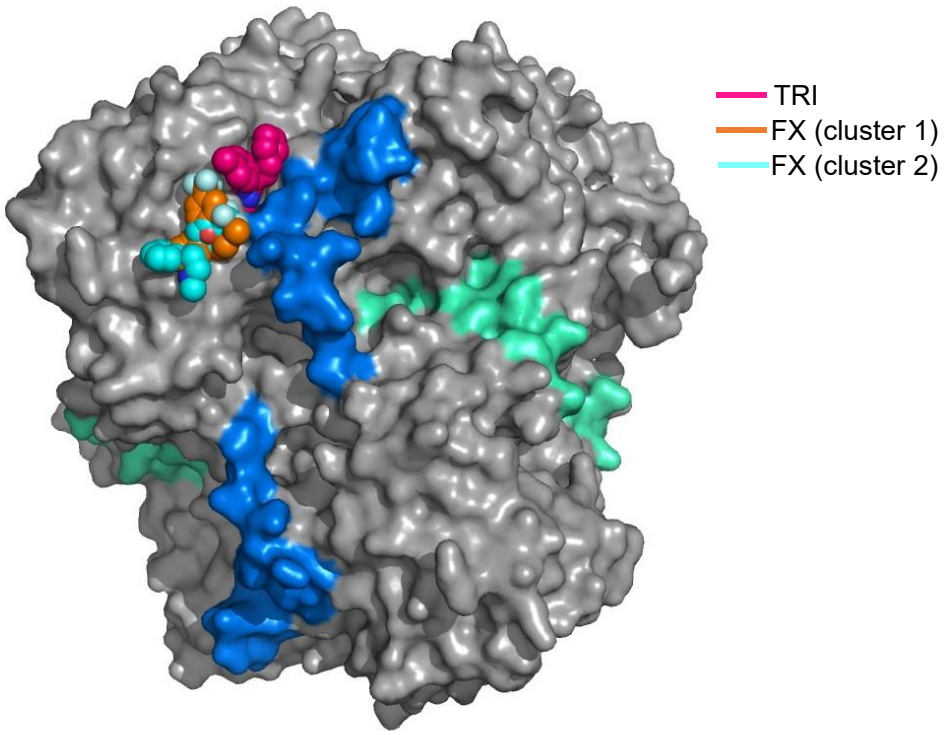


Figure 3.7: Representation of the binding sites of TRI and FX ligands in LDH-B. TRI ligand is shown in pink, FX ligand (cluster 1) in orange and FX ligand (cluster 2) in cyan. Tetramerization arms are represented in blue and 22-amino acid α-helix in green

As observed in Figure 3.7, both ligands are positioned close to the tetramerization arm (blue surface), triprolidine close to the residues Ala9 and Val 11 and fluoxetine close to Glu13. The LDH-B-TRI complex structure is composed of four monomers in the ASU, with one TRI molecule found in monomer A. The LDH-B-FX structure is composed of 12 monomers in the ASU with eight FX molecules present in the complex. The absence of ligands in some monomers of the LDH-B-TRI and LDH-B-FX structures, as well as the occupancy for these ligands (77% for TRI2 and between 60 and 70% for FX) can be explained by the low affinity of ligands for LDH-B (K_d (TRI) > 1 mM[79] and K_d (FX) > 1 mM (data not published)). Moreover, each protein complex was crystallized in the presence of ligands (TRI or FX) at a concentration of 1.2 mM (ligand/protein ratio: 3/1), corresponding to three times the protein concentration. Therefore, it may turn out that the protein may not have been saturated with ligands, resulting in a partial occupation in the protein. Some crystallization assays with a higher ligand/protein ratio (6 times and 100 times for TRI and FX, see Appendix 3A for details) were performed but did not result

in protein crystals. Indeed, the use of a higher ligand concentration, previously identified to have a destabilizing effect on LDH-B, could disrupt the crystal packing of the protein.

TRI and FX ligands have a low affinity for tetrameric LDH, compared to the dimeric LDH-B. Indeed, the screening strategy aims at identifying compounds interacting at the tetrameric interface of LDH. Therefore, the selected compounds should have a lower affinity for the tetrameric enzyme compared to the dimeric LDH-Btr.

In presence of triprolidine ligand, tetrameric LDH presents a destabilization profile (in nanoDSF) similar to the one observed with macrocycle 7[79]. Macrocycle 7 is a stapled peptide designed and highlighted for competing with the N-terminal part of LDH tetramerization arm (residue Ala1 to Ile8)[59]. Therefore, it has been suggested that the TRI ligand could interact in the same way. Crystallographic results correlate with *in vitro* data regarding the binding site of triprolidine. Indeed, the structure of the LDH-B-TRI complex revealed that the ligand binds near the N-term part of the tetramerization arm (Figs. 3.3 and 3.7).

Concerning the fluoxetine ligand, experiments are ongoing. Nevertheless, a competition assay between triprolidine and fluoxetine has been performed and shows no competition between the two ligands (data not published). Crystallographic results correlate with this data: two distinct binding sites have been identified for TRI and FX ligands.

To further characterize the binding site of the TRI and FX ligands, structures of LDH-Btr-TRI and LDH-Btr-FX complexes are necessary. Crystallization assays were performed but did not result in protein crystals formation, although TRI and FX ligands have a stabilizing effect for LDH-Btr. Details of LDH-Btr crystallization assays in presence of TRI and FX ligands are given in Appendix 3A.

General discussion, conclusions & outlooks

1. General discussion and conclusions

This work focused on the study of the Lactate Dehydrogenase B (LDH-B), one of the two main isoforms of LDH, catalyzing the interconversion of pyruvate into L-lactate in the presence of the cofactor NAD at the end of the glycolytic pathway. LDH is at the core of lactate metabolism and was highlighted as a promising therapeutic target for the development of novel cancer therapies. A common approach to target a protein is through the active site. However, because of its active site nature, inhibition of LDH remains challenging. An emerging strategy, based on the active tetrameric quaternary structure of LDH, has been proposed by targeting the oligomeric interface to induce the tetramer disruption. In order to design PPI modulators targeting this interface, more insights regarding the quaternary structure of LDH are required.

1.1. Structural characterization of the tetramerization sites 1 and 2

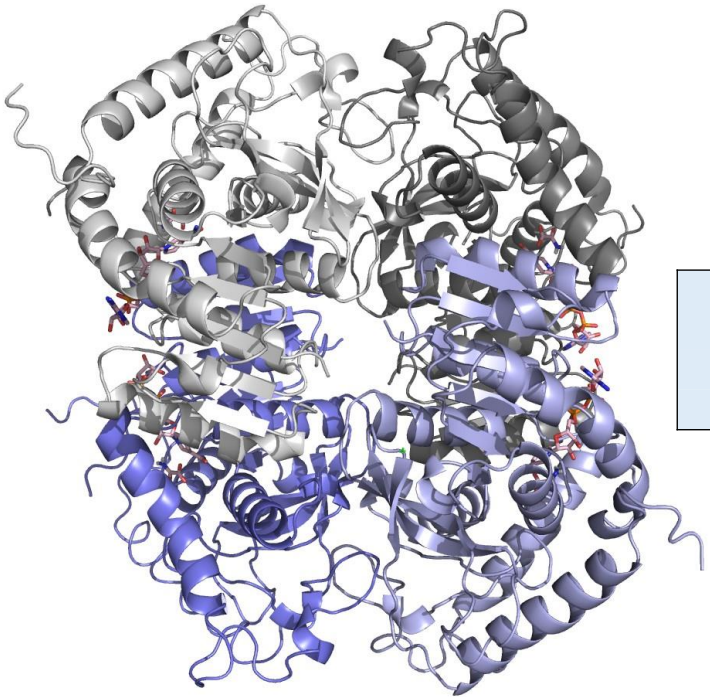
One of the main objectives of this work consisted in a better understanding of the tetramerization process of LDH-B, via the structural characterization of its tetrameric interface, for which two tetramerization sites (tetramerization sites 1 and 2) have been previously identified[50,59].

For this purpose, **part I** consisted in characterizing the tetramerization site 1 via the structural analysis of a truncated form of LDH-B (LDH-Btr) where the 19 N-terminal residues, corresponding to the tetramerization arm, were deleted. First, using the 6xHis-LDH-Btr protein, a crystallization strategy was implemented, based on improving the crystallizability of the protein by acting on the compositional and conformational stabilities of the recombinant protein. Nevertheless, this strategy did not lead to LDH-Btr protein crystals. To overcome the crystallization problem, cofactor and substrate analog were added in crystallization assays and a new plasmid construct coding for LDH-Btr-6xHis was designed, with the following two major changes:

- The position of the 6xHis-tag at the C-terminal extremity of the protein
- Modification of the cleavage site nature, by preferring a TEV site for this second plasmid construct

The modifications made to perform the new crystallization assays, led to protein crystals formation and the solving of the first LDH-Btr structure at a resolution of 2.98 Å (Fig. D.1). This has highlighted that different approaches can be used for promoting the obtention of crystallizable protein systems. The diversity of the available experimental

approaches, as well as the development of IT tools (for predicting the crystallizability of a protein system from the amino acid sequence), allow to improve chances of protein crystallization. Nevertheless, protein crystallization remains still an empirical method based on a trial-and-error procedure.



Resolution (Å)	2.98
Space group	$P2_12_12_1$
a, b, c (Å)	71.53, 85.66, 207.10
α, β, γ (°)	90.00, 90.00, 90.00

Figure D.1: Representation of the crystallographic structure of LDH-Btr

One of the main pieces of information highlighted from the crystallographic structure of the truncated protein concerns the oligomerization state of LDH-Btr. The asymmetric unit of LDH-Btr structure is composed of a tetramer, although the truncation of the tetramerization arm leads to a dimeric protein in solution. The tetramer obtained by crystallography confirmed the presence of an equilibrium between the dimeric and tetrameric forms of LDH-Btr, previously identified in solution[50]. Indeed, the dimeric protein can be self-assembled into a "weak tetramer" in solution at higher concentrations. Therefore, it seems more likely to obtain a tetramer by crystallography, since the protein concentrations used for crystallization assays are high (> 240 µM). The equilibrium is shifted toward the tetrameric form. The assemblage of LDH-Btr into a tetramer is still possible due to the presence of a second tetramerization site (tetramerization site 2) that was discussed in part II.

In addition, obtaining a crystallographic structure of the LDH-Btr protein is an interesting tool for the design and study of new modulators targeting the tetrameric interface of LDH.

Indeed, with the tetramerization arm being truncated, the tetramerization site 1 is accessible for interaction with potential PPI modulators (peptides or small molecules).

In the continuity of characterizing the LDH-B tetrameric interface, **part II** consisted in studying the tetramerization site 2. First, a structural study of selected LDH-B mutants (LDH-B^{D64A}, LDH-B^{H67A} and LDH-B^{F71A}) was performed by X-ray crystallography to further understand the effect of the mutation on the protein system, and consequently the role of each mutated residue in the LDH-B tetramer stability. The different mutants were chosen because of their distinct profiles (oligomerization state, thermal and chemical stabilities) based on the work of Thabault *et al.*[50]. Their work highlighted the presence of four hotspots (defined as residues essential for the stability of the tetrameric LDH-B protein), Glu61, Asp64, Leu70 and Phe71, at the tetrameric interface of LDH-B in solution. Secondly, this structural study was complemented by a molecular dynamics (MD) approach, i.e., with Steered Molecular Dynamics (SMD) method, which aims to better characterize the tetrameric interface through the tetramer self-disassembly of LDH-B into dimers. This approach focused on the 22-amino acid α-helix interacting with the tetramerization site 2.

At the end of this part II, the molecular dynamics simulations have confirmed the four hotspots, previously identified in solution. Indeed, this MD approach suggested that Glu61, Asp64, Leu70, Phe71, but also His67 are more involved in the stability of the tetramer. The role of these residues in the assembly of the LDH-B tetramer has been assessed from an energetic point of view, but also via the SMD dissociation trajectories of the tetramer when the hotspots are mutated into alanine. Furthermore, the crystallographic data, in combination with the molecular dynamics results, suggested that Asp64 (D64) is involved in the formation of a stable tetramer via the stabilization of the D64-interacting α-helix.

Results of parts I and II, presented in this work, highlighted that the two approaches, i.e., molecular dynamics and crystallography, are complementary techniques for the structural characterization of the tetramerization site 2. Steered Molecular Dynamics is an appropriate approach for *in silico* drug design, despite some limitations. Indeed, the choice of the force field, the direction of the applied external force vector or even the magnitude of the force can introduce some errors in simulation results[129]. However, SMD remains a good technique that can be used as a high throughput screening for hotspot

identification to guide mutant selection for experimental characterization. In addition, this technique is interesting for auto-association studies of protein systems.

Concerning protein crystallography, X-Ray Diffraction (XRD) is an approach widely used for the determination of three-dimensional structures of protein. Although dimers of LDH-B were observed in solution for some mutants and the truncated LDH-B (LDH-Btr and LDH-B^{F71A}), tetramers of proteins were obtained during crystallization. This underlines one limit of using crystallography to study protein oligomerization and calls for additional biophysical experimental approaches. Obtaining new structural data through other techniques than XRD, such as cryogenic Electron Microscopy (cryo-EM) would allow a better characterization of the LDH-B oligomerization states (dimeric form) that was observed in the solution. Cryo-EM is a technique that relies on the use of a protein solution that is flash-frozen on a support (grid) in liquid ethane. Frozen protein particles are then pictured by electron microscopy. Bi-dimensional images of particles are classified following the different orientations of the protein (2D classification). At the end, the classification of all these particles allows to generate a 3D volume map to construct the final protein structure[136,137].

Crystallography and cryo-EM could be used in a complementary way. Table D.1 presents some usage advantages/drawbacks of XRD and cryo-EM.

Table D.1: Comparison of X-Ray Diffraction (XRD) and cryogenic Electron Microscopy (cryo-EM) techniques for the resolution of three-dimensional protein structures

Techniques	Advantages	Drawbacks	Resolution
XRD	- High resolution - Large range of molecular weight (MW) - Well developed	- Sample purity - Crystallizable system - Solid crystalline state - Conditions optimization	High
Cryo-EM	- Sample amount - Close to the native state	- Medium/low resolution - High MW system - Costly - Conditions optimization	Medium/low

XRD is a widely used and well-developed method. It allows the determination of 3D structures at high resolution. Nevertheless, it requires high-purity protein samples, as well as identifying the conditions for protein crystallization[138,139]. In addition, XRD

requires high protein concentrations, which are not always representative of the native protein oligomerization state in the solution.

Unlike XRD, cryo-EM allows studying protein systems in solution using a lower sample concentration. Those conditions are closer to the native states of protein (lower concentration required). It allows obtaining structures at medium to low resolution, even if some high-resolution structures have been resolved in some particular cases. One of the main difficulties of cryo-EM is its applicability to systems with a MW < 100 kDa, which represents most of the proteins known so far in the proteome (75% of proteins have a MW < 50 kDa). Nevertheless, these difficulties have recently been overcome using nano-, macro- or mega- bodies, which allowed to study protein systems with a MW < 100 kDa[140]. Therefore, dimeric forms of LDH-B, such as LDH-Btr and LDH-B^{F71A} could be studied by cryo-EM.

- The equilibrium between the dimeric and tetrameric forms would be shifted towards the dimer due to the low protein concentration required.
- To solve the problem of having a monodispersed sample for cryo-EM, Size Exclusion Chromatography (SEC) can be used to isolate one protein population (tetramer/dimer).
- To overcome the cryo-EM size limitation for protein systems < 100 kDa, LDH-B targeting nanobodies could be generated[140]. A nanobody (Nb) is a 15 kDa protein derived from Llama antibodies. Nanobodies targeting the LDH protein can be fused to a scaffold protein, such as Maltose-Binding Protein (MBP) or the adhesin domain of *Helicobacter pylori* (HopQ) or even the *Escherichia coli* K12 glucosidase (YgjK) to produce Macrobody or Megabody (Mb), respectively (~50-100 kDa). The MW of the newly formed protein complex is then increased and can be suitable for cryo-EM.

1.2. Structural study of LDH-B-small molecule complexes

The main objective of **part III** was the characterization of the binding mode of small molecules interacting at the tetrameric interface of LDH-B via the study of LDH-B-small molecule complexes using X-ray crystallography. The tested molecules were previously identified by the UCLouvain collaborators and have shown an interesting profile to prevent the tetramer formation of LDH-B protein[79].

This work allowed to obtain two crystallographic structures of LDH-B in complex with triprolidine (TRI) and Fluoxetine (FX) compounds: LDH-B-TRI and LDH-B-FX. Therefore, based on these structural data, two distinct binding sites close to the tetramerization site could be identified (Fig. D.2).

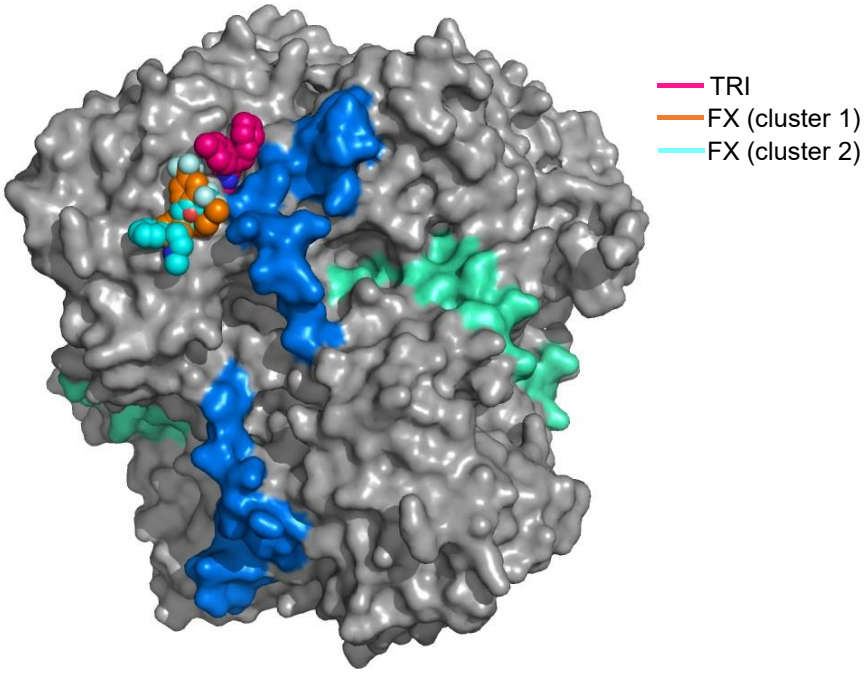


Figure D.2: Representation of the binding site of TRI and FX ligands in LDH-B. TRI ligand is shown in pink, FX ligand (cluster 1) in orange and FX ligand (cluster 2) in cyan. Tetramerization arms are represented in blue and 22-amino acid α-helix in green

The structure of the LDH-B-TRI complex revealed that the ligand binds near the N-term part of the tetramerization arm (close to the residues Ala9 and Val11) (Fig. D.2 in pink). This result can be correlated with the experimental data. Indeed, in presence of TRI compound, tetrameric LDH presents a thermal destabilization profile similar to the one observed with macrocycle 7[79]. As mentioned in the introduction (see section 5.2), macrocycle 7 is a stapled peptide designed and highlighted for competing with the N-terminal part of the tetramerization arm (residue Ala1 to Ile8)[59]. In addition, a competition assay between triprolidine and macrocycle 7 on LDH-Btr has been performed and shows a competition between the two molecules. Therefore, it has been suggested that the TRI compound could interact in the same way as macrocycle 7.

The structure of the LDH-B-FX complex revealed that the compound binds near the tetramerization arm (close to Glu13). Two clusters of FX have been identified through the two different orientations adopted by the compound (Fig. D.2 in orange and cyan). Concerning the experimental data generated by the UCLouvain collaborators,

experiments are ongoing. Nevertheless, a competition assay between triprolidine and fluoxetine has been performed and didn't show competition between the two compounds (data not published). Again, crystallographic results correlate with this data since two distinct binding sites have been identified for TRI and FX molecules.

The binding sites of TRI and FX ligands could be further characterized by obtaining crystallographic structures of LDH-Btr-TRI/FX complexes. Indeed, the presence of the tetramerization arm could potentially compete for the binding of these molecules (binding sites close to the arm) and obtaining structures of complexes with LDH-Btr could reveal the real interaction sites of these molecules. In addition, the cryo-EM technique could be used to reveal some potential conformational changes in the quaternary structure of the protein in presence of the compound. Nevertheless, XRD remains the method of choice to identify the binding site of the molecule at the tetrameric interface of LDH-B from an atomic point of view.

2. Outlooks

At the end of this work, several outlooks can be foreseen. First, the crystallographic conditions of the LDH-Btr protein could be optimized aiming to have easier access to LDH-Btr crystals, as well as to improve the crystallographic structure resolution. Indeed, the analysis of LDH-Btr crystals has allowed solving a structure at a medium resolution (2.98 Å). Therefore, obtaining a crystallographic structure at higher resolution will allow to obtain more precision on the LDH-Btr structure, which can be used as an interesting tool for the design and study of new modulators. Moreover, these optimized crystallization conditions can provide a good starting point for the co-crystallization conditions of LDH-Btr-small molecule complexes (molecules binding at the tetrameric interface of LDH-B).

In addition, it would be interesting to further characterize the dimeric state of LDH-B using cryo-EM. This technique should allow studying the three-dimensional structure of LDH-B in a close to native dimeric state (using mutants or LDH-Btr protein which are dimeric in solution) and will provide information leading to a better understanding of the LDH-B tetramerization process.

Concerning the characterization of the tetrameric interface of LDH-B, it could be interesting to complete the analysis of the tetramerization site 2 by studying the Glu61 and Leu70 residues as hotspots by obtaining crystallographic structures of corresponding

mutant proteins (LDH-B^{E61A} and LDH-B^{L70A}). The structural analysis of the mutants could provide further insight into the role of Glu61 et Leu70 residues in the formation of a stable LDH-B tetramer. In this work, the tetramerization site 1 has been characterized via the study of the LDH-B truncated form (LDH-Btr). This study could be complemented via the analysis of three hotspots that have been observed in solution by the collaborators (Leu3, Leu7 and Ile8)[59]. Therefore, the same methodology used for the characterization of tetramerization site 2 (using X-ray crystallography and molecular dynamics approaches) can be applied to tetramerization site 1.

Finally, for part III consisting in the structural study of LDH-B-small molecule complexes, several perspectives can be envisaged. First, to further characterize the binding site of molecules (such as fluoxetine and triprolidine) interacting at the tetrameric interface of LDH-B, it would be interesting to obtain small molecule-LDH-Btr complexes by XRD. Indeed, the truncation of the tetramerization arm of LDH-Btr makes the tetramerization site 1 more accessible for small molecules. In addition, these complexes can be also studied by Steered Molecular Dynamics to further characterize the binding mode of the compounds to LDH-B. This study could also be extended to other small molecules or peptides identified in solution by the UCLouvain collaborators as promising compounds to target the tetrameric interface of LDH protein.

Materials and methods

1. Recombinant LDH-B expression and purification

1.1. Transformation of competent bacteria

Transformed bacteria (*E. coli* Rosetta BL21 (DE3)) with pET-28a(+) (coding for 6xHis-LDH-Btr) was provided by Prof. R. Frédérick's group (UCLouvain). An aliquot of competent *E. coli* BL21 (DE3) bacteria was thawed on ice, and one µL of plasmid (pET-24(+) concentrated to 200 $ng.\mu L^{-1}$) was added to 50 µL of bacteria. The bacteria/plasmid mixture was then incubated on ice for 30 minutes. The transformation was carried out by heat shock for one minute and 15 seconds at 42°C in a water bath. Thereafter, bacteria were incubated for two minutes on ice. One mL of Lysogeny Broth (LB) Lennox medium was added to bacteria, and cells were incubated for one hour at 37°C under agitation. Then, bacteria were spread on a LB-agar plate (without antibiotics) using beads. The plate was incubated over-night (o/n) at 37°C to let bacteria grow in colonies. Finally, from a bacteria colony, a preculture of 10 mL was carried out and transformed bacteria were conserved in 20% (v/v) of glycerol (= glycerol stock) at -80°C.

1.2. Bacteria culture and protein induction

10 mL of LB medium, supplemented with plasmid-specific antibiotic, were inoculated with 50 µL of transformed bacteria (kanamycin (KAN): final concentration (C_f) = 50 $\mu g.mL^{-1}$ and chloramphenicol (CAM): C_f = 34 $\mu g.mL^{-1}$). KAN was used for bacteria transformed with pET-24a(+) plasmid construction and KAN + CAM were used for bacteria transformed with pET28a(+). Preculture was incubated o/n at 37°C under agitation.

The bacterial culture was performed by inoculating 300 mL LB medium with 10 mL of preculture supplemented with specific antibiotics (as mentioned before). The culture medium was incubated at 37°C under agitation (175 RPM). Bacterial growth was followed by an optical density (OD) measure at 600 nm. When an OD of 0.6 was reached, the protein of interest was induced by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG) (C_f = 1 or 0.5 mM for bacteria transformed with pET-28a(+) and pET-24a(+) plasmid construction, respectively). The bacterial culture was then incubated at 20°C for 20 hours under agitation (175 RPM). Bacterial pellets were recovered by centrifugation (30 minutes, 4°C, 5 000 RPM – rotor JA-20, Beckman Coulter Avanti J-E centrifuge). Bacteria were then resuspended in lysis buffer (50 mM Tris pH 8.5, 300 mM NaCl, 10 mM $MgCl_2$, 5 mM imidazole, 5% (v/v) glycerol, anti-proteases (cOmplete™, EDTA-free Protease Inhibitor Cocktail Tablets, Roche – 1 tablet/50 mL solution)) and cells were

disrupted by sonication (10 x 30 seconds spaced with 30 seconds resting time – Branson Sonifier®, power 7: 20 W). The supernatant containing the protein of interest was recovered after centrifugation (30 minutes, 4°C, 10 000 RPM – rotor JA-14, Beckman Coulter Avanti J-E centrifuge) and was then filtered through 0.45 µm and 0.20 µm filters. Finally, one µL of B-mercaptoethanol was added per mL of supernatant solution.

1.3. Macromolecule purification

a) Immobilized Metal Affinity Chromatography (IMAC)

The His-tagged proteins (6xHis-LDH-Btr, LDH-Btr-6xHis, LDH-B, LDH-B^{D64A}, LDH-B^{H67A}, LDH-B^{F71A}) were first purified by IMAC (His-Trap FF crude column – GE Healthcare). After loading the protein, the column was washed with buffer A (50 mM Tris pH 8.5, 300 mM NaCl, 30 mM imidazole, 10 mM $MgCl_2$, 5% (v/v) glycerol). The protein was then eluted with buffer B containing 250 mM imidazole using a gradient over 45 minutes for 6xHis-LDH-Btr, LDH-Btr-6xHis, LDH-B^{D64A} and LDH-B^{F71A} proteins (buffer B: 50 mM Tris pH 8.5, 300 mM NaCl, 250 mM imidazole, 10 mM $MgCl_2$, 5% (v/v) glycerol). The protein buffer was then exchanged with buffer C using a desalting column (PD10 Sephadex™ G-25 – GE Healthcare) or by dialysis (o/n), before being used for other experiments or flash-freezed and stored at -80°C. This buffer depends on the protein used, as well as on the experiment performed afterward. A list of used buffers is reported in Table M.1.

Table M.1: List of buffers used for proteins purified by IMAC

Protein	Next step	Buffer	Buffer exchange method
6xHis-LDH-Btr	Crystallization assays	50 mM phosphate buffer pH 7.5, 100 mM NaCl, 10% (v/v) glycerol, 1 mM DTT	Dialysis
	Thrombin cleavage	50 mM Tris pH 8.5, 300 mM NaCl, 10 mM $MgCl_2$, 5% (v/v) glycerol	PD10
	Ion exchange chromatography	25 mM Tris-HCl pH 7.5, 10 mM NaCl, 1 mM DTT, 5% (v/v) glycerol	Dialysis / PD10
LDH-Btr-6xHis/ LDH-$B^{mutants}$/ LDH-B	Crystallization assays	50 mM Tris-HCl pH 7.6, 150 mM NaCl, 5% (v/v) glycerol or 25 mM HEPES pH 7.5, 150 mM NaCl, 5% (v/v) glycerol	PD10
	Size exclusion chromatography		

b) Ion exchange chromatography

Once purified by IMAC, a second purification step by affinity chromatography (HiTrap Q HP column – GE Healthcare) can be performed (for 6xHis-LDH-Btr protein). After loading of protein on the column (buffer A: 25 mM Tris-HCl pH 7.5, 10 mM NaCl, 5% (v/v) glycerol), the protein was eluted using a salt gradient in NaCl over 30 minutes (buffer B: 25 mM Tris-HCl pH 7.5, 350 mM NaCl, 5% (v/v) glycerol). The protein buffer was then exchanged with buffer C (50 mM phosphate buffer pH 7.5, 100 mM NaCl, 10% (v/v) glycerol) using a desalting column (PD10).

c) Size exclusion chromatography

This purification step was performed after ion exchange chromatography (for 6xHis-LDH-Btr) or IMAC (for LDH-Btr-6xHis, LDH-B^{D64A}) using a gel filtration column (HiLoad 16/60 Superdex 200 pg – GE Healthcare). 10 mg of protein were loaded onto the column per run. Protein recovered after purification was concentrated with an amicon (cut-off: 10 kDa) by centrifugation (4°C, 6 300 RPM – rotor FA-35-6-30, Eppendorf 5430 R centrifuge). The protein buffer was then exchanged (see Table M.1) using a desalting column or by dialysis (o/n), before being used for other experiments or flash-freezed and stored at -80°C.

1.4. Thrombin cleavage

a) Thrombin cleavage optimization

An optimization of the thrombin cleavage was performed to determine the optimal cleavage conditions. The test was carried out using 6xHis-LDH-Btr purified by IMAC concentrated at 1.5 $mg.mL^{-1}$ in buffer 25 mM Tris-HCl pH 7.5, 10 mM NaCl, 10 mM $MgCl_2$ and 10% (v/v) glycerol. The protein was incubated at 4°C with thrombin (different tested concentrations: 1U, 5U, 10U and 25U/mg of protein) for different incubation times (2h, 5h, 6h, and 20h). For each incubation time, 10 µg of protein were collected. Samples were then analyzed on SDS-PAGE (10% acrylamide).

b) Thrombin cleavage and purification of cleaved LDH-Btr ($LDH\text{-}Btr_{cleaved}$)

The thrombin cleavage was performed using the optimized conditions (5U thrombin/mg of protein with 20h of incubation time) with 6xHis-LDH-Btr purified by IMAC (IMAC 1) concentrated at 1.5 $mg.mL^{-1}$ in buffer 50 mM Tris pH 8.5, 10 mM $MgCl_2$, 300 mM NaCl and 5% (v/v) glycerol. Once the cleavage was done, a second IMAC (IMAC 2) was performed. The protein buffer was exchanged with buffer A (50 mM Na phosphate pH 7.5, 150 mM NaCl, 30 mM imidazole and 5% (v/v) glycerol). The $LDH\text{-}Btr_{cleaved}$ (in presence of thrombin) was recovered in Flow Through (FT), while the uncleaved protein and 6xHis-tag were eluted with buffer B (50 mM Na phosphate pH 7.5, 150 mM NaCl, 250 mM imidazole and 5% (v/v) glycerol). After IMAC 2, the cleaved protein was exchanged again with buffer C (50 mM Na phosphate pH 7.5, 150 mM NaCl and 5% (v/v) glycerol) and a last purification step using benzamidine column (HiTrap® Benzamidine Fast Flow column – GE Healthcare) was performed to separate $LDH\text{-}Btr_{cleaved}$ from thrombin. The cleaved protein was eluted with buffer D (isocratic elution) (50 mM Na phosphate pH 7.5, 1 M NaCl and 5% (v/v) glycerol). Thrombin was eluted with buffer E (buffer D supplemented with 20 mM p-aminobenzamidine). The final protein buffer corresponds to buffer F (50 mM phosphate buffer pH 7.5, 100 mM NaCl).

2. Crystallization

2.1. Crystallization assays of LDH-B proteins

Crystallization of different LDH-B proteins was performed (6xHis-LDH-Btr, LDH-Btr-6xHis, $LDH\text{-}B^{D64A}$, $LDH\text{-}B^{H67A}$, and $LDH\text{-}B^{F71A}$). Concerning the crystallization assays of 6xHis-LDH-Btr and LDH-Btr-6xHis, the details are shown in the

results section (see Table 1.4 for 6xHis-LDH-Btr and Table 1.6 for LDH-Btr-6xHis). Crystallization assays of the three mutants of LDH-B are reported in Table M.2.

Table M.2: Crystallization assays of LDH-B^{D64A}, LDH-B^{H67A}, and LDH-B^{F71A} – screenings of crystallization conditions with mutants purified by IMAC (screenings 1-2, 4-6) and size exclusion chromatography (screening 3). Composition of the protein buffer Hepes: 25 mM Hepes pH 7.5, 150 mM NaCl, 5% (v/v) glycerol

Protein	Screening	Protein conc. (mg.mL^{-1})	Kit	Drop*	T (°C)	Protein buffer	Ligands	Method
LDH-B^{D64A}	1	8, 18, 26, 35	JCSG-plus™ (MD1-37)	1/1	RT	Hepes	/	Sitting drop
	2	9, 18, 22, 25.5	PACT premier™ (MD)	1/1	RT	Hepes	/	Sitting drop
	3	9, 13, 18	JCSG-plus™ (MD1-37)	1/1	RT	Hepes	Without or OXM (1.5x**) & NADH (1.5x**)	Sitting drop
LDH-B^{H67A}	4	8, 18.5, 24, 30	JCSG-plus™ (MD1-37)	1/1	RT	Hepes	/	Sitting drop
LDH-B^{F71A}	5	5, 8, 12.5, 15	JCSG-plus™ (MD1-37)	1/1	RT	Hepes	/	Sitting drop
	6	8, 13.5, 18, 22	PACT premier™ (MD1-29)	1/1	RT	Hepes	/	Sitting drop

* Mother liquor/protein ratios ** Ligand concentration (ligand/protein ratio)

Different screenings of crystallization conditions were performed for the three mutants (LDH-B^{D64A}, LDH-B^{H67A}, and LDH-B^{F71A}) by varying several parameters and using different screening kits (MD1-37 and MD1-29 from Molecular Dimensions[††]). Crystals were obtained in some screening conditions in 96-well plates. Some of them were repeated and optimized with the hanging-drop vapor-diffusion method to refine the crystallization conditions.

2.2. Crystallization of LDH-B proteins

Different forms of LDH-B were crystallized (LDH-Btr-6xHis, LDH-B^{D64A}, LDH-B^{H67A}, and LDH-B^{F71A}). The crystals of LDH-B proteins were obtained using the sitting-drop vapor-diffusion method in different crystallization conditions:

[††]https://www.moleculardimensions.com/

- **LDH-Btr-6xHis**: the crystals were obtained from a protein solution concentrated at 9 mg.mL^{-1} (protein buffer: 50 mM Tris-HCl pH 7.6, 150 mM NaCl, 5% (v/v) glycerol). The protein was previously incubated with NADH and oxamic acid (ratio protein/ligand: 1/3) for at least one hour on ice. The method used was the sitting-drop vapor-diffusion method in 96-well plates at 293 K. Crystals were obtained after 10 days in 0.2 M lithium sulfate, 0.1 M sodium acetate pH 4.5 and 50% (v/v) PEG 400.
- **LDH-B^{D64A}**: the crystals were obtained from a protein solution concentrated at 8 mg.mL^{-1} (protein buffer: 25 mM HEPES pH 7.5, 150 mM NaCl, 5% (v/v) glycerol). The method used was the hanging-drop vapor-diffusion method in 96-well plates at 293 K. Crystals were obtained in 24% (w/v) PEG 1500 and 20% (v/v) glycerol.
- **LDH-B^{H67A}**: the crystals were obtained from a protein solution concentrated at 8 mg.mL^{-1} (protein buffer: 25 mM HEPES pH 7.5, 150 mM NaCl, 5% (v/v) glycerol). The method used is the sitting-drop vapor-diffusion method in 96-well plates at 293 K. Crystals were obtained after 12 days in 24% (w/v) PEG 1500 and 20% (v/v) glycerol.
- **LDH-B^{F71A}**: the crystals were obtained from a protein solution concentrated at 12.5 mg.mL^{-1} (protein buffer: 25 mM HEPES pH 7.5, 150 mM NaCl, 5% (v/v) glycerol). The method used is the sitting-drop vapor-diffusion method in 96-well plates at 293 K. Crystals were obtained in 0.2 M lithium sulfate, 0.1 M Tris pH 8.5 and 40% (v/v) PEG 400.

The different crystals were flash-freezed in liquid nitrogen. Data collection was performed at Soleil synchrotron (Gif-sur-Yvette, France) on PROXIMA-1 and PROXIMA-2A beamlines at 100 K, using a Dectris EIGER X 16M detector at a single wavelength of 0.98011 (for LDH-B^{D64A} and LDH-B^{H67A}) or 0.97857 (for LDH-B^{F71A}). Data were processed with XDS[120] and the structures were solved by molecular replacement with Phaser[121] with a monomer of LDH-B (PDB entry: 1I0Z)[65]. Model building and refinement were performed with Phenix[122] and Coot[123].

2.3. Co-crystallization of LDH-B-ligand complexes

Co-crystallization of different LDH-B-ligand complexes were performed (LDH-B-MAP, LDH-B-TRI, LDH-B-IMI, LDH-B-FX, LDH-Btr-FX and LDH-Btr-TRI). The crystallization assay details are reported in Appendix 3.A. Crystals were obtained in some

screening conditions in 96-well plates. Some of them were repeated and optimized with the hanging-drop vapor-diffusion method to refine the crystallization conditions.

- **LDH-B-TRI**: the crystals were obtained from a protein solution concentrated at 15 mg.mL^{-1} (protein buffer: 50 mM Tris-HCl pH 7.5, 150 mM NaCl). The protein was previously incubated with triprolidine (TRI) (ratio protein/ligand: 1/3), NADH (2.5 mM) and oxamic acid (1 mM) for at least one hour on ice. The method used was the sitting-drop vapor-diffusion method in 96-well plates at 293 K. Crystals were obtained in 24% (w/v) PEG 1500 and 20% (v/v) glycerol.
- **LDH-B-FX**: the crystals were obtained from a protein solution concentrated at 15 mg.mL^{-1} (protein buffer: 50 mM Tris-HCl pH 7.5, 150 mM NaCl). The protein was previously incubated with Fluoxetine (FX) (ratio protein/ligand: 1/3), NADH (2.5 mM) and oxamic acid (1 mM) for at least one hour on ice. The method used was the sitting-drop vapor-diffusion method in 96-well plates at 293 K. Crystals were obtained in 0.2 M ammonium acetate, 0.1 M Bis-Tris pH 5.5 and 25% (w/v) PEG 3350.

The different crystals were flash-freezed in liquid nitrogen. Data collection was performed at Soleil synchrotron (Gif-sur-Yvette, France) on PROXIMA-1 and PROXIMA-2A beamlines at 100 K, using a Dectris EIGER X 16M detector at a single wavelength of 0.97857 (for LDH-B-TRI complex) or 0.98011 (for LDH-B-FX complex). Data were processed with XDS[120] and the structures were solved by molecular replacement with Phaser[121] with a monomer of LDH-B in presence of NADH and OXM (PDB entry: 1I0Z)[65]. Model building and refinement were performed with Phenix[122] and Coot[123]. TRI and FX ligands were added to the structure using eLBOW[124].

3. Differential Scanning Fluorimetry (DSF)

The thermal stability of LDH-B in different conditions was evaluated by DSF using the Protein Thermal Shift Dye kit (ThermoFisher Scientific). The experiments were performed in a 96-well plate (incubated on ice to avoid the denaturation of the protein) and each well (V_f: 20 µL) was composed as follows: 5 µL DSF buffer, 12.5 µL protein solution (C_f: 5 µg of protein/well) and 2.5 µL diluted dye (8x concentrated). The DSF buffer and the Sypro^{TM} Orange dye (1 000x concentrated) were provided in the commercial kit. Once the wells were filled and homogenized, the plate was sealed and centrifuged (3 minutes, 4°C, 3 000 RPM – rotor A-2-MTP, Eppendorf 5430 R centrifuge). Then, a temperature gradient was performed: the plate was heated up to 98°C (increment:

1°C/minute and stabilization at 25°C and 98°C (2 minutes)) with a StepOnePlus™ Real-Time PCR system (Applied Biosystem) and the ROX filter is applied for fluorescence measure. The data were then processed using Excel and GraphPad Prism (version 5.04, GraphPad Software).

For the characterization of LDH-B forms (LDH-B, 6xHis-LDH-Btr, and LDH-Btr$_{cleaved}$), the T_m was determined by the first derivative calculated from the sigmoidal curve. Concerning the buffer condition optimization, the protein solution (V_f: 12.5 µL) is composed as follows: 2.5 µL of protein (5 µg), 5 µL tested buffer (4x concentrated), and 5 µL salt solution (4x concentrated). The ΔT_m is calculated as the difference between the T_m (tested buffer) and the T_m (control buffer – 50 mM phosphate buffer pH 7.5, 100 mM NaCl). Concerning the additive screening, the protein solution (V_f: 12.5 µL) is composed as follows: 2.5 µL of protein (5 µg), 8 µL of buffer (100 mM Bicine pH 8.3, 500 mM NaCl), and 2 µL of additive (10x concentrated). The ΔT_m is calculated as the difference between the T_m (protein with test additive) and the T_m (protein without additive). All DSF data were processed using Excel and GraphPad Prism (version 5.04, GraphPad Software).

4. Molecular dynamics simulations

4.1. Molecular Dynamics simulations of crystallographic structures

Molecular Dynamics (MD) simulations were performed starting with the crystallographic structures of LDH-B^{D64A}, LDH-B^{H67A}, LDH-B^{F71A} (without any ligand), and LDH-B (PDB entry: 1I0Z). The unrefined part of LDH-B^{D64A} structure was built by homology using Prime in Maestro software (version 12.9)[141,142]. The MD simulations were performed using GROMACS 2016 (version 5.1.4)[143] with the CHARMM27 force field[144] and CMAP corrections. The protein system was placed in a cubic box with a distance of at least 2.0 nm between the protein and the box. Then, the protein was solvated with a hybrid solvent, using all-atom TIP3P water molecules (shell of 1.0 nm around the protein) while the rest of the box was filled with coarse-grained SIRAH water molecules[145–147]. Chloride and sodium ions were added to neutralize the total charge of the protein system and reach a concentration of 0.1 M in NaCl. The MD trajectories were generated under Periodic Boundary Conditions (PBC) using the Particle Mesh Ewald (PME) method and a distance cut-off of 1.2 nm was applied for the Coulomb and the vdW interactions. The V-rescale and Parrinello-Rahman algorithms were used to fix the temperature and the pressure, respectively. All MD calculations were carried out with a time step of 2 fs. A

first minimization was carried out using the steepest descent algorithm with an energy step size of 0.01 nm and a force convergence criterion of 1000 kJ.mol^{-1}.nm^{-1}. The equilibration stage was designed so as to progressively increase the temperature, as shown in Table M.3.

Table M.3: Description of the different equilibration steps

Temperature (K)	Time (ps)	System
50	10	NVT
150	20	NVT
300	20	NVT

Then, the final equilibration step (110 ns) was performed at 300 K and 1 bar. Finally, a 300 ns-production stage was carried out.

From the MD simulations generated for the protein systems LDH-B^{D64A}, LDH-B^{H67A}, LDH-B^{F71A}, and LDH-B, RMSD profiles of the 22-amino acids α-helix and the tetramerization site 2 during the 300 ns-production were analyzed. In addition, experimental (from crystallographic structure) and MD C_α RMSF profiles were studied. The experimental RMSF profiles were generated from the B-factor of each residue[128]:

$$RMSF_i^2 = \frac{3B_i}{8\pi^2}$$

4.2. Steered Molecular Dynamics

Steered Molecular Dynamics (SMD) simulations were performed for the available LDH-B crystallographic structure (PDB entry: 1I0Z) in the presence of ligands (NADH and OXM). Several mutated protein systems were generated from the LDH-B structure where the single-point mutation is inserted using the Pymol software[125]. The same software as described above was used as well as the solvation procedure. To allow the separation of the two dimers, the protein system was aligned along the y-axis and the box length along the same axis was enlarged by 5.0 nm. The ligand topology files were generated with SwissParam software[148]. Position constraints were applied on the C_α atoms of the fixed AC dimer, except for the residues constituting the tetramerization arm and the 22-amino acid α-helix and those at the interface of the two dimers. In addition, angle constraints were applied for the mobile BD dimer, except for the same residues cited previously for position constraints. A first minimization was processed using the steepest descent algorithm with an energy step size of 0.01 nm and a force convergence criterion of 1000 kJ.mol^{-1}.nm^{-1}. The equilibration stage was designed so as to

progressively increase the temperature, as mentioned in Table M.3, and an additional equilibration step (20 ns) was performed at 300 K and 1 bar. Finally, the SMD simulation was performed at 300 K and 1 bar by applying an external force on the center of mass of BD dimer (k = 500 and $kJ.mol^{-1}.nm^{-1}$ and $v_{pull} = 0.005\ nm.ps^{-1}$)

Bibliography

1. Ponomarenko, E. A. *et al.* The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.* (2016).

2. Proteomes . Homo sapiens (Human). Available at: https://www.uniprot.org/proteomes/UP000005640. (Accessed: 11th October 2022)

3. Cozzone, A. J. Proteins: Fundamental Chemical Properties. in *Encyclopedia of Life Sciences* (ed. John Wiley & Sons, L.) 1–10 (2002).

4. Sun, P. D., Foster, C. E. & Boyington, J. C. Overview of protein structural and functional folds. *Curr. Protoc. protein Sci.* **35**, 17.1.1-17.1.189 (2004).

5. Nelson, D. L. & Cox, M. M. *Lehninger - Principles of biochemistry*. (2008).

6. Huttlin, E. L. *et al.* Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).

7. Thabault, L., Liberelle, M. & Frédérick, R. Targeting protein self-association in drug design. *Drug Discov. Today* **26**, 1148–1163 (2021).

8. Cardinale, D. *et al.* Homodimeric Enzymes as Drug Targets. *Curr. Med. Chem.* **17**, 826–846 (2010).

9. Ivanov, A. A., Khuri, F. R. & Fu, H. Targeting protein-protein interactions as an anticancer strategy. *Trends Pharmacol. Sci.* **34**, 393–400 (2013).

10. Rosell, M. & Fernández-Recio, J. Hot-spot analysis for drug discovery targeting protein-protein interactions. *Expert Opin. Drug Discov.* **13**, 327–338 (2018).

11. Sheng, C., Dong, G., Miao, Z., Zhang, W. & Wang, W. State-of-the-art strategies for targeting protein-protein interactions by small-molecule inhibitors. *Chem. Soc. Rev.* **44**, 8238–8259 (2015).

12. Bakail, M. & Ochsenbein, F. Targeting protein-protein interactions, a wide open field for drug design. *Comptes Rendus Chim.* **19**, 19–27 (2016).

13. Zhang, M. Q. & Wilkinson, B. Drug discovery beyond the 'rule-of-five'. *Curr. Opin. Biotechnol.* **18**, 478–488 (2007).

14. Modell, A. E., Blosser, S. L. & Arora, P. S. Systematic Targeting of Protein-Protein Interactions Approaches to Targeting Protein-Protein Interactions. *Trends Pharmacol Sci* **37**, 702–713 (2016).

15. Zinzalla, G. & Thurston, D. E. Targeting protein-protein interactions for therapeutic intervention: A challenge for the future. *Future Med. Chem.* **1**, 65–93 (2009).

16. Akçay, G. *et al.* Inhibition of Mcl-1 through covalent modification of a noncatalytic lysine side chain. *Nat. Chem. Biol.* **12**, 931–936 (2016).

17. Cheng, S. S., Yang, G. J., Wang, W., Leung, C. H. & Ma, D. L. The design and development of covalent protein-protein interaction inhibitors for cancer treatment. *J. Hematol. Oncol.* **13**, 1–14 (2020).

18. Li, B., Rong, D. & Wang, Y. Targeting Protein-Protein Interaction with Covalent Small-Molecule Inhibitors. *Curr. Top. Med. Chem.* **19**, 1872–1876 (2019).

19. Chan, A. M., Goodis, C. C., Pommier, E. G. & Fletcher, S. Recent applications of covalent chemistries in protein-protein interaction inhibitors. *RSC Med. Chem.* **13**, 921–928 (2022).

20. Lu, H. *et al.* Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Signal Transduct. Target. Ther.* **5**, 213 (2020).

21. Tariq, S., Tariq, S., Khan, M., Azhar, A. & Baig, M. Venetoclax in the Treatment of Chronic Lymphocytic Leukemia: Evidence, Expectations, and Future Prospects. *Cureus* **12**, (2020).

22. Deeks, E. D. Venetoclax: First Global Approval. *Drugs* **76**, 979–987 (2016).

23. Weaver, B. A. How Taxol/paclitaxel kills cancer cells. *Mol. Biol. Cell* **25**, 2677–2681 (2014).

24. Nadal-Bufi, F. *et al.* Designed β-Hairpins Inhibit LDH5 Oligomerization and Enzymatic Activity. *J. Med. Chem.* **64**, 3767–3779 (2021).

25. Craik, D. J., Fairlie, D. P., Liras, S. & Price, D. The Future of Peptide-based Drugs. *Chem. Biol. Drug Des.* **81**, 136–147 (2013).

26. Jauset, T. & Beaulieu, M. E. Bioactive cell penetrating peptides and proteins in cancer: a bright future ahead. *Curr. Opin. Pharmacol.* **47**, 133–140 (2019).

27. Howl, J. & Jones, S. A new biology of cell penetrating peptides. *Pept. Sci.* **113**, e24154 (2021).

28. Carvajal, L. A. *et al.* Dual inhibition of MDMX and MDM2 as a therapeutic strategy in leukemia. *Sci. Transl. Med.* **10**, eaao3003 (2018).

29. Chène, P. Inhibiting the p53-MDM2 interaction: An important target for cancer therapy. *Nat. Rev. Cancer* **3**, 102–109 (2003).

30. Li, S. *et al.* Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **7**, 301–311 (2005).

31. European Medicines Agency. Erbitux, INN-cetuximab. 1-37.

32. Syed, Y. Y. Durvalumab: First Global Approval. *Drugs* **77**, 1369–1376 (2017).

33. Antonia, S. J. *et al.* Durvalumab after Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **377**, 1919–1929 (2017).

34. Fischer, G., Rossmann, M. & Hyvönen, M. Alternative modulation of protein-protein interactions by small molecules. *Curr. Opin. Biotechnol.* **35**, 78–85 (2015).

35. Hassell, A. M. *et al.* Crystallization of protein-ligand complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 72–79 (2006).

36. Müller, I. Guidelines for the successful generation of protein-ligand complex crystals. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 79–92 (2017).

37. Li, Z. *et al.* The OncoPPi network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies. *Nat. Commun.* **8**, 14356 (2017).

38. Matos, B., Howl, J., Jerónimo, C. & Fardilha, M. The disruption of protein-protein interactions as a therapeutic strategy for prostate cancer. *Pharmacol. Res.* **161**, 105145 (2020).

39. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the warburg effect: The metabolic requirements of cell proliferation. *Science.* **324**, 1029–1033 (2009).

40. Kalyanaraman, B. Teaching the basics of cancer metabolism: Developing antitumor strategies by exploiting the differences between normal and cancer cell metabolism. *Redox Biol.* **12**, 833–842 (2017).

41. Vazquez, A., Liu, J., Zhou, Y. & Oltvai, Z. N. Catabolic efficiency of aerobic glycolysis: The Warburg effect revisited. *BMC Syst. Biol.* **4**, 1–9 (2010).

42. Fiume, L., Manerba, M., Vettraino, M. & Di Stefano, G. Inhibition of lactate dehydrogenase activity as an approach to cancer therapy. *Future Med. Chem.* **6**, 429–445 (2014).

43. McClelland, G. B., Khanna, S., González, G. F., Butz, C. E. & Brooks, G. A. Peroxisomal membrane monocarboxylate transporters: Evidence for a redox shuttle system? *Biochem. Biophys. Res. Commun.* **304**, 130–135 (2003).

44. Schueren, F. *et al.* Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. *Elife* **3**, e03640 (2014).

45. Kim, J. W. & Dang, C. V. Multifaceted roles of glycolytic enzymes. *Trends Biochem. Sci.* **30**, 142–150 (2005).

46. Rebucci, M. & Michiels, C. Molecular aspects of cancer cell resistance to chemotherapy. *Biochem. Pharmacol.* **85**, 1219–1226 (2013).

47. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).

48. Liberti, M. V. & Locasale, J. W. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends Biochem. Sci.* **41**, 211–218 (2016).

49. DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G. & Thompson, C. B. The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. *Cell Metab.* **7**, 11–20 (2008).

50. Thabault, L. *et al.* Discovery of a novel lactate dehydrogenase tetramerization domain using epitope mapping and peptides. *J. Biol. Chem.* **296**, 1–12 (2021).

51. Sonveaux, P. *et al.* Targeting lactate-fueled respiration selectively kills hypoxic tumor cells in mice. *J. Clin. Invest.* **118**, 3930–3942 (2008).

52. Brisson, L. *et al.* Lactate Dehydrogenase B Controls Lysosome Activity and Autophagy in Cancer. *Cancer Cell* **30**, 418–431 (2016).

53. Danhier, P. *et al.* Cancer metabolism in space and time: Beyond the Warburg effect. *Biochim. Biophys. Acta - Bioenerg.* **1858**, 556–572 (2017).

54. White, E. Deconvoluting the context-dependent role for autophagy in cancer. *Nat. Rev. Cancer* **12**, 401–410 (2012).

55. Chiarugi, A., Dölle, C., Felici, R. & Ziegler, M. The NAD metabolome - A key determinant of cancer cell biology. *Nat. Rev. Cancer* **12**, 741–752 (2012).

56. Sonveaux, P. *et al.* Targeting the lactate transporter MCT1 in endothelial cells inhibits lactate-induced HIF-1 activation and tumor angiogenesis. *PLoS One* **7**, e33418 (2012).

57. de Saedeleer, C. J. *et al.* Lactate Activates HIF-1 in Oxidative but Not in Warburg-Phenotype Human Tumor Cells. *PLoS One* **7**, e46571 (2012).

58. Estrella, V. *et al.* Acidity generated by the tumor microenvironment drives local invasion. *Cancer Res.* **73**, 1524–1535 (2013).

59. Thabault, L. *et al.* Interrogating the lactate dehydrogenase tetramerization site using (stapled) peptides. *J. Med. Chem.* **63**, 4628–4643 (2020).

60. Végran, F., Boidot, R., Michiels, C., Sonveaux, P. & Feron, O. Lactate influx through the endothelial cell monocarboxylate transporter MCT1 supports an NF-kB/IL-8 pathway that drives tumor angiogenesis. *Cancer Res.* **71**, 2550–2560 (2011).

61. Pugh, C. W. & Ratcliffe, P. J. Regulation of angiogenesis by hypoxia: Role of the HIF system. *Nat. Med.* **9**, 677–684 (2003).

62. Colen, C. B. *et al.* Metabolic targeting of lactate efflux by malignant glioma inhibits invasiveness and induces necrosis: An in vivo study1. *Neoplasia* **13**, 620–632 (2011).

63. Izumi, H. *et al.* Monocarboxylate transporters 1 and 4 are involved in the invasion activity of human lung cancer cells. *Cancer Sci.* **102**, 1007–1013 (2011).

64. Valvona, C. J., Fillmore, H. L., Nunn, P. B. & Pilkington, G. J. The Regulation and Function of Lactate Dehydrogenase A: Therapeutic Potential in Brain Tumor. *Brain Pathol.* **26**, 3–17 (2016).

65. Read, J. A., Winter, V. J., Eszes, C. M., Sessions, R. B. & Brady, R. L. Structural basis for altered activity of M- and H-isozyme forms of human lactate dehydrogenase. *Proteins Struct. Funct. Genet.* **43**, 175–185 (2001).

66. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).

67. Adams, M. J. *et al.* Structure of Lactate Dehydrogenase at 2.8 A Resolution. *Nature* **227**, 1098–1103 (1970).

68. Cameron, A. *et al.* Identification and activity of a series of azole-based compounds with lactate dehydrogenase-directed anti-malarial activity. *J. Biol. Chem.* **279**, 31429–31439 (2004).

69. Shibata, S. *et al.* Identification of the first highly selective inhibitor of human lactate dehydrogenase B. *Sci. Rep.* **11**, 1–12 (2021).

70. Medvedev, K. E., Kinch, L. N., Schaeffer, R. D. & Grishin, N. V. Functional analysis of rossmann-like domains reveals convergent evolution of topology and reaction pathways. *PLoS Comput. Biol.* **15**, e1007569 (2019).

71. Shin, W.-H. 55 Years of the Rossmann Fold. *Methods Mol. Biol.* **1958**, 1–13 (2019).

72. Schmidt, R. K. & Gready, J. E. Molecular dynamics simulations of L-lactate dehydrogenase: Conformation of a mobile loop and influence of the tetrameric protein environment. *J. Mol. Model.* **5**, 153–168 (1999).

73. El Hawrani, A. S., Moreton, K. M., Sessions, R. B., Clarke, A. R. & Holbrook, J. J. Engineering surface loops of proteins - a preferred strategy for obtaining new enzyme function. *Trends Biotechnol.* **12**, 207–211 (1994).

74. Świderek, K. & Paneth, P. Importance of the lactate dehydrogenase quaternary structure in theoretical calculations. *J. Phys. Chem. B* **114**, 3393–3397 (2010).

75. Opitz, U., Rudolph, R., Jaenicke, R., Ericsson, L. & Neurath, H. Proteolytic Dimers of Porcine Muscle Lactate Dehydrogenase: Characterization, Folding, and Reconstitution of the Truncated and Nicked Polypeptide Chain. *Biochemistry* **26**, 1399–1406 (1987).

76. Jaenicke, R., Vogel, W. & Rudolph, R. Dimeric Intermediates in the Dissociation of Lactic Dehydrogenase. *Eur. J. Biochem.* **114**, 525–531 (1981).

77. Yamamoto, S. & Storey, K. B. Dissociation-Association of lactate dehydrogenase Isozymes: Influences on the formation of tetramers versus dimers of M4-LDH and H4-LDH. *Int. J. Biochem.* **20**, 1261–1265 (1988).

78. Jafary, F., Ganjalikhany, M. R., Moradi, A., Hemati, M. & Jafari, S. Novel Peptide Inhibitors for Lactate Dehydrogenase A (LDHA): A Survey to Inhibit LDHA Activity via Disruption of Protein-Protein Interaction. *Sci. Rep.* **9**, 1–13 (2019).

79. Thabault, L. *et al.* Discovery of small molecules interacting at lactate dehydrogenases tetrameric interface using a biophysical screening cascade. *Eur. J. Med. Chem.* **230**, (2022).

80. Oshima, N. *et al.* Dynamic Imaging of LDH Inhibition in Tumors Reveals Rapid In Vivo Metabolic Rewiring and Vulnerability to Combination Therapy. *Cell Rep.* **30**, 1798-1810.e4 (2020).

81. Fantin, V. R., St-Pierre, J. & Leder, P. Attenuation of LDH-A expression uncovers a link between glycolysis, mitochondrial physiology, and tumor maintenance. *Cancer Cell* **9**, 425–434 (2006).

82. Xie, H. *et al.* Targeting lactate dehydrogenase-A inhibits tumorigenesis and tumor progression in mouse models of lung cancer and impacts tumor-initiating cells. *Cell Metab.* **19**, 795–809 (2014).

83. Le, A. *et al.* Inhibition of lactate dehydrogenase A induces oxidative stress and inhibits tumor progression. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2037–2042 (2010).

84. Rani, R. & Kumar, V. Recent Update on Human Lactate Dehydrogenase Enzyme 5 (hLDH5) Inhibitors: A Promising Approach for Cancer Chemotherapy. *J. Med. Chem.* **59**, 487–496 (2016).

85. Granchi, C., Bertini, S., Macchia, M. & Minutolo, F. Inhibitors of Lactate Dehydrogenase Isoforms and their Therapeutic Potentials. *Curr. Med. Chem.* **17**, 672–697 (2010).

86. Ward, R. A. *et al.* Design and synthesis of novel lactate dehydrogenase a inhibitors by fragment-based lead generation. *J. Med. Chem.* **55**, 3285–3306 (2012).

87. Billiard, J. *et al.* Quinoline 3-sulfonamides inhibit lactate dehydrogenase A and reverse aerobic glycolysis in cancer cells. *Cancer Metab.* **1**, 1–17 (2013).

88. Boudreau, A. *et al.* Metabolic plasticity underpins innate and acquired resistance to LDHA inhibition. *Nat. Chem. Biol.* **12**, 779–786 (2016).

89. Di Magno, L. *et al.* Discovery of novel human lactate dehydrogenase inhibitors: Structure-based virtual screening studies and biological assessment. *Eur. J. Med. Chem.* **240**, 114605 (2022).

90. Dragovich, P. S. *et al.* Identification of substituted 3-hydroxy-2-mercaptocyclohex-2-enones as potent inhibitors of human lactate dehydrogenase. *Bioorganic Med. Chem. Lett.* **24**, 3764–3771 (2014).

91. Dragovich, P. S. *et al.* Identification of substituted 2-thio-6-oxo-1,6-dihydropyrimidines as inhibitors of human lactate dehydrogenase. *Bioorganic Med. Chem. Lett.* **23**, 3186–3194 (2013).

92. Rupiani, S. *et al.* Identification of N-acylhydrazone derivatives as novel lactate dehydrogenase A inhibitors. *Eur. J. Med. Chem.* **101**, 63–70 (2015).

93. Kohlmann, A. *et al.* Fragment growing and linking lead to novel nanomolar lactate dehydrogenase inhibitors. *J. Med. Chem.* **56**, 1023–1040 (2013).

94. Altamimi, A. M. S. *et al.* Symmetric molecules with 1,4-triazole moieties as potent inhibitors of tumour-associated lactate dehydrogenase-A. *J. Enzyme Inhib. Med. Chem.* **33**, 147–150 (2018).

95. Zhou, Y. *et al.* Development of novel human lactate dehydrogenase A inhibitors: High-throughput screening, synthesis, and biological evaluations. *Eur. J. Med. Chem.* **177**, 105–115 (2019).

96. Ždralević, M. *et al.* Double genetic disruption of lactate dehydrogenases A and B is required to ablate the "Warburg effect" restricting tumor growth to oxidative metabolism. *J. Biol. Chem.* **293**, 15947–15961 (2018).

97. Rai, G. *et al.* Discovery and Optimization of Potent, Cell-Active Pyrazole-Based Inhibitors of Lactate Dehydrogenase (LDH). *J. Med. Chem.* **60**, 9184–9204 (2017).

98. Rai, G. *et al.* Pyrazole-Based Lactate Dehydrogenase Inhibitors with Optimized Cell Activity and Pharmacokinetic Properties. *J. Med. Chem.* **63**, 10984–11011 (2020).

99. Manerba, M. *et al.* Galloflavin (CAS 568-80-9): A Novel Inhibitor of Lactate Dehydrogenase. *ChemMedChem* **7**, 311–317 (2012).

100. Friberg, A. *et al.* Structural Evidence for Isoform-Selective Allosteric Inhibition of Lactate Dehydrogenase A. *ACS Omega* **5**, 13034–13041 (2020).

101. Ramírez-Alvarado, M., Blanco, F. J. & Serrano, L. De novo design and structural analysis of a model β-hairpin peptide system. *Nat. Struct. Biol.* **3**, 604–612 (1996).

102. Rupp, B. Origin and use of crystallization phase diagrams. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **71**, 247–260 (2015).

103. Deller, M. C., Kong, L. & Rupp, B. Protein stability: A crystallographer's perspective. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **72**, 72–95 (2016).

104. Bucher, M. H., Evdokimov, A. G. & Waugh, D. S. Differential effects of short affinity tags on the crystallization of Pyrococcus furiosus maltodextrin-binding protein. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 392–397 (2002).

105. Gao, K., Oerlemans, R. & Groves, M. R. Theory and applications of differential scanning fluorimetry in early-stage drug discovery. *Biophys. Rev.* **12**, 85–104 (2020).

106. Boivin, S., Kozak, S. & Meijers, R. Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr. Purif.* **91**, 192–206 (2013).

107. Reinhard, L., Mayerhofer, H., Geerlof, A., Mueller-Dieckmann, J. & Weiss, M. S. Optimization of protein buffer cocktails using Thermofluor. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **69**, 209–214 (2013).

108. Santos, S. P. *et al.* Thermofluor-based optimization strategy for the stabilization and crystallization of Campylobacter jejuni desulforubrerythrin. *Protein Expr. Purif.* **81**, 193–200 (2012).

109. Goldberg, M. E., Expert-Bezançon, N., Vuillard, L. & Rabilloud, T. Non-detergent sulphobetaines: A new class of molecules that facilitate in vitro protein renaturation. *Fold. Des.* **1**, 21–27 (1996).

110. Vuillard, L., Madern, D., Franzetti, B. & Rabilloud, T. Halophilic protein stabilization by the mild solubilizing agents nondetergent sulfobetaines. *Analytical Biochemistry* **230**, 290–294 (1995).

111. Vuillard, L., Braun-Breton, C. & Rabilloud, T. Non-detergent sulphobetaines: A new class of mild solubilization agents for protein purification. *Biochem. J.* **305**, 337–343 (1995).

112. Vuillard, L., Rabilloud, T., Leberman, R., Berthet-Colominas, C. & Cusack, S. A new additive for protein crystallization. *FEBS Lett.* **353**, 294–296 (1994).

113. Liao, Y. T., Manson, A. C., DeLyser, M. R., Noid, W. G. & Cremer, P. S. Trimethylamine N-oxide stabilizes proteins via a distinct mechanism compared with betaine and glycine. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2479–2484 (2017).

114. Slabinski, L. *et al.* XtalPred: A web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403–3405 (2007).

115. Jahandideh, S., Jaroszewski, L. & Godzik, A. Improving the chances of successful protein structure determination with a random forest classifier. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 627–635 (2014).

116. Slabinski, L. *et al.* The challenge of protein structure determination-lessons from structural genomics. *Protein Sci.* **16**, 2472–2482 (2007).

117. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **337**, 635–645 (2004).

118. Guruprasad, K., Reddy, B. V. B. & Pandit, M. W. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng. Des. Sel.* **4**, 155–161 (1990).

119. Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M. & Lundegaard, C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct. Biol.* **9**, 1–10 (2009).

120. Kabsch, W. Biological crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **D66**, 125–132 (2010).

121. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

122. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 861–877 (2019).

123. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

124. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. Electronic ligand builder and optimization workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, 1074–1080 (2009).

125. The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.

126. Derewenda, Z. S. Application of protein engineering to enhance crystallizability and improve crystal properties. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 604–615 (2010).

127. Ruggiero, A., Smaldone, G., Squeglia, F. & Berisio, R. Enhanced Crystallizability by Protein Engineering Approaches: A General Overview. *Protein Pept. Lett.* **19**, 732–742 (2012).

128. Kuzmanic, A. & Zagrovic, B. Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys. J.* **98**, 861–871 (2010).

129. Do, P. C., Lee, E. H. & Le, L. Steered Molecular Dynamics Simulation in Rational Drug Design. *J. Chem. Inf. Model.* **58**, 1473–1482 (2018).

130. Zhang, J. long, Zheng, Q. chuan & Zhang, H. xing. Unbinding of glucose from human pulmonary surfactant protein D studied by steered molecular dynamics simulations. *Chem. Phys. Lett.* **484**, 338–343 (2010).

131. Isralewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D. & Schulten, K. Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Model.* **19**, 13–25 (2001).

132. Gao, Q. *et al.* Molecular dynamics simulation and steered molecular dynamics simulation on irisin dimers. *J. Mol. Model.* **24**, (2018).

133. Miño, G., Baez, M. & Gutierrez, G. Effect of mutation at the interface of Trp-repressor dimeric protein: A steered molecular dynamics simulation. *Eur. Biophys. J.* **42**, 683–690 (2013).

134. Patel, J. S., Berteotti, A., Ronsisvalle, S., Rocchia, W. & Cavalli, A. Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5. *J. Chem. Inf. Model.* **54**, 470–480 (2014).

135. Shen, M. *et al.* Steered molecular dynamics simulations on the binding of the appendant structure and helix-β2 in domain-swapped human cystatin C dimer. *J. Biomol. Struct. Dyn.* **30**, 652–661 (2012).

136. Ravikumar, A., Gopnarayan, M. N., Subramaniam, S. & Srinivasan, N. Comparison of side-chain dispersion in protein structures determined by cryo-EM and X-ray crystallography. *IUCrJ* **9**, 98–103 (2022).

137. Glaeser, R. M. Proteins, interfaces, and cryo-EM grids. *Curr. Opin. Colloid Interface Sci.* **34**, 1–8 (2018).

138. Blundell, T. L., Jhoti, H. & Abell, C. High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov.* **1**, 45–54 (2002).

139. Ilari, A. & Savino, C. Protein structure determination by X-ray crystallography. *Methods Mol. Biol.* **452**, 63–87 (2008).

140. Wentinck, K., Gogou, C. & Meijer, D. H. Putting on molecular weight: Enabling cryo-EM structure determination of sub-100-kDa proteins. *Curr. Res. Struct. Biol.* **4**, 332–337 (2022).

141. Jacobson, M. P., Friesner, R. A., Xiang, Z. & Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **320**, 597–608 (2002).

142. Jacobson, M. P. *et al.* A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins Struct. Funct. Genet.* **55**, 351–367 (2004).

143. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**–**2**, 19–25 (2015).

144. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545–1614 (2009).

145. Darre, L., Tek, A., Baaden, M. & Pantano, S. Mixing atomistic and coarse grain solvation models for MD simulations : let WT4 handle the bulk. *J. Chem. Theory Comput.* **8**, 3880–3894 (2012).

146. Darré, L. *et al.* SIRAH: A structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J. Chem. Theory Comput.* **11**, 723–739 (2015).

147. Gonzalez, H. C., Darré, L. & Pantano, S. Transferable mixing of atomistic and coarse-grained water models. *J. Phys. Chem. B* **117**, 14438–14448 (2013).

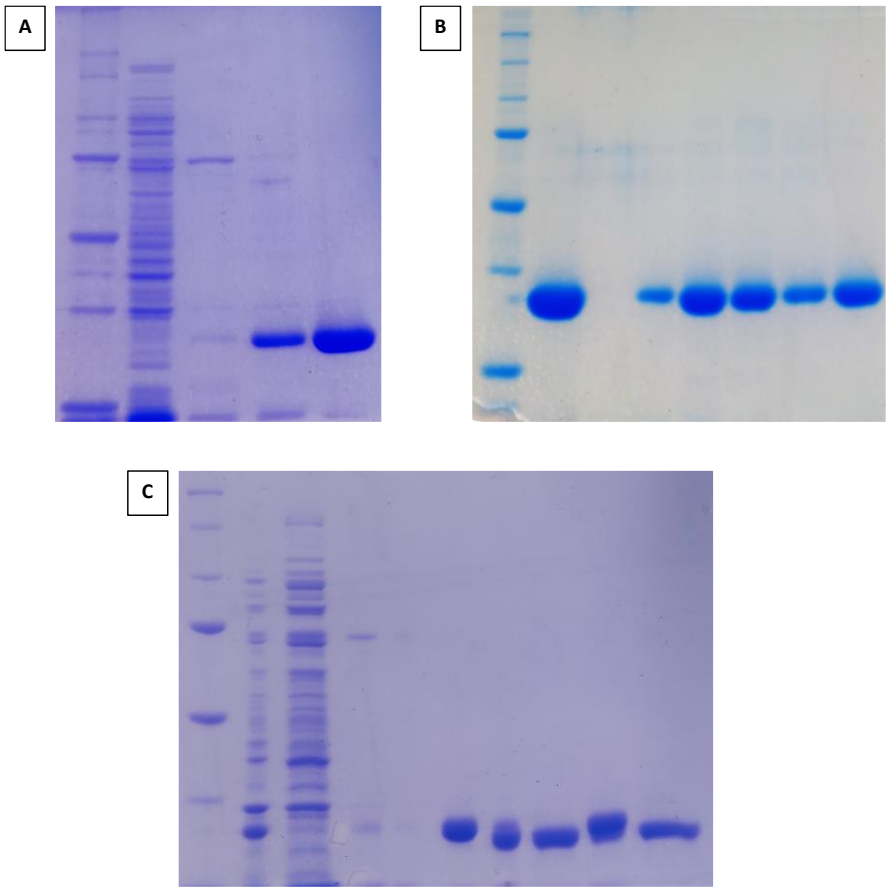
148. Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O. SwissParam: A Fast Force Field Generation Tool for Small Organic Molecules. *J. Comput. Chem.* **32**, 2359–2368 (2011).

149. BIOVA, Dassault Systèmes, Discovery Studio Visualizer, Version 18.1, San Diego: Dassault Sytèmes, 2018.

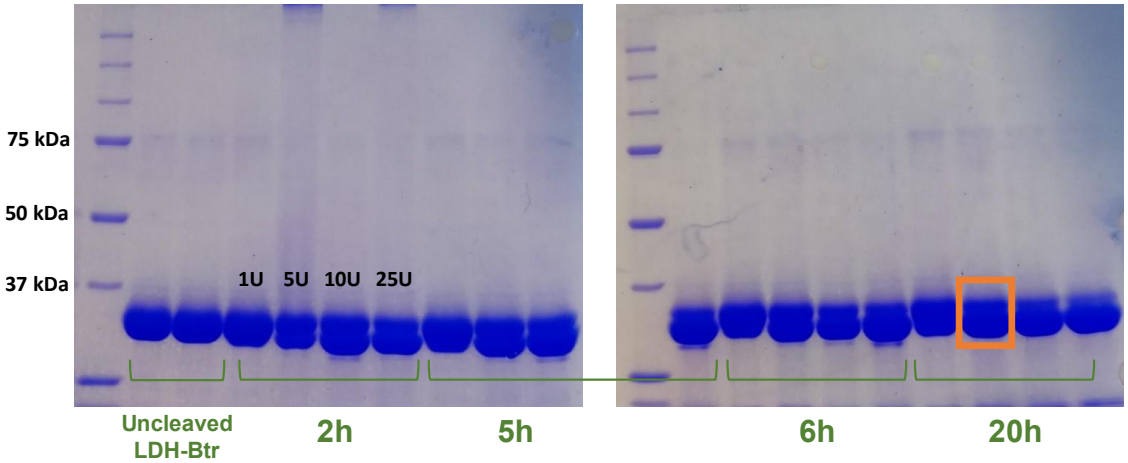
Appendices

Appendix 1

Structural study of truncated LDH-B (LDH-Btr) – Characterization of the tetramerization site 1



Appendix 1.A: Representation of full SDS-PAGE presented in part I - (A) Gel presented in Figure 1.3 – (B) Gel presented in Figure 1.4 – (C) Gel presented in Figure 1.5



Appendix 1.B: Thrombin cleavage optimization of LDH-Btr. SDS-PAGE (10% acrylamide) – LDH-Btr samples after thrombin cleavage under the different tested conditions: thrombin concentration (1U to 25U/mg protein) and incubation time (2 hours to 20 hours – green lines) – selected condition: 5U/mg protein and 20 hours of incubation (orange box)

Appendix 1.C: List of the 17 selected additives (Hampton Research kit (HR2-428)) screened by DSF

	Additives	Test concentration
C34	Glycine	0.1 M
C44	Urea	0.01 M
C45	B-Nicotinamide adenine dinucleotide hydrate	0.01 M
C46	Adenosine-5'-triphosphate disodium salt hydrate	0.01 M
C49	Ethylenediaminetetraacetic acid disodium salt dihydrate	0.01 M
C54	D-(+)-Glycose monohydrate	3% (w/v)
C55	Sucrose	3% (w/v)
C56	Xylitol	3% (w/v)
C57	D-Sorbitol	3% (w/v)
C59	D-(+)-Trehalose dihydrate	3% (w/v)
C60	D-(+)-Galactose	3% (w/v)
C62	Glycerol	3% (v/v)
C63	NDSB-195	0.3 M
C73	Trimethylamine N-oxide dihydrate	3% (w/v)
C88	Acetonitrile	4% (v/v)
C91	Ethyl acetate	0.5% (v/v)
C93	Dichloromethane	0.025% (v/v)

Appendix 2

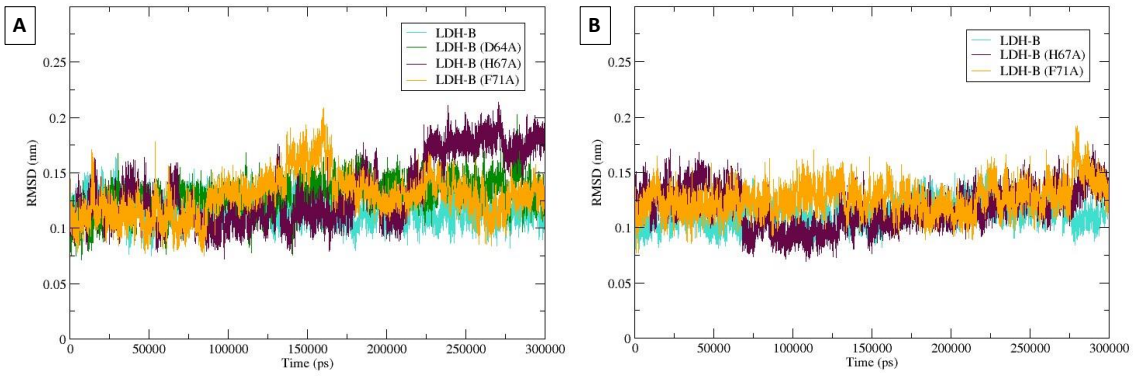
Structural study of LDH-B mutants – Characterization of the tetramerization site 2

Appendix 2.A: Data collection and refinement statistics for LDH-B^{D64A}, LDH-B^{H67A} and LDH-B^{F71A} structures. Values for the highest-resolution shell are given brackets

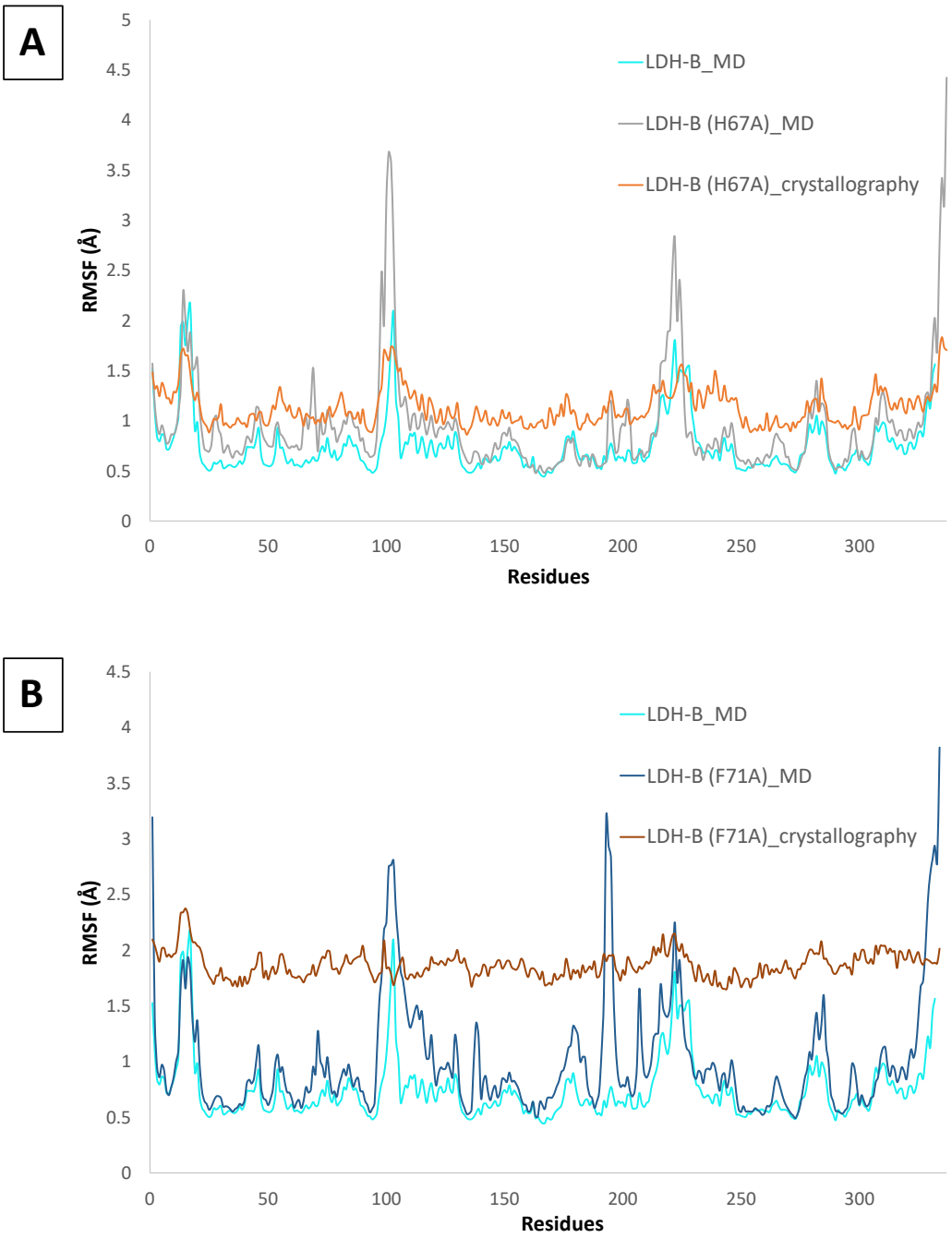
	LDH-B^{D64A}	**LDH-B^{H67A}**	**LDH-B^{F71A}**
Wavelength	0.98011	0.98011	0.97857
Resolution range (Å)	47.95 – 2.00 (2.07 – 2.00)	46.31 – 2.39 (2.48 – 2.39)	46.20 – 3.22 (3.34 – 3.22)
Space group	$P2_1$	$P2_12_12_1$	R32
a, b, c (Å)	98.06, 68.27, 111.81	88.95, 130.21, 131.77	194.49, 194.49, 384.32
α, β, γ (°)	90, 103.319, 90	90, 90, 90	90, 90, 120
Total reflections	519131 (50041)	831569 (76368)	937257 (83932)
Unique reflections	97234 (1843)	61223 (4985)	45342 (4350)
Multiplicity	5.3 (5.2)	13.6 (12.7)	20.7 (19.3)
Completeness (%)	73.06 (19.09)	93.34 (82.37)	99.58 (96.69)
Mean I/σ(I)	5.58 (0.77)	12.74 (1.70)	16.34 (1.95)
Wilson B factor (Å^2)	19.63	31.26	95.42
R_{merge}	0.2229 (1.818)	0.2608 (1.653)	0.1911 (1.716)
R_{meas}	0.2473 (2.022)	0.2709 (1.722)	0.1959 (1.762)
R_{pim}	0.1053 (0.8731)	0.07289 (0.4780)	0.04285 (0.3941)
$CC_{1/2}$	0.992 (0.385)	0.996 (0.667)	0.999 (0.681)
CC*	0.998 (0.745)	0.999 (0.895)	1.000 (0.900)
Reflections used in refinement	71183 (1843)	57192 (4985)	45321 (4348)
Reflections used for R_{free}	3421 (84)	1860 (153)	2266 (217)
R_{work}	0.2435 (0.2602)	0.1653 (0.2519)	0.2208 (0.3040)
R_{free}	0.2936 (0.3132)	0.2281 (0.3322)	0.2714 (0.3790)
CC_{work}	0.881 (0.761)	0.959 (0.738)	0.953 (0.725)
CC_{free}	0.849 (0.672)	0.911 (0.585)	0.956 (0.504)
Number of non-hydrogen atoms			
Total	10565	10903	15654
Macromolecules	10072	10307	15369
Ligands	85	56	156
Solvent	458	572	129
Protein residues	1312	1343	2003
R.m.s. deviations			
Bonds (Å)	0.004	0.008	0.003
Angles (°)	0.62	0.96	0.66
Ramachandran plot			
Favored (%)	97.45	96.10	95.53
Allowed (%)	2.47	3.75	4.17
Outliers (%)	0.08	0.15	0.30
Rotamer outliers (%)	0.79	1.12	0.00
Clashscore	16.59	12.17	8.52
Average B factors (Å^2)			
Overall	22.27	36.42	86.35
Macromolecules	22.17	36.30	86.51
Ligands	38.21	44.54	89.05
Solvent	23.19	38.22	62.94

Appendix 2.B: Interactions details for the LDH-B (PDB entry: 1I0Z) and mutant (LDH-B^{H67A}, LDH-B^{D64A} and LDH-B^{F71A}) structures. Interactions are reported for monomer A of each structure. Distances (Å) reported in the table is measured between the donor and acceptor atoms. The interactions were generated with Discovery Studio Visualizer[149]

	Partner 1	Partner 2	Type	d (Å)
LDH-B	Asp64 (A) [NH]	Gly60 (A) [C=O]	Hydrogen bond	2.94
	Gly68 (A) [NH]	Asp64 (A) [C=O]	Hydrogen bond	3.37
	H_2O (421)	Asp64 (A) [C-O^-]	Hydrogen bond	2.86
	H_2O (418)	Asp64 (A) [C-O^-]	Hydrogen bond	2.97
	Lys245 (B) [NH_3^+]	Asp64 (A) [C-O^-]	Hydrogen bond	2.85
	Asn249 (B) [NH]	Asp64 (A) [C-O^-]	Hydrogen bond	2.86
	Ala251 (B) [NH]	Asp64 (A) [C-O^-]	Hydrogen bond	3.18
	Lys245 (B) [NH_3^+]	Asp64 (A) [C-O^-]	Salt bridge	2.85
	His67 (A) [NH]	Met63 (A) [C=O]	Hydrogen bond	2.87
	H_2O (457)	His67 (A) [C=O]	Hydrogen bond	2.65
	H_2O (421)	His67 (A) [N]	Hydrogen bond	2.83
	Phe71 (A) [NH]	Gly68 (A) [C=O]	Hydrogen bond	2.90
	Leu254 – Ser255 (B)	Phe71 (A) [phenyl]	Amide-Pi stacked	4.29
LDH-B^{H67A}	Ala67 (A) [NH]	Met63 (A) [C=O]	Hydrogen bond	2.93
	H_2O (85)	Ala67 (A) [C=O]	Hydrogen bond	2.71
LDH-B^{D64A}	Ala64 (A) [NH]	Gly60 (A) [C=O]	Hydrogen bond	2.89
LDH-B^{F71A}	Ala71 (A) [NH]	Gly68 (A) [C=O]	Hydrogen bond	3.25



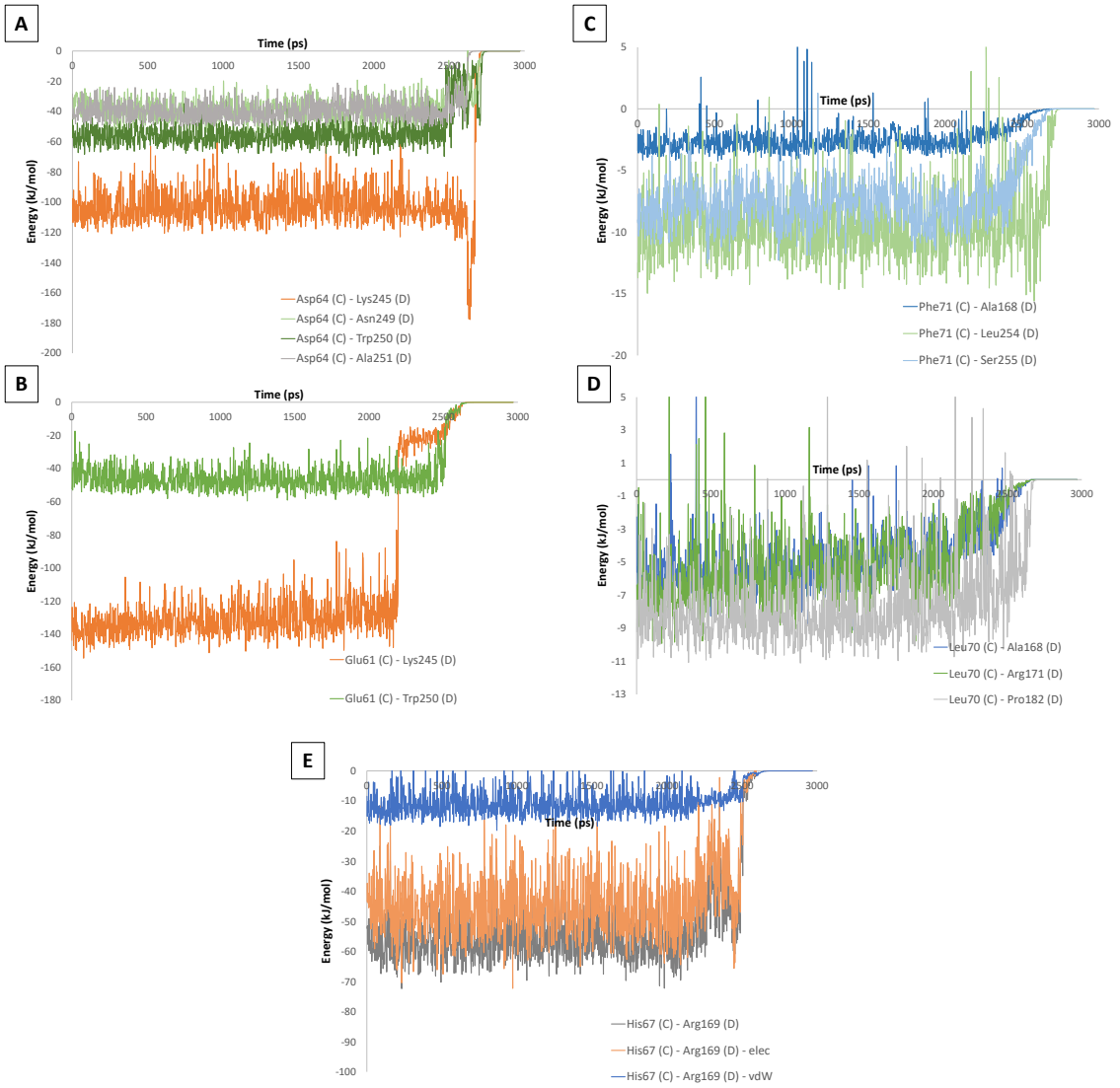
Appendix 2.C: (A) RMSD profiles of the 22-amino acid α-helix (monomer A) of LDH-B (cyan), LDH-B^{D64A} (green), LDH-B^{H67A} (brown) and LDH-B^{F71A} (orange) during the 300 ns MD simulation at 300 K and 1 bar – (B) RMSD profiles of the tetramerization site 2 of LDH-B (cyan), LDH-B^{H67A} (brown) and LDH-B^{F71A} (orange) during the 300 ns MD simulation at 300 K and 1 bar



Appendix 2.D: (A) Experimental (from crystallographic structure – in orange) and MD (in grey) RMSF profile of LDH-B^{H67A} C_α and MD RMSF profile of LDH-B C_α (in cyan) – (B) Experimental (from crystallographic structure – in brown) and MD (in blue) RMSF profile of LDH-B^{f71a} and MD RMSF profile of LDH-B C_α (in cyan)

Appendix 2.E-1: Mean interaction energy of the residue from the 22-amino acid α-helix with BD dimer calculated over the first 2000 ps of the SMD trajectory. The different contributions to the total interaction energy ($kJ.mol^{-1}$) for monomer C are presented: short-range Coulomb energy contribution ($kJ.mol^{-1}$) and short-range Lennard-Jones energy contribution ($kJ.mol^{-1}$). The total interaction energy ($kJ.mol^{-1}$) corresponds to the sum of Coulomb and Lennard-Jones energy contributions

	Coulomb contribution ($kJ.mol^{-1}$)	**Lennard-Jones contribution ($kJ.mol^{-1}$)**	**Total interaction energy ($kJ.mol^{-1}$)**
Leu54	-0.04 ± 0.01	-0.30 ± 0.03	-0.34 ± 0.03
Glu55	-0.47 ± 0.03	-1.01 ± 0.03	-1.48 ± 0.04
Asp56	-5.63 ± 0.40	-13.33 ± 0.16	-18.96 ± 0.43
Lys57	6.24 ± 0.58	-9.95 ± 0.97	-3.71 ± 1.13
Leu58	-0.26 ± 0.01	-1.19 ± 0.04	-1.45 ± 0.04
Lys59	1.85 ± 0.09	-10.18 ± 0.32	-8.33 ± 0.33
Gly60	-10.08 ± 1.40	-18.91 ± 0.26	-28.99 ± 1.42
Glu61	-176.78 ± 2.50	-3.75 ± 1.40	-180.53 ± 2.87
Met62	-0.86 ± 0.08	-2.27 ± 0.01	-3.13 ± 0.08
Met63	0.13 ± 0.10	-26.44 ± 0.31	-26.31 ± 0.33
Asp64	-282.43 ± 0.86	-23.69 ± 0.08	-306.12 ± 0.86
Leu65	-0.81 ± 0.05	-16.72 ± 0.31	-17.53 ± 0.31
Gln66	-0.10 ± 0.69	-15.63 ± 0.51	-15.73 ± 0.86
His67	-48.95 ± 0.53	-43.51 ± 0.38	-92.46 ± 0.65
Gly68	-0.04 ± 0.04	-13.37 ± 0.08	-13.41 ± 0.09
Ser69	-3.64 ± 0.47	-14.11 ± 0.71	-17.75 ± 0.85
Leu70	0.28 ± 0.23	-38.94 ± 0.75	-38.66 ± 0.78
Phe71	-9.24 ± 0.49	-42.68 ± 0.38	-51.92 ± 0.62
Leu72	-6.38 ± 0.74	-9.63 ± 0.21	-16.01 ± 0.77
Gln73	-16.25 ± 2.50	-22.31 ± 1.60	-38.56 ± 2.97
Thr74	2.32 ± 0.82	-4.94 ± 0.38	-2.62 ± 0.90
Pro75	-2.89 ± 0.26	-18.01 ± 0.93	-20.90 ± 0.97



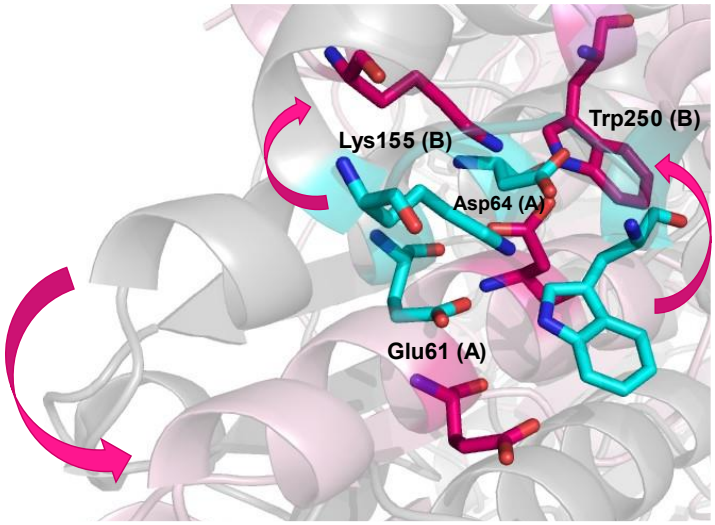
Appendix 2.E-2: Energy profiles for interactions established between Glu61 (C), Asp64 (C), His67 (C), Leu70 (C), and Phe71 (C) residues and the BD dimer – (A) Energy profiles (total interaction energy ($kJ.mol^{-1}$)) for interactions between Asp64 (C) and the BD dimer – (B) Energy profiles (total interaction energy ($kJ.mol^{-1}$)) for interactions between Glu61 (C) and the BD dimer – (C) Energy profiles (total interaction energy ($kJ.mol^{-1}$)) for interactions between Phe71 (C) and the BD dimer – (D) Energy profiles (total interaction energy ($kJ.mol^{-1}$)) for interactions between Leu70 (C) and the BD dimer – (E) Energy profiles (total interaction energy ($kJ.mol^{-1}$) in grey, Coulomb energy contribution in orange ($kJ.mol^{-1}$) and Lennard-Jones energy contribution ($kJ.mol^{-1}$) in blue) for the interaction between His67 (C) and the BD dimer

Appendix 2.E-3: Mean interaction energy for different interactions between Glu61 (C), Asp64 (C), His67 (C), Leu70 (C), and Phe71 (C) from the 22-amino acid α-helix and the BD dimer. The energy values were calculated over the first 2000 ps of SMD trajectory

	Coulomb contribution ($kJ.mol^{-1}$)	Lennard-Jones contribution ($kJ.mol^{-1}$)	Total energy ($kJ.mol^{-1}$)
Glu61 (C) – Lys245 (D)	-134.51 ± 2.40	2.90 ± 0.89	-131.61 ± 2.56
Glu61 (C) – Trp250 (D)	-44.98 ± 0.56	-1.52 ± 0.56	-46.50 ± 0.79
Asp64 (C) – Lys245 (D)	-106.81 ± 1.20	4.60 ± 0.21	-102.21 ± 1.22
Asp64 (C) – Asn249 (D)	-39.63 ± 0.57	0.40 ± 0.34	-39.23 ± 0.66
Asp64 (C) – Trp250 (D)	-47.99 ± 0.22	-8.23 ± 0.24	-56.22 ± 0.33
Asp64 (C) – Ala251 (D)	-35.51 ± 0.79	-4.37 ± 0.17	-39.88 ± 0.81
His67 (C) – Arg169 (D)	-45.39 ± 0.38	-11.25 ± 0.14	-56.64 ± 0.40
Leu70 (C) – Pro182 (D)	0.31 ± 0.10	-8.53 ± 0.07	-8.22 ± 0.12
Leu70 (C) – Arg171 (D)	4.05 ± 0.18	-9.83 ± 0.43	-5.78 ± 0.47
Leu70 (C) – Ala168 (D)	-0.29 ± 0.03	-5.02 ± 0.16	-5.31 ± 0.16
Phe71 (C) – Ala168 (D)	0.69 ± 0.02	-3.38 ± 0.06	-2.69 ± 0.06
Phe71 (C) – Leu254 (D)	1.69 ± 0.11	-11.49 ± 0.10	-9.80 ± 0.15
Phe71 (C) – Ser255 (D)	-0.45 ± 0.06	-7.34 ± 0.08	-7.79 ± 0.10

Appendix 2.E-4: Characterization of vdW interactions between Leu70 (C)/Phe71 (C) and BD dimer

	Interactions	$t_{interaction\ disappears}$ (ps)	$t_{energy\ reached\ 0}$ (ps)
Cluster 1	Leu70 (C) – Pro182 (D)	2200	2722
	Leu70 (C) – Arg171 (D)	2144	2696
	Leu70 (C) – Ala168 (D)	2134	2708
	Phe71 (C) – Ala168 (D)	2168	2702
Cluster 2	Phe71 (C) – Leu254 (D)	2652	2768
	Phe71 (C) – Ser255 (D)	2466	2736



Appendix 2.F: Rearrangement of the Glu61 (A) and Asp64 (A) residue environment during the SMD simulation (t = 0 ps in cyan and t = 2726 ps in pink)

Appendix 3

Structural study of LDH-B-ligand complexes – Identification of molecules interacting at the tetrameric interface

Appendix 3.A: Summary table of LDH-B(tr)-ligand complex crystallization assays – Crystallization assays for LDH-B are shown in orange in the table and for LDH-Btr in green. * LDH-B crystallization conditions (PDB entry: 1I0Z) correspond to 21% (w/v) PEG 8000, 100 mM Na-HEPES pH 7.5. ** Values in brackets are ligand concentration (ligand/protein ratio). Abbreviations used for ligands: maprotiline (MAP), triprolidine (TRI), imipramine (IMI), fluoxetine (FX), nicotinamide adenine dinucleotide (NADH) and oxamate (OXM)

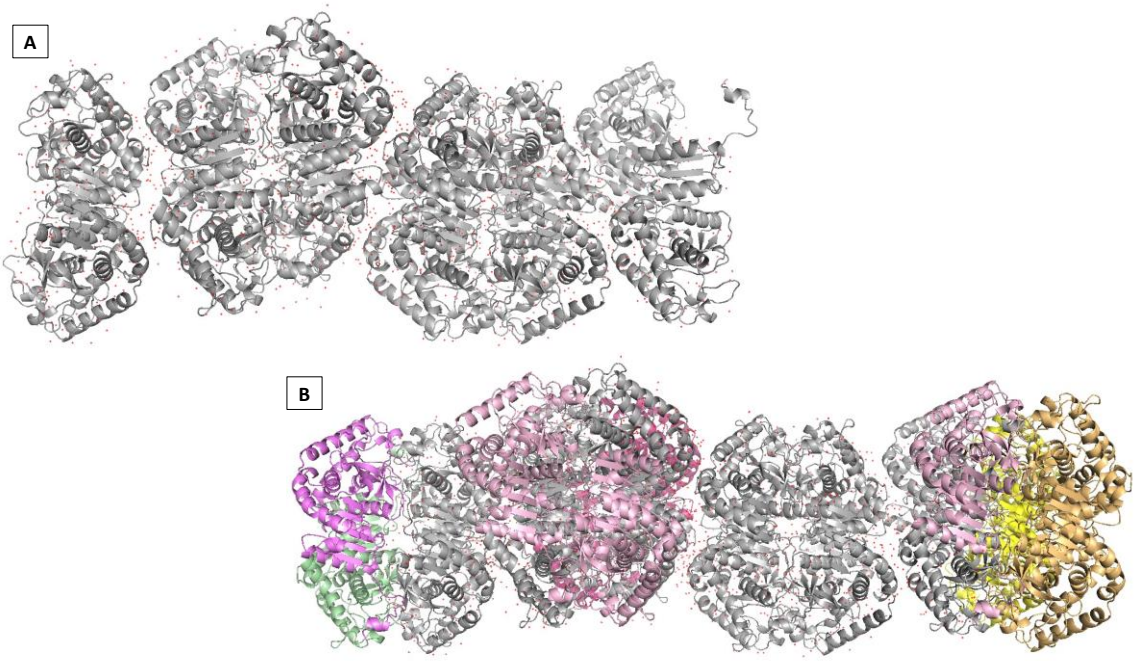
Screening	**Protein concentration ($mg.mL^{-1}$)**	**Kit**	**Ratio (condition/protein)**	**Temperature (°C)**	**Protein buffer**	**Ligands**	**Method**
1	15 - 17	LDH-B crystallization conditions (PDB entry: 1I0Z)*	1/1	RT	50 mM HEPES pH 7.5	MAP/TRI (3x**) + 2.5 mM NADH and 1 mM OXM or MAP/TRI (100x**) + 2.5 mM NADH and 1 mM OXM or MAP/TRI (100x**)	Hanging drop
2	14 - 15	JCSG-plus™ (MD1-37)	1/1	RT	50 mM Tris-HCl pH 7.5, 150 mM NaCl	MAP/TRI/IMI/FX (3x**) + 2.5 mM NADH + 1 mM OXM or MAP/TRI/IMI/FX (3x**)	Sitting drop
3	10		1/1	RT		FX (6x**) + 1.7 mM NADH + 0.7 mM OXM or FX (6x**)	Sitting drop
4	15	JCSG-plus™ (MD1-37)	1/1	RT	50 mM Tris-HCl pH 7.5, 150 mM NaCl	FX (x3**) + 2.5 mM NADH + 1 mM OXM or FX (3x**)	Sitting drop
5	10		1/1	RT		FX/TRI (6x**) + 1.6 mM NADH + 0.7 mM OXM or FX/TRI (6x**)	Sitting drop

Appendix 3.B: Data collection and refinement statistics for LDH-B-TRI and LDH-B-FX complex structures. Values for the highest-resolution shell are given in brackets

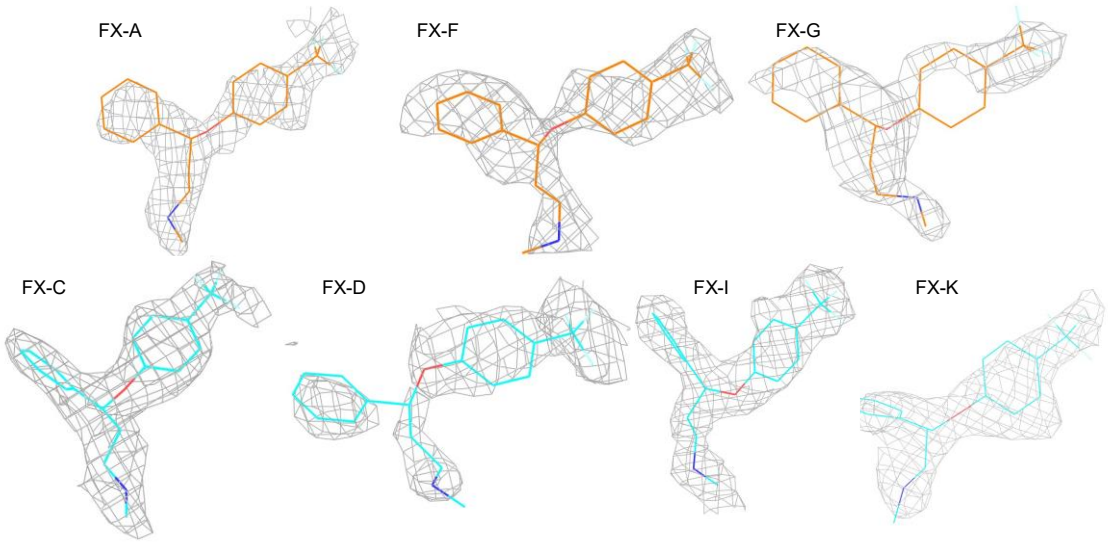
	LDH-B - TRI	**LDH-B - FX**
Wavelength	0.97857	0.98011
Resolution range (Å)	46.22 – 2.56 (2.653 – 2.562)	48.68 – 2.08 (2.149 – 2.08)
Space group	$P2_12_12_1$	$P2_1$
a, b, c (Å)	89.28, 130.34, 131.10	66.53, 413.96, 85.30
α, β, γ (°)	90, 90, 90	90, 109.57, 90
Total reflections	683583 (66139)	1835441 (173084)
Unique reflections	49797 (4801)	259434 (25018)
Multiplicity	13.7 (13.8)	7.1 (6.9)
Completeness (%)	99.75 (97.92)	99.35 (94.64)
Mean I/σ(I)	13.84 (2.42)	9.23 (1.30)
Wilson B factor (Å^2)	46.78	34.37
R_{merge}	0.2501 (1.191)	0.4705 (1.504)
R_{meas}	0.2599 (1.237)	0.5076 (1.623)
R_{pim}	0.06997 (0.33)	0.1894 (0.6036)
$CC_{1/2}$	0.994 (0.78)	0.956 (0.778)
CC*	0.999 (0.936)	0.989 (0.936)
Reflections used in refinement	49786 (4800)	258863 (24633)
Reflections used for R_{free}	2488 (240)	12943 (1232)
R_{work}	0.1759 (0.3658)	0.2099 (0.3141)
R_{free}	0.2458 (0.4670)	0.2637 (0.3771)
CC_{work}	0.965 (0.845)	0.953 (0.874)
CC_{free}	0.923 (0.676)	0.927 (0.764)
Number of non-hydrogen atoms		
Total	10976	33256
Macromolecules	10284	30768
Ligands	133	1133
Solvent	559	1355
Protein residues	1337	4000
R.m.s. deviations		
Bonds (Å)	0.008	0.008
Angles (°)	1.01	0.95
Ramachandran plot		
Favored (%)	96.16	95.77
Allowed (%)	3.76	4.07
Outliers (%)	0.08	0.15
Rotamer outliers (%)	1.55	1.30
Clashscore	6.73	7.64
Average B factors (Å^2)		
Overall	44.26	43.01
Macromolecules	44.01	42.96
Ligands	54.77	44.37
Solvent	46.38	43.00

Appendix 3.C: Interactions details for LDH-B-TRI and LDH-B-FX complex structures. Interactions for FX ligands from cluster 1 in orange and cluster 2 in cyan. Distances (Å) reported for the hydrogen bonds in the table are measured between the donor (D) and acceptor (A) atoms. Distances (Å) reported for the halogen bonds correspond to the distance between Fluor and A atoms. Distances influenced by the thermal displacement and the rotation of -CF_3 group. Angles (°) are defined by D, Fluor, and A atoms. Distances (Å) reported for the pi-pi stacking interactions correspond to the distance between the centroids of the aromatic rings. Angles (°) (γ and θ_2) are defined as the angle between the planes of the two aromatic rings and the deviation to the normal to the aromatic plane of the ligand, respectively. The interactions were generated with Discovery Studio Visualizer[149]

	Partner 1	**Partner 2**	**Type**	**d (Å)**	**Angle (°)**
TRI2	Lys155 (C) [NH_3^+]	TRI2 [pyridine cycle]	Hydrogen bond	2.95	
	Lys155 (C) [NH_3^+]	TRI2	Pi-cation	4.47	
FX-A	Trp148 (A)	FX [cycle F]	Pi-pi stacking (T-shape)	5.24	75.33 (γ) 8.66 (θ_2)
	FX [NH_2^+]	Glu13 (H) [C-O^-]	Salt bridge	5.25	
	FX [F]	Glu286 (A) [C-O^-]	Halogen bond	1.93	166.56
FX-B	FX [NH_2^+]	Glu13 (G) [C-O^-]	Salt bridge	3.98	
	FX [F]	Glu 286 (B) [C-O^-]	Halogen bond	1.94	155.34
FX-F	FX [NH_2^+]	Glu13 (I) [C-O^-]	Salt bridge	2.36	
	Lys155 (F) [NH_3^+]	FX [cycle F]	Pi-cation	3.92	
	Pro10 (I) [C=O]	FX [F]	Halogen bond	3.62	146.94
FX-G	FX [NH_2^+]	Glu13 (B) [C-O^-]	Salt bridge	3.59	
	Lys155 (G) [NH_3^+]	FX [cycle F]	Pi-cation	3.50	
FX-C	Lys155 (C) [NH_3^+]	FX [cycle F]	Pi-cation	4.50	
	Lys155 (C) [NH]	FX [O]	Hydrogen bond	2.36	
	FX [NH_2^+]	Gly152 (C)	Hydrogen bond	2.83	
	H_2O (1232)	FX [F]	Hydrogen bond	2.17	
FX-D	FX [NH_2^+]	Glu15(F) [C-O^-]	Salt bridge	5.30	
	FX [NH_2^+]	Gly152(D) [C=O]	Hydrogen bond	2.74	
	Trp148(D)	FX [cycle]	Pi-pi stacking	3.79	20.19(γ)
				4.29	19.04 (θ_2)
					20.35 (γ)
					32.61 (θ_2)
	His156 (D) [NH]	FX [F]	Hydrogen bond	3.59	
FX-I	FX [NH_2^+]	Glu15 (C) [C-O^-]	Salt bridge	5.51	
	H_2O (995)	FX [F]	Hydrogen bond	2.36	
FX-K	Lys155 (K) [NH_3^+]	FX [cycle F]	Pi-cation	4.86	
	FX [NH_2^+]	Gly152(K) [C=O]	Hydrogen bond	3.23	
	Lys155 (K) [NH_3^+]	FX [F]	Hydrogen bond	2.73	



Appendix 3.D: (A) Representation of asymmetric unit of LDH-B-FX complex structure – (B) Representation of LDH-B-FX complex structure (in grey) and crystallographic neighbors generated by symmetry (colored)



Appendix 3.E: Representation of the 2Fo-Fc map in grey mesh (σ = 0.7 and carve = 1.6) for FX ligands (cluster 1 in orange and cluster 2 in cyan)