

**UNIVERSITÀ DEGLI STUDI
DI MODENA E REGGIO EMILIA**

Dottorato di ricerca in Molecular and Regenerative Medicine

Ciclo XXXII

**O-phosphoethanolamine phospho-lyase:
from structural characterization to biomedical applications**

Candidato: Chiara Vettrano

Relatore (Tutor): Dr. Carlo Augusto Bortolotti

Correlatore (Co-Tutor): Dr. Emilio Parisini

Coordinatore del Corso di Dottorato: Prof. Michele De Luca

Table of Contents

List of Figures	6
List of Tables	10
Abstract	13
Sommario.....	15
Chapter 1. General introduction to PLP dependent enzymes.....	19
1. Pyridoxal 5'-phosphate (PLP) dependent enzymes	20
Chapter 2. Structural characterization of the human O-phosphoethanolamine phospho-lyase ..	25
2.1. Introduction	26
2.3. Materials and methods	27
2.3.1. Gene cloning.....	27
2.3.2. Protein production	29
2.3.4. Enzyme activity.....	30
2.3.5. UV-Visible spectroscopic measurements	30
2.3.6. Quaternary structure assessment	30
2.3.7. Protein crystallization.....	30
2.3.8. Data collection and processing.....	31
2.3.9. Structure solution and refinement.....	31
2.3.10. Multiple sequence alignment and protein-ligand interactions representations.....	31
2.4. Results and Discussion.....	31
2.4.1 Protein production	31
2.4.1 Protein crystallization.....	32
2.4.2. Data collection and processing.....	32
2.4.3. Structure solution and refinement.....	33
2.4.4. Overall Structure	34
2.4.5. Active site	38

2.5. Concluding remarks	41
Chapter 3. ETNPPL-based biosensor for PEA detection	42
3.1. Introduction	43
3.2. General principles of biosensors	46
3.2.1. Electrochemical biosensors	47
3.2.2. Field-effect transistors (FETs)	48
3.2.2.1. Electrolyte-gated field-effect transistors (EGOFETs)	49
3.3. Materials and methods	50
3.3.2. Electrolyte gated FET (EGOFET) fabrication	50
3.3.3. Instrumentation.....	50
3.3.4. His-tagged ETNPPL immobilization on gold gates	51
3.3.4.1. Absorption evaluation for His-tagged hETNPPL on gold surface	51
3.3.4.2. Biosensing experiments for His-tagged ETNPPL on Au surface	52
3.4. Results and discussion	53
3.5. Concluding remarks	57
Chapter 4. TpVar1-mediated delivery of human O-Phosphoethanolamine phospho-lyase in hepatic carcinoma cell line.....	59
4.1. Introduction	60
4.2. Material and Methods	62
4.2.1. Gene cloning	62
4.2.2. Protein production	63
4.2.3. Activity assays and kinetic analysis.....	63
4.2.4. Spectroscopic Measurements	64
4.2.5. Quaternary structure assessment	64
4.2.6. Circular dichroism spectroscopy.....	64
4.3. Results and discussion	65
4.4. Concluding remarks	68

Chapter 5. Structure-based approaches to the discovery of modulators of E-Cadherin-Mediated Cell–Cell Adhesion	69
5.1. Introduction	70
5.1.1. Cadherins as pharmaceutical targets	73
5.2. Exploring cadherin-modulator interactions using NMR and computational studies.....	76
5.3. Discovery of cadherin modulators by structure-based virtual screening	79
5.4. Materials and methods	87
5.4.1 Gene cloning.....	88
5.4.2 Protein production	88
Chapter 6. Experimental procedures.....	90
6.1. Heterologous protein expression.....	91
6.2. Chromatography techniques	93
6.3. Crystallization techniques.....	95
6.4. X-ray diffraction	98
6.5. Structure solution and refinement.....	101
6.5.1 Structure solution methods.....	102
6.6. X-ray sources.....	103
6.7. Immobilization methods.....	104
References:	107
Appendix I.....	121

List of Figures

Figure 1. Chemical structures of vitamin B6 and of pyridoxal-5'-phosphate (PLP).....	20
Figure 2. Pyridoxal-5'-phosphate linkage with enzyme and substrate to form internal and external aldimine, respectively.	21
Figure 3. Resonance structures of carbanionic intermediates. The quinonoid structure is shown on the right. Adapted from Toney M.D. (2011) ⁵	21
Figure 4. PLP-dependent enzymes classification according to Grishin et al. Examples of 5 PLP-dependent enzymes belonging to the main fold-types. Adapted from Rocha J. F. et al., (2019) ²²	22
Figure 5. Overview of the reaction catalyzed by ETNPPL.....	23
Figure 6. Proposed mechanism for the 1,2-elimination reaction catalyzed by ETNPPL, based on the known enzymatic mechanisms of other PLP-dependent lyases ³³ . After entering the active site and forming an initial Michaelis complex (ES I), PEA can attack the Schiff base formed by PLP with the active site lysine (internal aldimine) and – passing through an unstable gem-diamine intermediate (ES II) – leads to breaking up of internal aldimine and formation of external aldimine (ES III). This process is commonly termed transaldimination. After external aldimine formation, deprotonation at C α produces a PLP-stabilized carbanion or quinonoid intermediate (ES IV). Proton abstraction from the amino carbon, and the subsequent elimination of phosphate yields the PLP derivative of ethyleneamine (ES V). The eventual hydrolysis of ethyleneamine, forms acetaldehyde and ammonia. Adapted from Schioli D. et al. (2015) ³²	24
Figure 7. 10% SDS-PAGE assessing hETNPPL purity. Washes and elution fractions from NiNTA purification step (left). Elution fractions from size exclusion chromatography (right).	31
Figure 8. Overall structure of the hETNPPL PMP-bound form. Cartoon representation of the hEtnppl dimer. Chains A, and B are colored in blue and cyan, respectively.	35
Figure 9. DLS analysis of hETNPPL. The enzyme in solution features a hydrodynamic diameter (Size = 10 nm) that is consistent with a homodimer.....	35
Figure 10. Comparison of the hETNPPL dimer (blue/cyan) (PDB code: 6tor) and A1RDF1 from <i>Arthrobacter aureescens</i> TC1 (white) (PDB code: 5g4i). The α -helices and the β -strands are depicted as cylinders and arrows, respectively. The bound cofactor molecule is shown as grey sticks and circled.	36
Figure 11. Structure of the chain A of the hETNPPL. The N-terminal lobe (residues 1-85), large domain (residues 86-301) and C-terminal lobe (residues 302-499) are colored in cyan, magenta and green, respectively. The bound cofactor molecule is represented as grey sticks (left). Topology diagram of hETNPPL. Helices are indicated as rectangles and β -strands by arrows (right).....	36
Figure 12. PMP cofactor in the active site. (a) PMP molecule and K278 residue are represented in yellow and blue sticks, respectively. The 2Fo-Fc electron density map is depicted as gray mesh at 1 σ . Water molecules are shown as red spheres. (b) The absorption spectrum of hETNPPL shows a prominent peak at 410nm, which is typical of bound PLP, and a second peak centered at approximately 330 nm region which is indicative of the presence of bound pyridoxamine-5'-phosphate (PMP).	37
Figure 13. Dimer interface in the hETNPPL structure. (a) Surface representation of the dimer subunits. The critically important residues forming interface contacts are coloured in blue (chain A) and cyan (chain B). (b) LigPlot representation	

of the interaction between the two subunits. Hydrogen bonds are indicated by dashed lines, while those residues that form hydrophobic interactions are shown in spokes.....	38
Figure 14. Alignment of hETNPPL with A1RDF, identical and similar residues are boxed in red and yellow background, respectively. Catalytic lysine residue is indicated by a black arrow. Yellow and green arrows indicate the residues involved in binding the cofactor and in the PEA phosphate recognition, respectively.	39
Figure 15. Comparison of hETNPPL and A1RDF1 active site. Critically important residues of hETNPPL (green) and A1RDF1 (white) active site, and molecule bound are presented in sticks, and are numbered in the corresponding color. Residue from the opposite monomer are marked (*). Water molecules in our structure are shown as red spheres. (a) Superposition of hETNPPL to PLP molecule bound in the active site of A1RDF1 (PDB code: 5g4i). (b) Superposition of hETNPPL to PLP-PEA complex bound to the active site of A1RDF1 (PDB code: 5g4i).	40
Figure 16. Schematic representation of a typical biosensor. Adapted from Grieshaber et al. (2008) ⁸⁰	47
Figure 17. General scheme of an organic field-effect transistor (OFET) (a) and an electrolyte-gated organic field-effect transistor (EGOFET) (b). Adapted from Kergoat et al. (2012) ⁹³	49
Figure 18. Illustration of the compact and diffuse layers corresponding to EGOFET. Adapted from Kergoat et al. (2012) ⁹³	50
Figure 19. Schematic overview of the biosensor.....	51
Figure 20. Overview of the reaction scheme for hETNPPL (Bioreceptor) with its substrate PEA (Analyte), and the used coupled with ADH.....	52
Figure 21. Gold (Au) gates (left). Biosensor setup (right)	53
Figure 22. Measured activity for free enzyme (positive control), unwashed surface, washed surface, and the two washing solutions.....	53
Figure 23. Enzymatic activity ($\mu\text{mol}/\text{min}$) calculated for the free enzyme, the immobilized enzyme and the two washing solutions.....	54
Figure 24. Output characteristics of the device together with transfer curves recorded at different source-drain voltages.	55
Figure 25. A total of six transfer curves recorded for 1 mM concentration of PEA.....	55
Figure 26. Calibration curve of the biosensor.	56
Figure 27. Calibration curve in the presence of BSA.	57
Figure 28. 10% SDS-PAGE assessing TpVar1-hETNPPL purity. Flow through, washes and elution fractions from the NiNTA purification step (left). Comparison between the final purified TpVar1- hETNPPL and hETNPPL (right).....	65
Figure 29. Spectral properties of hETNPPL and TpVar1- Etnppl. (A) Absorption spectra of 10 μM hEtnppl (dashed line) and TpVar1 (solid line). (B) CD spectra of 2.5 μM hETNPPL (dashed line) and TpVar1-hETNPPL (solid line).	66
Figure 30. DLS analysis of hETNPPL and TpVar1-hETNPPL. The hydrodynamic diameter of both TpVar1-hETNPPL and hETNPPL is about 10 nm.	66
Figure 31. Kinetic characterization of hETNPPL and TpVar1- hETNPPL. The PEA phospho-lyase reaction catalyzed by recombinant hETNPPL (●) and TpVar1-ETNPPL (○). Rate of the elimination reaction as a function of the concentration of PEA at 30°C, pH 8.0. Fitting of the PEA data points to the Michaelis-Menten equation yielded the following values for the kinetic parameters: $K_m = 1.10 \pm 0.13 \text{ mM}$, $k_{cat} = 227 \pm 7$, $k_{cat}/K_m (\text{s}^{-1}\text{M}^{-1}) = 2.06 \times 10^5 \pm 5.4 \times 10^4$ for hETNPPL and K_m	

$=0.98 \pm 0.16$ mM and $K_{cat} = 95 \pm 4$ kcat/Km $=0.97 \times 10^5 \pm 2.5 \times 10^4$ for TpVar1. (B) pH-dependence of hETNPPL (●) and TpVar1- hETNPPL (○).	67
Figure 32. Schematic view of cadherin-mediated junctions (left). Schematic view merged with PDBs structures (right). Adapted from Brash et al. (2012) ¹²⁸	70
Figure 33. The cadherin superfamily.....	71
Figure 34. Schematic representation of the cadherin dimerization mechanism. The strand-swap dimer formation involves mutual insertion of the Trp2 side chain in the binding pocket of the partner molecule (right). The dynamic cadherin dimerization trajectory features a crucial intermediate configuration that is referred to as the X-dimer (center). Three calcium ions are shown as orange dots. Adapted from Dalle Vedove et al. (2019) ¹⁴⁴	72
Figure 35. Structure of ADH-1 and FR159. Adapted from Doro et al. (2015) ¹⁶⁰	74
Figure 36. Crystal structure of the human E-cadherin-(Val3)-EC1EC2 fragment (residues 3–213) in complex with FR159 (PDB code 4zte) (left). Hydrophobic interactions between the ligand and the protein. (right). Adapted from Nardone, V. et al. (2016) ¹³¹	75
Figure 37. The binding epitope of FR159 in the presence of wild type E-cadherin-EC1-EC2. The dots on the structure indicate the ligand binding epitope at different temperatures (left, blu 283K, green 290K and red 298K). The histogram shows the comparison of absolute STD % at different temperatures (right). Adapted from Civera et al. (2019) ¹⁶⁶	77
Figure 38. Docking binding mode of FR159 into E-cadherin. The ligand (grey carbon atoms) is superimposed to the DWVI sequence (light blue) of the E-cadherin X-ray crystal structure (PDB code: 3Q2V). Residues of the binding pocket interacting with the ligand are labelled. Adapted from Civera et al. (2019) ¹⁶⁶	78
Figure 39. AS11, AS9, and, to a lesser extent, AS8 regulate cell–cell adhesion. BxPC-3 E-cadherin/P-cadherin cells were incubated in the presence of either 0.1% DMSO or 0.05 mM AS2, AS8, AS9, AS11, or LC11 and allowed to form spheroids in suspension for 24 h. (A) Representative pictures of spheroids. Scale bar: 1000 μ m. (B) The spheroid area was measured by phase-contrast microscopy and analyzed by ImageJ. Values represent the mean of two experiments performed in octuplet. A $p < 0.05$ was considered statistically significant and is indicated by ‘***’ when $p < 0.001$. Adapted from Dalle Vedove et al. (2019) ¹⁴⁴	80
Figure 40. Cell surface cadherin extinction was assessed by both immunofluorescence and western blot. (A) Using RNA interference, E- and P-cadherin were stably knocked down in human pancreatic tumor BxPC-3 cells expressing both cadherins (see the Materials and Methods section for experimental details). Images were captured and analyzed using a SP5 Leica confocal microscope equipped with LAS AF Lite software. Scale bar: 25 μ m. (B) BxPC-3 E-cadherin/P-cadherin, BxPC-3 E-cadherin, and BxPC-3 P-cadherin cells were lysed, and the expressions of both P-cadherin and E-cadherin were detected by western blot. Adapted from Dalle Vedove et al. (2019) ¹⁴⁴	82
Figure 41. Both AS11 and AS9 target E-cadherin. BxPC-3 E-cadherin cells (P-cadherin depletion) (A) and BxPC-3 P-cadherin cells (E-cadherin depletion) (B) were incubated in the presence of 0.05 mM (E-cadherin cells) or 0.1 mM (P-cadherin cells) AS2, AS8, AS9, AS11, or LC11 and allowed to aggregate in suspension for 24 h. The spheroid area was measured by phase-contrast microscopy and analyzed by ImageJ. A $p < 0.05$ was considered statistically significant and is indicated by ‘***’ when $p < 0.001$ and ‘*’ when $p < 0.05$. See the Materials and Methods section for details on the statistical analysis. Adapted from Dalle Vedove et al. (2019) ¹⁴⁴	83
Figure 42. AS11 is more potent than AS9 to impair cell–cell interaction. BxPC-3 E-cadherin/P-cadherin cells (no cadherin depletion) (A) and BxPC-3 E-cadherin cells (P-cadherin depletion) (B) were incubated in the presence of AS9 or AS11 at	

various concentrations and allowed to form spheroids in suspension for 24 h. The spheroid area was measured by phase-contrast microscopy and analyzed by ImageJ. A $p < 0.05$ was considered statistically significant and is indicated by ‘***’ when $p < 0.001$. See the Materials and Methods section for details on the statistical analysis. Adapted from Dalle Vedove et al. (2019)¹⁴⁴. 83

Figure 43. BxPC-3 E-cadherin/P-cadherin cells were treated with 0.1% DMSO or 0.05 mM of inhibitor and allowed to form aggregates for 24 h. Cell viability was then assessed by a trypan blue exclusion test. Black arrows indicate dead cells. Scale bar: 400 μ m. Adapted from Dalle Vedove et al. (2019)¹⁴⁴. 84

Figure 44. 3D representation of the theoretical binding mode for AS9 (A) and AS11 (B). 2D ligand interaction diagram of the theoretical binding mode for AS9 (C) and AS11 (D). Hydrogen bond interactions are shown as pink arrows. Positive and negative charged amino acids are represented in blue and red, respectively. Residues involved in hydrophobic or polar interactions are shown in green and light blue, respectively. Ligand-exposed fractions are indicated as a gray, circular shadow. Adapted from Dalle Vedove et al. (2019)¹⁴⁴. 85

Figure 45. AS11, but not AS9, regulates cell invasion. BxPC-3 E-cadherin/P-cadherin cells (no cadherin depletion), BxPC-3 E-cadherin cells (P-cadherin depletion), and BxPC-3 P-cadherin cells (E-cadherin depletion) were allowed to form spheroids for 72 h. Spheroids were then embedded in type I collagen. After embedding, followed by a 24 h incubation in the presence of either 0.2% DMSO or 0.1 mM AS9 or AS11, the spheroid area was observed by phase contrast microscopy (A). Scale Bar: 500 μ m. (B): the spheroid area was measured using ImageJ. Results are expressed as the percentage of invasion of treated cells compared to control cells. Adapted from Dalle Vedove et al. (2019)¹⁴⁴. 86

Figure 46. Spheroid area ratio t48/t24 for AS9 and AS11. BxPC-3 E-cadherin/P-cadherin cells (no cadherin depletion) were allowed to form spheroids for 72 h. Spheroids were then embedded in type I collagen and incubated for 48 h in the presence of either 0.1% DMSO or AS9 (A) or AS11 (B) concentrations ranging from 0.01 to 0.5 mM. The spheroid area was observed by phase-contrast microscopy and measured after 24 and 48 h incubation. See the Materials and Methods section for details on the statistical analysis. A $p < 0.05$ was considered statistically significant and is indicated by ‘***’ when $p < 0.001$ and ‘*’ when $p < 0.05$. See the Materials and Methods section for details on the statistical analysis. Adapted from Dalle Vedove et al. (2019)¹⁴⁴. 87

Figure 47. The general components of a chromatographic system, as illustrated here by using a column to separate two chemicals, A and B. 93

Figure 48. Binding of the Histidine tag to bivalent metal ions. Two neighboring histidines from the His-tagged protein form a complex with the bivalent metal ions. 94

Figure 49. Schematic depiction of two common vapor diffusion techniques: hanging drop (left) and sitting drop (right). A drop with protein solution is situated either on top (hanging drop) or next to (sitting drop) a reservoir containing precipitant. As the water from the protein solution vaporizes, it transfers to the reservoir and crystallization occurs. . 96

Figure 50. The phase diagram for the crystallization of macromolecules. The diagram can be divided into two main regions, the undersaturation and the supersaturation region. The solid line separating them denotes maximum solubility at different concentrations of a precipitant (salt or polymers). The region of supersaturation is further divided into the metastable and labile regions. The metastable region corresponds to the conditions at which nucleation can occur, while the labile region, albeit not suitable for the nucleation process, can still promote to crystal growth from preformed nuclei introduced from outside the system using the so-called seeding technique. Crystals can only be grown from a supersaturated solution. The final region, which features extreme supersaturation, is denoted the precipitation region,

where the aggregation of the protein to form an amorphous precipitate is most probable. Adapted from McPherson et al. (2014).	97
Figure 51. Schematic depiction of Bragg's law (left) and the Ewald sphere (right).	100
Figure 52. Swiss Light Source (SLS) synchrotron of the Paul Scherrer Institute in Villigen (Switzerland). Pictures of SLS synchrotron ring (left) and X06DA-PXIII beamline (right).	104
Figure 53. Enzyme immobilization methods.	105

List of Tables

Table 1. Macromolecule-production information.	28
Table 2. Purification table for wild type hETNPPL.	32
Table 3. Crystallization conditions of hETNPPL.	32
Table 4. Data collection and processing. Statistics for the highest-resolution shell are shown in parentheses.	33
Table 5. Structure refinement. Statistics for the highest-resolution shell are shown in parentheses.	34
Table 6. The quantified free amino acids in urine samples of two studied groups using the LC-ESI-MS/MS method. <i>P</i> values for the comparison of the variables between two groups were calculated according to Mann-Whitney <i>U</i> test, Student's <i>t</i> -test or Welch's <i>F</i> test. AUC values were obtained in univariate ROC curve analyses. Adapted from Dereziński P. et al. (2017) ⁷⁴ .	45
Table 7. TpVar1-hETNPPL production.	62
Table 8. Purification table for TpVar1-hETNPPL.	65
Table 9. The steady-state kinetic parameters for the reactions catalyzed by hETNPPL and TpVar1- hETNPPL.	67
Table 10. Minimal medium composition.	88
Table 11. Crystal systems.	99
Table 12. List of the tested compounds.	121

Abstract

Enzymes have been used as biocatalysts in many productive sectors for more than a hundred years. However, over the last 50 years, since the advent of recombinant DNA technologies, they have found a much wider and diverse range of applications, spanning across virtually any kind of human activities. In particular, because of their remarkable properties, enzymes have become the major choice for medical diagnostic applications and are being increasingly utilized for both the detection and the treatment of a broad variety of diseases.

The purpose of this thesis is to study and exploit the high substrate specificity of a very peculiar human pyridoxal 5'-phosphate (PLP)-dependent enzyme, named O-phosphoethanolamine phospholyase (ETNPPL) for two possible applications, a diagnostic and a therapeutic one. The enzyme promotes the irreversible degradation of O-phosphoethanolamine (PEA) to acetaldehyde, phosphate and ammonia. Physiologically, ETNPPL is involved in the phospholipid metabolism, since PEA is the precursor of phosphatidylethanolamine (PE) in the CDP-ethanolamine-Kennedy pathway.

The physico-chemical characterization of ETNPPL reported in the literature suggests a mechanism of action that is very similar to other PLP-dependent enzymes, implying that the same residues are present in the active site. However, a structural characterization of human ETNPPL (hETNPPL) has not yet been reported in the literature.

The first issue that I addressed in this thesis was to obtain the high-resolution 3D-structure of hETNPPL by X-ray crystallography in order to validate the mechanism of action proposed in the literature. To this aim, starting from the cloning of the DNA fragment encoding for the hETNPPL into a suitable expression vector, I expressed and purified the enzyme. After assessing the enzymatic activity of the purified protein sample, I carried out crystallization experiments and I obtained good quality crystals for X-ray data collection. I determined the structure of hETNPPL (EC 4.2.3.2), in complex with pyridoxamine-5'-phosphate (PMP), at 2.05 Å resolution by molecular replacement using the structure of A1RDF1 from *Arthrobacter aurescens* TC1 (PDB 5g4i) as the search model.

Once I have completed the structural characterization of the enzyme and I have fully rationalized its high substrate specificity for phosphoethanolamine (PEA), I set out to investigate one of its possible practical applications by designing and working towards the development of a biosensor for the selective detection of PEA in the urine. This part of the work was done in collaboration with the group of Dr. Mario Caironi (Italian Institute of Technology, Milano). PEA levels in the urine of prostate cancer patients have been reported to decrease with respect to healthy patients. Therefore, an

ETNPPL-based assay for the measurement of PEA concentrations would be most useful for prostate cancer screening purposes. To date, the early detection of prostate cancer involves the prostate-specific antigen (PSA) blood test. However, this approach is known to produce high rates of false positive results since increased PSA levels are often found in men with benign prostatic hyperplasia (BPH), prostatitis and prostate injury. For this reason, the need for new biomarkers that can be used for diagnostic purposes either alone or in combination with standard PSA tests appears evident.

A method for the rapid and quantitative detection of this analyte would be of relevance for the diagnosis of such disease and, at the same time, would represent a potent validation tool of PEA as a *bona fide* biomarker for prostate cancer.

A second aim of this thesis work was to develop a delivery system to transport ETNPPL into hepatic carcinoma cells (HCC), where a down regulation of the enzyme has been associated to an increased rate of *de novo* lipogenesis (DNL). DNL ability is possessed only by liver, adipose and lactating breast tissue once the organism enters adulthood. However, due to their constantly replicating nature, cancer cells employ DNL as an alternative living pathway in order to meet their structural and energy-providing lipids requirements. Since PEA is one of the four major phospholipid precursors, decreased levels of ETNPPL should favour the availability of fatty acids that are crucial for cancer cell proliferation. The goal of this part of my work is to develop an ETNPPL delivery system that allows the investigation of the effect of restoring cellular levels of ETNPPL on the lipid reservoirs in the cells and, as a consequence, on the proliferation of cancer cells.

Finally, the last project described in this thesis was carried out in close collaboration with the groups of Prof. Laura Belvisi (University of Milano, Milano), Prof. A. Cavalli (Italian Institute of Technology, Genova) and Prof. Frédéric André (Aix-Marseille Université, France). This work dealt with the discovery and investigation of small-molecule modulators of cadherin-mediated cell-cell adhesion. While cadherin aberrant expression is known to correlate with many pathological conditions, only a limited arsenal of cadherin homo-dimerization inhibitors have been developed to date, the majority of them being expensive and often unspecific monoclonal antibodies. Therefore, the identification of drug-like molecules that can be effective in modulating cadherin-mediated cell-cell adhesion would likely help pave the way for the development of viable therapeutic strategies in many pathological settings.

Sommario

Le prime applicazioni industriali di enzimi risalgono a più di cento anni fa. Tuttavia, è solo negli ultimi 50 anni che, grazie all'avvento delle tecnologie del DNA ricombinante, il loro impiego ha subito una forte crescita in moltissimi settori industriali. In particolar modo, grazie alle loro interessanti proprietà, gli enzimi stanno trovando grande impiego in campo medico, principalmente nella diagnostica e nel trattamento di un sempre crescente numero di patologie.

Lo scopo di questa tesi è quello di studiare l'elevata specificità di substrato di un enzima umano PLP-dipendente, detto O-fosfoetanolamina fosfo-liasi (ETNPPL), per lo sviluppo di due possibili applicazioni, un'applicazione diagnostica ed una terapeutica. ETNPPL catalizza una reazione irreversibile in cui si ha la degradazione della fosfoetanolamina (PEA) con formazione di acetaldeide, ioni fosfato e ammonio. La caratterizzazione fisico-chimica di questo enzima, già riportata in letteratura, suggerisce un meccanismo di azione simile a quello di altri enzimi PLP-dipendenti. Tale ipotesi implica che nel sito attivo dell'enzima siano presenti gli stessi residui aminoacidici. Tuttavia, ad oggi, una conferma di questa ipotesi mediante caratterizzazione strutturale non è stata ancora ottenuta.

Il primo obiettivo di questo lavoro di tesi è stato quello di ottenere la struttura tridimensionale ad alta risoluzione dell'enzima umano hETNPPL attraverso cristallografia a raggi-X. Per questo scopo, è stato effettuato il clonaggio del gene codificante per la hETNPPL umana all'interno di un vettore che ha permesso l'espressione eterologa dell'enzima in cellule di *E.coli*. L'enzima così prodotto è stato purificato mediante cromatografia in fase liquida e, una volta testata la sua attività enzimatica, è stato cristallizzato per gli esperimenti di cristallografia ai raggi-X. La struttura 3D dell'enzima hETNPPL è stata infine determinata con una risoluzione di 2.05 Å.

Una volta razionalizzato a livello molecolare il meccanismo alla base dell'elevata specificità dell'enzima per il suo substrato naturale, la fosfoetanolamina (PEA), si è deciso di studiarne un'applicazione pratica tramite la progettazione e lo sviluppo di un biosensore in grado di rilevare selettivamente tale substrato. Questa parte del lavoro è stata sviluppata in collaborazione col gruppo del Dr. Mario Caironi (Istituto Italiano di Tecnologia, Milano). L'analisi si colloca in un più ampio contesto legato allo screening del tumore alla prostata che, ad oggi, viene effettuato utilizzando principalmente il test della PSA. Dal momento che livelli elevati di PSA possono essere rinvenuti anche in presenza di altre patologie, il test della PSA comporta un alto numero di falsi positivi. Per questo motivo appare evidente la necessità di ricercare nuovi biomarkers del tumore alla prostata che siano in grado di aumentare l'attendibilità dei risultati dello screening, eventualmente anche supportando il test della PSA.

Recentemente la PEA è stata indicata come possibile biomarker del tumore alla prostata. I livelli di PEA nelle urine di pazienti affetti da questa patologia risultano infatti significativamente inferiori rispetto a quelli delle persone sane. Di conseguenza, un metodo rapido e quantitativo per l'identificazione di PEA nelle urine, basato sull'affinità dell'enzima ETNPPL per il suo substrato naturale, potrebbe essere particolarmente rilevante per la diagnosi della malattia e, al tempo stesso potrebbe essere utilizzato come metodo per validare la PEA come biomarker del tumore alla prostata.

Il secondo obiettivo di questo lavoro di tesi è stato quello di sviluppare un sistema di trasporto di hETNPPL all'interno di cellule di carcinoma epatico (HCC). Questo studio si basa sulla recente osservazione che bassi livelli di espressione di hETNPPL all'interno di queste cellule sono in relazione ad una aumentata lipogenesi. Tale aumento è uno dei meccanismi alternativi che le cellule tumorali utilizzano per la propria sopravvivenza. Infatti, queste cellule hanno una velocità di crescita e proliferazione molto elevata che deve essere sostenuta da una continua fornitura di lipidi sia strutturali che energetici. Dal momento che la fosfoetanolamina (PEA), substrato naturale di hETNPPL, è un precursore fosfolipidico, una diminuzione dei livelli dell'enzima dovrebbero favorire la disponibilità lipidica.

L'obiettivo di questo lavoro è quello ingegnerizzare hETNPPL in modo da permettere il trasporto dell'enzima all'interno di cellule HCC e di poter investigare gli effetti di una sua eventuale ri-somministrazione sulla proliferazione delle cellule tumorali.

Infine, un ulteriore progetto descritto in questa tesi riguarda la ricerca e lo studio di piccole molecole in grado di inibire una particolare classe di proteine di adesione cellulare note come caderine, le quali svolgono un ruolo preminente nello sviluppo di gravi patologie quali cancro o malattie legate a processi infiammatori. Lo studio è stato condotto in collaborazione con i gruppi della Prof. Laura Belvisi (Università di Milano), del Prof. A. Cavalli (Istituto Italiano di Tecnologia, Genova) e del Prof. Frédéric André (Aix-Marseille Université, Francia). Fino ad ora sono state identificate solo poche molecole in grado di modulare il processo di adesione cellulare mediato dalle caderine. Poiché la maggior parte di esse sono costosi e spesso aspecifici anticorpi monoclonali, l'identificazione di piccole molecole che possano inibire o modulare il processo di omodimerizzazione delle caderine potrebbe rappresentare un nuovo e promettente approccio terapeutico in molti stati patologici.

Chapter 1. General introduction to PLP dependent enzymes

1. Pyridoxal 5'-phosphate (PLP) dependent enzymes

Pyridoxal 5'-phosphate (PLP) is the active phosphorylated form of vitamin B₆. It is used in almost all the organisms by a variety of different enzymes, thus representing the most versatile organic cofactor in biology¹⁻⁵.

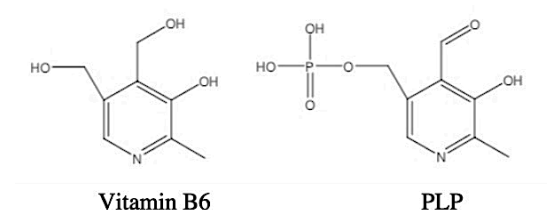


Figure 1. Chemical structures of vitamin B6 and of pyridoxal-5'-phosphate (PLP).

PLP-dependent enzymes catalyze a wide variety of reactions involving amino compounds, such as racemization, decarboxylation, transamination, retro-aldol cleavage, α -elimination and β - or γ -replacement reactions, among others^{6,7}. Since these enzymes are often involved in many key cellular processes and metabolic pathways, their dysregulation has been often correlated to several pathological conditions. For instance, deficiency of alanine-glyoxylate aminotransferase, is involved in the primary hyperoxaluria type I, a rare metabolic disorder characterized by the formation of calcium oxalate stones mainly in the urinary tract⁸, while the downregulation of the O-phosphoethanolamine phospho-lyase (ETNPPL) has been associated with several clinicopathological features such as alpha fetoprotein (AFP) serum level and T stage. Furthermore, ETNPPL has been demonstrated to play a crucial role in the lipid metabolism of HCC tissue, negatively regulating the lipogenesis of hepatocarcinoma cells (HCC)⁹. Moreover, its aberrant expression is thought to contribute in schizophrenia and bipolar disorder¹⁰⁻¹². The deficiency of L-3,4-dihydroxyphenylalanine (L- dopa) decarboxylase (DDC), the enzyme that decarboxylates L-dopa and 5-hydroxytryptophan to produce the two neurotransmitters dopamine and serotonin respectively, leads to development delay, abnormal movement, and other neurotransmitter-related symptoms¹³. These few examples are only a small portion of PLP-dependent enzymes whose dysregulation has a clinical implication.

Due to their physiological relevance, PLP-dependent enzymes are currently being the subject of numerous structure-function relationship studies many of which focus on how the active site residues facilitate PLP to increase the reaction specificity.

A key feature that has already emerged from the investigations of this broad class of enzymes is that despite their functional diversity, reactions catalyzed by PLP-dependent enzymes are characterized by a series of common mechanistic features related to the PLP specific composition. All PLP-dependent enzyme-catalyzed reactions begin with a Schiff base-exchange reaction (transimination). Indeed, with rare exceptions, in which PLP binds to the active site through the phosphate group⁴, in all PLP enzymes the aldehyde group of the cofactor is covalently linked, via an imine bond, to the ϵ -amino group of an active-site lysine residue to form an “internal” aldimine. In the subsequent step, the amino group of the incoming substrate replaces the ϵ -amino group of the lysine, forming a coenzyme-substrate Schiff Base (external aldimine) (**Figure 2**). This “external” aldimine is the common central intermediate for all PLP-catalyzed reactions. Starting from this intermediate and depending on which bond at C α is cleaved, a differentiation in terms of reaction specificity can occur.

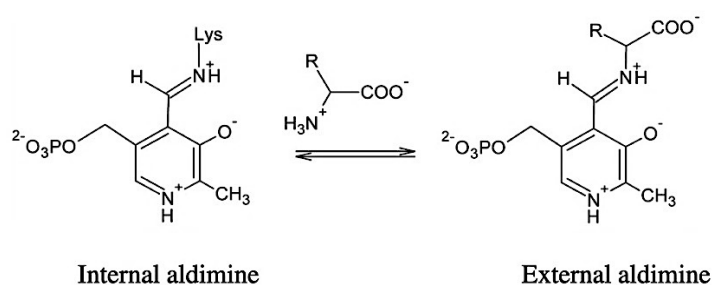


Figure 2. Pyridoxal-5'-phosphate linkage with enzyme and substrate to form internal and external aldimine, respectively.

In addition to the aldehyde group, also the heteroaromatic pyridine ring of PLP has a fundamental function. Indeed, with the exceptions of the aminomutase family¹⁴, pyridine ring enables the stabilization of carbanionic intermediate formed by proton abstraction from the carbon attached to the Schiff base nitrogen⁵. The carbanionic intermediate is commonly called “quinonoid” intermediate because of the quinone-like structure of one of its resonance forms (**Figure 3**).

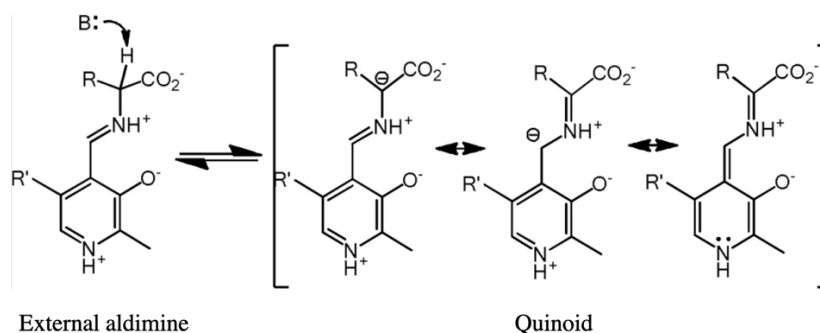


Figure 3. Resonance structures of carbanionic intermediates. The quinonoid structure is shown on the right. Adapted from *Toney M.D. (2011)*⁵.

The role of the other two functional groups of PLP, the hydroxyl and phosphate groups, is rather uncertain. The hydroxyl group could function as a proton donor or acceptor while the role of the phosphate group of PLP has been proposed only for few enzymes. For example in the reaction catalyzed by glycogen phosphorylase, the phosphate group seems to act as a general acid/base for accepting or donating a proton^{15,16}. Recently, the role of the phosphate group in PLP was also investigated in serine palmitoyltransferase, where the substitution of pyridoxal 5'-phosphate with pyridoxal, lowered more than 10-fold the enzyme activity¹⁷, suggesting the importance of the phosphate group in the intermediate formation and stereospecific orientation of carbanionic intermediates.

To date, many PLP-dependent enzymes have been structurally characterized. According to Grishin *et al.*, they were grouped in 5 main fold-types¹⁸ (**Figure 4**). Subsequently, two additional fold-types were introduced to classify some enzyme families that could not be categorized into these main classes^{18–21}.

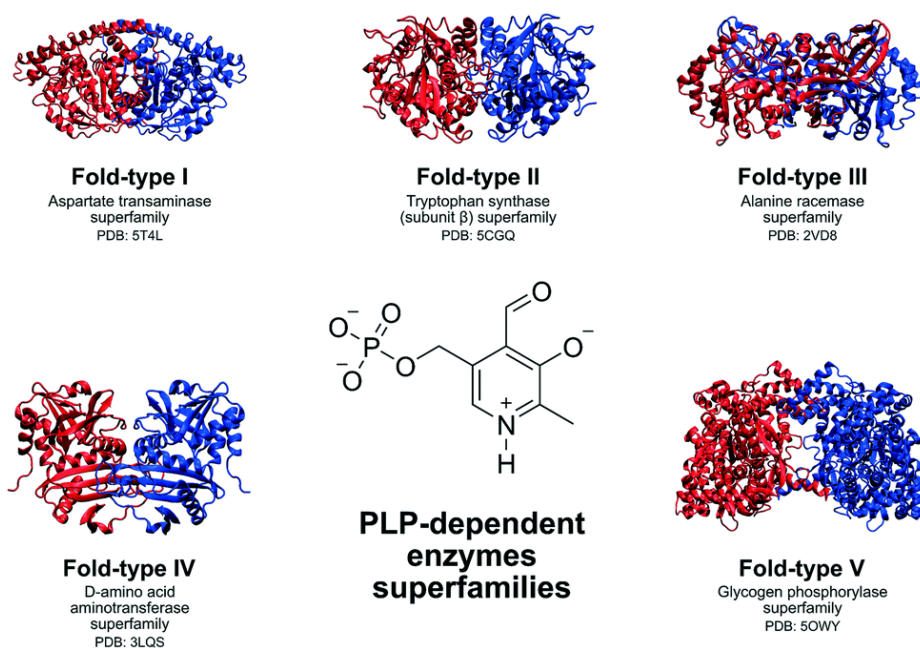


Figure 4. PLP-dependent enzymes classification according to Grishin *et al.* Examples of 5 PLP-dependent enzymes belonging to the main fold-types. Adapted from Rocha J. F. *et al.*, (2019)²².

The work described in this thesis focuses on a subfamily of pyridoxal phosphate (PLP)-dependent enzymes belonging to the broader fold-type I structural group, named subgroup-II aminotransferases (AT-II)²³ or class-III aminotransferases¹⁸.

Among the different classes, fold-type I enzymes is the group that features the highest degree of functional and structural diversity. It includes aminotransferases, decarboxylase groups II and III^{18,24} and enzymes with α -, β - or γ -elimination activity²⁵.

This subclass can be represented by the two prototypical enzymes, ornithine δ -transaminase and γ -aminobutyrate (GABA) transaminase^{2,23,26,27}, which are responsible for the transference of an amino group from an amino donor to an acceptor. Generally, in most AT-II, the preferred amino group acceptors are either α -ketoglutarate (α -KG) or pyruvate, whereas the amino group donors are molecules with an amino group located distal to a carboxylate and usually at the end of an alkyl chain. The latter are conventionally called ω -amines, and their transamination leads to production of an aldehyde (or rarely a ketone), as shown in the following equation:



Starting from the first ω -amine-specific transaminases discovered, numerous studies have been carried out on this subgroup, providing a great amount of biochemical and structural information. Although the number of known ω -amine-specific transaminases has grown steadily over time²⁸, many studies have also identified some strictly related AT-II members that do not act as transaminases. Among them, the human O-phosphoethanolamine phospho-lyase (ETNPPL) is one of the most peculiar enzymes acting on terminal amines. The enzyme is also called alanine-glyoxylate aminotransferase 2-like 1 (AGXT2L1)¹² precisely because of its high sequence identity (36%) with alanine-glyoxylate aminotransferase 2 (AGXT2). However, despite their close resemblance, ETNPPL has been demonstrated to act as a lyase rather than an aminotransferase, promoting the irreversible degradation of O-phosphoethanolamine (PEA) to acetaldehyde, phosphate and ammonia^{29–31} (**Figure 5**).

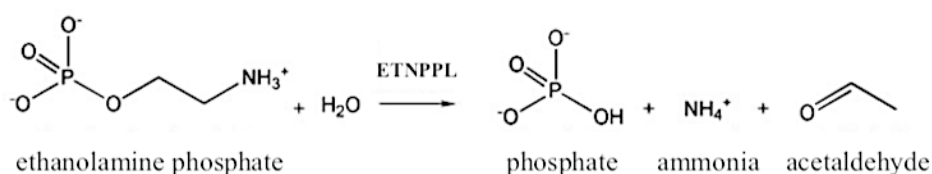


Figure 5. Overview of the reaction catalyzed by ETNPPL.

During the last years, a series of studies have been performed on this particular enzyme. The physico-chemical characterization of ETNPPL suggests a mechanism of action³² that is very similar to other PLP-dependent enzymes (**Figure 6**). However, a structural characterization of human ETNPPL (hETNPPL) has not yet been reported in the literature.

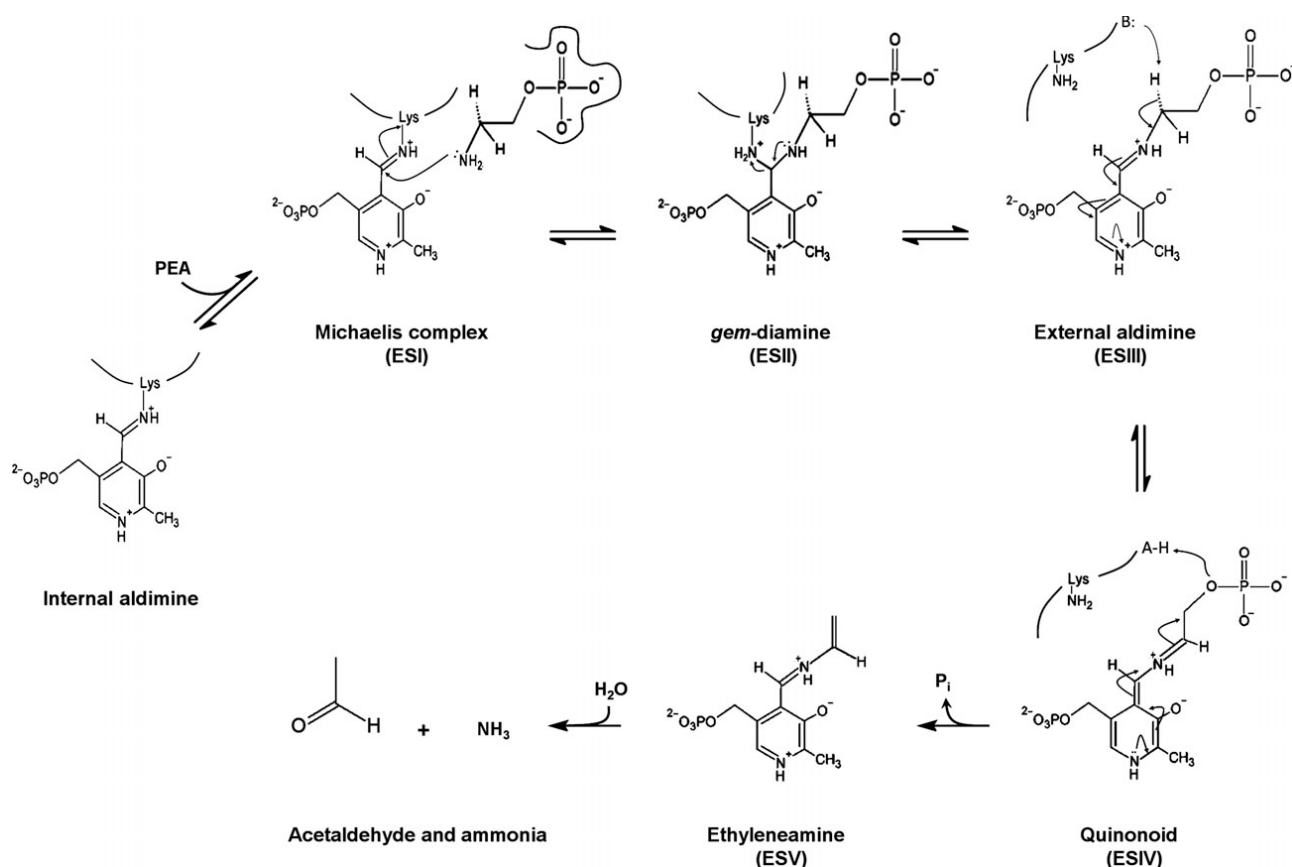


Figure 6. Proposed mechanism for the 1,2-elimination reaction catalyzed by ETNPPL, based on the known enzymatic mechanisms of other PLP-dependent lyases³³. After entering the active site and forming an initial Michaelis complex (ES I), PEA can attack the Schiff base formed by PLP with the active site lysine (internal aldimine) and – passing through an unstable gem-diamine intermediate (ES II) – leads to breaking up of internal aldimine and formation of external aldimine (ES III). This process is commonly termed transaldimination. After external aldimine formation, deprotonation at C α produces a PLP-stabilized carbanion or quinonoid intermediate (ES IV). Proton abstraction from the amino carbon, and the subsequent elimination of phosphate yields the PLP derivative of ethyleneamine (ES V). The eventual hydrolysis of ethyleneamine, forms acetaldehyde and ammonia. Adapted from Schioli D. *et al.* (2015)³².

In the following chapter, the high-resolution 3D-structure of hETNPPL determined by X-ray crystallography allows to complete the biochemical characterization of the enzyme thus rationalizing for the first time the proposed mechanism of action.

Chapter 2. Structural characterization of the human O-phosphoethanolamine phospho-lyase

2.1. Introduction

Enzymes dependent on the cofactor pyridoxal 5'-phosphate (PLP) are catalytically very versatile, being responsible for example of nearly all metabolic transamination, decarboxylation, racemization, β - and γ -elimination reactions on amino acids⁵. In addition, PLP enzymes play key roles in the metabolism of sphingolipids³⁴ and carbohydrates¹⁵ as well as in the biosynthesis or degradation of other compounds such as neurotransmitters, nucleotides and heme¹⁹. Despite this variety of roles, PLP-dependent enzymes are often structurally similar, as they belong to just a small number of independent evolutionary lineages¹⁹. Accordingly, for these enzymes, prediction of activity based on sequence is not straightforward. Furthermore, given their manifold cellular functions, even for those enzymes whose catalytic activity is established, the actual biological role may be difficult to demarcate.

One exemplary case is the human O-phosphoethanolamine phospho-lyase (hEtnppl, EC 4.2.3.2). This enzyme was originally named AGXT2L1 (AGXT2-like 1) based on its homology to alanine glyoxylate aminotransferase 2 (AGXT2), a mitochondrial PLP-dependent enzyme that transaminates a variety of substrates³⁵ while also acting as a lyase on L-cysteine conjugates³⁶. Unlike AGXT2, however, hETNPPL is almost completely inactive as a transaminase^{12,37}. Instead, it was shown to catalyze the irreversible 1,2-elimination of O-phosphoethanolamine (PEA) to produce acetaldehyde, phosphate and ammonia²⁹.

Moreover, a thorough investigation of the substrate selectivity of the enzyme has provided evidence of a strong discrimination against PEA-like amino acid compounds^{12,32}.

In light of these data, it has been speculated that hETNPPL may be involved in phospholipid metabolism, since in the CDP-ethanolamine-Kennedy pathway³⁸ PEA is the precursor of phosphatidylethanolamine (PE), one of the four main membrane phospholipids. In a previous work, it has been demonstrated that the knock-down of hETNPPL disrupts the normal synthesis of lipids in Hepatocellular carcinoma (HCC) and Cholangiocellular carcinoma (CCA) tissues⁹, promoting an unusual activated lipogenesis (called *de novo* lipogenesis, DNL) that plays an essential role in supplying phospholipids during tumorigenesis^{39,40}. The same study provided an *in silico* screening of the expression levels of hETNPPL in other types of tumors, demonstrating that the expression of the enzyme decreases both in ovarian cancer and in breast cancer tissues. These data imply that hETNPPL has a PEA catabolic activity and regulates the synthesis of PE in the CDP-ethanolamine-Kennedy pathway. In addition, since the enzyme is mainly expressed in the liver and in the brain, it is not surprising that a dysregulated expression of the enzyme appears to have clinical significance also in some severe mental conditions. Indeed, earlier works had showed that hETNPPL is the most consistently

upregulated gene in neuropsychiatric diseases such as schizophrenia and bipolar disorder^{10,41}. Although the catalytic function of hETNPPL is known²⁹ and a scheme for its reaction mechanism has been proposed³² (**Figure 6**), no structural information regarding this enzyme has ever been provided.

In the following sections, I will describe the work I have done to obtain the high resolution crystal structure of hETNPPL. This part of my thesis work will be the subject of a publication that is currently in preparation (Vettraino C, Peracchi A, Donini S, Parisini E. Structural characterization of the human O-phosphoethanolamine phospho-lyase. To be submitted).

2.3. Materials and methods

2.3.1. Gene cloning

To construct a pET23a-hETNPPL plasmid, the complete *hetnppl* gene (NP_112569.2) was PCR amplified from a pET28Cpo-hETNPPL vector³⁷. The amplification product was sub-cloned into the pET23a vector (Novagen) using the NdeI and XhoI restriction sites and fused at its C-terminus with a 6His-tag (**Table 1**). All the primers used were supplied by Eurofins scientific or Metabion as lyophilized powder, while restriction enzymes were purchased from Thermo-Scientific. After both PCR and digestion with restriction enzymes, the DNA fragments were purified from agarose gel (1% in 40 mM Tris, 20mM acetic acid and 1mM EDTA "TAE" buffer) using commercial DNA extraction kits (purchased from either Promega or Macherey - Nagel). DNA products were dissolved in MilliQ water and validated by DNA sequencing.

Both the DNA fragment obtained after the PCR and the pET23a plasmid were digested by incubation with the NdeI and XhoI restriction enzymes at 37°C for 2 hours. In order to prevent the possible recircularization of partially digested vectors, a dephosphorylation reaction was carried out by incubating the digested vector after purification for 20 min at 37°C with FastAP Alkaline Phosphatase (Thermo Scientific). A further incubation at 75° for 5 min allowed to inactivate the enzyme. The subsequent ligation reaction was done overnight at 4°C using a 1:4 vector to insert molar ratio. In order to increase the amount of the modified plasmid, *E. coli* XL1B cells (Invitrogen) were transformed with the ligation products using the heat shock method. Briefly, 50 µl of XL1B competent cells were incubated with 10 µl of the ligated mixture for 20 minutes on ice, then left for 90 seconds in a water bath at 42°C (to allow the DNA to diffuse into the cells) and subsequently moved back on ice for 2 minutes. The mixture was added to 950 µl of LB broth (10 g/l NaCl, 10 g/l of tryptone enzymatic digest, 5 g/l yeast extract), incubated at 37°C with orbital shaking at 180 rpm for 45 minutes, pelleted for 6 minutes at 8000 rpm, resuspended in 100 µl of LB broth and finally spread on

LB/agar (same as the broth plus 18 g/l of LB-Agar) plates supplemented with 100 µg/ml of ampicillin. The plates were then incubated overnight at 37°C. To assess whether if the cloning was successful, single colonies were picked from the plate and grown overnight in 10 ml of LB broth supplemented with 100 µg/ml of ampicillin. The cells were then pelleted by centrifugation at 4500 rpm (4°C) and the plasmid was extracted using a Plasmid Miniprep kit (QUIAGEN). In order to visually check the correct insertion of the gene into the vector, the obtained plasmid was digested with the restriction enzymes (NdeI and XhoI) and run on 1% agarose gel. Those plasmids that showed a relevant band in this test were sent to a sequencing facility to confirm the nature of the inserted gene.

Table 1. Macromolecule-production information

Source organism	<i>H. sapiens</i>
DNA source	pET28Cpo-hETNPPL
Forward*	Tttttc <u>atatgt</u> gcgagctgtacagtaag
Reverse*	Tttctc <u>gag</u> gtgtcttgagcctcttact
Expression vector	pET23a
Expression host	<i>E. coli</i> (DE3) Rosetta
Complete amino-acid sequence of the construct produced	MCELYSKRDTLGLRKKHIGPSCKVFFASDPIK- IVRAQRQYMFDENGEQYLDGINNVAHVGH- CHPGVVKAALKQMELLNTNSRFLHDNIVEYAKRLSATLPEK LSVCYFTNSGSEANDLALRLARQFRGHQDVITLD- HAYHGHLSLIEISPYK- FQKGKDVKKEFVHVAPTPDITYRGKYREDHADSASAYADEVK KIIEDAHNSGRKIAAFIAESMQSCGGQIIP- PAGYFQKVAEYVHGAGGVFI- ADEVQVGFGFRVGKHFWSFQMYGEDFVPDIVTMGKPMGNHGP VACVVTTKEIAEAFSSSGMEYFNTYGG- NPVSCAVGLAVLDIENEDLQGNAKRVGNYL- TELLKKQKAKHTLIGDIRGIGLFIGIDLVKDHLKRTPATAE AQHIIYKMKEKRVLLSADGPHRNVLKIKPPMCFTEED- AKFMVDQLDRILTVLEEAMGTKTESVTSENTPCKTKMLKEA HI- ELLRDSTTDSKENPSRKRNGMCTDTHSLLSKRLKT <u>LEHHHH</u> <u>HH</u>
Uniprot identifier	Q8TBG4

*: primer sequence (5'-3') used in this study, the restriction sites for NdeI (forward) and XhoI (reverse) are underlined. °: The 8-residues His 6 tag added to the C-terminus of the native sequence is underlined.

2.3.2. Protein production

Protein expression

hETNPPL was expressed in *E. coli* (DE3) Rosetta cells (Invitrogen) using the same transformation protocol described in the cloning section. Standard production was carried out in 1 l of LB broth supplemented with 50 µg/ml of ampicillin. Once the optical density of the solution ($\lambda=600$ nm) reached 0.6, protein expression was induced with 0.4 mM IPTG. After overnight expression at 25°C and 180 rpm, cells were harvested by centrifugation for 10 minutes at 10000 rpm and the resulting pellet, if not immediately processed, was stored at -80°C.

Protein extraction and purification

Cells were resuspended in Buffer A (50 mM Tris HCl pH 8.0, 50 mM NaCl, 0.5 mM β -mercapto-ethanol, 10 µM PLP, 5% Glycerol) to a final concentration of 0.2 g/ml. The suspension was supplemented with 140 µl/l of DNase (Sigma Aldrich), 14 ml/l of PMSF (Phenylmethylsulfonyl fluoride, a broad range protease inhibitor, Sigma Aldrich) and 0.5 mM β -mercapto-ethanol (β meEtOH, an oxidation scavenger that limits free radical damage, Sigma Aldrich).

Cells were then lysed by sonication using a Branson sonifier and centrifuged for 30 minutes at 15000 rpm and 4°C in order to separate the soluble from the insoluble fraction.

As an initial purification step, Ni-NTA affinity chromatography was performed to exploit the histidine tag that was added at the C-terminal end of the protein in the cloning step. The soluble fraction of the cell lysate was incubated with 1.5 ml of Ni-NTA-agarose beads suspension (QIAGEN) for 1 hour on an orbital shaker, then it was flowed through a ten milliliters disposable column and washed with 10 ml of Buffer B (50 mM Tris HCl pH 8.0, 50 mM NaCl, 0.5 mM β meEtOH, 10 µM PLP). In order to remove those proteins that remained bound to the column in a non-specific manner, two intermediate washes with 15 ml of buffer C (buffer B + 50 mM imidazole) and 3 ml of buffer D (buffer B + 80 mM imidazole) respectively were carried out. Finally, hETNPPL was eluted in 10 fractions using a total of 15 ml of buffer E (buffer B + 0.3 M imidazole).

The positive fractions were pooled, concentrated using an Amicon Ultra-4 centrifugal filter with a molecular weight cut-off of 10 KDa and loaded onto a HiPrep 26/60 Sephacryl S100HR size exclusion column (GE Healthcare) pre-equilibrated with Buffer B (50 mM Tris buffer pH 8.0, 50 mM NaCl). The purified protein was collected, concentrated to 10 mg/mL and stored at -80 °C. Unless otherwise stated all the purification steps described above were performed at RT. Protein

concentration was evaluated by Bradford assay⁴² using bovine serum albumin (Sigma) as the standard for building the calibration curve. Final sample purity was assessed by SDS-PAGE.

2.3.4. Enzyme activity

Enzymatic activity was determined following the amount of acetaldehyde generated as a product of the hETNPPL activity using a spectrophotometric assay coupled with alcohol dehydrogenase (ADH)^{29,32}. Since ADH catalyzes the reduction of acetaldehyde to ethanol in the presence of reduced nicotinamide adenine dinucleotide (NADH), the activity of hETNPPL was determined spectrophotometrically by measuring the absorbance of NADH at 340 nm. The assay was adapted to support a high throughput format in a 96 transparent polystyrene plates from Grainer Bio. The 300 μ L reaction mixture contained 25 mM Tris HCl pH 8.0, 10 mM phosphoethanolamine (PEA), 0.2 mM (NADH), and 15 U of yeast ADH (all the components were purchased from Sigma Aldrich). After 1 minute of pre-incubation, the reaction was started by adding the enzyme at a final concentration of 1 μ M, and the decrease in absorbance at 340 nm ($6220 \text{ M}^{-1} \text{ cm}^{-1}$) was monitored using a Spark10 M plate reader (Tecan). The reaction was carried out at 30 °C. One unit (U) is defined as the amount of enzyme required to produce 1 μ mol of acetaldehyde per minute, and specific activity was expressed as U mg^{-1} of enzyme.

2.3.5. UV-Visible spectroscopic measurements

Absorption spectra of hETNPPL (10 μ M final concentration) was collected at 25 °C in 50 mM Tris HCl pH 8.0, using a Cary 50 spectrophotometer (Varian), with 1 cm path-length quartz cuvettes.

2.3.6. Quaternary structure assessment

Dynamic light scattering (DLS) analysis was performed on a Zetasizer Nano ZS device (Malvern Instruments) at 25 °C with enzyme samples diluted to 10 μ M in 25 mM Tris HCl pH 8.0. Each sample was measured 10 times for 10 s in a 10 mm path length quartz cuvette. Data were collected and analyzed using Zeta Software (Malvern). All data were analyzed using the Zetasizer software v.7.12 (Malvern Instruments).

2.3.7. Protein crystallization

Crystallization experiments were set up using VDX or VDXm 24-well plates with siliconized glass circle slides, purchased from Hampton Research. The initial crystallization conditions were screened around those described for A1RDF1 crystals⁴³ by varying pH and ammonium sulfate concentration.

2.3.8. Data collection and processing

Diffraction data were collected at the Swiss Light Source of the Paul Scherrer Institute (Villigen, Switzerland) using the X06DA-PXIII beamline. Diffraction images were processed with iMosflm⁴⁴ and scaled using SCALA from the CCP4 program suite⁴⁵.

2.3.9. Structure solution and refinement

The structure of hETNPPL was determined by molecular replacement using PHASER⁴⁷ from the PHENIX program suite⁴⁸, and the structure of the phosphoethanolamine lyase from *Arthrobacter aureus* TC1 as the search model (PDB code: 5g4i;). Model building was done using AUTOBUILD⁴⁹, while refinement was carried out using REFMAC5⁵⁰ and PHENIX⁵¹ alternated with manual model building in COOT⁵². Water molecules were added both automatically using PHENIX and manually from visual inspection of the electron density map. The stereochemistry quality of the structure was assessed with the program PROCHECK⁵³. All the figures containing molecular graphics were generated using PyMOL.

2.3.10. Multiple sequence alignment and protein-ligand interactions representations

Sequence alignment was generated using Toffee Espresso⁵⁵ and visualized with the hETNPPL crystal structure using ESPript⁵⁶. Interaction diagrams were prepared using the program LIGPLOT⁵⁷.

2.4. Results and Discussion

2.4.1 Protein production

The purification protocol used for hETNPPL yielded more than 20 mg of pure protein starting from 1 L of bacterial culture (details are reported in **Table 2**).

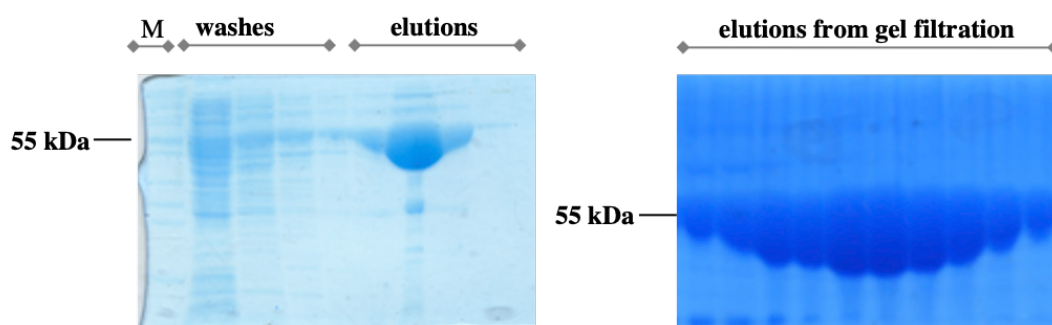


Figure 7. 10% SDS-PAGE assessing hETNPPL purity. Washes and elution fractions from NiNTA purification step (left). Elution fractions from size exclusion chromatography (right).

Table 2. Purification table for wild type hETNPPL

Operation step	Total protein (mg)	Total Activity (units)	Specific Activity (units/mg)	Fold Purification	Yield (%)
Crude Lysate	2350	282	0.12	1	100
NiNTA	167	20	0.12	1	7
S100	99.3	15	0.15	1.25	5
hETNPPL	98.1	75	0.77	6.4	26.5

2.4.1 Protein crystallization

The best crystals were grown by mixing 2 μ l of enzyme at a concentration of 10 mg/mL with an equal volume of a crystallization buffer containing 20% PEG 3350, 100 mM Bis Tris propane pH 7.0, 200 mM ammonium sulfate, and 3% MPD (**Table 3**). Crystals were soaked for 1 minute in the crystallization buffer supplemented with 25% glycerol and flash frozen in liquid nitrogen.

Table 3. Crystallization conditions of hETNPPL.

Method	Vapor diffusion
Temperature (K)	298
Protein concentration (mg·ml ⁻¹)	10-12
Buffer composition of protein solution	50 mM Tris-HCl pH 8.0, 50 mM NaCl
Reservoir solution composition	20% PEG 3350, 100 mM BTP pH 7.0, 200 mM (NH ₄) ₂ SO ₄ , 3% MPD
Volume and ratio of drop	4 μ l (1:1 ratio)
Volume of reservoir (μ l)	500

2.4.2. Data collection and processing

The crystals, which belong to space group H3 with the following unit cell parameters $a = 138.00$ Å $b = 138.00$ Å, $c = 121.87$ Å, $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$, diffracted to a resolution of 2.05 Å. The asymmetric unit contained two protein molecules, which form the functional catalytic dimer⁴⁶. Data collection and merging statistics are shown in **Table 4**.

Table 4. Data collection and processing. Statistics for the highest-resolution shell are shown in parentheses.

Diffraction source	X06DA-PXIII, SLS-PSI
Wavelength (Å)	1.00
Temperature (K)	100
Detector	Pilatus 6M
Crystal to detector distance (mm)	235
Rotation range per image (°)	0.1
Total Range per image (°)	180
Exposure time per image (s)	0.5
Space group	H 3
a, b, c (Å)	138.00, 138.00, 121.87
α , β , γ (°)	90, 90, 120
Resolution range (Å)	69.09 - 2.05 (2.16 - 2.05)
Total reflections	262785
No. of unique reflections	54318
CC 1/2	0.992 (0.771)
Completeness (%)	100.00 (100.00)
Multiplicity	4.9 (4.7)
Mean $I/\sigma(I)$	10.1 (3.6)
R_{meas}	0.151 (0.567)
R_{pim}	0.068 (0.258)
Overall B factor from Wilson plot-factor (Å ²)	16.39

2.4.3. Structure solution and refinement

The final structure of hETNPPL in the presence of PMP-bound to the active site in both the subunits was solved at 2.05 Å. Despite a certain lack of homogeneity in the quality of the electronic density map, residues from Ile31 to Thr442 (chain A) and Lys443 (chain B) could be placed without ambiguity, while no clear electron density was visible for residues Lys156-Val160, and Gly303-Phe307. Refinement statistics are given in **Table 5**. The coordinates and structure factors for the PMP-bound form of hEtnppl have been deposited in the Protein Data Bank⁵⁴ with the PDB code 6tor.

Table 5. Structure refinement. Statistics for the highest-resolution shell are shown in parentheses.

Resolution range (Å)	42.66 - 2.05 (2.123 - 2.05)
Completeness (%)	100.00 (100.00)
No. of reflections (work/test)	54318/5432
Final R_{cryst}	0.1668 (0.1991)
Final R_{free}	0.2122 (0.2611)
No. of non-H atoms	
Protein	6296
Ligands	68
Water	627
Total	6991
R.m.s deviations	
Bonds (Å)	0.009
Angles (°)	1.13
Average B factors (Å ²)	
Overall	20.30
Protein	19.50
Water	26.8
Ligand	32.5
Ramachandran plot	
Favoured regions (%)	99.5
Outliers (%)	0.5

2.4.4. Overall Structure

The crystal structure of hETNPPL was solved at 2.05 Å resolution with a final crystallographic R_{work} of 16.6% and a R_{free} of 21.2% (**Table 4**). The structure shows the presence of a protein homodimer in the asymmetric unit, as expected for a fold type I PLP-dependent enzyme (**Figure 8**).

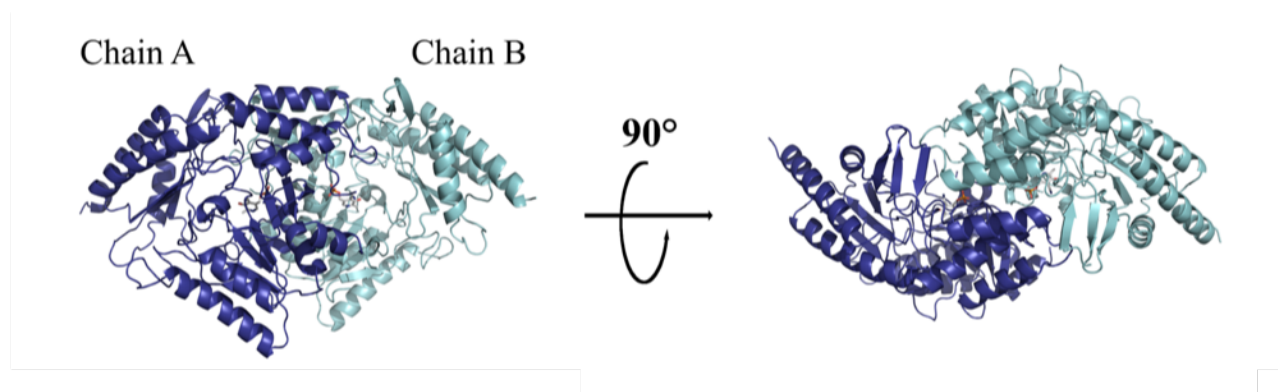


Figure 8. Overall structure of the hETNPPL PMP-bound form. Cartoon representation of the hETnppl dimer. Chains A, and B are colored in blue and cyan, respectively.

This is consistent with DLS analysis, which features a hydrodynamic diameter of about 10 nm for the dimer in solution (**Figure 9**).

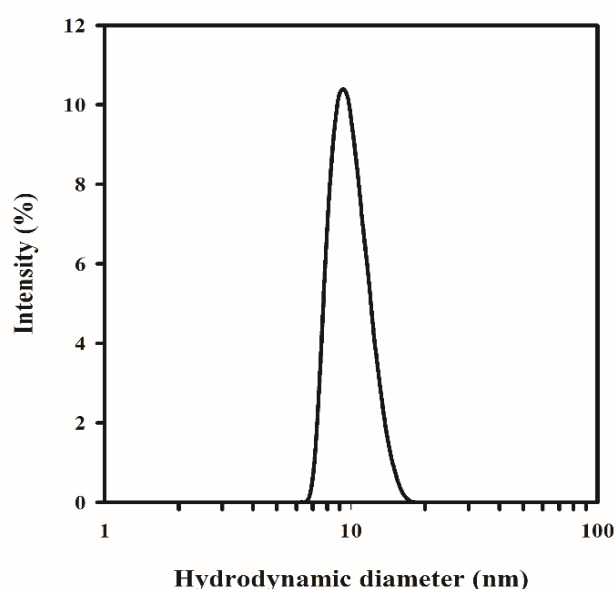


Figure 9. DLS analysis of hETNPPL. The enzyme in solution features a hydrodynamic diameter (Size = 10 nm) that is consistent with a homodimer.

Further superimposition of hETNPPL and A1RDF1 (PDB code: 5g4i; 39% sequence identity) reveals that, overall, the two phospho-lyases share a similar three-dimensional structure, with an r.m.s.d. of 1.54 Å for 654 Cα atoms (**Figure 10**).

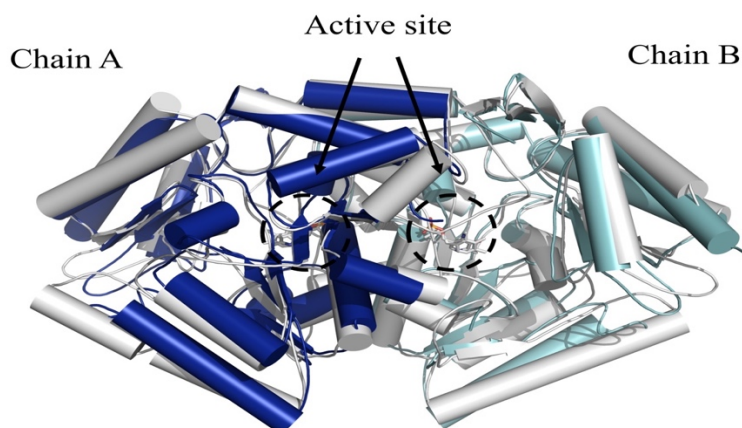


Figure 10. Comparison of the hETNPPL dimer (blue/cyan) (PDB code: 6tor) and A1RDF1 from *Arthrobacter aurescens* TC1 (white) (PDB code: 5g4i). The α -helices and the β -strands are depicted as cylinders and arrows, respectively. The bound cofactor molecule is shown as grey sticks and circled.

Likewise, the overall structural conformation of the two chains of hETNPPL (A and B) is very similar, with an r.m.s.d. of 0.019 Å. Each monomer consists of a small domain and a large domain (**Figure 11**). The small domain features an N-terminal and a C-terminal lobe. The N-terminal lobe (residues 1–85) is composed by an antiparallel β -sheet (β 1-3) and one α -helix (α 1) connected via a short loop to the large domain, or PLP-binding domain, (residues 86–301), which is the conserved location for the cofactor in the active site of fold type 1 PLP-dependent enzymes²⁰.

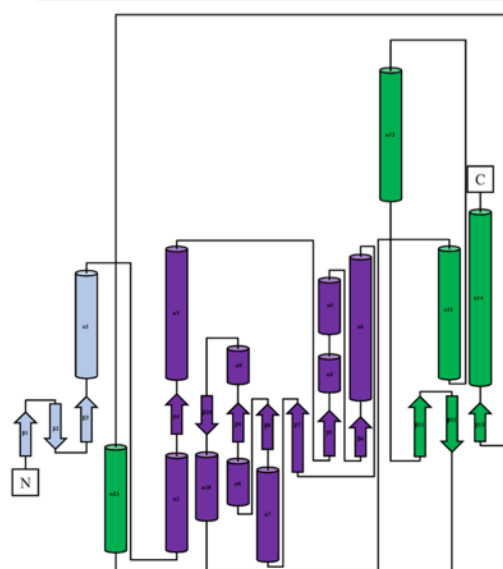
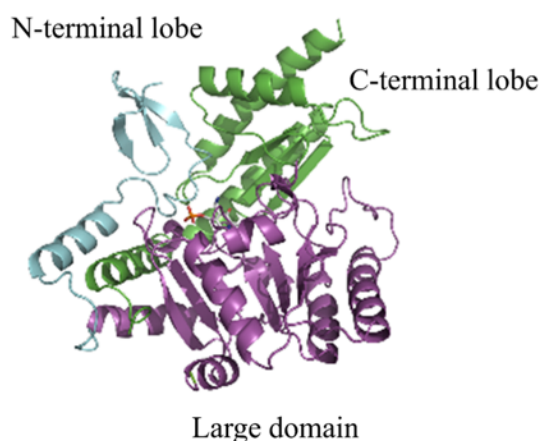


Figure 11. Structure of the chain A of the hETNPPL. The N-terminal lobe (residues 1–85), large domain (residues 86–301) and C-terminal lobe (residues 302–499) are colored in cyan, magenta and green, respectively. The bound cofactor molecule is represented as grey sticks (left). Topology diagram of hETNPPL. Helices are indicated as rectangles and β -strands by arrows (right).

Interestingly, no clear continuous density was visible in either subunit between the catalytic lysine, K278 and the cofactor, thus indicating that each subunit binds a pyridoxamine 5'-phosphate PMP rather than a PLP molecule (**Figure 12a**).

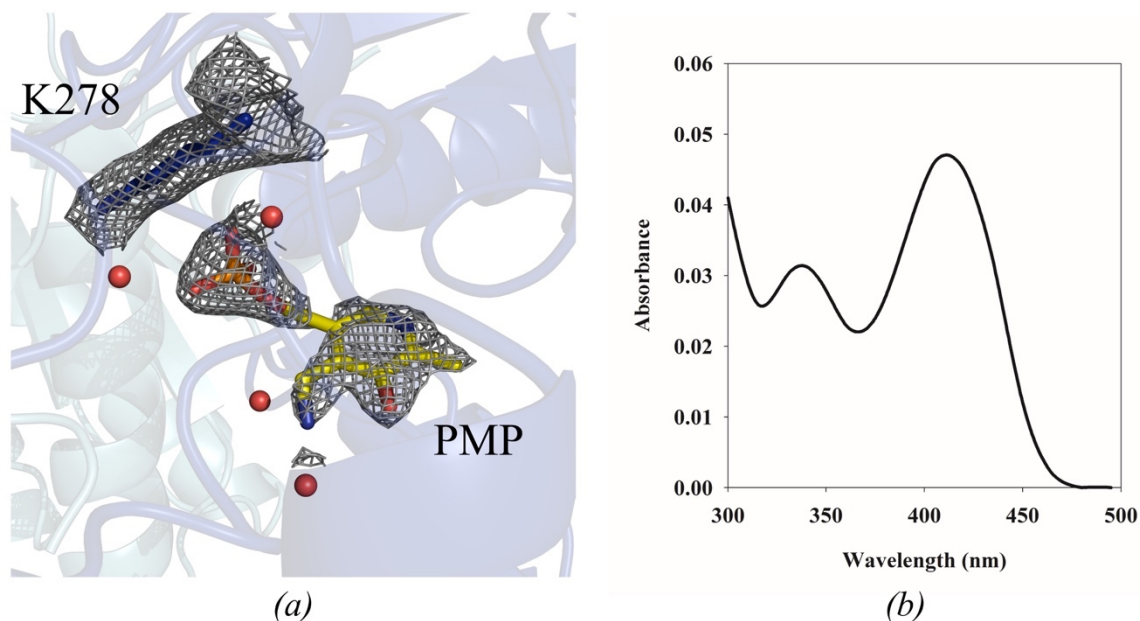


Figure 12. PMP cofactor in the active site. (a) PMP molecule and K278 residue are represented in yellow and blue sticks, respectively. The 2Fo-Fc electron density map is depicted as gray mesh at 1 σ . Water molecules are shown as red spheres. (b) The absorption spectrum of hETNPPL shows a prominent peak at 410nm, which is typical of bound PLP, and a second peak centered at approximately 330 nm region which is indicative of the presence of bound pyridoxamine-5'-phosphate (PMP).

PLP-binding domain consists of a twisted β -sheet (β 4–10), where the first six strands run parallel to each other while the β 10 strand runs antiparallel, and nine α -helices (α 2–8), adopting an α - β - α sandwich fold. On one side, the β -sheet is shielded from the solvent by five helices (α 2, α 6, α 7, α 8, α 10), while four helices (α 3, α 4, α 5, α 9) are involved in the intermolecular interface between the two subunits generating the active site. The C-terminal lobe (residues 302–499) contains a three-stranded antiparallel β -sheet (β 11–13), and four α -helices (α 11–14), all located on the exposed side of the β -sheet. The α 11 helix is inserted between the N-terminal lobe (α 1) and the large domain (α 2) and is involved in formation of the cofactor binding site. The buried area between the two chains forming the dimer is 6363.8 \AA^2 according to PISA⁵⁸. Head-to-head arrangement in the dimeric structure (**Figure 8a**) is stabilized by various polar interactions between the two subunits (**Figure 13a**). Based on

PISA calculations, the key residues forming interface contacts belong to the N-terminal lobe and to the large domain (**Figure 13b**).

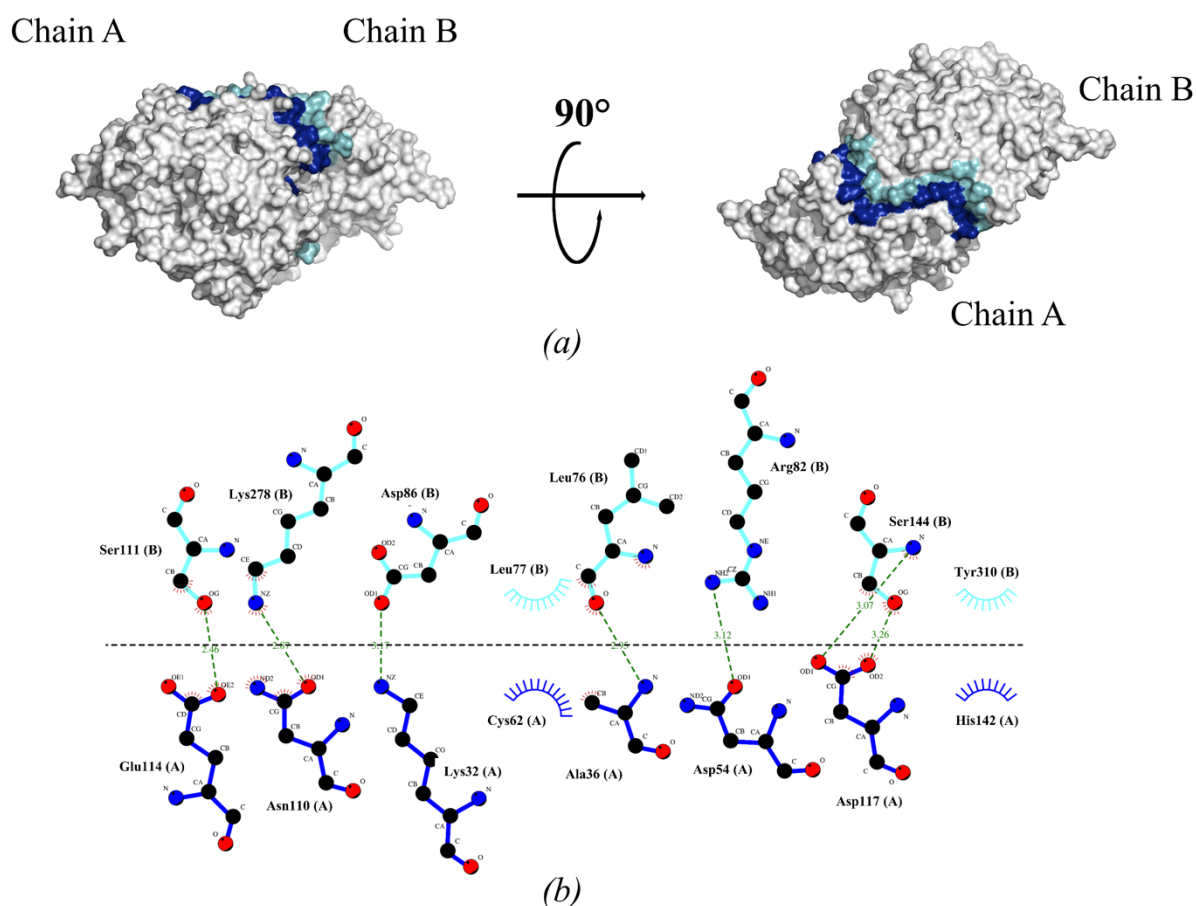


Figure 13. Dimer interface in the hETNPPL structure. (a) Surface representation of the dimer subunits. The critically important residues forming interface contacts are coloured in blue (chain A) and cyan (chain B). (b) LigPlot representation of the interaction between the two subunits. Hydrogen bonds are indicated by dashed lines, while those residues that form hydrophobic interactions are shown in spokes.

2.4.5. Active site

The active site of hETNPPL is located at the interface between the two subunits of the dimer, each subunit making crucial contributions to the active site. Most of the residues that line the catalytic pocket correspond to those that are found in the close structural homolog A1RDF1 from *Arthrobacter aurescens* TC1, strongly suggesting that these enzymes share a common mechanism (**Figure 14**).

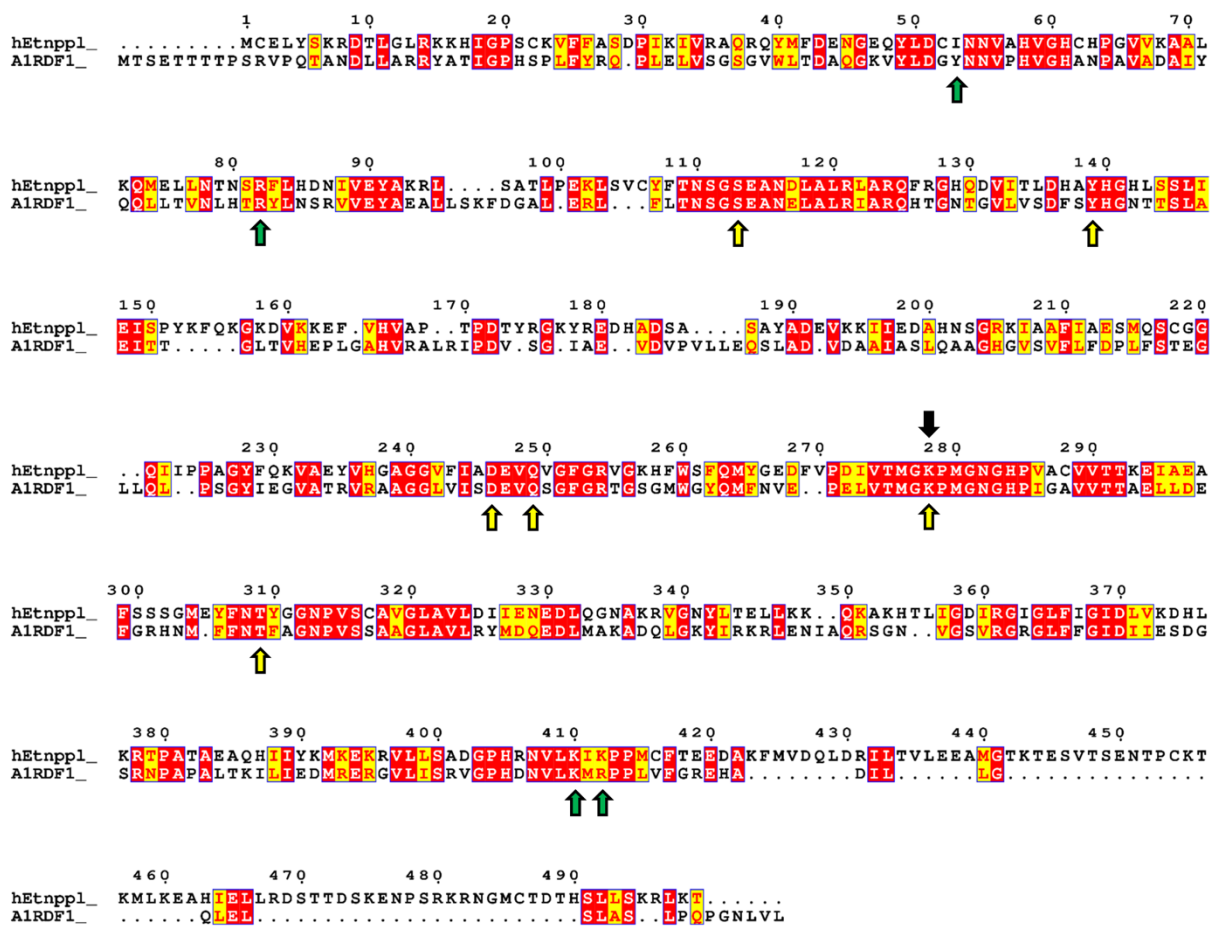


Figure 14. Alignment of hETNPPL with A1RDF, identical and similar residues are boxed in red and yellow background, respectively. Catalytic lysine residue is indicated by a black arrow. Yellow and green arrows indicate the residues involved in binding the cofactor and in the PEA phosphate recognition, respectively.

The cofactor binds essentially in the same position in hETNPPL and in A1RDF1 from *Arthrobacter aureescens* TC1 in complex with PLP (internal aldimine) and PLP-PEA (external aldimine). However, in the crystal structure of hETNPPL the cofactor does not covalently bind to the ϵ -amino group of the adjacent K278 residue, suggesting that the cofactor is PMP (**Figure 12a**). To confirm this conclusion, we did a UV-Visible spectrometric analysis of the enzyme in order to obtain its absorption spectrum. A previous study demonstrated that the hETNPPL PMP-bound form has an absorption peak at approximately 330 nm in the presence of a small, anionic ω -amine such 3-aminopropylphosphonate (3-APP). Herein, hETNPPL exhibited an absorption peak at 410 nm (**Figure 12b**), in agreement with the PLP-bound form, and an additional peak at 330 nm, characteristic for the presence PMP, confirming that the PMP-bound form was present in the sample and that may indeed provide the crystal structure of hETNPPL. A comparison of the hETNPPL-PMP complex with the reported structure in

complex with PLP (internal aldimine) for A1RDF1⁴³ confirms that almost all the interactions previously described are conserved (**Figure 15a**).

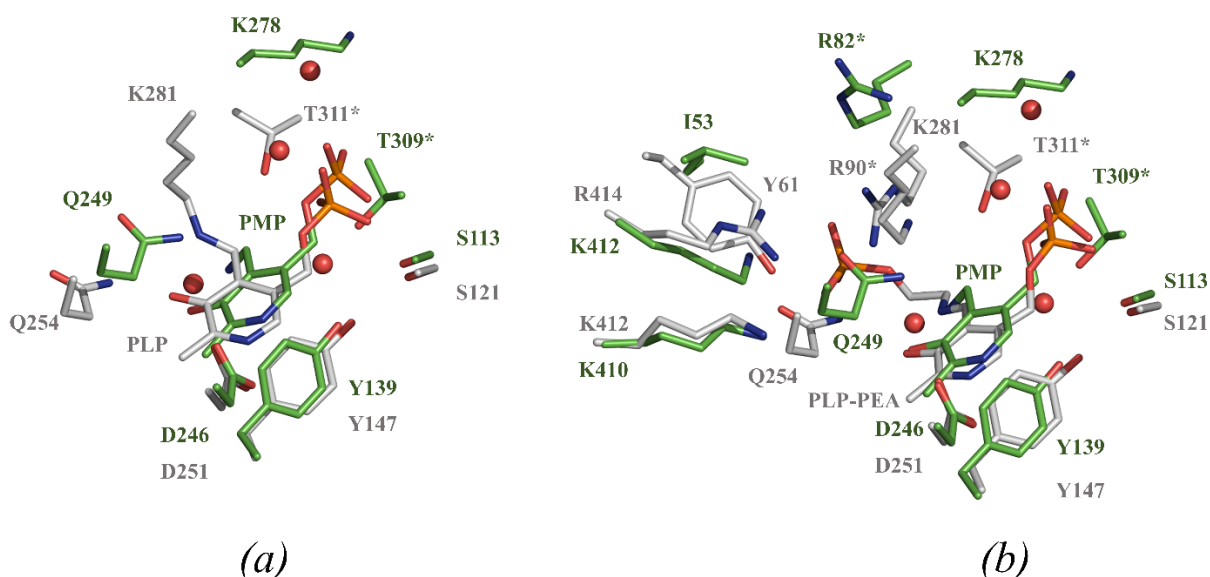


Figure 15. Comparison of hETNPPL and A1RDF1 active site. Critically important residues of hETNPPL (green) and A1RDF1 (white) active site, and molecule bound are presented in sticks, and are numbered in the corresponding color. Residue from the opposite monomer are marked (*). Water molecules in our structure are shown as red spheres. (a) Superposition of hETNPPL to PLP molecule bound in the active site of A1RDF1 (PDB code: 5g4i). (b) Superposition of hETNPPL to PLP-PEA complex bound to the active site of A1RDF1 (PDB code: 5g4i).

Specifically, the oxygen atoms in the phosphate group of PMP make stabilizing contacts with side-chains of one subunit: OP1 interacts with the side chain OH atom of S113, while OP2 and OP3 form water-mediated contacts to the NH1 atom and the ϵ -amino group of the catalytic lysine K278, respectively. The nitrogen atom on the PMP ring forms a hydrogen bond with the main chain amide of D246, while the hydroxyl group of the PMP forms a hydrogen bond with the side chain amine of Q249. Moreover, PMP forms hydrophobic interactions with nearby hydrophobic residues, especially Y139. Despite these structural similarities, a significant change can be observed for the loop Ser300-Asn313 that connects the large domain and the C-terminal lobe from the other subunit (B), which moves away from the active site. Notably, this loop contains a conserved Thr309 (Thr311 in A1RDF1) that contributes to the stabilization of the cofactor through the interaction of its side-chain hydroxy group with the PLP phosphate from the opposite subunit. Although the electron density of this loop could be only partially assigned, the side-chain hydroxy group of Thr309 in subunit B was clearly visible in the map. Owing to its ~ 5 Å distance from the PLP phosphate, it is very unlikely that this residue contributes to the binding of PLP. This finding does not exclude the role of Thr309

in binding the cofactor, but it may suggest some degree of flexibility in this loop, which in the external aldimine is free from any interaction with the cofactor molecule. Additionally, a comparison of the structure of the PLP-PEA complex in A1RDF1 with the hETNPPL structure reveals that all the critical residues in the active site are conserved (**Figure 15b**). Similarly to A1RDF1 (**20**), the hETnppl active site shows a cluster of positively charged residues (K410, K412 from one subunit, and R82 from the other) that are involved in the coordination of the negatively charged phosphate group of the PEA substrate, differing by only one residue in the hETNPPL active site (I53) compared to A1RDF1 (Y61). It is interesting to note that the presence of PMP in the active site (external aldimine) of hETNPPL alter the position of K278, which, compared with the K281 position in A1RDF1 in the internal and external aldimine, shifts away from the PMP cofactor. As described above, this suggests that in the external aldimine, the catalytic lysine may be free from any interaction.

2.5. Concluding remarks

In conclusion, I determined the crystal structure of a human O-phosphoethanolamine phospho-lyase at 2.05 Å resolution. The structure shows high structural homology with the crystal structure of the phospho-lyase from *Arthrobacter aurescens* *TC1*, which is to date the only phospho-lyase structure reported in the literature⁴³. While confirming, as previously described⁴³, the presence of a cluster of positively charged residues for the recognition of the phosphate moiety of the substrate, the structure of the active site of hETNPPL suggests a striking mobility of the catalytic K278 residue. Future work will focus on the fine details of the relevance of the conformational changes observed here to the catalytic cycle of the enzymes. This structure provides information that is crucial for understanding not only the observed specificity of hETNPPL, but also its role in phospholipid metabolism alterations that are associated with neuropsychiatric diseases.

Chapter 3. ETNPPL-based biosensor for PEA detection

3.1. Introduction

Prostate cancer (PCa) is one of the most common cancer in men worldwide, and the second leading cause of male cancer-related deaths in the United States, behind only lung cancer⁵⁹.

The etiology of prostate cancer has not yet been completely understood. Aside the likely contribution of other causes, age, race, family history, and genetics are well-established factors in the onset and the development of the disease. Usually prostate cancer grows slowly and is initially confined to the prostate gland, where it may not cause serious harm. However, certain types of prostate cancers are very aggressive and can spread quickly. Currently, there is wide range of possible treatments whose choice and effectiveness depend, in particular, on the stage of the disease. In general, the earlier the detection — especially when cancer is still confined to the prostate gland — the more likely it is for a man to undergo successful treatment.

To date, early detection screening of prostate cancer involves digital rectal examination (DRE) and prostate-specific antigen (PSA) test. The latter was introduced in the 1980s to replace the first test employed in PC, the prostatic acid phosphatase (PAP) test, which was used for nearly 50 years⁶⁰.

PSA is a protein encoded by the prostate-specific gene kallikrein 3 (*KLK3*) located on chromosome 19q13.4 belonging to the serine proteases family⁶¹. Mature PSA is obtained by two proteolytic cleavages of the inactive precursors pre-proenzyme PSA (pre-proPSA) and (proenzyme PSA) proPSA and then is secreted into semen⁶¹. Under normal conditions, only low levels of PSA can be detected in blood while raised levels of serum PSA are often found in PCa patients⁶².

Initially, since the evidences that PSA level increased with advancing clinical stage⁶³, the protein was used as a biomarker for monitoring the progression of patients and for detecting disease recurrence after curative treatments⁶³. Subsequently, in combination with DRE and other clinical findings, PSA test started to be employed as diagnostic tool for the early detection of PCa⁶².

However, as increased PSA levels are often also found in men with benign prostatic hyperplasia (BPH), prostatitis and prostate injury^{64–66}, its use as a biomarker generally produces high rates of false positive results. Clearly, the medical and economic implications of these confounding elements are considerable. After its introduction in the diagnostic field, PSA started to be used as a biomarker in screening tests also among asymptomatic men. As a result, more men were being diagnosed with prostate cancer, the majority of them having low-grade tumors — which are unlikely to cause significant symptoms or mortality⁶⁷— or benign conditions such as inflammation or hyperplasia. This “overdiagnosis” led to unnecessary treatments and to invasive biopsy procedures.

To address all these problems and increase the accuracy of diagnosis, the importance of searching for additional non-invasive prostate cancer biomarkers, possibly to be used in combination with the PSA measurement, has become evident.

In recent years, numerous candidate compounds have been proposed^{66,68–71}. Among them, particular attention has been paid to low-molecular weight metabolites such as free amino acids present in biological fluids^{72–74}. In a recent study on PCa biomarkers, 42 proteinogenic and non-proteinogenic free amino acid profiles in different biological fluids of both healthy and prostate cancer patients were analyzed by using a complex analytical-bioinformatic strategy⁷⁴. This study confirmed the high potential of amino acids as PCa biomarkers and also provided strong evidence that a good PCa biomarker candidate can be found within the ethanolamine metabolic pathway, a finding that confirmed data that were obtained by analyzing PCa tissues⁷⁵. In particular, this work indicated that both ethanolamine and phosphoethanolamine occur at lower concentration in biologic fluids of PCa patients compared to the healthy controls. Both of them are small molecules — present at varying concentrations in bodily fluids — involved in the Kennedy metabolic pathway as precursors and degradation products of one of the 4 major membrane phospholipids, the phosphatidylethanolamine (PE). Because of their role in the Kennedy pathway, ethanolamine-containing metabolites are thought to provide information about cellular proliferation, apoptosis, and the activity of key enzymes, hence correlate with cancer presence, aggressiveness, and therapeutic response⁷⁶.

Between the two metabolites reported above, the phosphoethanolamine appears more interesting because, in principle, it could serve as a biomarker in urine of PCa patients. Indeed, the difference in the level of PEA in urine — as determined in urine samples after normalization against the urinary creatinine concentration — was found to be statistically significant (**Table 6**).

Based on the assumption that PEA is a possible PCa biomarker in urine, the goal of this work is to develop a hETNPPL-based biosensing system that may increase the accuracy of diagnosis in PCa patients by supplementing or, in time, replacing PSA test. This work was done in collaboration with the group of Dr. Mario Caironi (Italian Institute of technology, Milano).

The biosensing system of choice, which we selected with the double aim of rapidly and quantitatively detecting PEA as well as validating it as a prostate cancer biomarker, belongs to the family of organic field-effect transistor (OFET) sensors. Interestingly, these types of biosensors usually feature high sensing performance levels and the low-cost of fabrication, which make them an ideal choice for the proposed applications.

Table 6. The quantified free amino acids in urine samples of two studied groups using the LC-ESI-MS/MS method. P values for the comparison of the variables between two groups were calculated according to Mann-Whitney U test, Student's t-test or Welch's F test. AUC values were obtained in univariate ROC curve analyses. Adapted from *Dereziński P. et al. (2017)*⁷⁴.

Amino acid	Concentration in urine samples [10 μ M amino acid / M creatinine]						p value	AUC
	The prostate cancer group (n = 49)			The control group (n = 40)				
	Average	Median	Range	Average	Median	Range		
1-methylhistidine	2516.9	1138.7	124.6 - 16491.3	4777.7	2731.4	138.0 - 24075.1	0.010	0.660
3-methylhistidine	1558.8	1525.9	364.0 - 3190.0	2147.8	1841.4	764.3 - 6658.4	0.004	0.679
alanine	2270.2	1573.1	266.5 - 7452.4	2501.1	1997.5	477.5 - 7384.6	0.523	0.540
arginine	161.5	100.4	21.6 - 952.3	278.2	238.9	69.6 - 878.5	< 0.001	0.831
argininosuccinic acid	111.5	85.1	16.7 - 451.0	108.2	87.6	38.3 - 288.6	0.695	0.524
asparagine	637.7	604.3	123.4 - 1620.4	1018.5	900.7	392.8 - 2804.3	< 0.001	0.773
aspartic acid	22.0	19.0	2.3 - 69.2	17.3	10.6	4.1 - 76.3	0.165	0.587
carnosine	118.5	65.3	3.8 - 1101.0	125.8	86.9	18.6 - 560.4	0.316	0.562
citrulline	42.3	29.2	5.7 - 167.3	62.0	47.0	11.3 - 422.3	0.008	0.664
cystathionine	97.6	96.3	4.9 - 269.4	279.9	162.2	37.3 - 1923.2	< 0.001	0.764
cystine	400.7	349.2	38.3 - 1816.3	584.8	445.1	189.0 - 2385.6	0.003	0.685
ethanolamine	2440.2	2597.8	847.0 - 4838.3	4103.6	3800.6	2021.2 - 7776.4	< 0.001	0.858
glutamic acid	89.0	83.1	12.5 - 239.8	145.3	114.1	38.3 - 623.5	0.002	0.692
glutamine	2781.4	2677.3	97.0 - 6676.7	4238.6	3889.5	1505.0 - 10003.7	< 0.001	0.736
glycine	8527.7	7533.5	1050.8 - 19514.0	10731.4	8648.7	2952.8 - 44161.0	0.148	0.590
histidine	3829.7	3060.1	119.2 - 9470.3	5773.9	5603.4	2142.9 - 12610.5	0.001	0.712
homocitrulline	120.4	122.9	19.0 - 338.8	275.7	196.5	67.7 - 1136.1	< 0.001	0.839
hydroxyproline	31.2	13.5	1.9 - 167.4	43.8	17.1	3.1 - 387.0	0.243	0.572
isoleucine	93.9	81.0	13.9 - 267.2	124.8	101.1	49.0 - 367.8	0.008	0.664
leucine	218.1	188.0	31.9 - 708.1	283.5	237.7	120.6 - 681.7	0.005	0.673
lysine	905.5	423.1	51.3 - 7879.6	1684.3	1047.9	262.8 - 10773.2	< 0.001	0.738
methionine	63.6	58.1	2.8 - 184.6	96.7	86.4	37.1 - 212.6	< 0.001	0.764
ornithine	124.2	104.9	12.5 - 427.9	186.9	146.4	49.1 - 689.9	0.002	0.689
phenylalanine	338.2	280.8	43.0 - 997.0	476.7	364.0	215.3 - 1089.0	0.001	0.708
phosphoethanolamine	113.5	97.9	7.2 - 299.7	308.3	263.6	41.2 - 1313.6	< 0.001	0.879
proline	91.1	79.3	11.8 - 288.7	83.4	60.8	21.9 - 287.4	0.188	0.581
sarcosine	12.7	7.3	0.8 - 101.5	19.2	11.5	2.5 - 80.6	0.056	0.619
serine	2843.2	2513.9	512.9 - 5925.1	3527.1	3108.8	1180.9 - 7353.6	0.018	0.646
taurine	6605.8	5836.1	560.5 - 16462.5	6227.6	3790.1	931.5 - 40322.8	0.032	0.633
threonine	957.9	691.1	47.1 - 3227.7	1016.9	920.3	377.0 - 2323.3	0.145	0.590
tryptophan	466.9	414.5	37.4 - 1159.9	665.2	547.3	272.7 - 1631.4	0.002	0.689
tyrosine	537.6	438.8	34.7 - 1372.7	820.2	642.5	350.1 - 2369.7	< 0.001	0.723
valine	342.2	301.9	44.4 - 1033.9	366.8	324.5	164.8 - 970.4	0.274	0.568
α -aminoadipic acid	212.6	151.7	4.2 - 584.3	312.3	253.8	111.8 - 1034.0	0.001	0.704
α -amino-n-butyric acid	110.1	105.9	2.8 - 237.2	114.3	106.6	23.2 - 253.4	0.689	0.525
β -alanine	288.5	162.0	13.7 - 3113.6	237.1	173.1	26.1 - 1385.6	0.898	0.508
β -aminoisobutyric acid	1639.6	866.2	209.9 - 9144.3	1716.0	871.6	115.0 - 9943.9	0.795	0.516
γ -amino-n-butyric acid	12.6	12.2	3.8 - 33.1	30.3	29.4	1.3 - 62.4	< 0.001	0.932
δ -hydroxylysine	28.4	21.1	5.7 - 161.7	102.2	47.7	13.1 - 898.7	< 0.001	0.796

Once the analyte, the bio-recognition molecule and a transducer that translates a binding event into a measurable signal were selected, the development of a first prototype of biosensor involved the following steps:

1. The design of the biosensor based on measurement range;
2. The selection of a suitable immobilization method;
3. Stability test and calibration curves;

All of these points, as well as the classification and the characteristics of the main types of biosensors, are described in the following paragraphs.

3.2. General principles of biosensors

A biosensor is an analytical device that incorporates a biological recognition element connected to a physical transducer to generate a signal that is proportional to the concentration of a specific target analyte⁷⁷. There are different ways to categorize biosensors; the main two are based on the type of signal transduction and biorecognition elements. By and large, a biosensor comprises: a) a biorecognition element that specifically binds to the analyte; b) an interface where a specific biological event takes place; c) a transducer that converts the biological response to a detectable signal; d) a computer software that converts the signal to a physical parameter that describes the process being investigated; e) an interface to the human operator (**Figure 16**).

In order to develop a successful biosensor, the first condition that must be met is a high selectivity of the biological recognition element for the target. Generally, this is accomplished by immobilizing on the sensor substrate a biological recognition element that has a specific binding affinity for the desired molecule⁷⁸. Typical biological elements include enzymes, antibodies, nucleic acids, whole cells, and receptors^{79,80}. They can interact in two different ways with the analyte: by forming a complex (affinity sensor) or by reacting with it to form a new biochemical reaction product (catalytic sensor)⁸¹.

Due to their high specificity for the substrate and the high turnover rates of biocatalysts, enzymes are currently the most commonly used biological recognition elements in diagnostic devices⁸².

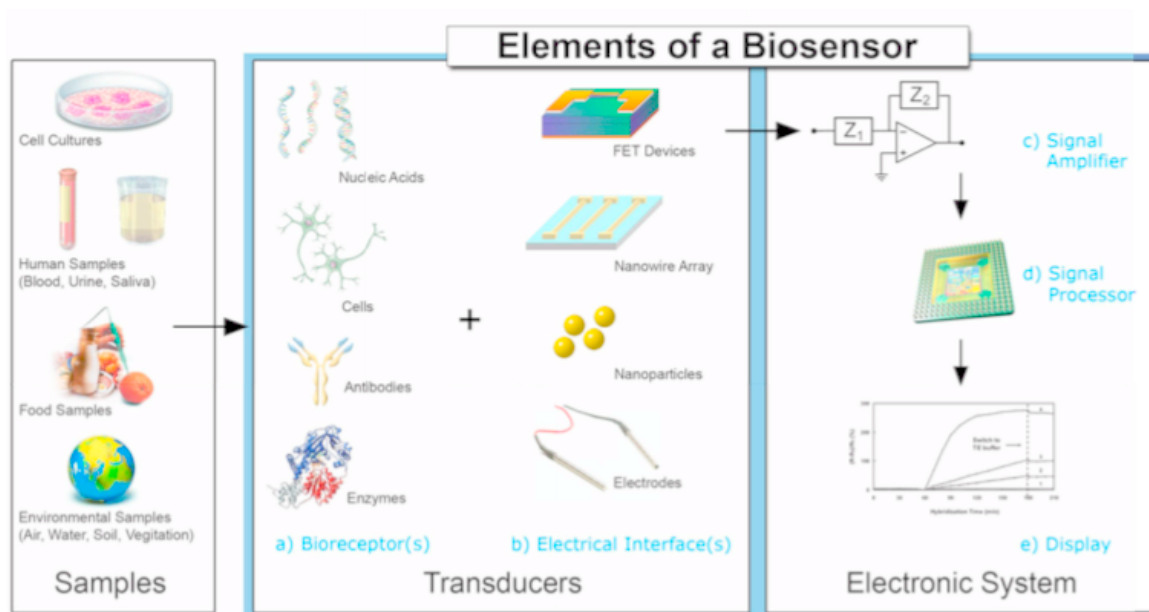


Figure 16. Schematic representation of a typical biosensor. Adapted from *Grieshaber et al. (2008)*⁸⁰.

In addition to the different types of possible biorecognition elements, biosensors can be classified based on the nature of the signal transduction being recorded. Depending on the physicochemical change resulting from the sensing event, different methods of transduction are possible⁸³. Based on these methods, biosensors can be divided in electrochemical, optical, bioluminescent, mass-based, piezoelectric and calorimetric.

3.2.1. Electrochemical biosensors

Among the wide variety of biosensors available, electrochemical biosensors are the most extensively used due to their low production costs and their ease of use⁸⁴. Moreover, electrochemical biosensors have been shown to work in complex matrices, such as serum, urine, blood and milk^{85,86} and they are compatible with modern miniaturization/microfabrication technologies⁷⁷. The basic principle for electrochemical biosensors is that during the bio-recognition process, a certain number of ions or electrons are produced or consumed, thus affecting measurable electrical properties of the solution, such as electric current or potential. The electrochemical signal produced is proportional to the amount of analyte that is present in solution. Typically, in this class of biosensors, the reaction under investigation would either generate a measurable current (*amperometric*) or measurable potential (*potentiometric*) or alter the conductive properties of a medium between electrodes (*conductometric*)^{78,87}.

Potentiometric biosensors measure the difference of potential (voltage) between a working electrode and a reference electrode in an electrochemical cell when zero or no significant current flows

between them^{78,88}. They work by detecting concentration changes for a specific ion. This variation can correlate with the concentration of a specific analyte that is present in a sample. The relationship between the concentration of the analyte and the potential is governed by the Nernst equation:

$$E = E^0 - \frac{RT}{nF} \ln Q$$

where E is the measured potential, E^0 is a constant potential contribution to the cell, R the universal gas constant, T the absolute temperature (K), n is the charge number of the electrode reaction, F is the Faraday constant and Q is the ratio of ion concentration at the anode to ion concentration at the cathode⁸⁹.

Conductimetric biosensors measure the ability of ions in solution to carry current between the electrodes when an alternating potential is applied between them. Because enzymatic reactions often involve changes in the ionic concentration, thus altering the electrical conductivity of the solution, these devices have been often associated with enzymes⁸⁸. However, the applicability of conductimetric devices in biosensing is limited by the requirement to measure small conductivity changes in media of high ionic strength⁸⁸.

Amperometric biosensors are based on the direct detection of current changes produced by the oxidation or reduction of electroactive species at the working electrode (i.e., gold, carbon, platinum, etc.). Usually, the current is measured at a constant potential and this is referred to as *amperometry*. If a current is measured during controlled variations of the potential, this is referred to as *voltammetry*. In both cases, changes that occur in the current are directly proportional to the concentration of the analyte in the sample⁷⁸. Amperometric biosensors are characterized by a high sensitivity and allow direct correlation of the measured signal with the concentration of the analyte. In addition, they can be easily miniaturized so that they can operate with small sample volumes, even in complex matrices^{90,91}.

3.2.2. Field-effect transistors (FETs)

In addition to the classical methods, other types of electrochemical detection techniques based on smart systems are currently being used in order to increase the applicability of biosensors in clinical diagnostics. In particular, there is growing interest in the *field-effect* (FET) technology, in which biosensors are interfaced with electronics to obtain very highly sensitive devices.

A FET is a type of transistor consisting of three metallic conducting electrodes: source (S), drain (D) and gate (G), a very thin insulating layer (dielectric) and a semiconductor into which charge carriers flow. When a voltage is applied to the gate, the charge carrier density in the semiconductor between

source and drain is modified, thus modulating the source-drain current. In organic FETs (OFETs), the gate is separated from the semiconductor by an insulator such as oxide, polymer, self-assembled monolayer (**Figure 17a**). Depending on the configuration and doping of the semiconducting material, the presence of a positive or negative potential at the gate electrode would either attract charge carriers (*e.g.* electrons) or repel charge carriers in the conduction channel. This would either fill or empty the depletion region of charge carriers and thus form or deform the effective electrical dimensions of the conducting channel. This controls the conductance between the source and drain electrodes.

3.2.2.1. Electrolyte-gated field-effect transistors (EGOFETs)

A very interesting variation of OFETs is represented by the electrolyte gated FETs (EGOFETs) in which the conductivity of the semiconducting channel is modulated by a solid or a liquid electrolyte placed between the semiconductor and the gate. As this class of devices displays a higher gate capacitance than other types of OFETs, it requires much smaller voltages (<1 V, *versus* >10 V or even higher)⁹².

In the most common configuration – top-gate, bottom-contact – the gate electrode is immersed in the electrolyte whereas the source and drain electrodes, isolated from the electrolyte, provide electrical contact to the channel (**Figure 17b**).

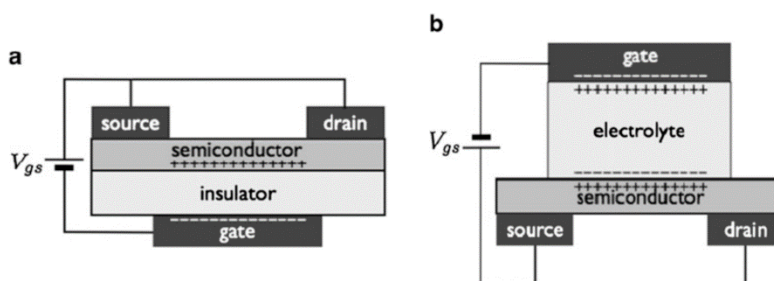


Figure 17. General scheme of an organic field-effect transistor (OFET) (a) and an electrolyte-gated organic field-effect transistor (EGOFET) (b). Adapted from *Kergoat et al. (2012)*⁹³.

In an EGOFET made with an appropriate semiconductor, upon positive polarization of the gate, the anions and cations of the electrolyte accumulate at the gate/electrolyte and at the electrolyte/OSC interface respectively, resulting in the formation of an electrical double layer (EDL) at both interfaces. The EDL is composed of a monolayer of ions, named Helmholtz layer (HL) and of a diffuse layer (DL) in which the ionic concentration starts to decrease going towards the bulk electrolyte. If the gate is negatively polarized, as shown in **Figure 18**, the excess of electrons on its surface will produce the accumulation of cations at the gate/electrolyte interface; at the same time, accumulation of anions at the electrolyte/semiconductor interface produces accumulations of holes in the topmost layer of the semiconductor^{94,95}, which causes the OSC to become conducting. It has been demonstrated that even

very low operating potentials can cause the formation of an EDL, sufficient to generate a locally high electrical field localized at the electrolyte/semiconductor interface, and therefore a high charge carrier density⁹³.

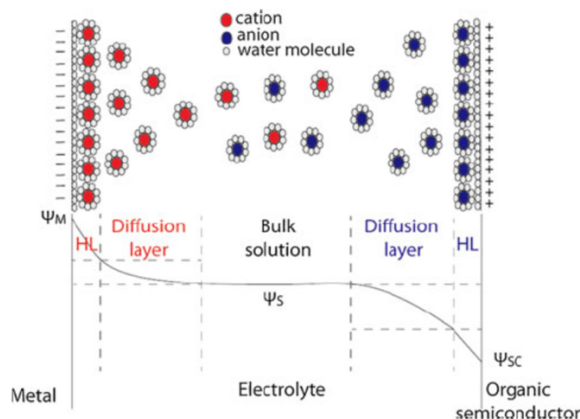


Figure 18. Illustration of the compact and diffuse layers corresponding to EGOFET. Adapted from *Kergoat et al. (2012)*⁹³.

3.3. Materials and methods

The biosensing system used in this work belongs to the EGOFET devices family. The device was fabricated and characterized by the group of Prof. Mario Caironi (Italian Institute of Technology, Milano).

3.3.2. Electrolyte gated FET (EGOFET) fabrication

The water gated field effect transistor was fabricated in a bottom contact, top gate architecture. 15nm thick gold contacts were fabricated by combination of lift-off photolithography and thermal evaporation on corning glass substrate. The channel length of the transistor was 5μm. Semiconducting single wall carbon nanotubes (SWCNT) ink was spin-coated on the contacts and annealed. Milli-Q water was used to gate the active area of the transistor. A 50 nm thick gold coating on polyethylene naphthalate (PEN) substrate (130 μm) was used as the floating gate of the transistor.

3.3.3. Instrumentation

Figure 19 shows a schematic overview of the device used as an enzymatic biosensor for the detection of PEA. A Semiconductor Parameter Analyzer (SPA) was used to obtain output and transfer characteristics of the transistor.

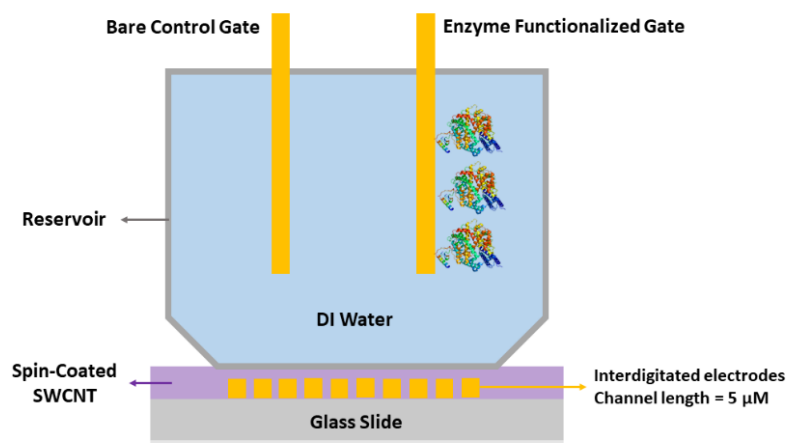


Figure 19. Schematic overview of the biosensor.

3.3.4. His-tagged ETNPPL immobilization on gold gates

In order to detect PEA, His-tagged ETNPPL produced according to the protocol described previously, was employed to functionalize the gold (Au) gate. The immobilization method used in the biosensor construction consisted in the direct adsorption of His-tagged ETNPPL on Au surfaces. We selected this method, as a first approach, for its simplicity and high potential and because among all the immobilization methods it is the one that best preserves the enzymatic activity⁹⁶.

The immobilization was performed manually, by drop casting 40 μl of a solution containing 200 μg of His-tagged ETNPPL on each substrate and leaving them to dry at room temperature for 20 hours, under fume hood.

3.3.4.1. Absorption evaluation for His-tagged hETNPPL on gold surface

Immobilization of His-tagged hETNPPL on a gold surface has been examined using a spectrophotometric assay. Briefly, the amount of enzyme immobilized on each surface was determined by following the amount of acetaldehyde produced as a product of the enzyme using an assay coupled with alcohol dehydrogenase (ADH;

Figure 20). Since ADH catalyzes the reduction of acetaldehyde to ethanol in the presence of reduced nicotinamide adenine dinucleotide (NADH), the activity of hETNPPL was determined spectrophotometrically by measuring the absorbance of NADH at 340 nm. The immobilized enzymatic activity on the Au surfaces was assayed by immersing the carrier discs individually in a 1 mL solution containing Tris buffer (25 mM, pH 7.0), ADH (15U), and NADH (0.5 mM) at 25 °C in a 24 well plate.

After two washing steps in the same solution for 5 minutes, enzyme activity was tested. Reactions were started by adding PEA at a final concentration of 5 mM, and the decrease in absorbance at 340nm ($\text{NADH } \epsilon_{340} = 6220 \text{ M}^{-1} \text{ cm}^{-1}$) was monitored in a Spark 10M (Tecan) (Figure 1B). As a positive control, the same reaction was also done without any surface, by adding the enzyme (40 μl) directly in solution.

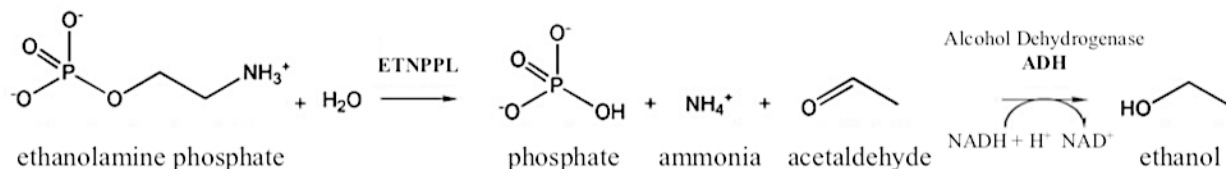


Figure 20. Overview of the reaction scheme for hETNPPL (Bioreceptor) with its substrate PEA (Analyte), and the used coupled with ADH.

3.3.4.2. Biosensing experiments for His-tagged ETNPPL on Au surface

Biosensing measurements were carried out by plunging the Au-electrode in PEA solution with known concentration for 30s. After this incubation time, the electrode was placed as the Fgate of the biosensor. The transfer curve of the Fgate and immediately after it, the transfer response of the control gate (Cgate) were recorded. A total of six concentrations were tested. For each concentration, three different experiments were carried out by using physically different gates. The final calibration curve was plotted based on the following equation at $V_g = -0.6\text{V}$:

$$\frac{\Delta I}{I_0} = \frac{I_{Fgate} - I_{Cgate}}{I_{Cgate}}$$

The selected gate voltage was the point of maximum transconductance for the Cgate. This gate was always fixed during the whole experiments.

As a negative control, the device was interrogated in the presence of Bovine Serum Albumin (BSA) instead of the enzyme on the gate electrode and in the presence of the target molecule, PEA, in the solution.

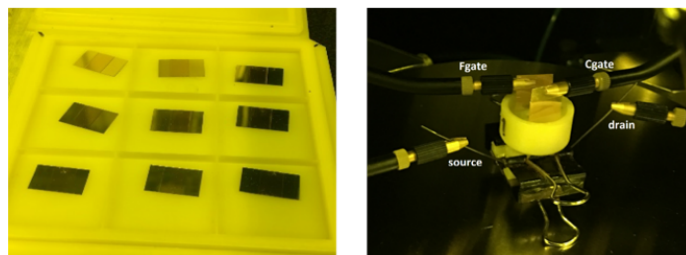


Figure 21. Gold (Au) gates (left). Biosensor setup (right)

3.4. Results and discussion

His-tagged ETNPPL immobilization on gold gates

Since the variation of absorbance at 340nm over time (minutes) can be converted to μmol of product generated per minute, the amount of enzyme deposited after two wash steps can be measured by extrapolating the slopes obtained in the spectrophotometric assay (**Figure 22**).

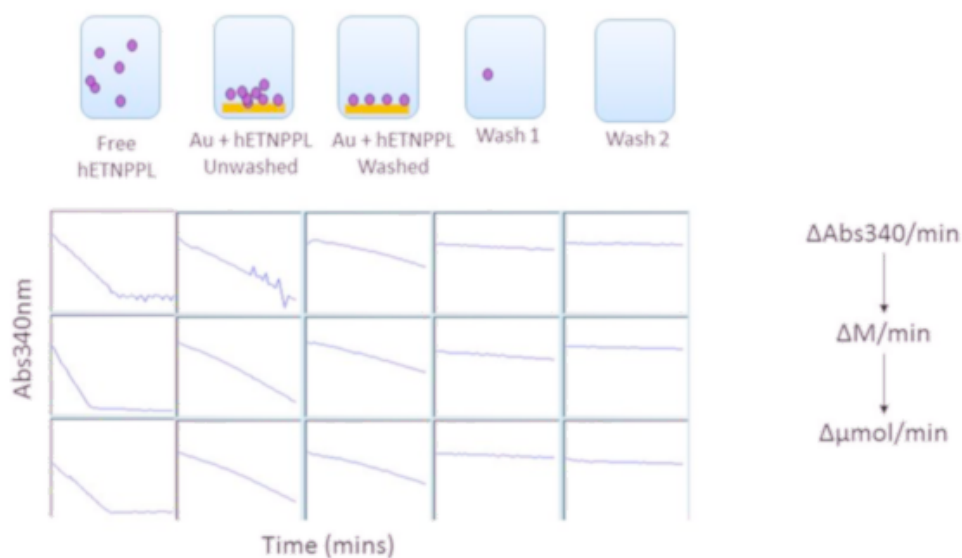


Figure 22. Measured activity for free enzyme (positive control), unwashed surface, washed surface, and the two washing solutions.

Absorption evaluation for His-tagged hETNPPL on gold surface

As expected, the free enzyme is around 4-times more active than the unwashed enzyme (free enzyme= 0.1 $\mu\text{mol}/\text{min}$ vs. unwashed=0.023 $\mu\text{mol}/\text{min}$). **Figure 23** shows that after two washing steps the washed surface maintained significant enzymatic activity, although a small amount of activity can be detected in both wash 1 and wash 2 (washed=0.0124 $\mu\text{mol}/\text{min}$, wash 1=0.0024 $\mu\text{mol}/\text{min}$, wash 2=0.00097 $\mu\text{mol}/\text{min}$).

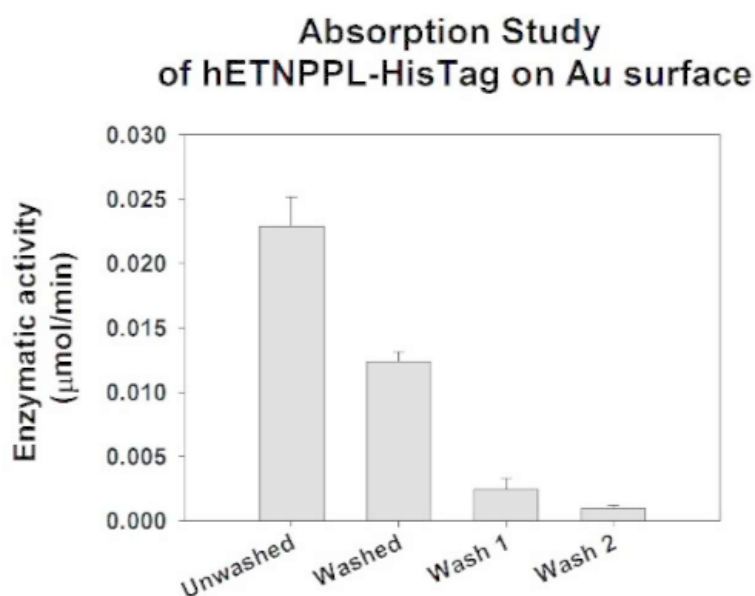


Figure 23. Enzymatic activity ($\mu\text{mol}/\text{min}$) calculated for the free enzyme, the immobilized enzyme and the two washing solutions.

Despite the reduced enzymatic activity of the immobilized enzyme with respect the free enzyme, the overall results show that His-tagged hETNPPL was successfully immobilized on gold surface. Comparing the enzymatic activity between the unwashed and washed surface, a first quantification of the real amount of the immobilized enzyme can be done by using the following equations:

$$\frac{0.023 \mu\text{mol}/\text{min} \text{ (Unwashed activity)}}{0.0124 \mu\text{mol}/\text{min} \text{ (Washed activity)}} = 1.85 \text{ (fraction of lost enzyme)}$$

$$\frac{200 \mu\text{g} \text{ (deposited enzyme)}}{1.85 \text{ (fraction of lost enzyme)}} = 108 \mu\text{g} \text{ immobilized enzyme}$$

Biosensing experiments for His-tagged ETNPPL on Au surface

Figure 24 shows the output characteristic of the device together with transfer curves obtained at different drain-source voltage (V_{DS}). Maximum transconductance was achieved at $V_{DS}=-0.5V$. Therefore, it was selected as the operation voltage for biosensing.

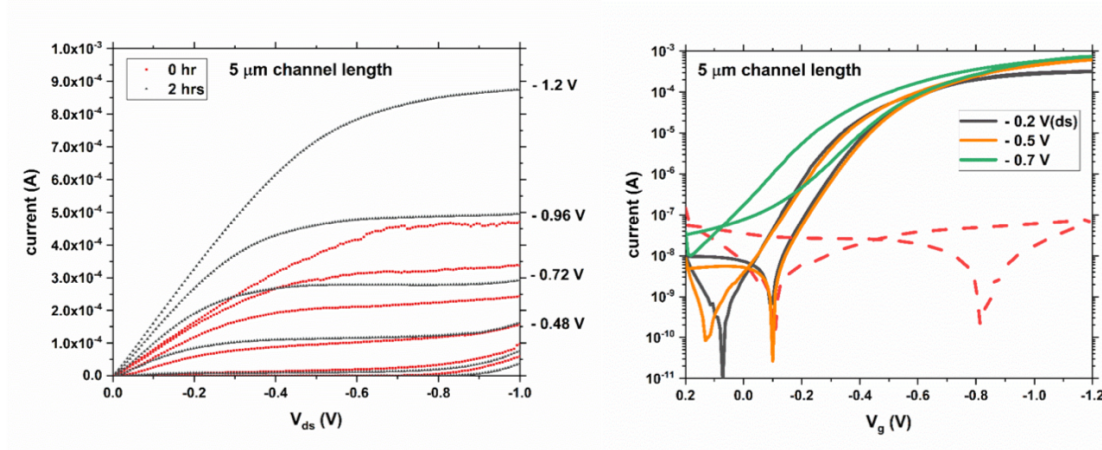


Figure 24. Output characteristics of the device together with transfer curves recorded at different source-drain voltages.

Overall, in the course of the experiment, a total of 36 transfer curves were recorded. As an example, **Figure 25** demonstrates all the six transfer curves (only forward scan) obtained for PEA concentration of 1 mM . Three curves were recorded from three physically different Fgates while the other three curves were obtained from the fixed control gate of the biosensor; each recorded immediately after one of the Fgates.

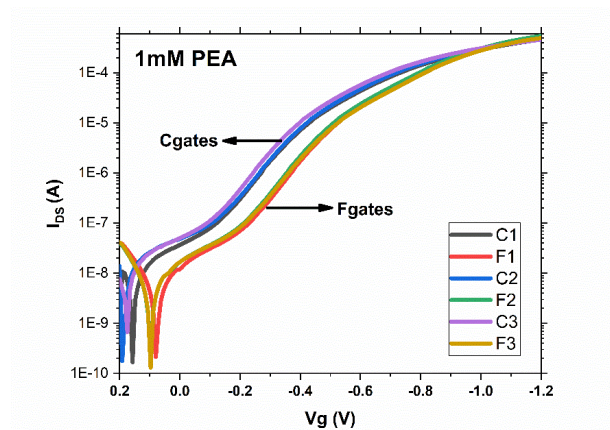


Figure 25. A total of six transfer curves recorded for 1 mM concentration of PEA.

Calibration curve

The calibration curve obtained from the biosensor is shown in **Figure 26**. The concentration range of PEA for the calibration standards was 0–1mM. A hyperbola function known as Michaelis-Menten model in enzyme kinetics was used to fit the experimental data.

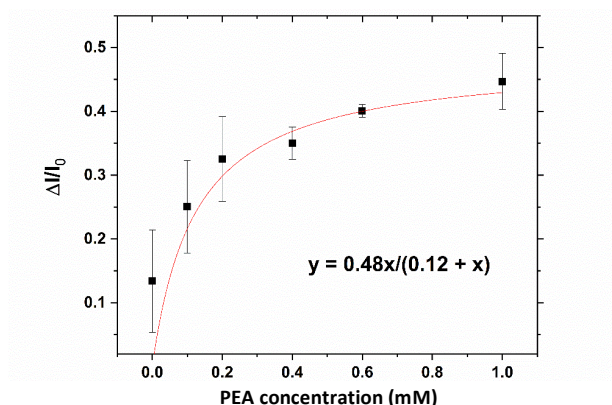


Figure 26. Calibration curve of the biosensor.

As shown, the response of the hETNPPL-based biosensor revealed a typical immunoassay behavior, where saturation and non-sensitive zones were clearly identified. The reduced chi-square of the fit was 0.97.

Regarding the relatively large standard deviations in the final results, a possible cause could be related to the Fgates, which were manually cut to the size of 1×1.5 cm² and whose active area (0.5×1 cm²) was confined by Kapton tape. This manual process could cause differences in the gate electroactive area, which can affect the final results. Moreover, surface functionalization by drop casting is another possible source of variation in the size of the protein-covered area of the gates.

Finally, another reason for the observed variability could be that the PEA is processed too fast by the enzyme, and the products are released by the active site during the analysis thus reducing the precision of the measurements.

Negative control

As shown in **Figure 27**, when BSA is deposited on a gold surface instead of hETNPPL, current change is almost zero for all the PEA concentrations. This is expected, as BSA usually serves as a blocking moiety for biosensors. The overall result shows that in the absence of the enzyme, neither the gold electrode work function nor the capacitance of the Helmholtz double layer change.

Moreover, these results suggest that BSA can be used to block PEA adsorption on the surface of the electrode in our future experiments.

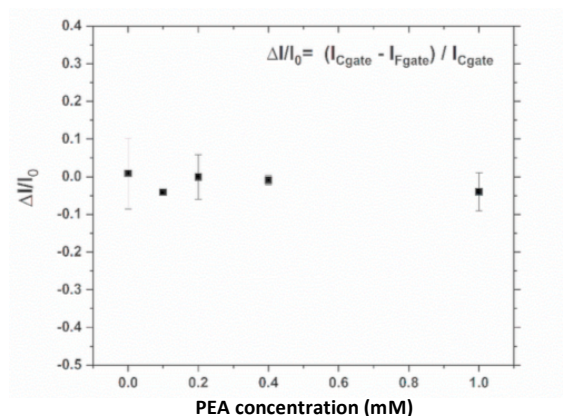


Figure 27. Calibration curve in the presence of BSA.

3.5. Concluding remarks

Since the discovery, in 1930, of prostatic acid phosphatase (PAP) as a clinical marker for prostate cancer (PCa) progression, the clinical management of this disease has extensively employed biomarkers. Prostate-specific antigen (PSA) is the blood biomarker actually in use in PCa screening. However, owing to its inherent limitations, intensive efforts are being taken to find a more valid alternative.

In a recent study on PCa biomarkers, the phosphoethanolamine (PEA) emerged as one of the most promising candidates. However, a confirmation of this finding would require further investigation.

Here, I reported the first prototype version of an ETNPPL-based EGOFET biosensor for the specific detection of PEA in solution. This system, designed with the double aim to detect the PEA and to validate it as a possible PCa biomarker, would be ideal for diagnostic applications because of the safe and easy means of measurement and because it has a high sensitivity being based on an enzyme that is specific for the selected biomarker.

Overall, the preliminary results shown in this work demonstrate that the measurement setup is capable of detecting the changes of PEA concentration of the solution. Obviously, despite the feasibility of PEA detection against O-phosphoethanolamine phospho-lyase, further optimization to increase the performance and the stability of the system are needed. Moreover, a negative control test based on interfering proteins found in urine should be carried on.

Future experiments will aim to minimize the standard deviations observed in the measurements. The first step will consist in mutating the crucial residues present in the active site of the hETNPPL so that the substrate binds to the enzyme but is no longer processed or at least, that the reaction is slowed down. This should increase the stability of the measurement. Moreover, in the next experiments, simulated urine and automated deposition techniques, such as enzyme printing, will be tested. Finally, BSA will be used in the next trials as blocking agent.

After system optimization using the experiments listed above, the range of PEA concentration will be lowered from mM to μM and nM. It will be helpful to define the lower detection limit and the overall working range of the sensor.

Chapter 4. TpVar1-mediated delivery of human O-Phosphoethanolamine phospho-lyase in hepatic carcinoma cell line

4.1. Introduction

The Kennedy pathway is one of the predominant mechanisms that mammalian cells employ to synthesize phosphatidylethanolamine (PE), which is one of the four major components of the phospholipid membrane. In this pathway, an important precursor of PE is phosphoethanolamine (PEA), which is formed by the action of ethanolamine kinase on free ethanolamine. Because of PEA involvement in the PE synthesis, it is believed that an aberrant expression of hETNPPL – which is able to recognize and process PEA – may affect the fatty acid and lipid metabolic processes. To support this hypothesis, a down-regulation of the enzyme has been recently correlated to increased lipogenesis in Hepato-Carcinoma Cells (HCC) tissues⁹.

Increased lipogenesis is one of the hallmarks of metabolic reprogramming in cancer cells⁹⁷. Indeed, owing to their uncontrolled rapid growth and division, cancer cells need to provide a continuous supply of both structural and energy-providing lipids⁹⁸. This need is what prompts a wide variety of tumors to activate or upregulate—as in the case of liver tumor tissue— *de novo* lipogenesis (DNL)⁹⁹.

DNL in health and disease

De novo lipogenesis (DNL) is a highly regulated metabolic pathway that, in normal conditions, converts excess of carbohydrate into fatty acids (FAs), which are then esterified to storage triacylglycerols (TGs). TGs can be then mobilized for β -oxidation, membrane biogenesis, protein modification, synthesis of lipoproteins, steroids and other lipid mediators, and so on¹⁰⁰.

TGs are commonly stored into special compartment called lipid droplets (LDs), which are cytoplasmatic organelles present in most eukaryotic cell types. These reservoirs consist of core lipid components surrounded by an amphipathic lipid layer with integral proteins¹⁰¹. Among these proteins, perilipins play critical roles in regulating the storage and utilization of lipids allowing for the droplets to be enzymatically processed¹⁰².

With the exception of adipocytes, LDs in cancer cells are more abundant than in the other cell types. Despite the same high content of LDs, some differences are present between adipocytes and cancer cells. The main difference arises from the different use that the two cells make of LDs. Indeed, adipocytes use these LDs as emergency reserves, while cancer cells use these reservoirs continuously, in a more dynamic way. Due to these differences, LDs in adipose cells are large and few in number, while in cancer cells they are both smaller and more abundant. Because of their high surface to volume ratio, the latter can be degraded and formed more rapidly.

In addition to this, the characteristic biochemical properties of lipids produced via DNL make cancer cells more resistant to oxidative stress-induced cell death¹⁰³. Indeed, DNL enriches the cancer cell membranes with saturated and/or mono-unsaturated FAs¹⁰³, which are more stable to lipid peroxidation than the polyunsaturated FAs that are normally produced. Moreover, saturated lipids pack more densely, thus hampering the correct uptake of drugs.

Because of the correlation between a down regulation of hETNPPL and an increase in DNL in HCC⁹, the investigation of the possible effects of enzyme re-administration on LDs to contrast DNL within this cell type is believed to be of potential therapeutic interest. Indeed, this approach may be relevant for tumour treatment, since a reduction of DNL could decrease the resistance of HCC to cancer therapy.

As a part of my PhD work on the hETNPPL enzyme, in collaboration with the group of Dr. Maria Rosa Antognazza (Italian Institute of Technology, Milano) I set out to investigate if the exogenous administration of hETNPPL in HCC would affect the lipid reservoirs in HCC cells. The work can be divided in two part.

In the first part, I constructed and purified a mutant of the hETNPPL that would likely be capable of crossing the plasma membrane, thus allowing enzyme delivery inside the cell. This mutant exploits the shuttling properties of a specific peptide belonging to the class of cell penetrating peptides (CPPs), which are protein transduction domains, typically comprising 5–30 amino acids that can pass through cell membranes delivering protein into the cells. The CPP-peptide I selected derives from a 9-residue long domain rich in basic amino acids of the HIV-1 Tat protein. During the last years, this peptide (**RKKRRQRRR**) has been shown to possess the ability to cross biological membranes together with cargo proteins of different nature in a process termed protein transduction^{104,105}. In this work, since the evidence that positively charged residues favor the uptake of CPPs¹⁰⁶, I added an additional arginine residue to the traditional Tat-peptide to obtain the TpVar1-hETNPPL mutant. In order to exclude any possible interference of the CPP with the enzyme activity, I performed all the biochemical characterization of the TpVar1-hETNPPL and then I compared these data with those of wild type hETNPPL.

The second part of the work, which includes the evaluation of the changes in lipid content of HCC before and after TpVar1-hETNPPL uptake, is still ongoing. Here, upon enzyme uptake, the LDs will be visually assessed using confocal microscopy.

4.2. Material and Methods

4.2.1. Gene cloning

To construct the pET23a-TpVar1- hETNPPL plasmid, I performed four subsequent PCR amplification runs using the pET23a- hETNPPL plasmid as the initial template in order to add at the 5' of hETNPPL coding sequence the entire TpVar1 cell penetrating peptide sequence. I then sub-cloned the final PCR product into the pET23a expression vector with the same restriction enzymes as described above (**Table 7**). The resulting construct was finally validated by DNA sequencing.

Table 7. TpVar1-hETNPPL production.

Source organism	<i>H. sapiens</i>
DNA source	pET23a-hETNPPL
Forward-1*	cgctcgcggt tgcgagctgtacagtaag
Forward-2*	cggcaacgccg tcg cggt tgcgagc
Forward-3*	cgcaagaaacg tcg cgcaacgccg tcgc
Forward-4*	gatgatcatatg cgccgcaagaaacg tcg cgcaac
Reverse*	tttctcgagtggtcttgagcctcttact
Expression vector	pET23a
Expression host	<i>E. coli</i> (DE3) Rosetta
Complete amino-acid sequence of the construct produced ^{o a}	MRKKRRQRRR GCELYSKRDTLGLRKKHIGPSCKVF- FASDPIKIVRAQRQYMFDENGEQYLD- CINNVAHVGHCHPGVVKAALKQMELLNNTNSRFLHDNIVEYA KRLSATLPEKLSVCYFTNSGSEANDLALRLARQFR- GHQDVITLDHAYHGHLSSLIEI- SPYKFQKGKDVKKEFVHVAPTPDITYRGKYREDHADSASAYA DEVKKI IEDAHNSGRKIAAFIAESMQSCGGQIIP- PAGYFQKVAEYVHGAGGVFIA- DEVQVGFGFRVGKHFWSFQMYGEDFVPDIVTMGKPMGNHPV ACVVTTEKIAEAFSSSGMEYFNTYGGNPVSCAVGLAVL- DIIENEDLQGNARVGNYL- TELLKKQKAKHTLIGDIRGIGLFIGIDLVDHLKTRTPATAE AQHIIYKMKEKRVLLSADGPHRNVLKIKPPMCFTEEDAK- FMVDQLDRILTVLEEAM- GTKTESVTSENTPCKTKMLKEAHIELLRDSTTDSKENPSRK RNGMCTDTHSLLSKRLKT LEHHHHHH

*: primer sequence (5'-3') used in this study, the restriction sites for NdeI (forward) and XhoI (reverse) are underlined. ^o: The 8-residues His6 tag added to the C-terminus of the protein sequence is underlined. ^a: sequence codifying for Tpvvar1 is in bold.

4.2.2. Protein production

Protein expression

TpVar1-hETNPPL was expressed in *E. coli* (DE3) Rosetta cells. Bacterial cells were grown in the Lauria Bertani (LB) medium with 50 µg/ml ampicillin at 37 °C. At an optical density (OD₆₀₀) of 0.6, expression of the fusion proteins was induced with IPTG (isopropyl β-D-1-thiogalactoside; 0.4 mM final concentration). After overnight expression at 25°C and 180 rpm, cells were harvested by centrifugation for 10 minutes at 10000 rpm and the resulting pellet, if not immediately processed, was stored at -80°C.

Protein extraction and purification

Cells were re-suspended in Buffer A (50 mM sodium phosphate pH 7.0, 300 mM NaCl, 5 mM 2-mercaptoethanol, 40 µM PLP, 5 % Glycerol) supplemented with protease inhibitor cocktail, and DNase I, and disrupted by sonication on ice. The lysate was clarified by centrifugation, and the soluble fraction was loaded onto a Ni-NTA column equilibrated with Buffer A. Column was first thoroughly washed using ten column volumes (CV) of Buffer A supplemented with 10 mM imidazole, then a second washing step was done with 2 CV of Buffer A supplemented with 50 mM imidazole, and finally the target enzyme was eluted with same buffer supplemented with 300 mM imidazole. The positive fractions were pooled, concentrated to 1 mg/mL using an Amicon Ultra-4 centrifugal filter with a molecular weight cut-off of 10 KDa and dialyzed at 4°C against Buffer B (50 mM sodium phosphate pH 7.0, 300 mM NaCl, 1 mM DTT, 4 µM PLP, 5 % Glycerol) to remove imidazole. Unless otherwise stated, all the purification steps described above were performed at RT. Protein concentration was quantified by Bradford assay using bovine serum albumin (Sigma) as the standard. Sample purity was assessed by 10% SDS-PAGE.

4.2.3. Activity assays and kinetic analysis

Enzymatic activity was determined following the amount of acetaldehyde produced as a product of the hETNPPL activity using a spectrophotometric assay coupled with alcohol dehydrogenase (ADH), as previously described (Chapter 2). In this work, the assay was adapted to support high throughput format in a 96 transparent polystyrene plates from Grainer Bio. The 300 µL reaction mixture contained 25 mM Tris HCl pH 8.0, 10 mM PEA, 0.2 mM NADH, and 15 U of yeast ADH (all the components were purchased from Sigma Aldrich). After 1 minute of pre-incubation, the reaction was started by adding the enzyme at a final concentration of 1 µM, and the decrease in absorbance at 340 nm (6220 M⁻¹ cm⁻¹) was monitored in a Spark10 M (Tecan). Reactions were carried out at 30 °C. One unit (U) is defined as the amount of enzyme required to produce 1 µmol of acetaldehyde per minute,

and specific activity was expressed as U mg⁻¹ of enzyme. Data were analyzed by nonlinear least-squares fitting to the appropriate kinetic equation using SIGMA PLOT (Systat Software Inc., Richmond, CA, USA). The apparent steady-state parameters for hETNPPL over the PEA were determined by varying the substrate concentration from 0.1 to 20 mM. Data were fitted by non-linear least-square analysis using a Michaelis-Menten equation for hyperbolic substrate kinetics:

$$v = V_{\max} S / K_m S$$

in which v , V_{\max} , S , and K_m represent the steady state reaction rate, maximum reaction rate, substrate concentration, and Michaelis-Menten constant for the substrate, respectively. The turnover number k_{cat} was calculated from the maximum reaction rate, and the catalytic efficiency was calculated as the ratio of k_{cat} (s⁻¹) to K_m (M):

$$k_{\text{cat}} \text{ (s}^{-1}\text{)} = V_{\max} \text{ (}\mu\text{M s}^{-1}\text{)} / (\text{enzyme concentration, } \mu\text{M})$$

To evaluate the enzyme activity of hETNPPL and TpVar1-hETNPPL mutant at different pH values, the assay described above was performed over a pH range of 5.0-10.0 using the following buffer (25mM): Sodium Citrate (pH 5.0), MES-NaOH (pH 6.0), Tris-HCl (pH 7.0-10.0).

4.2.4. Spectroscopic Measurements

Absorption spectra of hETNPPL and of the TpVar1-hETNPPL mutant (10 μ M final concentration) were collected at 25 °C in 50 mM Tris HCl pH 8.0, in the absence and in the presence of 10 mM PEA, or 100 mM phosphate, using a Cary 50 spectrophotometer (Varian), with 1 cm path-length quartz cuvettes.

4.2.5. Quaternary structure assessment

Dynamic light scattering (DLS) measurements were done to assess the oligomeric state of the purified enzymes using a Zetasizer Nano ZS device (Malvern Instruments) at 25 °C with enzyme samples diluted to 10 μ M in 25 mM Tris HCl pH 8.0. Each sample was measured 10 times for 10 s in a 10 mm path length quartz cuvette.

4.2.6. Circular dichroism spectroscopy

To assess the secondary structure content of the enzymes, circular dichroism (CD) spectra were recorded with a Jasco J715 spectropolarimeter equipped with a Peltier thermostatic cell set at 20°C, by using 2 mm path-length quartz cuvettes, in 20 mM NaPO₄ buffer pH 7.0. Enzyme concentration was 10 μ M. Scans were recorded in the region between 300 nm and 550 nm, at a speed of 50 nm/min with a bandwidth of 2 nm and a sensitivity of 20 millidegrees.

4.3. Results and discussion

Protein production

The purification protocol used for TpVar1- hETNPPL mutant was adapted through several optimization cycles because the addition of the TpVar1 N-terminal tail to the protein caused aggregation problems. The final protocol yielded more than 2 mg of pure protein starting from 1 L of bacterial culture. As expected, the purified TpVar1-hETNPPL protein migrates with a slightly higher apparent molecular mass than hETNPPL on SDS-PAGE (**Figure 28**).

Table 8. Purification table for TpVar1-hETNPPL.

Operation step	Total protein (mg)	Total activity (units)	Specific activity (units/mg)	Fold purification	Yield (%)
Crude Lysate	440	4,4	0,01	1	100
NiNTA	40	4	0,1	10	90
Dialysis	10	2,8	0,28	28	63

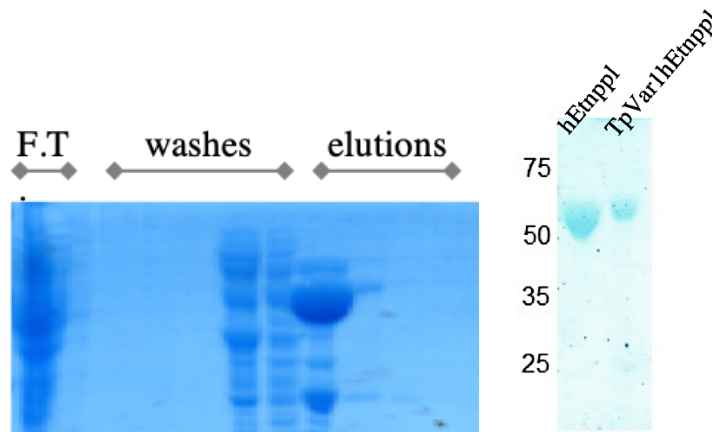


Figure 28. 10% SDS-PAGE assessing TpVar1-hETNPPL purity. Flow through, washes and elution fractions from the NiNTA purification step (left). Comparison between the final purified TpVar1-hETNPPL and hETNPPL (right).

Biochemical characterization

As shown in **Figure 29**, the spectral properties of hETNPPL and TpVar1- Etnppl are very similar, with the only exception of the peak at 330nm in the absorption spectra of the hETNPPL, which is characteristic of PMP and can be attributed to the slow reaction of PLP with amino compounds present in the Tris buffer. CD spectra in the far-UV region of TpVar1- hETNPPL is identical to those of hETNPPL, indicating that the addition of the TpVar1 peptide at the N-terminal end of the protein does not alter the secondary structure composition of hETNPPL.

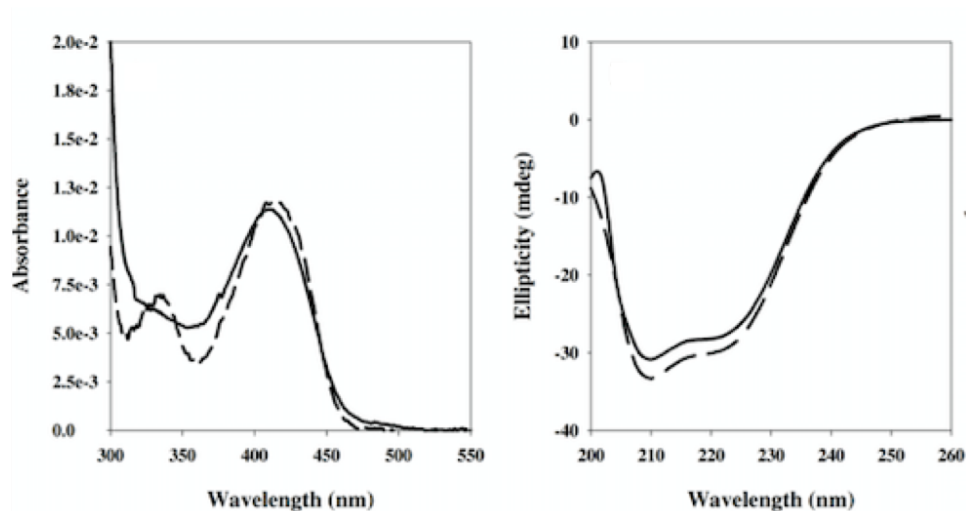


Figure 29. Spectral properties of hETNPPL and TpVar1- Etnppl. (A) Absorption spectra of 10 μM hEtnppl (dashed line) and TpVar1 (solid line). (B) CD spectra of 2.5 μM hETNPPL (dashed line) and TpVar1-hETNPPL (solid line).

The DLS analysis shows a hydrodynamic diameter of about 10 nm for both constructs (**Figure 30**), indicating that also the dimeric structure of hETNPPL is preserved in the hETNPPL mutant.

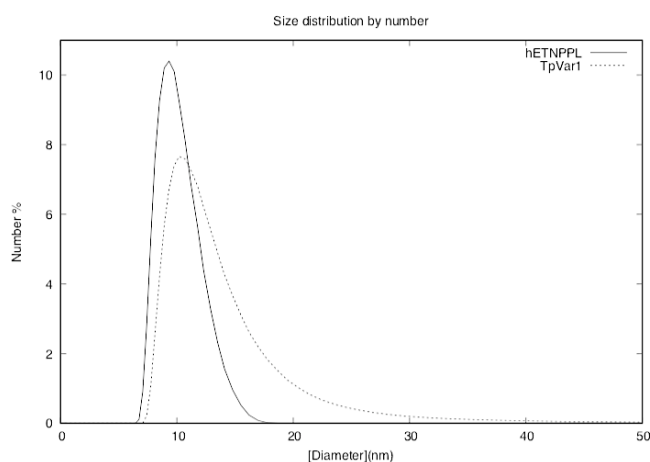


Figure 30. DLS analysis of hETNPPL and TpVar1-hETNPPL. The hydrodynamic diameter of both TpVar1-hETNPPL and hETNPPL is about 10 nm.

The steady-state kinetic parameters for the reaction catalyzed by TpVar1-hETNPPL have been measured and compared with those of the untagged hETNPPL (**Figure 31**). As reported in **Table 9**, although a nearly two-fold reduction of the k_{cat} and of the k_{cat}/K_m values for TpVar1-hETNPPL were observed, K_m was essentially unchanged (1 mM), confirming that TpVar1-hETNPPL retains a significant catalytic activity.

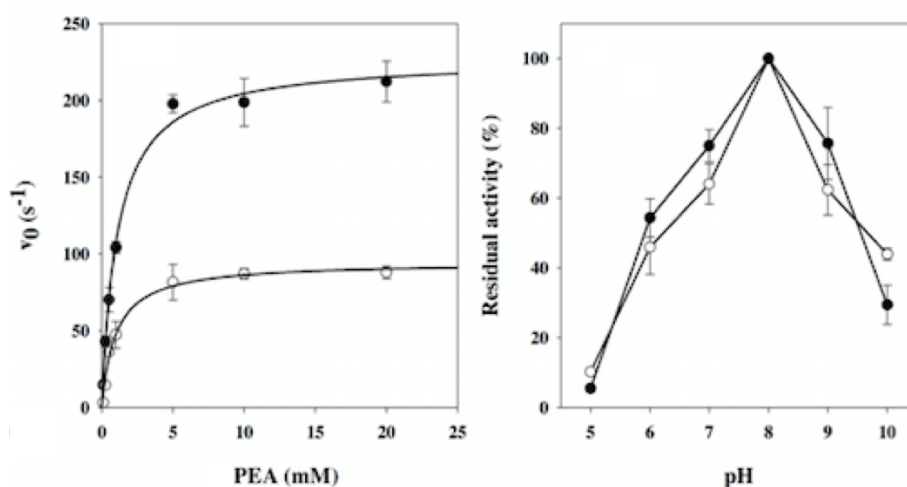


Figure 31. Kinetic characterization of hETNPPL and TpVar1- hETNPPL. The PEA phospho-lyase reaction catalyzed by recombinant hETNPPL (●) and TpVar1-ETNPPL (○). Rate of the elimination reaction as a function of the concentration of PEA at 30°C, pH 8.0. Fitting of the PEA data points to the Michaelis-Menten equation yielded the following values for the kinetic parameters: $K_m = 1.10 \pm 0.13$ mM, $k_{\text{cat}} = 227 \pm 7$, k_{cat}/K_m (s⁻¹M⁻¹) = $2.06 \times 10^5 \pm 5.4 \times 10^4$ for hETNPPL and $K_m = 0.98 \pm 0.16$ mM and $K_{\text{cat}} = 95 \pm 4$ $k_{\text{cat}}/K_m = 0.97 \times 10^5 \pm 2.5 \times 10^4$ for TpVar1. (B) pH-dependence of hETNPPL (●) and TpVar1- hETNPPL (○).

Table 9. The steady-state kinetic parameters for the reactions catalyzed by hETNPPL and TpVar1-hETNPPL.

	k_{cat} (s ⁻¹)	K_m (mM)	k_{cat}/K_m (s ⁻¹ M ⁻¹)
hETNPPL	227 ± 7	1.10 ± 0.13	$2.06 \times 10^5 \pm 5.4 \times 10^4$
TpVar1- hETNPPL	95 ± 4	0.98 ± 0.16	$0.97 \times 10^5 \pm 2.5 \times 10^4$

4.4. Concluding remarks

HCC is the most common malignant type of primary liver cancer and the third most common cause of cancer death worldwide^{107,108}. HCC is extremely resistant to chemotherapy and the only curative treatments that is currently available, the kinase inhibitor Sorafenib (Nexavar®), is effective in a minority of patients^{109,110}. Surgical resection or transplantation and palliative care are often the best options^{108,111}. Because of this, new and better strategies for the treatment of HCC are needed.

Since DNL is commonly upregulated in liver tumors, it is currently receiving considerable attention, in particular as a therapeutic target for the treatment of HCC patients^{112–116}. DNL generates lipids whose biochemical properties make them highly resistant to drug therapy. For this reason, a possible therapeutic approach in HCC treatment could consist in reducing DNL rate, thus limiting lipid availability. With this purpose, the exogenous administration of hETNPPL, which should be capable of interfering with the lipid pathway, could be used as adjuvant therapy in HCC treatment, helping to decrease the DNL rate.

One of the bottlenecks of an efficient enzyme administration is the crossing of the cellular membrane¹¹⁷. There are many systems to overcome this problem. Among them, particular attention has been paid to the so called cell penetrating peptides (CPPs), which are small peptides, mostly cationic in nature, whose shuttling properties have been exploited in a variety of enzyme replacement therapy study^{118–122}.

In this work, we initiated a protein transduction approach to deliver recombinant hETNPPL to HCC. In order to obtain TpVar1-hETNPPL fusion protein, I added the fragment encoding the TpVar1 sequence to the gene encoding hETNPPL by multiple PCR cycles and I cloned the PCR product in a suitable bacterial expression vector. I developed a protocol to obtain a good amount of pure protein. Then, I performed the biochemical characterization and compared the results with those of the hETNPPL to evaluate possible alterations in the activity or in the conformation due to the addition of the TpVar1 protein basic domain to the enzyme.

The biochemical characterization of the tagged mutant reveals that, despite the addition of the CPP, the enzyme is still able to catalyze PEA degradation. Moreover, no substantial differences in the overall conformation of the two enzymes were observed, indicating that TpVar1-hETNPPL can be used for the cellular experiments and that, in principle, it can be tested in an enzyme replacement therapy for HCC.

Chapter 5. Structure-based approaches to the discovery of modulators of E-Cadherin-Mediated Cell–Cell Adhesion

5.1. Introduction

The cadherins are a large family of transmembrane calcium-dependent cell adhesion proteins that mediate adherens junction, playing a crucial role in tissue morphogenesis and stability^{123,124}. They also take part in the signaling process that regulates cytoskeletal organization, cell cycle progression and differentiation^{125,126}. Moreover, they provide tissues with specific mechanical properties such as elasticity and the ability to withstand mechanical stress^{124,127}.

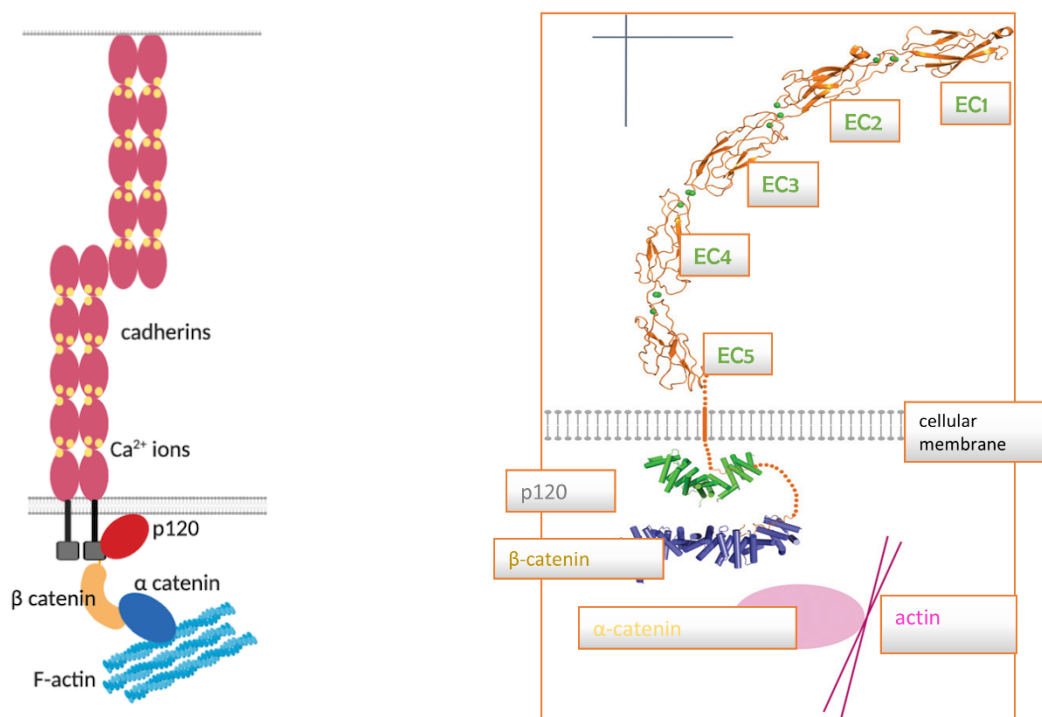


Figure 32. Schematic view of cadherin-mediated junctions (left). Schematic view merged with PDBs structures (right). Adapted from *Brash et al. (2012)*¹²⁸.

Cadherins superfamily comprises a wide range of cell adhesion molecules (**Figure 33**). Among them, the so-called ‘classical cadherins’, found in the adherens junction of epithelial cells, were the first and, to date, also the most studied cadherins^{129,130}. They can be further divided into type I or type II depending on selected sequence features.

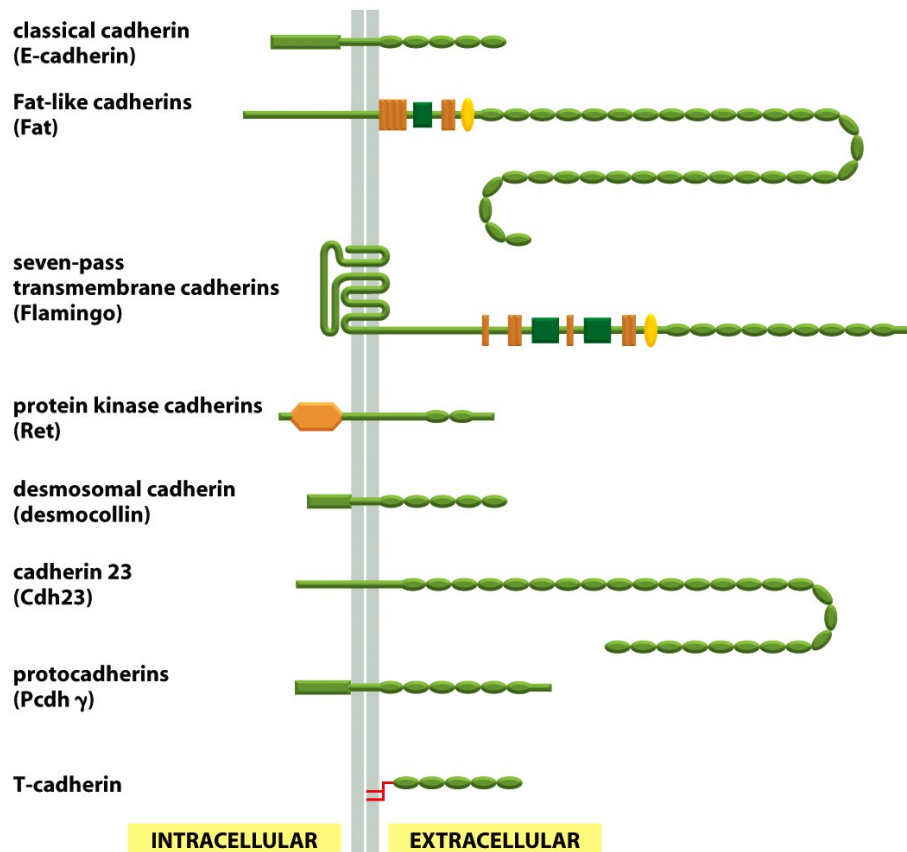


Figure 33. The cadherin superfamily.

The different members of the classical cadherins subfamily are conventionally named on the basis of the tissue in which they were first discovered. Following this convention, E-cadherin corresponds to cadherin found in epithelial tissue, N-cadherin in neuronal, R-cadherin in retinal, P-cadherin in placental, and so on.

All classical cadherin family members share a high degree of homology and structural similarity¹³¹. Their overall architecture features an elongated extracellular portion organized in five immunoglobulin-like domains (ECs) arranged in tandem, a single-pass transmembrane helix and an intracellular portion consisting of 150 amino acids, which is the most highly conserved portion of these proteins¹³². The dynamic association of cadherin intracellular tail with the cytoplasmic molecules p120 catenin and β -catenin, which in turn form a complex with α -catenin, allows communication between cadherins and the actin cytoskeleton¹³³. It is known that the binding between cadherins and these cytoplasmatic molecules is crucial for the folding of the cadherin cytoplasmic domain¹³⁴.

In addition to these general features, three calcium ions are found at the interface between each pair of ECs. These ions, which provide rigidity to the protein, are essential for protein activity¹³⁵.

Over the last two decades, a combination of structural and mutational studies conducted on classical cadherins has allowed the identification and the detailed characterization of a highly dynamic homo-dimerization mechanism that ultimately leads to the formation of a homo-dimeric complex, which is usually referred to as the “strand-swap” dimer^{125,126,136–142}. This mechanism consists of the opening of the N-terminal adhesion arm and the insertion of the side chain of the conserved tryptophan at position 2 (Trp2) into an acceptor pocket of the outermost extracellular domain (EC1) of the partner molecule. A crucial intermediate state, commonly referred to as the X-dimer (**Figure 34**), is involved in the binding event. This weakly adhesive dimeric conformation promotes strand-swap dimer formation by bringing the adhesion arms of two interacting cadherins in close proximity, thus lowering the energetic barrier associated with the arm swap¹⁴³.

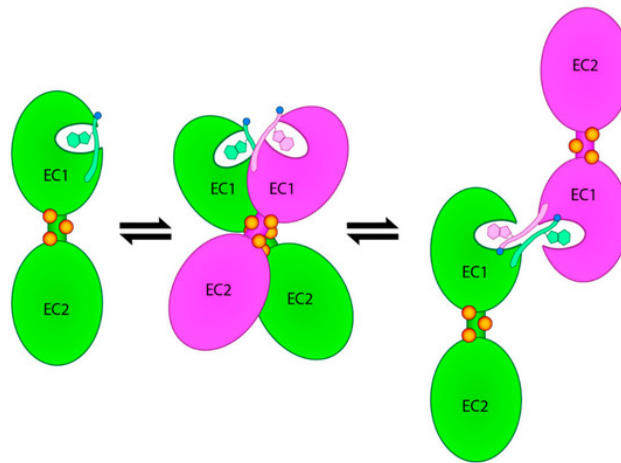


Figure 34. Schematic representation of the cadherin dimerization mechanism. The strand-swap dimer formation involves mutual insertion of the Trp2 side chain in the binding pocket of the partner molecule (right). The dynamic cadherin dimerization trajectory features a crucial intermediate configuration that is referred to as the X-dimer (center). Three calcium ions are shown as orange dots. Adapted from *Dalle Vedove et al. (2019)*¹⁴⁴.

The homophilic interaction occurring at the level of the EC1 of two cadherins protruding from two opposing cells to form the strand-dimer is commonly described as a *trans* interaction. However, further stabilizing contacts, mostly hydrophobic in nature, are found in all strand-swap structures between the EC1 of a cadherin and the EC2 of neighboring molecules present on the same cell, to form the so-called *cis* interaction. This lateral assembly supports and favors cadherin adhesion¹⁴⁵. No indication that the *trans* interaction would form in the absence of lateral *cis* interaction has ever been obtained, thus suggesting that the lateral self-assembly mechanism is key to the successful formation of the *trans* homo-dimer. Overall, the combination of these interactions forms a dynamic zipper-like

structure at the cell–cell contact, ultimately resulting in the formation of an adherens junction between two neighboring cells.

This general homo-dimerization multistep mechanism is shared by all members of the classical cadherin family. However, although all the different cadherins show a high sequence homology, cadherin association is very specific as it almost exclusively occurs between identical members of the family¹²³.

5.1.1. Cadherins as pharmaceutical targets

Over the years, many studies have confirmed the involvement of this class of adhesion molecules in numerous pathological conditions. In particular, classical cadherin dysfunction has been often correlated with cancer progression and metastasis¹⁴⁶. During cancer progression and metastasis, a switch in expression from epithelial E-cadherin (CDH1) to neuronal N-cadherin (CDH2) is considered a molecular hallmark of epithelial-to mesenchymal transition (EMT). In the majority of carcinomas E-cadherin is down-regulated suggesting a tumor repression role. However, in some epithelial ovarian cancer (EOC), E-cadherin expression is retained, thus favoring the proliferation of malignant cells¹⁴⁷. In addition to E- and N-cadherin, the aberrant expression of P-cadherin (CDH3) and cadherin-11 (CDH11) have also been correlated to different types of cancer such as malignant melanoma, breast, lung, gastric, pancreatic and colorectal cancer^{148–153}. Moreover, CDH11 plays a crucial role in the development of the synovium, a layer of cells that surrounds the bones' articulating surfaces and provides lubrication for the cartilage¹⁵⁴. Studies done using monoclonal antibodies against cadherin-11 in mouse models of inflammatory arthritis have demonstrated that cadherin-11 inhibition reduces cytokine production by synovial fibroblasts, thus contrasting their pathological behavior¹⁵⁴. Moreover, both N- and VE-cadherin are involved in blood vessel formation¹⁵⁵, which is a requirement for the growth of solid tumors¹⁵⁶.

Due to their involvement in cancer and other pathological conditions, many members of the classical cadherin family are currently being considered interesting pharmaceutical targets. In this context, a valid therapeutic approach consists in targeting the cadherin adhesive interface with small molecules that may inhibit or modulate cadherin-mediated cell adhesion. However, a rational approach to the design of small molecule inhibitors of cadherin homo-dimerization has, so far, been hampered by the cadherin's intrinsic dynamic behavior, which drives the protein's dimerization process through an extremely variable and energetically shallow conformational landscape.

The first example of the development of a small molecule directed against cadherin homodimerization is ADH-1 (**Figure 35**), a cyclic peptide based on the His79-Ala80-Val81 conserved sequence of the EC1 domain, that has been approved for entering phase I clinical trial as an anticancer agent for

patients with advanced solid tumors (ClinicalTrials.gov Identifier: NCT00225550, NCT00264433, NCT00390676, NCT00265057, NCT00421811, and NCT01825603)^{157–159}.

ADH-1 has been found to inhibit N-cadherin-mediated processes with a modest activity, which was associated mainly with the disruption of cadherin-mediated cell-cell contacts of metastatic cells with the surrounding endothelial cells. However, no information regarding its actual binding mode and inhibition mechanism has ever been reported, thus hampering the design of more effective cadherin homodimerization inhibitors.

Recently, Doro and colleagues designed and tested a library of peptidomimetic compounds based on the N-terminal sequence of the E/N-cadherin adhesion arm (DWVI). In these compounds the central dipeptide unit (Trp2-Val3) was substituted by a scaffold bearing an aromatic group which mimics the indole moiety of the Trp2 residue that is crucially involved in the cadherin homo-dimerization mechanism¹⁶⁰. Among this library, the FR159 (**Figure 35**) resulted the more promising compounds, inhibiting N- and E-cadherin homophilic adhesion at 1 mM concentration more efficiently of ADH-1.

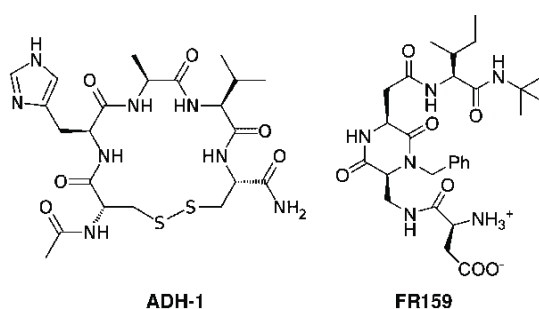


Figure 35. Structure of ADH-1 and FR159. Adapted from *Doro et al. (2015)*¹⁶⁰.

Interestingly, the high resolution crystal structure of the complex between the deleted form of E-cadherin EC1-EC2 domain (lacking the N-terminal Asp1-Trp2 residues) and the FR159 (PDB code: 4zte)¹³¹, showed that the actual binding region of the inhibitor is not within the Trp2 cavity, as initially postulated in the design phase (**Figure 36**). Indeed, the FR159 ligand was found to bind at the level of the adhesion arms of the two interacting cadherins, across the weakly adhesive X-dimer conformation, a binding mode that prevents strand dimer formation by blocking the opening of the adhesion arm at the level of the Pro5-Pro6 portion of the adhesion arm (the “elbow” or “hinge” of the six amino acid-long N-terminal strand exchange portion)¹³¹.

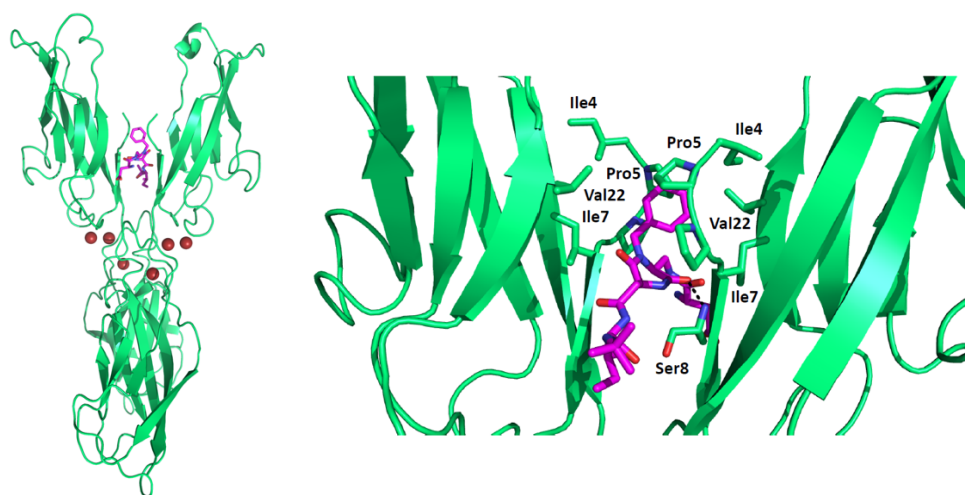


Figure 36. Crystal structure of the human E-cadherin-(Val3)-EC1EC2 fragment (residues 3–213) in complex with FR159 (PDB code 4zte) (left). Hydrophobic interactions between the ligand and the protein. (right). Adapted from *Nardone, V. et al. (2016)*¹³¹.

As shown in **Figure 36**, the peptidomimetic ligand is mostly stabilized by hydrophobic contacts such as those involving, in particular, the phenyl ring and the side chains of residues Ile4, Pro5, Ile7, Leu21, and Val22 from both cadherin molecules. Interestingly, the residues involved in ligand stabilization are conserved across most type I classical cadherins.

This unexpected binding mode has been attributed to the large conformational changes occurring in the course of the complex cadherin dimerization process. Hence, it is very likely that during the process the ligand may also bind, transiently, to different surface areas of the protein. Nevertheless, this new binding interface represents without any doubts a region where suitable ligands can interfere with the cadherin homo-dimerization process.

In this respect, as a continuation of the work about the complex between E-cadherin EC1-EC2 and the FR159 that was carried out in our laboratory ¹³¹, during my PhD, I took part in two different studies that developed from there. The first study consisted in a hybrid approach whereby NMR (Nuclear Magnetic Resonance) and computational techniques were utilized to provide a possible dynamic picture of the binding event between E-cadherin EC1-EC2 and the FR159.

The second study was related to a structure-based virtual screening approach for the identification of commercially available drug-like molecules that may act similarly to FR159, albeit with a higher potency.

These two studies were published in:

Civera M, Vasile F, Potenza D, Colombo C, Parente S, **Vettrai** C, Prosdocimi T, Parisini E, Belvisi L. Exploring E-cadherin-peptidomimetics interaction using NMR and computational studies. (2019) *PLoS Comput. Biol.* 15 (6): e1007041.

Dalle Vedove A, Falchi F, Donini S, Dobric A, Germain S, Di Martino GP, Prosdocimi T, **Vettrai** C, Torretta A, Cavalli A, Rigot V, André F, Parisini E. Structure-based virtual screening allows the identification of efficient modulators of E-cadherin-mediated cell-cell adhesion. (2019) *Int. J. Mol. Sci.* 20(14): E3404.

5.2. Exploring cadherin-modulator interactions using NMR and computational studies.

Owing to the complexity of the E-cadherin homo-dimerization process, to better elucidate the binding event between E-cadherin EC1-EC2 and FR159, we set out to analyze the interaction between the two species in solution using a combination of NMR (Nuclear Magnetic Resonance) and computational techniques, in collaboration with the group of Prof. Laura Belvisi (University of Milano).

In this study, ligand-based NMR techniques (Saturation Transfer Difference, STD, and transferred NOE, tr-NOESY)^{161,162}, which were performed in the Belvisi group, were used to assess binding occurrence and to identify the binding epitope of the ligands at different temperatures and on two different E-cadherin constructs (the wild type E-cadherin EC1EC2 and the V3-EC1EC2 construct). The NMR data thus obtained were then analyzed computationally in order to rationalize the atomic details of possible peptidomimetic-cadherin interactions over time.

Briefly, STD-NMR is a method based on Overhauser effect^{163–165}, which is a spin-spin coupling phenomena that provides information on the reciprocal distance of the sampled nuclei, allowing the extrapolation of the 3D structure of the molecule, as long as its structure formula is known. The method relies on the selective irradiation of the protein, which allows magnetization to be transferred to the bound ligand, which is present in large excess relative to the protein. Because of the binding equilibrium, the saturated ligand is displaced in solution. Since ligand protons closer to the protein are more likely to become highly saturated, they show the strongest signal in the mono-dimensional STD spectrum. Therefore, the modulation of the ligand signal intensity is used as an epitope-mapping method to describe the target-ligand interactions that occur over time.

In order to investigate a wide range of possible interactions between FR159 and E-cadherin, STD-NMR experiments were carried out in the presence of two slightly different E-cadherin constructs, the wild type-E-cadherin-EC1-EC2 and the deleted form of E-cadherin (Val3)-EC1-EC2. The latter, which lacks the two N-terminal residue Asp1 and Trp2, was used to solve the crystal structure of E-

cadherin in complex with FR159. I prepared both proteins according to the protocol described in the material and methods section.

The results of STD-NMR showed a different binding epitope of the ligand and also a different temperature effect on STD signals for the two different protein constructs, suggesting an involvement of the Asp1-Trp2 sequence during the dynamic binding events that occur over time (**Figure 37**).

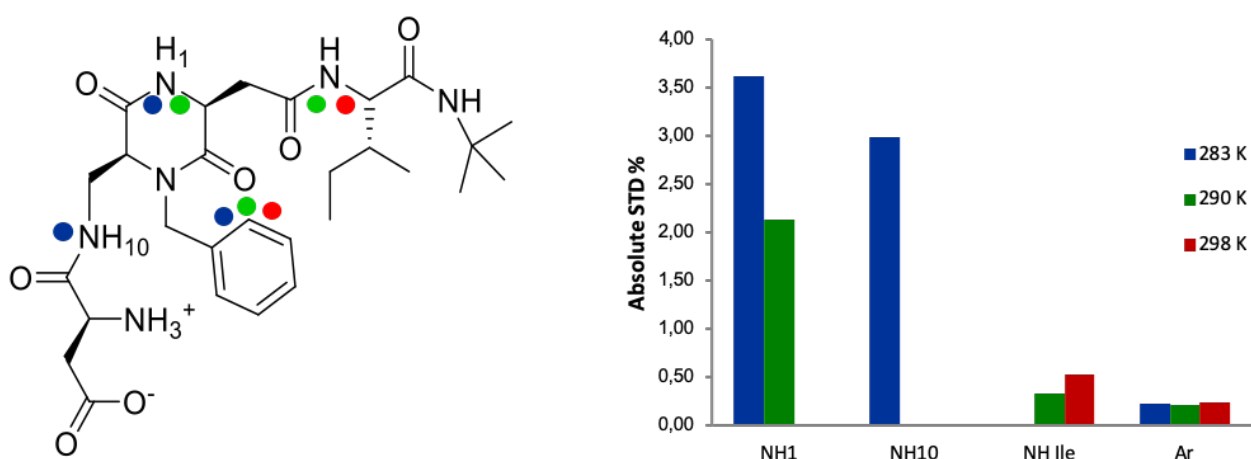


Figure 37. The binding epitope of FR159 in the presence of wild type E-cadherin-EC1-EC2. The dots on the structure indicate the ligand binding epitope at different temperatures (left, blu 283K, green 290K and red 298K). The histogram shows the comparison of absolute STD % at different temperatures (right). Adapted from *Civera et al. (2019)*¹⁶⁶.

The ligand binding mode was further investigated by MD simulations starting from the corresponding docking poses of FR159 into a wild type-E-cadherin EC1 model (**Figure 38**). Results of these calculations showed that the starting ligand binding mode was quite conserved during the MD runs. Indeed, the aromatic group remained docked into the Trp2 pocket, also at 300 K and 320 K. This result supports the presence of the aromatic hydrogens STD signal (detected at all temperatures). By contrast, protein flexibility increased with temperature, especially for the adhesive arm residues.

Overall, MD results on ligand-cadherin binding were in agreement with STD spectra, suggesting that the binding interactions between the protein and the inhibitor are not constant over time and depend on the protein flexibility, especially the adhesive arm, an active part of the binding site that participate in the interaction with the ligand.

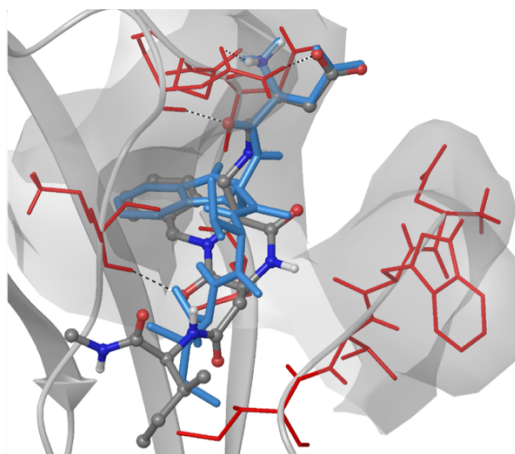


Figure 38. Docking binding mode of FR159 into E-cadherin. The ligand (grey carbon atoms) is superimposed to the DWVI sequence (light blue) of the E-cadherin X-ray crystal structure (PDB code: 3Q2V). Residues of the binding pocket interacting with the ligand are labelled. Adapted from *Civera et al. (2019)*¹⁶⁶.

This kind of behavior was expected. Indeed, in our case the inhibitor interacts with a surface formed by two protein molecules, mostly through weak hydrophobic interaction rather than with the active site of an enzyme. Moreover, because of the highly dynamic behavior of the protein, the interface between two interacting cadherin is subjected to many conformational changes in the course of the homo-dimerization process.

In order to map the protein residues that are involved in the interaction with the inhibitor and clarify its mechanism of action we also decided to perform, as a complement to the STD-NMR experiments, HSQC-NMR experiments with labelled E-cadherin constructs (wild type vs. deleted forms). This technique exploits a J-coupling effect to transfer the nuclear spin polarization and enhance the NMR signal and a correlation between two different types of nuclei, such as ^1H - ^{13}C or ^1H - ^{15}N . This latter is mostly used with proteins, where the presence of amides can be exploited. In our case, by comparing the spectrum of the protein alone with the spectrum of the complex, we expected to be able to identify the shift of specific amide peaks and therefore, which residues interact with our compounds. To this end, according to the procedure described in the material and methods section, we produced both human E-cadherin EC1-EC2 and human E-cadherin (Val3) EC1- EC2 samples enriched with ^{15}N .

These HSQC-NMR experiments are currently still ongoing.

5.3. Discovery of cadherin modulators by structure-based virtual screening

The recent crystal structure of E-cadherin-FR159 complex (PDB code: 4zte)¹³¹, which was determined in our lab, allowed the identification of an unexpected druggable interface, paving the way for the rational design of new modulators of E-cadherin-mediated cell–cell adhesion. During my PhD, we used a structure-based virtual screening approach to validate the new druggable interface and, at the same time, to identify more potent and specific modulators of E-cadherin-mediated cell–cell adhesion.

In collaboration with the group of Prof. Andrea Cavalli (Italian Institute of Technology, Genova), we conducted a virtual screening (VS) study on two libraries of commercially available drug-like compounds. Based on their similarity with FR159, some of the compounds that we identified via this method were expected to bind E-cadherin using the same interface identified in the E-cadherin-FR159 complex. At first, 1000 of the best-scored compounds were visually inspected. A similarity cluster analysis reduced the number of candidate compounds to 200. Of these, 15 compounds (one for each cluster) were selected to be tested experimentally. The selected compounds were then tested *in vitro* in the group of Prof. Frédéric André (Aix-Marseille Université, France). To investigate the effect of the candidate inhibitors on cadherin-dependent cell–cell adhesion, a spheroid assay was performed on human pancreatic tumor BxPC-3 cells, which express both E- and P-cadherin. In this test, cells dispersed in a dense media are allowed to aggregate into spheroids. Since cells use cadherins to aggregate, the morphological changes of the resulting spheroids can be related to the activity of the tested molecules.

At the same time, in our laboratory I set up a series of co-crystallization experiments on the top-scoring molecules using the same approach as described for FR159¹³¹.

Cell-cell adhesion assay

At first, working with the lower active concentration determined for the FR159 ligand¹⁶⁰ (1 mM), we found that 5 of the 15 identified compounds (AS2, AS8, AS9, AS11, and LC11) counteracted BxPC-3 E-cadherin/P-cadherin ability to form spheroids. The structural formulas and commercial codes of the selected molecules are listed in Appendix I (Pag. 121). Subsequently, due to solubility problems, we decided to lower the inhibitor concentration to 0.05 mM. At this concentration, only three molecules (AS11, AS9, and, to a lesser extent, AS8) displayed anti-cell–cell adhesion activity (**Figure 39A**) promoting formation of spheroids that were significantly less compact than when cells were treated with DMSO (**Figure 39B**).

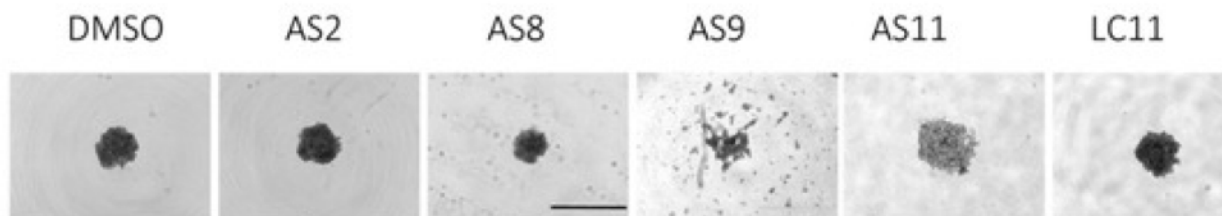
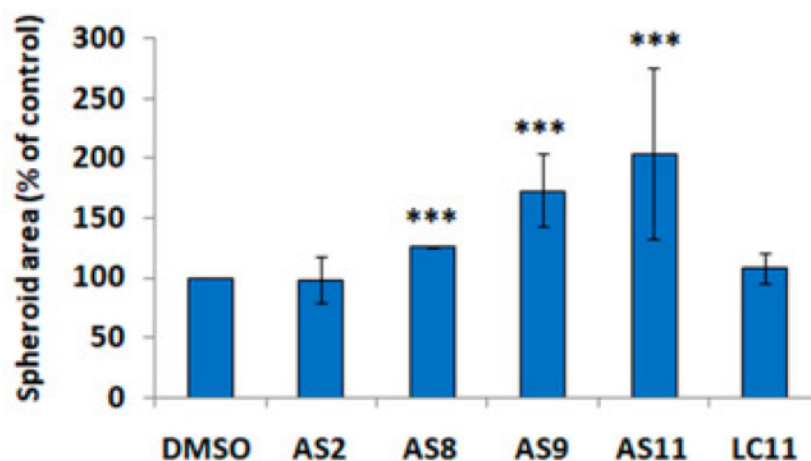
A**E-Cadh/P-Cadh Cells****B**

Figure 39. AS11, AS9, and, to a lesser extent, AS8 regulate cell–cell adhesion. BxPC-3 E-cadherin/P-cadherin cells were incubated in the presence of either 0.1% DMSO or 0.05 mM AS2, AS8, AS9, AS11, or LC11 and allowed to form spheroids in suspension for 24 h. (A) Representative pictures of spheroids. Scale bar: 1000 μ m. (B) The spheroid area was measured by phase-contrast microscopy and analyzed by ImageJ. Values represent the mean of two experiments performed in octuplet. A $p < 0.05$ was considered statistically significant and is indicated by ‘***’ when $p < 0.001$. Adapted from *Dalle Vedove et al. (2019)*¹⁴⁴.

At this stage, I set out to investigate the direct interaction between cadherins and AS9 or AS11 using either isothermal titration calorimetry (ITC) or microscale thermophoresis (MST). Both of them are physical techniques used to study the binding of small molecules to proteins.

Briefly, MST assay is based on thermophoresis, the movement of molecules in a temperature gradient induced by an infrared laser. In this technique, the molecules are distributed homogenously with an

initial fluorescence signal. When the IR laser is activated, the fluorescent signal is decreased. With the turnoff of the IR-laser, the molecule diffusion is back, solely driven by mass diffusion. The trace difference between a fluorescent molecule binding with or without non-fluorescent ligands indicates a binding signal. ITC measures the heat exchange associated with molecular interactions at a constant temperature. It directly determines the binding affinity (K_a), enthalpy changes (ΔH), and binding stoichiometry (n) of the interaction between two or more species in solution. The experimental methodology involves performing several titrant injections from a syringe (usually the ligand) into the solution (usually the macromolecule) in the cell, while maintaining the system at isobaric, quasi-isothermal conditions. As the first injection is made, the microcalorimeter measures all heat released until the binding reaction has reached equilibrium. The quantity of heat measured is in direct proportion to the amount of binding.

Unfortunately, all the attempts to obtain a complete dose-response curve using either ITC or MST in the presence and in the absence of the two ligands have not been successful. Indeed, the DW deletion in the cadherin-EC1-EC2 fragment generates a truncated mutant whose X-dimer conformation features a $K_D > 660 \mu M$ ¹⁶⁷ and, at the concentration values needed to acquire a complete and reliable dose-response the protein sample shows gelification or even precipitation, thus preventing us from completing the experiments.

To overcome this inconvenient, we decided to follow a different strategy using RNA interference. Our approach consisted in silencing each of the two cadherins individually and perform a cell–cell adhesion assay using BxPC-3 E-cadherin or BxPC-3 P-cadherin (**Figure 40**).

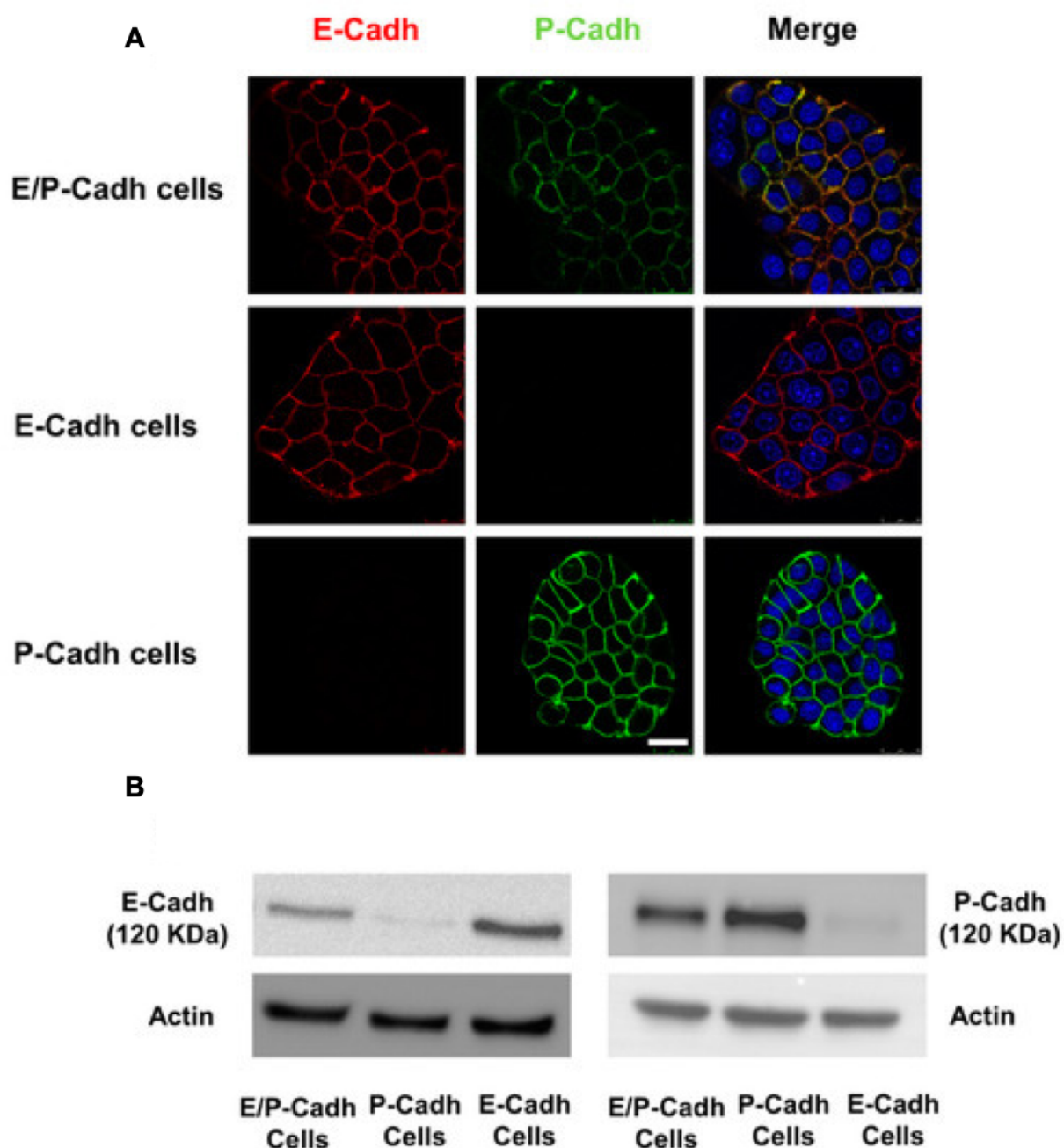


Figure 40. Cell surface cadherin extinction was assessed by both immunofluorescence and western blot. (A) Using RNA interference, E- and P-cadherin were stably knocked down in human pancreatic tumor BxPC-3 cells expressing both cadherins (see the Materials and Methods section for experimental details). Images were captured and analyzed using a SP5 Leica confocal microscope equipped with LAS AF Lite software. Scale bar: 25 μ m. (B) BxPC-3 E-cadherin/P-cadherin, BxPC-3 E-cadherin, and BxPC-3 P-cadherin cells were lysed, and the expressions of both P-cadherin and E-cadherin were detected by western blot. Adapted from *Dalle Vedove et al. (2019)*¹⁴⁴.

Results showed that AS9 and AS11 were able to inhibit BxPC-3 E-cadherin cell–cell adhesion at a 50 μ M concentration (**Figure 41A** and **Figure 42B**), while they had no effect on cells expressing only P-cadherin (**Figure 41B**). Moreover, AS9 affected BxPC-3 E-cadherin cell aggregation less

efficiently than AS11 for whichever concentration was used (**Figure 42**). Both AS9 and AS11 influenced cell–cell adhesion without affecting cell viability, as observed by the trypan blue exclusion assay (**Figure 43**).

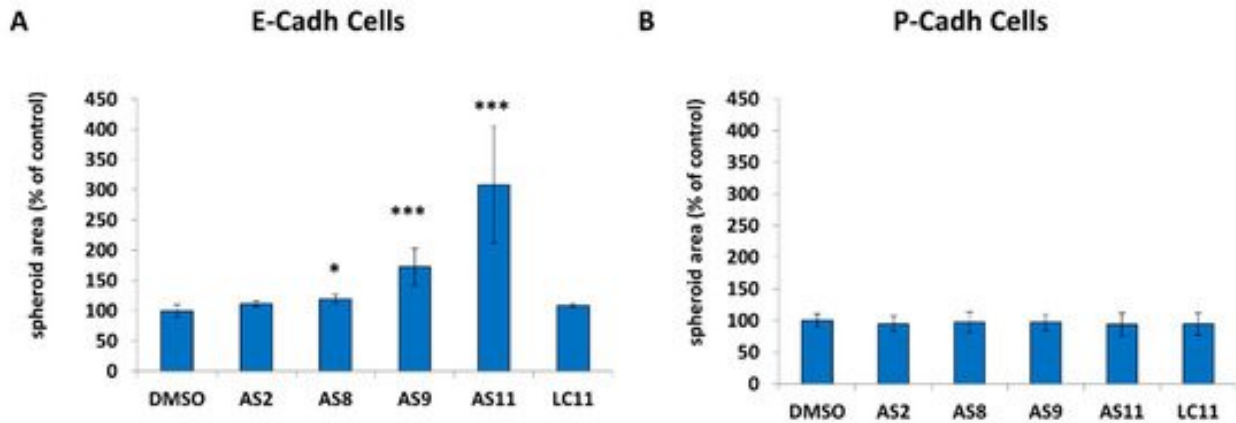


Figure 41. Both AS11 and AS9 target E-cadherin. BxPC-3 E-cadherin cells (P-cadherin depletion) (A) and BxPC-3 P-cadherin cells (E-cadherin depletion) (B) were incubated in the presence of 0.05 mM (E-cadherin cells) or 0.1 mM (P-cadherin cells) AS2, AS8, AS9, AS11, or LC11 and allowed to aggregate in suspension for 24 h. The spheroid area was measured by phase-contrast microscopy and analyzed by ImageJ. A $p < 0.05$ was considered statistically significant and is indicated by ‘***’ when $p < 0.001$ and ‘*’ when $p < 0.05$. See the Materials and Methods section for details on the statistical analysis. Adapted from *Dalle Vedove et al. (2019)*¹⁴⁴.

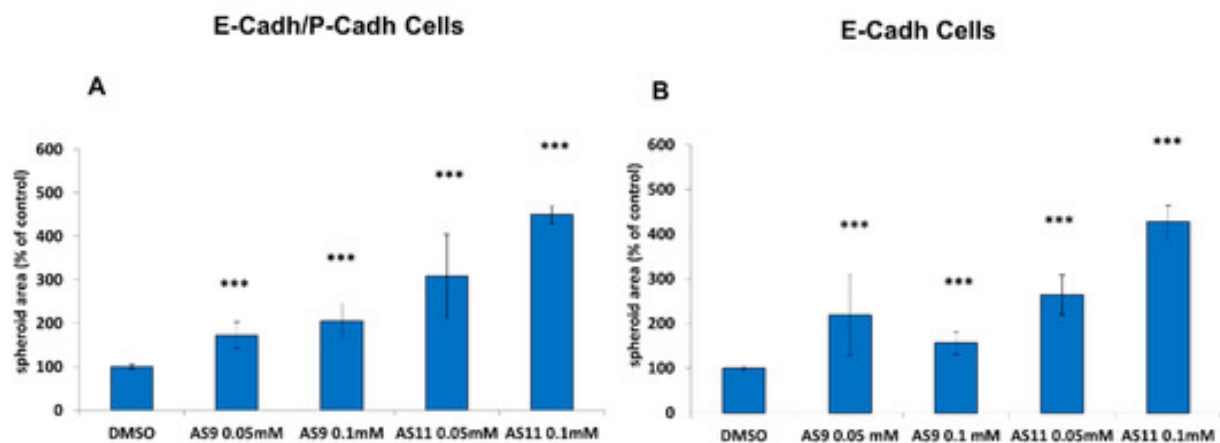


Figure 42. AS11 is more potent than AS9 to impair cell–cell interaction. BxPC-3 E-cadherin/P-cadherin cells (no cadherin depletion) (A) and BxPC-3 E-cadherin cells (P-cadherin depletion) (B) were incubated in the presence of AS9 or AS11 at various concentrations and allowed to form spheroids in suspension for 24 h. The spheroid area was measured by phase-contrast microscopy and analyzed by ImageJ. A $p < 0.05$ was considered statistically significant and is indicated by ‘***’ when $p < 0.001$. See the Materials and Methods section for details on the statistical analysis. Adapted from *Dalle Vedove et al. (2019)*¹⁴⁴.

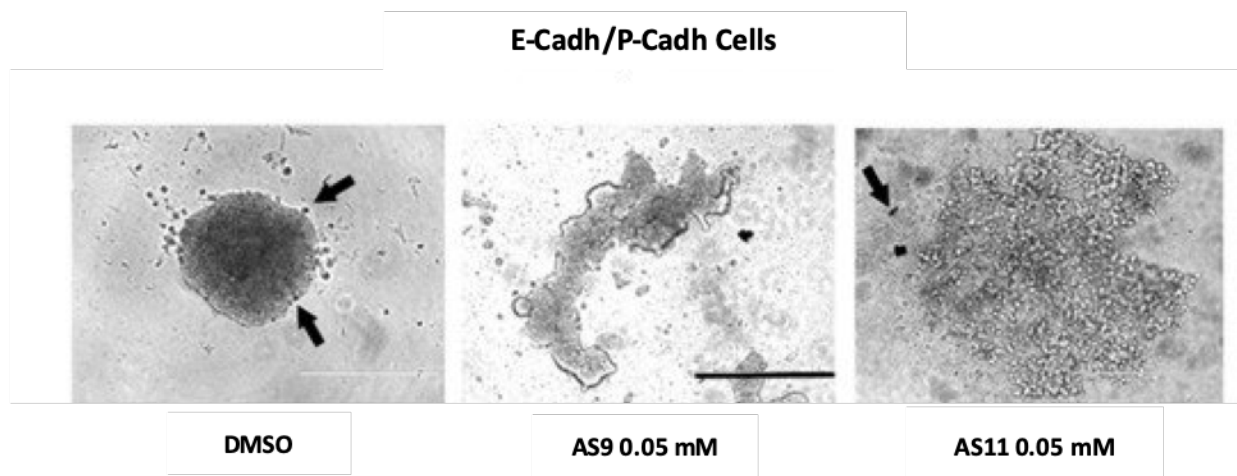


Figure 43. BxPC-3 E-cadherin/P-cadherin cells were treated with 0.1% DMSO or 0.05 mM of inhibitor and allowed to form aggregates for 24 h. Cell viability was then assessed by a trypan blue exclusion test. Black arrows indicate dead cells. Scale bar: 400 μ m. Adapted from *Dalle Vedove et al. (2019)*¹⁴⁴.

Interestingly, all our numerous attempts to crystallize either AS9 or AS11 in complex with E-cadherin following the same approach used for FR159 failed, most likely because of the much higher potency of these two compounds relative to FR159, which makes them incompatible with the formation of the stable hydrophobic pocket where FR159 was found to bind the E-cadherin X-dimer¹³¹. The theoretical binding mode for the two most potent inhibitors identified in this study is shown in **Figure 44**.

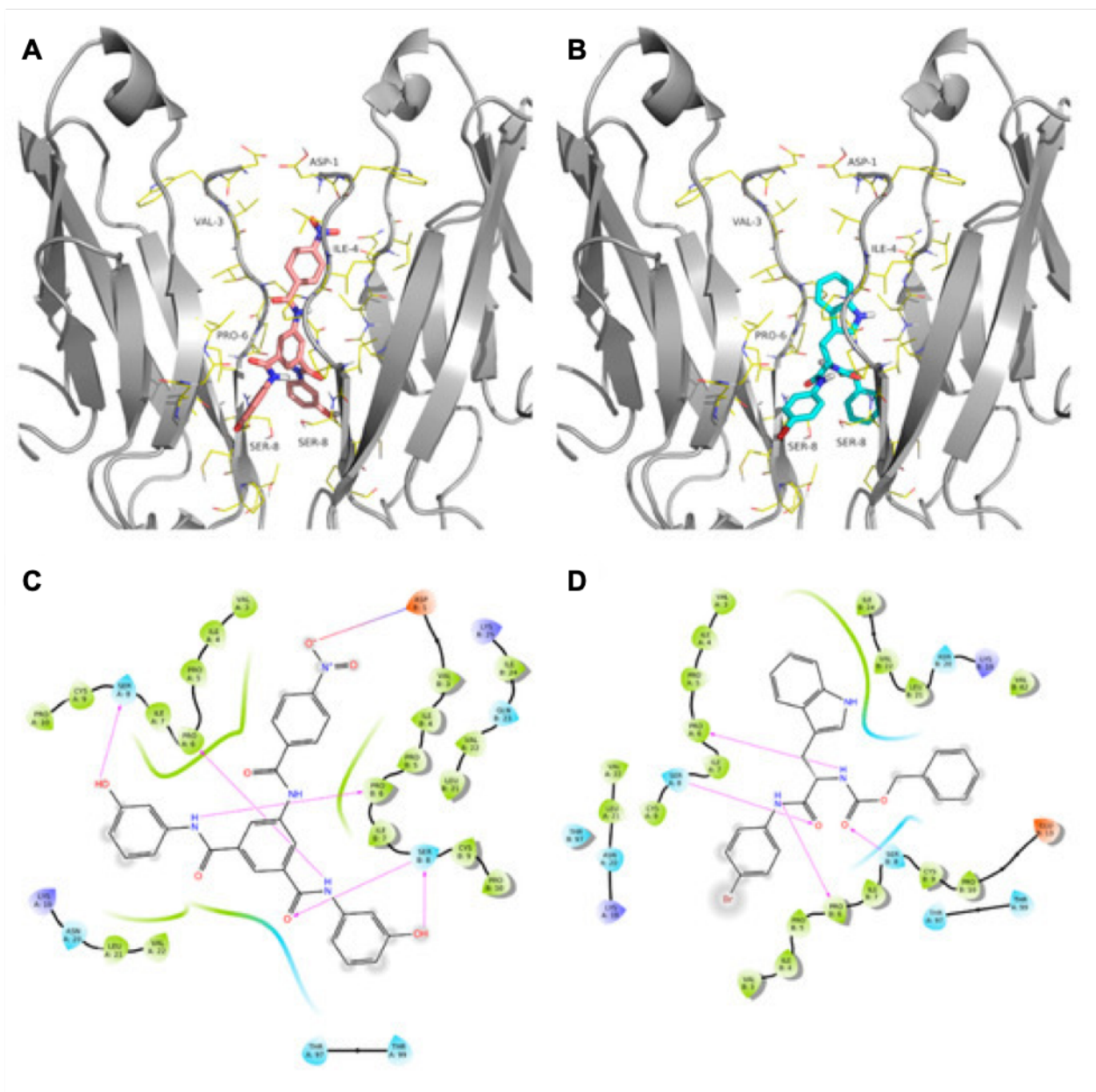


Figure 44. 3D representation of the theoretical binding mode for AS9 (A) and AS11 (B). 2D ligand interaction diagram of the theoretical binding mode for AS9 (C) and AS11 (D). Hydrogen bond interactions are shown as pink arrows. Positive and negative charged amino acids are represented in blue and red, respectively. Residues involved in hydrophobic or polar interactions are shown in green and light blue, respectively. Ligand-exposed fractions are indicated as a gray, circular shadow. Adapted from *Dalle Vedove et al. (2019)*¹⁴⁴.

Cell Invasion

Since both E-cadherin and P-cadherin are involved in pancreatic cancer cell invasion¹⁶⁸, we set out to analyze the impact of both AS9 and AS11 compounds on the invasive capacity of the three human pancreatic cell models. As shown in **Figure 45A**, BxPC-3 E-cadherin/P-cadherin organized in spheroids invaded a 3D type I collagen gel. Interestingly, AS11, but not AS9, was found to block the

invasive capacity of E-cadherin-expressing cells. However, neither AS9 nor AS11 decreased the invasive capacity of P-cadherin-expressing cells (**Figure 45B**).

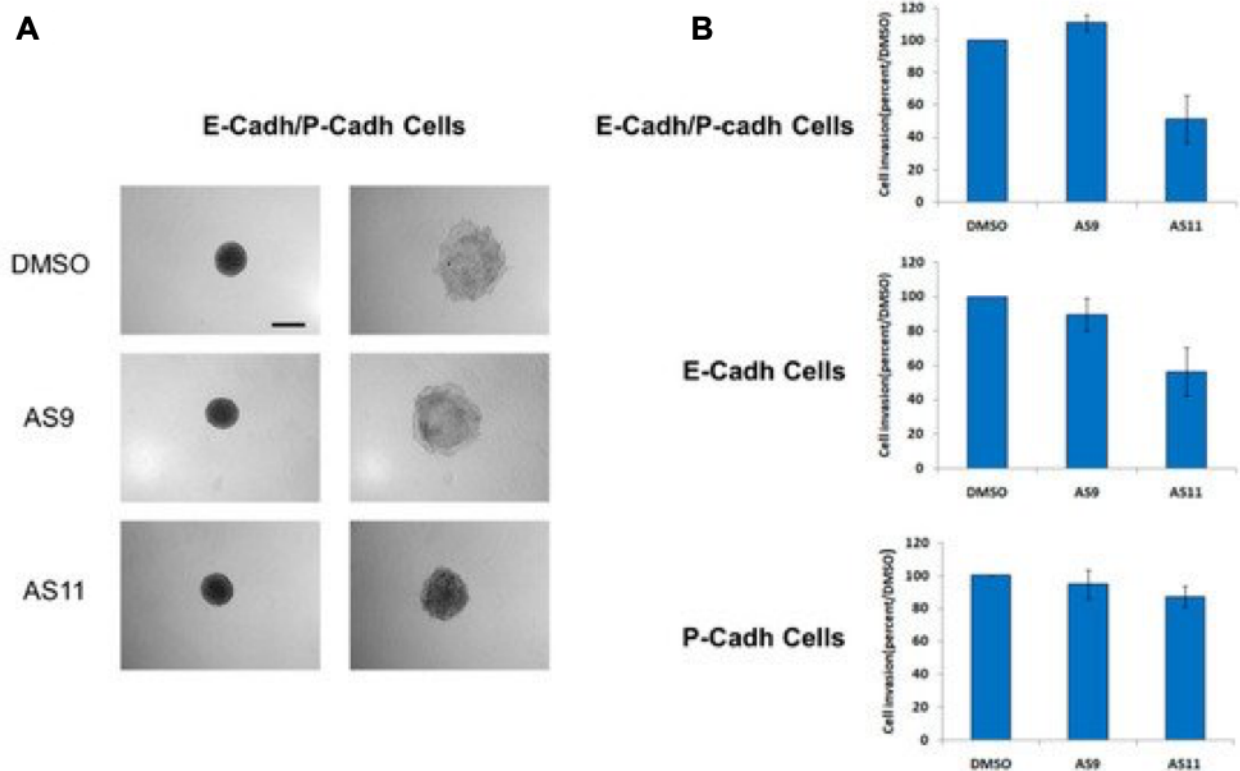


Figure 45. AS11, but not AS9, regulates cell invasion. BxPC-3 E-cadherin/P-cadherin cells (no cadherin depletion), BxPC-3 E-cadherin cells (P-cadherin depletion), and BxPC-3 P-cadherin cells (E-cadherin depletion) were allowed to form spheroids for 72 h. Spheroids were then embedded in type I collagen. After embedding, followed by a 24 h incubation in the presence of either 0.2% DMSO or 0.1 mM AS9 or AS11, the spheroid area was observed by phase contrast microscopy (A). Scale Bar: 500 μ m. (B): the spheroid area was measured using ImageJ. Results are expressed as the percentage of invasion of treated cells compared to control cells. Adapted from *Dalle Vedove et al. (2019)*¹⁴⁴.

When AS11 was used, this blockage of cell invasion effect could be observed for a concentration as low as 0.05 mM. (**Figure 46**). Taken together, these results confirm that AS11 selectively targets E-cadherin rather than P-cadherin. Since depletion of one of the both cadherins expressed leads to a decrease in cell invasion, this indicates that AS11 could be considered as an anti-invasive compound for pancreatic cancer cells.

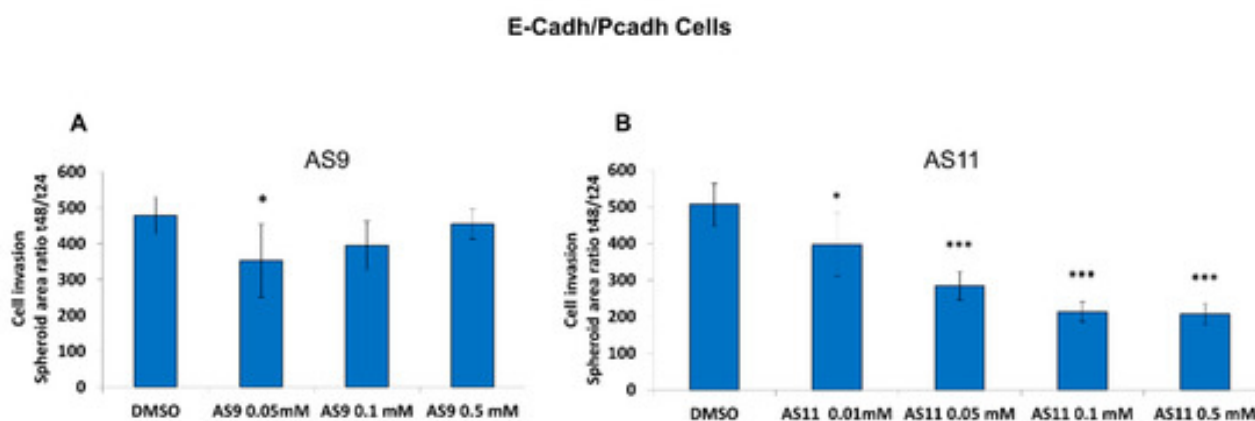


Figure 46. Spheroid area ratio t48/t24 for AS9 and AS11. BxPC-3 E-cadherin/P-cadherin cells (no cadherin depletion) were allowed to form spheroids for 72 h. Spheroids were then embedded in type I collagen and incubated for 48 h in the presence of either 0.1% DMSO or AS9 (A) or AS11 (B) concentrations ranging from 0.01 to 0.5 mM. The spheroid area was observed by phase-contrast microscopy and measured after 24 and 48 h incubation. See the Materials and Methods section for details on the statistical analysis. A $p < 0.05$ was considered statistically significant and is indicated by ‘***’ when $p < 0.001$ and ‘*’ when $p < 0.05$. See the Materials and Methods section for details on the statistical analysis. Adapted from *Dalle Vedove et al. (2019)*¹⁴⁴.

The results showed in this part of the work demonstrate that, by using a structure-based virtual screening approach, we were able to identify selective E-cadherin modulators with potential pharmaceutical efficacy. Also, the functional data shown here validate the druggable interface previously identified by X-ray crystallography demonstrating the importance of the crystal structure of the complex between the E-cadherin-EC1EC2 fragment and the FR159¹³¹. Indeed, this structure led to the identification of new druggable cadherin pocket, paving the way for future structure-based drug development studies aimed at the selective modulation of cadherin-mediated cell–cell adhesion.

5.4. Materials and methods

I produced two different protein fragments: E-cadherin (Val3) EC1-EC2, E-cadherin EC1-EC2, both from Homo sapiens. Moreover, in order to complete NMR studies, I produced the same fragments also as ¹⁵N labelled protein.

The protocol used for the E-cadherin (Val3) EC1-EC2 production is the same used for the structural characterization of the E-cadherin-FR159 complex¹³¹, but an additional dialysis against a 1 mM CaCl₂ aqueous solution was done prior to the NMR experiments. This precaution was adopted in order to prevent the precipitation of PBS (phosphate-buffer saline) in the presence of the Ca²⁺ ions that are required to ensure cadherin activity. The protocol used for the production of ¹⁵N labelled proteins is

the same as for the unlabelled ones, with the only exception that cell growth and protein production were carried out in a minimal medium, supplemented with 10% of ^{15}N NH_4Cl (10 g/l) as a source of ^{15}N (**Table 10**).

Table 10. Minimal medium composition.

10x M9 salts:	20%	1 M	1 M	100 mg/ml	100x
	glucose solution	MgSO_4	CaCl_2	ampicillin	MEM vitamins
67.8 g/l Na_2HPO_4					
30.0 g/l KH_2PO_4					
5.0 g/l NaCl					
100 ml/L	15 ml/l	1 ml/l	0.1 ml/l	1 ml/l	10 ml/l

5.4.1 Gene cloning

Two different genes encoding for the EC1-EC2 portion of human E-cadherin, one including residues 1–213 and one lacking the first two N-terminal residues (Asp1-Trp2) (i.e. including residues 3–213) have previously been cloned separately into two pET-3a expression vectors (Novagen) using the NdeI and BlnI restriction sites. In both cases, fragments were fused at their N terminus to a 6His-tag, a spacer peptide (Ser-Ser-Gly-His-Ile), and the enterokinase recognition site (Asp-Asp-Asp-Asp-Lys). All the primers used were supplied by Eurofins scientific or Metabion as lyophilized powder, while restriction enzymes were purchased from Thermo-Scientific.

5.4.2 Protein production

Protein expression

Both constructs were expressed in *E. coli* (BL21(DE3)pLysS cells (Invitrogen). Briefly, 50 μl of *E. coli* (BL21(DE3)pLysS cells were incubated with 1 μl of pET-3a expression vector for 30 minutes on ice, then left for 90 seconds at 42°C (to allow the DNA to diffuse into the cells) and subsequently moved back on ice for 2 minutes. The mixture was added to 950 μl of LB broth (10 g/l NaCl , 10 g/l of tryptone enzymatic digest, 5 g/l yeast extract), incubated at 37°C with orbital shaking at 180 rpm for 45 minutes, pelleted for 6 minutes at 8000 rpm, resuspended in 100 μl of LB broth and finally spread on LB/agar (same as the broth plus 18 g/l of LB-Agar) plates supplemented with 100 $\mu\text{g}/\text{ml}$ of ampicillin. The plates were then incubated overnight at 37°C. Standard production was carried out

in 1 l of LB broth supplemented with 50 µg/ml of ampicillin. Once the optical density of the solution ($\lambda=600$ nm) reached 0.6, protein expression was induced with 0.5 mM IPTG. After overnight expression at 25°C in an orbital shaker at 180 rpm, cells were harvested by centrifugation for 10 minutes at 10000 rpm and the resulting pellet, if not immediately processed, were stored at -80°C.

Protein extraction and purification

Cells were resuspended in TBS buffer (Tris-buffered saline, 50mM Tris, 150 mM NaCl, pH 7.4) to a final concentration of 0.2 g/ml. The suspension was supplemented with 140 µl/l of DNase (Sigma Aldrich), 14 ml/l of PMSF (Phenylmethylsulfonyl fluoride, a broad range protease inhibitor, Sigma Aldrich) and 112 µl/l β -mercapto-ethanol (an oxidation scavenger that limits free radical damage, Sigma Aldrich). Cells were then lysed in TBS, pH 7.4, and 2 mM CaCl_2 by sonication using a Branson sonifier and centrifuged for 40 minutes at 18000 rpm and 4°C in order to separate the soluble from the insoluble fraction. The two cell lysates were purified at first by Ni-affinity chromatography. The soluble fraction of the cell lysate was incubated with 1 ml of Ni-NTA-agarose beads suspension (QIAGEN) for 1 hour on an orbital shaker, then it was flowed through a ten milliliters disposable column and washed with 10 ml of TBS plus 2 mM CaCl_2 , and then eluted in 5 fractions using a total of 8 ml of the same buffer supplemented with 0.4 M imidazole. The eluted fractions were then collected and further purified by gel filtration using a Sephacryl 100 HR HiPrep 26/60 size exclusion column (GE Healthcare).

Before the digestion, both proteins were concentrated to about 1.7 mg/ml and dialyzed in TBS buffer + 20 mM CaCl_2 using centrifugal filter units (Merck Millipore). The protein solutions were then supplemented with 1 µl/ ml of Enterokinase (New England Biolabs) and let react for 3-4 days at 25°C on an orbital shaker. The reaction mixtures were then purified by Ni-affinity chromatography to remove all traces of the cleaved 6His-tag and any residual uncleaved protein.

The two flow-through fractions were then collected and further purified by size exclusion chromatography with TBS + 1 mM CaCl_2 . Protein concentration was evaluated by Bradford assay⁴² using bovine serum albumin (Sigma) as the standard for building the calibration curve. Final sample purity was assessed by 10% SDS-PAGE.

Chapter 6. Experimental procedures

6.1. Heterologous protein expression

In the 70s, the advent of recombinant DNA technology revolutionized the development in biology opening many possibilities concerning the large-scale expression of heterologous proteins, thus overcoming all the limitations related to the extraction and purification of proteins from their natural sources. In addition to the large-scale production, DNA technology made possible also the introduction of specific mutations on gene fragments or the production of tagged or chimeric proteins. As a result of these innovations, a wide range of therapeutic products began to be produced by modifying microorganisms, animals, and plants to yield medically useful substances.

Modern recombinant DNA technology uses four main protein expression systems: bacteria, yeast, insect or mammalian systems. Their choice depends on the characteristics of the protein that is to be produced. The most widely used expression systems is the bacterial one, in particular the *E. coli* system. Its popularity relates to its ease of manipulation; it is easily cultured, grows fast in inexpensive media and produces high yields of recombinant protein. However, since bacterial cells are not equipped to accomplish particular folding architectures and post-translational modifications such as disulfide bond formation between two neighboring cysteines, or the glycosylation or phosphorylation reactions that are often necessary for the correct activity of a protein, many eukaryotic proteins need to be expressed in more complex systems.

Since the human O-phosphoethanolamine-phospho lyase (hETNPPL) and the E-Cadherin fragments produced in this thesis do not need any post translational modification to explain their activity, the expression host used for their large-scale production is *E. coli*.

In particular, in order to replicate the plasmids assembled in the laboratory, I used the two *E. coli* strains, XL1B and DH5 α (Invitrogen). These particular strains prevent the expression of endonucleases, a feature that is crucial when PCR-derived DNA is used. Indeed, unlike the genetic material of bacterial cells, PCR-derived DNA does not contain methylations at specific sites, which is why it would be recognized by the host organism as foreign and degraded by endonucleases.

In addition to these cells, I also used the Rosetta and BL21(DE3) *E. coli* strains to express the proteins of interest. These two strains, lacking *lon* and *ompT* proteases, are lysogens of λ DE3 meaning that they carry the gene for T7 RNA polymerase under control of the lacUV5 promoter. They are commonly used for production of recombinant proteins from genes cloned in particular vectors, named pET vectors. In these systems, protein expression is induced by adding Isopropyl- β -thio-D-galactopyranoside (IPTG) to the cell medium. Since is possible to adjust levels of protein expression by

changing the concentration of IPTG, this feature allows to enhance the proper folding and therefore the solubility and activity of difficult target proteins.

Apart from the common features, Rosetta strain, which is a BL21 *lacZY* derivative, was chosen because it allows to enhance the expression of eukaryotic proteins featuring codons that are rarely used in *E. coli* such as AGG, AGA, AUA, CUA, CCC, GGA.

As previously mentioned, in addition to the host strain, the other fundamental component of an expression systems is the expression vector that is needed to introduce the gene encoding the protein of interest into the host. The expression vectors used in this thesis, which are called pET plasmids, are the simplest and most frequently chosen vectors for bacterial systems. They can be selected on the basis of many characteristic features such as, copy number, restriction sites, tags or fusion proteins encoded, selection mechanism, and promoter region.

pET vectors are characterized by the presence of the *lacI* gene encoding for the lac repressor (LacI) and the T7 promoter region, which is located upstream of the lac operator and the lac repressor sequences and controls the gene transcription. The lac repressor protein (LacI) is sensitive to the presence of lactose. In the absence of this disaccharide, the lac repressor binds to the operator sequence on DNA, blocking access of T7 RNA polymerase to the promoter site and thus preventing transcription of the gene encoding the target protein.

When lactose binds to LacI, it induces conformational changes in the protein structure that greatly reduce its affinity for DNA thus allowing the transcription of the gene encoding the protein of interest. Experimentally, protein expression is induced by adding to the cell media IPTG, which is a structural analog of lactose. IPTG acts like lactose, but unlike this molecule is not part of any metabolic pathways, therefore it will not be recognized and hydrolyzed by the cell.

Another important feature of the pET vectors used in this thesis is the presence of the gene encoding for resistance to ampicillin. This gene acts as a selectable marker conferring to those cells that have been successfully transformed with the vector the ability to grow in a medium supplemented with ampicillin that would otherwise be toxic.

6.2. Chromatography techniques

Chromatography is an analytical technique that is commonly used for separating mixtures of chemical substances into their individual components. There is a great variety of chromatography techniques, such as liquid chromatography (encompassing size-exclusion chromatography, ion-exchange chromatography and affinity chromatography) and gas chromatography. Despite this variety, separation is always attained based on differential interactions between a mobile and a stationary phase. The main components and of a typical chromatographic system are illustrated in **Figure 47**.

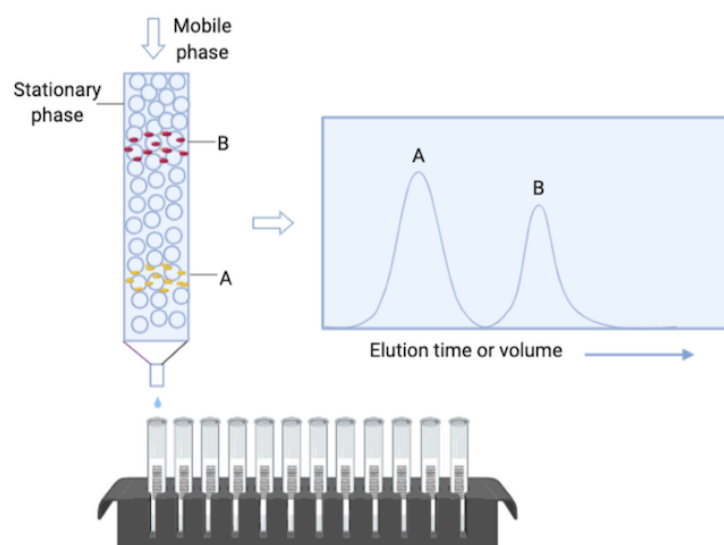


Figure 47. The general components of a chromatographic system, as illustrated here by using a column to separate two chemicals, A and B.

The stationary phase is fixed within the system by a support. Upon sample application, the mobile phase travels through the stationary phase carrying sample components with it. The components that have the strongest interactions with the stationary phase will be retained for a longer time into the column than components that have weaker interactions with the stationary phase. The difference in the rate of travel into the chromatographic system leads to the final separation of the different chemical components of a mixture.

Affinity chromatography is one of the most powerful chromatographic techniques, often used in molecular biology laboratories as a first purification step because of its good selectivity and high capacity. Affinity chromatography exploits the affinity of the target protein for the stationary phase, which is usually made of polymeric gel functionalized with a molecule or coordinated metal ions,

such as Ni^{2+} or Co^{2+} ions. In this work, I used a metal affinity chromatography called NiNTA chromatography because Ni^{2+} ions are immobilized onto polymeric resin beads through a nitrilotriacetic (NTA) linker. Since the high affinity of Ni^{2+} ions for histidine residues, the 6 His-tag fused to the proteins at the cloning stage allows them to anchor to the stationary phase (**Figure 48**). Owing to the reversible nature of this interaction, the immobilization step is generally followed by an intermediate washing step and by an elution step. The latter involves flowing into the column an imidazole solution that, through a competitive reaction with the histidine residues, releases the tagged protein from the Ni^{2+} beads, thus allowing it to be recovered in the elution fractions.

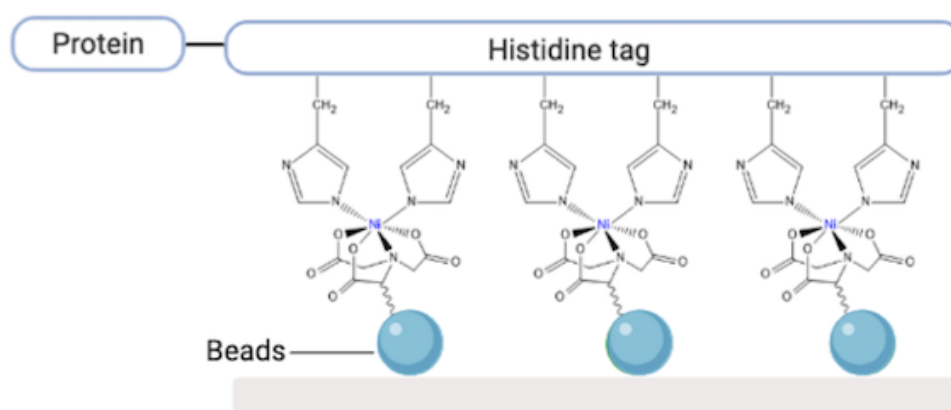


Figure 48. Binding of the Histidine tag to bivalent metal ions. Two neighboring histidines from the His-tagged protein form a complex with the bivalent metal ions.

Size-exclusion chromatography (SEC) also known as gel filtration (GF), is the simplest and mildest chromatography method used for protein purification. SEC allows the components of a mixture to be physically discriminated by their size as they pass through a column packed with a resin. The resin consists of a porous matrix of spherical particles. As a sample travels down the column, molecules larger than the pores flow through it without diffusing into the beads, while the smallest molecules remain trapped in them for a longer time. Hence, the smaller the molecules the longer is their elution time.

Ion exchange chromatography is also frequently used as chromatographic method in biochemistry. It enables the separation of ionizable molecules based on their total charge. In this technique, a protein mixture is loaded onto a column charged with either a positively or a negatively stationary phase. The column is usually pre-equilibrated with a specific buffer in order to adjust pH and ionic strength

values to promote the binding of the target protein to the resin. Once the protein solution flows through the column, the target molecule binds to the resin by displacing the counter ions of the stationary phase. Protein elution can be achieved by gradually changing the pH or the ionic strength of the elution buffer. This allows the proteins to be separated according to the strength of their binding.

6.3. Crystallization techniques

One of the main focuses of my PhD work has been the characterization of the 3D structure of the human O-phosphoethanolamine phospho-lyase (hETNPPL) by single crystal X-ray crystallography. This kind of analysis provides a correlation between amino-acidic composition and tridimensional folding of a macromolecule and its function. Moreover, it represents the starting point for the development of the hETNPPL-based biosensor that I described in Chapter 3.

X-ray crystallography is currently the most favored technique for structure determination of biological macromolecules. Indeed, it is instrumental for the understanding of molecular mechanisms as well as the specificity of any protein–ligand interaction. As such, it plays a crucial role also in structure-based drug design and in guiding site-directed mutagenesis studies.

The major bottleneck of X-ray crystallography consists in the production of crystals of suitable quality and size for X-ray diffraction. Indeed, countless crystallographic projects in laboratories around the world have failed because of problems occurring in the crystallization phase. The principle of protein crystallization is to bring the molecule solution to a supersaturation state in which the protein molecules can start packing together and form ordered tridimensional arrays, thus promoting nucleation and, if specific conditions are met, crystal growth.

Many variables can influence a crystallization process. Among them is the choice of precipitant, its concentration, the chemical nature of the buffer, its pH, the protein concentration, the temperature, the crystallization technique, and the possible inclusion of additives. Since no *a priori* knowledge of a successful crystallization condition for a specific protein is possible, initial experiments are usually based on a “trial and error” procedure. In order to cover the wider possible range of variables, commercially available “crystal screen” consisting of tens of different solutions varying widely in precipitant, buffer, pH, and salt, known as a sparse matrix are often used¹⁶⁹. To increase the chances of success and reduce systematic errors, crystallization robots are also used today to screen more than several thousands of parameters. In this way, it is often possible to obtain a first indication as to which conditions may need to be fine-tuned and further improved to move from initial microcrystals to good

quality crystals. Such improvement can be obtained by changing many different parameters, such as precipitant or protein concentrations, temperature, pH and others.

Crystallization experiments are usually set up using the vapor diffusion method (**Figure 49**) in which a drop, containing a 1:1 mix of unsaturated precipitant and protein solution, is placed in a sealed well containing a reservoir with precipitant in higher concentrations. The sealed environment allows water vapor to migrate from the droplet to the reservoir to reach the equilibrium.

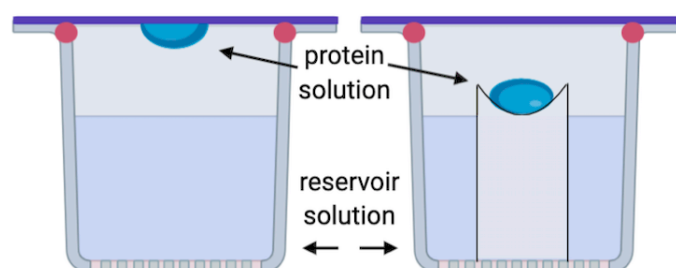


Figure 49. Schematic depiction of two common vapor diffusion techniques: hanging drop (left) and sitting drop (right). A drop with protein solution is situated either on top (hanging drop) or next to (sitting drop) a reservoir containing precipitant. As the water from the protein solution vaporizes, it transfers to the reservoir and crystallization occurs.

As water leaves the drop, the protein undergoes an increase in concentration. In this way the system moves from the stable area through the supersaturated region of the solubility diagram, where nucleation and eventually crystal formation can occur (**Figure 50**).

The vapor diffusion technique has been used for decades, mostly through the ‘hanging drop’ method. The sitting drop technique is an analogous method that offers the advantage that can be easily adapted to high-throughput technologies. Experimentally, both techniques involve the use of the same closed environment. The hanging drop method is set up with a drop of 1–10 μl suspended from a glass coverslip over the reservoir solution, while in sitting drop technique the drop is placed on a small bridge inserted into the reservoir.

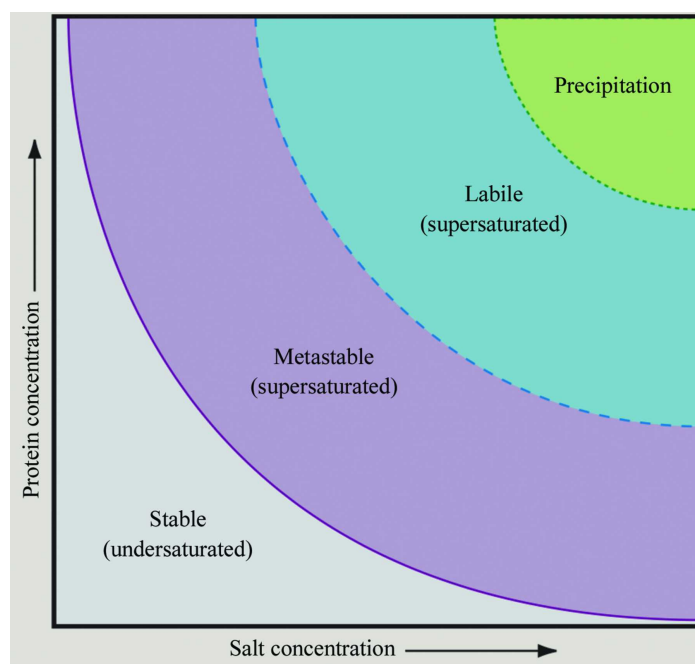


Figure 50. The phase diagram for the crystallization of macromolecules. The diagram can be divided into two main regions, the undersaturation and the supersaturation region. The solid line separating them denotes maximum solubility at different concentrations of a precipitant (salt or polymers). The region of supersaturation is further divided into the metastable and labile regions. The metastable region corresponds to the conditions at which nucleation can occur, while the labile region, albeit not suitable for the nucleation process, can still promote to crystal growth from preformed nuclei introduced from outside the system using the so-called seeding technique. Crystals can only be grown from a supersaturated solution. The final region, which features extreme supersaturation, is denoted the precipitation region, where the aggregation of the protein to form an amorphous precipitate is most probable. Adapted from *McPherson et al. (2014)*.

Once crystals of sufficient quality have been obtained, they are fished using a small loop and cryo-cooled to protect them from radiation damage¹⁷⁰. In order to prevent the formation of ice crystals caused by the flash cooling, which would interfere with the protein diffraction pattern, prior to freezing, crystals are soaked in a solution containing a cryoprotectant that allows the system to undergo a glassy transition. Cryoprotectant solution are usually prepared by supplemented the crystallization solution with the 20-30% of compounds such as glycerol, sucrose, 2-Methyl-2,4-pentanediol (MPD) or Polyethylene glycol (PEG) 400. Crystals are then extracted from the cryo-protectant solution using a fiber loop and mounted on a goniometric head for data collection.

6.4. X-ray diffraction

X-rays are electromagnetic waves with a wavelength ranging from 0.01 to 10 nm, discovered in 1895 by the German physicist W. C. Röntgen. The potential applications of X-rays came in 1912 from the hand of Max von Laue, who revealed the wave nature of X-rays and from the subsequent developments in 1913 of the English physicist Sir W.H. Bragg and his son Sir W.L. Bragg, who showed that X-ray diffraction could be used for the determination of the atomic structure of matter. These findings paved the way for the X-ray crystallography. However, it was only 45 years later that this technique was used to determine the first protein structure. This was the structure of myoglobin, which gave the authors, Max Perutz and John Kendrew the Chemistry Nobel Prize in 1962. Since then, many other researchers providing informative protein crystallographic structures have been awarded the Nobel Prize.

X-rays are generated when accelerated electrons collide with the atoms of the electrode knocking an electron out of an inner shell, thus allowing electrons from higher energy states to fill the vacancy. This causes the emission of photons with precise energies determined by the electron energy levels involved in the transition. These photons constitute the X-ray radiation.

In X-ray crystallography, this radiation is used to determine the atomic and molecular structure of a crystal, in which the highly ordered structure causes a beam of incident X-rays to diffract in many specific directions.

By definition, a crystal is a regular array of atoms or molecules. The basic repeating unit from which the crystal is constructed is called the unit cell. Its linear and angular dimensions are indicated as a , b , and c , α , β , and γ . By applying a well-defined set of geometrical operations, the unit cell can be repeated in space to generate a univocal crystal lattice whose origin is arbitrary chosen. However, since there are many choices of repeating blocks in any given lattice, crystallographers have chosen the following criteria for selecting unit cells. By convention, the unit cell edges are chosen to be right-handed, to have the highest possible symmetry that is compatible with the arrangements of the molecules in the crystal, and to have the smallest cell volume. If other symmetry considerations do not override, then the cell is chosen so that $a \leq b \leq c$, and α , β , and γ all $< 90^\circ$ or all $\geq 90^\circ$. This type of cell is referred to as the reduced cell.

By combining all possible symmetry elements, it is possible to generate 230 space groups, but not all of these are permitted for proteins because of the chirality of amino acids. Indeed, some of them such as inversion centers, mirror planes and glide planes are incompatible as they would reverse the chirality of molecules. Therefore, the number of the possible space groups is reduced to 65.

Space groups and crystals are divided into seven crystal systems according to their point groups, and into seven lattice systems according to their Bravais lattice (centered lattices). The 7 crystal systems are listed in **Table 11**.

Table 11. Crystal systems.

Sr. No.	Crystal system	Axial length of Unit Cell	Inter axial angles	Numbers of Lattice in the system
1	Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$	3
2	Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$	2
3	Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$	4
4	Monoclinic	$a \neq b \neq c$	$\alpha = \beta = 90^\circ \neq \gamma$	2
5	Triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90^\circ$	1
6	Trigonal	$a = b = c$	$\alpha = \beta = \gamma < 120^\circ \neq 90^\circ$	1
7	Hexagonal	$a = b \neq c$	$\alpha = \beta = 90^\circ, \text{ and } \gamma = 120^\circ$	1

A key contribution in the development of X-ray crystallography was the work of the Braggs (father and son), who in 1913, also using experimental data produced by von Laue, realized that crystalline solids at certain specific wavelengths and incident angles, produce intense peaks of reflected X-ray radiation. By modeling the crystal as a set of discrete parallel planes separated by a constant distance d , W. L. Bragg proposed that the incident X-ray radiation would produce a diffraction peak (Bragg peak) if their reflections off the various planes interfered constructively.

In response to their discovery the two scientists provided also a mathematical description for the experimental observations by formulating the Bragg's law:

$$n\lambda = 2d \sin\theta$$

where n is an integer number, λ is the wavelength of the incident beam, d is the interplanar distance and θ is the incidence angle. The formula relates the spacing between each member of a family of

parallel lattice planes and the angles of incidence at which these planes produce the most intense reflections of X-rays.

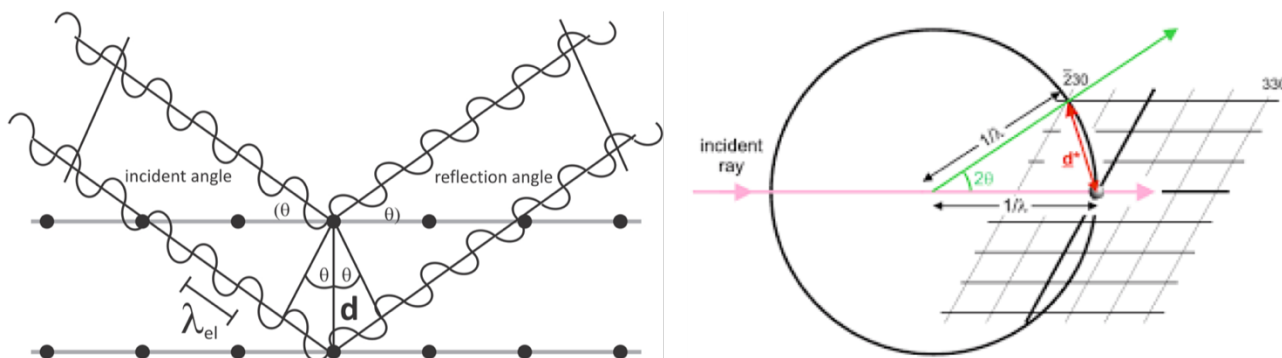


Figure 51. Schematic depiction of Bragg's law (left) and the Ewald sphere (right).

The properties of Bragg's law can be easily visualized by means of a geometrical construction named Ewald sphere (**Figure 51**), which uses of the reciprocal lattice concept to demonstrate the relationship between λ , θ and the crystal orientation. This geometric construction is a sphere of radius $1/\lambda$, with the crystal placed at the center. The origin of the reciprocal lattice lies in the transmitted beam, at the edge of the Ewald sphere. In order to have reflections (diffraction maxima) the Bragg equation in vector form, must be satisfied. This condition occurs whenever a reciprocal lattice point lies exactly on the Ewald sphere. Since this event is not likely, in order to move more reciprocal lattice points through the Ewald sphere and hence to acquire a complete set of data, it is necessary to move the crystal and the detector in the course of the diffraction experiment.

Every plane of the reciprocal space can be identified by a set of three indices (h, k, l). Experimentally, once a complete dataset is obtained, the first step of the processing consists in the assignment of a set of these indexes to all the reflections of the diffraction pattern by a process named indexing. At this stage, the cell parameters ($a, b, c, \alpha, \beta, \gamma$) and the space group of the crystal are determined.

A more complex process is needed to determine the shapes of the molecules in the crystal, which depends on a complex mathematical relationship between the intensities and the electron density map. Indeed, unlike the geometric parameters which only depend on the reciprocal disposition of the atoms in the unit cell, the intensity of the reflections depends both on the number and the nature of the atoms in the crystal lattice. Each atom of the crystallographic plane originating diffraction signals gives rise to a secondary wave. By summing the amplitude and phases of these waves, a complex number,

called the structure factor $F(hkl)$, can be obtained. This number is characterized by a modulus, which is directly proportional to the square root of the amplitude of the diffracted waves and a phase.

The structure factors are used to calculate the electron density map that describes the shape of the molecule being investigated.

Unfortunately, a direct reconstruction of the structural image of the molecule is hindered by the fact that experimentally it is only possible to measure the amplitude of the diffraction pattern spots, while the phase information is lost: this is the ‘phase problem’ of X-ray crystallography. Since without this information it is not possible to reconstruct the electron density in the unit cell, the problem must be overcome. Several approaches can be used to recover the lost phases.

6.5. Structure solution and refinement

Despite the variety of methods that can be employed to overcome the ‘phase problem’, all of them are affected by the use of essential simplifications. As a result, the obtained structure is only a model of the real structure adopted by the molecule in the crystal. In order to approximate as much as possible the real shape of the molecule of interest, refinement and optimization cycles need to be implemented in the overall structure determination process.

In small-molecule crystallography, some basic assumptions about atoms give rise to relationships between the amplitudes from which phase information can be extracted. In protein crystallography, these *ab initio* methods can only be used in very rare cases in which there are data at 1.2 Å or higher resolution. In the majority of cases, phases are derived either by finding the positions of heavy atoms that are present into the protein or that have been introduced (methods such as MIR, SIR, MAD, SAD) or by using the atomic coordinates of a structurally similar protein (molecular replacement).

The techniques exploiting heavy atoms are very useful to solve the crystal structure of proteins for which no prior information on their tridimensional fold is available. However, they require an extensive manipulation of the crystal and they often result in damaging the crystal itself, thus compromising the entire diffraction experiment. Moreover, these techniques often require additional steps to introduce specific elements or residues (such as, for instance, cysteines or seleno-methionines) to the protein sequence. Molecular replacement (MR), does not present these disadvantages; its only limitation is that it can be used only when a model homologous to the protein of interest, is available. However, since over the last 40 years, the number of proteins that have been structurally characterized has increased exponentially, the possibility that an unknown structure has some features in common with a known protein is high. Indeed, to date more than 60% of deposited macromolecular structures have been solved by MR⁵⁴.

6.5.1 Structure solution methods

Single and multiple isomorphous replacement (SIR and MIR) are normally used when no closely related structure is available and requires at least two data sets: one native set from the protein crystal and at least one derivative set from the protein crystal containing heavy atoms. The most important feature of this technique is the presence of heavy metal ions, which are usually incorporated in the protein molecule by soaking the crystal in a solution containing a heavy metal salt without perturbing its structure: that is, the two structures must be ‘isomorphous’.

The heavy atoms introduced in the crystal will generate a set of reflections that are more intense than those of other atoms forming the molecule. When compared, the differences between the data sets of native crystal and a crystal containing the heavy metal ions give information about the position of the heavy atoms that have been incorporated in the structure. Computer refinement of the heavy atom parameters is carried out and these parameters can be used as a starting point to determine the protein phase angles. Together with the amplitudes, these can be used to calculate the structure factors, which can then be subjected to rounds of refinement, allowing complete structure determination.

Single/Multiple Wavelength Anomalous Dispersion (SAD/MAD) have become a popular alternative to multiple isomorphous replacement experiments following the introduction of tunable X-ray beamlines at synchrotrons¹⁷¹. These methods are based on the dependence of the atomic scattering on the X-ray wavelength used in the diffraction experiment. This method can exploit heavy atoms but under certain conditions, it can also work elements such as sulphur and selenium. Depending on the atoms whose anomalous behavior is exploited, different wavelengths are used to collect the dataset. The signal obtained is then used to calculate their structure factors and, finally to obtain whole structure.

Molecular replacement (MR) is the method of choice when the molecule under study is similar to another molecule (the model) whose structure has been previously determined. As sequence similarity often correlates with structure similarity, the conventional approach to finding structurally similar models consists in using the amino acid sequence of the target to search against those of a set of known structures.

MR consists of placing the homologous structure in the cell of the investigated crystal using translation and rotation functions and comparing the calculated structure factors and the observed values from the diffraction. The agreement between them is used to identify the orientation of the model that most closely matches that of the unknown molecule in the crystal. The same process is then repeated, placing the model at every possible position in the unit cell to identify the correct translation. Once

the correct orientation and translation has been identified, the model can be used to calculate phases for all the structure factors. These phases are an excellent approximation of the phases of the measured signal and are used together with the experimental intensities to obtain a first electron density map. This map is the starting point for the refinement process, which aims at minimizing the difference between observed and calculated amplitudes.

The quality of the fit of a model to the diffraction data is assessed using many indicators. Among them, the most important is the R value, which measures the discrepancy between the observed (F_o) and calculated (F_c) structure-factor amplitudes:

$$R = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$$

Depending on the resolution and quality of the diffraction data, well-refined structures have R-factors below 20–25 percent. However, a low R value does not necessarily correspond to a good model. Indeed, by increasing the number of adjustable parameters used to describe the model, this R value can be lowered. Several cases are known in which models with serious errors were refined to ‘acceptable’ R values^{172–174}. To overcome this problem, in 1992, Axel T. Brunger introduced a method of statistical cross-validation^{173,175}. The method consists in excluding a randomly chosen subset of reflections from refinement and using these reflections only for the calculation of a “free” R-factor using the same formula given above. If refinement is progressing correctly, the free R-factor will drop below 30 percent.

6.6. X-ray sources

A typical in-house X-ray diffractometer is composed by a sealed tube containing a cathode that emits electrons by thermionic effect. These electrons are then accelerated towards a target by applying a voltage. As a result of the collision, some electrons located in the core levels of the target material, usually K or L shells, are dislodged and, electrons set at higher levels fill the generated vacancies. During this electronic transition a certain amount of energy is released. This energy constitutes the X-ray radiation produced by the diffractometer.

Generally, the intensity of the X-ray radiation that is generated with an in-house diffractometer is too weak to provide structural information about macromolecules. In this case, synchrotrons represent the X-ray source of choice. In synchrotrons, electrons circulate under vacuum into large circular machines at a speed nearing that of light. The storage rings are formed by a combination of curved bending magnets and straight sections called insertion devices. The monochromatic radiation emitted

in the insertion devices, which is very intense and intrinsically collimated normally is the one used in protein crystal diffraction experiments.

During my PhD, data collections were performed at the X06DA- PXIII beamline of the Swiss Light Source (SLS) synchrotron located at the Paul Scherrer Institute in Villigen (Switzerland, **Figure 52**). The specific beamline I used is the X06DA-PXIII. This beamline features a focused beam of 80 microns x 45 microns (h x v), with the possibility of fine-tuning the wavelength at any value in the range 2.07 to 0.71 Å.



Figure 52. Swiss Light Source (SLS) synchrotron of the Paul Scherrer Institute in Villigen (Switzerland). Pictures of SLS synchrotron ring (left) and X06DA-PXIII beamline (right).

6.7. Immobilization methods

The chosen immobilization method for the bio-recognition element is one of the main factors affecting the good performance of a biosensor. In order to obtain high sensitivity, stability, reproducibility and short response time, the native structure of the enzyme as well as its biological activity and its orientation should be preserved. There are a few immobilization techniques: adsorption, entrapment, cross-linking and covalent attachment (**Figure 53**). While adsorption and entrapment are physical methods, cross-linking and covalent bonding can be considered as chemical methods of immobilization¹⁷⁶.

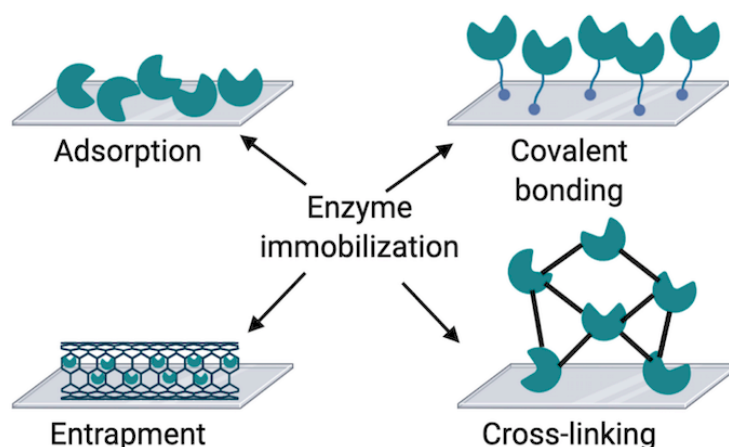


Figure 53. Enzyme immobilization methods.

Physical Immobilization Direct adsorption of enzymes on solid supports is the easiest and cheapest immobilization method used in the enzyme-based biosensor construction. The adsorption is usually done by placing a fixed amount of enzyme in contact with solid surfaces and allowing it to dry for a fixed period of time. After incubation, washing steps are done to remove unadsorbed enzyme molecules. In order to prevent loss of enzyme activity, biocompatible materials and solutions are typically employed. However, since adsorption mechanisms are based on weak bonds such as electrostatic, hydrophobic interactions and van der Waal's forces, changes in temperature, pH or ionic strength may result in enzyme desorption/leaching¹⁷⁷. Moreover, since immobilized enzymes do not usually possess a homogeneous orientation, it is possible to partially or completely block the access to the active sites. Moreover, non-specific adsorption of contaminants on the surface can also occur, thus interfering with the signal.

In order to prevent desorption, the physical entrapment of enzymes within a polymeric network could be performed. The general procedure for this immobilization method consists at first in the mixing of enzymes into a monomer solution and then in the polymerization of the monomer solution using different methods. Depending on the type of entrapment, a number of procedures such as electropolymerization, photopolymerization, sol-gel and others can be used. Overall, physical entrapment confers a good stability to enzymes and allows control of the pH, the polarity or the amphiphilicity of the microenvironment. Nevertheless, this technique can diminish or even block the active sites needed for the immunorecognition event.

Chemical Immobilization A valid alternative to physical methods consists in using chemical immobilization, which offers the advantage of controlling the orientation of the immobilized species through specific groups that are present in the biomolecules. However, since these methods involve

the use of chemical reactions to attach bio-recognition molecules on solid supports, denaturation problems can occur.

Cross-linking is an irreversible immobilization method that is obtained by adding bifunctional agents, such as for instance glutaraldehyde, to the enzymes. Immobilization occurs through the formation of intermolecular cross-links between the enzyme molecules by covalent bonds. Among the advantages of this method, there is the possibility to optimize the microenvironment by adding stabilizing agents and the minimization of enzyme leakage. However, a limitation of the method is the possibility for enzymes to undergo conformational changes and lose activity due to the use of glutaraldehyde.

Covalent attachment is one of the most widely used immobilization method, whereby surfaces are activated using specific reagents, such as glutaraldehyde or carbodiimide, in order to bind functional groups that are present on enzyme covalently. The enzyme functional groups that are typically utilized for covalent attachment include: amino group, carboxylic group, phenolic group, sulfhydryl group, thiol group, imidazole group, indole group and hydroxyl group¹⁷⁸. This type of immobilization prevents enzyme leakage from the support and provides high uniformity and good control of the amount of immobilized enzyme. However, owing to the chemical modifications thus introduced, there is a high risk of enzyme denaturation.

References:

1. John, R. A. Pyridoxal phosphate-dependent enzymes. *Biochim. Biophys. Acta (BBA)/Protein Struct. Mol.* **1248**, 81–96 (1995).
2. Mehta, P. K. & Christen, P. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Adv. Enzymol. Relat. Areas Mol. Biol.* **74**, 129–184 (2000).
3. Percudani, R. & Peracchi, A. A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep.* **4**, 850–854 (2003).
4. Eliot, A. C. & Kirsch, J. F. Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations. *Annu. Rev. Biochem.* **73**, 383–415 (2004).
5. Toney, M. D. Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochim. Biophys. Acta - Proteins Proteomics* **1814**, 1407–1418 (2011).
6. Fernandes, H. S., Ramos, M. J. & Cerqueira, N. M. F. S. A. Catalytic Mechanism of the Serine Hydroxymethyltransferase: A Computational ONIOM QM/MM Study. *ACS Catal.* **8**, 10096–10110 (2018).
7. Fernandes, H. S., Ramos, M. J. & Cerqueira, N. M. F. S. A. The Catalytic Mechanism of the Pyridoxal-5'-phosphate-Dependent Enzyme, Histidine Decarboxylase: A Computational Study. *Chem. - A Eur. J.* **23**, 9162–9173 (2017).
8. Roncador, A., Oppici, E., Montioli, R., Maset, F. & Cellini, B. TAT-Mediated Delivery of Human Alanine:Glyoxylate Aminotransferase in a Cellular Model of Primary Hyperoxaluria Type I. *Int. J. Pept. Res. Ther.* (2013). doi:10.1007/s10989-012-9333-9
9. Ding, Q. *et al.* AGXT2L1 is down-regulated in hepatocellular carcinoma and associated with abnormal lipogenesis. *J. Clin. Pathol.* (2016). doi:10.1136/jclinpath-2015-203042
10. McQuillin, A., Rizig, M. & Gurling, H. M. D. A microarray gene expression study of the molecular pharmacology of lithium carbonate on mouse brain mRNA to understand the neurobiology of mood stabilization and treatment of bipolar affective disorder. *Pharmacogenet. Genomics* **17**, 605–617 (2007).
11. Stankiewicz, A. M. *et al.* Social stress increases expression of hemoglobin genes in mouse prefrontal cortex. *BMC Neurosci.* **15**, (2014).
12. Schirotti, D., Cirrincione, S., Donini, S. & Peracchi, A. Strict reaction and substrate specificity

- of AGXT2L1, the human O-phosphoethanolamine phospho-lyase. *IUBMB Life* **65**, 645–650 (2013).
13. Cellini, B. Biochemical and Computational Approaches to Improve the Clinical Treatment of Dopa Decarboxylase-Related Diseases: An Overview. *Open Biochem. J.* **6**, 131–138 (2012).
 14. Frey, P. A. Radical Mechanisms of Enzymatic Catalysis. *Annu. Rev. Biochem.* **70**, 121–148 (2001).
 15. Livanova, N. B., Chebotareva, N. A., Eronina, T. B. & Kurganov, B. I. Review: Pyridoxal 5'-phosphate as a catalytic and conformational cofactor of muscle glycogen phosphorylase b. *Biokhimiya* **67**, 1317–1327 (2002).
 16. Helmreich, E. J. M. How pyridoxal 5'-phosphate could function in glycogen phosphorylase catalysis. *BioFactors* **3**, 159–172 (1992).
 17. Beattie, A. E. *et al.* Reconstitution of the pyridoxal 5'-phosphate (PLP) dependent enzyme serine palmitoyltransferase (SPT) with pyridoxal reveals a crucial role for the phosphate during catalysis. *Chem. Commun.* **49**, 7058–7060 (2013).
 18. Grishin, N. V., Phillips, M. A. & Goldsmith, E. J. Modeling of the spatial structure of eukaryotic ornithine decarboxylases. *Protein Sci.* **4**, 1291–1304 (1995).
 19. Percudani, R. & Peracchi, A. The B6 database: A tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families. *BMC Bioinformatics* **10**, 273 (2009).
 20. Jansonius, J. N. Structure, evolution and action of vitamin B6-dependent enzymes. *Curr. Opin. Struct. Biol.* **8**, 759–769 (1998).
 21. Schneider, G., Käck, H. & Lindqvist, Y. The manifold of vitamin B6 dependent enzymes. *Structure* **8**, (2000).
 22. Rocha, J. F., Pina, A. F., Sousa, S. F. & Cerqueira, N. M. F. S. A. PLP-dependent enzymes as important biocatalysts for the pharmaceutical, chemical and food industries: A structural and mechanistic perspective. *Catal. Sci. Technol.* **9**, 4864–4876 (2019).
 23. MEHTA, P. K., HALE, T. I. & CHRISTEN, P. Aminotransferases: demonstration of homology and division into evolutionary subgroups. *Eur. J. Biochem.* **214**, 549–561 (1993).
 24. Steffen-Munsberg, F. *et al.* Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications. *Biotechnol. Adv.* **33**, 566–604 (2015).

25. Kappes, B., Tews, I., Binter, A. & MacHeroux, P. PLP-dependent enzymes as potential drug targets for protozoan diseases. *Biochim. Biophys. Acta - Proteins Proteomics* **1814**, 1567–1576 (2011).
26. Christen, P. & Mehta, P. K. From cofactor to enzymes. The molecular evolution of pyridoxal-s'-phosphate-dependent enzymes. *Chem. Rec.* **1**, 436–447 (2001).
27. Jensen, R. A. & Gu, W. Evolutionary recruitment of biochemically specialized subdivisions of family I within the protein superfamily of aminotransferases. *J. Bacteriol.* **178**, 2161–2171 (1996).
28. Rausch, C., Lerchner, A., Schiefner, A. & Skerra, A. Crystal structure of the ω-aminotransferase from *Paracoccus denitrificans* and its phylogenetic relationship with other class III amino-transferases that have biotechnological potential. *Proteins Struct. Funct. Bioinforma.* **81**, 774–787 (2013).
29. Veiga-da-Cunha, M., Hadi, F., Balligand, T., Stroobant, V. & Van Schaftingen, E. Molecular identification of hydroxylysine kinase and of ammoniophospholyases acting on 5-phosphohydroxy-L-lysine and phosphoethanolamine. *J. Biol. Chem.* **287**, 7246–7255 (2012).
30. Sprinson, D. B. & Weliky, I. The conversion of ethanolamine to acetate in mammalian tissues. *Biochem. Biophys. Res. Commun.* **36**, 866–870 (1969).
31. Fleshood, H. L. & Pitot, H. C. The metabolism of O-phosphorylethanolamine in animal tissues. II. Metabolic regulation of O-phosphorylethanolamine phospholyase in vivo. *Arch. Biochem. Biophys.* **141**, 423–429 (1970).
32. Schirotli, D., Ronda, L. & Peracchi, A. Kinetic characterization of the human O-phosphoethanolamine phospho-lyase reveals unconventional features of this specialized pyridoxal phosphate-dependent lyase. *FEBS J.* **282**, 183–199 (2015).
33. Bender, D. Transaminases: Edited by P. Christen and D.E. Metzler Wiley-Interscience; New York, 1985 643 pages. £158.00. *FEBS Lett.* **190**, 179–180 (1985).
34. Yard, B. A. *et al.* The Structure of Serine Palmitoyltransferase; Gateway to Sphingolipid Biosynthesis. *J. Mol. Biol.* **370**, 870–886 (2007).
35. Kontani, Y., Kaneko, M., Kikugawa, M., Fujimoto, S. & Tamaki, N. Identity of D-3-aminoisobutyrate-pyruvate aminotransferase with alanine-glyoxylate aminotransferase 2. *BBA - Gen. Subj.* **1156**, 161–166 (1993).
36. Cooper, A. J. L., Krasnikov, B. F., Okuno, E. & Jeitner, T. M. L-Alanine-glyoxylate

aminotransferase II of rat kidney and liver mitochondria possesses cysteine S-conjugate β -lyase activity: A contributing factor to the nephrotoxicity/hepatotoxicity of halogenated alkenes? *Biochem. J.* **376**, 169–178 (2003).

37. Donini, S. *et al.* Recombinant production of eight human cytosolic aminotransferases and assessment of their potential involvement in glyoxylate metabolism. *Biochem. J.* (2009). doi:10.1042/BJ20090748
38. EUGENE P. KENNEDY AND SAMUEL B. WEISS. In the eugene p. kennedy. *J Biol Chem* **222**, 193–214 (1956).
39. Menendez, J. A. & Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat. Rev. Cancer* **7**, 763–777 (2007).
40. Mounier, C., Bouraoui, L. & Rassart, E. Lipogenesis in cancer progression (review). *Int. J. Oncol.* **45**, 485–492 (2014).
41. Shao, L. & Vawter, M. P. Shared Gene Expression Alterations in Schizophrenia and Bipolar Disorder. *Biol. Psychiatry* **64**, 89–97 (2008).
42. Hammond, J. B. W. & Kruger, N. J. The Bradford Method for Protein Quantitation BT - New Protein Techniques. 25–32 (1988). doi:10.1385/0-89603-126-8:25
43. Cuetos, A. *et al.* Structural Basis for Phospholyase Activity of a Class III Transaminase Homologue. *ChemBioChem* **17**, 2308–2311 (2016).
44. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 271–281 (2011).
45. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763 (1994).
46. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).
47. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
48. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
49. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 61–69 (2007).

50. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
51. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).
52. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
53. Vaguine, A. A., Richelle, J. & Wodak, S. J. SFCHECK: A unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 191–205 (1999).
54. Berman, H. M. The Protein Data Bank / Biopython. *Presentation* **28**, 235–242 (2000).
55. Notredame, C., Higgins, D. G. & Heringa, J. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
56. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, (2014).
57. No Title. doi:10.1093/protein/8.2.127
58. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).
59. Thorne, H. *et al.* Decreased Prostate Cancer-Specific Survival of Men with BRCA 2 Mutations from Multiple Breast Cancer Families. *Cancer Prev. Res.* **4**, 1002–1010 (2011).
60. Lowe, F. C. & Trauzzi, S. J. Prostatic acid phosphatase in 1993: Its limited clinical utility. *Urol. Clin. North Am.* **20**, 589–595 (1993).
61. Balk, S. P., Ko, Y. J. & Bubley, G. J. Biology of prostate-specific antigen. *J. Clin. Oncol.* **21**, 383–391 (2003).
62. Catalona, W. J. *et al.* Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N. Engl. J. Med.* **324**, 1156–1161 (1991).
63. Stamey, T. A. *et al.* Prostate-Specific Antigen as a Serum Marker for Adenocarcinoma of the Prostate. *N. Engl. J. Med.* **317**, 909–916 (1987).
64. Velonas, V., Woo, H., Remedios, C. & Assinder, S. Current Status of Biomarkers for Prostate Cancer. *Int. J. Mol. Sci.* **14**, 11034–11060 (2013).
65. Pavlou, M. & Diamandis, E. P. The Search for New Prostate Cancer Biomarkers Continues.

Clinical Chemistry **55**, 1277–1279 (2009).

66. Bohm, L. *et al.* Plasma sarcosine does not distinguish early and advanced stages of prostate cancer. *South African Med. J.* **102**, 677 (2012).
67. Potosky, A. L., Feuer, E. J. & Levin, D. L. Impact of screening on incidence and mortality of prostate cancer in the United States. *Epidemiol. Rev.* **23**, 181–186 (2001).
68. Trock, B. J. Application of metabolomics to prostate cancer. *Urol. Oncol. Semin. Orig. Investig.* **29**, 572–581 (2011).
69. Kim, Y. & Kislinger, T. Novel approaches for the identification of biomarkers of aggressive prostate cancer. *Genome Med.* **5**, (2013).
70. Miyagi, Y. *et al.* Plasma free amino acid profiling of five types of cancer patients and its application for early detection. *PLoS One* **6**, (2011).
71. Rigau, M. *et al.* The present and future of prostate cancer urine biomarkers. *Int. J. Mol. Sci.* **14**, 12620–12649 (2013).
72. Shamsipur, M., Naseri, M. T. & Babri, M. Quantification of candidate prostate cancer metabolite biomarkers in urine using dispersive derivatization liquid-liquid microextraction followed by gas and liquid chromatography-mass spectrometry. *J. Pharm. Biomed. Anal.* **81–82**, 65–75 (2013).
73. Heger, Z. *et al.* Determination of common urine substances as an assay for improving prostate carcinoma diagnostics. *Oncol. Rep.* **31**, 1846–1854 (2014).
74. Dereziński, P., Klupczynska, A., Sawicki, W., Pałka, J. A. & Kokot, Z. J. Amino acid profiles of serum and urine in search for prostate cancer biomarkers: A pilot study. *Int. J. Med. Sci.* (2017). doi:10.7150/ijms.15783
75. Swanson, M. G. *et al.* Quantification of choline- and ethanolamine-containing metabolites in human prostate tissues using ¹H HR-MAS total correlation spectroscopy. *Magn. Reson. Med.* **60**, 33–40 (2008).
76. Leach, M. O. *et al.* Measurements of human breast cancer using magnetic resonance spectroscopy: A review of clinical measurements and a report of localized ³¹P measurements of response to treatment. *NMR Biomed.* **11**, 314–340 (1998).
77. Touhami, A. Biosensors and Nanobiosensors: Design and Applications. *Nanomedicine* 374–400 (2015).

78. Chaubey, A. & Malhotra, B. D. Review Mediated biosensors. *Biosens. Bioelectron.* **7**, 441–456 (2002).
79. Nguyen, H. H., Lee, S. H., Lee, U. J., Fermin, C. D. & Kim, M. Immobilized enzymes in biosensor applications. *Materials (Basel)*. **12**, 1–34 (2019).
80. Grieshaber, D., MacKenzie, R., Vörös, J. & Reimhult, E. Electrochemical Biosensors - Sensor Principles and Architectures. *Sensors* **8**, 1400–1458 (2008).
81. Patel, P. D. (Bio)sensors for measurement of analytes implicated in food safety: A review. *TrAC - Trends Anal. Chem.* **21**, 96–115 (2002).
82. Sun, G. *et al.* Chemical Sensors. **705**, 13 (1988).
83. Wilson, J. *Sensor Technology Hand book*. (1999).
84. Ding, C., Zhao, F., Zhang, M. & Zhang, S. Hybridization biosensor using 2,9-dimethyl-1,10-phenantroline cobalt as electrochemical indicator for detection of hepatitis B virus DNA. *Bioelectrochemistry* **72**, 28–33 (2008).
85. Caygill, R. L. *et al.* Novel impedimetric immunosensor for the detection and quantitation of Adenovirus using reduced antibody fragments immobilized onto a conducting copolymer surface. *Biosens. Bioelectron.* **32**, 104–110 (2012).
86. Conzuelo, F., Gamella, M., Campuzano, S., Reviejo, A. J. & Pingarrón, J. M. Disposable amperometric magneto-immunosensor for direct detection of tetracyclines antibiotics residues in milk. *Anal. Chim. Acta* **737**, 29–36 (2012).
87. Moina, C. & Ybarr, G. Fundamentals and Applications of Immunosensors. *Adv. Immunoass. Technol.* (2012). doi:10.5772/36947
88. D’Orazio, P. Biosensors in clinical chemistry. *Clin. Chim. Acta* **334**, 41–69 (2003).
89. Cöpel, W. Biosensoren: Biosensors: Theory and Applications. Von D. G. Buerk. Technomic Publishing Company, Lancaster, 1993. 219 S., geb., SFr. 187,-. ISBN 0-87762-975-7. . *Nachrichten aus Chemie, Tech. und Lab.* **41**, 1156–1157 (1993).
90. Neumega (IL-11) label. *FDA label* **112**, 211–212 (1966).
91. Bollella, P. & Gorton, L. Enzyme based amperometric biosensors. *Curr. Opin. Electrochem.* **10**, 157–173 (2018).
92. Wang, D., Noël, V. & Piro, B. Electrolytic gated organic field-effect transistors for application in biosensors—A review. *Electron.* **5**, (2016).

93. Kergoat, L., Piro, B., Berggren, M., Horowitz, G. & Pham, M. C. Advances in organic transistor-based biosensors: From organic electrochemical transistors to electrolyte-gated organic field-effect transistors. *Anal. Bioanal. Chem.* **402**, 1813–1826 (2012).
94. Stern, E. *et al.* Label-free immunodetection with CMOS-compatible semiconducting nanowires. *Nature* **445**, 519–522 (2007).
95. Stern, E. *et al.* Importance of the debye screening length on nanowire field effect transistor sensors. *Nano Lett.* **7**, 3405–3409 (2007).
96. Huang, L. & Cheng, Z. M. Immobilization of lipase on chemically modified bimodal ceramic foams for olive oil hydrolysis. *Chem. Eng. J.* **144**, 103–109 (2008).
97. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
98. Baenke, F., Peck, B., Miess, H. & Schulze, A. Hooked on fat: The role of lipid synthesis in cancer metabolism and tumour development. *DMM Dis. Model. Mech.* **6**, 1353–1363 (2013).
99. Sunami, Y., Rebelo, A. & Kleeff, J. Lipid metabolism and lipid droplets in pancreatic cancer and stellate cells. *Cancers (Basel)*. **10**, (2018).
100. Ameer, F., Scanduzzi, L., Hasnain, S., Kalbacher, H. & Zaidi, N. De novo lipogenesis in health and disease. *Metabolism: Clinical and Experimental* (2014). doi:10.1016/j.metabol.2014.04.003
101. Koizume, S. & Miyagi, Y. Lipid droplets: A key cellular organelle associated with cancer cell survival under normoxia and hypoxia. *Int. J. Mol. Sci.* **17**, (2016).
102. Fujimoto, T. & Parton, R. G. Not just fat: The structure and function of the lipid droplet. *Cold Spring Harb. Perspect. Biol.* **3**, 1–17 (2011).
103. Rysman, E. *et al.* De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer Res.* **70**, 8117–8126 (2010).
104. Fittipaldi, A. *et al.* Cell Membrane Lipid Rafts Mediate Caveolar Endocytosis of HIV-1 Tat Fusion Proteins. *J. Biol. Chem.* **278**, 34141–34149 (2003).
105. van den Berg, A. & Dowdy, S. F. Protein transduction domain delivery of therapeutic macromolecules. *Curr. Opin. Biotechnol.* **22**, 888–893 (2011).
106. Guidotti, G., Brambilla, L. & Rossi, D. Cell-Penetrating Peptides: From Basic Research to

Clinics. Trends Pharmacol. Sci. **38**, 406–424 (2017).

107. Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **127**, 2893–2917 (2010).
108. Thomas, M. B. *et al.* Hepatocellular carcinoma: Consensus Recommendations of the National Cancer Institute Clinical Trials Planning Meeting. *J. Clin. Oncol.* **28**, 3994–4005 (2010).
109. Arao, T. *et al.* FGF3/FGF4 amplification and multiple lung metastases in responders to sorafenib in hepatocellular carcinoma. *Hepatology* **57**, 1407–1415 (2013).
110. Llovet, J. M. *et al.* Sorafenib in advanced hepatocellular carcinoma. *N. Engl. J. Med.* **359**, 378–390 (2008).
111. Cabibbo, G. *et al.* A meta-analysis of survival rates of untreated patients in randomized clinical trials of hepatocellular carcinoma. *Hepatology* **51**, 1274–1283 (2010).
112. Calvisi, D. F. *et al.* Increased lipogenesis, induced by AKT-mTORC1-RPS6 signaling, promotes development of human hepatocellular carcinoma. *Gastroenterology* **140**, 1071–1083.e5 (2011).
113. Yamashita, T. *et al.* Activation of lipogenic pathway correlates with cell proliferation and poor prognosis in hepatocellular carcinoma. *J. Hepatol.* **50**, 100–110 (2009).
114. Yahagi, N. *et al.* Co-ordinate activation of lipogenic enzymes in hepatocellular carcinoma. *Eur. J. Cancer* **41**, 1316–1322 (2005).
115. Li, C. *et al.* SREBP-1 has a prognostic role and contributes to invasion and metastasis in human hepatocellular carcinoma. *Int. J. Mol. Sci.* **15**, 7124–7138 (2014).
116. Nelson, M. E. *et al.* Inhibition of hepatic lipogenesis enhances liver tumorigenesis by increasing antioxidant defence and promoting cell survival. *Nat. Commun.* **8**, (2017).
117. Fittipaldi, A. & Giacca, M. Transcellular protein transduction using the Tat protein of HIV-1. *Adv. Drug Deliv. Rev.* **57**, 597–608 (2005).
118. Roncador, A., Oppici, E., Montioli, R., Maset, F. & Cellini, B. TAT-mediated delivery of human alanine: Glyoxylate aminotransferase in a cellular model of primary hyperoxaluria type i. *Int. J. Pept. Res. Ther.* **19**, 175–184 (2013).
119. Rapoport, M., Salman, L., Sabag, O., Patel, M. S. & Lorberboum-Galski, H. Successful TAT-mediated enzyme replacement therapy in a mouse model of mitochondrial E3 deficiency. *J. Mol. Med.* **89**, 161–170 (2011).

120. Vyas, P. M. *et al.* A TAT-frataxin fusion protein increases lifespan and cardiac function in a conditional Friedreich's ataxia mouse model. *Hum. Mol. Genet.* **21**, 1230–1247 (2012).
121. Toro, A. & Grunebaum, E. TAT-mediated intracellular delivery of purine nucleoside phosphorylase corrects its deficiency in mice. *J. Clin. Invest.* **116**, 2717–2726 (2006).
122. Toro, A., Paiva, M., Ackerley, C. & Grunebaum, E. Intracellular delivery of purine nucleoside phosphorylase (PNP) fused to protein transduction domain corrects PNP deficiency in vitro. *Cell. Immunol.* **240**, 107–115 (2006).
123. Leckband, D. & Sivasankar, S. Cadherin recognition and adhesion. *Curr. Opin. Cell Biol.* **24**, 620–627 (2012).
124. Gumbiner, B. M. Regulation of cadherin-mediated adhesion in morphogenesis. *Nat. Rev. Mol. Cell Biol.* **6**, 622–634 (2005).
125. Leckband, D. E. & de Rooij, J. Cadherin Adhesion and Mechanotransduction. *Annu. Rev. Cell Dev. Biol.* **30**, 291–315 (2014).
126. Shapiro, L. & Weis, W. I. Structure and biochemistry of cadherins and catenins. *Cold Spring Harb. Perspect. Biol.* **1**, (2009).
127. Sivasankar, S. Tuning the kinetics of cadherin adhesion. *J. Invest. Dermatol.* **133**, 2318–2323 (2013).
128. Brasch, J., Harrison, O. J., Honig, B. & Shapiro, L. Thinking outside the cell: How cadherins drive adhesion. *Trends Cell Biol.* **22**, 299–310 (2012).
129. Schuh, R. *et al.* Molecular cloning of the mouse cell adhesion molecule uvomorulin: cDNA contains a B1-related sequence. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 1364–1368 (1986).
130. Van Roy, F. & Berx, G. The cell-cell adhesion molecule E-cadherin. *Cell. Mol. Life Sci.* **65**, 3756–3788 (2008).
131. Nardone, V. *et al.* Crystal Structure of Human E-Cadherin-EC1EC2 in Complex with a Peptidomimetic Competitive Inhibitor of Cadherin Homophilic Interaction. *J. Med. Chem.* **59**, 5089–5094 (2016).
132. Nollet, F., Kools, P. & Van Roy, F. Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members. *J. Mol. Biol.* **299**, 551–572 (2000).
133. Shapiro, L. & Weis, W. I. Structure and biochemistry of cadherins and catenins. *Cold Spring*

Harb. Perspect. Biol. **1**, (2009).

134. Huber, A. H. & Weis, W. I. The structure of the β -catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by β -catenin. *Cell* **105**, 391–402 (2001).
135. Ciatto, C. *et al.* T-cadherin structures reveal a novel adhesive binding mechanism. *Nat. Struct. Mol. Biol.* **17**, 339–347 (2010).
136. Parisini, E., Higgins, J. M. G., Liu, J. huan, Brenner, M. B. & Wang, J. huai. The Crystal Structure of Human E-cadherin Domains 1 and 2, and Comparison with other Cadherins in the Context of Adhesion Mechanism. *J. Mol. Biol.* **373**, 401–411 (2007).
137. Shapiro, L., Kwong, P. D., Fannon, A. M., Colman, D. R. & Hendrickson, W. A. Considerations on the folding topology and evolutionary origin of cadherin domains. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6793–6797 (1995).
138. Nagar, B., Overduin, M., Ikura, M. & Rini, J. M. Structural basis of calcium-induced E-cadherin rigidification and dimerization. *Nature* **380**, 360–364 (1996).
139. Boggon, T. J. *et al.* C-cadherin ectodomain structure and implications for cell adhesion mechanisms. *Science* (80-.). **296**, 1308–1313 (2002).
140. Pertz, O. A new crystal structure, Ca^{2+} dependence and mutational analysis reveal molecular details of E-cadherin homoassociation. *EMBO J.* **18**, 1738–1747 (1999).
141. Parisini, E., Higgins, J. M. G., Liu, J. huan, Brenner, M. B. & Wang, J. huai. The Crystal Structure of Human E-cadherin Domains 1 and 2, and Comparison with other Cadherins in the Context of Adhesion Mechanism. *J. Mol. Biol.* **373**, 401–411 (2007).
142. Patel, S. D. *et al.* Type II Cadherin Ectodomain Structures: Implications for Classical Cadherin Specificity. *Cell* **124**, 1255–1268 (2006).
143. Li, Y. *et al.* Mechanism of E-cadherin dimerization probed by NMR relaxation dispersion. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16462–16467 (2013).
144. Vedove, A. D. *et al.* Structure-based virtual screening allows the identification of efficient modulators of E-cadherin-mediated cell-cell adhesion. *Int. J. Mol. Sci.* **20**, (2019).
145. Harrison, O. J. *et al.* The extracellular architecture of adherens junctions revealed by crystal structures of type i cadherins. *Structure* **19**, 244–256 (2011).
146. Berx, G. & van Roy, F. Involvement of members of the cadherin superfamily in cancer. *Cold Spring Harb. Perspect. Biol.* **1**, (2009).

147. De Santis, G., Miotti, S., Mazzi, M., Canevari, S. & Tomassetti, A. E-cadherin directly contributes to PI3K/AKT activation by engaging the PI3K-p85 regulatory subunit to adherens junctions of ovarian carcinoma cells. *Oncogene* **28**, 1206–1217 (2009).
148. Zhang, C. C. *et al.* PF-03732010: A fully human monoclonal antibody against P-cadherin with antitumor and antimetastatic activity. *Clin. Cancer Res.* **16**, 5177–5188 (2010).
149. Taniuchi, K. *et al.* Overexpressed P-cadherin/CDH3 promotes motility of pancreatic cancer cells by interacting with p120ctn and activating Rho-family GTPases. *Cancer Res.* **65**, 3092–3099 (2005).
150. Hazan, R. B., Qiao, R., Keren, R., Badano, I. & Suyama, K. Cadherin switch in tumor progression. *Ann. N. Y. Acad. Sci.* **1014**, 155–163 (2004).
151. Altucci, L. *et al.* Rexinoid-triggered differentiation and tumor-selective apoptosis of acute myeloid leukemia by protein kinase A-mediated desubordination of retinoid X receptor. *Cancer Res.* **65**, 8754–8765 (2005).
152. Smacchia, M. P. *et al.* A case of GH deficiency and beta-thalassemia. *Minerva Endocrinol.* **37**, 201–209 (2012).
153. Paredes, J. *et al.* Epithelial E- and P-cadherins: Role and clinical significance in cancer. *Biochim. Biophys. Acta - Rev. Cancer* **1826**, 297–311 (2012).
154. Lee, D. M. *et al.* Cadherin-11 in synovial lining formation and pathology in arthritis. *Science (80-.).* **315**, 1006–1010 (2007).
155. Blaschuk, O. W. & Rowlands, T. M. Cadherins as modulators of angiogenesis and the structural integrity of blood vessels. *Cancer Metastasis Rev.* **19**, 1–5 (2000).
156. Vasudev, N. S. & Reynolds, A. R. Anti-angiogenic therapy for cancer: Current progress, unresolved questions and future directions. *Angiogenesis* **17**, 471–494 (2014).
157. Yarom, N. *et al.* Phase I Clinical Trial of Exherin (ADH-1) in Patients with Advanced Solid Tumors. *Curr. Clin. Pharmacol.* **8**, 81–88 (2013).
158. Perotti, A. *et al.* Clinical and pharmacological phase I evaluation of ExherinTM (ADH-1), a selective anti-N-cadherin peptide in patients with N-cadherin-expressing solid tumours. *Ann. Oncol.* **20**, 741–745 (2009).
159. Blaschuk, O. W. Discovery and development of N-cadherin antagonists. *Cell Tissue Res.* **348**, 309–313 (2012).

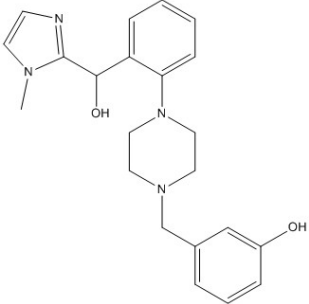
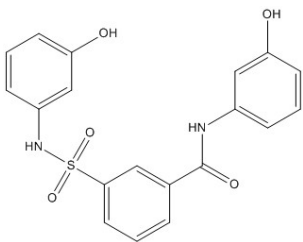
160. Doro, F. *et al.* Computational design of novel peptidomimetic inhibitors of cadherin homophilic interactions. *Org. Biomol. Chem.* **13**, 2570–2573 (2015).
161. Mayer, M. & Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* **123**, 6108–6117 (2001).
162. Meyer, B. & Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew. Chemie - Int. Ed.* **42**, 864–890 (2003).
163. Vasile, F., Rossi, D., Collina, S. & Potenza, D. Diffusion-ordered spectroscopy and saturation transfer difference NMR spectroscopy studies of selective interactions between ELAV protein fragments and an mRNA target. *European J. Org. Chem.* **2014**, 6399–6404 (2014).
164. Vasile, F. *et al.* NMR interaction studies of Neu5Ac- α -(2,6)-Gal- β -(1-4)-GlcNAc with influenza-virus hemagglutinin expressed in transfected human cells. *Glycobiology* **28**, 42–49 (2018).
165. Potenza, D. *et al.* A NMR and computational study of Smac mimics targeting both the BIR2 and BIR3 domains in XIAP protein. *Org. Biomol. Chem.* **10**, 3278–3287 (2012).
166. Civera, M. *et al.* Exploring E-cadherin-peptidomimetics interaction using NMR and computational studies. *PLoS Comput. Biol.* **15**, (2019).
167. Harrison, O. J. *et al.* Two-step adhesive binding by classical cadherins. *Nat. Struct. Mol. Biol.* **17**, 348–357 (2010).
168. Siret, C. *et al.* Cadherin-1 and cadherin-3 cooperation determines the aggressiveness of pancreatic ductal adenocarcinoma. *Br. J. Cancer* **118**, 546–557 (2018).
169. Jancarik, J. & Kim, S. H. Sparse matrix sampling. A screening method for crystallization of proteins. *J. Appl. Crystallogr.* **24**, 409–411 (1991).
170. Garman, E. F. & Schneider, T. R. Macromolecular Cryocrystallography. *J. Appl. Crystallogr.* **30**, 211–237 (1997).
171. Hendrickson, W. A. & Ogata, C. M. Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* **276**, 494–523 (1997).
172. Brändén, C. I. & Alwyn Jones, T. Between objectivity and subjectivity. *Nature* **343**, 687–689 (1990).
173. Brünger, A. T. Free R value: A novel statistical quantity for assessing the accuracy of crystal

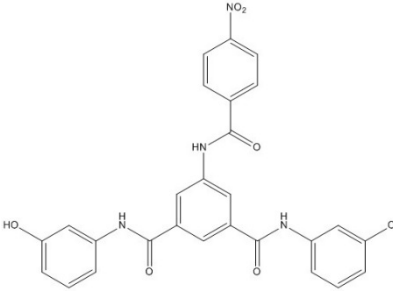
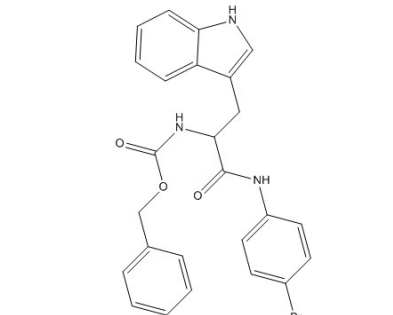
structures. *Nature* **355**, 472–475 (1992).

174. Kleywegt, G. J. & Jones, T. A. Model building and refinement practice. *Methods Enzymol.* **277**, 208–230 (1997).
175. Brünger, A. T. Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **49**, 24–36 (1993).
176. Magalhães, J. M. C. S. & Machado, A. A. S. C. Urea potentiometric biosensor based on urease immobilized on chitosan membranes. *Talanta* **47**, 183–191 (1998).
177. Mohamad, N. R., Marzuki, N. H. C., Buang, N. A., Huyop, F. & Wahab, R. A. An overview of technologies for immobilization of enzymes and surface analysis techniques for immobilized enzymes. *Biotechnol. Biotechnol. Equip.* **29**, 205–220 (2015).
178. Microbial Enzymes and Biotransformations. *Microb. Enzym. Biotransformations* (2005). doi:10.1385/1592598463

Appendix I

Table 12. List of the tested compounds.

Company (ID#)	Internal code	Chemical structure
Asinex (AEM14687298)	AS2	
Asinex (BAS00093476)	AS8	

<p>Asinex (BAS00132635)</p>	<p>AS9</p>	
<p>Asinex (BAS00602705)</p>	<p>AS11</p>	
<p>Life Chemicals (F2762-0527)</p>	<p>LC11</p>	