

Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN
NANOSCIENZE PER LA MEDICINA E PER L'AMBIENTE

Ciclo 34

Settore Concorsuale: 03/B1 - FONDAMENTI DELLE SCIENZE CHIMICHE E SISTEMI INORGANICI

Settore Scientifico Disciplinare: CHIM/03 - CHIMICA GENERALE E INORGANICA

E-CADHERIN AND CHOLINE KINASE: TWO CHALLENGING DRUG
DISCOVERY TARGETS

Presentata da: Archimede Torretta

Coordinatore Dottorato

Dario Braga

Supervisore

Emilio Parisini

Co-supervisore

Dario Braga

Esame finale anno 2022

Alla mia sposa ed alla mia famiglia
À mon épouse et ma famille

Acknowledgments

First of all, I would like to thank from the bottom of my heart my PhD supervisor, Dr. Emilio Parisini, for the time and patience he dedicated to me and to my scientific and personal development. I am deeply grateful for his efforts in helping me writing this manuscript, advising, and mentoring me.

I would also thank Dr. Marina Veronesi, Prof. Frédéric André, Prof. Luisa Carlota López-Cara, and their research teams for their contributions to the research projects I am about to present and for their willingness to listening and discussing.

Finally, my thanks go to my spouse, family, and friends, who continuously supported me over these years.

Abstract

The data presented in this thesis was generated using molecular biology, protein chemistry and X-ray crystallography techniques. However, while the methodologies employed are essentially the same, the research work presented here refers to two different proteins, which are part of different research projects in the laboratory. For this reason, the content of this thesis is divided in two independent parts, each provided with an introduction and a general overview of the research topic and state-of-the-art, a materials and methods section discussing the techniques used and the protocols followed, and a section where the results are presented and discussed in detail.

The first half of the thesis deals with the structural characterization of the complex between human E-cadherin and three different small molecule potential inhibitors identified via a fragment-based drug discovery (FBDD) screening campaign that was conducted using a library of commercially available small fluorinated chemical fragments. For this screening phase, we used ^{19}F -NMR as readout. The NMR experiments were done by our collaborator Dr. Marina Veronesi at the D3 PharmaChemistry division of the Italian Institute of Technology (IIT) in Genova (Italy). Functional cell adhesion assays to validate the inhibitory effects of the fragments thus identified were carried out in collaboration with Prof. Frédéric André at the University of Marseille (France).

The second half of the thesis describes the structural characterization of *Plasmodium falciparum* Choline Kinase (PfChoK), an important pharmaceutical target in the fight against malaria, as well as the biochemical characterization of a library of potential inhibitors of PfChoK. These inhibitors were synthesized in the group of Prof. Luisa Carlota López-Cara at the Department of Pharmaceutical and Organic Chemistry of the University of Granada (Spain) in the framework of an ongoing collaboration between the two groups.

List of Published Papers

- F. Aguilar Troyano, **A. Torretta**, G. Rubbini, A. Fasiolo, P. Navarro, M. Carrasco, G. Pérez-Moreno, C. Bosch Navarrete, D. Gonzalez-Pacanowska, E. Parisini, L. López-Cara **“New Compounds with Bioisosteric Replacement of Classic Choline Kinase Inhibitors Show Potent Antiplasmodial Activity”** (2021) *Pharmaceutics*
- A. Gautieri, F. Rigoldi, **A. Torretta**, A. Redaelli, E. Parisini **“In silico engineering of enzyme access tunnels”** (2022) *Enzyme Engineering, Methods in Molecular Biology Series*, Vol. 2397
- C. Brullo, F. Rapetti, S. Abbate, T. Prosdocimi, **A. Torretta**, M. Semrau, M. Massa, S. Alfei, P. Storici, E. Parisini, O. Bruno **“Design, synthesis, biological evaluation and structural characterization of novel GEBR library PDE4D inhibitors”**, *Journal of European Medicinal Chemistry*, (2021)
- F. Rigoldi, S. Donini, **A. Torretta**, A. Carbone, A. Redaelli, T. Bandiera, E. Parisini, A. Gautieri **“Rational backbone redesign of fructosyl peptide oxydase to widen its active site access tunnel”** (2020) *Biotechnology and Bioengineering*
- **Torretta**, L. C. Lopez-Cara, E. Parisini **“Crystal Structure of the Apo and the ADP-Bound Form of Choline Kinase from Plasmodium falciparum”** (2020) *Crystals*
- V. Cavalloro, K. Russo, F: Vasile, L. Pignataro, **A. Torretta**, S. Donini, M. S. Semrau, P. Storici, D: Rossi, F: Rapetti, C: Brullo, E. Parisini, O. Bruno, S. Collina **“Insight into GEBR-32a: Chiral Resolution, Absolute Configuration and Enantioference in PDE4D Inhibition”** (2020) *Molecules*
- Dalle Vedove, F. Falchi, S. Donini, A: Dobric, S: Germain, GP. Martino, T. Prosdocimi, C: Vettraino, **A. Torretta**, A. Cavalli, V. Rigot, F. André, E. Parisini **“Structure-Based Virtual Screening Allows the Identifcation of Efcient Modulators of E-Cadherin-Mediated Cell-Cell Adhesion”** (2019) *International Journal of Molecular Sciences*

Index

ACKNOWLEDGMENTS	II
ABSTRACT	III
LIST OF PUBLISHED PAPERS	IV
INDEX OF ABBREVIATIONS	VII
INDEX OF FIGURES	IX
INDEX OF TABLES	XI
DRUGGING THE UNDRUGGABLE: TOWARD THE DEVELOPMENT OF SELECTIVE MODULATORS OF E-CADHERIN-MEDIATED CELL-CELL ADHESION	12
Introduction	13
The Cadherin Superfamily	14
The Major Cadherin Family	17
The Adherens Junction and E-cadherin	19
Structural Determinants of E-cadherin <i>Cis</i> and <i>Trans</i> Interactions	24
The E-cadherin Homo-dimerization Process	28
E-cadherin and the Cancer Phenotype	34
The Limitations of Conventional Therapies and the Possible Role of E-cadherin as a Pharmaceutical Target	36
Drug Discovery on E-cadherin	38
Aim of the Project	42
Materials and Methods	43
Cloning, Protein Expression and Purification of E-cadherin and V3-E-cadherin Constructs	43
NMR Fragment-Based Screening	44
Crystal structures of the V3-E-cadherin-small fragment complexes	44
X-Ray Diffraction Data Acquisition and Structure Determination	45
Cell Adhesion and Cell Invasion Assays	47
Results	48
E-cadherin and V3-E-cadherin Production and Purification	48
Hit Compounds from the ¹⁹ F NMR-based Screening	51
ARNs and V3-E-cadherin EC1-EC2 Co-Crystallization Experiments	55
Structure Determination of the V3-E-cadherin EC1-EC2:ARN complex	56
Cell Adhesion and Cell Invasion Assays	60
Discussion	65
Binding Mode of the ARN Compounds	65
Description of the ARN1512 Interaction Mode	69
Description of the ARN1577 Interaction Mode	71
Description of the ARN1883 Interaction Mode	73
Cell Adhesion and Cell Invasion Assays	75

Conclusions	77
BIOISOSTERIC REPLACEMENT OF CLASSICAL CHOLINE KINASE INHIBITORS: ANTITUMORAL AND ANTIPLASMODIAL ACTIVITY	78
Introduction	79
The Protein Kinase Superfamily	79
The Eukaryotic Protein Kinase Fold	82
The Structure of Choline Kinase	83
The Biological Significance of Human Choline Kinase	88
Implications of Choline Kinase in Pathological Conditions	92
Human Choline Kinase α Inhibition	96
The Biological Significance of <i>Plasmodium falciparum</i> Choline Kinase	105
Human Choline Kinase α Inhibitors Repurposed as <i>Plasmodium falciparum</i> Choline Kinase Inhibitors	107
Aim of the Project	108
Materials and Methods	109
Cloning, Protein Expression and Purification of human Choline Kinase	109
Cloning, Protein Expression and Purification of <i>Plasmodium falciparum</i> Choline Kinase	109
25 μ M Fractional Activity Screening	110
Enzymatic Assay for the IC ₅₀ Calculation	113
Enzymatic Assay for the Determination of the Inhibition Mode	113
Analytical Size-exclusion Chromatography	114
Crystallization	114
X-Ray Diffraction Data Acquisition and Structure Determination	115
Results	117
Human ChoK α and PfChoK Production and Purification	117
PfChoK Crystallization Experiments	120
PfChoK Oligomerization State	122
25 μ M Fractional Activity Screening	123
IC ₅₀ Determination	125
Determination of the Inhibition Mode	128
Discussion	134
Structure of the Apo and ADP-bound form of <i>P. falciparum</i> Choline Kinase	134
Antiplasmodial Activity and <i>Ex Vivo</i> Potency	142
Determination of the Inhibition Mode	145
Conclusions and Future Perspectives	146
BIBLIOGRAPHY	147

Index of Abbreviations

The list of abbreviation used in this work is listed in order of appearance.

<i>Abbreviation</i>	<i>In full</i>
<i>EC</i>	<i>Extracellular cadherin</i>
<i>CDH</i>	<i>Major cadherins</i>
<i>PCDH</i>	<i>Protocadherins</i>
<i>CDHR</i>	<i>Cadherin-related molecules</i>
<i>cPCDH</i>	<i>Clustered protocadherins</i>
<i>ncPCDH</i>	<i>Non-clustered protocadherins</i>
<i>Ksp</i>	<i>Kidney-specific</i>
<i>LI</i>	<i>Liver-intestine</i>
<i>EGF</i>	<i>Epidermal growth factor</i>
<i>HRM</i>	<i>Hormone receptor motif</i>
<i>G-protein-coupled receptor proteolytic site</i>	<i>GPS</i>
<i>EMT</i>	<i>Epithelial-to-mesenchymal transition</i>
<i>MET</i>	<i>Mesenchymal-to-epithelial transition</i>
<i>MDCK</i>	<i>Madin-Darby Canine Kidney</i>
<i>HAV motif</i>	<i>His79-Ala80-Val81 motif</i>
<i>PGT motif</i>	<i>Pro123-Gly124-Thr125 motif</i>
<i>V3-E-cadherin</i>	<i>E-cadherin construct lacking Asp1 and Trp2</i>
<i>OD₆₀₀</i>	<i>Optical density measured at 600 nm</i>
<i>IPTG</i>	<i>Isopropyl -β-D-1-thiogalactopyranoside</i>
<i>PMSF</i>	<i>Phenylmethylsulphonyl fluoride</i>
<i>Ni-NTA</i>	<i>Nickel nitroacetic acid</i>
<i>LEF</i>	<i>Local Environmental Fluorine</i>
<i>PKs</i>	<i>Protein Kinases</i>
<i>ePKs</i>	<i>eukaryotic Protein Kinases</i>
<i>aPKs</i>	<i>atypical Protein Kinases</i>

ChoK	Choline Kinase
PC	phosphatidylcholine
PE	phosphatidylethanolamine
Hc-3	Hemicholinium-3
<i>PfChoK</i>	<i>Plasmodium falciparum</i> Choline Kinase
<i>PEG</i>	<i>Polyethylene glycol</i>
<i>TCEP</i>	Tris(2-carboxyethyl)phosphine

Index of Figures

Figure 1: Extracellular cadherin domain	14
Figure 2: General scheme of a prototypical cadherin.....	15
Figure 3: Schematic representation of the main features of the major cadherin subfamilies.....	18
Figure 4: The adherens junction and the process of tumor formation.....	19
Figure 5: The structure of the adherens junction is regulated by the extracellular portion of classical cadherins.....	20
Figure 6: The effect of calcium binding	22
Figure 7: Structural determinants of the trans interaction of E-cadherin	25
Figure 8: Structural determinants of the cis interaction in E-cadherin	26
Figure 9: Structural determinants of the X-dimer	29
Figure 10: Detail of P-cadherin in closed conformation	31
Figure 11: Cadherin homodimerization model.....	32
Figure 12: FR159 V3-E-cadherin co-crystal structure	39
Figure 13: Chemical structures of AS8, AS9 and AS11	41
Figure 14: SDS-PAGE of the purification profile of E-cadherin EC1-2 after Ni-NTA affinity chromatography	48
Figure 15: Size-exclusion chromatogram and corresponding SDS-PAGE gel	49
Figure 16: SDS-PAGE profile of the enterokinase digested fraction.....	49
Figure 17: Summary SDS-PAGE gel showing E-cadherin EC1-2 purification profile	50
Figure 18: NMR spectra recorded in the presence of 25 compounds and of 1, 2, and 4 μ M wt E-cadherin EC1-EC2.....	51
Figure 19: Line broadening of the ^{19}F NMR signal.....	52
Figure 20: Line broadening of ARN1512.....	52
Figure 21: NMR dose-response curve of ARN1534	53
Figure 22: Crystal morphologies of V3-E-cadherin EC1-2 co-crystallized with the ARN compounds	55
Figure 23: Structural alignment between E-cadherin X-dimer and the three obtained co-crystal structures.....	56
Figure 24: Omit maps of ARN1577	57
Figure 25: Omit maps of ARN1512	57
Figure 26: : Omit maps of ARN1883	58
Figure 27: 2Fo-Fc electron density of the ARN ligands	58
Figure 28: Effect of ARN1512, ARN1883, ARN1577, ARN1879, and ARN2068 on cell-cell adhesion.....	61
Figure 29: ARN 1512 reduces cell-cell adhesion in a concentration-dependent fashion.....	61
Figure 30: Results of the RNA interference on BxPC-3 cells.....	62
Figure 31: Effects of ARN1512 and ARN1883 targeting either E-cadherin or P-cadherin.....	62
Figure 32: N-cadherin expression and its adhesive function in MiaPaca-2 cells	63
Figure 33: Effect of ARN1512 and ARN1883 on MiaPaca-2 N-cadherin cells	63
Figure 34: Effects of ARN1512 and ARN1883 on cell invasion.....	64
Figure 35: ARN1512 and AS11 target E-cadherin in an independent manner	64
Figure 36: Summary of the structural details of the three co-crystal structures.....	66
Figure 37: Structural superimposition between the three co-crystal structures, E-cadherin strand swap and X-dimer.....	67
Figure 38: Electrostatic surface calculated by Chimera of the empty and compound occupied pockets.....	68
Figure 39: Details of the interaction of ARN1512_A with E-cadherin acceptor pocket.....	70
Figure 40: Details of the interaction of ARN1577_Z with E-cadherin acceptor pocket.....	72
Figure 41: Details of the interaction of ARN1883 with E-cadherin acceptor pocket	74
Figure 42: Sequence differences between E- and N-cadherin.....	76
Figure 43: The Human Kinome.....	81
Figure 44: Structural features of the crystal structure of C. Elegans ChoK	84
Figure 45: Interactions of the ADP molecule with human ChoK	85
Figure 46: Interactions of the phosphorylcholine molecule with human ChoK.....	86

Figure 47: The Kennedy pathway	89
Figure 48: Key enzymes involved in the choline metabolism.....	93
Figure 49: Oncogenic signaling pathways and choline metabolism	93
Figure 50: Binding mode of Hc-3.....	97
Figure 51: Binding mode of TCD-717	100
Figure 52: Binding mode of compound 1.....	101
Figure 53: Binding mode of compound 2.....	102
Figure 54: Binding mode of compounds 5, 12 and 14	102
Figure 55: Binding mode of compound in PDBID 5FTG	104
Figure 56: SDS-PAGE of the purification profile of PfChoK after Ni-NTA affinity chromatography.....	117
Figure 57: Size-exclusion chromatogram and corresponding SDS-PAGE gel	118
Figure 58: Summary SDS-PAGE gel showing the PfChoK purification profile	119
Figure 59: Crystal morphologies of PfChoK apoenzyme and ADP-bound form.....	120
Figure 60: Overall fold of the PfChoK apoenzyme (PDBID: 6YXS) and the ADP-bound PfChoK enzyme (PDBID: 6YXT)	121
Figure 61: Results of the analytical size-exclusion chromatography	122
Figure 62: Percentual fractional activity at 25 μ M compound concentration	123
Figure 63: IC ₅₀ curves of all the tested compounds.....	127
Figure 64: IC ₅₀ curves of all the tested compounds calculated at a choline concentration equals to its Michaelis-Menten constant.....	129
Figure 65: Fractional activity calculated with a constant compound and one substrate concentrations, while varying the concentration of the other..	132
Figure 66: Sequence alignment of the full sequences of <i>P. falciparum</i> choline kinase (PfChoK), <i>C. Elegans</i> ChoK-A2 and human ChoK α 1	134
Figure 67: Two 180-degree rotated views of the PfChoK apoenzyme structure (PDB code: 6YXS)	135
Figure 68: Structural features of PfChoK.....	137
Figure 69: Magnesium ion coordination in the apoenzyme (PDB code: 6YXS) (A) and in the ADP-bound form (PDB code: 6YXT) (B).....	138
Figure 70: Stereodiagram of the ADP binding mode	138
Figure 71: Structural superimposition of ADP molecules and ADP molecules with magnesium ions	139
Figure 72: Structure alignment of the PfChoK (green cartoon) and the human ChoK α 1 (blue and orange cartoons) choline binding sites	140

Index of Tables

Table 1: Official nomenclature of the 114 human genes encoding cadherin molecules.....	16
Table 2: Cadherin switching during oncogenesis.....	36
Table 3: Data collection and refinement statistics for the three co-crystal structures.....	46
Table 4: Results of the ¹⁹ F NMR-based screening.....	54
Table 5: Michaelis-Menten constants for the two ChoK isoforms. Data obtained from the BRENDA database [168].....	88
Table 6: Scheme of the synthesized library.....	111
Table 7: Scheme of the synthesized diphenyl disulfide and 1,2-bisphenylthioethane linker compounds.....	112
Table 8: Data collection and refinement statistics for 6YXS and 6YXT.....	116
Table 9: Inhibitory effect of the compounds from the library, excluding the one featuring the sulfur-containing linker. N.d., not determined. The CK prefix is omitted for brevity.....	124
Table 10: Summary of the IC ₅₀ values calculated from the curves displayed in Figure 47.....	127
Table 11: Summary of the K _i and α parameters calculated from the inhibition curves in Figure 64 and Figure 65.....	133

*DRUGGING THE UNDRUGGABLE:
TOWARD THE DEVELOPMENT OF SELECTIVE
MODULATORS OF E-CADHERIN-MEDIATED
CELL-CELL ADHESION*

Introduction

In the course of the evolution from single cell to multicellular organisms, the appearance of adhesion molecules marked a fundamental step toward the formation of organized multicellular structures, held together by cell-cell and cell-matrix contacts. In any newborn mammal, the critical role of adhesion molecules and cell-cell junctions becomes apparent from the very first stages of development. Indeed, the formation of specialized cell-cell junctions occurs during the first rounds of cleavage of embryonic stem cells and contributes to the formation of the blastocyst, which will later develop into entire complex organisms.

Adherens junctions allow the formation of a compact two-dimensional layer of cells to form tissues that not only act as protective barriers against the environment, but also affect the spatial distribution of biologically active molecules (nutrients, signaling molecules, growth factors...) and allow the mechanical coupling between cells. The ability of cells to distribute mechanical stresses makes them capable of responding to external stimuli while maintaining their functionality.

Cell adhesion molecules are crucially involved in tissue organization, development, homeostasis, and re-modelling. Hence, their altered expression levels or their functional impairment is often associated with pathological conditions [1].

Four different protein families are classified as adhesion molecules, each having a different function: immunoglobulins mediate cell-cell recognition in the immune system cells; integrins mediate cell-matrix interactions; selectins are involved in the recognition and binding of sugar moieties; cadherins mediate cell-cell adhesion and the formation of adherens junctions [2].

The focus of this thesis is on the cadherin superfamily, with particular reference to its prototypic member E-cadherin.

The Cadherin Superfamily

Cadherins are a large superfamily of calcium-dependent membrane glycoproteins that are involved in cellular adhesion and recognition. Each member of the superfamily is characterized by the tandem arrangement of a variable number of extracellular cadherin (EC) domains and by the conserved DRE, DXNDN, and DXD calcium-binding motifs in their sequence. All cadherins are composed of an extracellular portion, also referred to as the ectodomain, a transmembrane region and a cytoplasmatic domain, the latter featuring specific binding regions for adaptor proteins such as α - and β -catenin, which are responsible for linking cadherins on the surface of the cell to the actin cytoskeleton. The domain topology of an EC repeat is showed in Figure 1. It is an immunoglobulin-like fold consisting of a “Greek-key” β sandwich motif formed by seven β strands arranged in a circular fashion and a characteristic β helix between strand C and D. Overall, each EC repeat forms a globular domain and three calcium ions are found at the level of the linker region within each EC tandem repeat (Figure 2).

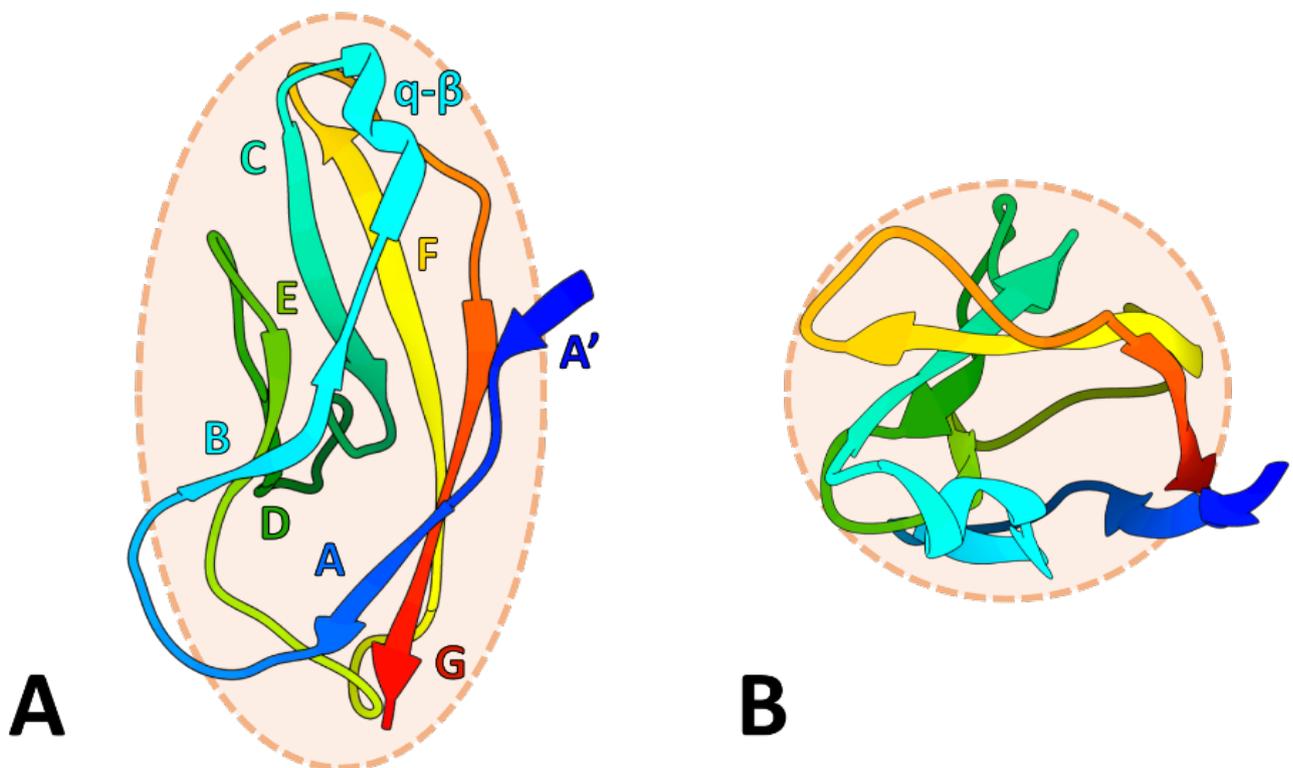


Figure 1: Extracellular cadherin domain

Panel A depicts the side view of the EC domain, while panel B its top view. The domain is drawn in cartoon representation colored in rainbow with a blue N-terminal and a red C-terminal. Each strand is labelled with an alphabetical letter. The label “q-β” indicates the β helix featured between strand C and D. The coordinates used for the images are taken from the crystal structure of E-cadherin (PDBID: 2O72).

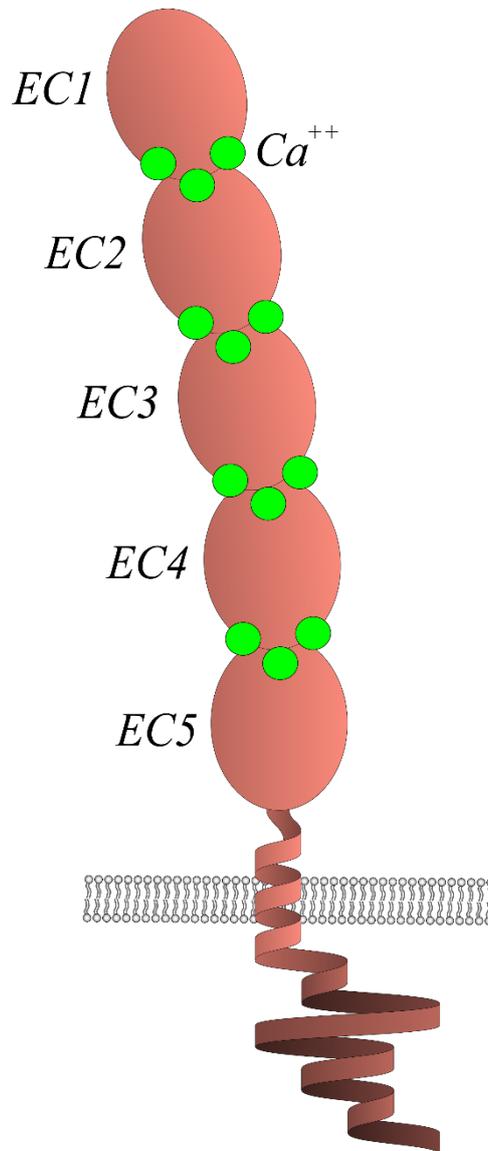


Figure 2: General scheme of a prototypical cadherin

The drawing shows the overall structure of a prototypical cadherin molecule, featuring several EC domains (five in the classical cadherins). Three calcium ions are found at the interface in each tandem repeat (green dots).

So far, 114 different human genes have been classified as cadherin-encoding genes (Table 1). These genes have been further divided into three families based on their function, structure and evolutionary history: major cadherins (CDH), protocadherins (PCDH) and cadherin-related molecules (CDHR) [3], [4]. In the following, while I will discuss the CDH family in detail, I will only briefly introduce the PCDH and CDHR families, as these latter ones are not the focus of my thesis.

Table 1: Official nomenclature of the 114 human genes encoding cadherin molecules. Adapted from [3].

Family	Subfamily	Official gene names (HGNC symbol, common alias)
Major cadherins (CDH) 32 members	Type I classical cadherins	cadherin 1 (CDH1 , E-cadherin), cadherin 2 (CDH2 , N-cadherin), cadherin 3 (CDH3 , P-cadherin), cadherin 4 (CDH4 , R-cadherin), cadherin 15 (CDH15 , M-cadherin)
	Type II classical cadherins	cadherin 5 (CDH5 , VE-cadherin), cadherin 6 (CDH6 , K-cadherin), cadherin 7 (CDH7), cadherin 8 (CDH8), cadherin 9 (CDH9 , T1-cadherin), cadherin 10 (CDH10 , T2-cadherin), cadherin 11 (CDH11 , OB-cadherin), cadherin 12 (CDH12 , N-cadherin 2), cadherin 18 (CDH18), cadherin 19 (CDH19), cadherin 20 (CDH20), cadherin 22 (CDH22), cadherin 24 (CDH24)
	7D cadherins	cadherin 16 (CDH16 , Ksp-cadherin), cadherin 17 (CDH17 , LI-cadherin)
	Desmosomal cadherins	desmocollin 1 (DSC1), desmocollin 2 (DSC2), desmocollin 3 (DSC3), desmoglein 1 (DSG1), desmoglein 2 (DSG2), desmoglein 3 (DSG3), desmoglein 4 (DSG4)
	Flamingo or CELSR cadherins	cadherin EGF LAG seven-pass G-type receptor 1 (CELSR1), cadherin EGF LAG seven-pass G-type receptor 2 (CELSR2), cadherin EGF LAG seven-pass G-type receptor 3 (CELSR3)
Protocadherins (PCDH) 65 members	Clustered protocadherins	protocadherin alpha cluster (PCDHA@: PCDHAC1 , PCDHAC2 , PCDHA1 up to PCDHA14), protocadherin beta cluster (PCDHB@: PCDHB1 up to PCDHB16), protocadherin gamma cluster (PCDHG@: PCDHGA1 up to PCDHGA12 , PCDHGB1 up to PCDHGB7 , PCDHGC3 , PCDHGC4 , PCDHGC5)
	Non-clustered protocadherins	protocadherin 1 (PCDH1), protocadherin 7 (PCDH7), protocadherin 8 (PCDH8), protocadherin 9 (PCDH9), protocadherin 10 (PCDH10), protocadherin 11 X-linked and Y-linked (PCDH11X and PCDH11Y), protocadherin 12 (PCDH12), protocadherin 17 (PCDH17), protocadherin 18 (PCDH18), protocadherin 19 (PCDH19), protocadherin 20 (PCDH20)
Cadherin-related (CDHR) 17 members	–	cadherin related family member 1 (CDHR1 , PCDH21), cadherin related family member 2 (CDHR2 , PCDH24), cadherin related family member 3 (CDHR3 , CDH28), cadherin related family member 4 (CDHR4 , CDH29), cadherin related family member 5 (CDHR5 , MU-PCDH), protocadherin related 15 (PCDH15 , CDHR15), ret proto-oncogene (RET , CDHR16), cadherin related 23 (CDH23 , CDHR23)
	Dachsous	dachsous cadherin-related 1 (DCHS1 , CDHR6), dachsous cadherin-related 2 (DCHS2 , CDHR7)
	FAT	FAT atypical cadherin 1 (FAT1 , CDHR8), FAT atypical cadherin 2 (FAT2 , CDHR9), FAT atypical cadherin 3 (FAT3 , CDHR10), FAT atypical cadherin 4 (FAT4 , CDHR11)
	Calsyntenins	calsyntenin 1 (CLSTN1 , CDHR12), calsyntenin 2 (CLSTN2 , CDHR13), calsyntenin 3 (CLSTN3 , CDHR14)

PCDH are mainly expressed in the central nervous system of vertebrates. They were identified by Sano et al. in 1993 and they feature either six or seven EC repeats, a transmembrane domain and a cytoplasmic tail [5]. In 1999, Wu et al. discovered that tandemly arranged gene arrays of PCDH were organized as three clusters on human chromosome 5q31 and named the three clusters as α -, β - and γ -PCDH [6]. Later, this subfamily was called clustered PCDH (cPCDH) to distinguish their members from the other family members, which are called non-clustered PCDH (ncPCDH) because their coding sequences were found to be distributed throughout the human genome [7].

PCDH expression is important for the development of the central nervous system. The diverse expression of different cPCDH isoforms contributes to the creation of diversity on neuronal cell membranes, hence to specificity, creating functional diversification between neurons. Furthermore, cPCDH are involved in dendrite and axon directional growth, synapse formation and neuronal survival. NcPCDH have a region-dependent expression profile contributing to the specification of a neuron identity and, like cPCDH, play a role in dendritic, axonal and synaptic development [8]- [9]. The CDHR family gathers the most varied members of the cadherin superfamily and many of their functions are still unclear. For example, it is not known whether calsyntenins, which are composed of only two EC repeats, mediate adhesion or are secreted ligands or chaperones. Others, like Dachsous and FAT, on the other hand, are among the longest cadherins known to date as they feature up to 34 EC repeats. They are known to regulate planar cell polarity and proliferation [3].

The Major Cadherin Family

The CDH family is divided into type I and type II classical cadherins, 7D cadherins, desmosomal cadherins, Flamingo or CELSR cadherins, and some solitary members such as cadherin 23 (H-cadherin). With the exception of the 7D and Flamingo cadherins, the other subfamilies possess five EC repeats. Conventionally, they are numbered from the outmost (EC1) to the membrane proximal (EC5) domain. The two members of the 7D cadherin subfamily, cadherin 16 (kidney-specific (Ksp) cadherin) and cadherin 17 (liver-intestine (LI) cadherin), originated after chromosomal duplication during vertebrate evolution and are expressed in kidney and intestine epithelia, where they mediate cell adhesion [10]. Interestingly, unlike classical cadherins, the two 7D cadherin members do not form complexes with catenins because they display shorter cytoplasmic domains (around 20 amino acids); however, even without interacting with the actin cytoskeleton, they can still mediate adhesion [11]. The CELSR cadherin subfamily comprises three members, CELSR1, CELSR2 and CELSR3, which are characterized by a large ectodomain featuring nine EC repeats, several cysteine-rich epidermal growth factor (EGF) repeat-like domains, two laminin G-type domains, one hormone receptor motif (HRM), and a G-protein-coupled receptor proteolytic site (GPS). Further features include a seven-pass transmembrane domain and a cytoplasmic domain. CELSR1 is involved in establishing and maintaining the epithelial cell-polarity of epithelial tissues while CELSR2 and CELSR3 play a role in neural development, especially neuronal migration and axon development and guidance in the central and peripheral nervous system [12].

Type I and type II classical cadherins as well as desmosomal cadherins share a great deal of similarity in their molecular structure and in their homophilic dimerization mechanism. Indeed, their extracellular domain consist of five EC repeats and they employ the two outermost EC repeats, EC1 and EC2, to dimerize through strictly homophilic interactions (type I classical cadherins and desmosomal cadherins) or through partly promiscuous heterophilic interactions involving different family members (type II cadherins) [13].

Type I classical cadherins and desmosomal cadherins employ the N-terminal EC1 repeat to interact in *trans* with an identical cadherin molecule protruding from an opposing cell. Here, a highly conserved tryptophan residue in position 2 docks inside a hydrophobic pocket of the partner molecule forming a so-called strand swap dimer. This interaction is usually coupled with a so-called *cis* interaction, whereby a second cadherin molecule, seemingly protruding from the same cell as the reference protein, interacts via its second EC repeat, EC2, at the level of the EC1 of a strand-swapped cadherin [14]. This mechanism will be discussed in detail in the paragraph where the structural features involved in the dimerization process are described. Type I classical cadherins are responsible

for the formation of adherens junctions between epithelial cells, which occurs via the synergic cooperation between *trans* and *cis* interactions [15]. Likewise, desmosomal cadherins are related to the formation of desmosomes, which are another type of specialized cell-cell junctions that connect epithelial cells [16].

Type II classical cadherins are characterized by the presence of two conserved tryptophan residues (in positions 2 and 4 of the N-terminal adhesion arm) that insert in a large hydrophobic pocket within the EC1 of the partner molecule in the homo-dimerization process [17]. Type II cadherins are expressed combinatorially in the nervous system and thanks to their ability to interact both homophilically and heterophilically with one another [13], they contribute to the development of neural circuits and to the spatial organization of neurons [18].

A schematic representation of the main features of the major cadherin subfamilies is depicted in Figure 3. In the following, the discussion will focus on the structure and functions of the adherens junction and on the prototypical member of the classical cadherin subfamily, E-cadherin, with particular reference to its possible druggability.

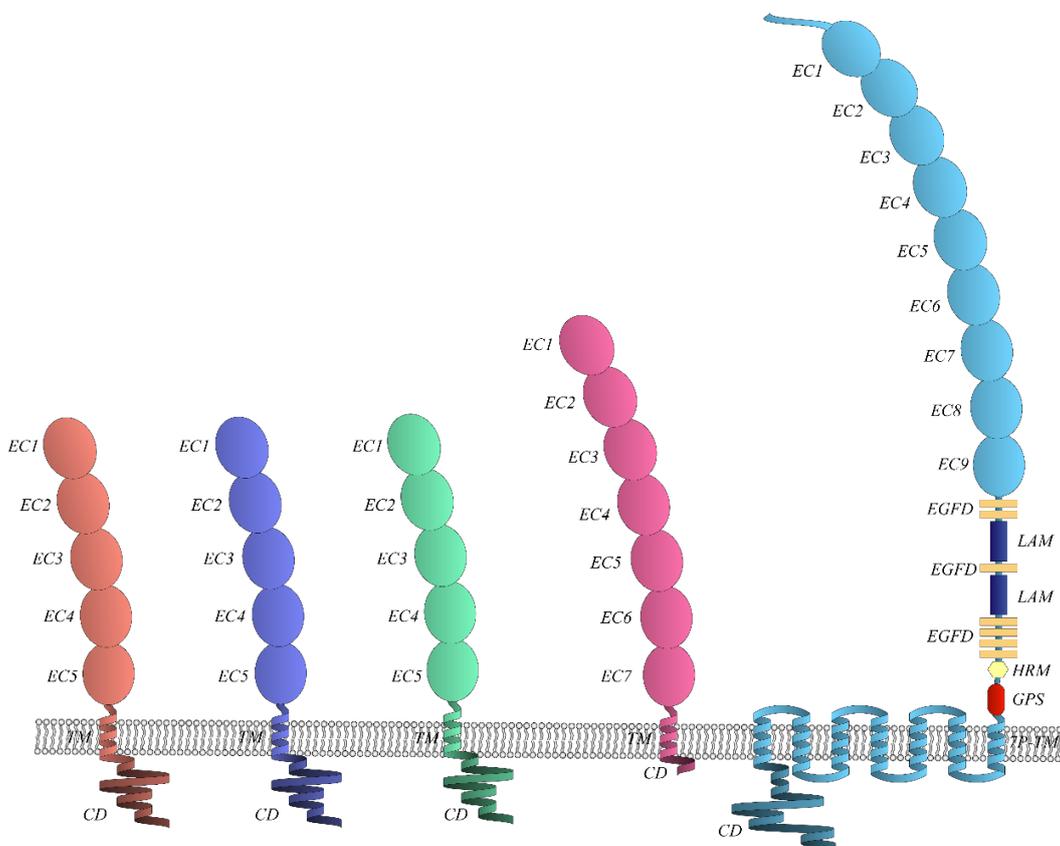


Figure 3: Schematic representation of the main features of the major cadherin subfamilies
 Here, a drawing highlighting the main structural features of the major cadherin subfamilies is shown. Starting from left, type I classical cadherins (red), desmosomal cadherins (blue), type II cadherins (green), 7D cadherins (pink) and CELSR cadherins (light blue) are depicted. The cell membrane is shown in grey and EC repeats are numbered from the outmost to the most proximal to the cell membrane. The abbreviation in the drawing stand for: TM, transmembrane domain; 7P-TM, seven-pass transmembrane domain; CD, cytoplasmic domain; GPS, G-protein-coupled receptor proteolytic site; HRM, hormone receptor motif; EGFD, cysteine-rich epidermal growth factor repeat-like domains; and LAM, laminin G-type domain.

The Adherens Junction and E-cadherin

Immunohistological studies have shown that while E-cadherin is ubiquitously distributed in all epithelial tissues, it is particularly abundant in the adherens junctions between epithelial cells, where it mediates cellular assembly [19], [20], [21]. Interestingly, an aberrant E-cadherin expression pattern often correlates with malignancy. Indeed, E-cadherin upregulation and/or downregulation, absence or mutations are associated with tumor growth, progression, and invasion. In this process, cancer cells undergo an epithelial-to-mesenchymal transition (EMT) and migrate from the tumor primary site into the blood stream. Finally, they undergo a reverse transition, the so-called mesenchymal-to-epithelial transition (MET), which restores their adhesive properties and allow them to adhere to other organs, thus establishing metastases in different regions of the body (Figure 4) [22], [23].

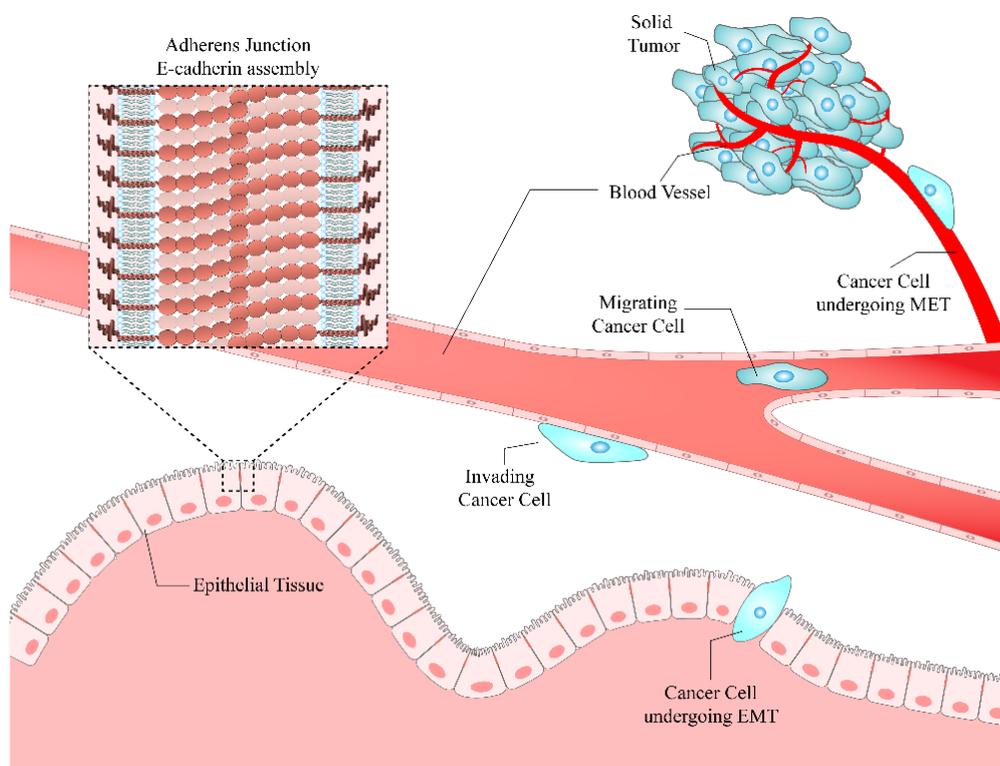


Figure 4: The adherens junction and the process of tumor formation

The figure depicts a simplified epithelial tissue sheet and the process of formation of a solid tumor: a cancer cell detaches from the epithelial sheet by undergoing EMT and it is able to migrate toward a blood vessel where the process of invasion begins. The migrating cancer cell eventually exits the blood vessel and undergoing a MET regresses to its original epithelial phenotype starting the process that ultimately leads to the formation of a solid tumor.

E-cadherin-mediated adherens junction formation is a key event in the development of epithelial tissues. Indeed, after the first few rounds of cell divisions, which take place during the early stages of development, the establishment of the apico-basal polarity and the formation of E-cadherin-mediated adherens junctions, i.e. the formation of the first epithelial tissue of an organism, mark the transition from morula to blastula [24].

Adhesive forces provided by cadherins at the adherens junctions contribute to the prevention of tissue dissociation into their cellular components and their inherent dynamicity provides tissues with the ability to withstand mechanical stresses while, at the same time, retaining cellular mobility [25]. As dynamic entities, adherens junctions confer adequate mechanical properties to the tissues. Moreover, they ensure the renewal process and the healing of wounds through cell movements and rearrangements.

The adherens junction is an extremely complex cell-cell interface. E-cadherin is a transmembrane protein that provides a linkage between cells while communicating with the underlying cellular cytoskeleton. The dynamic oligomerization of the E-cadherin ectodomain determines the formation, the rearrangement and the dissolution of adherens junctions [26]. The E-cadherin ectodomain is the major determinant of the overall architecture of the adherens junction, as it regulates its characteristic zipper-like structure and the spacing between cells (Figure 5: The structure of the adherens junction is regulated by the extracellular portion of classical cadherins (Figure 5) [14].

In this context, the initial study on C-cadherin, the ortholog of E-cadherin in *Xenopus Laevis*, revealed that the supramolecular arrangement of the full ectodomain (Figure 5, panel A) found in the crystal structure is compatible with the hypothesis that cadherin-mediated cell-cell adhesion is guided and organized by a complex array of so-called *cis* and *trans* intermolecular interactions. Moreover, it suggested that one layer of the C-cadherin EC1-EC5 portion in the crystal represents an accurate, albeit only partial, model for understanding the role of classical cadherins in cell adhesion and their activation mechanism [27].

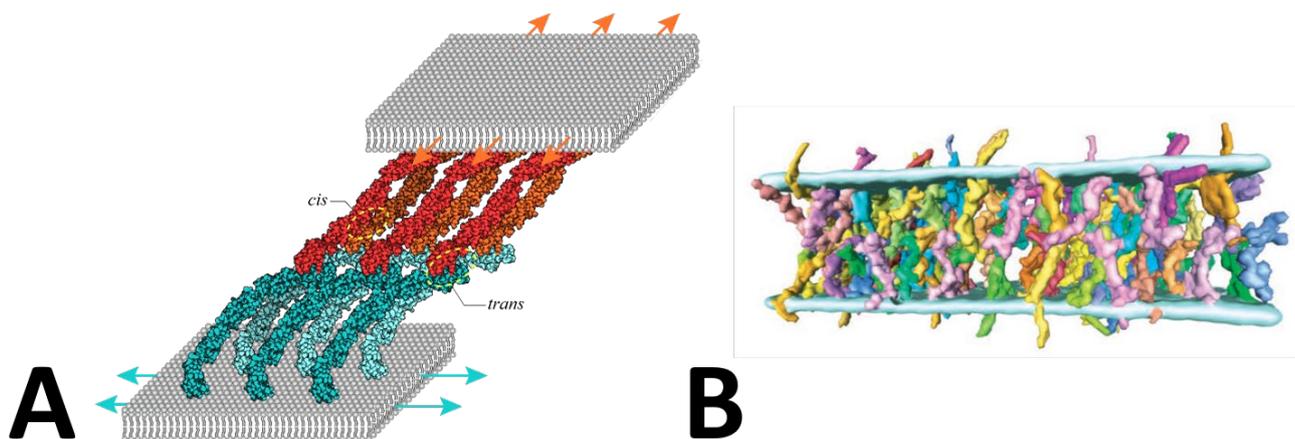


Figure 5: The structure of the adherens junction is regulated by the extracellular portion of classical cadherins. In the panel A, cadherin molecules of the same array contact each other in *cis* and at the same time establish a *trans* interaction with a molecule protruding from the opposite cell membrane (light blue molecules with red ones and cyan molecules with orange ones). Arrows exiting from the cellular membrane (depicted in gray) stand for the continuation of the array. Adapted from PDBID: 1L3W. Panel B depicts an electron tomography of a desmosome, where desmosomal cadherins are observed in multiple orientations. Adapted from [28].

Indeed, the observed distance between the EC5s of two interacting proteins in the crystal is remarkably similar to the average width of an adherens junction, as observed and measured by electron tomography studies [27].

As highlighted in panel A of Figure 5, the zipper-like structure of an adherens junction is formed by parallel arrays of cadherin molecules that are protruding from the same cell (*cis* interaction). In turn, they contact partner molecules forming *trans* interactions with cadherins that seemingly protrude from the opposing cell. Both interactions involve mostly the EC1 and the EC2 repeats.

However, although this extremely ordered two-dimensional crystal lattice has been actually observed in mature desmosomes [28], it is not completely accountable for the overall structure of the adherens junction. Indeed, in another electron tomography study [29], the authors demonstrated that cadherin molecules adopt a great variety of different conformations in cell-cell junctions and this enormous conformational freedom is one of the factor contributing to the dynamic properties of cadherin-mediated cell-cell junctions (Figure 5 panel B).

The great conformational flexibility of cadherin ectodomains can also be explained by the presence of three calcium ions at the interface between each EC tandem repeat, where they are coordinated by conserved Asp residues. Several studies [30], [31], [32] demonstrated that calcium chelation is essential for the dimerization of classical cadherins because it provides rigidity to the ectodomains and the necessary strain to position the EC repeats at the right relative orientation. At the same time, calcium chelation leaves enough conformational freedom to the protein to adjust the angles between its different EC domains, thus maximizing the likelihood of finding a homo-dimerization partner among those proteins that protrude from the neighboring cell (Figure 6, panel A).

In the absence of calcium, cadherins lose their structural rigidity and collapse onto themselves, finally being degraded by proteases. Interestingly, the effect of the absence of calcium on classical cadherins has not only been assessed functionally but it has also been shown by X-ray crystallography. Indeed, the EC1-EC2 portion of mouse H-cadherin was crystallized without calcium (PDB code 3K5R). The structure shows that the two EC domains collapse on one another, thus causing the ectodomain to lose its characteristic banana-shape and its functionality (Figure 6, panel B) [33].

The formation and the maturation of the adherens junction are complex processes that relate to the cadherin dimerization mechanism and to the supramolecular dynamic assembly of the E-cadherin ectodomains. Based on pioneering studies on Madin-Darby Canine Kidney (MDCK) epithelial cells [34], we can divide this overall process into three phases.

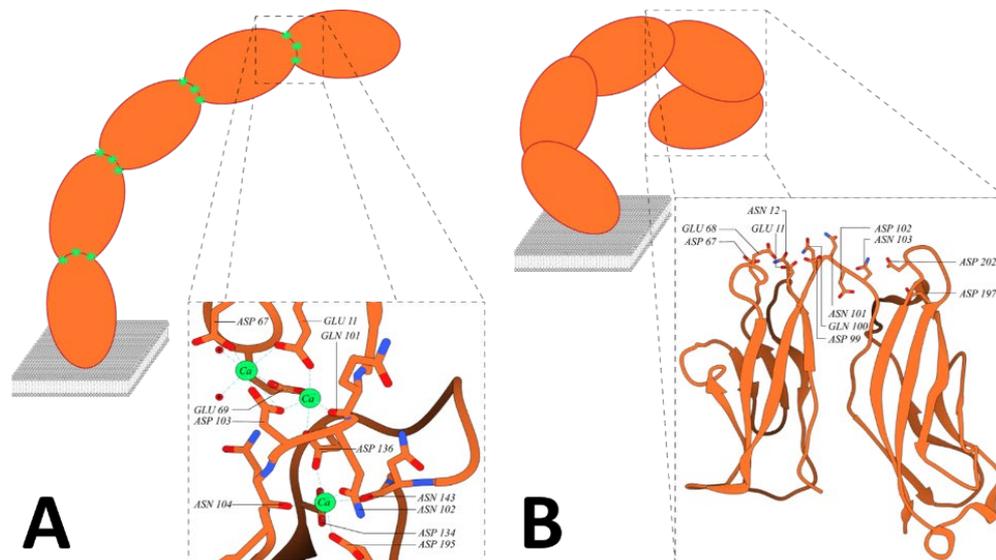


Figure 6: The effect of calcium binding

On top of panel A, a schematic representation of a calcium activated cadherin molecule is shown and, on the bottom, the typical conformation of a calcium-binding region is highlighted. In this case, the interface between EC1 and EC2 of human E-cadherin is shown as example. Each residue that binds calcium is drawn in sticks and labelled with its three-letter code name and its residue number. Calcium ions are drawn in green spheres and water molecules binding the calcium on the top are drawn as small red spheres and labelled with an asterisk. The figure was drawn starting from PDBID: 2O72.

On top of panel B, a probable conformation for calcium-free cadherin ectodomain is drawn and on the bottom the structure of calcium-free mouse H-cadherin (PDBID: 3K5R) is drawn as cartoon with the conserved calcium-binding amino acids shown in sticks. Carbon atoms are drawn in orange, oxygen in red and nitrogen in blue.

In the first phase, which is usually referred to as the exploratory phase, E-cadherin clusters assemble onto the cell membrane due to 1) the clustering forces provided by homophilic *cis* interactions acting via the passive trap mechanism [35] and 2) the intracellular dimerization of catenins, which bind to the E-cadherin cytoplasmic domain. Catenins stabilize E-cadherin clusters and, at the same time, contact actin bundles and filaments, providing further stabilization [36]. In this phase, lamellipodia-mediated contacts are populated by adhesion spots or *puncta*, in which E-cadherin begins to localize. As such, they represent the initial contacts between cells.

In the second phase, the so-called adhesion plaque is formed by the coalescence of *puncta* and by subsequent extension through the zippering of the plasma membranes. At this stage, the actin cytoskeleton rearranges because it is stimulated by the mechanical tensions that exist between adhered cells [37]. The newly formed adherens junction is connected to microtubules, which run radially and parallel to the apico-basal axis, through a plethora of adaptor proteins that mediate the linkage to the E-cadherin cytoplasmic domain [38]. The connection between adherens junctions and microtubules provides a route for the myosin- and dynein-mediated active transport of E-cadherin-containing vesicles to the adherens junction and this cooperative interaction contributes to the regulation of cell-cell adhesion [39]. The steady-state levels of E-cadherin molecules in dynamic clusters at the adherens junction is preserved by a combination of fission, fusion, endocytosis, recycling, and exocytosis of clusters and molecule pools [40].

The final phase is the condensation phase, in which the formation of multiple adherens junctions in epithelial cells leads to the formation of larger cell-cell contacts and to the redistribution of mechanical stresses via the maximization of the contact area between adhered cells.

Clearly, the E-cadherin dimerization process is only one of the many events that lead to the formation of the adherens junction. However, understanding its stress- and conformational-dependent adhesive properties allows the elucidation of its crucial role in the genesis of cell-cell contacts.

Structural Determinants of E-cadherin *Cis* and *Trans* Interactions

As previously stated, the organized, zipper-like array of molecules that characterizes adherens junctions is determined by the cooperation between *cis* and *trans* interactions in the ectodomains of classical cadherins. These interactions occur at the level of the two outmost EC repeats, as shown by the crystal structures of the whole ectodomain of C-cadherin of *Xenopus Laevis* [27] (PDBID: 1L3W) as well as E- (PDBID: 3Q2V) and N-cadherin (PDBID: 3Q2W) of *Mus Musculus* [14]. Since the heterologous expression of these long constructs is difficult in bacteria and usually requires the use of a mammalian expression system, which is typically characterized by a lower yield compared to the bacterial expression system, further crystallographic studies have focused on the minimal construct that is capable of mediating homo-dimerization, i.e. the EC1-EC2 fragment. Indeed, the simultaneous presence of *cis* and *trans* interactions has been confirmed for instance by Parisini et al. [41] in the high resolution strand-swapped structure of the EC1-EC2 domain of human E-cadherin (PDBID: 2O72). This minimal construct has significant advantages: it can be heterologously expressed in bacteria and with a significantly higher yield of soluble protein while, at the same time, resulting in a crystal structures where the nature of classical cadherin *cis* and *trans* interactions is fully preserved. The *trans* interaction results from a domain swapping mechanism, a protein dimerization mechanism that is common in nature and consists in the exchange of one domain from one protein with an identical one from the partner protein, whereby both swapped domains become stabilized by the same interface in the dimer. When two identical monomers come into contact, domain swapping is triggered by conformational changes that occurs at the level of the portion of the molecule that connect the swapping domain to the body of the protein, the so-called hinge loop, and the overall process leads to the stabilization of the swapped domains in the partner molecules [42]. The lengths and structures of the swapping domains can vary significantly among different proteins and range from entire globular domain to single α helices or β strands.

In all classical cadherins, the swapping domain is composed of the first six N-terminal residues of the protein that are contained in the first β strand of the EC1. This is usually referred to as the adhesion arm. The highly conserved Trp2 residue fits into a $\sim 800 \text{ \AA}^2$ wide, highly conserved hydrophobic pocket in the EC1 of the partner molecule. Here, it is stabilized by van der Waals interactions with the residues that line the adhesion pocket (Ile24, Ser26, Tyr36, Ser78, Ala80, Glu89, Met92) and by a hydrogen bond between the nitrogen atom of the Trp2 side chain (indole) and the carbonyl oxygen of Asp90 (Figure 7, panel A). At the N-terminus, the Asp1 residue is involved in multiple interactions: a salt bridge with Glu89 of the partner molecule, a hydrogen bond between its backbone carbonyl group and the backbone amide group of Asn27, and a hydrogen bond between the N-terminal nitrogen

with the side chain of Asn27 (Figure 7, panel A). In its swapped conformation, the arm is also stabilized by a hydrogen bond between Val3 and Lys25 of the partner molecule. This interaction mimics the contacts formed by the adhesion arm in the monomeric cadherin structure, when the arm folds against the protein's own body (closed conformation) and the Trp2 docks inside the protein's own acceptor pocket. Further stabilization is due to two water-mediated hydrogen bonds that are formed by the backbone amide group of Ile4 and the backbone carbonyl group of Pro5 with the cadherin body with the backbone carbonyl of Glu93 and the backbone amide group of Leu95, respectively (Figure 7, panel B). Moreover, the adhesion arm of E-cadherin is characterized by a diproline motif situated in position 4 and 5. When the adhesion arm is in its swapped conformation, such motif is strained because it features a conformational angle of -86° , which is significantly outside the energetically favorable -55° to -75° range [41]. Since the proline residues cannot act as hydrogen bonding donors/acceptors, they prevent the formation of a stable intermolecular conformation of the adhesion arm [43]. Hence, the diproline motif can be regarded as a particular hinge loop and its conformational strain might be thought of as a “spring force” that unlocks the Trp2 residue from its acceptor pocket.

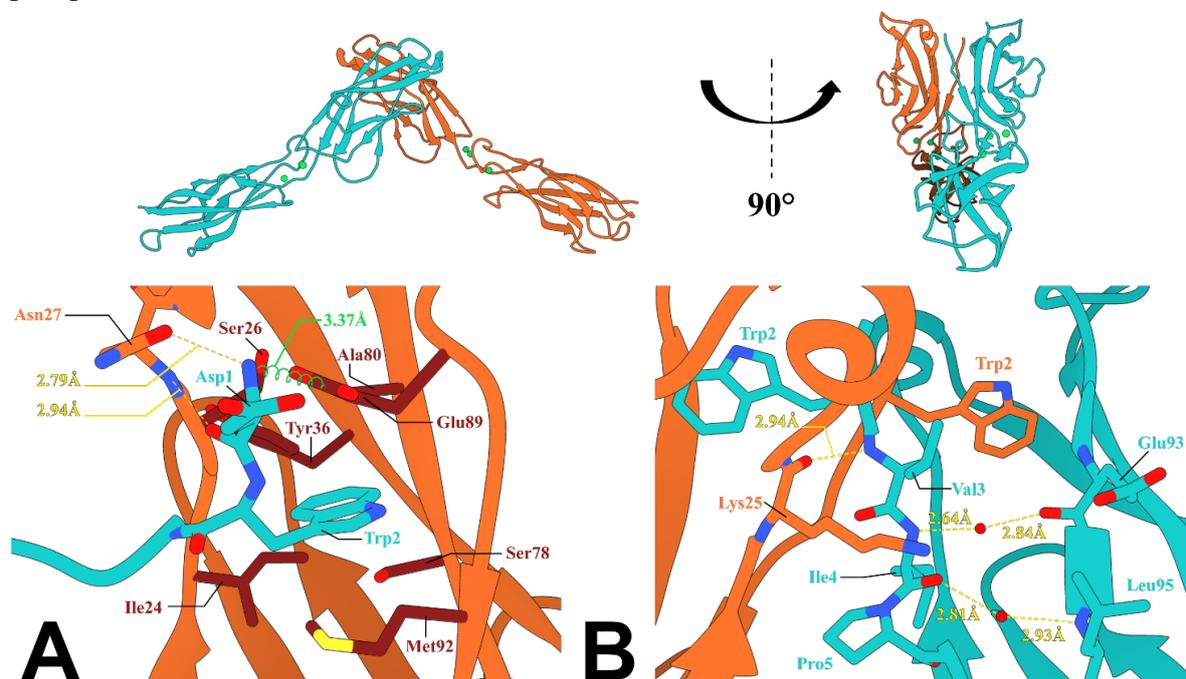


Figure 7: Structural determinants of the trans interaction of E-cadherin

The top side of the figure shows the cartoon model obtained from the crystal structure of E-cadherin EC1-2 in strand swap conformation (PDBID:2O72), where the two proteins in the dimer are colored in orange and cyan with green spheres as calcium atoms. The interactions depicted in the two underlying panels take place in the two EC1.

Panel A highlights the interactions of Trp2 and Asp1 in the acceptor pocket of the partner molecule. Here, the hydrophobic residues that compose the pocket are colored in dark red, the cadherin body in orange and the swapped adhesion arm in cyan with red oxygens, blue nitrogens, and yellow sulphurs. All residues are labelled according to their three-letter code and sequence number. Hydrogen bonds are indicated by yellow dashed lines and accompanied by their corresponding distances, while the green spring corresponds to a salt-bridge whose distance is indicated on top of it.

Panel B depicts the stabilizing interactions that the swapped adhesion arm establishes with its partner molecule and with the cadherin body from which it protrudes. Here the color-code of panel A is still applied and water molecules are indicated as small red spheres.

The *cis* interaction (Figure 8, top panel) involves the formation of an interface between two regions, one situated in the EC1 domain and one in the EC2 domain of two E-cadherin proteins seemingly protruding from the same cell. Multiple interfacing residues mediate the *cis* interaction in EC1 occupying an extended region of circa 500 Å². They are: Phe35, Tyr36, Ser37, Ile38, Thr39, Asp44, Arg55, His79, Ala80, Val81, Ser82, Ser83, Asn84, Gly85, Asn86, Ala87, and Pro91; while in EC2 they are: Gly120, Leu122, Pro123, Gly124, Asp159, Asn161, Thr164, Ile165, Asn166, Arg167, Asn168, Val171, Ser173, Val174, Val175, Thr176, Thr177, and Gly178. At the center of these interfaces lie the conserved His79-Ala80-Val81 (HAV) motif in the EC1 and the Pro123-Gly124-Thr125 (PGT) motif in the EC2.

As shown in panel A and panel B of Figure 8, both HAV and PGT motifs are situated in hydrophobic regions that are surrounded by polar residues. Here, Pro123 forms a characteristic kink that fits into the HAV motif and establishes stabilizing hydrophobic interactions, while the surrounding polar regions feature hydrogen bonds (e.g. between the carbonyl group of Ser83 and the side chain of Thr164 or between the side chain of Asn84 and the backbone amide of Arg167) and water-mediated contacts (Figure 8, panel C).

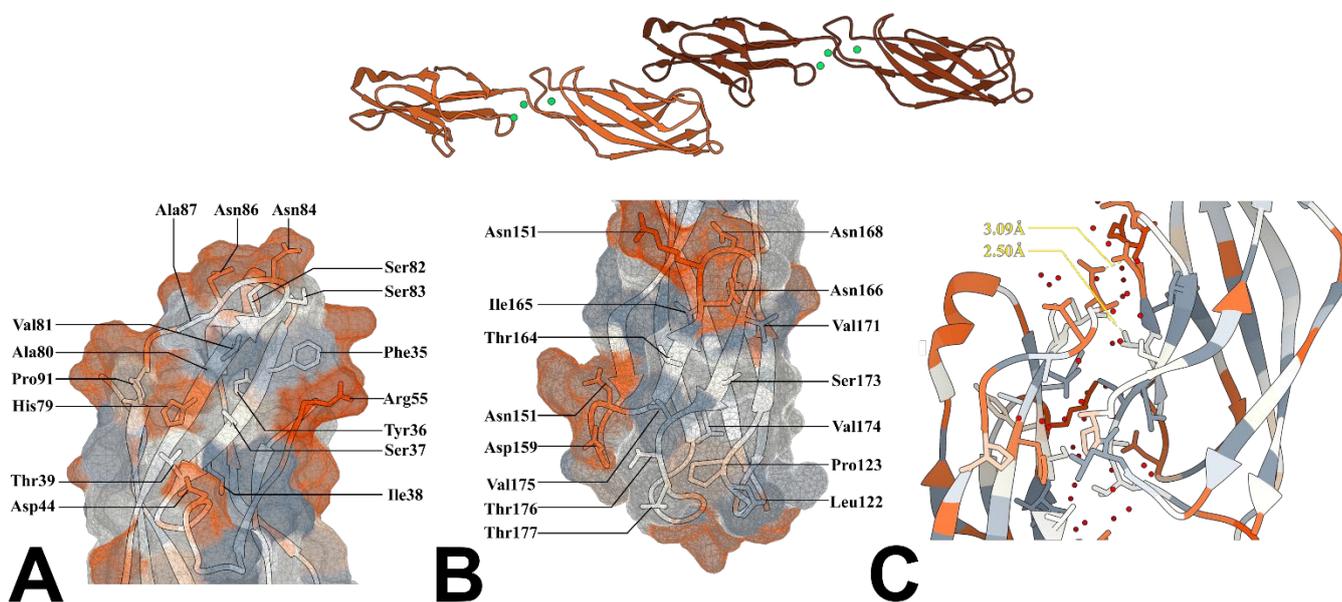


Figure 8: Structural determinants of the *cis* interaction in E-cadherin

The top side of the figure shows the cartoon model obtained from the crystal structure of E-cadherin EC1-2 in strand swap conformation (PDBID:2O72), where the two proteins are obtained through crystal symmetry and are colored in orange and brown with green spheres as calcium atoms. The interactions depicted in the two underlying panels take place in one EC1 and one EC2.

Panels A and B highlight the two interfaces in respectively EC1 and EC2 that are involved in the *cis* interaction. Both protein bodies are shown in cartoon model with the respective surface superimposed. The residues involved in the interaction are drawn in sticks and the color code used for the residues and the surface is based on the Kyte and Doolittle hydrophobicity scale using dark gray for the most hydrophobic residues and orange red for the least hydrophobic. Each interacting residue is labelled with its three-letter code and sequence number.

Panel C depicts the two interfacing while interacting and uses the same color code as the previous panels. Water molecules are represented as small red spheres and hydrogen bonds are drawn with yellow dashed lines with their corresponding distance.

Interestingly, mutating the HAV motif by inverting it into VAH does not affect the adhesive ability of E-cadherin in *in vitro* cell assays, despite the HAV motif being totally conserved among classical cadherins, [44]. However, Harrison et al. [14] have shown that site-directed mutagenesis of either Val81 in EC1 or Leu175 in EC2 into Asp resulted in the abolishment of *cis* interaction in crystal structures, due to the perturbation of the hydrophobic nature of the contact. These data confirm the importance of the *cis* interaction in the supramolecular assembly of E-cadherin molecules.

The E-cadherin Homo-dimerization Process

In order for the cadherin homo-dimerization process to occur, the energy barrier that is associated with strand swapping event needs to be overcome. This results in a slow process that, rather than featuring an instant on/off mechanism, requires E-cadherin molecules to transit between several intermediate conformational states before reaching the final strand-swap dimer conformation [45]. Historically, the exploration and the elucidation of this complex mechanism began with the crystallization of an E-cadherin molecule featuring an amino acid N-terminal extension of a few amino acids resulting from the cleavage of the His-tag that was used for purification purposes [46]. In the cell, E-cadherin is produced as an inactive protein whose 100 residue-long N-terminal portion folds into a prodomain that is subsequently cleaved by proteases during protein maturation in the Golgi apparatus [46]. Before prodomain cleavage, the protein stays inactive and in a close conformation, featuring the Trp2 residue docked into the protein's own binding pocket. Indeed, the bulky prodomain portion prevents the protein from undergoing adhesion arm opening and proceeding to strand swap dimer formation. Similarly, it has been found that in all classical cadherins the N-terminal addition of amino acid sequences of any length provide a bulky extension that results in the hindering of the strand-swap (*trans*) interaction. In this context, depending on the type of N-terminal extension, Trp2 has been found to be either docked inside its own acceptor pocket (PDBID: 1FF5, [30]) or crystallographically disordered, i.e. non-visible in the electron density map (PDBID: 1EDH, [47]). In all cases, the corresponding crystal structures feature two E-cadherin molecules arranged in an X-shaped dimer whereby they contact each other at the level of the calcium-binding domains, which are located near the linker regions between the EC1 and the EC2 of the two proteins. Owing to the shape of this type of dimer, this conformational state is usually referred to as the X-dimer (Figure 9). Although the X-dimer was initially considered to be a crystallographic artifact because of the presence of an artificial N-terminal amino acid extension in the protein, subsequent structural studies demonstrated that the X-dimer is a *bona fide* and biologically relevant adhesive conformation. Indeed, it was shown that other classical cadherins [46], [48] interact homophilically to form X-dimers; interestingly, H-cadherin, which lacks the conserved Trp2 residue that is found in all classical cadherins, has been shown to be nevertheless able to mediate adhesion [33]. Furthermore, the adhesive properties of the X-dimer have also been characterized by atomic force microscopy (AFM) in a study published by Rakshit et al. [49] where it was also demonstrated that, to reach the strand swap conformation, cadherins pass through the X-dimer conformation, which allow the lowering of the energy barrier that is associated with the strand swapping event.

Therefore, the so-called X-dimer conformation can be considered as an intermediate checkpoint in the cadherin dimerization process in which the EC domains of the two interacting molecules come in close contact and conveniently orient themselves so that strand swapping becomes more energetically favorable.

The crystal structure of human E-cadherin in the X-dimer conformation (PDBID: 4ZT1) revealed that the binding involves the same amino acids on the two protomers [50]. These residues, which are distributed across an 800 Å² surface area, are: Pro5, Pro6, Ser8, Pro10, Glu11, Asn12, Glu13, Lys14, Gly15, Lys19, Leu21, Val22, Thr99, Asp100, Gln101, Asn102, Asp103, Lys105, Asp138, Val139, Asn140, Thr141, Tyr142, Asn143, Leu196, Glu199, Gly200, and Leu201. Overall, the interaction is fundamentally hydrophobic in nature, with contributions from water-mediated contacts and hydrogen bonds (Figure 9, panel A).

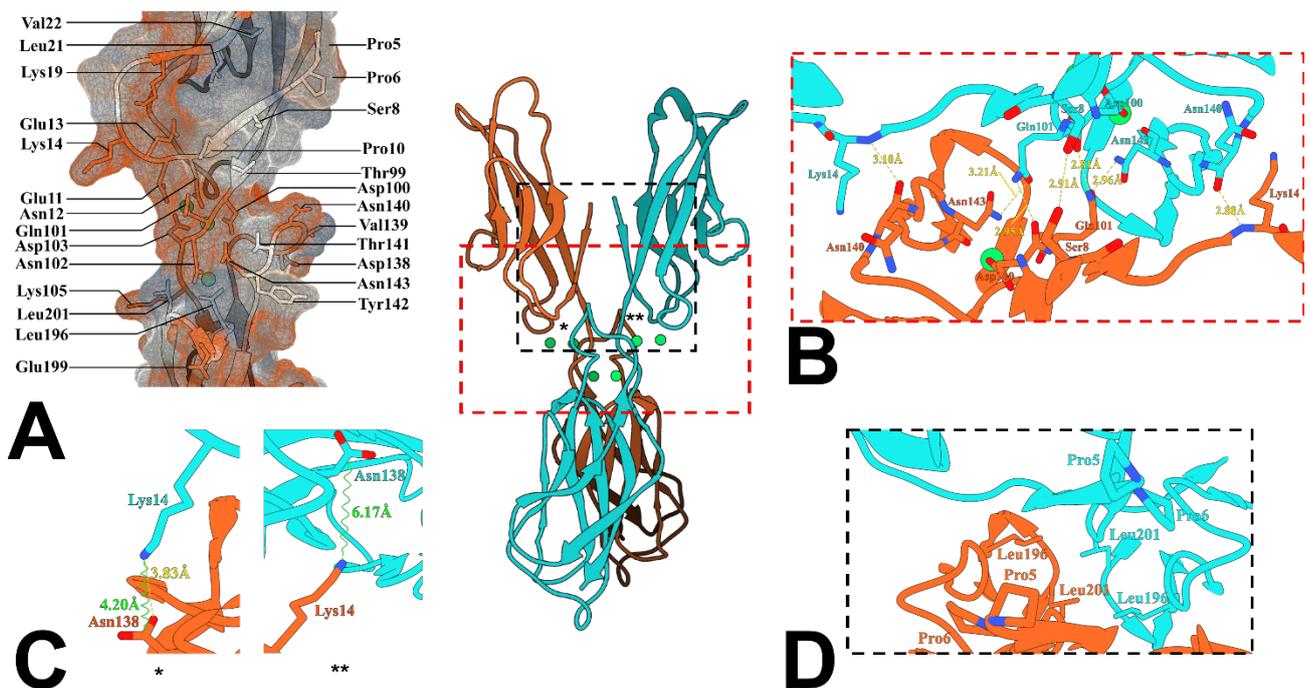


Figure 9: Structural determinants of the X-dimer

The central side of the picture shows the X-dimer conformation observed in the crystal structure of a truncated form of E-cadherin (v3-E-cadherin) (PDBID: 4ZT1), where the two interacting proteins are drawn in cartoon model and are colored in orange and cyan with green calcium ions. The truncation involved the removal of the first two amino acids at the N-terminus of the protein.

Panel A shows the residues that form the X-dimer interface. Each residue is labelled with its three-letter code and sequence number, and it is drawn in sticks. The surface of the protein is also shown with a color-code based on the Kyte and Doolittle hydrophobicity scale using dark gray for the most hydrophobic residues and orange red for the least hydrophobic.

Panel B depicts the hydrogen bonding network described in the text. Each hydrogen bond is drawn with yellow dashed lines and the measured distance is indicated. Each interacting residue is drawn with sticks, the protein body in cartoon model and the color code is the same as the central panel, the red box encompasses roughly the considered region of the dimer.

Panel C shows the details of the interactions of Lys 14 and Asp138 in the two monomers showing that Lys14 from monomer B is much closer to Asp138 in monomer A compared to the other couple. Hydrogen bonds are indicated with yellow dashed lines and coulomb interactions with green springs. Each interaction mode is accompanied by its measured distance colored with the same color code. Each interacting residue is drawn with sticks, the protein body in cartoon model and the color code is the same as the central panel. Asterisks refer to the position of the represented interaction in the crystal structure.

Panel D shows the hydrophobic interactions described in the text. Each interacting residue is drawn with sticks, the protein body in cartoon model, and the color code is the same as the central panel, the black box encompasses roughly the considered region of the dimer.

As the two protomers form a dimer featuring a pseudo C2 symmetry axis (Figure 9, left panel: the two proteomers are drawn in cartoon models and colored orange and cyan), the hydrogen bonding pattern across the dimer is mainly symmetric. Stabilizing interactions between the side chain of Ser8 of monomer A and its counterpart from monomer B, between Asn140 backbone carbonyl group and the backbone amide group of Lys14, between Asp100 backbone carbonyl group and the side chain of Gln101, and between Gln101 side chain and the side chain of Asn143 are clearly identifiable (Figure 9, panel B).

Notably, the only asymmetric interaction is the one made by Lys14, whose side chain in monomer A is at hydrogen bonding distance with the side chain of Asp138, while in monomer B it is farther from Asp138, at a distance that is however compatible with the formation of a salt-bridge (Figure 9, panel C).

Furthermore, the amino acid pairs Leu196-Leu201 and Pro5-Pro6, which are symmetrically oriented and interact via van der Waals forces, constitute the only hydrophobic residues involved in the X-dimer interface (Figure 9, panel D).

While the characterization of the X-dimer conformation clarified the most important intermediate step in the E-cadherin dimerization process, the starting point of the whole dimerization process remained structurally uncharacterized until the structure of human P-cadherin in closed conformation was determined by Dalle Vedove et al. in 2014 [51]. P-cadherin is another Type I classical cadherin that shares 67% sequence identity to E-cadherin and, so far, is the only classical cadherin that naturally crystallizes as a closed monomer. The k_D measured by ultracentrifugation of its EC1-EC2 fragment is 31 μM , a value that is compatible with the standard concentration used in crystallization experiments [52]. In fact, in ultracentrifugation experiments the E-cadherin EC1-EC2 construct displays a k_D value of 97 μM , which excludes the possibility to crystallize it in monomeric form because the standard concentration used in crystallization experiments is lower than the k_D . This difference can be ascribed to the lack of the diproline motif in the adhesion arm of human P-cadherin. This allows P-cadherin to crystallize in the closed or in the strand swap conformation, depending on the crystallographic conditions. Indeed, the introduction of the diproline motif in the adhesion arm of P-cadherin forces the crystallographic structure to be in the strand swap conformation [45], [51].

Besides providing structural information on the starting point in the classical cadherin dimerization process, the crystal structure of human P-cadherin (PDB ID: 4OY9) reveals further crucial functional information. Indeed, the human P-cadherin structure allowed to infer that in solution, prior to the strand swapping event, classical cadherins feature two interconverting conformational states: one characterized by the adhesion arm stabilized in its acceptor pocket and the other with the adhesion

arm in an open (albeit not strand swapped) conformation. In fact, the side chain of Trp2 is found in two alternative conformations inside the acceptor pocket (Figure 10, panel A and B). In one conformation, the nitrogen in the Trp2 side chain forms a hydrogen bond with the backbone carbonyl of Asp90, as also seen in the E-cadherin strand swap dimer; in the other, it interacts with the backbone carbonyl group of Lys25, which is situated at the opposite side of the acceptor pocket. Since the distance between the two residues that sandwich Trp2, Glu89 and Met92, is not compatible with a rotation of its side chain inside the pocket, it must be concluded that cadherin monomers exist as interconverting conformations. In essence, the adhesion arm continuously enters and exits the acceptor pocket, thus offering the indole the possibility to re-enter with any of the two possible orientations. The speed of this conformational interconversion is much higher than the speed of the overall dimerization process, which has been estimated to be in the order of seconds [26].

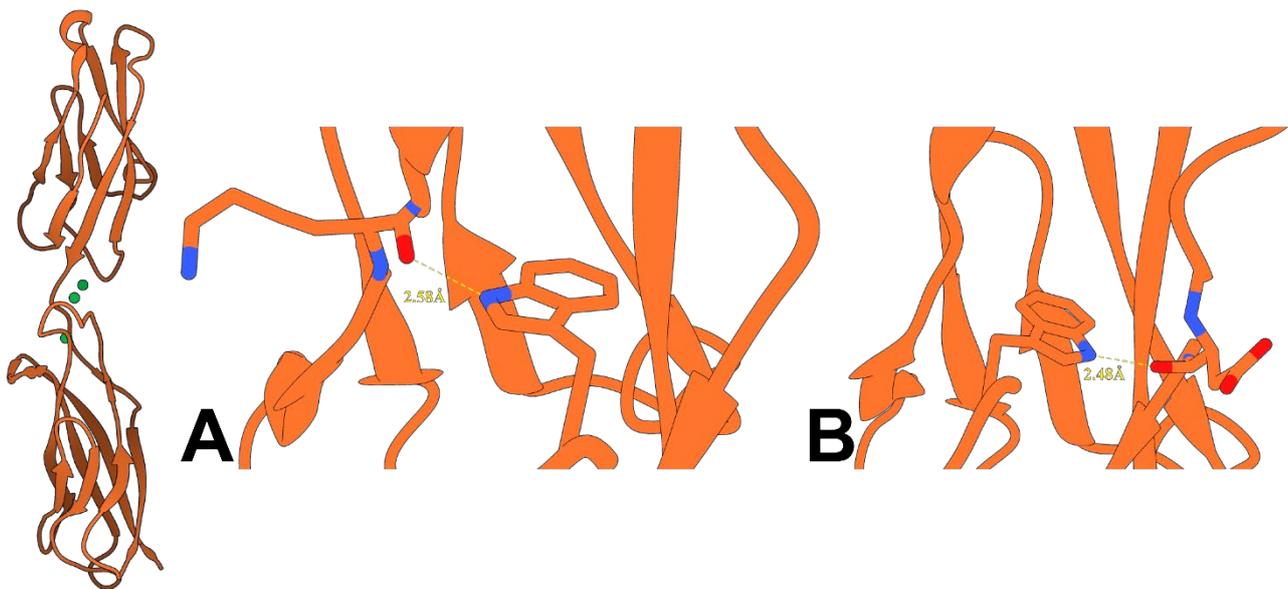


Figure 10: Detail of P-cadherin in closed conformation

The left side of the figure shows the crystal structure of P-cadherin in closed conformation represented as cartoon model with the protein colored in orange and the calcium ions represented as green spheres.

Panel A shows the hydrogen bonding interaction between Trp2 and Lys25. The two residues are drawn in sticks, while the hydrogen bond is represented with a yellow dashed line and labelled with the measured distance.

Panel B shows the hydrogen bonding interaction between Trp2 and Asp90. The two residues are drawn in sticks, while the hydrogen bond is represented with a yellow dashed line and labelled with the measured distance.

Thanks to the possibility to map a variety of conformation states provided by several different available crystal structures of P-cadherin and to the extensive structural data previously collected on E-cadherin, we are now able to elaborate the following model for the dimerization of classical cadherins (Figure 11).

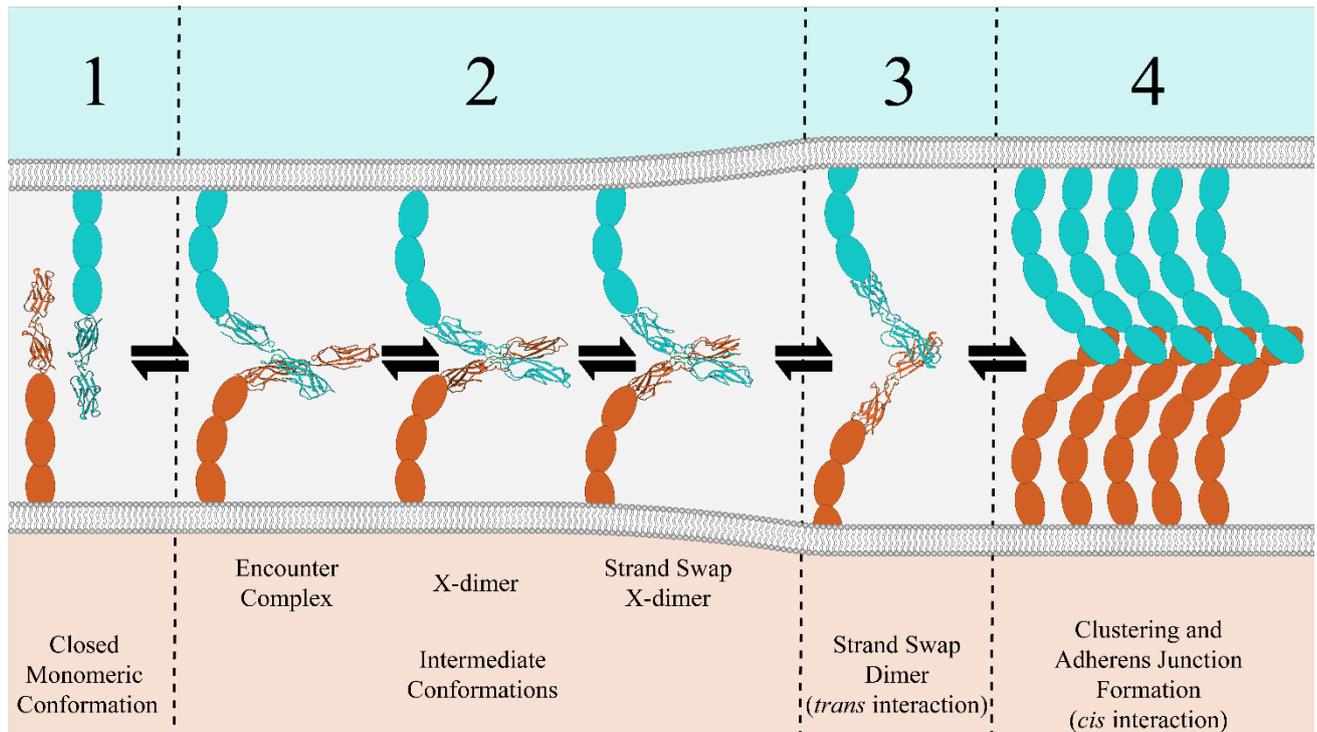


Figure 11: Cadherin homodimerization model

The picture depicts the current model for the cadherin dimerization process. Cadherin molecules are drawn in orange and cyan, with EC1 and EC2 drawn as cartoon models obtained from the deposited crystal structures (PDBIDs: 4OY9, 4ZMW, 4ZT1, 4ZMQ, 2O72), while EC3 to 5 as colored ellipses, with the exception of stage 4, where EC1 and EC2 structures were substituted by ellipses for clarity. The plasma membrane bilayer is schematically drawn in gray.

The picture shows the pathway of cadherin homodimerization and adherens junction formation by going through different conformational states in dynamic equilibrium with one another.

In stage 1, monomeric cadherins meet at the interfaces between cells, their first approach results in the formation of an encounter complex that over time evolves into the X-dimer. All the X-dimer conformations are collected in stage 2 that is terminated when the adhesion arms are exchanged. From this moment onward, the cadherin bodies begin to shift and reorient and eventually they establish the strand swap dimer (*trans* interaction).

Over time, strand swap dimers cluster on the cell membrane surface of adhering cells (*cis* interaction) and this event marks the formation of a stable cell-cell contact and the formation of the adherens junction that matures when more strand swapped dimers coalesce in the early junction interacting through lateral clustering.

1. Monomeric cadherins that are primed for adhesion (i.e. after prodomain cleavage) exist in two different conformations, one with the Trp2 docked into the protein's own acceptor pocket and the whole adhesion arm folded in a closed conformation, the other with the Trp2 outside the EC1 acceptor pocket and the adhesion arm taking up an open conformation, as described by Dalle Vedove et al. [51].
2. When two cadherin molecules approach each other, they start interacting through the calcium-binding regions to form the so-called X-dimer. In this conformation, the two EC1 repeats are put into close contact and, thanks to the hydrophobic environment at the bottom of the adhesion arms, strand swap can occur. In turn, the X-dimer conformation can be considered as an ensemble of different conformations that mark the progressive approach of EC1 domains, as it was determined by the different X-dimer conformations crystallized by Kudo et al. [45]. In fact, in the structure named enc-X-dimer (PDBID: 4ZMW) the contact between

P-cadherin molecules is at the level of the EC2 domains and it can be regarded as the initial contact between cadherin molecules. After this initial contact, P-cadherin shifts toward the conventional X-dimer conformation, where the interaction between the two proteins occurs almost exclusively at the level of the calcium-binding regions (PDBID: 4ZMX). At this point, cadherin molecules further proceed in their dimerization process by forming yet another intermediate state in which the X-dimer conformation is still present, albeit with the adhesion arms of the two protomers swapped (ss-X-dimer, PDBID: 4ZMQ). This latter state explains why the X-dimer is said to “catalyze” the formation of the strand swap despite the fact that cadherins are not enzymes: the progressive approach between cadherin molecules to take up the X-dimer conformations generates the necessary conformational changes that ultimately lower the energy barrier associated with the exchange of the adhesion arms. In a similar fashion, enzymes orient their substrates to catalyze biochemical reactions by lowering their activation energy.

3. Once the adhesion arms have been exchanged and the Trp2 residues have been accommodated in the partner acceptor pockets (*trans* interaction), the cadherin bodies begin a twisting rearrangement movement perpendicular to the scissoring direction defined by the formation of the X-dimer and this movement leads to the formation of the final strand swap dimer [53]. Here, as initially noted by the analysis of human E-cadherin strand swap dimer by Parisini et al. in [41] and subsequently in P-cadherin by Kudo et al. in [45], water molecules stabilize the opened conformation of the adhesion arm. Indeed, their contributions in terms of hydrogen bonding is highly relevant for the stabilization of the strand swap dimer, which otherwise would tend to rapidly dissociate.
 4. The lateral clustering of strand swap dimers through *cis* interactions provides further stabilization to the strand swap dimer. Henceforth, arrays of cadherin strand swapped dimers can stabilize one another by clustering together, thus providing sufficient resistance to the mechanical stresses that the cell is subjected to by responding as a single entity within an adherens junction.
-

E-cadherin and the Cancer Phenotype

The formation of E-cadherin-mediated cell-cell contacts marks a crucial step in the development of epithelial tissues, the most simple, primordial, and widespread tissues in most organisms. The organization of cells in epithelial tissues stands at the basis of the process of developing, remodeling and protecting organs against external hazards. Moreover, the epithelium acts as a physical barrier to prevent dehydration, while contributing to the integrity and to the shaping of different organs.

The correct segregation of different cell populations relies on the formation of selective cell-cell contacts. The formation and the dissolution of such cell-cell junctions hold equal importance for the formation of patterns of different cells. Indeed, by dissolving the adherens junctions, epithelial cells become capable of movement and can invade different regions of a developing body by acquiring a motile phenotype that grants the possibility of reorganizing and creating new cellular structures and patterns in order to develop adult organs and tissues [24], [54].

The process of the acquisition of a motile phenotype is called epithelial-to-mesenchymal transition (EMT), while the inverse process is the MET. A complex network of intracellular, intercellular and microenvironmental cues that can translate into the downregulation or the upregulation of epithelial and mesenchymal markers orchestrates both processes.

Despite being held together by specialized cellular junctions such as the adherens junction, epithelial cells retain the ability to move and rearrange inside the two-dimensional epithelial layer in which they lie. On the other hand, mesenchymal cells contact their neighbors only focally and are generally highly motile [55].

The ability to undergo EMTs and/or METs is an evolutionary requirement that defines both the natural processes of development and regeneration (type I EMTs) and the pathological ones of fibrosis (type II EMTs), tumor growth and cancer progression (type III EMTs) [56].

According to the U.S. National Cancer Institute, the origin of almost 90% of cancer cases are carcinomas, i.e. malignancies of the epithelial tissue [57]; this high percentage is not surprising since epithelial tissues are widely spread in the human body as they form the skin, the covering and lining of all organs and internal passageways.

Even though the precise molecular details vary from case to case, the overall sequence of events that lead to type III EMTs and METs in carcinomas can be considered essentially the same as those that govern the physiological processes of development and wound healing. In type III EMTs, epithelial cancer cells disassemble the intercellular and cell-matrix contacts and reorganize their cytoskeleton to form lamellipodia and filopodia and be able to onset migration. As they acquire a motile phenotype, new cell-matrix contacts are formed at the leading edge to provide anchorage to the cell body and the

pericellular matrix is degraded. In particular, the motile phenotype is provided by changes in the type of adhesion molecules [58]. Eventually, cancer cells can intravasate into a blood or lymphatic vessel and after circulating in the blood system, they extravasate from it when they reach a new location [59]. Despite the vast majority of cancer cells that escape into the systemic circulation either die, fail to invade distant sites or lie dormant for extended periods of time, only a small population actually recapitulates the EMT through a MET, regains its epithelial phenotype by forming new cell-cell and cell-matrix contacts and establishes a new tumor growth site [60], [61], [62].

The tumor microenvironment is characterized by abnormal conditions such as hypoxia, low pH, inflammation as well as scarce nutrient concentrations and these factors contribute to eliciting a series of responses that go in the direction of cancer cell metabolic adaptation (Warburg and post-Warburg effects), epigenetic alterations and dedifferentiation [63]. All these phenotypic modifications characterizing cancer cells are related in complex ways to signaling pathways and promote signaling cascades that affect the expression of classical cadherins and in particular E-cadherin. Indeed, E-cadherin is generally downregulated to dismantle cell-cell contacts in the preparatory phase of EMTs and then upregulated during METs and solid tumor formation [63], [64].

In the context of oncogenesis, E-cadherin has been extensively described both as a tumor suppressor gene and as a pro-oncotic gene, depending on the cellular context and malignancy [65].

Examples of epithelial tumor cells that usually lose E-cadherin expression and/or function during their progression toward malignancy are pancreatic cancer [66], gastric cancer [22], [67], [68], [69], [70], melanoma [71], prostate cancer [72], [73], [74], solid and ascitic sarcoma [75] and colorectal cancer [76]. Here, the tumor suppressor role of E-cadherin is attested by the fact that rescuing its expression leads to the inhibition of invasiveness and eventually to tumor cell death.

In turn, the pro-oncotic behavior of E-cadherin is observed for example in ovarian cancer [77], [78], [79] and inflammatory invasive breast cancer [80], [81], where E-cadherin overexpression is the main cause of solid tumor growth and compaction.

The alteration of the expression profile of E-cadherin in tumor progression can occur via different mechanisms, namely increased endocytosis and proteolytic degradation, loss of heterozygosity in E-cadherin gene, epigenetic modifications and silencing, and inactivating mutations [82].

Moreover, it has been observed that tumors often recapitulate a process reminiscent of the switching in adhesion molecules that takes place during normal embryonic development. This process has been named “cadherin switching” and consists in the change in the expression profiles of cadherin molecules (Table 2). Although it was initially thought that cadherin switching proceeded through the exchange of E-cadherin to N-cadherin, further evidences in the literature suggested that a general rule

could not be established, as the change does not exclusively involves N-cadherin, but also other cadherin molecules. Moreover, the expression profiles of the normal cadherin set of a cell can be affected, not limiting the cadherin switching phenomenon to the downregulation of E-cadherin [83].

Table 2: Cadherin switching during oncogenesis. Adapted from [83].

Switch	Example	Reference
From E-cadherin to N-cadherin	Melanoma	(Li et al., 2001)
	TGF β -induced EMT in mammary epithelial cells	(Maeda et al., 2005; Miettinen et al., 1994; Piek et al., 1999)
	Prostate cancer	(Gravdal et al., 2007; Jaggi et al., 2006; Tomita et al., 2000)
From E-cadherin to T-cadherin	Breast cancer	(Han et al., 1999; Hazan et al., 2000; Nieman et al., 1999)
	Pancreatic cancer	(Hotz et al., 2007; Nakajima et al., 2004)
From E-cadherin to P-cadherin	Hepatocellular carcinoma	(Riou et al., 2006)
	Pancreatic cancer	(Taniuchi et al., 2005)
From E-cadherin to cadherin 11	Gastric cancer	(Shimoyama and Hirohashi, 1991)
	Prostate cancer	(Bussemakers et al., 2000; Tomita et al., 2000)
	Breast cancer	(Pishvaian et al., 1999)
From E- and P-cadherin to N-cadherin	Oral squamous cell carcinoma	(Chen et al., 2004; Islam et al., 1996; Pyo et al., 2007)
From N-cadherin to E-cadherin	Ovarian cancer	(Patel, I. et al., 2003; Wong et al., 1999)

The Limitations of Conventional Therapies and the Possible Role of E-cadherin as a Pharmaceutical Target

The unregulated proliferation and survival of cancer cells within a solid tumor results in the production of gradients of nutrients, byproducts, and acidity because of the increased intervascular distances. In fact, the formation of neo-vasculature is necessary for tumor growth, but the developing blood vessels cannot keep pace with the growth rate of an expanding tumor and therefore develop as highly chaotic, i.e. lacking hierarchy and regularity. Since the speed and flow direction in the neo-vasculature are unstable because of the fragility of these vessels, blocked red blood cells and lymphocytes or the very tumor cells growing into them are the likely cause of blockages in the vasculature. These factors ultimately contribute to the development of a starving, hypoxic and highly acidic environment, where the unstable and dysfunctional neo-vasculature is not able to ensure suitable gradients and concentrations of anticancer drugs inside the solid tumor. In turn, the solid tumor develops therapeutic resistances over time [84] and the hypoxic environment greatly diminishes the effects of radiation therapy since the creation of reactive oxygen species that damage tumor cells is based on the presence of oxygen molecules [85].

The factors that mainly affect the delivery of chemotherapeutics through the bloodstream and their effectiveness on cancer cells are:

1. The limited availability of a suitable diffusion gradient because of the leaky and aberrant tumor neo-vasculature.
2. The variability in tumor vessel perfusion characterized by irregular periods of opening and closing.
3. The microenvironmental conditions such as low oxygen and glucose, high acidity, and irregular blood flow, coupled to the genetic instability of tumor cells might contribute to the production of mutated tumor cell lines that are more aggressive and insensitive to treatment.
4. The ability of tumor cells to express drug-transporters that can secrete chemotherapeutics outside tumor masses further decrease intratumoral drug concentrations.

Therefore, it is apparent that pharmacokinetics and biological variables (dose, drug metabolism, excretion rates and drug sequestration) are the main supply factors to take into account, while the intratumoral distribution of drugs is determined by the physicochemical properties of the active molecule and represent an important clinical parameter that, however, is often hard to evaluate with precision.

Given the importance of E-cadherin in the formation of the adherens junction and its role in tumor onset and progression and provided that conventional therapies present the above-mentioned limitations, we might attribute the possible role of pharmacological target to E-cadherin.

This role is clearly not the one of target for an antitumoral therapy *per se*, as it might be pointed out that targeting the homo-dimerization of E-cadherin in tumors might stimulate them to acquire a motile phenotype and prime them for invasiveness. However, given the limited possibility for chemotherapeutics of penetrating solid tumors and of reaching effective intratumoral concentrations, we might think of E-cadherin antagonists as molecules that act in cooperation with conventional antitumoral drugs in such a way that they could enhance their effectiveness by increasing their permeation and penetration inside tumor masses. As they interfere with E-cadherin function, at a suitable concentration they might be able to loosen solid tumors cell-cell contacts, thus modifying the diffusive properties of the tumor environment.

Finally, not only do potential E-cadherin antagonists represent viable cell adhesion modulators, but also their possible use might not be limited to pharmacology and clinical treatments. In fact, they might find application as chemical tools in future academic studies on the properties and functions of adherens junctions, such as for instance in mechanobiology studies.

Drug Discovery on E-cadherin

To date, despite its relevance as a tumor suppressor gene, pro-oncotic and potential biomarker of malignancy, E-cadherin is still a poorly explored pharmacological target since the greatest efforts towards the discovery and the development of classical cadherin binding molecules have been mainly directed towards N- and P-cadherin [86], [87].

In 1990, Blaschuk et al. [88] demonstrated that synthetic peptides that contained the HAV motif in their sequence were able to abrogate E-cadherin dependent mouse embryo compaction and to interfere with N-cadherin-mediated neurite outgrowth in astrocytes. This initial discovery and its further development [88], [89], [90] paved the way for the formulation of the anti-adhesion drug model, which is based on the notion that the destabilization of the adhesive function of classical cadherins can help fight cancer progression by promoting the apoptosis of cancer cells [91]. However, based on the considerations expressed in the previous section, it is evident that this vision is largely surpassed by the observation that tumor cells require, and employ based on their needs, adhesion and anti-adhesion. Therefore, anti-adhesion drug molecules might still find practical applications in the contexts where their dissociative action acts in coordination with chemotherapeutic drugs to enhance their cytotoxic power.

The identification of N- and E-cadherin antagonists [86], [90], allowed the development of the first N-cadherin antagonist, ADH-1 (ExherinTM), which has been demonstrated to be effective against tumor vasculogenesis, and thus can determine tumor cell apoptosis as well as metastasis development [92], [93], [94], [95].

ADH-1 is a cyclic peptide composed of the HAV motif flanked by two cysteine residues that has been shown to possess anti-cancer activity, especially in gynecological cancers. It has been tested in phase I clinical trials where it has shown an acceptable toxicity profile, a mean half-life of 20 minutes in the bloodstream and a mean terminal phase of 2.2 hours [96], [97], [98].

Despite the promising results of phase I clinical trials, ADH-1 was demonstrated to have no perceivable effect in melanoma cases when combined with melphalan after 12 weeks of clinical trials compared to patient that were administered with melphalan only [99].

To date, the exact mechanism of action of ADH-1 is still unclear as the only published paper on the topic is a computational study [100] where ADH-1 was docked into the N-cadherin adhesion pocket. However, given the rigid nature of the docking process and the highly dynamic dimerization mechanism of classical cadherins, these computational results do not provide a direct evidence of the binding mechanism and do not consider other binding sites or multiple conformational states of N-cadherin.

Recently, Mrozik et al. [101] used the structure of ADH-1 to perform a small molecule library screen and identified a novel mimetic compound that was characterized *in vitro* and *in vivo* in mouse models of multiple myeloma. The compound was found to increase the activity of bortezomib in a synergistic way.

If we exclude a publication of Senoo et al. [102] where an SPR-based assay was used to discover small molecules that are able to inhibit P-cadherin adhesion *in vitro*, the drug discovery approaches on P-cadherin have mainly relied on the use of monoclonal antibodies. Such antibodies are either able to interfere with P-cadherin adhesive functions *in vitro* [103] or consist of anti-P-cadherin antibodies conjugated with a microtubule-disrupting agent to deliver antitumoral drugs to cancer cells [104].

The drug discovery process on E-cadherin started from an *in silico* virtual screening based on the tetrapeptide sequence DWVI of its adhesion arm [105]. This study aimed at the individuation of peptidomimetic compounds with general formula NH_3^+ -Asp-scaffold-Ile-NHCH₃ that could interfere with E-cadherin dimerization and modulate cadherin mediated cell-cell adhesion. From this screening, three peptidomimetic compounds that showed comparable and better results than ADH-1 at 1 and 2 mM concentrations in epithelial ovarian cancer cell lines were identified. In particular, the best performing molecule, named FR159 (Figure 12, panel B), is the first small molecule to be crystallized with E-cadherin (PDBID: 4ZTE, [50]).

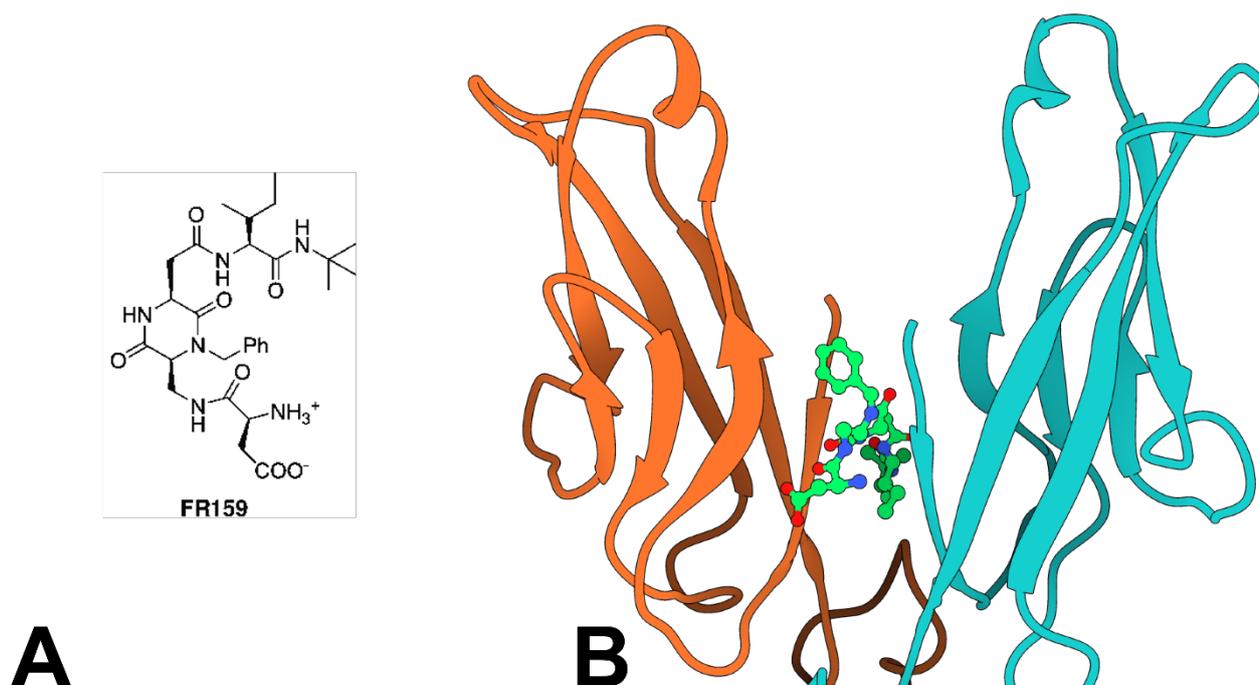


Figure 12: FR159 V3-E-cadherin co-crystal structure

Panel A shows the chemical structure of FR159 (Ph stands for phenyl, adapted from [106]), while panel B depicts FR159 that binds the E-cadherin X-dimer conformation at the level of the hydrophobic pocket at the bottom of the adhesion arm. Cadherin molecules are shown in orange and cyan cartoons, while FR159 is drawn in ball and sticks with green carbon atoms, blue nitrogens, and red oxygens. The picture was obtained from the co-crystal structure deposited in with PDBID: 4ZTE.

Contrary to expectations, FR159 was not found to be bound to the protein with the phenyl ring fitting inside the adhesion pocket, but to interact with the two E-cadherin molecules in the X-dimer conformation at the level of the hydrophobic pocket formed near the diproline motifs (Figure 12, panel B). Here, the phenyl ring of the peptidomimetic compound fits into the hydrophobic pocket lined by residues Ile4, Pro5, Pro6, Ile7 and Val22. Here, the tert-butyl moiety is rather disordered and makes van der Waals contacts with Thr97 and Pro10. At the same time, its Asp side chain forms water-mediated hydrogen bonds with the carbonyl group of Glu13, and the carbonyl group in the heterocycle forms a hydrogen bond with the amide of Ser8.

Obtaining co-crystal structures with the intact E-cadherin adhesion arm was not possible and FR159 was co-crystallized with a truncated version of E-cadherin (named from here on V3-E-cadherin because it has been obtained through the elimination of the Asp1 and Trp2 residues). Interestingly, given the pseudo-C2 symmetry of E-cadherin X-dimer crystals, FR159 has been found in two alternative conformations, which have been refined with 50% occupancy each.

In a later study [106], a high-throughput docking screening has been performed on two sets of commercially available compounds that were docked to the crystal structure of human E-cadherin-EC1EC2 in X-dimer conformation. Molecular dynamics simulations were used to inspect the 1000 best-scored compounds visually. Then, based on a similarity cluster analysis, this initial pool was reduced to 200 candidate compounds, and finally 15 of them were selected to be tested experimentally, as representative of each obtained cluster.

To investigate the impact of the virtual screening-derived library of small molecules on cadherin-dependent cell–cell adhesion, the ability to interfere with the capacity of adenocarcinoma cells (BxPC-3) expressing E-cadherin and P-cadherin and forming compact spheroids was analyzed. Using an inhibitor concentration of 1 mM, i.e., the lower active concentration determined for FR159, 5 of the 15 identified compounds (AS2, AS8, AS9, AS11, and LC11) were found to affect BxPC-3 cell–cell adhesion, although some of them showed solubility problems.

To avoid solubility issues, cell-cell adhesion assays were performed at an inhibitor concentration of 0.05 mM: at this concentration, three molecules (AS11, AS9, and, to a lesser extent, AS8, shown in Figure 13) retained anti-cell-cell adhesion activity.

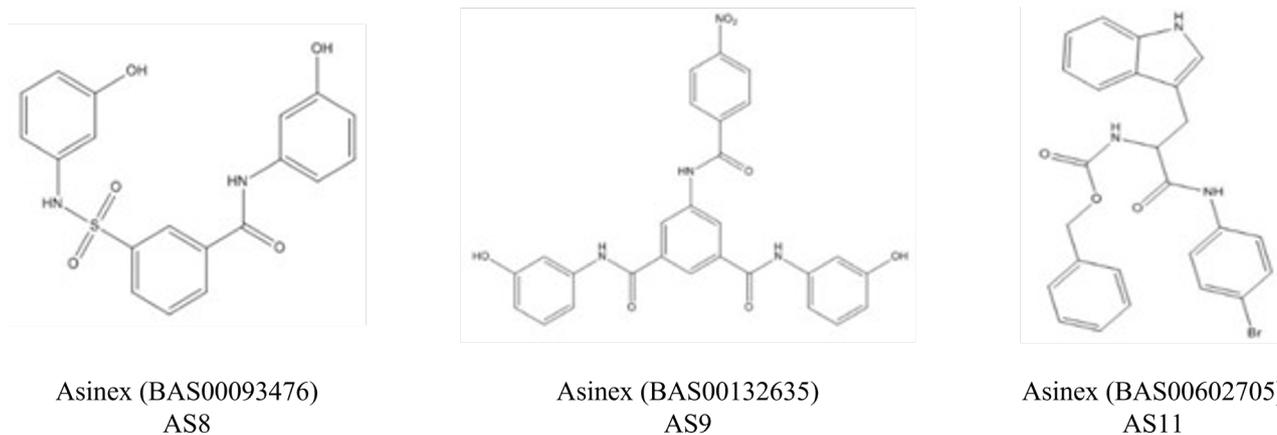


Figure 13: Chemical structures of AS8, AS9 and AS11

The quantification of the spheroid areas confirmed that treatment with AS11, AS9 and AS8 promoted the formation of less compact spheroids than when cells were treated with DMSO (~ 100 %, 75 % and 25 % more compared with DMSO for AS11, AS9, and AS8 respectively). To determine which of the two cadherins is impacted by the compounds, P-cadherin or E-cadherin were stably knocked down by RNA interference, and cell-cell adhesion assays were performed on the obtained cell models. Both AS09 and AS11 efficiently impaired BxPC-3 E-cadherin cell-cell adhesion at 0.05 mM, while at the same concentration they had no effect on cells expressing only P-cadherin, and without affecting cell viability, as demonstrated by trypan blue exclusion assay.

AS9 affected BxPC-3 E-cadherin cell aggregation less efficiently than AS11 for whichever concentration was used. Indeed, while both AS9 and AS11 showed a clear dose-response relationship when tested at different concentrations (0.05 and 0.1 mM), based on the area of the spheroids formed by both BxPC-3 E-cadherin/P-cadherin and BxPC-3 E-cadherin cells after 24 h of incubation with the two compounds, AS11 showed greater potency than AS9.

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-dimer.

The successful identification of the first E-cadherin inhibitors has led us to the conclusion that, despite their conformational flexibility, cadherins are druggable proteins and has encouraged us explore further approaches to identify novel modulators of E-cadherin-mediated cell-cell adhesion. Our latest efforts are described in the following sections.

Aim of the Project

The aim of the project is to identify through a fragment-based drug discovery approach (FBDD) novel small chemical fragments as potential E-cadherin binders and to characterize their binding mode via X-ray crystallography. To this end, I expressed and purified wild-type E-cadherin EC1-EC2 protein. The protein served as target in a FBDD screening using a library of small fluorinated chemical fragments and ^{19}F -NMR spectra as the readout method. The FBDD screening was performed by Dr. Marina Veronesi at the D3 PharmaChemistry division of the Italian Institute of Technology (IIT) in Genova. The screening allowed the identification of five potential binders (hits), which were then used in the following crystallization experiments with the truncated form of E-cadherin lacking the first two amino acids at the N-terminus (V3-E-cadherin-EC1-EC2) that I produced for X-ray crystallography purposes. I obtained three high resolution crystal structures of different cadherin-small fragment complexes explaining the binding interaction of the hit compounds. The fragments were also tested in cell adhesion and cell invasion assays in order 1) to measure their ability to modulate E-cadherin-mediated cell-cell adhesion and 2) to assess their specificity toward different members of the classical cadherin family. Cell adhesion and cell invasion assays were performed in the laboratory of Prof. Frédéric André at the University of Marseille (France). The details of the experiments and the results are described in the following sections.

Materials and Methods

Cloning, Protein Expression and Purification of E-cadherin and V3-E-cadherin Constructs

The two DNA constructs of wild-type E-cadherin EC1-EC2 and V3-E-cadherin EC1-E2, a construct lacking Asp1 and Trp2 at the N-terminus, were already present in the lab as they had been previously cloned into pET-3a expression vectors and used in other publications of Dr. Parisini's group [41], [50]. Both E-cadherin fragments are fused at their N-terminus to a 6xHis-tag, a spacer peptide (Ser-Ser-Gly-His-Ile) and the enterokinase recognition site (Asp-Asp-Asp-Asp-Lys).

The two constructs were used to transform BL21(DE3) pLysS *E. coli* cells (Invitrogen, Carlsbad, CA, USA). The transformed cells were cultured at 37°C until OD₆₀₀ was equal to 0.6. Protein expression was then induced with 1 mM of IPTG, and cells were cultured overnight at 25°C and 180 rpm.

Cells were separated from the exhausted medium by centrifugation at 10,000 rpm and 4°C. Cell pellets were then resuspended in a 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM CaCl₂ buffer supplemented with 0.2 mM PMSF (Sigma Aldrich, St. Luis, MO, USA), DNase (Sigma Aldrich, St. Luis, MO, USA), 0.5 mM β-mercaptoethanol (Sigma Aldrich, St. Luis, MO, USA), sonicated and further centrifuged at 15,000 rpm and 4°C to separate the soluble fractions from the cellular debris. Cell lysates were loaded into a Ni-NTA column (Qiagen, Venlo, The Netherlands) for an affinity purification run. After incubation, the column was extensively washed with 40 column volumes (CV) of 50 mM Tris/HCl pH 7.5, 150 mM NaCl, 2 mM CaCl₂ and 1 CV of 50 mM Tris/HCl pH 7.5, 150 mM NaCl, 2 mM CaCl₂, 10 mM imidazole buffer. Finally, the His tagged protein was eluted from the column using a 50 mM Tris/HCl pH 7.5, 150 mM NaCl, 2 mM CaCl₂, 400 mM imidazole elution buffer.

A second purification step was done using a HiPrep 26/60 Sephacryl 100 HR size-exclusion chromatography column (GE Healthcare, Little Chalfont, Buckinghamshire, UK) previously equilibrated with the running buffer: 50 mM Tris/HCl pH 7.5, 150 mM NaCl, 2 mM CaCl₂.

Further CaCl₂ was then added to the protein to reach 20 mM final concentration prior to digestion with enterokinase (New England Biolabs, Ipswich, MA, USA) overnight at room temperature. The digested fractions were separated from the undigested ones and from the cleaved His-tag by passing the sample over a Ni-NTA column following the same procedure described for the first purification step.

NMR Fragment-Based Screening

According with the works of Vulpetti [107] and Dalvit [108] we decided to perform a ^{19}F NMR binding screening on the wt human E-Cadherin EC1-EC2 testing the Local Environmental Fluorine (LEF) compounds library in house at the D3 PharmaChemistry division of the Italian Institute of Technology (IIT) in Genova. The library was previously built clustering about 7000 commercial fluorinated compounds as function of the fluorine environment and selecting about 1200 commercial fluorinated compounds (^{19}F -cpds) containing CF or CF_3 moieties. Out of the 1200 ordered compounds, about 800 compounds passed the SPAM filter [109] showing solubility $> 300 \mu\text{M}$ in PBS (pH 7.4) with 10% D_2O and $5 \mu\text{M}$ EDTA, purity $> 75\%$ and lack of aggregation, which are crucial features for obtaining reliable hits and avoid false positives and/or negatives in the binding assays. These 800 compounds constituted the LEF-library.

All NMR screening experiments were recorded at 298 K with a Bruker FT NMR Avance III 600 MHz spectrometer, equipped with a 5 mm CryoProbe QCI $^1\text{H}/^{19}\text{F}-^{13}\text{C}/^{15}\text{N}-\text{D}$ quadruple resonance, a shielded z-gradient coil, and an automatic sample changer SampleJet NMR system. All NMR samples were in 50 mM Tris HCl pH 8.0, 150 mM NaCl, 2 mM CaCl_2 , 10% D_2O for the lock signal using at maximum 2% of $\text{DMSO-}d_6$. For each samples a 1D ^{19}F and two ^{19}F - R_2 filter have been recorded, all with an acquisition time of 0.8 s, relaxation delay (d1) of 5 s, number of scans of 128 for CF_3 mixtures and 256 for CF mixtures, and a spectral width of 80 ppm and of 40 ppm for the CF and CF_3 mixtures, respectively. The R_2 filter experiments were recorded with the Carr-Purcell-Meibom-Gill [110] scheme with a time interval of 23.5 ms between the 180° pulses and with different total lengths (94 and 282 ms). The spectra were acquired with proton decoupling using the Waltz-16 composite pulse sequence with a 90° pulse of $120 \mu\text{s}$. Chemical shifts are referenced to the CFCl_3 signal in water.

Crystal structures of the V3-E-cadherin-small fragment complexes

Co-crystals of V3-E-cadherin and the best hits from the FBDD screening, ARN1577, ARN1883 and ARN1512, were obtained by the hanging drop vapor diffusion method using VDXm 24-well plates and siliconized glass coverslips (Hampton Research, Aliso Viejo, California, USA).

In each experiment, $1 \mu\text{L}$ of a 10 mg/mL V3-E-cadherin sample solution was mixed with an equal volume of crystallization buffer and with $1 \mu\text{L}$ of a 6 mM solution of ARN1512 or ARN1577 or, as in the case of ARN1883, with $1 \mu\text{L}$ of a 100 mM stock solution of the compound. The droplet was left equilibrating against a 1 mL reservoir at room temperature, yielding crystals of suitable dimensions for X-ray data collection within a few days.

Crystals grew in all the tested conditions comprising varying concentrations of ammonium sulfate ranging from 1.1 M to 1.6 M, 80 mM CaCl₂ and 0.1 M of Tris-HCl pH 7.5. Crystals were flash-frozen in 25% (v/v) glycerol for X-ray diffraction experiments.

X-Ray Diffraction Data Acquisition and Structure Determination

High-resolution datasets were collected for the three complex structures at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) using a $\lambda = 1.000 \text{ \AA}$ (beamline X06DA-PXIII).

Diffraction images were processed using iMosflm [111] and scaled with AIMLESS in the CCP4 program suite (Oxford, UK) [112]. Molecular replacement and model refinement were done using Phenix (Berkeley, CA, USA) [113] and the V3-E-cadherin X-dimer structure (PDBID: 4ZT1) as starting model. During refinements, manual adjustments to the model were performed with Coot [114].

Data collection and refinement statistics for the two crystal structures are shown in Table 3. Molecular graphics and analyses were done using UCSF Chimera (University of San Francisco, San Francisco, CA, USA) [115] and Pymol (Schrödinger Inc., New York, NY, USA) [116].

Table 3: Data collection and refinement statistics for the three co-crystal structures

Crystallographic Table			
Compound	<i>ARN1512</i>	<i>ARN1577</i>	<i>ARN1883</i>
Protein Data Bank ID	<i>7QW0</i>	<i>7QW1</i>	<i>7QW2</i>
Wavelength	<i>1.000</i>	<i>1.000</i>	<i>1.000</i>
Resolution range	<i>51.9-1.75 (1.81-1.75)</i>	<i>39.68-1.5 (1.55-1.5)</i>	<i>38.48-1.65 (1.71-1.65)</i>
Space group	<i>C 1 2 1</i>	<i>C 1 2 1</i>	<i>C 1 2 1</i>
Cell dimensions			
a, b, c (Å)	<i>120.44, 77.1, 72.73</i>	<i>121.86, 79.36, 73.03</i>	<i>121.37, 76.96, 72.64</i>
α, β, γ (°)	<i>90, 110.43, 90</i>	<i>90, 114.41, 90</i>	<i>90, 110.7, 90</i>
Total reflections	<i>122,553 (11963)</i>	<i>201,131 (20113)</i>	<i>149,526 (14953)</i>
Unique reflections	<i>61,431 (5986)</i>	<i>101,191 (10105)</i>	<i>75,151 (7487)</i>
Multiplicity	<i>6.1 (5.6)</i>	<i>6.1 (5.8)</i>	<i>6.1 (5.6)</i>
Completeness (%)	<i>98.1 (96.8)</i>	<i>100.0 (100.0)</i>	<i>100.0 (100.0)</i>
Mean I/sigma(I)	<i>5.4 (1.4)</i>	<i>10.3 (0.9)</i>	<i>9.8 (1.0)</i>
Wilson B-factor	<i>21.92</i>	<i>20.29</i>	<i>18.17</i>
R_{merge}	<i>0.208 (1.32)</i>	<i>0.080 (1.96)</i>	<i>0.126 (1.99)</i>
CC_{1/2}	<i>0.969 (0.635)</i>	<i>0.999 (0.541)</i>	<i>0.998 (0.468)</i>
Reflections used in refinement	<i>61,392 (5986)</i>	<i>101,091 (10077)</i>	<i>75,117 (7481)</i>
Reflections used for R_{free}	<i>3091 (308)</i>	<i>5178 (508)</i>	<i>3771 (344)</i>
R_{work}	<i>0.1728 (0.2762)</i>	<i>0.1995 (0.4302)</i>	<i>0.1916 (0.3320)</i>
R_{free}	<i>0.2023 (0.3235)</i>	<i>0.2257 (0.4529)</i>	<i>0.2171 (0.3403)</i>
Total number of non-H atoms	<i>3919</i>	<i>3881</i>	<i>3788</i>
macromolecules	<i>3248</i>	<i>3262</i>	<i>3219</i>
ligands	<i>128</i>	<i>96</i>	<i>82</i>
solvent	<i>543</i>	<i>523</i>	<i>487</i>
Protein residues	<i>421</i>	<i>420</i>	<i>420</i>
RMS			
bonds (Å), angles (°)	<i>0.017, 1.51</i>	<i>0.014, 1.38</i>	<i>0.015, 1.44</i>
Ramachandran plot			
favored, allowed, outliers (%)	<i>97.12, 2.88, 0.00</i>	<i>97.36, 2.64, 0.00</i>	<i>97.36, 2.64, 0.00</i>
Rotamer outliers (%)	<i>0.27</i>	<i>0.26</i>	<i>0.27</i>
Clashscore	<i>4.19</i>	<i>3.75</i>	<i>3.39</i>
Average B-factor	<i>33.75</i>	<i>36.20</i>	<i>31.66</i>
macromolecules	<i>31.64</i>	<i>34.62</i>	<i>29.74</i>
ligands	<i>54.17</i>	<i>61.36</i>	<i>69.40</i>
solvent	<i>41.57</i>	<i>41.46</i>	<i>37.96</i>
<i>Statistics for the highest-resolution shell are shown in parentheses.</i>			

Cell Adhesion and Cell Invasion Assays

The human pancreatic BxPC-3 cell line was routinely cultured in DMEM/10% fetal calf serum (FCS) as previously described [117] and authenticated using short tandem repeat (STR) profiling (ATCC). BxPC-3 cells were used as a model system since these cells express high levels of both E-cadherin and P-cadherin at cell–cell contacts. E- and P-cadherin were stably knocked down in the BxPC-3 cell line by RNA interference using mission shRNA lentiviral transduction particles (Sigma, St Quentin Fallavier, France). The generated stable BxPC-3 cell lines were called E-cadherin/P-cadherin (no cadherin depletion), E-cadherin (P-cadherin depletion), and P-cadherin (E-cadherin depletion). Cell surface cadherin extinction was assessed by both immunofluorescence and western blot for immunofluorescence detection,

The human pancreatic MiaPaca cell line was routinely cultured in RPMI 10% fetal calf serum (FCS). MiaPaca cells do not express any cadherin. We generated N-cadherin expressing cells by transfecting cells with a plasmid encoding EGFP tagged N-cadherin (Addgene). Mock cells transfected with plasmid encoding EGFP. Stable clones were obtained after limited dilution and genistein selection. All cell lines were cultured in the laboratory for no more than 10 passages and were tested for *Mycoplasma* every 3 weeks.

A spheroid formation assay was used to investigate the effect of inhibitors on cadherin-mediated cell–cell adhesion properties. Isolated BxPC-3 cells were seeded onto U-bottom untreated tissue culture 96-well plates at a concentration of 5000 cells per well in 100 μ l DMEM containing 10% FCS and 0.24 % methylcellulose supplemented or not with inhibitors. Spheroids were then incubated for 24 h in a temperature- and CO₂-controlled chamber. Images were taken using an Olympus microscope equipped with a \times 4 objective. The size of the spheroids was quantified by measuring the area occupied by cells using image J software [118]. Each tested condition was done in octuplets repeated at least 2 times.

In cell invasion assays, cells were cultured for 72 h as described above to obtain compact spheroids. The latter were embedded into 1.6 mg/ml of bovine collagen I matrix (Advanced Biomatrix) diluted in 0.12% methylcellulose and covered with serum-free medium. Spheroids were then incubated for 24 h in a temperature and CO₂ controlled chamber. Images were captured 48 h after embedding using an Olympus microscope equipped with a \times 4 objective. The size of the spheroids was quantified by measuring the area occupied by cells using ImageJ software [118]. Each tested condition was done at least in sextuplet repeated 2 times

Results

E-cadherin and V3-E-cadherin Production and Purification

The productions of the E-cadherin EC1-EC2 and V3-E-cadherin EC1-EC2 proteins followed the same protocol. Hence, for simplicity, here I will show the SDS-PAGE gels and the chromatograms that were obtained for the E-cadherin EC1-E2 construct; indeed, those of the V3-E-cadherin EC1-EC2 construct are very similar, the latter being a truncated version of the former as it lacks the first two amino acids.

Overnight expression of both constructs resulted in the isolation of an average value of almost 4 grams per liter of cellular pellets after centrifugation. After the extraction of the soluble fractions, the first protein purification step was done via Ni-NTA affinity chromatography and resulted in a relatively pure sample, as assessed by SDS-PAGE analysis.

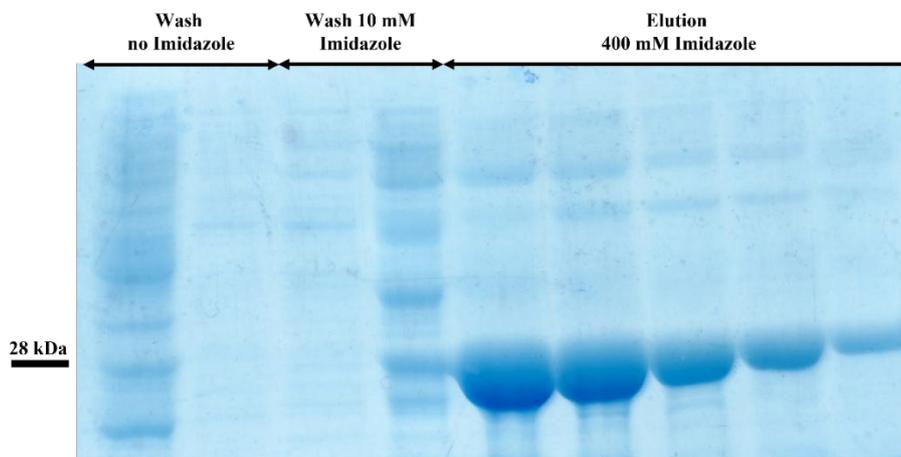


Figure 14: SDS-PAGE of the purification profile of E-cadherin EC1-2 after Ni-NTA affinity chromatography. A 28 kDa marker line is drawn on the left, while on top the content of each lane is labelled according to the different phase in the affinity chromatography process: lanes 1 and 2 correspond to the beginning and the end of the preliminary washing without imidazole, lanes 3 and 4 to the beginning and the end of the 10 mM imidazole wash, and the lanes from 5 to 9 to the first fractions of the 400 mM imidazole elution, where it is evident the presence of an enriched spot at ~30 kDa mass weight corresponding to the E-cadherin EC1-2 fragment.

Figure 14 shows the profile of the first protein purification step: clearly, the two washing steps (lane 1 to 4) removed the majority of contaminants from the cell lysate and the elution resulted in fractions enriched in the band with a molecular weight of roughly 30 kDa, which could readily be identified as E-cadherin EC1-EC2. Indeed, despite having a theoretical molecular weight of ~25 kDa, E-cadherin EC1-EC2 is known to migrate in SDS-PAGE as a 30 kDa protein. Following the Ni-NTA affinity chromatography, the isolated proteins were further purified via size-exclusion chromatography.

The elution profiles for both constructs was very similar and consisted of a two-peak chromatogram where the first peak corresponded to E-cadherin molecules that co-eluted with a higher molecular weight contaminant, while the second peak was predominant and corresponded to the substantially pure E-cadherin or V3-E-cadherin constructs, as assessed by SDS-PAGE (Figure 15).

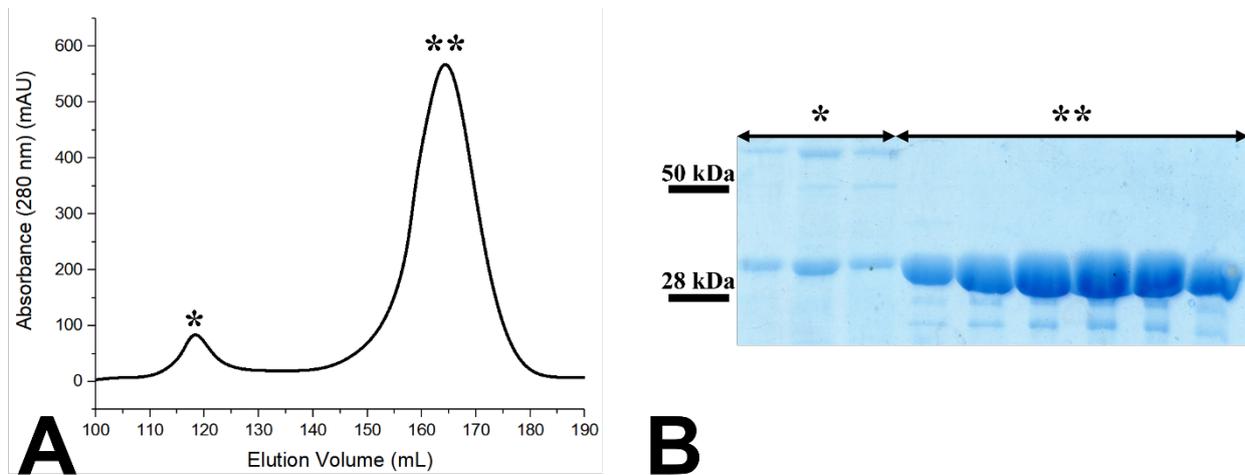


Figure 15: Size-exclusion chromatogram and corresponding SDS-PAGE gel

In panel A, the two-peaks chromatogram of E-cadherin EC1-2 fragment is displayed. On the horizontal axis is displayed the elution volume in mL, while on the vertical axis the absorbance at 280 nm is shown in milli arbitrary units (mAU).

In panel B, the corresponding SDS-PAGE gel showing the purification profile of the size-exclusion chromatography. Lanes from 1 to 3 are relative to the first elution peak (marked with one asterisk in both panels) and the remaining lanes from 4 to 9 correspond to fractions eluted under the second peak (marked with two asterisks in both panels). Mass weights of 28 kDa and 50 kDa are depicted on the left side of the SDS-PAGE gel.

The purified sample was digested with enterokinase overnight at room temperature to remove the N-terminal His-tag and produce an E-cadherin fragment that is primed for adhesion.

Upon termination of the digestion reaction, the sample was loaded onto a Ni-NTA affinity chromatography column and the flow through containing the His-tag-free digested protein was collected for further processing (Figure 16).

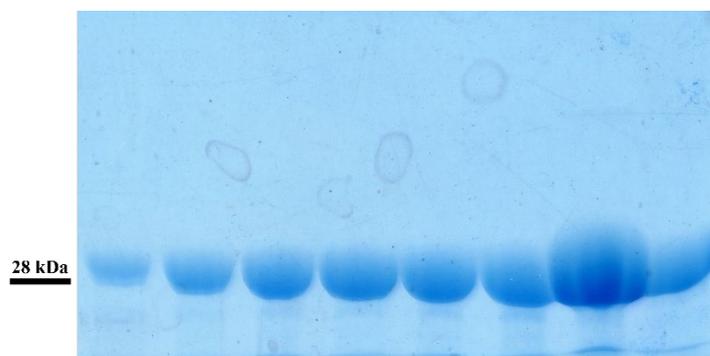


Figure 16: SDS-PAGE profile of the enterokinase digested fraction

Lanes from 1 to 8 are here shown to depict the purification profile of the Ni-NTA unbound fraction (lanes 1-6) and the eluted fraction (lanes 7-8). On the left side of the gel, the 28 kDa marker line is drawn.

The purified fractions were pooled and concentrated at 10 mg/ml. The protein thus obtained was used both for FBDD experiments (E-cadherin EC1-EC2) and in co-crystallization experiments with the fragments identified via FBDD (V3-E-cadherin EC1-EC2).

To summarize the purification profile, the SDS-PAGE gel in Figure 17 shows that the majority of recombinant protein remained in the soluble fraction after sonication and that the target proteins were obtained at a suitable purity for crystallization experiments.

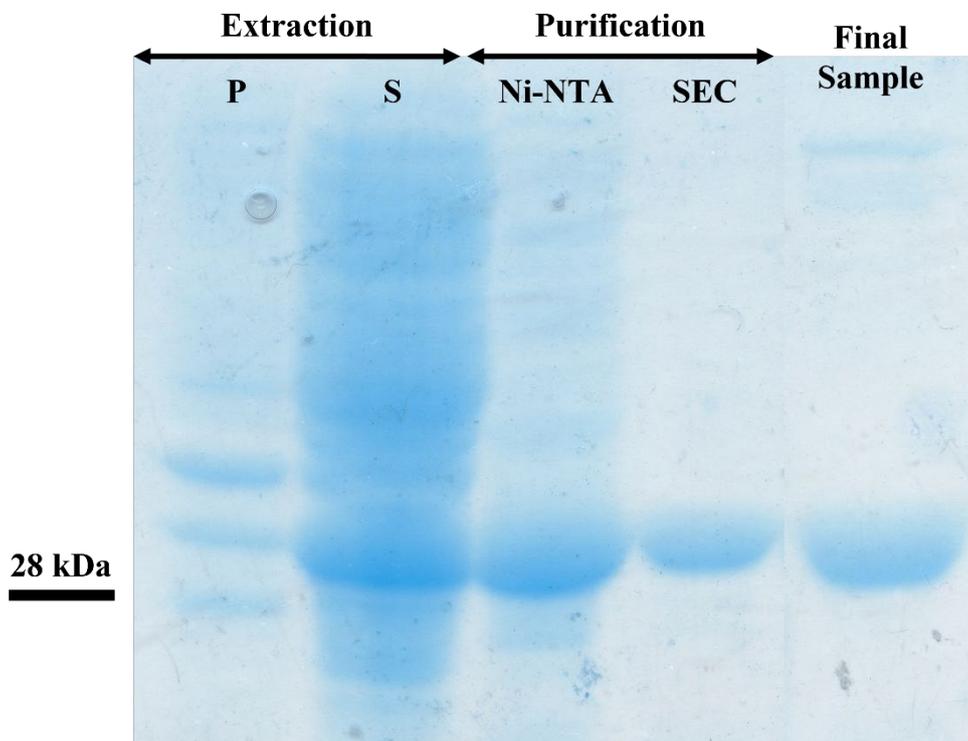


Figure 17: Summary SDS-PAGE gel showing E-cadherin EC1-2 purification profile. Lanes 1 and 2 are labelled P and S that indicate the pellet and soluble fractions obtained after the extraction by sonication, respectively. Lanes 4 and 5 are labelled Ni-NTA and SEC to indicate samples obtained during the first and second purification steps during Ni-NTA affinity chromatography and size-exclusion chromatography, respectively. The last lane, correspond to the final sample. On the left side of the gel, the 28 kDa mass weight line is drawn.

Hit Compounds from the ^{19}F NMR-based Screening

For the sake of clarity, here I will only briefly report and describe the results of the ^{19}F NMR-based small fragment screening using wt E-cadherin EC1-EC2 as target, which was done by Dr. Marina Veronesi using a library of small fluorinated chemical fragments at the D3 PharmaChemistry division of the IIT in Genova.

Firstly, the stability and the possible aggregation state of the wt E-cadherin EC1-EC2 was assessed by NMR over two days. To this end, two mixtures of 25 compounds each were tested in the absence and in the presence of 1, 2, 4 μM wt E-cadherin EC1-EC2 and recording the NMR spectra every 12 hours. In the presence of aggregated protein, it is expected that most of the compounds show a binding effect: in our experiment, the substantially identical NMR spectra recorded at different concentrations of the protein (Figure 18) indicate that the protein is not aggregated. Hence, since the protein appeared to be stable in our experimental condition, it could be safely used in FBDD screening using NMR.

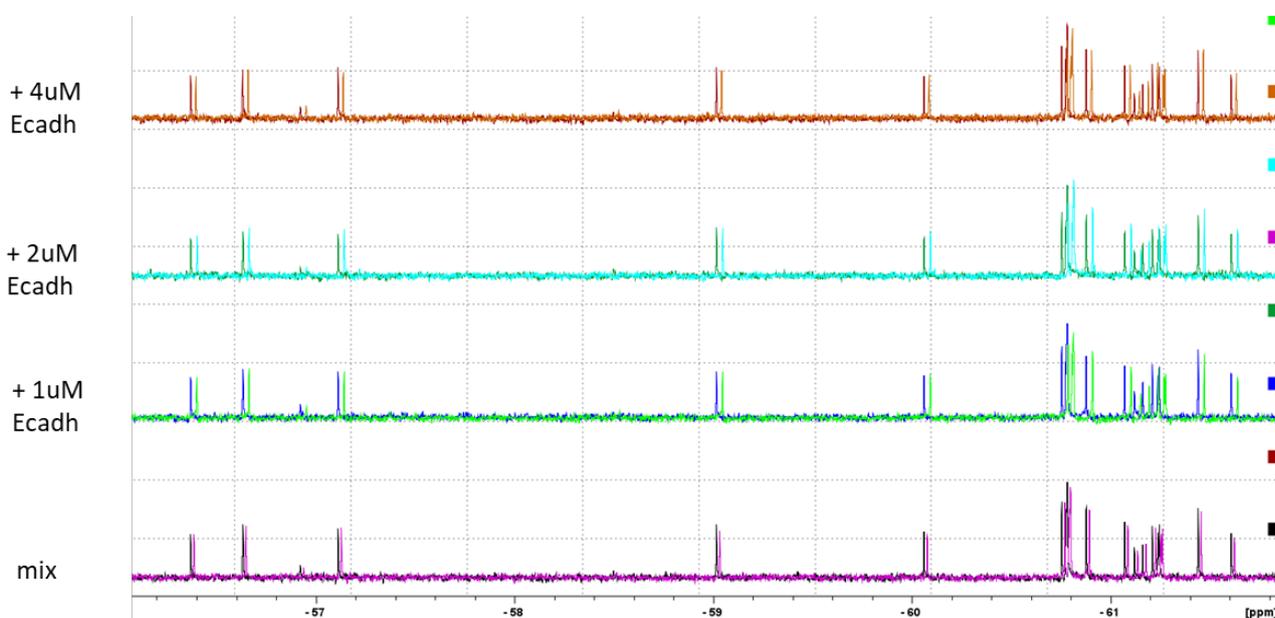


Figure 18: NMR spectra recorded in the presence of 25 compounds and of 1, 2, and 4 μM wt E-cadherin EC1-EC2. The picture shows the NMR spectra a mix of 25 compounds recorded in the presence and absence of wt E-cadherin EC1-2 at different concentrations, showing that the NMR signal is unchanged over time and indicating the absence of aggregation of the protein.

Therefore, we first screened the LEF library in mixtures of 25 compounds each, at the concentration of 20 μM (CF_3 labeled compounds) and 40 μM (CF labeled compounds) in the presence and in the absence of 2 μM wt E-cadherin EC1-EC2. About 50 compounds showed a line broadening of their ^{19}F NMR signal in the presence of protein (Figure 19), indicating a binding event.

^{19}F $T_2\rho$ binding experiments

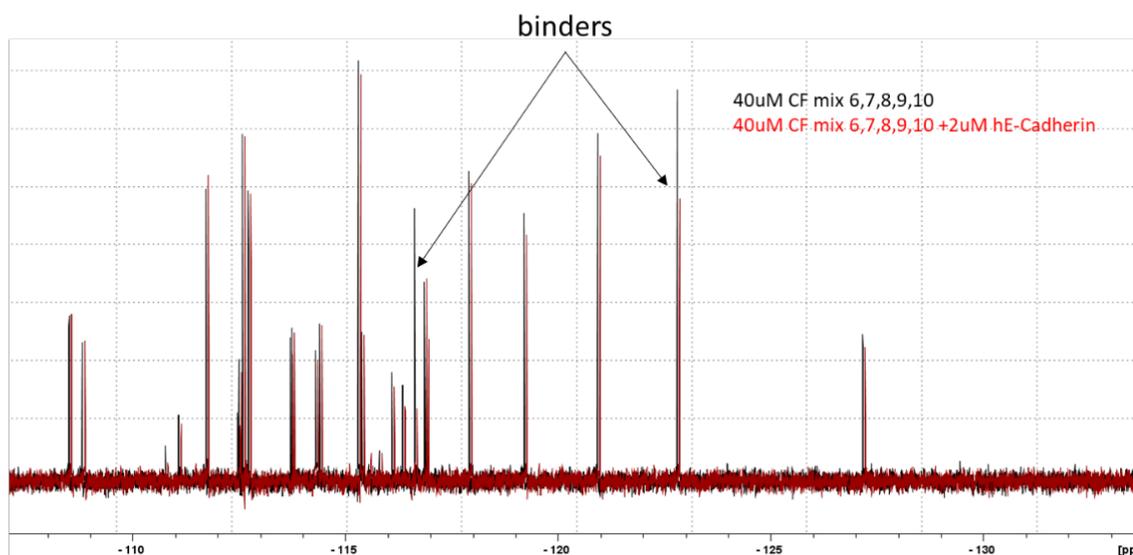


Figure 19: Line broadening of the ^{19}F NMR signal

To confirm that the compounds were hits, the putative binding compounds were then deconvoluted and the individual compounds were tested in the presence of two non-binding compounds (negative control) in the same experimental condition of the screening. As depicted in Figure 20 the compound ARN1512 showed an evident line broadening of its ^{19}F NMR signal in the presence of 2 μM wt E-cadherin EC1-EC2 whereas no variation in the NMR signals was visible for the two negative controls, indicating that ARN1552 binds to wt E-cadherin EC1-EC2.

^{19}F $T_2\rho$ binding experiments, single compound

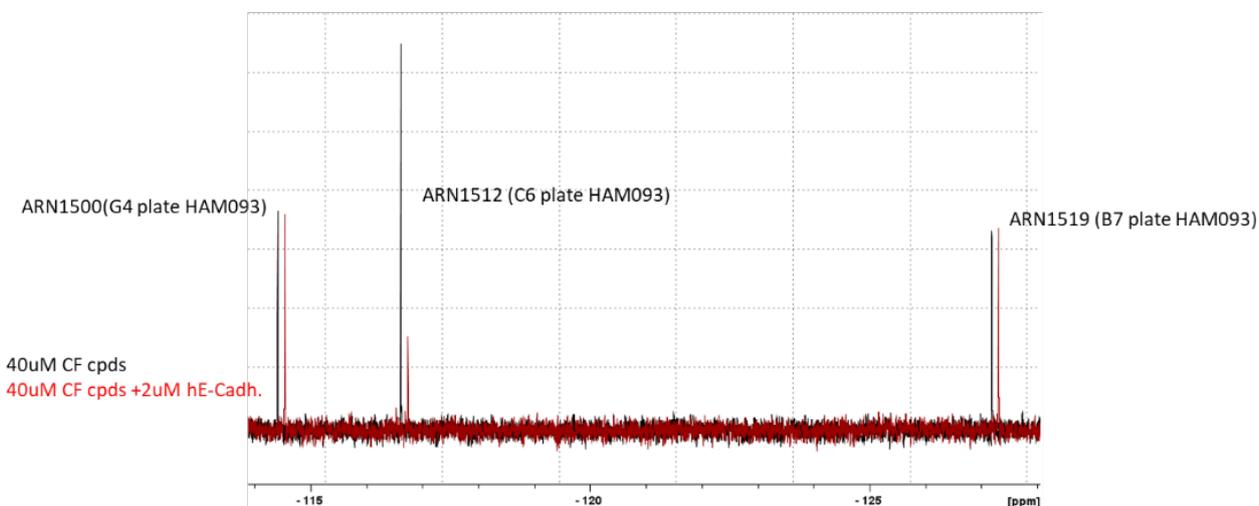


Figure 20: Line broadening of ARN1512

As further confirmation of this binding event, six compounds were tested in the presence of an increasing concentration of wt E-cadherin EC1-EC2, to test a possible dose-response effect. In general, the line broadening of the ^{19}F signal of a binder is proportional to $[\text{EL}]/[\text{L}_{\text{TOT}}]$ ratio, where EL is the concentration of the Ligand-Protein complex and L_{TOT} is the total concentration of the Ligand. Four of these six compounds showed a clear dose effect response (an example in Figure 21) and were therefore considered real hits and used for further experiments, whereas the other two, not showing a dose response effect were discarded. The confirmation of the other putative hits is ongoing.

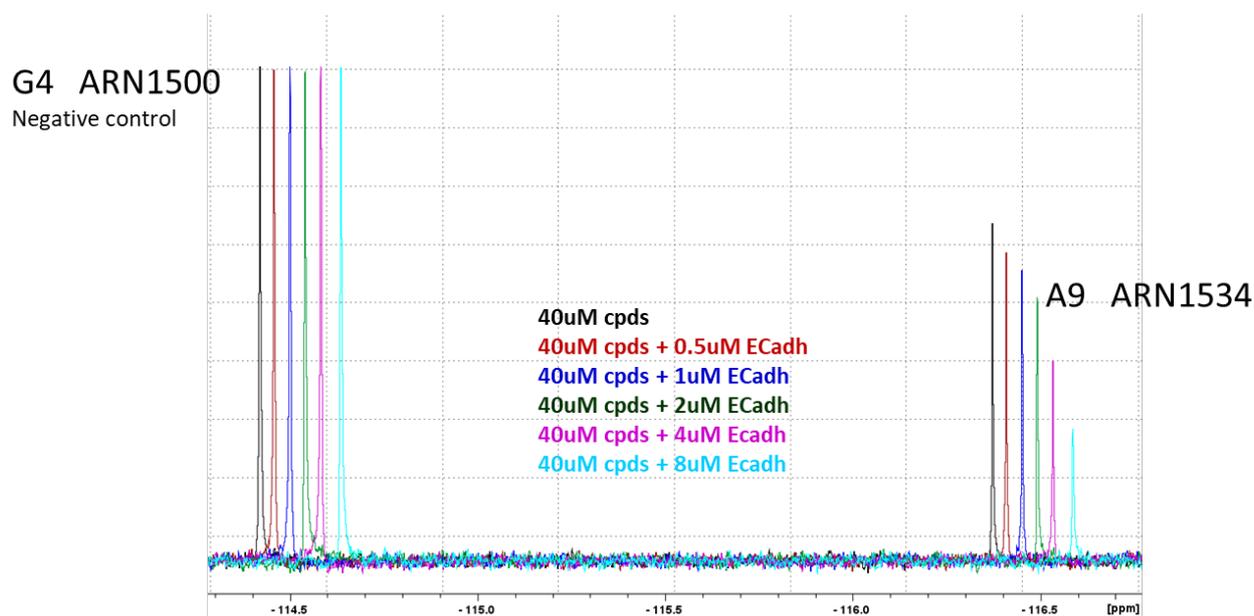
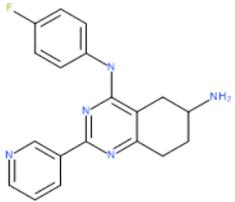
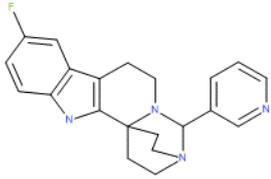


Figure 21: NMR dose-response curve of ARN1534

Table 4 reports in a synthetic way the results of the hits characterized by the greatest chemical shift during the screening. It also lists some properties of the hit compounds such as identifier, mass weight, and molecular structure and some NMR-screening parameters such as the effect of the hit compounds in mixture, the effect of the compound in single composite-pulse decoupling and the ^{19}F chemical shift.

Table 4: Results of the ^{19}F NMR-based screening.

Compound Identifier	Mass Weight	Molecular Structure	^{19}F -NMR Effect in Mixtures	^{19}F -NMR Effect Single Composite-Pulse Decoupling	^{19}F Chemical Shift
ARN1577	289.352		Medium shift signal	Confirmed weak-medium signal	-122.83
ARN1512	408.306		Strong shift signal	Confirmed strong signal	-116.64
ARN1883	348.423		Medium-strong shift signal	Confirmed medium-strong signal	-123.86

Based on these results, we set out to identify the mode of binding of these hits with E-cadherin by X-ray crystallography. To this end, we used the V3-E-cadherin-EC1-EC2 construct, i.e. the protein lacking the first two aminoacids at the N-terminus. The use of the V3-E-cadherin EC1-EC2 fragment reflects the same strategy that allowed the crystallization of E-cadherin with FR159, the first crystal structure of a complex of a E-cadherin extracellular portion with a small molecule [50]. In essence, this strategy provides two major advantages: 1) it leaves the Trp2 pocket constitutively open, thus eliminating the competition between the Trp2 and the fragment for binding into the Trp2 pocket, a potentially crucial binding site for the inhibition of the cadherin dimerization mechanism. 2) it greatly limits the conformational space that can be explored by the protein, i.e. it limits its conformational flexibility, as the lack of the first two residues impair the formation of the strand exchange dimer and allows the highly dynamic E-cadherin protein to shuttle only between its monomeric state and its X-dimer state.

ARNs and V3-E-cadherin EC1-EC2 Co-Crystallization Experiments

Co-crystals of the three ARN compounds and V3-E-cadherin EC1-E2 grew in 1.1 M to 1.6 M $(\text{NH}_4)_2\text{SO}_4$, 80 mM CaCl_2 , and 0.1 M of Tris-HCl pH 7.5. Representative images of the grown crystals are shown in Figure 22.

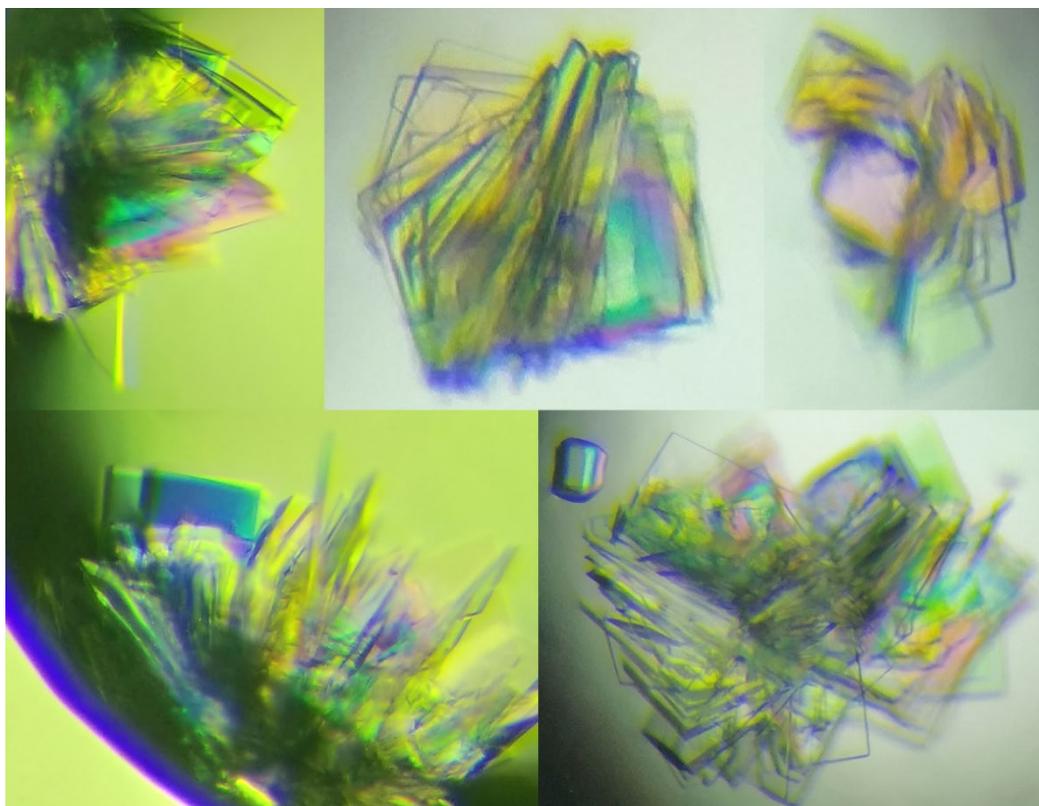


Figure 22: Crystal morphologies of V3-E-cadherin EC1-2 co-crystallized with the ARN compounds
The picture shows the crystal morphologies for the co-crystals obtained during the co-crystallization experiments. Plate-like crystals grew as clusters that could easily be separated in order to obtain single crystals for the successive diffraction experiments.

All crystals usually grew forming clusters and had roughly the same plate-like rectangular morphology. Despite their apparent multiplicity, each cluster turned out to be composed of single crystals, which, upon gentle touch, could be readily separated into single crystals suitable for X-ray diffraction experiments. Crystals were cryo-protected with 25% glycerol prior to data collection.

Structure Determination of the V3-E-cadherin EC1-EC2:ARN complex

From the crystals thus obtained, numerous high-resolution datasets were collected at the Paul Scherrer Institute in Switzerland. The best three datasets in terms of resolution cut-off, completeness, intensity-to-noise ratio, ligand occupancy and quality of the electron density map were refined until convergence.

As expected, in all crystals the V3-E-cadherin EC1-EC2 construct adopted the X-dimer conformation. Their structure showed little to no difference in terms of RMSD (0.41 Å, 1.07 Å, and 0.85 Å for the complexes with ARN1512, ARN1577 and ARN1883, respectively) compared to the unligated E-cadherin EC1-EC2 structure that is available in the protein databank (PDBID: 4ZT1), as shown in Figure 23.

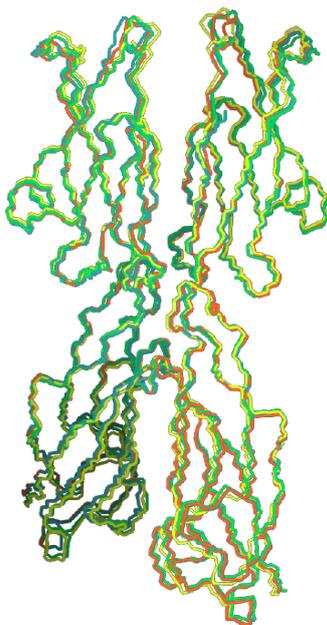


Figure 23: Structural alignment between E-cadherin X-dimer and the three obtained co-crystal structures
 Structural superimposition of the E-cadherin X-dimer (PDBID: 4ZT1) with the three ARN co-crystal structures. The backbone atoms of the four aligned proteins are drawn in wires and each protein is colored differently: V3-E-cadherin X-dimer is colored in orange, while the co-crystal structures of ARN1512_A, ARN1577_Z, and ARN1883 in cyan, green, and yellow, respectively.

Unlike FR159, which was found to bind into a hydrophobic pocket at the bottom of the adhesion arm that only forms, transiently, as the protein goes through its X-dimer conformation, the ARN compounds were all found to bind inside the acceptor pocket that is usually occupied by the Trp2 residue, which is lacking in our crystallographic construct. To provide scientific evidence for their effective localization, the omit maps drawn at 2.0 sigma level are shown in Figure 25, Figure 24, and Figure 26.

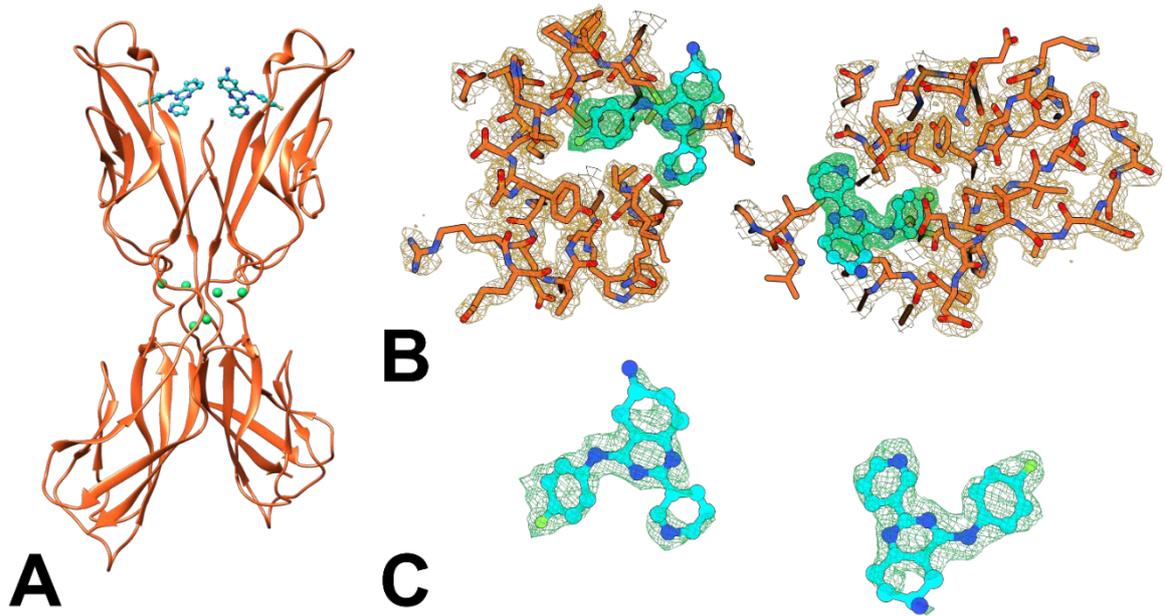


Figure 24: Omit maps of ARN1577

In panel A the V3-E-cadherin X-dimer in orange cartoon model with calcium ions in green and ARN1577 in cyan ball and stick is depicted.

Panel B display the section containing the adhesion pocket where ARN1577 has been found. The protein body is drawn in orange sticks with the corresponding electron density at 1.00 rmsd sigma level superimposed and colored in orange. The ligand is drawn in cyan ball and sticks and the electron density of its omit map is drawn at 2.00 rmsd sigma level and colored in green.

Panel C shows the details of the omit map electron density drawn in green and at 2.00 rmsd sigma level with the underlying model drawn in cyan ball and sticks.

Oxygen atoms are depicted in red, nitrogens in blue, phosphoruses in yellow and fluorines in green.

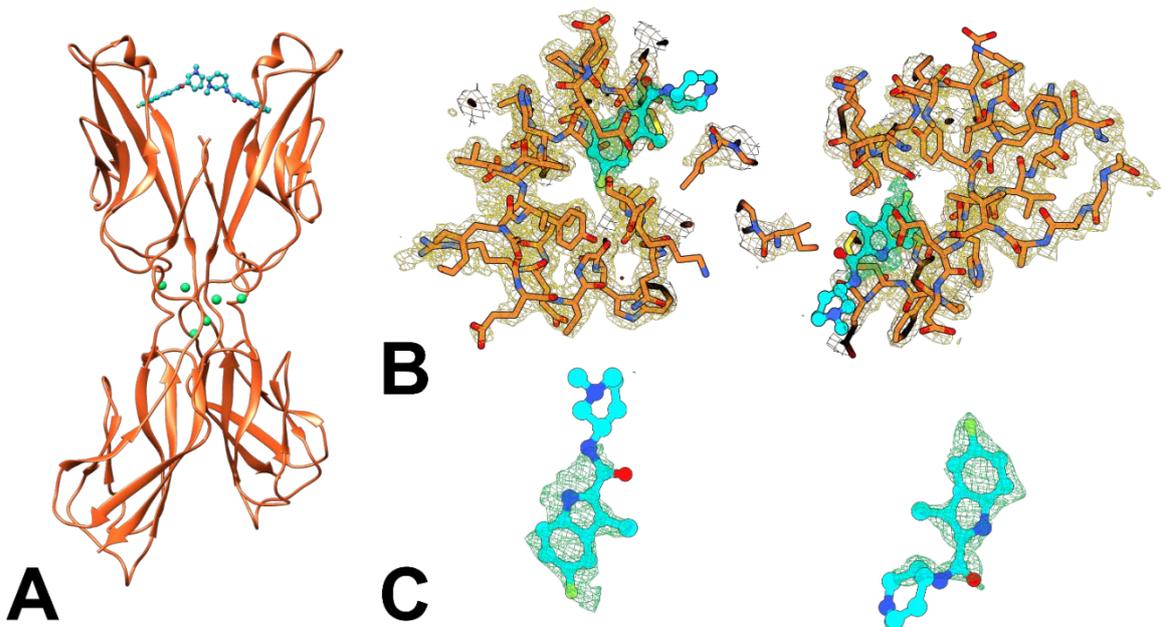


Figure 25: Omit maps of ARN1512

In panel A the V3-E-cadherin X-dimer in orange cartoon model with calcium ions in green and ARN1512 in cyan ball and stick is depicted.

Panel B display the section containing the adhesion pocket where ARN1512 has been found. The protein body is drawn in orange sticks with the corresponding electron density at 1.00 rmsd sigma level superimposed and colored in orange. The ligand is drawn in cyan ball and sticks and the electron density of its omit map is drawn at 2.00 rmsd sigma level and colored in green.

Panel C shows the details of the omit map electron density drawn in green and at 2.00 rmsd sigma level with the underlying model drawn in cyan ball and sticks.

Oxygen atoms are depicted in red, nitrogens in blue, phosphoruses in yellow and fluorines in green.

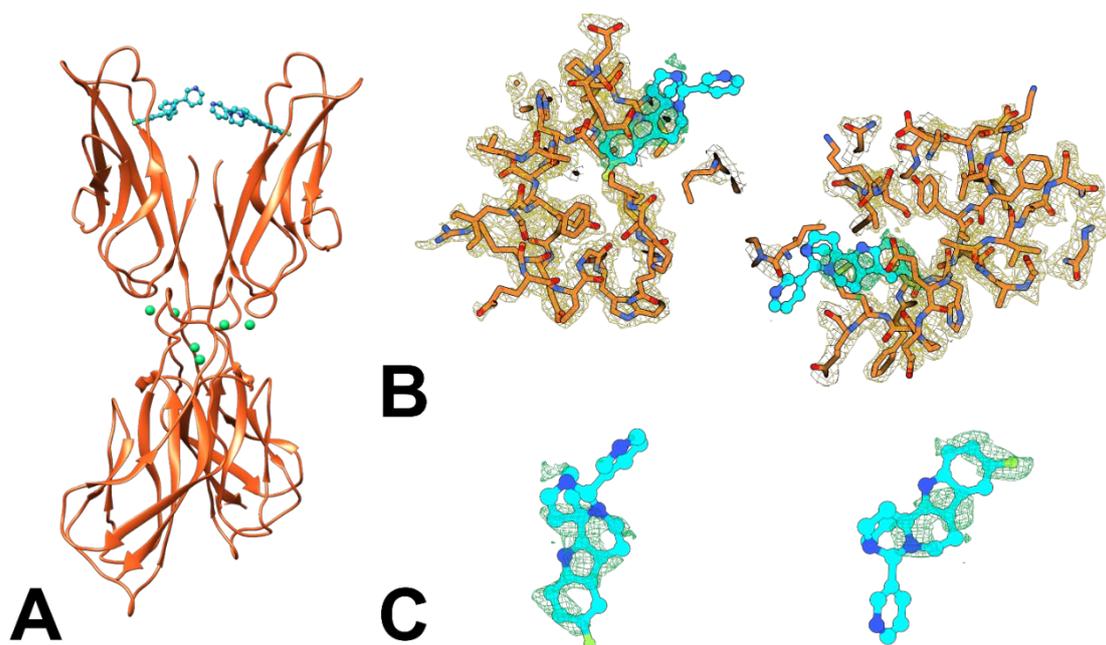


Figure 26: : Omit maps of ARN1883

In panel A the V3-E-cadherin X-dimer in orange cartoon model with calcium ions in green and ARN1883 in cyan ball and stick is depicted.

Panel B display the section containing the adhesion pocket where ARN1883 has been found. The protein body is drawn in orange sticks with the corresponding electron density at 1.00 rmsd sigma level superimposed and colored in orange. The ligand is drawn in cyan ball and sticks and the electron density of its omit map is drawn at 2.00 rmsd sigma level and colored in green.

Panel C shows the details of the omit map electron density drawn in green and at 2.00 rmsd sigma level with the underlying model drawn in cyan ball and sticks.

Oxygen atoms are depicted in red, nitrogens in blue, phosphoruses in yellow and fluorines in green.

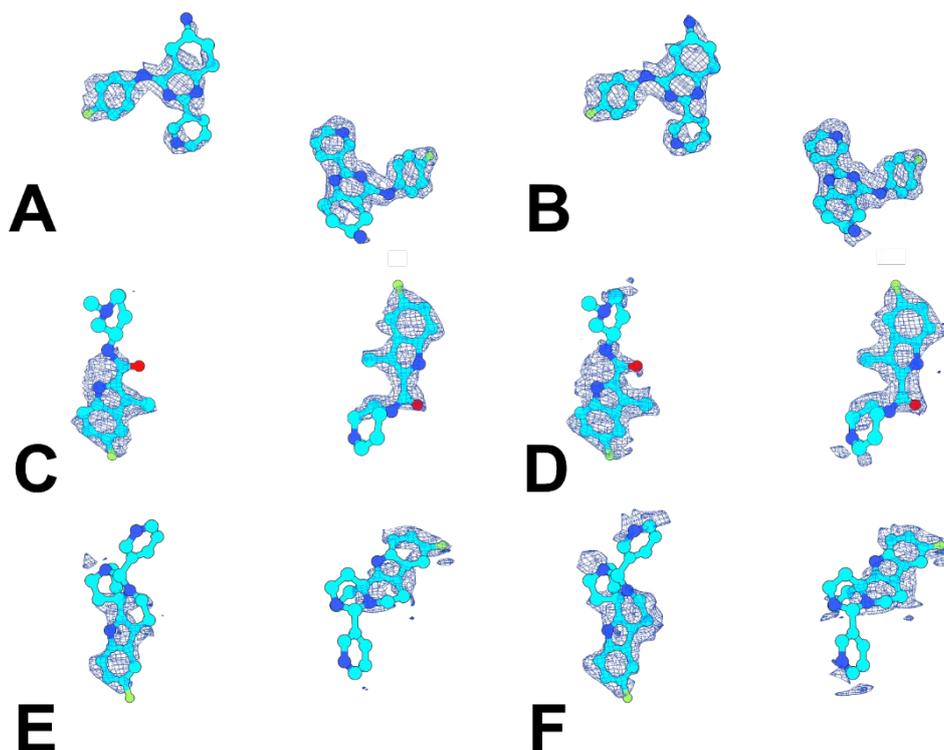


Figure 27: 2Fo-Fc electron density of the ARN ligands

The picture shows the calculated 2Fo-Fc electron density of the models in which the three ARN compounds have been inserted. All ligand electron densities are displayed with a blue mesh superimposed to the compound structures that are displayed in cyan ball and stick model with blue nitrogens and green fluorines.

Panels A, C, and E show the electron density at 0.7 rmsd sigma level, while panels B, D, and F at 0.5 rmsd sigma level.

The full models calculated upon ligand insertion feature a very good electron density coverage of the ligand molecules, whose fluorinated moiety can be fully appreciated in the electron density map at 0.7 sigma level (Figure 27 panels A, C, and E). Map coverage of the full ligand becomes almost complete at 0.5 rmsd sigma level (Figure 27, B, D, and F).

A detailed description of the mode of binding of the three binders for which we could determine the X-ray crystal structure is provided in the Discussion chapter.

Cell Adhesion and Cell Invasion Assays

To assess the ability of these fragments to modulate cell adhesion and/or cell invasion, our collaborator, Dr. Frédéric André' and his team at the Centre de Recherches en Oncologie biologique et Onco-pharmacologie of the University of Aix-Marseille (France), performed standard cell adhesion and cell adhesion functional assays.

The effects of various compounds on cell-cell adhesion of BxPC3 E-cadherin/P-cadherin cell were evaluated at 0.05 mM concentration and compared to 0.1 % DMSO control, and the measured spheroid area after 24 hours was calculated and used to compare the effect of each compound (Figure 28). The degree of spheroid compaction is caused by the expression of E- and P-cadherin on the surface of the cells; thus, when cadherin inhibition occurs, the spheroids tend to spread. Hence, an increase in the area of the spheroid is correlated with the inhibitory action of the compound. The data show that a statistically significant increase in the area is observed when ARN1512 and ARN1883 are used, while the variations in the spheroid areas for the other tested compounds is not significant. The best performing compounds, ARN1512, was also tested at different concentrations and Figure 29 shows that there is a statistically significant increase in the spheroid area at 0.05 mM, while at 0.01 mM the effect is not significant, indicating that the activity of ARN1512 is concentration-dependent. The effects of the two best performing compounds on cells expressing only either E- or P-cadherin was evaluated with a cell-cell adhesion assay employing RNA interference on BxPC-3 cells as described in [106] generating stable cell lines expressing either E-cadherin or P-cadherin (Figure 30). The anti-adhesion effect is maintained regardless of both E- and P-cadherin-mediated cell adhesion (Figure 31).

The effect of ARN1512 and ARN1883 on N-cadherin was evaluated with a cell-cell adhesion assay performed with MiaPaca-2 N-cadherin expressing cells, obtained as described in the Cell Adhesion and Invasion Assay section of the Material and Methods chapter (Figure 32).

As shown in Figure 33, ARN1512 has no effect on N-cadherin expressing cells, that form more compact spheroids compared to the not transfected MiaPaca-2 cells (mock cells) used as control in this experiment. Similar results were obtained for ARN1883 (personal communication, data not shown).

Preliminary data on invasion assays performed on BxPC3 E-cadherin/P-cadherin cells with ARN1512 and ARN1883 shows that both compounds inhibit cell invasion with a slightly better efficacy compared to AS11 (Figure 34). However, the graph in Figure 34 was obtained by performing two experiments in octuplets and therefore it is not possible to provide a complete statistical analysis to assess if the results are significant.

Moreover, the combined use of the most potent ARN compound, ARN1512, and AS11 both at 0.025 mM showed better results in terms of anti-adhesion efficacy relative to the two single compounds at 0.025 mM concentration (Figure 35). These data were obtained from two single experiment and therefore should not be considered as statistically significant yet. However, if combined with the structural data presented in this thesis and with the previous knowledge on AS11 [106], they provide an indication for the fact that the ARN compounds and AS11 have a different E-cadherin binding site mechanisms and the combined use of both compounds might display a synergistic effect.

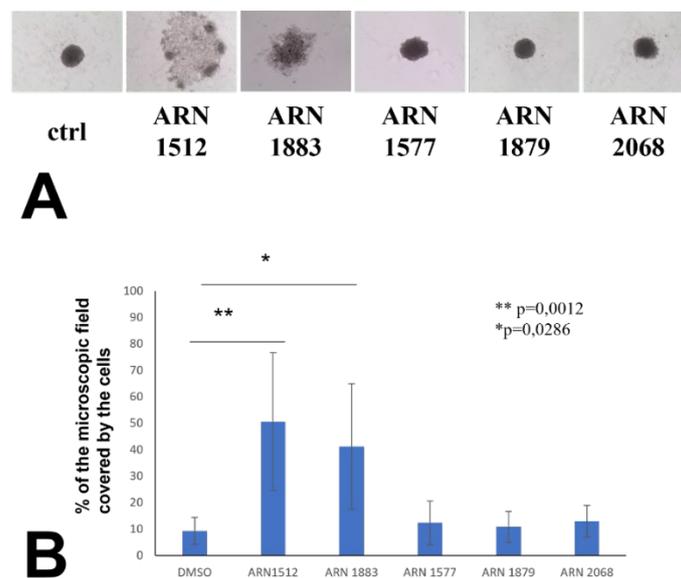


Figure 28: Effect of ARN1512, ARN1883, ARN1577, ARN1879, and ARN2068 on cell-cell adhesion.

In panel A, representative pictures of the spheroids after incubation of 24 hours with 0.05 mM compound and 0.1 % DMSO are shown. In panel B, the measured spheroid area is plotted against the experimental condition referring to it. Values represent the mean of 5 experiments performed at least in quintuplet. A $p < 0.05$ was considered as statistically significant and is indicated by (**) when $p < 0.005$ and (*) when $p < 0.05$.

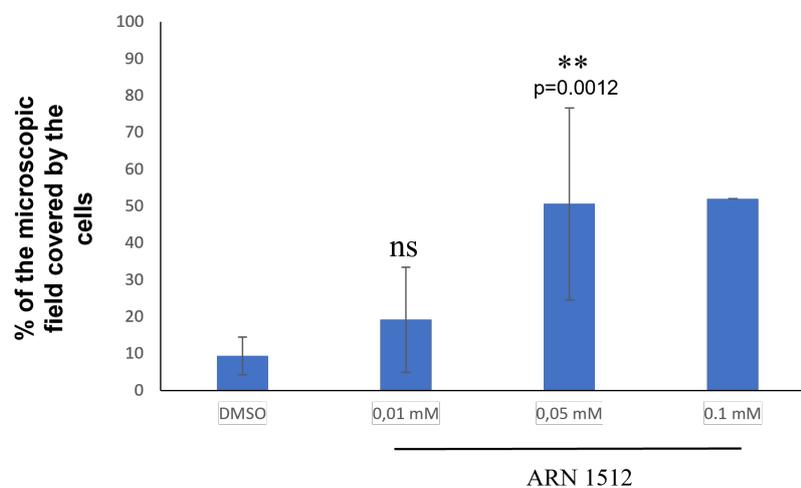


Figure 29: ARN 1512 reduces cell-cell adhesion in a concentration-dependent fashion

BxPC3 E-cadherin/P-cadherin cells were incubated in the presence of either 0.1% DMSO or various doses of ARN1512. Cells were allowed to form spheroids in suspension for 24 h. Values represent the mean spheroid area of 1 to 5 experiments performed in octuplets. Double asterisks indicate statistical significance with the reported p value, and ns stands for not significant.

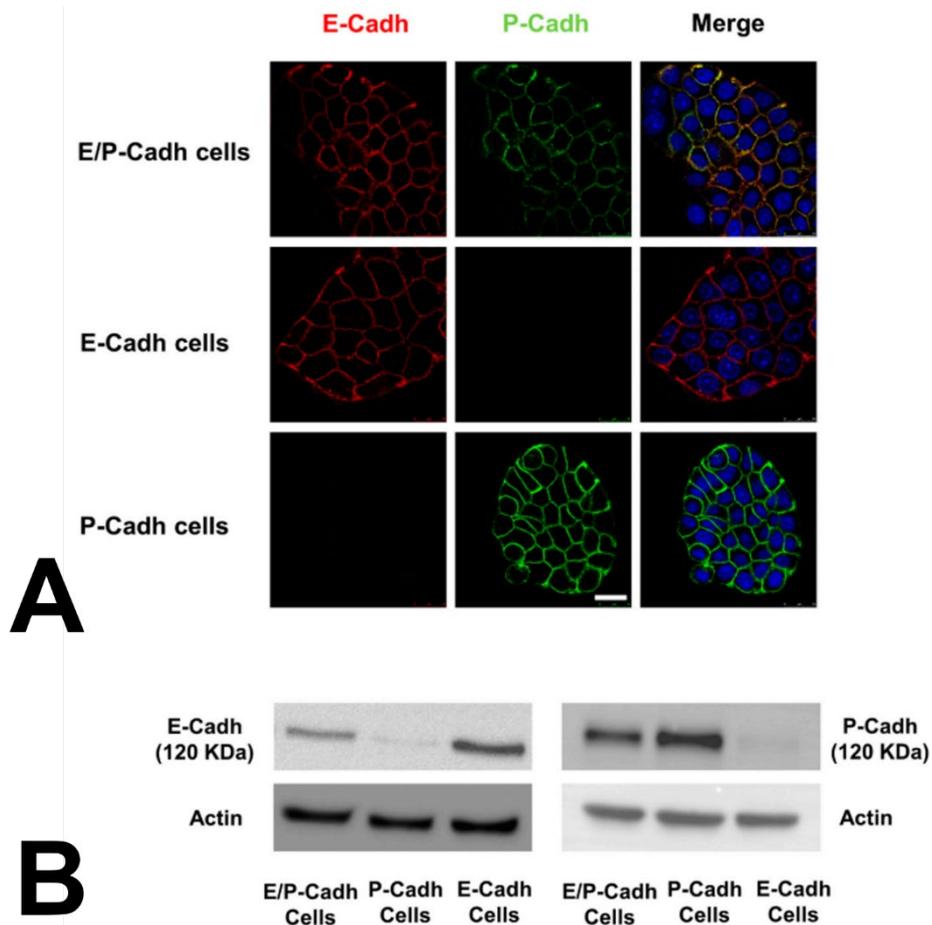


Figure 30: Results of the RNA interference on BxPC-3 cells

In panel A, using RNA interference, E- and P-cadherin were stably knocked down in BxPC-3 cells expressing both cadherins. Images were captured and analyzed using a SP5 Leica confocal microscope equipped with LAS AF Lite software. Scale bar: 25 μ m.

In panel 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 [106].

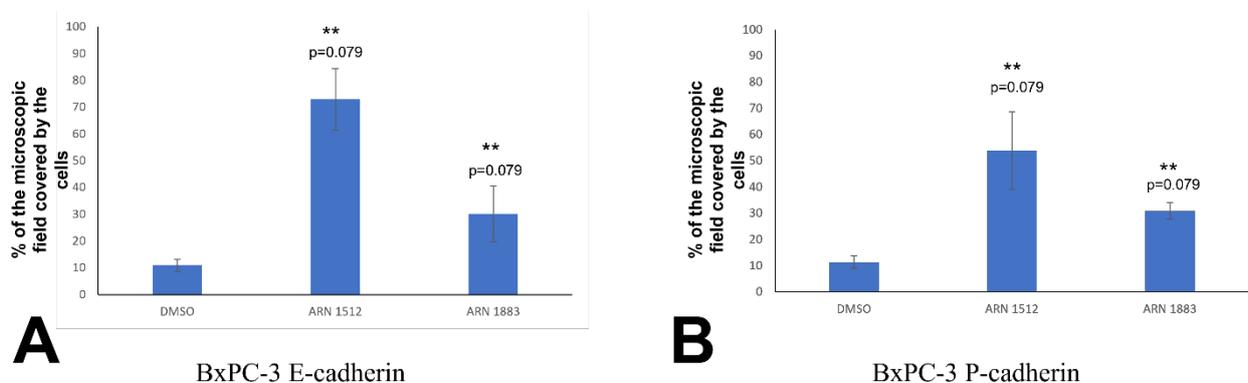


Figure 31: Effects of ARN1512 and ARN1883 targeting either E-cadherin or P-cadherin

BxPC-3 E-cadherin (A) and BxPC-3 P-cadherin cells (B) were incubated in the presence of 0.1% DMSO or 0.05 mM ARN 1883 or ARN 1512 for 24 h. The spheroid area was measured by phase-contrast microscopy and analyzed by ImageJ. Values represent the mean of 5 experiments performed in quintuplet. Double asterisks indicate statistical significance with the reported p value.

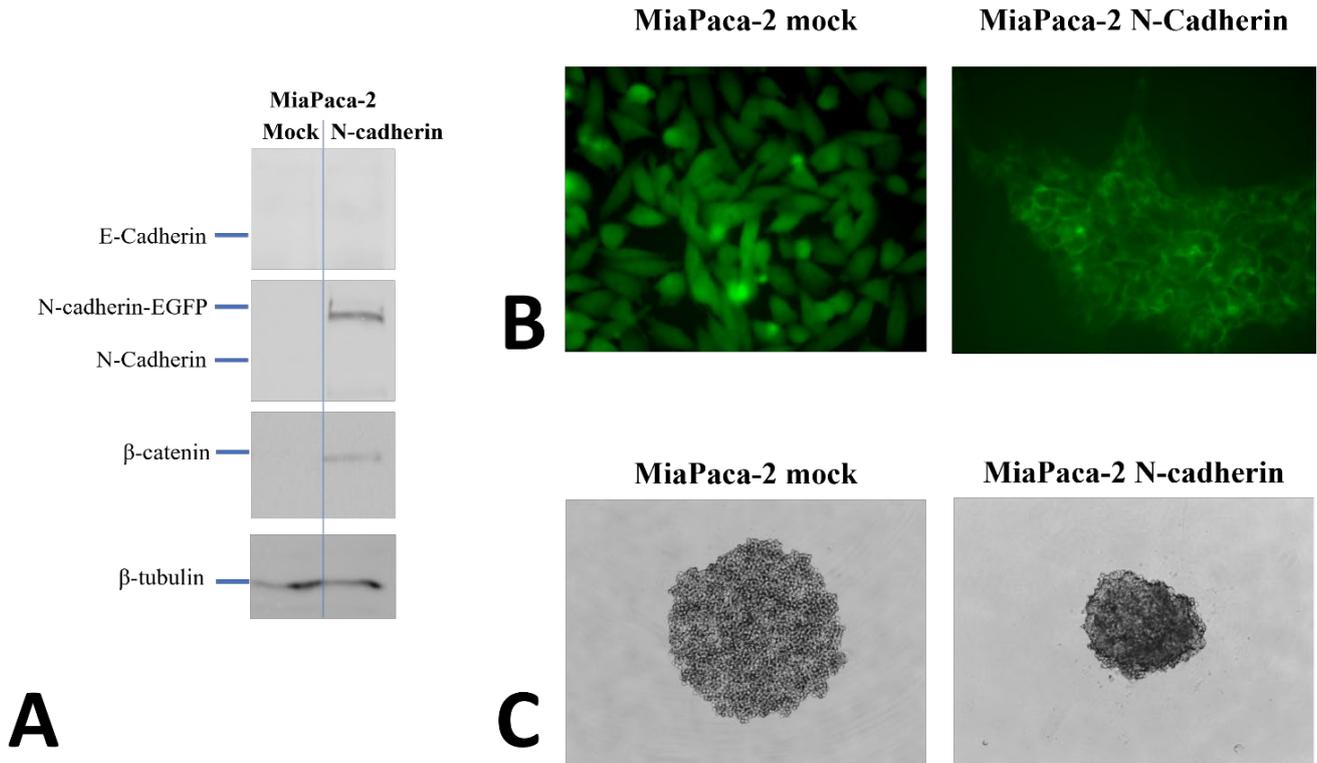


Figure 32: N-cadherin expression and its adhesive function in MiaPaca-2 cells
MiaPaca-2 cells were stably transfected with a N-cadherin-GFP containing vector (N-cadherin) or a GFP-containing vector (pEGFP). In panel A, the E-cadherin, N-cadherin and β catenin expression assessment by western blot by using specific antibodies is shown. Note that N-cadherin expression promotes β catenin expression. In panel B, the result of the cell surface expression assessed by fluorescence of GFP are shown. Images were captured using a LSM Zeiss confocal microscope equipped with Zen software. In panel C, the impact of N-cadherin on cell-cell adhesion is shown by using an aggregation assay as described in the Materials and Methods section. MiaPaca-2 and MiaPaca-2 N-cadherin cells were allowed to form spheroids in suspension for 24 h and images were captured using a LSM Zeiss confocal microscope equipped with Zen software.

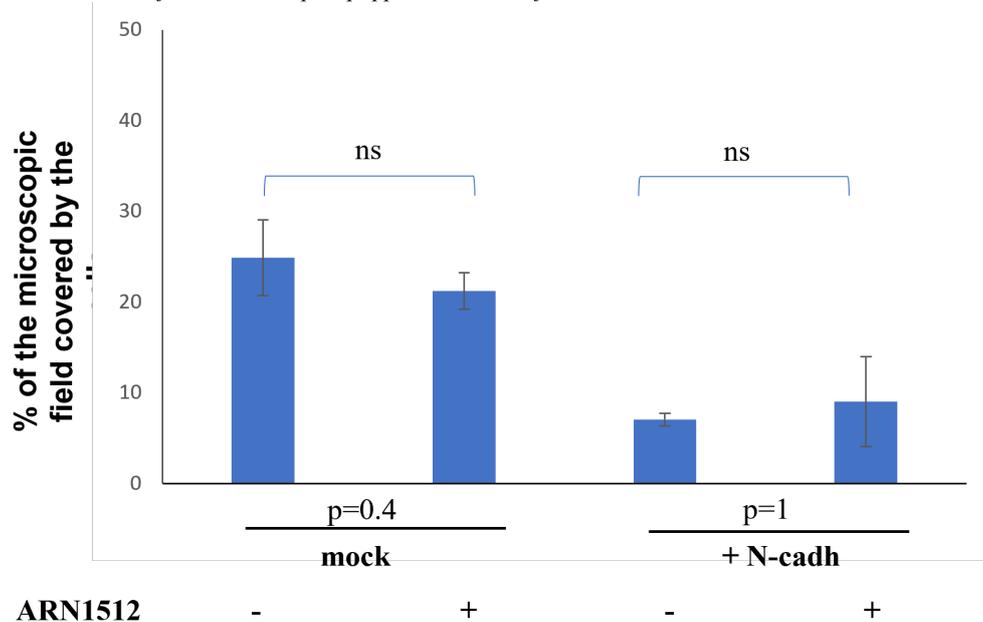


Figure 33: Effect of ARN1512 and ARN1883 on MiaPaca-2 N-cadherin cells
MiaPaca-2 (mock) and MiaPaca-2 N-cadherin cells were incubated in the presence of 0.1% DMSO or 0.05 mM ARN 1512 for 24 h. The spheroid area was measured by phase-contrast microscopy and analyzed by ImageJ. Values represent the mean of 5 experiments performed in octuplets. Corresponding p values are reported, and ns stands for not significant.

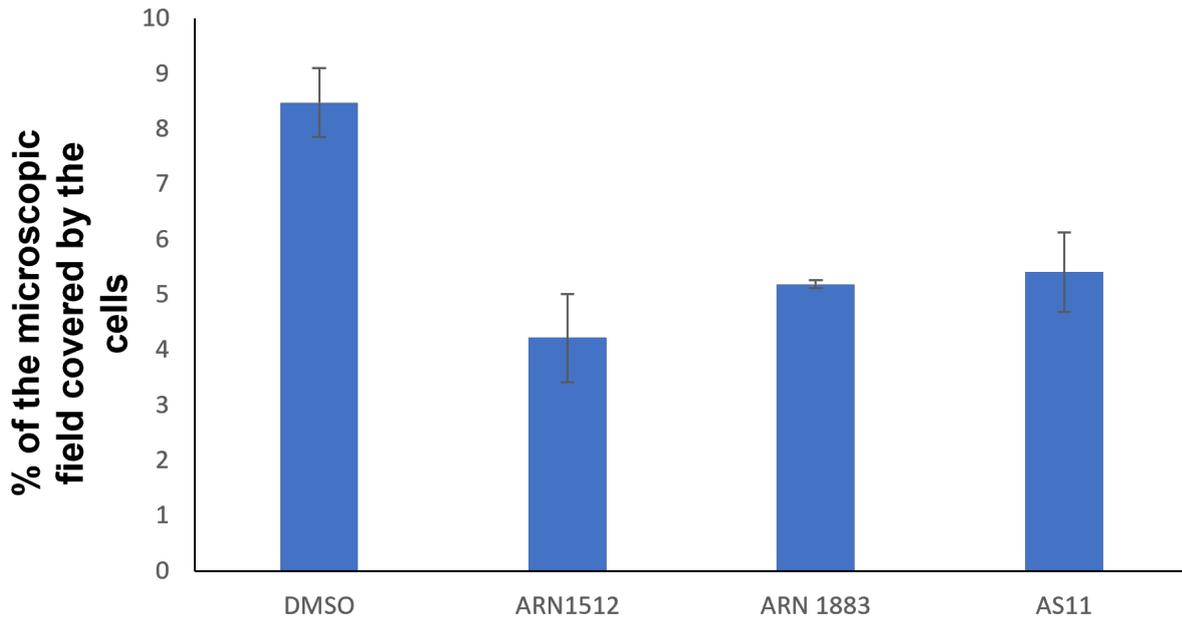


Figure 34: Effects of ARN1512 and ARN1883 on cell invasion

BxPC3 E-cadherin/P-cadherin cells were allowed to form spheroids for 72 h. Spheroids were embedded in type I collagen then incubated for 48 h in the presence of 0.1% DMSO or 0.05 mM ARN 1883, ARN 1512 or AS11. The spheroid area was observed by phase contrast microscopy 48 h after embedding. The spheroid area was evaluated using ImageJ software. Values represent the mean of 2 experiments performed in octuplets.

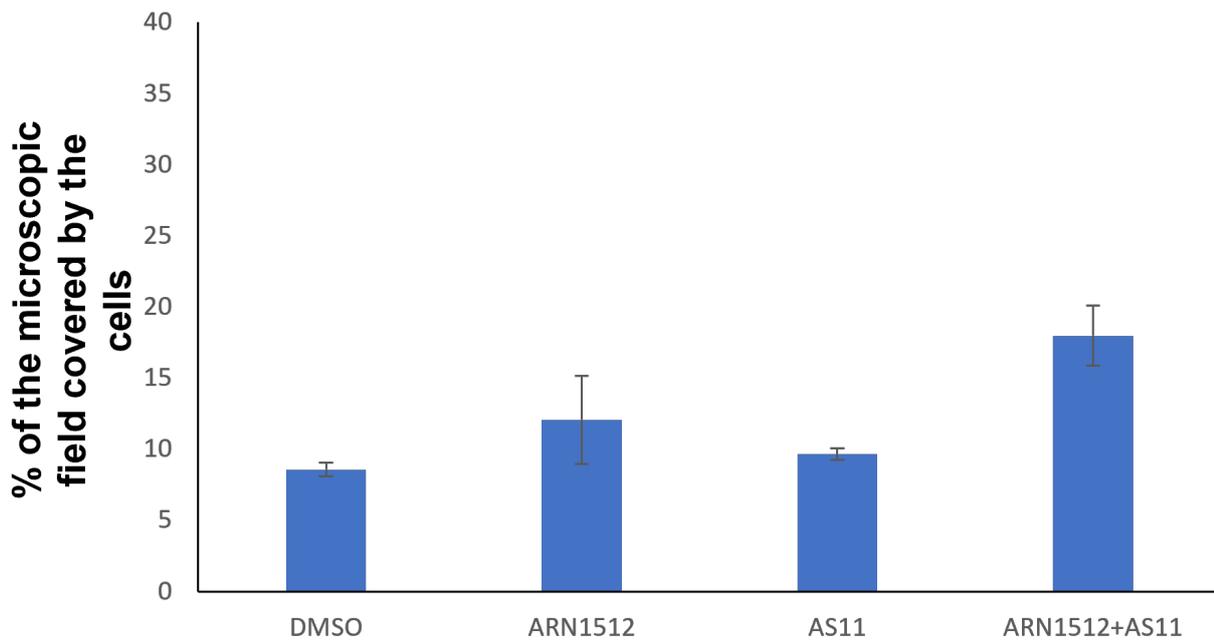


Figure 35: ARN1512 and AS11 target E-cadherin in an independent manner

BxPC-3 E-cadherin cells were incubated for 24 h in the presence of 0.1% DMSO, 0.025 mM ARN 1512, 0.025 mM AS11, or a combination of both 0.025 mM ARN1512 and 0.025 mM AS11. The spheroid area was measured by phase-contrast microscopy and analyzed by ImageJ. Values represent the mean of 2 experiments performed in octuplets.

Discussion

Binding Mode of the ARN Compounds

All three ARN compounds were found to bind both monomers in the X-dimer at the level of the adhesion pocket where, in wt-E-cadherin, Trp2 is located (Figure 36, panels A, D, and G). As shown in panels B, E, and H of Figure 36, the two monomers of all three structures superimpose within a very small rmsd value (1.05 Å, 0.87 Å, and 0.98 Å for the structures of ARN1512, ARN1577, and ARN1883, respectively). In particular, the rmsd values calculated from the superimposition of the two ligands in the asymmetric unit for the three structures gave similar results (0.41 Å, 1.87 Å, and 0.95 Å for the structures of ARN1512, ARN1577, and ARN1883, respectively; panels C, F, and I of Figure 36). The slightly higher value for the structural superimposition of ARN1577 is due to the fact that the “tail” composed by the amide bond linked to a methylpyrrole moiety is oriented in opposite directions in the crystallographic model. Moreover, the electron density of this portion of the molecules is extremely diffuse since it does not form any stabilizing contacts with the protein. Indeed, the alignment of the only 5-fluoro-3-methylindole moieties gives a rmsd value of 0.03 Å, indicating that they align almost perfectly.

In general, the fluorinated moiety (fluorobenzene moiety for ARN1512, 5-fluoro-3-methylindole moiety for ARN1577, and 7-fluoro-3,13,15-triazapentacyclononadeca-2(10),4(9),5,7-tetraene moiety for ARN1883) is always inserted into the Trp2 hydrophobic pocket, while the remaining part of the molecules are not stabilized by any contact with the other residues of the protein.

The binding of the three small chemical fragments to the protein does not generate conformational changes or rearrangements, neither globally, as attested by the rmsd values calculated by comparing the three structures with the unligated form of E-cadherin in a X-dimer conformation (PDBID: 4ZT1), nor locally in the residues lining the region of the adhesion pocket (Figure 37).

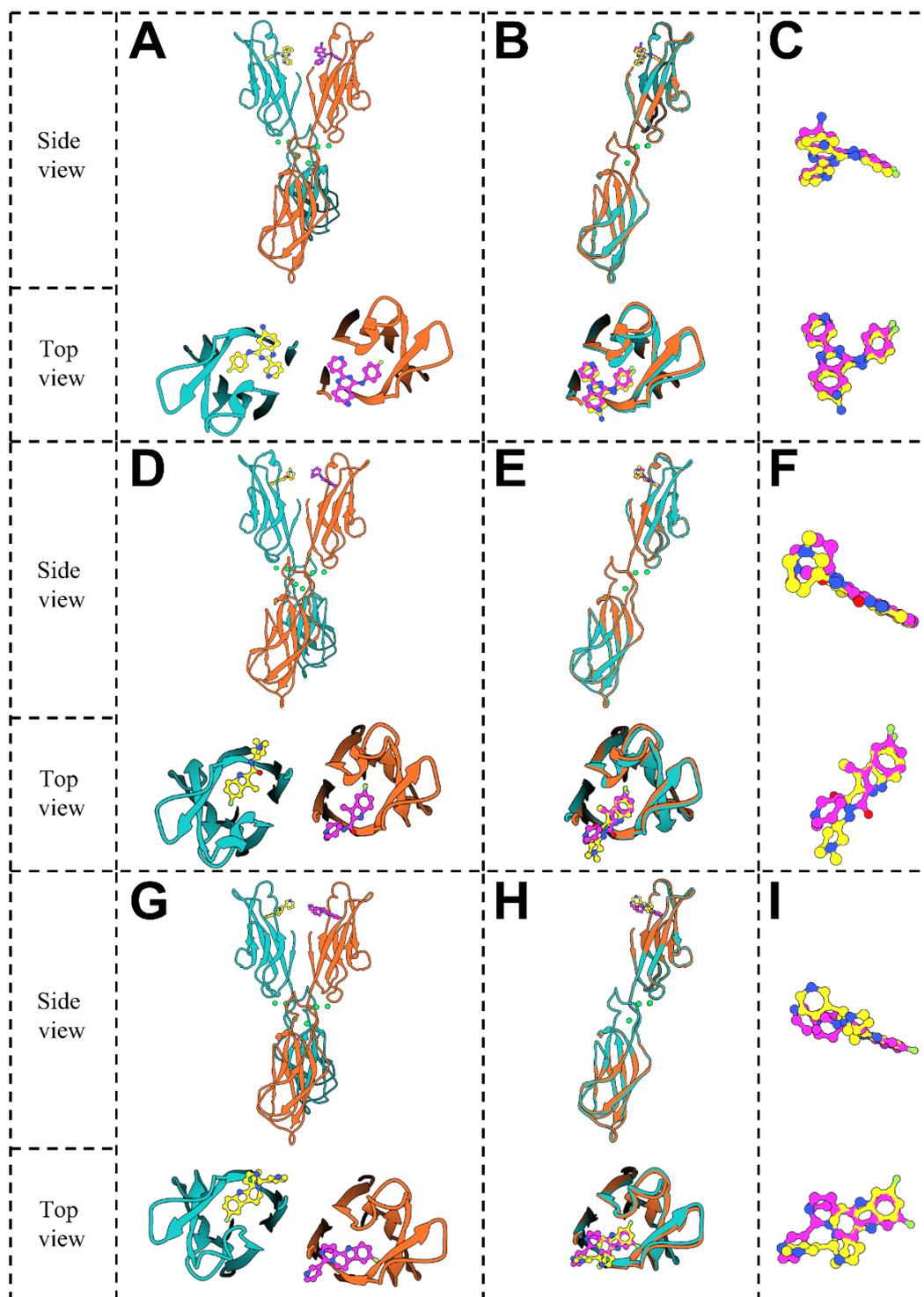


Figure 36: Summary of the structural details of the three co-crystal structures

The whole picture is divided in two views: one from the side and one from the bottom; and in different degrees of superimpositions.

Panels A, B, and C refer to ARN1512_A, panels D, E, and F to ARN1577_Z, and panels G, H, and I to ARN1883.

Panels A, D, and G show the co-crystal structures from the side and top point of views. The two X-dimer monomers are drawn in orange and cyan ribbon with the compounds drawn in ball and stick model with either yellow or magenta carbon atoms, blue nitrogens, red oxygens and green fluorines. Calcium ions in the side views are shown in green spheres.

Panels B, E, and H display the structural superimposition of the two monomers of the asymmetric unit and are colored according to the same color code.

Panels C, F, and I show the structural superimposition between the two compounds in the asymmetric unit and are colored according to the same color code.

It is worth noting that all the side chains of the residues that are located in a 3 Å spherical region centered around the ARN ligand possess the same side chain orientation, with the only exception of Met92, which occupies the adhesion pocket in the place of the Trp2. The difference in the position of the Met92 side chain in the three ARN structures makes the conformation of their adhesion pocket more similar to the one of the strand-swap E-cadherin dimer (PDBID: 2O72) than to that of the X-dimer (PDBID: 4ZT1). Indeed, the adhesion pocket is occupied by the ARN compounds in the case of the three co-crystal structures and by the Trp2 of the partner molecule in the case of the strand swap conformation. In the structure of all three complexes, Met92 is at the bottom of the pocket and providing a “floor” on which the fluorinated hydrophobic moiety of the compounds can lie, sandwiched between Met92 and Glu90 in the same way the natural Trp2 residue is stabilized by those two amino acids.

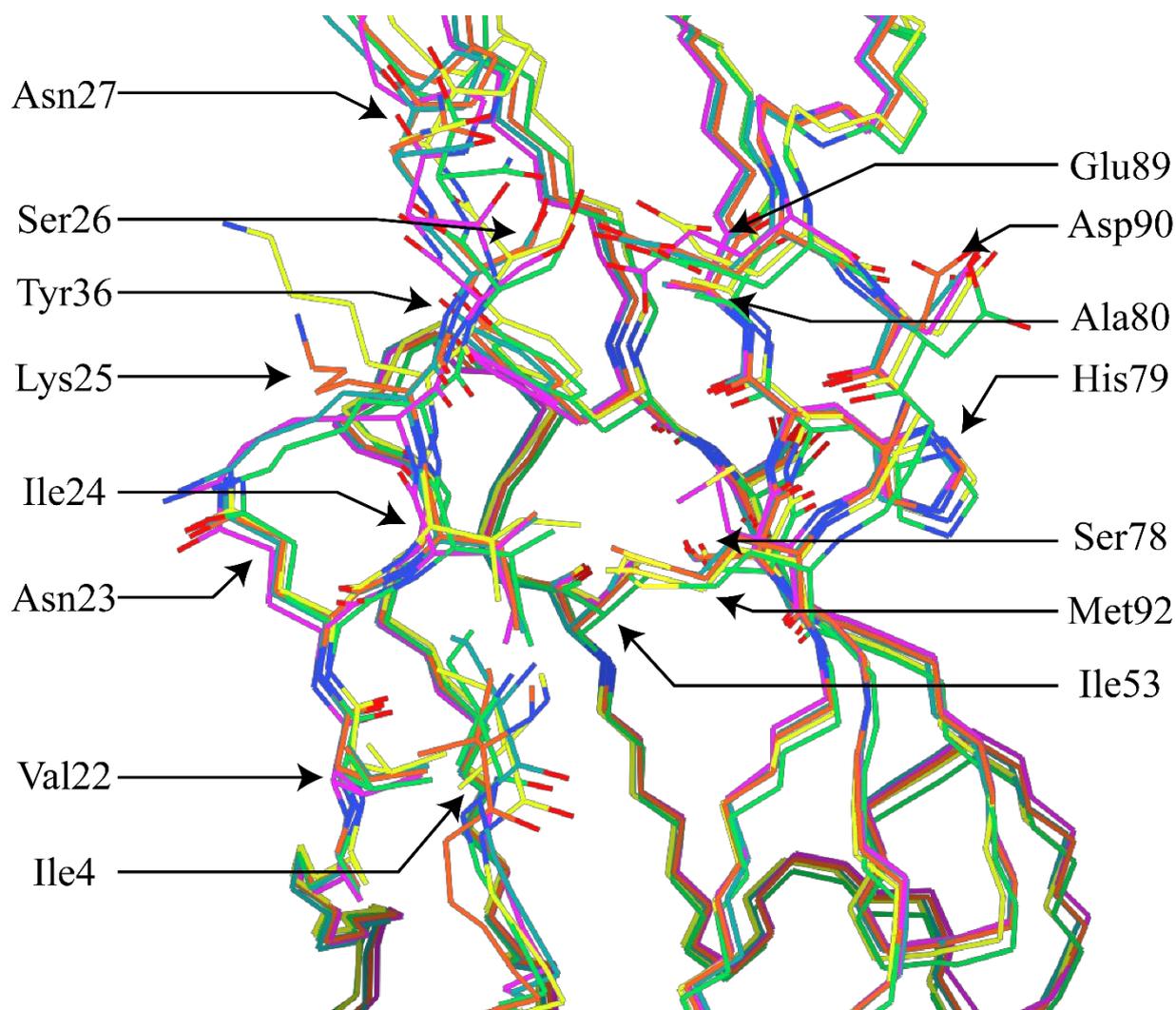


Figure 37: Structural superimposition between the three co-crystal structures, E-cadherin strand swap and X-dimer. The figure shows the structural superimposition between the ARN1512_A (orange), ARN_1577_Z (cyan), ARN1883 (green) co-crystal structures and E-cadherin strand swap (PDBID: 2O72, yellow) and X-dimer (PDBID: 4ZT1, magenta) drawn as wires with the residues in the 3 Å wide sphere drawn with wires and colored with carbon atoms of the color corresponding to the belonging structure, red oxygens, blue nitrogens, and yellow sulfurs. Each residue is labelled, and the label is connected to the structure by an arrow.

The binding of the ARN compounds to the acceptor pocket is particularly favored by the electrostatic profile of the EC1 domain. Indeed, if we look inside of the pocket from its entrance, we can appreciate either a nonpolar environment inside the adhesion pocket on the left hand-side and a negatively polarized/charged environment on the top and the right hand-side side of the pocket (Figure 38, panel A). The electrostatic surface of the ligands is complementary to the one of the EC1 pocket region because all compounds are mainly nonpolar, with a positively polarized region that is due to the presence of nitrogen atoms. This electrostatic complementarity allows the positioning of the nonpolar bulky fluorinated “head” of the compound inside the pocket and the “tail” in correspondence to the negatively charged region of the pocket (Figure 38, panels B, C, and D).

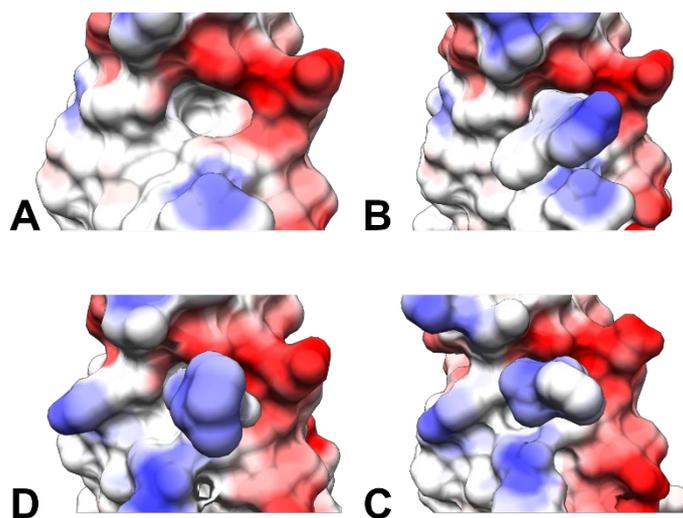


Figure 38: Electrostatic surface calculated by Chimera of the empty and compound occupied pockets. The figure shows the electrostatic potential of the adhesion pocket region when the pocket is empty (panel A), and when it is filled by ARN1512_A (panel B), ARN1577_Z (panel C), and ARN1883 (panel D).

White regions are attributed to nonpolar atoms, blue ones to positively charged/polarized atoms and red ones to negatively charged/polarized atoms.

All electrostatic surfaces have been calculated with the ‘Coulombic surface coloring’ function included in the Chimera software.

Finally, it is worth mentioning that, although the structural data presented here were obtained from a truncated form of E-cadherin, the absence of the first two N-terminal amino acids is not likely to influence the overall binding mode of the ARN compounds. Indeed, the conformational changes that are associated with the opening of the adhesion arm do not involve changes in the conformation of the residues that interact with the compounds within the adhesion pocket. The presence of Asp1 and Trp2 might have an effect on the strength and on the kinetics of the binding since they can provide additional sites for “tails” of the compounds to interact either in an attractive or repulsive way, depending on their chemical nature. However, an exact assessment of this issue cannot be made without additional biochemical data, which could be obtained by NMR or by calorimetry, or with the determination of the crystal structure of the complex between wt E-cadherin and the compounds. To this end, while considerable attempts to crystallize this latter complex have been made, they have so far been unsuccessful.

Description of the ARN1512 Interaction Mode

Figure 39 shows the interactions that ARN1512 makes with the E-cadherin monomers.

As previously discussed, the interaction between the ligand and the protein involves mainly the fluorobenzene moiety of the ligand and the acceptor pocket of the protein. The fluorinated aromatic ring forms hydrophobic interactions with the hydrophobic regions of the side chains of Gln23, Ile24, Tyr36, Ser78, Ala80, Glu89, Asp90, and Met92. In particular, the portion of the compound that more closely mimics Trp2 is sandwiched between Met92 and Glu89 in the same way as Trp2. Furthermore, the pyridine moiety of the ligand makes a hydrophobic contact with the side chain of Ile4 (Figure 39, panel A and B).

The compound is also stabilized by a hydrogen bonding between the nitrogen in the pyridine moiety and the backbone amide of Lys25 and between one nitrogen atom in the core of the compound and the sulfur of Met92, and by a water mediated contact with the backbone carbonyl of Asp90 (Figure 39, panels C and D).

Finally, the fluorine atom establishes a multipolar interaction with the backbone carbonyl group of Ala80 (Figure 39, panels E and F).

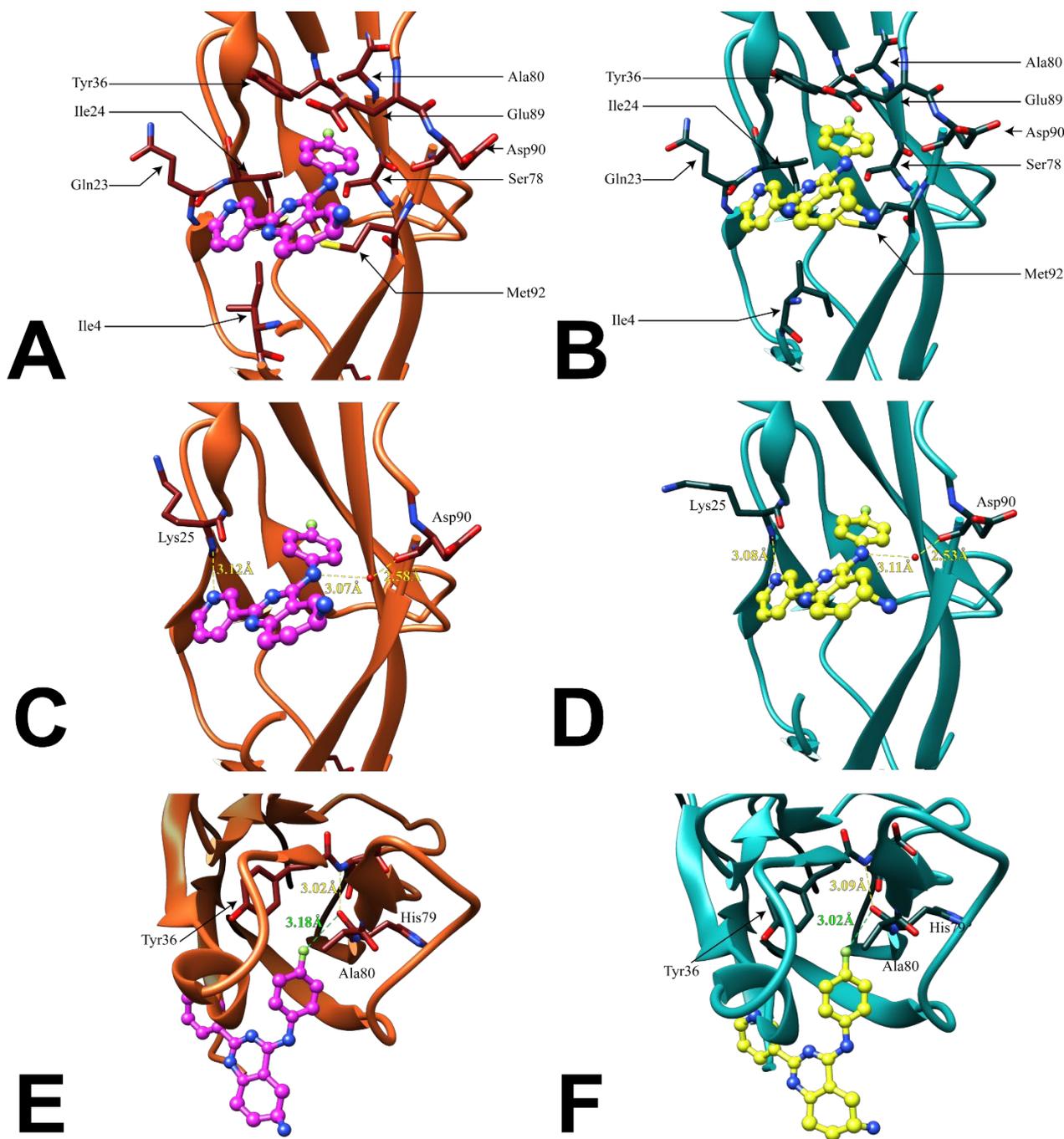


Figure 39: Details of the interaction of ARN1512_A with E-cadherin acceptor pocket

The picture shows the interactions of both ARN1512_A molecules in the two monomers: panels A, C, and E refer to monomer A, while panels B, D, and F to monomer B and they are colored according to the same color code of Figure 36.

Panels A and B show the hydrophobic interactions of the compounds inside the Trp2 acceptor pockets. Interacting residues are drawn in sticks and colored in either dark red or dark blue according to the belonging monomer, A or B, respectively.

Panels C and D highlight the hydrogen bonding interactions between ARN1512_A and the protein molecule. Interacting residues are labelled with their three-letter code and sequence number and hydrogen bonds are indicated with dashed lines and they are labelled with their measured distances. The water molecule that is involved in the water-mediated contact is drawn as a small red sphere.

Panels E and F display the multipolar interaction between the ARN1512_A fluorine atom and the carbonyl group of Ala80, that is involved with the amide group in hydrogen bonding. Multipolar interactions are drawn in green dashed lines and hydrogen bond with yellow ones. Each interaction is provided with the corresponding distance.

Description of the ARN1577 Interaction Mode

Figure 40 describes the interactions that ARN1577 makes with E-cadherin.

The ligand interacts with the protein mainly via the 5-fluoro-3-methylindole moiety, which forms hydrophobic contacts with the hydrophobic regions of the side chains of Ile24, Lys25, Tyr36, Ser78, His79, Ala80, Glu89, Asp90, and Met92 (Figure 40, panels A and B). In particular, the portion of the compound that more closely mimics Trp2 is sandwiched between Met92 and Glu89 in the same way of Trp2.

The nitrogen in the 5-fluoro-3-methylindole moiety forms a hydrogen bond with the backbone carbonyl group of Asp90 (Figure 40, panels C and D) and the fluorine atom makes a polar interaction with the aromatic ring of Tyr36 (Figure 40, panels E and F).

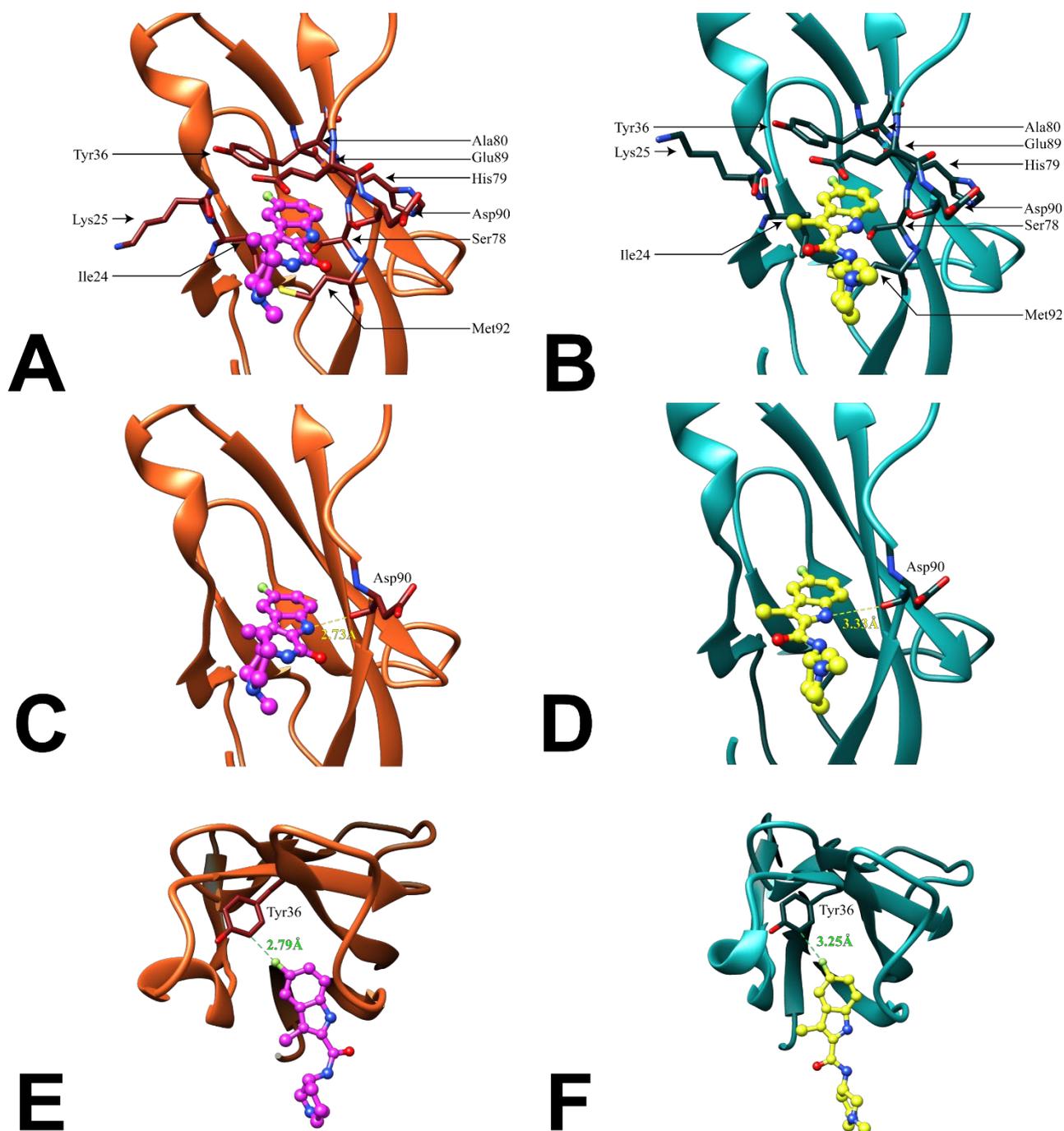


Figure 40: Details of the interaction of ARN1577_Z with E-cadherin acceptor pocket

The picture shows the interactions of both ARN1577_Z molecules in the two monomers: panels A, C, and E refer to monomer A, while panels B, D, and F to monomer B and they are colored according to the same color code of Figure 20.

Panels A and B show the hydrophobic interactions of the compounds inside the Trp2 acceptor pockets. Interacting residues are drawn in sticks and colored in either dark red or dark blue according to the belonging monomer, A or B, respectively.

Panels C and D highlight the hydrogen bonding interactions between ARN1577_Z and the protein molecule. Interacting residues are labelled with their three-letter code and sequence number and hydrogen bonds are indicated with dashed lines and they are labelled with their measured distances.

Panels E and F display the polar interaction between the ARN1577_Z fluorine atom and the side chain of Tyr36. Polar interactions are drawn in green dashed lines and hydrogen bond with yellow ones. Each interaction is provided with the corresponding distance.

Description of the ARN1883 Interaction Mode

Figure 41 shows the interaction mode of ARN1883 with E-cadherin.

As always, the fluorinated moiety is inserted in the acceptor pocket and makes hydrophobic contacts with the side chains and the nonpolar regions of Ile24, Lys25, Tyr36, Ser78, His79, Ala80, Glu89, Asp90, and Met92 (Figure 41, panels A and B). In particular, the portion of the compound that more closely mimics Trp2 is sandwiched between Met92 and Glu89 in the same way of Trp2.

Unlike the two previous co-crystal structures, ARN1883 has different orientations in the two E-cadherin pockets. In monomer A, the fluorine atom forms a hydrogen bond with a water molecule that, in turn, forms hydrogen bonds with the backbone carbonyl of Glu89 and with the backbone amide of Ala80 (Figure 41, panel C). In monomer B, the fluorine atom forms a polar contact with the aromatic ring of Tyr36 (Figure 41, panel D).

Furthermore, the hydrogen bonding pattern of the nitrogen in the condensed bicyclic rings is different from the two monomers since the atom has opposite orientations: in monomer A, the interacting residue is Lys25 with its backbone carbonyl (Figure 41, panel E), while in monomer B it is the backbone carbonyl of Asp90 (Figure 41, panel F). These different orientations, which are reminiscent of the double conformation of Trp2 in the crystal structure of human P-cadherin in closed conformation (PDBID: 4OY9), were the only possible ones that did not result in the appearance of negative density around the fluorine atom in the Fo-Fc map and in either one of the compounds during model refinement. While some attempts were made to model in each pocket these double conformations, these yielded poor results in terms of quality of the final electron density map in the region. However, as we observe the two orientations in the proteins, we cannot rule out the presence of both conformation at the same time in the pockets and that one is not traceable: this fact might be the cause for the diffuse and somehow fragmented electron density observed for ARN1883.

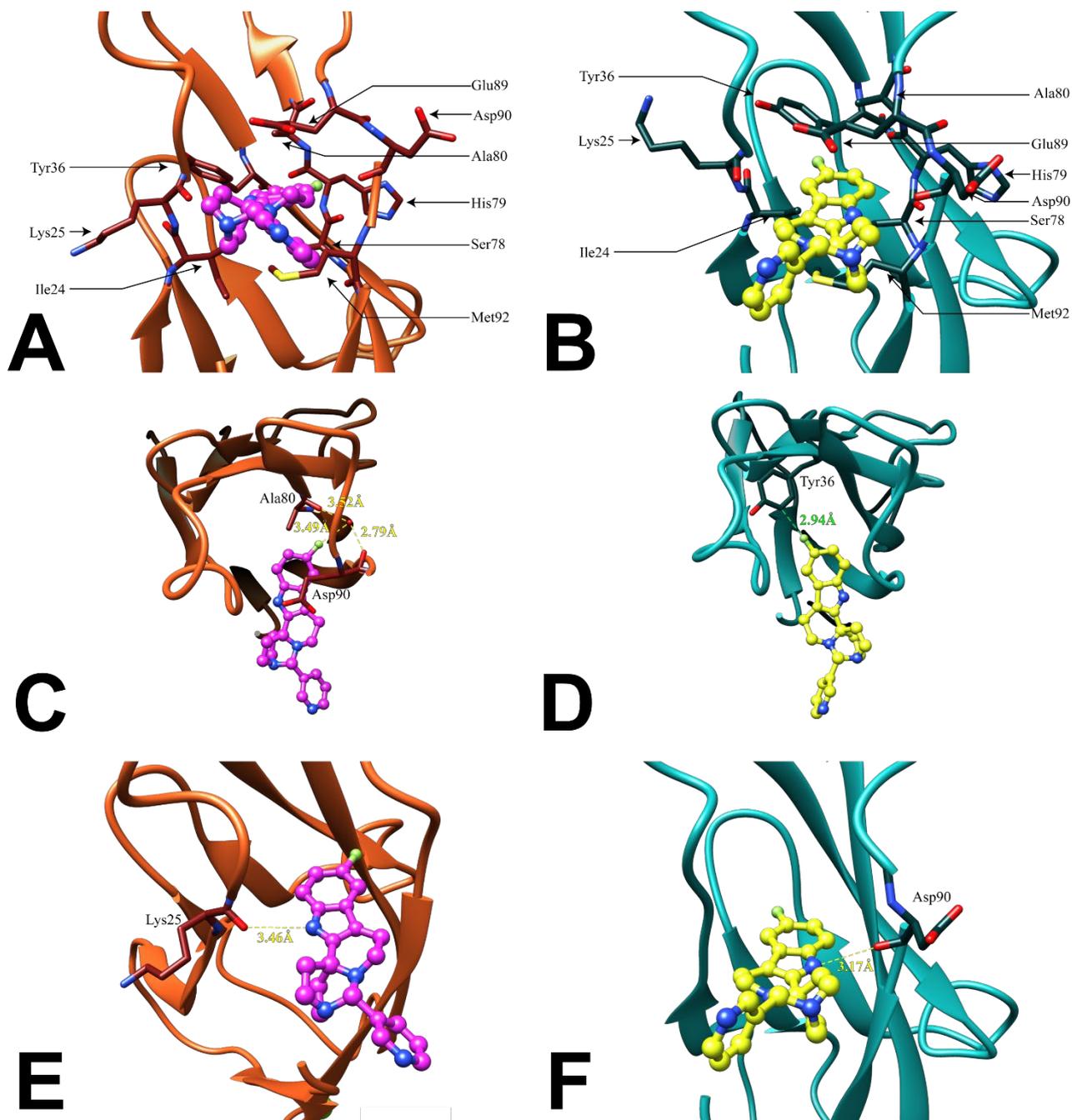


Figure 41: Details of the interaction of ARN1883 with E-cadherin acceptor pocket

The picture shows the interactions of both ARN1883 molecules in the two monomers: panels A, C, and E refer to monomer A, while panels B, D, and F to monomer B and they are colored according to the same color code of Figure 36.

Panels A and B show the hydrophobic interactions of the compounds inside the Trp2 acceptor pockets. Interacting residues are drawn in sticks and colored in either dark red or dark blue according to the belonging monomer, A or B, respectively.

Panels C and D highlight the different interaction modes of the ARN1883 fluorine atom: in monomer A, the fluorine atom is hydrogen bonding a water molecule that in turn acts as a bridge between Asp90 and Ala80, while in monomer B, it is observed to interact with the side chain of Tyr36 making a polar contact as it does ARN1577_Z. Interacting residues are drawn in ticks and colored according to the usual color code of Figure 36. Hydrogen bonds are indicated with yellow dashed lines and polar interactions with green ones, and both are labelled with the measured distances.

Panels E and F display the hydrogen bonding interactions between the ARN1577_Z and the backbone of either Lys 25 (monomer A, panel E) or Asp90 (monomer B, panel F). Hydrogen bonds are indicated with dashed lines, and they are labelled with their measured distances.

Cell Adhesion and Cell Invasion Assays

The cell-cell adhesion assay reported in Figure 28 and the invasion assay reported in Figure 34 clearly indicate that ARN1512 and ARN1883 are capable of inhibiting cell-cell adhesion and cell migration in BxPC3 cells with statistical significance at μM concentrations. Furthermore, the most potent compound of the series, ARN1512, is also able to exert its antiadhesive activity in a concentration dependent manner, as shown in Figure 29.

Given that BxPC3 cells were engineered to express both E- and P-cadherin and the good results obtained in cell-cell adhesion assays, prompted the possibility to evaluate whether the activity of the most potent compounds, ARN1512 and ARN1883, was due to the inhibition of either E- or P-cadherin.

Unlike AS9 and AS11, which displayed selectivity towards E-cadherin but no effect on P-cadherin, the establishment of BxPC-3 cells stably expressing either E- or P-cadherin through RNA interference revealed that, at least in the case of ARN1512 and ARN1883, there is no difference in the spheroid area increase when the two cadherins are separately targeted in BxPC-3 E-cadherin and BxPC-3 P-cadherin cell lines, and therefore, the antiadhesive effect of the compounds is exerted on both E- and P-cadherin-mediated cell-cell adhesion (Figure 31). Given the structural homology between the adhesion pockets of E- and P-cadherin, it is conceivable that the ARN compounds have similar effects on both proteins.

Interestingly, the cell-cell adhesion assay performed on MiaPaca2 cells expressing N-cadherin revealed that ARN1512 has no inhibitory effect on N-cadherin-mediated cell adhesion (Figure 33), despite the high sequence identity and structural homology with E-cadherin, indicating that this class of compounds are specific only towards E- and P-cadherin. The sequence alignment between E-, P- and N-cadherin may provide a possible explanation for this selective behavior. As shown in Figure 42, the most striking differences between the three is the presence of Ile and not Met in position 92, the “floor” of the adhesion pocket, and the presence of Arg in position 25 instead of Lys.

The first difference might be responsible for the loosening of the sandwiching interaction (previously described in detail in [41]) that holds in place both Trp2 and the ARN compounds. On the other hand, the second difference might be responsible for an unfavorable electrostatic surface: the presence of Arg24 and, above it, of Arg27 create a positively charged rim that might interfering with the positively charged compound. Molecular dynamics simulations are ongoing to test these hypotheses.

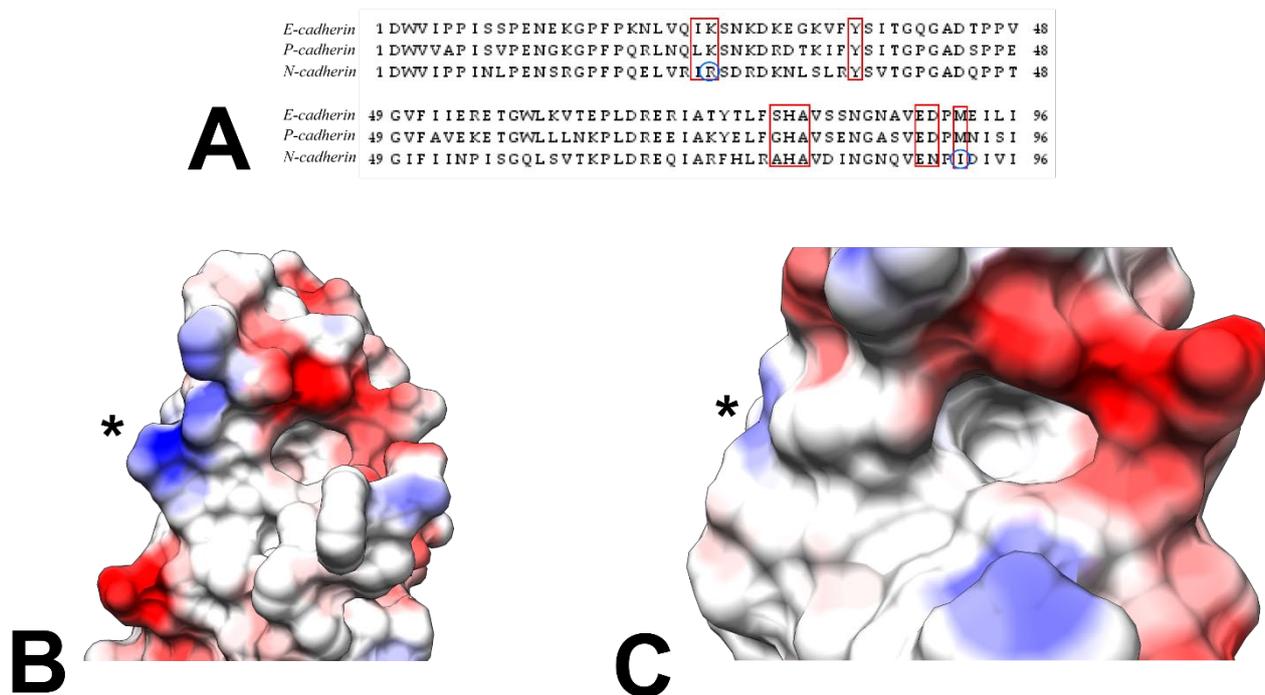


Figure 42: Sequence differences between E- and N-cadherin

In panel A, the sequence alignment of E-, P- and N-cadherin is displayed. Residues involved in the binding of the ARN compounds are contoured by red boxes and the most evident sequence differences are circled in blue.

In panel B and C, the electrostatic surfaces of N- and E-cadherin are respectively shown. The asterisk marks the position of either Arg24 or Lys24 depending on the considered protein.

Conclusions

Here, I reported the structural and functional characterization of three novel small molecule inhibitors of E-cadherin homo-dimerization.

For all the reasons described above, the ARN compounds can be regarded as “competitive” inhibitors of E-cadherin homo-dimerization. Of course, in this context the term competitive is used somewhat improperly as this is mostly an enzymology term utilized for inhibitors that hinder the access of a substrate to its catalytic site and E-cadherin is not an enzyme. However, if we think of the Trp2 residue as the “substrate” of the dimerization “reaction”, our comparison holds and the understanding of the mechanism of action of the ARN molecules might be easily explained and visualized.

While FR159 has been the first inhibitor ever to be co-crystallized with E-cadherin and has allowed the identification of a novel drugging site (i.e. the hydrophobic pocket that is formed when the protein is in the X-dimer conformation), these three ARN compounds are the first ever reported to be found to bind inside the adhesion pocket of E-cadherin. Such pocket has, for a long time and always elusively, being the actual target of many cadherin drug design studies. Hence, our data validate the cadherin adhesion pocket as a *bona fide* drugging site.

The functional data reported here demonstrate that ARN1512 and ARN1883 are the most potent compounds in *in vitro* cell assays and they effectively modulate BxPC-3 pancreatic adenocarcinoma cell-cell adhesion at micromolar concentrations, while showing no effect on N-cadherin-mediated cell-cell adhesion in another cell line of pancreatic adenocarcinoma (Miapaca-2). Their inhibitory effect is greater than that observed with AS11, which, to date, could be considered the gold standard. However, it remains to be established with statistical significance whether the ARN compounds are able to inhibit tumor cell invasion, if they outperform AS11 and if there is a synergistic effect with the combined use of the ARN compounds and AS11 provided that they have two different mechanisms of modulation of the cadherin dimerization process.

These issues are currently under investigation and therefore no definitive conclusion can be drawn so far. The structural data presented herein are the first of their kind and establish the ground for further developments that might lead to the improvement of the initial hits through successive rounds of structure-based drug design and to the discovery of other adhesion-pocket binding molecules through virtual screening of chemically similar molecules.

*BIOISOSTERIC REPLACEMENT OF CLASSICAL
CHOLINE KINASE INHIBITORS:
ANTITUMORAL AND ANTIPLASMODIAL
ACTIVITY*

Introduction

The Protein Kinase Superfamily

Protein Kinases (PKs) are a wide-spread class of enzymes that catalyze the formation of a phosphodiester bond between an amino acid of a protein that acts as an acceptor and a phosphate group that is generally provided by a donor molecule. These enzymes usually act in coordination with the phosphatases, a class of enzymes that mediate the opposite chemical reaction, i.e. the removal of a phosphate group and therefore the breakage of the phosphodiester bond.

Phosphorylation of proteins is a process that affects virtually every activity of cells, including proliferation, gene expression, transport across membranes, homeostasis, metabolism, motility, and apoptosis [119].

The process of protein phosphorylation/dephosphorylation is, arguably, the most common process of active regulation of the functional activity of proteins. Indeed, it is present in all organisms, ranging from the simplest prokaryotes [120], [121] to the more complex eukaryotic organisms [119], [122], [123]. Its ubiquitous presence might be ascribed to different factors, such as its high charge density, hydrophilicity, and propensity of the phosphate group to form salt bridges. These may be coupled with the low structural requirements that are involved with the creation of a phosphorylation site, i.e., a nucleophilic oxygen or nitrogen atom in a side chain of an amino acid surrounded by few other specific ones that act as recognition features. Moreover, this regulation mechanism provides an easy way to create complex signaling networks. Indeed, multiple phosphorylation sites can easily be accommodated in protein structures and can readily work as an ensemble to implement the biological equivalent of the logic gates that are nowadays present in our electronic equipment. In fact, phosphorylation in one site is able to affect allosterically the state of a protein and to interact in the same way with another phosphorylation site, providing a way to integrate multiple biochemical signals [124].

Almost all the phosphorylations that occur in the eukaryotic cells are mediated by a single superfamily of PKs, called eukaryotic Protein Kinases (ePKs), which share a conserved catalytic domain; however, several other proteins that lack sequence similarity to ePKs (albeit sharing some structural ones) are able to phosphorylate substrates and are called atypical Protein Kinases (aPKs) [123]. These latter are different from ePK since they generally phosphorylate small molecules instead of protein residues.

The first classification of PKs was done by Hanks et al. [125] in 1995. In their monumental work, they aligned the sequences of almost 400 ePKs belonging to different organisms, ranging from vertebrates to lower organisms, and established the first phylogenetic tree for the superfamily.

This phylogenetic tree is composed of five branches:

- The AGC group consists of basic amino acid-directed kinases that phosphorylate a Ser or a Thr residue located near an Arg or a Lys residue. This group is further classified into different families: the cyclic-nucleotide-dependent kinases, the protein kinase C family, the β -adrenergic receptor kinases, the ribosomal S6 kinases, and other close relatives.
- The CAMK group consists of basic amino acid-directed kinases that target Ser or Thr residues and are activated upon binding of calcium ions. It can be further divided into the calcium/calmodulin-dependent protein kinases, the Snf1/AMPK family, and some close relatives.
- The CMGC group consists of protein-serine kinases that phosphorylate their target substrates in Pro-rich regions. Members of this group are the cyclin-dependent kinases, Erk (MAP) kinases, the glycogen synthase 3 family, the casein kinases II, Cdk-like kinases, and other closely related members.
- The PTK group is composed of the protein tyrosine kinase family that is further divided into more than a dozen of membrane-spanning receptor families and nine nonreceptor families. The members of the PTK group can also be classified into conventional protein tyrosine kinases that phosphorylate exclusively Tyr residues and dual-specific protein tyrosine kinases that are able to attach the phosphate group also to Ser and Thr.
- Other unclassified members that displayed sequence similarity but were not assigned to any family at that time.

This initial classification has been more recently updated by Manning et al. [122] in 2002, where four additional groups have been added and aPKs have been included in the superfamily. The new groups are:

- The STE group composed by the MAPK cascade family.
 - The CK1 group composed by the casein kinase 1 family, the tau tubulin kinase family and the vaccina-related kinase family.
 - The TKL group composed of all the protein tyrosine kinase-like families that resemble both Tyr and Ser/Thr protein kinases. Examples for this group are the mixed-lineage kinases, LISK, intrleukin.1 receptor-related protein kinases, activitin kinases and TGF β receptor kinases.
 - The GYC group composed by the family of receptor guanylate cyclase kinases.
-

In total, in this work 518 PK genes were identified throughout the human genome: 478 were classified as ePK genes, while the remaining 40 ones as aPK genes. Figure 43 displays the phylogenetic tree of the Human Kinome.

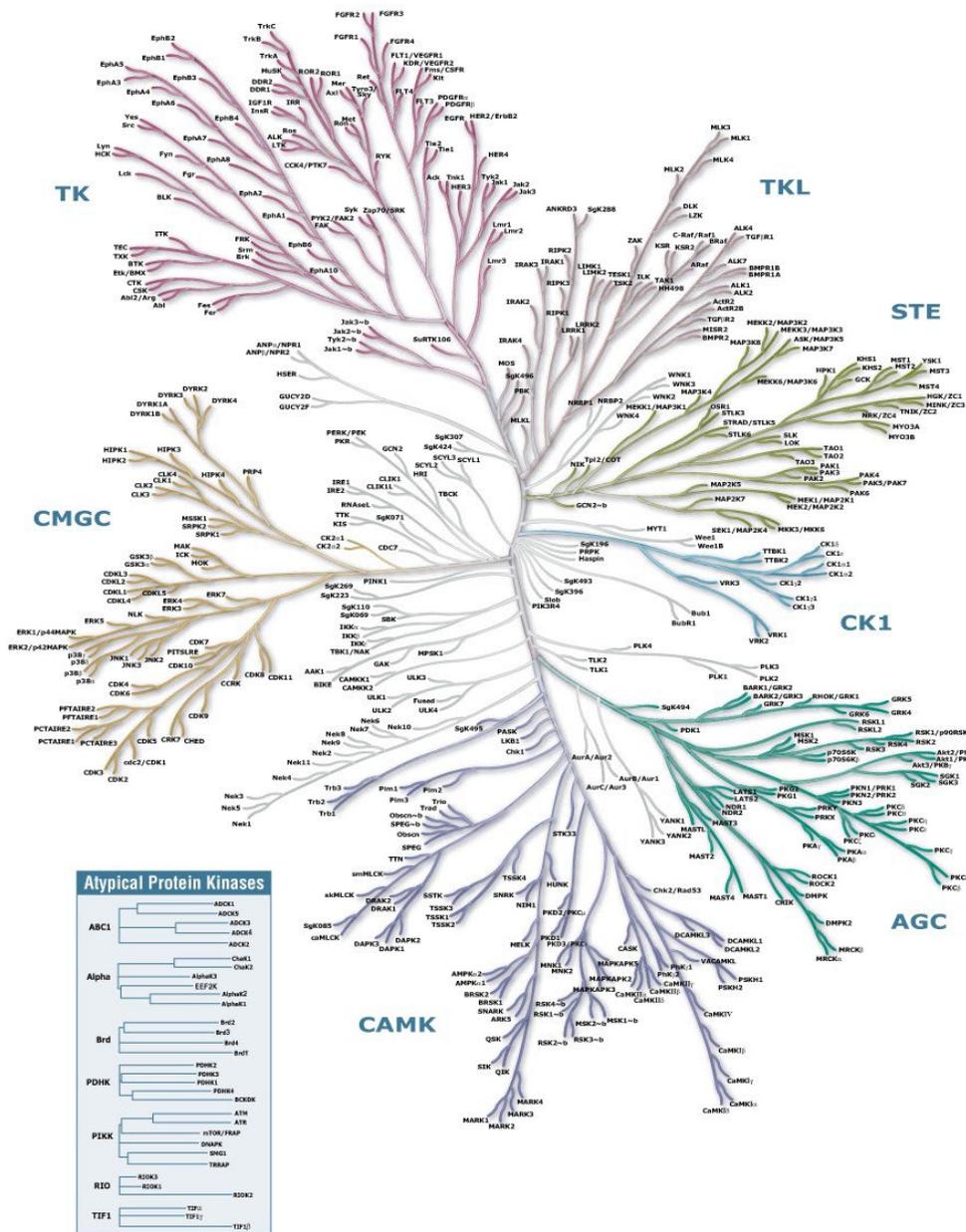


Figure 43: The Human Kinome

The picture displays the phylogenetic tree of the whole Human Kinome. In the main tree all the groups mentioned in the text are shown while the aPKs are depicted in a separate box in the bottom left of the picture. Illustration reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com).

The Eukaryotic Protein Kinase Fold

After the first protein kinase crystal structure was determined in 1991 by Knighton et al. [126], an increasing number of protein kinase structures have been deposited in the PDB Database and the picture that has emerged is that all ePKs share a common catalytic core folding topology [127], [128]. Despite aPKs showing little to no sequence similarity to ePKs, the same folding topology was later observed in the crystal structures of some aPKs [129], [130], [131], [132], leading to their inclusion in the superfamily classification.

The ePK fold is essentially the same: two lobes with the active site located in the cleft between the lobes. The N-terminal lobe is the smaller one and consists mainly of β strands, while the C-terminal lobe is larger and is mainly composed of α helices [133].

The most important structural motifs are located near the interface between the two lobes and are the nucleotide-binding site, the phosphate-binding site loop, the activation loop, and the substrate-binding site.

The nucleotide-binding site is the one involved in the binding of ATP and is mainly located in the smaller lobe where the β sheet region is found; it contributes to the correct orientation of the adenosine nucleotide that positions itself with the phosphate groups pointing outward toward the edge of the cleft. The phosphate-binding site is involved in the binding of the γ -phosphate group from ATP. Usually, this is a very flexible loop that contains the so-called glycine-rich motif, a structural motif that is highly conserved in all PKs and that features the sequence Gly-Thr-Gly-Ser-Phe-Gly.

The activation loop often contains additional phosphorylation sites and usually provides a suitable surface for binding the substrate. Finally, the substrate-binding site is contained in the C-terminal lobe and is involved in the binding and correct orientation of the substrate for the phosphotransfer reaction.

Despite their variable mechanism for controlling the activity of PKs [134], all protein kinases catalyze the same reaction: the transfer of a γ -phosphate group from ATP to a Ser, Thr, Tyr residue or a small molecule.

The kinase described in this thesis, Choline Kinase (ChoK), is classified as aPK and is present in the genome of various organisms, where it is mainly involved in lipid metabolism.

The Structure of Choline Kinase

The first experimental evidence of the existence of an enzyme that is capable of phosphorylating choline dates back to 1953, when Wittenberg et al. [135] demonstrated that crude yeast extracts and liver, brain, intestine, and kidney acetone extracts from several organisms were able to catalyze the phosphorylation of choline according to the reaction:



In the following years, the enzyme has been purified to homogeneity starting from different mammalian sources [136], [137], [138] and from *Saccharomyces cerevisiae* [139], [140] and the genes encoding for *bona fide* ChoKs were identified in mammals [141], [142], [143] and plants [144]. All these studies demonstrated that the presence of magnesium ions is necessary for the proceeding of the phosphotransfer reaction, and that, to a certain extent, ChoK displays the ability to phosphorylate ethanolamine too, despite the presence and existence of another enzyme that specifically phosphorylates ethanolamine (ethanolamine kinase) [138].

While the existence of ChoK has been known for several years, the first crystal structure was reported only in 2003, when Peisach et al. [145] determined the structure of the A-2 isoform of *C. Elegans* ChoK (PDBID: 1NW1). This enzyme is one of the seven proteins identified in the analysis of the *C. Elegans* genome and it belongs to the group of sequences that are the most similar to mammalian ChoKs (sequence identity ranging from 30 to 36% compared to mouse ChoK and depending on the sequence considered) [146].

This crystal structure revealed that *C. Elegans* ChoK isoform A-2 is a homodimeric enzyme constituted by two domains: a smaller N-terminal domain (152 amino acid long) and a larger C-terminal domain (271 amino acid long) connected by a short five-residue long linker strand. The overall architecture of this protein is the same as the one already found in ePKs catalytic domains, and its structure is very well superimposable to the structure of other ePKs [147].

As expected, the N-terminal domain is mainly composed of β strands (five in total) that form a twisted antiparallel β sheet with two helices inserted between the strands. One of these helices is involved in the formation of the dimer interface. On the other hand, the C-terminal domain is larger, and, aside from some β sheets that are found in hairpin loops, it is almost exclusively α helical. Peisach et al. divided it into three distinct regions: the central core, the insertion, and the C-terminal subdomains. The most important region is the central core subdomain because the most catalytically important motifs are located here.

Finally, the linker strand, which is bordered by two 90° turns, allows the N- and C-terminal domains to be placed in communication with each other.

Figure 44 shows the crystal structure of ChoK from *C. Elegans* highlighting the structural features discussed above.

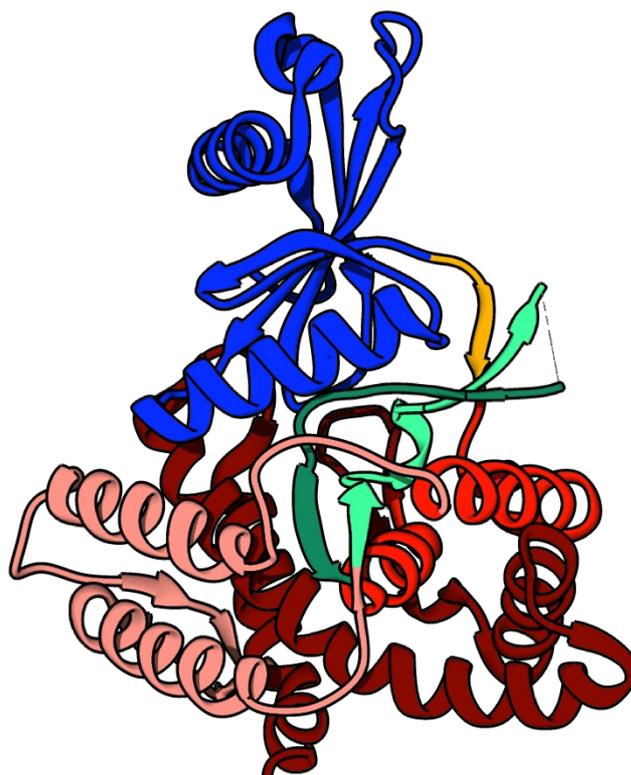


Figure 44: Structural features of the crystal structure of *C. Elegans* ChoK
The N-terminal region is shown in blue and the β strand linker in yellow. The C-terminal domain is further subdivided into the central core subdomain (Brenner's motif in light green and choline kinase motif in dark green), the insertion subdomain (pink) and the C-terminal subdomain (dark red).

Since the structure does not have any bound substrates, the authors could only speculate on the location of the ATP and choline binding sites, their hypotheses being based on the structure similarity between different ePKs. They identified a putative ATP binding site mainly located in the N-terminal domain, the phosphotransferase Brenner's motif, and a choline binding site in the C-terminal domain, as these are both highly conserved sequences.

More accurate information on the ATP and choline binding sites were provided in 2007 by Malito et al., as they determined the crystal structures of the isoform $\alpha 2$ of human ChoK in its apo- and product-binding (ADP and phosphorylcholine) forms (PDBIDs: 2CKO, 2CKP, and 2CKQ for respectively the apo-, ADP- and phosphorylcholine-bound forms) [148]. This human enzyme displays a sequence identity of 42% to the primary sequence of the *C. Elegans* ChoK, and, not surprisingly, it adopts the same overall folding topology and oligomerization state (homodimer). ADP was found to bind in a cleft between the N- and C-terminal domains and conserved residues

among the protein family from both domains participate in the adenosine nucleotide binding. Here, the adenine ring interacts with the backbone carbonyls of Gln207 and Ile209 and with the side chain of Glu206 via hydrogen bonds and it is sandwiched between Leu144 and Leu313, which help its positioning inside the pocket through hydrophobic interactions. All these interactions provide an explanation for the specificity of ChoK towards adenine nucleotides and are shown in the panel A of Figure 45. The ribose moiety of ADP forms hydrogen bonds with the backbone carbonyl group of Ser211 and the backbone amide of Arg213 (whose side chain is, however, not traceable in the electron density map and therefore it was modeled as Ala). An important contribution to the binding of ADP is provided also by the hydrophobic interaction with Leu214, which helps the correct positioning of the ribose (panel B of Figure 45). The α -phosphate group of ADP makes a salt bridge with the side chain of Arg146 and a hydrogen bond with Asp330. This latter residue is catalytically essential, and it is believed to be involved in the coordination of two magnesium ions that participate, as cofactors, in the phosphotransfer reaction; however, these ions are not present in the active site of the enzyme. The β -phosphate group hydrogen bonds the side chains of Arg117 and Asn122, these two residues being part of the so-called P-loop that is a necessary but not sufficient structural motif for phosphate-binding [149]. The interactions made by the two phosphate groups of ADP are depicted in panel C of Figure 45.

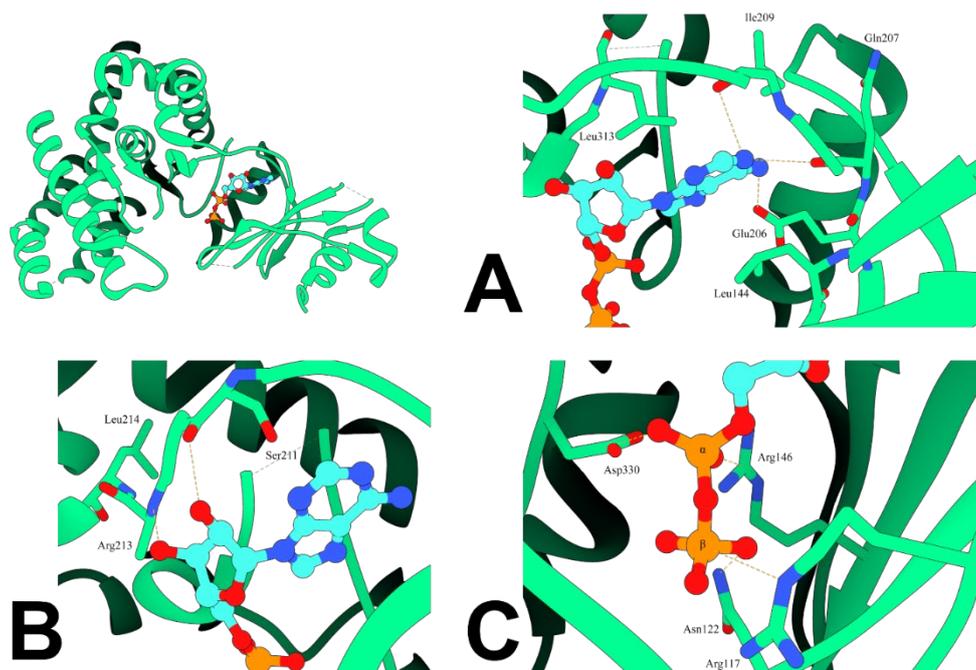


Figure 45: Interactions of the ADP molecule with human ChoK

On the top left there is a light green cartoon representation of PDBID 2CKP, with the ADP molecule drawn in ball and stick with light blue carbon atoms.

The interactions described in the text are depicted in panels A, B, and C, and they refer to the adenine, ribose, and phosphate groups respectively. Involved residues are indicated by their three name code and residue number, hydrogen bonding and electrostatic interactions are indicated by dashed orange lines and α and β phosphate groups are labelled accordingly.

The structure of ChoK with phosphorylcholine (PDBID: 2CKQ) allows the characterization of the choline binding site, which mainly involves some residues belonging to the Brenner's motif (residues from 302 to 311) and of the choline kinase motif (residues from 326 to 354). The architecture of the choline binding site consists of a deep hydrophobic pocket (Tyr333, Tyr354, Tyr440, Trp420, and Trp423) contoured by a series of negatively charged residues (Glu215, Glu217, Glu218, Glu309, Glu349, Asp353, Glu357, and Glu434). These charged residues steer the positively charged quaternary ammine into the choline binding site and stabilize it through electrostatic interactions, while the hydrophobic pocket participates in the correct orientation of the substrate (panel A of Figure 46). The phosphate group of phosphorylcholine forms hydrogen bonds with the side chains of Asp306, Gln308, and Asn311 and with the backbone amide of Leu120 (panel B of Figure 46). Water mediated hydrogen bonds are also present at the base of the choline binding site and involve two water molecules that put Asp306 in contact with Asn345 and Glu349 on the opposite side of the pocket.

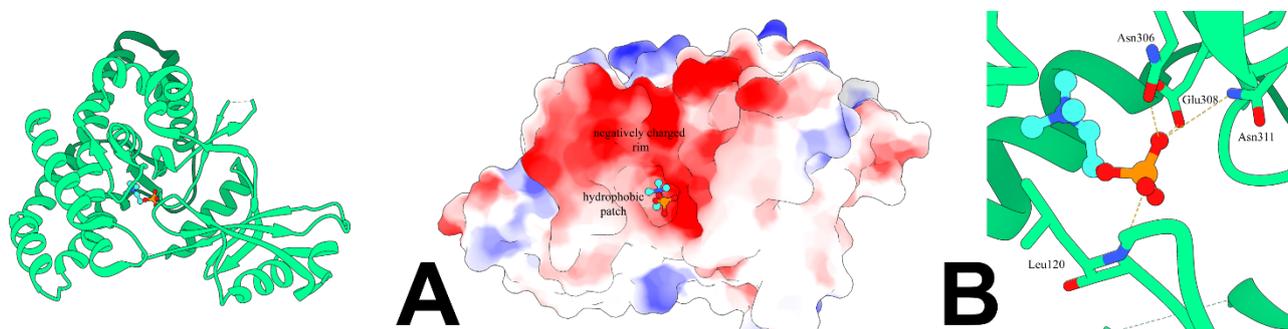


Figure 46: Interactions of the phosphorylcholine molecule with human ChoK
 On the left, a light green cartoon representation of PDBID 2CKQ, with the phosphorylcholine molecule drawn in ball and stick with light blue carbon atoms.
 The interactions described in the text are depicted in panels A and B. In panel A, the electrostatic surface of the choline binding site is displayed and the hydrophobic and negatively charged regions are labelled accordingly. In panel B, the interactions of the phosphate group are shown. The involved residues are indicated by their three-name code and residue number, and hydrogen bonding interactions are indicated by dashed orange lines.

The comparison between the ADP- and phosphorylcholine-bound forms of the enzyme shows that the enzyme undergoes important conformational changes upon substrate binding and catalysis as it has been observed in other ePKs [150]. In particular, the N-terminal domain undergoes a rotation of about 16° upon phosphorylcholine binding, the P-loop switches to a more closed conformation compared to the ADP-bound form, while the C-terminal domain retains the same orientation notwithstanding which product it binds.

These structural data helped in the definition of the reaction mechanism of ChoK, which has been established through the study of the substrates kinetics [151]. For most PKs, the transfer of the phosphate group from ATP to the substrate proceeds through the formation of a ternary complex that

involves both substrates at the same time [152], [153], [154], [155], [156], [157], [158]. However, some aPKs have been demonstrated to catalyze their reaction via a double displacement (ping-pong) mechanism [159], [160], [161]. In their work, Hudson et al. [151] established that the phosphotransfer reaction catalyzed by human ChoK proceeds with a ping-pong mechanism when the first product (ADP) must be dissociated before the second substrate (choline) can bind. Hence, the reaction must proceed through a phospho-enzyme intermediate.

The Biological Significance of Human Choline Kinase

According to the HUGO database [162], [163], the human genome contains two genes for ChoK, CHKA and CHKB, which encode for three isoforms: $\alpha 1$, $\alpha 2$ and β [164], [165]. Both genes are ubiquitously expressed in the cytoplasm of human cells where their active form is either homo- or heterodimeric. Interestingly, in mouse liver tissue the heterodimeric form is the most active, accounting for 60% of ChoK activity [166]. The isoform $\alpha 1$ differs from the $\alpha 2$ one by a twenty residue insertion that is absent in the $\alpha 2$ form, and it shares a 60% sequence identity with the isoform β . Both isoforms have the ability to phosphorylate either choline or ethanolamine. However, isoform β appears to be able to affect the production of phosphorylethanolamine [167]. The reported values for the kinetic parameters for the ChoK isoforms are reported in Table 5.

Table 5: Michaelis-Menten constants for the two ChoK isoforms. Data obtained from the BRENDA database [168]

Isoform	K_M ATP [mM]	K_M Choline [mM]	K_M Ethanolamine [mM]
Isoform α	0.4	0.2	12
Isoform β	0.4	0.57	2.9

In 1956, Kennedy and Weiss discovered that ChoK activity is related to lipid metabolism and, in particular, to the biosynthesis of two important phospholipids: phosphatidylcholine (PC) and phosphatidylethanolamine (PE) [169]. The two researchers discovered two related biochemical pathways (later called “Kennedy” pathway) that are ubiquitously found in eukaryotes and lead to the production of PC and PE through a three-step reaction pathway. Since PC and PE are required for the formation of two high-energy intermediates (CDP-ethanolamine for PE and CDP-choline for PC), the pathways are also named CDP-ethanolamine or CDP-choline, respectively [170].

The CDP-ethanolamine and CDP-choline pathways are highly similar, as they use a series of similar reactions that only differ in the involvement of two different initial substrates, either choline or ethanolamine (Figure 47).

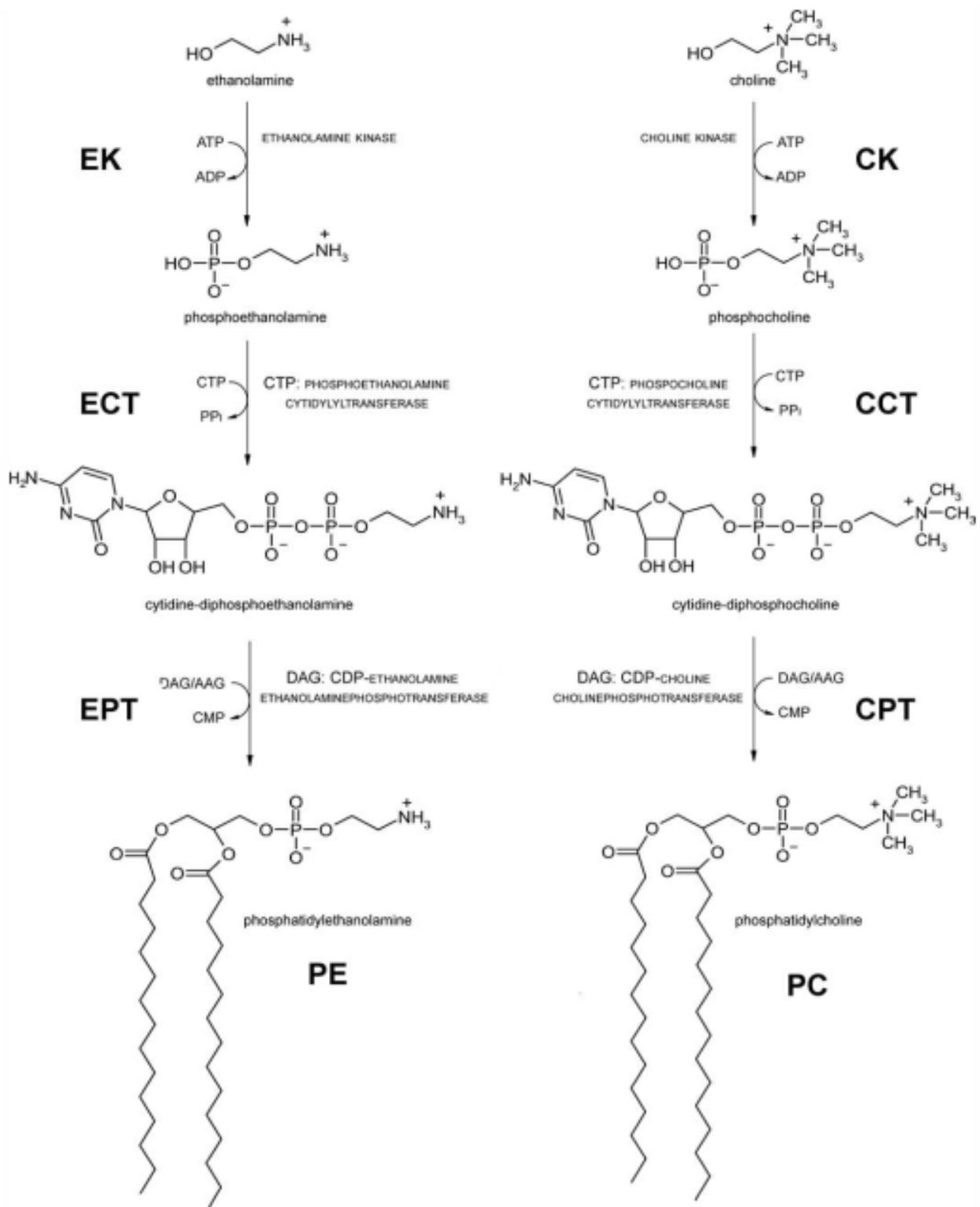


Figure 47: The Kennedy pathway
Schematic representation of the Kennedy pathway with both choline and ethanolamine as initial substrates. Adapted from [168].

The first enzymatic reaction consists of the ATP-dependent phosphorylation of either choline or ethanolamine that is performed by ChoK or Ethanolamine Kinase, respectively, and that lead to the formation of a phosphorylated substrate and ADP as a side-product [171]. Phosphorylcholine is a very important metabolite: it is very stable in physiological conditions, and it is essential for the induction of DNA synthesis, thus affecting the growth factor-stimulated cell proliferation [172]. The second step is considered to be rate-limiting for the pathway [173] and it involves a reaction catalyzed by either CTP:phosphorylethanolamine- or CTP:phosphorylcholine cytidylyltransferase. This results in the transfer of CDP to the substrate with the release of the phosphate group, forming the high-energy intermediates CDP-ethanolamine and CDP-choline [166].

The final step is performed by two enzymes (CDP-ethanolamine:1,2-diacylglycerol ethanolaminephosphotransferase or CDP-choline:1,2-diacylglycerol cholinephosphotransferase), which catalyze the transfer of ethanolamine or choline to two lipid anchors (which can be either diacylglycerol and alkyl-acylglycerol), releasing CMP as a byproduct and forming either PE or PC [171].

Beside their great similarity, the two pathways also feature a great degree of overlapping. For instance, despite being directed toward choline, ChoK has the ability to phosphorylate ethanolamine too and some phosphotransferases have been shown to be capable of using both substrates for the synthesis of PE and PC (see Table 5) [174], [175], [176]. Furthermore, these two pathways are connected by a third one, which enables the production of PC starting from PE. This third pathway is usually referred to as the PE methylation pathway and is mediated by PE methyltransferase, which catalyzes all the three methylations that are necessary for the conversion. However, the PE methylation pathway seems to have a significant effect on PC levels only in hepatocytes, where it accounts for almost 30% of the hepatocyte PC content [177].

PE and PC are two ubiquitously present phospholipids in human cell membranes, where they account for a variable range (20-30% and 30-60%, respectively) of the phospholipid content of various cell membranes [178].

From the molecular point of view, since PE is characterized by a small ethanolamine head group and long, unsaturated fatty acid tails, it adopts a conical shape. Lacking a charged head group and thanks to its shape, it induces a negative curvature in membranes, and its presence in cell membranes as a minor component compared to PC determines the formation of discontinuities in the fluid lipid bilayer. The discontinuities that are due to its presence have been reported to favor membrane fusion events and the molecular folding of some membrane proteins. Moreover, the reactive primary amino

group of its head is also able to form protein and lipid conjugates, affecting a substantial number of biological pathways [178], [179].

On the other hand, PC is a zwitterionic phospholipid that accounts for a considerable amount of the phospholipid mass of eukaryotic cell membranes. Its cylindrical molecular shape is one of the determinants of the formation of the membrane bilayer structure and its quaternary amine head is essential for PC function, since its replacement with secondary or primary amines has negative effects on cell viability. PC acts as cellular reservoir for second messengers such as fatty acids, diacylglycerol, phosphatic acid and lysophosphatidylcholine, which can be released in the cell upon arrival of a primary signal that activates the lipase enzymes. It is also a precursor of other phospholipids such as phosphatidylserine and sphingomyelin. It is not surprising that inhibiting its synthesis or depriving the cells of its precursors results in growth arrest and apoptosis, since cells are no longer capable of building new membranes and their signaling cascades are disturbed. Its scarce availability results in the inhibition of the secretion of some lipid-protein complexes and ultimately interferes with essential cellular functions [179], [180], [181], [182].

Implications of Choline Kinase in Pathological Conditions

Given the central role of ChoK in human lipid metabolism and the importance of the final products of the Kennedy pathway for the biosynthesis of the plasma membrane, it is not surprising that dysregulation of ChoK activity or perturbation of the choline metabolism can result in the onset of pathological conditions.

Over the years, abnormal choline metabolism has emerged as a relevant hallmark for malignancy [183] to such an extent that PET choline imaging has substituted the classical ^1H magnetic resonance spectroscopy in oncological applications, including detection and the possibility to evaluate the response to treatment [184], [185], [186], [187].

The molecular causes for the abnormal choline metabolism are to be looked for in the increased phosphorylcholine and cytoplasmatic free choline levels in cancer cells and solid tumors. These abnormal levels of metabolites have different and often concurring causes. For example, choline is transported into the tumor cells with a higher rate compared to normal cells [188], [189]. Moreover, there is an increased activity of PC-targeted phospholipases that breaks down PC into its single components, thus freeing them into the cell [190], [191], [192], [193]. Finally, an increased ChoK expression level and activity (especially of the isoforms α) [190], [191], [192], [194], [195], [196], [197].

A complete picture of the most relevant enzymes involved in the choline metabolism are presented in Figure 48, and, as it is shown in Figure 49, numerous oncogenic pathways share reciprocal interactions with them [198]. The two oncogenic signaling pathways that are able to affect both the choline metabolism-related enzymes activity and their transcription levels are the RAS pathway and the PI3K-AKT pathway.

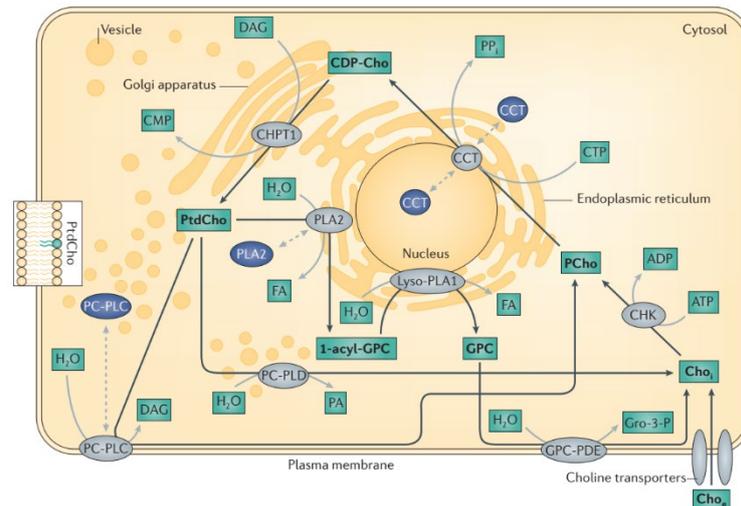


Figure 48: Key enzymes involved in the choline metabolism

Enzymes are indicated in grey shapes in the correspondence of the organelle they work in. Inactive enzyme forms are indicated in blue shapes; they are activated upon translocation to a different organelle or cellular compartment. Black arrows indicate metabolic reactions and substrates, and products are written in bold and indicated by green rectangles. Dashed arrows indicate the translocation to a different organelle or cellular compartment. Abbreviation list: CCT, CTP: phosphocholine cytidyltransferase; CDP-Cho, cytidinediphosphate-choline; CHKa, choline kinase- α ; Cho_e, extracellular free choline; Cho_i, intracellular free choline; CHPT1, diacylglycerol cholinephosphotransferase 1; CMP, cytidine monophosphate; CTP, cytidine triphosphate; FA, fatty acid; GPC, glycerophosphocholine; GPC-PDE, glycerophosphocholine phosphodiesterase; Gro-3-P, glycerol-3-phosphate; Lyso-PLA1, lysophospholipase A1; PCho, phosphocholine; PC-PLC, phosphatidylcholine-specific phospholipase C; PC-PLD, phosphatidylcholine-specific phospholipase D; PLA2, cytoplasmic phosphatidylcholine-specific phospholipase A2; PP_i, diphosphate. Adapted from [181].

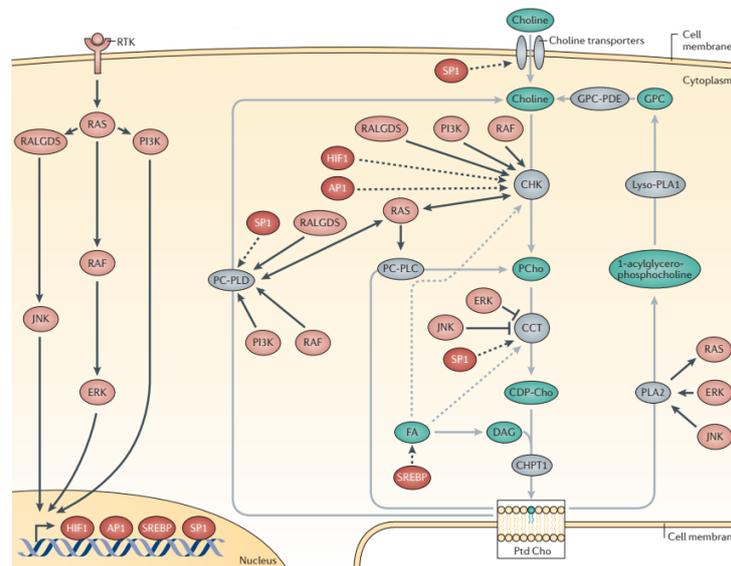


Figure 49: Oncogenic signaling pathways and choline metabolism

On the left side of the picture, oncogenic pathways that affect the choline metabolism (on the right side) are depicted. Grey arrows and grey shapes represent the enzymes involved with the choline metabolism, while red shapes represent the oncogenes. Dashed grey lines represent regulatory pathways in the choline metabolism. Black arrows indicate a connection between an oncogene and the choline metabolism; solid lines represent increased or decreased enzyme activity, while dashed ones indicate increased gene expression. Abbreviation list: API1, activator protein 1; CHK, choline kinase; CHPT1, diacylglycerol cholinephosphotransferase 1; CCT, CTP:phosphocholine cytidyltransferase; DAG, diacylglycerol; FA, fatty acid; GPC, glycerophosphocholine; HIF1, hypoxia-inducible factor 1; JNK, JUN N-terminal kinase; PC-PLC, phosphatidylcholine-specific phospholipase C; PC-PLD, phosphatidylcholine-specific phospholipase D; PLA2, phosphatidylcholine-specific phospholipase A2; PCho, phosphocholine; PtdCho, phosphatidylcholine; RALGDS, RAL GTPase guanine nucleotide dissociation stimulator; RTK, receptor tyrosine kinase; SREBP, sterol regulatory element binding protein- Adapted from [181].

Aberrant or increased levels of ChoK α expression and activity have been observed in a variety of human cancers: such as, breast [199], [200], [201], [202], [203], [204], lung [187], [196], [205], [206], colorectal [207], [208], [209], ovarian [191], [192], [210], [211], [212], endometrial [213], pancreatic [214], [215] and prostate [216], [217], [218], [219], [220], [221] cancer.

In many cases, siRNA-based downregulation or inhibition of the enzyme resulted in the arrest of tumor cell proliferation and decreased invasiveness and migration [195], [222], [223], [224]. However, the antiproliferative effect of the inhibition of ChoK α is not enough to determine the cancer cell death alone; but, when it is accompanied by other antitumoral therapies, it greatly increases their efficacy [225], [226].

It has been estimated that between 40 to 60% of human tumors are accompanied by ChoK α overexpression [227] and high levels of ChoK α in the cytoplasm of cancer cells are associated to poor prognosis [196]. Moreover, the overexpression of ChoK α can promote oncogenesis even in noncancerous cells and its non-catalytic role as scaffold for the assembly of protein signaling complexes promotes cancer cell survival [228].

If on one hand ChoK α plays a relevant role as a pro-oncogenic agent and its selective inhibition has antiproliferative effects on tumor cells, on the other, ChoK β has also been recently characterized as a novel potential therapeutic target in muscular dystrophy, bone deformities and narcolepsy [229].

In 2006, Sher et al. [230] were the first to show that a recessive mutation in the CHKB gene in mouse caused an intragenic deletion and therefore resulted in the production of an inactive ChoK β enzyme, causing rostrocaudal muscular dystrophy and neonatal forelimb bone deformity. The loss of ChoK β physiological activity resulted in an impairment of PC biosynthesis and the consequent accumulation of choline inside cells, decreased phosphorylcholine concentration, and an increased PC degradation. Mice presenting an aberrant ChoK β had abnormally large mitochondria that were characterized by a lower inner membrane potential and an aberrant cellular localization in muscle cells [231], [232].

Over the years, many CHKB mutations have been discovered and characterized in human cases of muscular dystrophy globally [233], [234], [235], [236], [237], [238], [239], [240], [241], [242], presenting characteristics and symptoms similar to the ones observed in mice (mitochondria enlargement and aberrant localization, loss of mitochondria or mitochondria functions, loss of components of the electron transport chain and progressive deterioration of muscle tissue).

A genome-wide association study also linked some genetic variants of CHKB to a predisposition and a susceptibility to narcolepsy [243] and other studies in mice [244], [245], linked the activity of ChoK β to ossification, bone formation and homeostasis and the loss of ChoK β has been connected to

bone defect formation and osteoporosis. While the molecular causes of this connection are not known yet, it seems that the lower levels of PC might be partially responsible for these pathologies.

To date, several small molecule compounds have been developed to target ChoK α to exploit the resulting antiproliferative and antitumoral effects of its inhibition. A special effort has also been directed towards the development of highly specific molecules that are able to target ChoK α but not ChoK β , since this isoform has not been implicated in the promotion of tumor formation and therefore its inhibition might impair the mitochondrial function and alter the normal physiology of muscle cells. Contrary to ChoK α , there are no data on ChoK β inhibitors in the scientific literature. Moreover, the emerging strategy to target the pathologies caused by ChoK β deficiency is to bypass it by externally providing CDP-choline, a product related to ChoK β catalytic cascade. Supplementation of CDP-choline inside ChoK β deficient mice determined an increase in the PC content inside muscles, partially rescued the animal from muscle weakness and could elicit a cell membrane repairing effect in aged animals [231], [246].

CDP-choline administration might also be an effective treatment for muscular dystrophy and for the mitigation of the effects of osteoporosis, since it has been shown to attenuate oxidative stress-induced cell death in human cardiac myocytes [247] and to restore to physiological levels the number of osteoclasts *in vitro* and *in vivo* in mouse models [245].

All the data here presented, highlight the importance of ChoK as new pharmacological target that might be important not only in the treatment of such pathologies but also as a diagnostic marker for early cancer detection.

Human Choline Kinase α Inhibition

There is a great deal of literature on ChoK α inhibition in tumor drug discovery [248], [249], [250]; Hence, for the sake of simplicity, here I will focus on the roughly 20 years of ChoK α inhibitors study that lead to the synthesis of the library of compounds that I worked on for this thesis.

The first ChoK α specific inhibitor that was discovered is Hemicholinium-3 (Hc-3) (Figure 50, panel A). It was initially identified as a mouse ChoK inhibitor by Spanner et al. in 1973 [251]. In their study on mouse brain, they determined that Hc-3 affected PC biosynthesis by targeting mouse ChoK (*in vitro* IC₅₀ of 20 μ M calculated on the mouse enzyme). Other studies [252], [253], [254], revealed that Hc-3 was also able to impair choline transport by acting as competitive inhibitors of choline transporters.

Hc-3 is a molecule composed by a biphenyl ring with two choline-like groups containing positively charged quaternary amine groups connected to its extremities. In solution, the structure of the two choline-like groups of Hc-3 closes on itself forming two oxazonium rings [255].

The inhibitory activity of Hc-3 for the human ChoK enzyme was characterized only later and the compound showed an IC₅₀ value of 500 μ M against the purified enzyme [256] and the ability to determine an antiproliferative effect with an EC₅₀ of 2.5 mM on human colon adenocarcinoma cells HT-29 [257]. Unfortunately, beside featuring a substantially low potency, Hc-3 shows also respiratory toxicity, since at effective concentrations it causes respiratory paralysis and failure in animal models, which are due to an impairment of the activity of cholinergic nerve terminals [255]. However, the structure of Hc-3 (PDBIDS: 3F2R) became the starting point for the design of a great variety of new ChoK inhibitors being synthesized to obtain higher potencies and to mitigate the toxic effects of Hc-3. In this process, the elucidation of the binding mode of Hc-3 on human ChoK isoforms has greatly helped in guiding the design of new compounds [258]. The first structural data on Hc-3 interactions with ChoK became available in 2010 and it showed that the compound is able to inhibit both ChoK α and ChoK β , with higher specificity towards the α isoform than the β one. The measured K_i value was 0.21, 0.23 and 116 μ M towards ChoK α 1, ChoK α 2 and ChoK β , respectively. Hence, the inhibitory effect was ~500 times higher towards the α isoforms than to the β one. Furthermore, the co-crystal structures of both isoforms revealed that the binding occurred in the same place, a groove in the C-terminal lobe near a cleft, and one oxazinium ring occupied the choline-binding site and the other was exposed to the solvent. In this region, the hydrophobic biphenyl group of Hc-3 interacted with a series of hydrophobic residues that lined the groove (Figure 50, panels B).

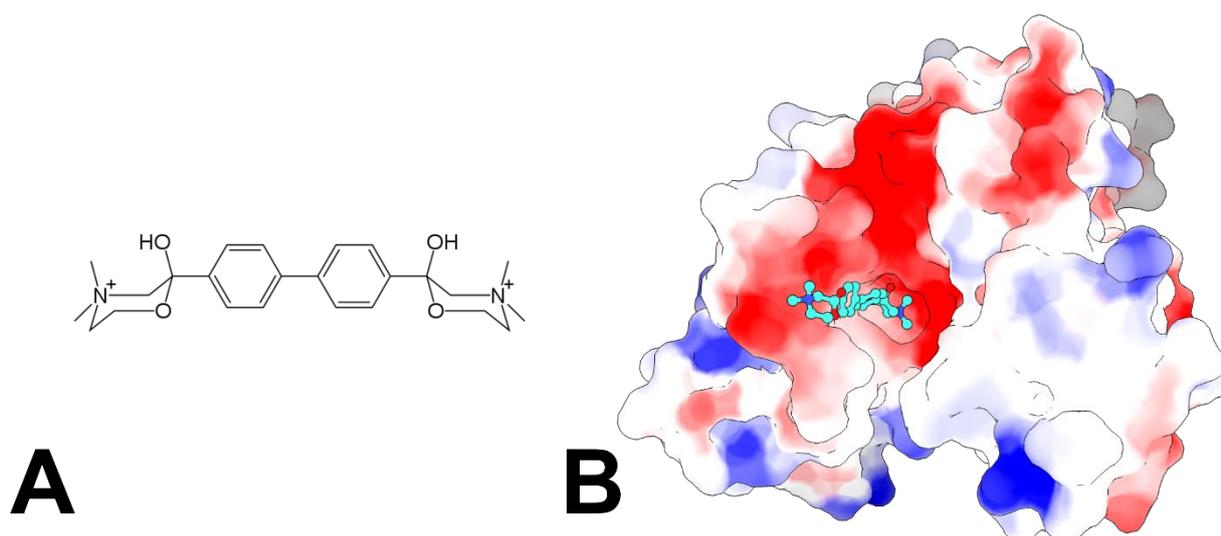


Figure 50: Binding mode of Hc-3

Panel A shows the molecular structure of Hc-3 in ball and stick model. Panel B depicts the crystal structure from PDBID 3F2R with the protein surface colored accordingly to the electrostatic potential and the Hc-3 molecule in ball and stick model.

Interestingly, contrary to the structure of the ChoK α complex, the structure of the ChoK β complex featured a phosphorylated Hc-3, leading the authors to conclude that, albeit through a slow process, ChoK β is able to phosphorylate Hc-3 while the α isoform is not. The authors suggested that small differences in the two isoforms might account for a greater flexibility of the Hc-3 binding groove in ChoK β than in ChoK α , which ultimately results in a better accommodation of the inhibitor in the choline-binding site, hence, in its phosphorylation.

An initial attempt to improve the potency of Hc-3 was made by Hernández-Alcoceba et al. in 1997, when they synthesized two different Hc-3-derived generations of compounds [259]. The first series of derivatives was composed of symmetrical bis-quaternary derivatives that displayed a bis-phenacyl moiety linked to different ammonium and pyridinium rings as terminal fragments, while the other series was characterized by an either rigid or semirigid central group containing aromatic rings and two terminal fragments such as pyridinium, quinolinium or isoquinolinium cations.

Compared to Hc-3, these compounds showed a 10-1000-fold increase in potency *in vitro* (IC₅₀ values ranging from 90 μ M to 1 μ M) and a 30-2500-fold increase in *ex vivo* antiproliferative potency in Ras-transformed NIH3T3 cell lines. This study demonstrated that terminal pyridinium groups provided the best results in terms of ChoK inhibition and antiproliferative effect, but also that the lipophilicity of the compounds played an important role in determining its potency. Follow up studies focusing on the possibility to substitute pyridinium and quinolinium rings at position 4 with different functional groups led to the development of an extensive library of symmetrical bis-quaternary compounds [260]. A quantitative structure-activity relationship analysis demonstrated that strong electron-donors

and highly lipophilic groups provided the most potent compounds in terms of antiproliferative effects on HT-29 cells.

The effects of mono-quaternary compounds based on the previously analyzed symmetric ChoK inhibitors were also evaluated, and it was demonstrated that these compounds were less potent ChoK inhibitors, often displaying also high toxicity. Therefore, it was concluded that, albeit not strictly necessary to have ChoK inhibitory effects, the presence of the second charge greatly enhanced the potency of the compound [261].

The effect of the distance between the two positively charged quaternary groups in symmetric compounds was also evaluated by Conejo-García et al. in [262] via the synthesis of different compounds with central moieties of increasing lengths. It was established that the optimal distance corresponded to the 3,3'-bisphenyl spacer. Moreover, it was observed that the potency of the compound increased with increasing length of the central linker and once it exceeded the optimal distance, the inhibitory potency dropped.

The effect of the number of charges on the head of these compounds was also considered [263]. Since monocationic compounds have been demonstrated to have low potency and poor antitumoral activity, a new library of bis- and triscationic compounds with substituted pyridinium groups was synthesized and their *ex vivo* potency and *in vitro* antiproliferative activity on the HT-29 cell line were evaluated. The study showed that while the introduction of the third charge determined a 7- to 35-fold increase in the *ex vivo* potency compared to the biscationic compounds, the loss in hydrophobicity of the triscationic compounds determined a drastic decrease in the antiproliferative activity of these compounds. The increase in their potency was attributed either to a statistical effect due to the presence of an additional binding group that results in the presence of more binding groups near the enzyme and ultimately a more favorable entropy, or to the presence of another binding site for the additional pyridinium group. Whatever the reason for the improved *ex vivo* potency of these compounds, their poor antiproliferative activity greatly hinders their possible use and further development.

Another strategy that was adopted to increase the efficacy of this library of compounds was to rigidify their molecular structure by taking the highly flexible conformation of these molecules and locking it into a fixed conformation via a cyclization reaction [264]. The activity of these compounds was also evaluated *ex vivo* and *in vitro* on HT-29 cells, but, despite their higher *ex vivo* inhibitory potency, they displayed modest *in vitro* antiproliferative activity.

Of all the synthesized compounds of the so-called first generation of Hc-3 derivatives, one called MN58b stood out with exceptionally high antiproliferative effects measured against a panel of

different human cancer cells (colon adenocarcinoma, carcinoma of the cervix, epidermoid carcinoma, breast cancer, histiocytic lymphoma, and chronic myeloid leukemia) at effective concentrations of less than 2.1 μM . Furthermore, the compound showed a well-tolerated toxicity profile in *in vivo* experiments of human tumor xenografts in mice, featuring also good antiproliferative effects [265]. MN58b showed a stronger specificity towards ChoK α (IC_{50} of 5 μM) compared to ChoK β (IC_{50} of 107.5 μM , ~ 22 times more potent on the α isoform) [167] and, while causing a reversible cell cycle arrest in normal cells, it caused an irreversible G₀-G₁ cell cycle arrest in tumor cells that caused oscillations in the availability of phospholipids (especially PC), loss of mitochondrial membrane potential and, ultimately, a promotion of apoptosis in malignant cells [266], [267].

The information and data obtained from the first Hc-3 derivatives allowed Lacal and colleagues to synthesize a new library of rationally designed compounds [268]. The compounds in this library showed potent *in vitro* antiproliferative and *in vivo* antitumoral activity against different human tumor xenografts in mice. The IC_{50} values against ChoK α and ChoK β were calculated from bacterial extracts expressing one of the two proteins and values in the range from 0.1 to 7 μM were obtained for ChoK α , while values in the range 33 to more than 50 μM were obtained for ChoK β . Hence, the compounds showed from ~ 7 to 400 times higher selectivity towards the α isoform. Against the different human tumor cell lines that were tested, the efficacy was high, and the toxicity in mice was low. Of all compounds, RSM-932A (IC_{50} of 1 and 33 μM for ChoK α and ChoK β , respectively) was selected as the best candidate in terms of potency and safety profile to be tested in the first phase I clinical trial on ChoK inhibition, where it was renamed TCD-717 [269].

Whereas no structural data are available on MN58b, in 2018 the co-crystal structure of TCD-717 with ChoK α was published by Kall et al. [270]. TCD-717 was shown to constrict conformational changes of ChoK α by binding across the N- and C-terminal lobe and determining several conformational changes in the side chains of the residues in its proximity (Figure 51). Interestingly, contrary to expectations, TCD-717 does not bind inside the choline-binding site but immediately under it, causing conformational changes in the hydrophobic amino acids that were observed to interact with Hc-3 biphenyl ring.



Figure 51: Binding mode of TCD-717

The picture shows the binding mode of TCD-717 that is drawn in ball and stick model: carbon atoms are cyan; nitrogens are blue and chlorines are green. The enzyme body is depicted in light green ribbon.

The main feature of all the compounds described so far is their symmetry: they are all characterized by symmetrical cationic heads. The second generation of ChoK α inhibitors was based on the synthesis of asymmetrical monocationic compounds that would be able to bind into the choline binding site with the charged head and to bind the ATP binding site with the other head, which needs to be designed in such a way to mimic ATP [271]. This family of compounds is composed by an adenine head to mimic ATP and either 4-dimethylamino- or 4-pyrrolidino-pyridinium charged group separated by linkers of different length (benzene, biphenyl, 1,2-diphenylethane and 1,4-diphenylbutan moieties). When tested, all the synthesized compounds presented good potency and antiproliferative activity on HepG2 cells (hepatocellular carcinoma). The crystal structure of the complex between ChoK α and one of these compounds (named compound 1 in this study) was determined (PDBID: 3ZM9) (Figure 52). As expected from computational studies, the compound was found to bind both substrate-binding sites, inducing an allosteric coupling characterized by a negative cooperativity in binding both ChoK α monomers and breaking the crystallographic symmetry between the two monomers [272].

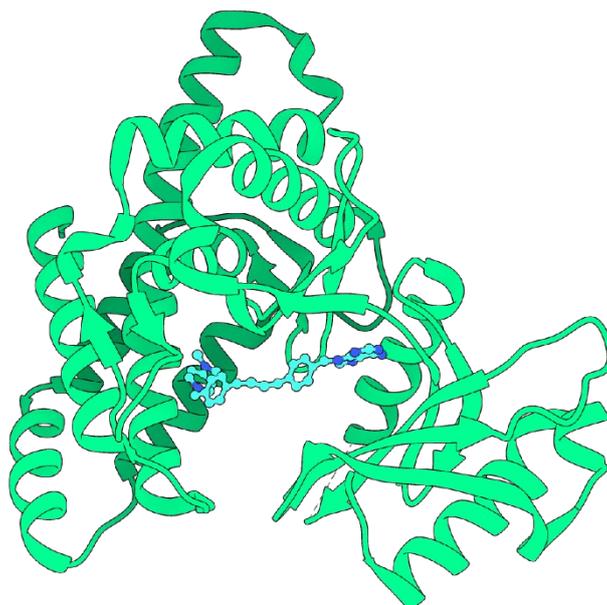


Figure 52: Binding mode of compound 1

The picture shows the binding mode of compound 1 that is drawn in ball and stick mode: carbon atoms are cyan, and nitrogens are blue. The enzyme body is depicted in light green ribbon.

Another asymmetrical compound of the series (named compound 2 in the study) was also co-crystallized with ChoK α (PDBID: 4BR3) [273]. The structure of this compound is characterized by a short linker featuring a biphenylic group that connects an adenine and a 1-4-(dimethylamino)pyridinium group. Its peculiarity is that in the crystal structure of the complex it is observed both in the choline binding site and in the ATP binding site, interacting with them via the two different head groups (Figure 53). The choline binding site is efficiently occupied by the 1-benzyl-4-(dimethylamino)pyridinium moiety, which interacts with Tyr333, Tyr354, Tyr423, Phe435 and Trp423 via hydrophobic, π - π , and cation- π interactions, while the ATP binding site is contacted by the adenine moiety in a ATP-like fashion. The chemical deconvolution approach of the study allowed the identification of the adenine and the 1-benzyl-4-(dimethylamino)pyridinium moieties as the most efficient binding fragments in the library of asymmetrical compounds.

However, the antiproliferative activity of asymmetrical compounds was lower than the antiproliferative activity of symmetrical ones. To explore a new strategy, a new library of asymmetrical biscationic compounds was synthesized using variable linker lengths and combining derivatives of the most efficient head groups, 4-(4-chloro-N-methylanilino)-pyridinium and 1-benzyl-4-(dimethylamino)pyridinium with 4-pyrrolidinopyridinium fragments, as heads [274].

The *ex vivo* value of IC₅₀ for this series of compounds is in the low micromolar range (less than 2.5 μ M) and three of them (compounds 5, 12 and 14) were co-crystallized with ChoK α (Figure 54).

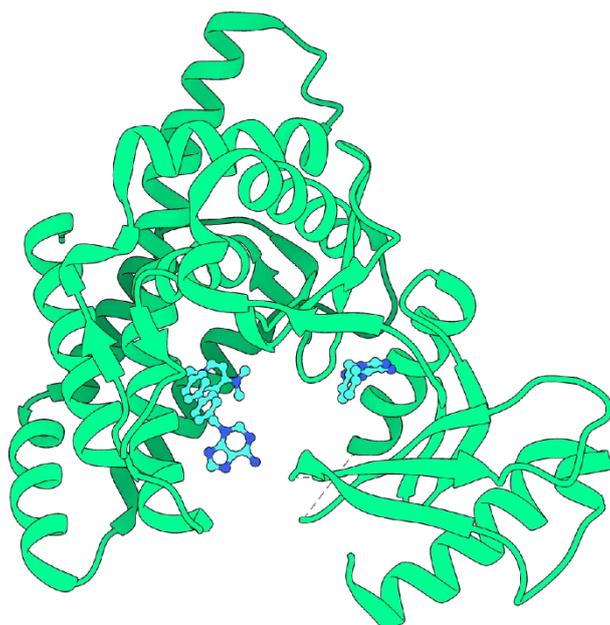


Figure 53: Binding mode of compound 2

The picture shows the binding mode of compound 2 that is drawn in ball and stick model: carbon atoms are cyan, and nitrogens are blue. The enzyme body is depicted in light green ribbon.

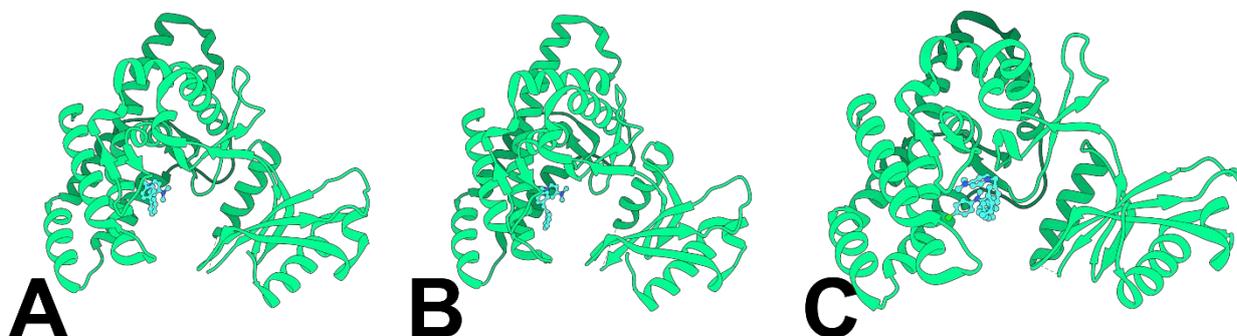


Figure 54: Binding mode of compounds 5, 12 and 14

The picture shows the binding mode of compounds 5 (A), 12 (B), and 14 (C) that are drawn in ball and stick model: carbon atoms are cyan, and nitrogens are blue. The enzyme body is depicted in light green ribbon.

Despite the suboptimal quality of the electron density maps in the compound-binding regions, in both 4CGA and 4CG9 it is possible to observe that the 1-benzyl-4-(dimethylamino)pyridinium moiety binds to the choline binding site in the same fashion as the compounds in Figure 52 and Figure 53 (Figure 54, panels A and B). Conversely, in 4CG8 the 4-chloro-N-methylaniline fragment binds deeply into the hydrophobic choline binding site, interacting with Trp248, Thr252, Tyr256, Tyr333, Leu419, Trp420, and Trp423, and affecting in particular the local conformation of Tyr256, Glu332, Tyr333, and Trp420. The compound assumes a twisted conformation thanks to the longer linker and to the 4-(dimethylamino)pyridinium moiety binding in a region just below Trp420 in the choline binding site, involving the hydrophobic residues Tyr333 and Tyr440 (Figure 54, panel C).

In vitro assays on HeLa cells demonstrate that this family of compounds is able to induce a cell cycle arrest, hence, to increase the number of apoptotic cells in cultures [274]. The compound in 4CG8 was the only one tested for selectivity towards ChoK β , featuring a 200-fold preference for the α isoform. A follow up paper [275] aimed at increasing the solubility of the compounds and at introducing hydrogen bonding interactions with the protein by adding electron donor or acceptor groups in linkers between biphenylic or bipyridinic rings through a bioisosteric replacement of carbon atoms with oxygen ones. A library of symmetrical and nonsymmetrical compounds with derivatives of the most successful cationic heads was synthesized and evaluated *ex vivo* and *in vitro* on a panel of nine different human cancer cell lines. The *ex vivo* IC₅₀ of the series ranged from 40 to 4.7 μ M and the *in vitro* GI₅₀ on the different cell lines ranged from low micromolar to nanomolar. Only one of them (PDBID: 5FTG) was successfully co-crystallized with ChoK α and shows the usual interaction of the 1-benzyl-4-(dimethylamino)pyridinium moiety with the choline binding site of the protein while the second head points away from the protein, interacting only with Phe361 and Ile433 (Figure 55).

In particular, this compound was shown to have higher sensitivity towards proliferating cells rather than quiescent ones and inhibited the cell growth through cell-cycle arrest. In a follow up paper [276], its antiproliferative activity was tested on a panel of eight human tumor cell lines and displayed GI₅₀ values encompassing the 500 and the 1 nM values, which are comparable to those of MN58b and TCD-717. It was established that the apoptotic effects on leukemia T-cells *in vitro* is due to an interference with MAPK and mTOR signaling.

The most recent series of compounds has been designed with a drug optimization strategy using symmetrical and asymmetrical compounds bearing the N,N-dimethylpyridinium moiety and a quinolinic ring as cationic heads using MN58b and TCD-717 as lead compounds. This series showed low toxicity in nonproliferating cells, *ex vivo* IC₅₀ values and *in vitro* GI₅₀ values in the low micromolar/high nanomolar range. The lead compound of the series was found to induce senescence in MDA-MB-231 cells, but did not promote apoptosis at concentration levels lower than 5 μ M.

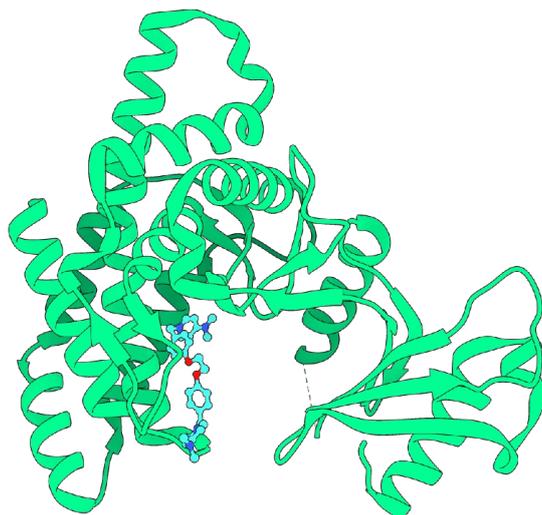


Figure 55: Binding mode of compound in PDBID 5FTG

The picture shows the binding mode of compound in 5FTG that is drawn in ball and stick model. The enzyme body is depicted in light green ribbon.

The Biological Significance of *Plasmodium falciparum* Choline Kinase

Malaria remains a major public health concern, especially in the African region, where most of the cases are reported. According to the latest “World Health Organization Malaria Report” [277], between 206 and 258 million cases were reported worldwide in 2018, resulting in an estimated death toll of 405,000, 67% of which occurred amongst children aged under 5, the most vulnerable population to the disease. Malaria is caused by the blood-parasites protozoa of the *Plasmodium* genus. Amongst the four human malaria parasites (*P. falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, and *Plasmodium malariae*) [278], *Plasmodium falciparum* is the deadliest and is accountable for the majority of the estimated cases [277].

Unfortunately, while conventional antimalarial remedies comprise artemisinin-based combination therapies, chloroquine, sulfadoxine-pyrimethamine, and amodiaquine [279], resistance to these conventional therapies has been reported worldwide [280], [281]. Hence, there is an urgent need for new antimalarial drugs to eradicate chemotherapy-resistant strains.

During their 48-hour long asexual intra-erythrocytic maturation cycle, *P. falciparum* merozoites go through different developmental stages (ring, trophozoite, and schizont state), which lead to 36 new daughter parasite cells [282]. Each of these cells is capable of beginning a new intra-erythrocytic life cycle, thus fostering the spread of the disease and causing the characteristic pathological symptoms of malaria (headache, fatigue, abdominal pain, fever, chills, perspiration, vomiting, metabolic acidosis, anemia, hypoglycemia, renal failure, and pulmonary edema) [279], [280], [283]. The intra-erythrocytic growth of the parasite is accompanied by substantial biochemical and biophysical changes in the host cell membrane. These changes are required in order to internalize large amounts of nutrients and to dispose of the waste products of the parasite metabolism [284], [285].

One of the most striking consequences of the blood-stage parasitization is the dramatic increase of the phospholipid content, especially PC and PE, inside the infected erythrocyte, which can reach up to five times its physiological level; this is a necessary requirement for the parasite membrane biosynthesis and, therefore, for its growth and replication [278], [286], [287], [288], [289], [290].

Interestingly, erythrocytes are enucleated cells that lack the enzymatic machinery for the *de novo* synthesis of phospholipids. Moreover, *P. falciparum* plasma membrane is mainly composed by PC and PE, which represent 40–50% and 35–40% of the total phospholipid content, respectively [291]. Hence, targeting the parasite lipid metabolism for pharmacological action constitutes an effective way of dealing with the spreading of the disease and with conventional therapy-resistant strains [278], [286], [287], [288], [289].

In *P. falciparum*, PC biosynthesis is operated by two distinct metabolic pathways: the Kennedy pathway and the serine decarboxylation-phosphoethanolamine methylation pathway [292], [293]. The former is responsible for 89% of PC synthesis, while the latter accounts for the remaining 11% [294]. Even in *P. falciparum*, the Kennedy pathway consists of a three-step enzymatic cascade, whereby choline is transformed into PC through the action of ChoK, CTP:phosphocholine cytidylyltransferase, and choline/ethanolamine phosphotransferase [170].

Similarly to its human counterpart, *P. falciparum* choline kinase (PfChoK) has also been demonstrated to be more selective to choline but also capable of phosphorylating ethanolamine [295], [296] and, likewise, it has been demonstrated that it also catalyze the phosphorylation of choline via a ping-pong mechanism [297]. PfChoK is expressed and localized in the cytoplasm of the parasite and displays K_m for choline and ATP of $145 \pm 20 \mu\text{M}$ and $2.5 \pm 0.3 \text{ mM}$, respectively [298].

Human Choline Kinase α Inhibitors Repurposed as *Plasmodium falciparum* Choline Kinase Inhibitors

Targeting the PC biosynthetic pathways by designing choline analogues has been shown to be lethal for the parasite in *in vitro* experiments, and ChoK has been proved to be a cytosolic marker for human erythrocyte-infection by *P. falciparum* [299], [300], [301], [302].

Over the years, several human ChoK inhibitors have been also tested on PfChoK and demonstrated that its selective inhibition causes the asexual intra-erythrocytic cycle of *P. falciparum* to arrest and results in the death of the parasite [297], [303], [304], [305], [306].

The choice to repurpose some human ChoK α inhibitors to inhibit the activity of PfChoK derived from the observation that the parasite enzyme catalytic regions share 67 % identity to the human ones [304]. One monocationic compound formed by a biphenyl linker connected to the cationic dimethylpyridinium moiety and an asymmetric biscationic one featuring the same linker and two cationic heads containing a combination of quinolinium and azepane moieties showed *ex vivo* IC₅₀ values against PfChoK in the micromolar range, comparable to those of MN58b and TCD-717 [304]. A subsequent library of compounds derived from second-generation inhibitors of human ChoK α featured IC₅₀ values in the low nanomolar range *in vitro* in erythrocytes infected with the parasite. The parasite cell cycle arrest was demonstrated to be due to the inhibition of PfChoK and the compounds bearing the dimethylpyridinium fragment displayed competitive inhibition with respect to choline. The drastic decrease of PfChoK activity was related to a low availability of PE and PC and therefore to parasite death [305].

Aim of the Project

This project has been conducted in collaboration with the research group of Dr. Luisa Carlota Lopez Cara of the Department of Pharmaceutical and Organic Chemistry of the University of Granada. This group synthesized a library of symmetrical biscationic choline kinase inhibitors featuring different pyridinium, quinolinium, thienopyridine and thienopyrimidine heads substituted at position 4 with pyrrolidinium, piperidium, azepane, N-methyl-aniline, or p-Chloro-N-methyl aniline substituents and biphenyl, bipyridinyl, bibenzyl, biphenethyl, diphenoxiethane, dibenzylsulfide and 1,2-bis(phenylthio)ethane linkers. The rationale of the synthesis of this new library is to find the best replacement for the quinoline and pyridine heads as cationic group and, at the same time, to increase the affinity for the target protein.

I was in charge of the production of both human ChoK α and PfChoK proteins, the former to be delivered to our collaborator in Spain for carrying out enzymatic assays on a different library of compounds, the latter for the biochemical and the structural characterization of the above-mentioned library against PfChoK, which I both did in our laboratory at the IIT in Milano.

Materials and Methods

Cloning, Protein Expression and Purification of human Choline Kinase

A truncated form of hChoK α 1 (Δ 75-457) cloned into a pET-28a vector (Invitrogen, Carlsbad, CA, USA) and featuring an N-terminal 6x His tag, was purchased from Genescript (Piscataway, NJ, USA) and used to transform *Escherichia coli* BL21 (DE3) Star cells (Invitrogen, Carlsbad, CA, USA).

The transformed cells were cultured in Luria-Bertani (LB) medium at 37°C until OD₆₀₀ = 0.6 and then protein expression was induced with 1mM IPTG.

After induction, cell cultures were incubated overnight at 20°C and 180 rpm. The cellular pellet was then separated from exhausted medium by centrifugation at 10,000 rpm, resuspended in 50 mM Tris/HCl pH 7.5, 500 mM NaCl, 0.2 mM PMSF (Sigma Aldrich, St. Luis, MO, USA), DNase (Sigma Aldrich, St. Luis, MO, USA), 0.5 mM β -mercaptoethanol (Sigma Aldrich, St. Luis, MO, USA) and sonicated. Finally, the soluble fraction was separated from the cell debris by centrifugation at 15,000 rpm and 4°C.

A two-step purification protocol was implemented to isolate the target enzyme. The first step consisted in a Ni-NTA affinity chromatography run: the cell lysate was incubated for 45 minutes with Ni-NTA agarose beads (Qiagen, Venlo, The Netherlands). After incubation, the column was extensively washed with 40 column volumes (CV) of 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 10 mM imidazole and 1 CV of 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 40 mM imidazole. Finally, the His tagged enzyme was eluted from the column with 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 400 mM imidazole.

The second step consisted in a size-exclusion chromatography run using a HiPrep 26/60 Sephacryl 100 HR column (GE Healthcare, Little Chalfont, Buckinghamshire, UK) previously equilibrated with 20 mM Tris/HCl pH 7.5, 150 mM NaCl running buffer.

Cloning, Protein Expression and Purification of *Plasmodium falciparum* Choline Kinase

To produce PfChoK, a construct featuring an N-terminal truncation (Δ 79-440) and a 6xHis tag was purchased from Genescript (Piscataway, NJ, USA) and cloned into a pET-28a vector.

The plasmid was used to transform *Escherichia coli* BL21 (DE3) Star cells (Invitrogen, Carlsbad, CA, USA) for protein production. The transformed bacteria were cultured in Luria-Bertani (LB) medium at 37°C until OD₆₀₀ = 0.6 and the expression was induced with 1 mM IPTG.

Cell cultures were incubated overnight at 20°C and 180 rpm and then they were centrifuged at 10,000 rpm to pellet the cells and separate them from exhausted medium. Following centrifugation, the pellet was resuspended in 50 mM Tris/HCl pH 7.5, 500 mM NaCl, 0.2 mM PMSF (Sigma Aldrich, St. Luis, MO, USA), DNase (Sigma Aldrich, St. Luis, MO, USA), 0.5 mM β -mercaptoethanol (Sigma Aldrich, St. Luis, MO, USA) and sonicated. The soluble cell lysate fraction was then separated from the cell debris by centrifugation at 15,000 rpm and 4°C.

The target enzyme was purified by a two-step purification protocol. In the first affinity chromatography step, the cell lysate was incubated for 45 minutes in a Ni-NTA agarose bead column (Qiagen, Venlo, The Netherlands). The column was then extensively washed with 40 column volumes (CV) of 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 10 mM imidazole and 1 CV of 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 40 mM imidazole. Finally, the His-tagged enzyme was eluted from the column with 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 400 mM imidazole.

A subsequent size-exclusion chromatography run was done using a HiPrep 26/60 Sephacryl 100 HR column (GE Healthcare, Little Chalfont, Buckinghamshire, UK) previously equilibrated with 20 mM Tris/HCl pH 7.5, 150 mM NaCl buffer, resulting in a highly pure sample. The final protein purification yield was 2.5 mg of recombinant enzyme per liter of bacterial culture.

25 μ M Fractional Activity Screening

All compounds were dissolved at 100 mM concentration in pure DMSO. A spectrophotometric 96-well plate endpoint assay was adapted for PfChoK from a previously described IC₅₀ assay [305]. The reaction mix was prepared as follows: 100 mM Tris/HCl pH 7.5, 100 mM KCl, 10 mM MgCl₂, 0.5 mM phosphoenolpyruvate, 0.25 mM NADH, 4 U of pyruvate kinase, 5 U lactate dehydrogenase, 2 mM ATP, 4 mM choline chloride for the positive control. A 0.2 % final concentration of DMSO was also added to the reaction mixture to provide another positive control to take into account the possible effect of DMSO. Finally, each tested compound was added to the reaction mix at a concentration of 25 μ M to perform the initial screening.

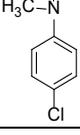
All the necessary components of the reaction mix for the assay were purchased from Sigma Aldrich (St. Luis, MO, USA) and the final assay volume of each experiment was 200 μ l.

The enzymatic reaction was initiated by adding the enzyme at a less than 10 nM concentration and the reaction rates were measured by monitoring the decrease in the absorbance at 340 nm for 20 minutes at room temperature using a Spark10M instrument (Tecan).

Fractional activity was calculated as a percentage obtained from the ratio between the positive DMSO control activity and the activity measured in the presence of the tested compound at a concentration of 25 μ M concentration.

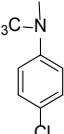
The list of the tested compounds is shown in Table 6 and Table 7.

Table 6: Scheme of the synthesized library

Compounds	Linker	Bioisosteric cationic head	4-substituent	
CKFa-M2	biphenyl	<i>thieno[3,2-b]pyridin-1-ium</i> 	Pyrrolidinyl 	
CKFa-M1		<i>thieno[3,2-d]pyrimidin-1-ium</i> 		
CKFa-M3		<i>thieno[2,3-d]pyrimidin-1-ium</i> 		
CKFg-9	biphenyl	<i>thieno[3,2-b]pyridin-1-ium</i>	Pyrrolidinyl 	
CKFa-21		<i>thieno[3,2-d]pyrimidin-1-ium</i>	Piperidinyl 	
CKFg-14		<i>thieno[2,3-d]pyrimidin-1-ium</i>		
CKFa-24		<i>thieno[3,2-d]pyrimidin-1-ium</i>	Azepanyl 	
CKFg-30		<i>thieno[2,3-d]pyrimidin-1-ium</i>		
CKFg-10		<i>thieno[3,2-b]pyridin-1-ium</i>	N-methyl-aniline 	
CKFa-22		<i>thieno[3,2-d]pyrimidin-1-ium</i>		
CKFg-18		<i>thieno[2,3-d]pyrimidin-1-ium</i>		
CKFP-1		biscationic	<i>thieno[3,2-d]pyrimidin-1-ium</i>	p-Chloro-N-methyl aniline 
CKFP-8			<i>thieno[2,3-d]pyrimidin-1-ium</i>	
CKFg-12	<i>thieno[3,2-b]pyridin-1-ium</i>		Pyrrolidinyl 	
CKFg-17	<i>thieno[3,2-d]pyrimidin-1-ium</i>			
CKFg-13	<i>thieno[2,3-d]pyrimidin-1-ium</i>			
CKFa-27	bipyridinyl	<i>thieno[3,2-d]pyrimidin-1-ium</i>	Piperidinyl 	
CKFg-32		<i>thieno[2,3-d]pyrimidin-1-ium</i>		
CKFa-26	biphenyl	<i>thieno[3,2-d]pyrimidin-1-ium</i>	Azepanyl 	
CKFg-20		<i>thieno[2,3-d]pyrimidin-1-ium</i>	Pyrrolidinyl 	
CKFg-11		<i>thieno[3,2-b]pyridin-1-ium</i>		
CKFg-16	<i>thieno[3,2-d]pyrimidin-1-ium</i>			
CKFg-15	biphenyl	<i>thieno[2,3-d]pyrimidin-1-ium</i>	Piperidinyl 	
CKFa-25		<i>thieno[3,2-d]pyrimidin-1-ium</i>		
CKFg-31		<i>thieno[2,3-d]pyrimidin-1-ium</i>		

CKFa-23		<i>thieno[3,2-d]pyrimidin-1-ium</i>	Azepanyl	
CKFg-19		<i>thieno[2,3-d]pyrimidin-1-ium</i>		
CKFf-1		<i>thieno[3,2-b]pyridin-1-ium</i>	Pyrrolidinyl	
CKFf-7		<i>thieno[3,2-d]pyrimidin-1-ium</i>		
CKFf-3	biphenethyl	<i>thieno[2,3-d]pyrimidin-1-ium</i>		
CKFa-33		<i>thieno[3,2-d]pyrimidin-1-ium</i>	Piperidinyl	
CKFf-6		<i>thieno[2,3-d]pyrimidin-1-ium</i>		
CKFa-29		<i>thieno[3,2-d]pyrimidin-1-ium</i>	Azepanyl	
CKFf-35		<i>thieno[2,3-d]pyrimidin-1-ium</i>		
CKFf-2		<i>thieno[3,2-b]pyridin-1-ium</i>	Pyrrolidinyl	
CKFf-8	diphenoxiethane	<i>thieno[3,2-d]pyrimidin-1-ium</i>		
CKFf-4		<i>thieno[2,3-d]pyrimidin-1-ium</i>		
CKFa-28		<i>thieno[3,2-d]pyrimidin-1-ium</i>	Piperidinyl	
CKFf-5		<i>thieno[2,3-d]pyrimidin-1-ium</i>		
CKFf-34		<i>thieno[3,2-d]pyrimidin-1-ium</i>	Azepanyl	
CKFf-36		<i>thieno[2,3-d]pyrimidin-1-ium</i>		

Table 7: Scheme of the synthesized diphenyl disulfide and 1,2-bisphenylthioethane linker compounds

Compounds	Linker	Bioisosteric cationic head	4-substituent	
CKEP-16		Pyridinium	Dimethylamino	
CKFP-2			Pyrrolidinyl	
CKFP-7	diphenyl disulfide	Quinolinium	<i>p</i> -Chloro- <i>N</i> -methyl aniline	
CKFP-4			Piperidinyl	
CKFP-3		<i>thieno[2,3-d]pyrimidin-1-ium</i>	Azepanyl	
CKFP-5		7-chloro-quinolinium	Pyrrolidinyl	
CKFP-16		<i>thieno[3,2-d]pyrimidin-1-ium</i>	Pyrrolidinyl	
PL46		1,2-	Pyridinium	Pyrrolidinyl
PL48	bis(phenylthio)ethane	7-chloro-quinolinium	Pyrrolidinyl	

Enzymatic Assay for the IC₅₀ Calculation

To assess PfChoK activity, a spectrophotometric 96-well plate assay was established based on a previously described assay [305]. The reaction mix was prepared as follows: 100 mM Tris/HCl pH 7.5, 100 mM KCl, 10 mM MgCl₂, 0.5 mM phosphoenolpyruvate, 0.25 mM NADH, 4 U of pyruvate kinase, 5 U lactate dehydrogenase, 2 mM ATP, 4 mM choline chloride and different concentrations of the selected compounds. The final assay volume was 200 µl for each experiment.

The enzymatic reaction was initiated by adding the enzyme at a less than 10 nM concentration and the reaction rates were measured by monitoring the decrease in the absorbance at 340 nm for 20 minutes at room temperature using a Spark10M instrument (Tecan).

Finally, the IC₅₀ values for each compound were determined by plotting the enzyme fractional activity (f.a.) against the logarithm of compound concentration (log[I]). Curve-fitting was performed with a dose-response curve from Origin (OriginLab Corporation, Northampton, MA, USA), using the following equation:

$$f.a. = \frac{1}{1 + 10^{\log(IC_{50}) - \log[I]}}$$

All data points were collected in triplicate and the final dose-response curves were plotted with MATLAB (MathWorks, Natick, MA, USA).

Enzymatic Assay for the Determination of the Inhibition Mode

The inhibition mechanism of the inhibitors with respect to the two substrates, choline and ATP, was determined using a method previously described by Lai et al. [307].

First, the IC₅₀ of each compound was determined at the K_M of choline using the previously described spectrophotometric assay. Then, using the same assay conditions, the concentration of one substrate was kept constant at saturating conditions (4 mM for choline and 2 mM for ATP) while varying the concentration of the other substrate and initial velocities were measured in the presence and in the absence of the inhibitor. By dividing the reaction velocity in the absence of the inhibitor with its corresponding velocity in the presence of the compound the percentage of inhibition (I%) was obtained. Inhibition percentages were plotted against the concentration of the varying substrate and fitted using a two-parameter-fit nonlinear regression algorithm manually implemented in Origin (OriginLab Corporation, Northampton, MA, USA) with the following equation:

$$I\% = 100 \cdot \left(\frac{\frac{[I]}{K_i} + \frac{[S][I]}{\alpha K_i K_s}}{1 + \frac{[S]}{K_i} + \frac{[I]}{K_i} + \frac{[S][I]}{\alpha K_i K_s}} \right)$$

where $[S]$ is the concentration of the varying substrate, $[I]$ is the inhibitor concentration, K_i is the dissociation constant that is assumed to be equal to the K_M and α is a constant that quantifies the effect that the inhibitor has on substrate binding and vice versa. K_i is the dissociation constant of the inhibitor.

Analytical Size-exclusion Chromatography

To determine the oligomerization state of PfChoK, an analytical size-exclusion chromatography run was done on a Superdex 200 Increase 10/300 GL column (GE Healthcare, Little Chalfont, Buckinghamshire, UK) using a 20 mM Tris HCl pH 7.5, 150 mM NaCl buffer. PfChoK was run through the column at three different concentrations (2, 1 and 0.5 mg/mL, respectively). Human Albumin (66 kDa) and eGFP (26 kDa) samples at 1 mg/mL, which are known to migrate as monomers and feature a single elution peak were used as markers of known molecular weight. For this analysis, the volume of the analyte was 500 μ L, the flow rate was set at 0.5 mL/min and the protein elution was followed by monitoring absorbance at 280 nm.

Crystallization

Crystals of the apo and of the ADP-bound form of the PfChoK enzyme were obtained by the hanging drop vapor diffusion method using VDXm 24-well plates with siliconized glass circle slides (Hampton Research, Aliso Viejo, California, USA). The screening for optimal crystallization conditions was done around the crystallization conditions reported in the PDB for the incomplete crystal structure of phosphocholine- and ADP-bound PfChoK (3FI8) (unpublished data). In each experiment, 1 μ L of a 10 mg/mL PfChoK sample solution was mixed with an equal volume of crystallization buffer and the droplet was left equilibrating against a 500 μ L reservoir at room temperature, resulting in crystals of suitable dimensions within one week. Crystals of the apoenzyme grew in a 16% (v/v) polyethylene glycol (PEG) 8000, 0.2 M NaCl, 0.1 M HEPES pH 7.5, 2 mM Tris(2-carboxyethyl)phosphine (TCEP), and 4 mM MgCl₂ crystallization buffer. Likewise, crystals of the ADP-bound form of the enzyme were obtained in 16% (v/v) polyethylene glycol (PEG) 4000, 0.2 M NaCl, 0.1 M HEPES pH 7.5, 2 mM Tris(2-carboxyethyl)phosphine (TCEP), 4 mM MgCl₂, and 2 mM ADP. Crystals of the apo and of the ADP-bound enzyme were then flash-frozen in 25% (v/v) glycerol and in 25% (v/v) ethylene glycol, respectively, for X-ray diffraction experiments.

X-Ray Diffraction Data Acquisition and Structure Determination

A 2.0 Å resolution dataset and a 2.2 Å dataset were collected for the apo and for the ADP-bound enzyme, respectively, at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) using a $\lambda = 1.000$ Å (beamline X06DA-PXIII). Diffraction images were processed using iMosflm [111] and scaled with SCALA in the CCP4 program suite (Oxford, UK) [112]. Molecular replacement and model refinement were done using Phenix (Berkeley, CA, USA) [113] and the incomplete crystal structure of phosphocholine- and ADP-bound PfChoK as the search and starting model (PDB code: 3FI8). During refinements, manual adjustments to the model were performed with Coot [114]. Data collection and refinement statistics for the two crystal structures are shown in Table 8. Molecular graphics and analyses were done using UCSF Chimera (University of San Francisco, San Francisco, CA, USA) [115] and Pymol (Schrödinger Inc., New York, NY, USA) [116].

Table 8: Data collection and refinement statistics for 6YXS and 6YXT

Crystallographic Table		
	<i>PfChoK apoenzyme</i>	<i>PfChoK ADP-bound form</i>
Protein Data Bank ID	6YXS	6YXT
Wavelength (Å)	1.00	1.00
Resolution range	57.53 - 2.00 (2.11 - 2.00)	68.00 - 2.20 (2.32 - 2.20)
Space group	$P 2_1 2_1 2_1$	$P 2_1 2_1 2_1$
Unit cell dimensions		
a, b, c (Å)	66.41, 68.82, 104.8	66.45, 68.00, 105.11
α, β, γ (°)	90, 90, 90	90, 90, 90
Total reflections	263600 (40520)	125859 (17449)
Unique reflections	33217 (4770)	24873 (3571)
Multiplicity	7.9 (8.5)	5.1 (4.9)
Completeness (%)	100.0 (100.0)	99.90 (100.0)
Mean I/σ(I)	8.00 (1.3)	6.5 (1.00)
Wilson B-factor	24.5	37.1
R_{merge}	0.203 (1.736)	0.146 (1.593)
$CC_{1/2}$	0.995 (0.429)	0.994 (0.486)
Reflections used in refinement	33149 (3242)	24706 (2339)
Reflections used for R_{free}	1739 (159)	1230 (114)
R_{work}	0.2063 (0.3247)	0.2305 (0.4671)
R_{free}	0.2385 (0.3812)	0.2553 (0.4983)
Total number of non-H atoms	3277	3165
macromolecules	3020	3028
ligands	53	41
solvent	204	96
Protein residues	359	359
RMS		
bonds (Å), angles (°)	0.003, 0.53	0.002, 0.66
Ramachandran plot		
favoured, allowed, outliers (%)	96.36, 2.80, 0.84	92.16, 5.60, 2.24
Rotamer outliers (%)	0.00	0.00
Clashscore	3.11	5.59
Average B-factor	38.25	57.15
macromolecules	37.84	57.32
ligands	49.42	50.23
solvent	41.36	54.66

Statistics for the highest-resolution shell are shown in parentheses.

Results

Human ChoK α and PfChoK Production and Purification

The productions of the human ChoK α and PfChoK enzymes followed the same protocol. Hence, for simplicity, here I will show only the SDS-PAGE gels and the chromatograms that were obtained for the PfChoK construct, as those of the ChoK α enzyme are very similar. Both constructs have been cloned as N-terminal truncated form ($\Delta 75-457$ and $\Delta 79-440$ for ChoK α and PfChoK, respectively). Overnight expression of both enzymes provided an average value of 15 grams of cellular pellets after centrifugation. After the extraction of the soluble fractions, the first protein purification step was done via Ni-NTA affinity chromatography and resulted in a relatively pure sample, as assessed by SDS-PAGE (Figure 56).

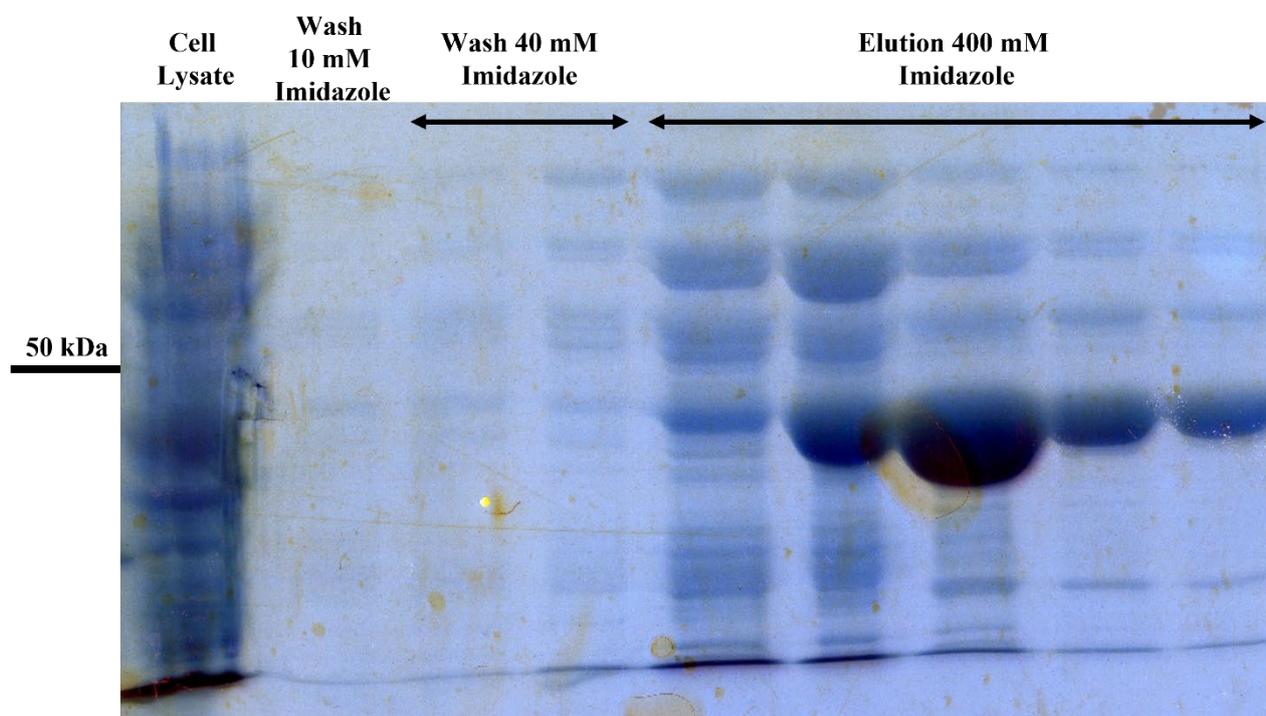


Figure 56: SDS-PAGE of the purification profile of PfChoK after Ni-NTA affinity chromatography
 A 50 kDa marker line is drawn on the left, while on top the content of each lane is labelled according to the different phase in the affinity chromatography process: lane 1 corresponds to the cell lysate, lane 2 to the end of the preliminary washing with 10 mM imidazole, lanes 3 and 4 to the end of the 40 mM imidazole wash, and the lanes from 5 to 9 to the first fractions of the 400 mM imidazole elution, where it is evident the presence of an enriched spot at ~45 kDa mass weight corresponding to PfChoK.

Figure 56 shows the profile of the first protein purification step: clearly, the two washing steps (lane 1 to 4) removed the majority of contaminants from the cell lysate and the elution resulted in fractions enriched in the band with a molecular weight of roughly 45 kDa, which could be readily be identified

as PfChoK. Following the Ni-NTA affinity chromatography, the isolated enzymes were further purified via size-exclusion chromatography.

The elution profiles for both enzymes were very similar and consisted of a two-peak chromatogram where the first peak corresponded to ChoK molecules that co-eluted with a higher molecular weight contaminant, while the second peak was predominant and corresponded to the substantially pure human ChoK α or PfChoK, as assessed by SDS-PAGE (Figure 57).

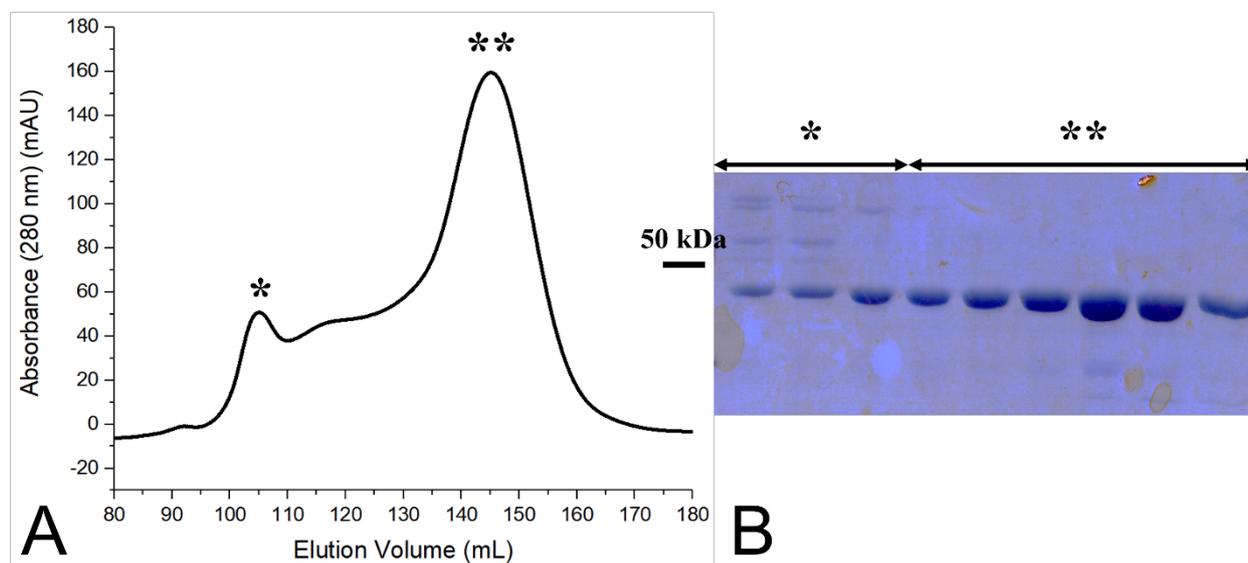


Figure 57: Size-exclusion chromatogram and corresponding SDS-PAGE gel
 In panel A, the two-peaks chromatogram of PfChoK is displayed. On the horizontal axis the elution volume is displayed in mL, while on the vertical axis the absorbance at 280 nm is shown in milli arbitrary units (mAU).
 In panel B, the corresponding SDS-PAGE gel showing the purification profile of the size-exclusion chromatography. Lanes from 1 to 3 are relative to the first elution peak (marked with one asterisk in both panels) and the remaining lanes from 4 to 9 correspond to fractions eluted under the second peak (marked with two asterisks in both panels). Mass weights of 50 kDa is depicted on the left side of the SDS-PAGE gel.

The purified fractions (indicated by the double asterisks mark in Figure 57) were pooled and concentrated at 10 mg/ml. The obtained human ChoK α was shipped to our collaborator at the University of Granada for further tests, while the PfChoK enzyme was further used in our lab at the IIT in Milano in crystallization and biochemical assays.

To summarize the purification profile, the SDS-PAGE gel in Figure 58 shows that the majority of recombinant protein remained in the soluble fraction after sonication and that the target enzymes were obtained at a suitable purity for both crystallization and biochemical experiments.

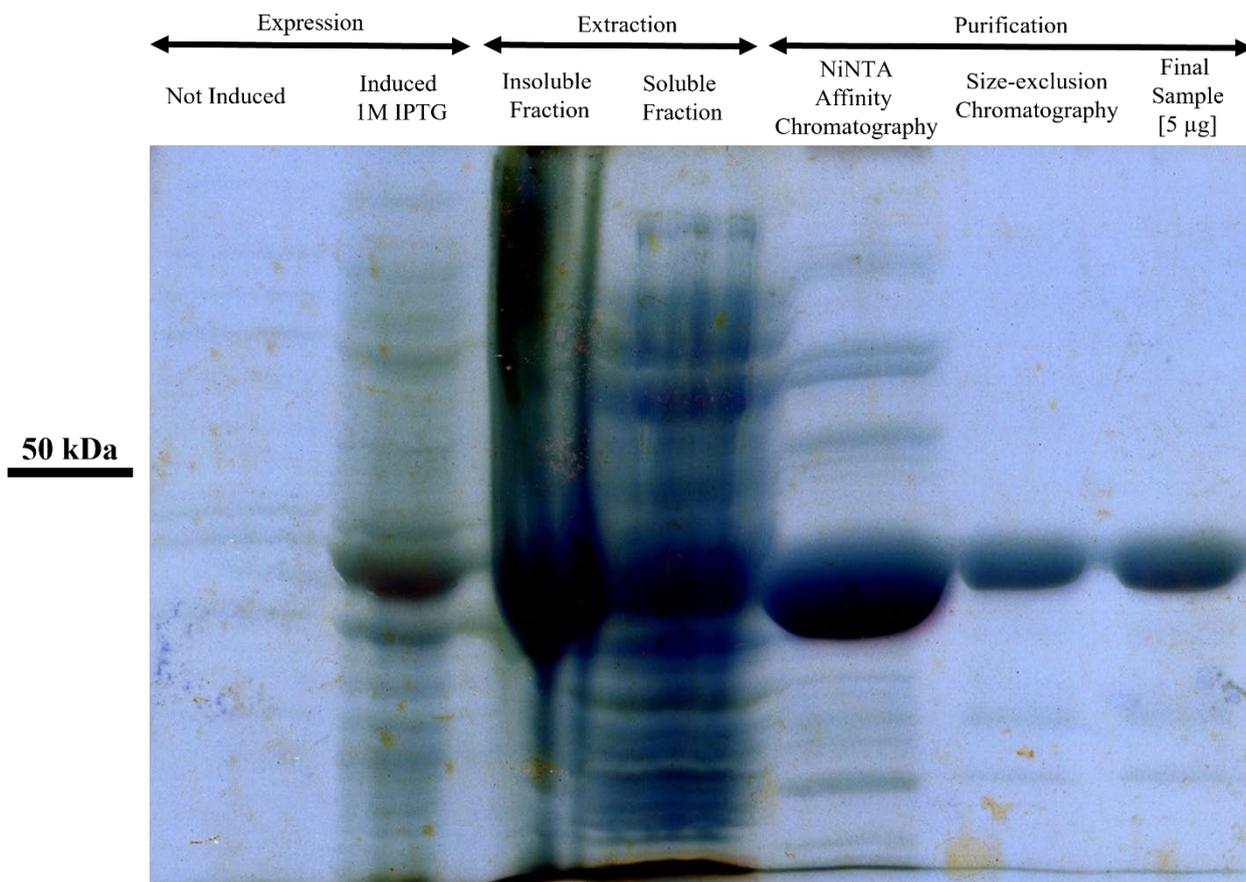


Figure 58: Summary SDS-PAGE gel showing the PfChoK purification profile. Lanes 1 and 2 refer to the expression phase where it is displayed the difference between the Not induced (lane 1) and the IPTG Induced (lane 2) fractions. Lanes 3 and 4 display the Insoluble and Soluble fractions after sonication, showing that despite some PfChoK remains insoluble and precipitates, the majority of the expressed recombinant enzyme remains in the soluble fraction. The three final lanes refer to the two step purification phase (lane 5, Ni-NTA Affinity Chromatography; lane 6, Size-exclusion Chromatography) and to the final sample.

The mass weight of 50 kDa is represented on the left side of the gel.

PfChoK Crystallization Experiments

Crystals of PfChoK and ADP-PfChoK grew in two similar conditions: 16% (v/v) PEG 8000, 0.2 M NaCl, 0.1 M HEPES pH 7.5, 2 mM TCEP, and 4 mM MgCl₂ for the apoenzyme and 16% PEG 4000, 0.2 M NaCl, 0.1 M HEPES pH 7.5, 2 mM TCEP, 4 mM MgCl₂, and 2 mM ADP for the ADP-bound structure.

The crystals of both structures displayed the same form and shape, and representative images of the grown crystals are depicted in Figure 59.

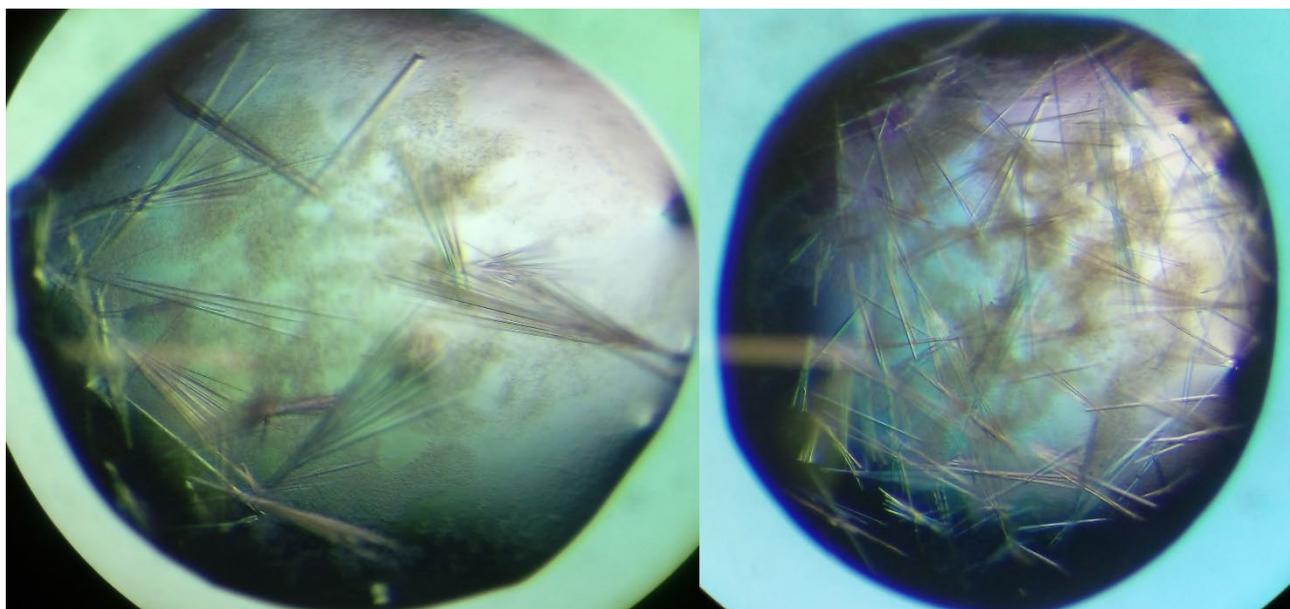


Figure 59: Crystal morphologies of PfChoK apoenzyme and ADP-bound form
The picture shows the crystal morphologies obtained during the crystallization experiments. Elongated pillar-like crystals grew as clusters originating from a common point, but each crystal could be easily singularly separated and cryo-protected.

All crystals usually grew forming clusters and had roughly the same elongated needle-like morphology. In each cluster, the single crystals could be readily separated from the others upon gentle touching and used in X-ray diffraction experiments. Crystals of the apoenzyme were cryo-protected with 25% (v/v) glycerol and crystals of the ADP-bound form with 25% (v/v) ethylene glycol prior to data collection.

Like other choline kinase eukaryotic enzymes, the structure of both the PfChoK apoenzyme and its ADP-bound form feature the typical globular and bilobate eukaryotic protein kinase fold (Figure 60), consisting of a 184-residue long N-terminal domain and a 245-residue long C-terminal domain that are connected by a short linker region.

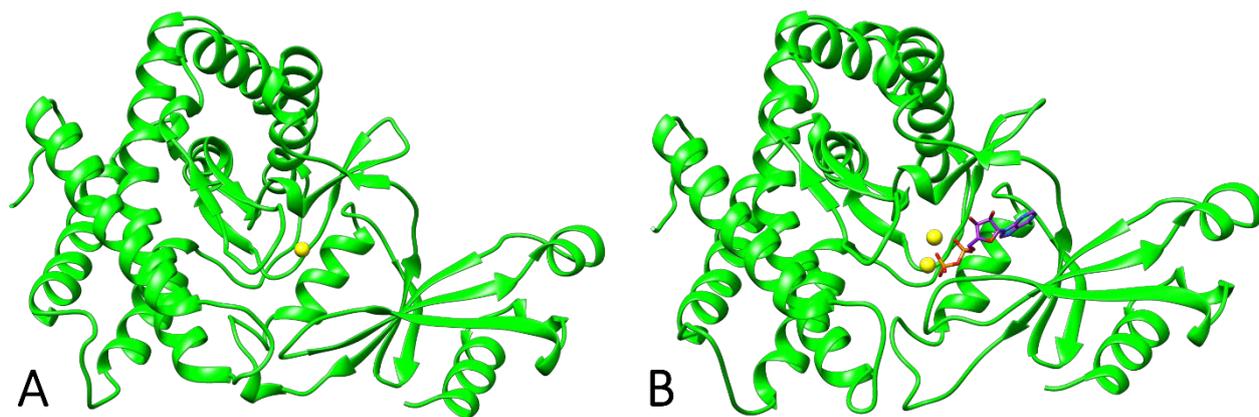


Figure 60: Overall fold of the PfChoK apoenzyme (PDBID: 6YXS) and the ADP-bound PfChoK enzyme (PDBID: 6YXT) Panel A refers to the apoenzyme, while panel B to the ADP-bound form. Proteins are shown as green cartoons, while magnesium ions are depicted as yellow spheres and the ADP molecule in panel B is shown in sticks with carbon atoms in purple, nitrogen atoms in blue, phosphorus atoms in orange and oxygen atoms in red.

PfChoK Oligomerization State

Since both the crystal structures feature only one protein in the asymmetric unit, an analytical size-exclusion chromatography was done to assess the monomeric state of PfChoK in solution. The results shows that the enzyme is essentially monomeric in solution (Figure 61), as derived from a comparison with eGFP and human Albumin that were used as known mass weight markers (27 and 66 kDa, respectively).

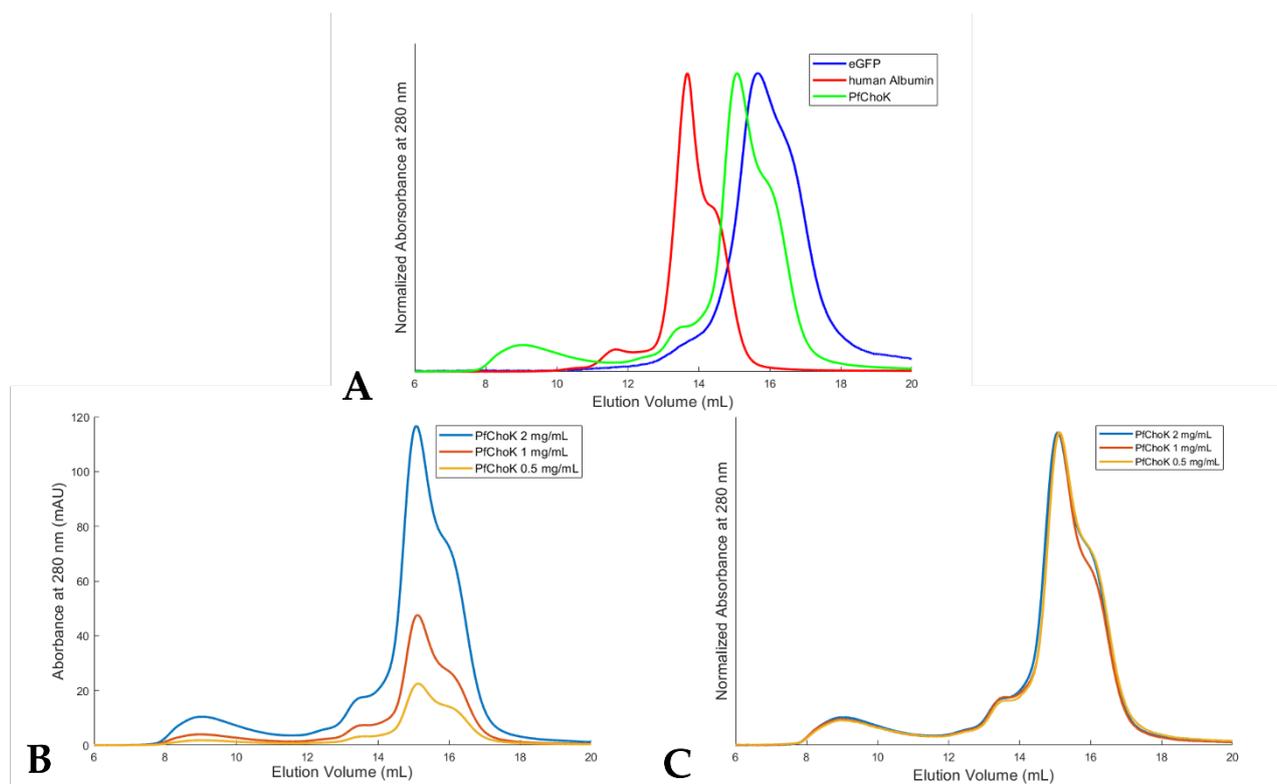


Figure 61: Results of the analytical size-exclusion chromatography

(A) Chromatograms of the normalized absorbance as a function of the elution volume of eGFP (blue line), human Albumin (red line) and PfChoK at the concentration of 2 mg/mL (green line) showing that the main PfChoK peak falls between the peaks of the two markers and therefore can be related to the molecular mass of the monomeric state (45 kDa). (B) Chromatograms of the PfChoK absorbance as a function of the elution volume at the three chosen concentrations (2 mg/mL as a blue line, 1 mg/mL as a red line and 0.5 mg/mL as an orange line) shows that the chromatogram morphology is the same regardless of the enzyme concentration. (C) Chromatograms of the normalized absorbance as a function of the elution volume at the three chosen concentrations (2 mg/mL as a blue line, 1 mg/mL as a red line and 0.5 mg/mL as an orange line) show that the relative height of the peaks remains the same at all the concentrations.

The shoulders present in the elution profiles are experimental artifacts due to the flow rate of the experiment or to different migration conformation of the same protein. They cannot be attributed to contaminants since after each chromatographic run, the elution profile was checked with an SDS-PAGE (data not shown).

25 μM Fractional Activity Screening

The results of the PfChoK inhibition screening performed at a concentration of 25 μM are shown in Figure 62 in a graph featuring the percentual fractional activity of each tested compound.

The results of this screening, together with the work performed by Dr. Lopez Cara's group on *P. falciparum* infected erythrocytes *in vitro* (which is briefly reported here in Table 9) allowed for the identification of the most potent molecules inside the library.

It has to be noted that the compounds featuring the diphenyl disulfide or the 1,2-bis(phenylthiol)ethane linker (respectively identified by the acronyms CKFP and PL in Figure 62) had better performances *ex vivo* compared to the other compounds in the library since in most cases they resulted in a more than 50% reduction in PfChoK residual activity. However, since this project is still ongoing, the *in vitro* potency on infected erythrocytes has not yet been fully characterized yet and this matter is still under investigation.

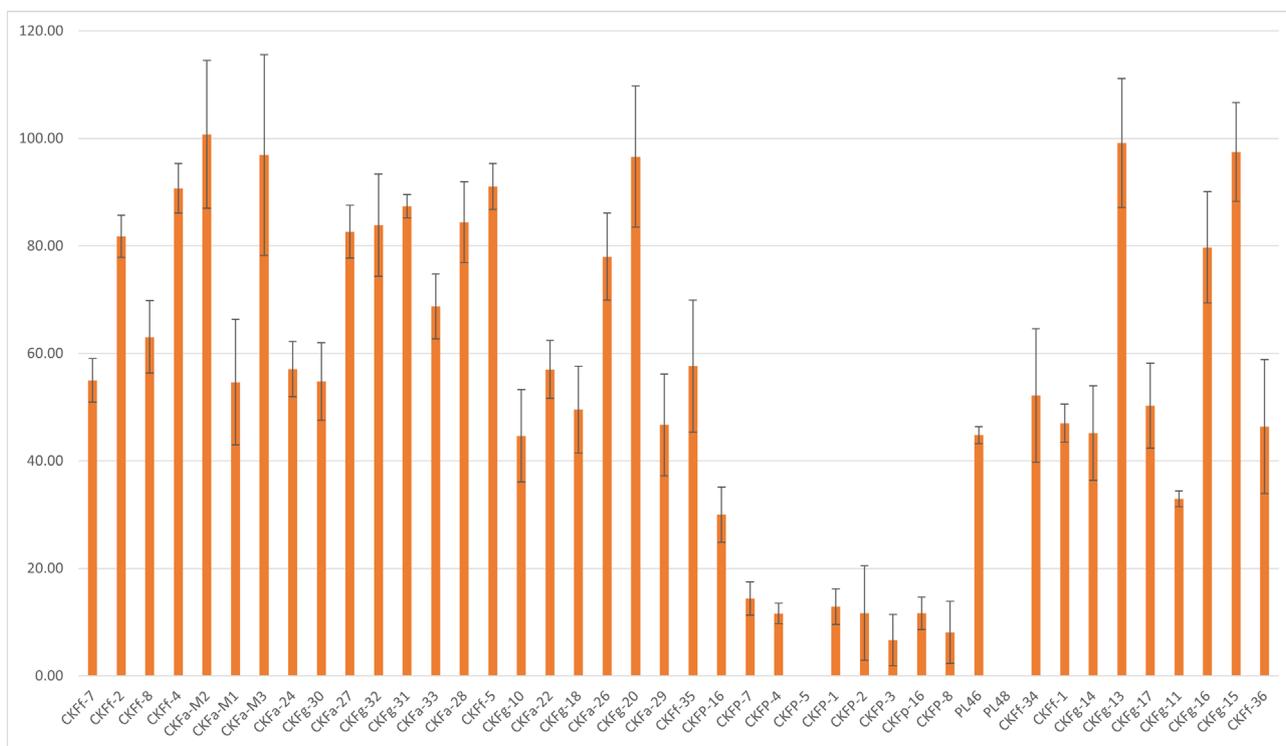


Figure 62: Percentual fractional activity at 25 μM compound concentration

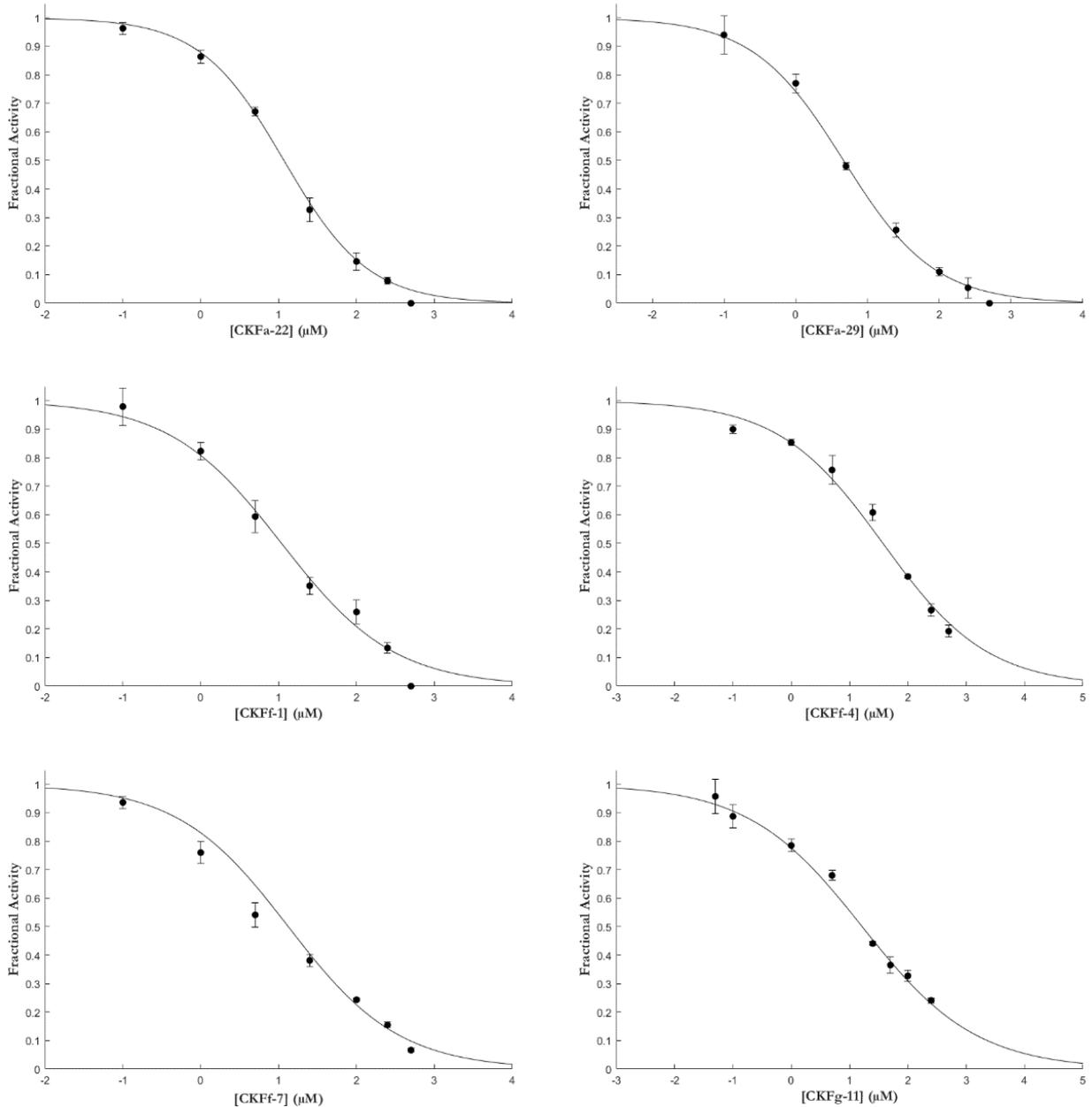
The graph shows on the horizontal axis the Percentual fractional activity of PfChoK in the presence of 25 μM of the tested compound and on the horizontal axis the name of the considered compound.

Table 9: Inhibitory effect of the compounds from the library, excluding the one featuring the sulfur-containing linker. N.d., not determined. The CK prefix is omitted for brevity

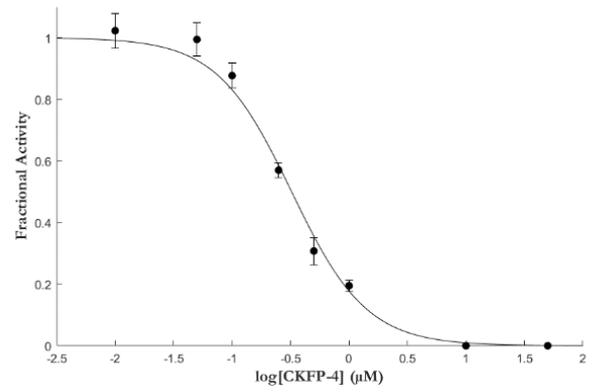
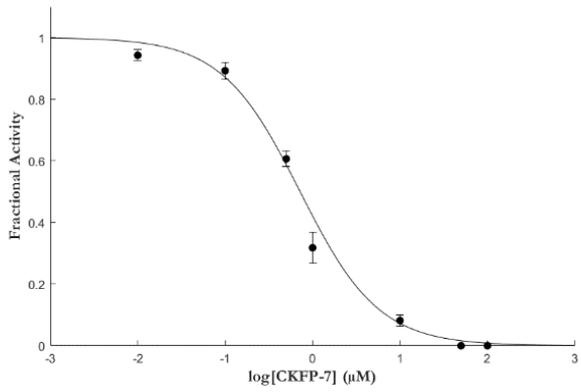
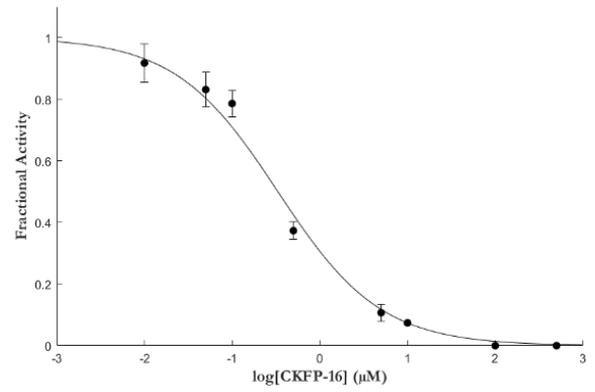
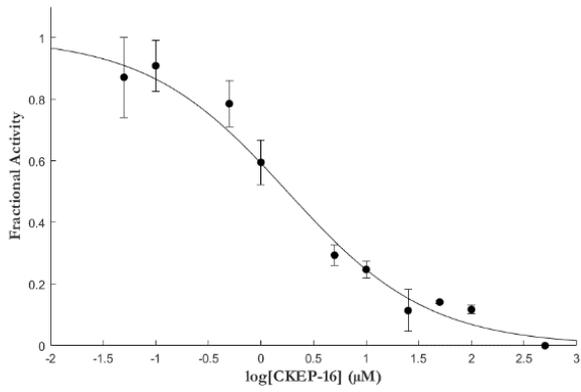
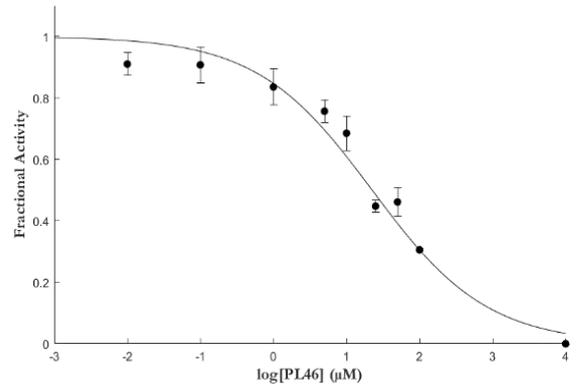
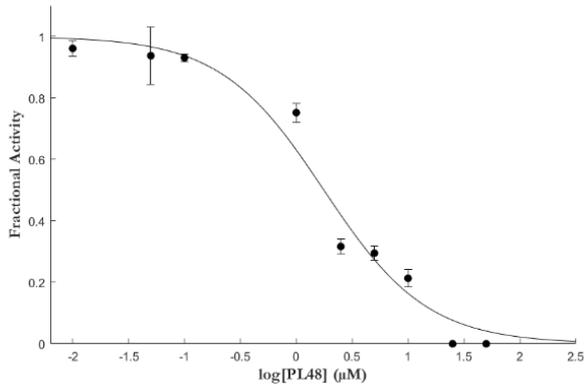
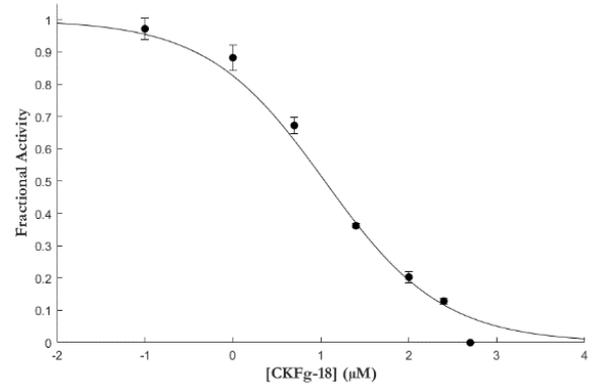
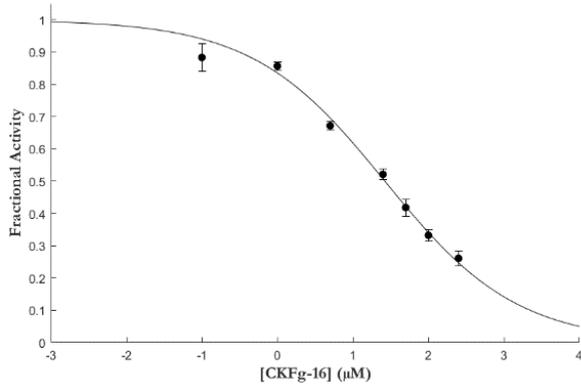
Compounds	Linker	Bioisoterics cationic head	4-substituent	In vitro GI ₅₀	PfCK Fractional Activity at 25 μM		
Fa-M2	monocationic	thieno[3,2-b]pyridin-1-ium	pyrrolidinyl	0,0489	100,75 ± 13,73		
Fa-M1		thieno[3,2-d]pyrimidin-1-ium		0,0410	54,65 ± 11,69		
Fa-M3		thieno[2,3-d]pyrimidin-1-ium		0,0351	96,91 ± 18,68		
Fg-9	biphenyl	thieno[3,2-b]pyridin-1-ium	pyrrolidinyl	0,1659	nd		
Fa-21		thieno[3,2-d]pyrimidin-1-ium		0,0137	nd		
Fg-14		thieno[2,3-d]pyrimidin-1-ium		0,0115	nd		
Fa-24		thieno[3,2-d]pyrimidin-1-ium		0,0578	57,08 ± 5,15		
Fg-30		thieno[2,3-d]pyrimidin-1-ium		---	54,77 ± 7,20		
Fg-10		thieno[3,2-b]pyridin-1-ium		piperidinyl	0,0530	44,66 ± 8,55	
Fa-22		thieno[3,2-d]pyrimidin-1-ium			0,0514	57,03 ± 5,37	
Fg-18		thieno[2,3-d]pyrimidin-1-ium			0,0604	49,53 ± 8,06	
FP-1		biphenyl		thieno[3,2-d]pyrimidin-1-ium	N-methyl-aniline	0,1422	12,88 ± 3,32
FP-8				thieno[2,3-d]pyrimidin-1-ium	p-Chloro-N-methylaniline	0,2816	8,10 ± 5,80
Fg-12	bipyridinyl	thieno[3,2-b]pyridin-1-ium	pyrrolidinyl	0,3049	nd		
Fg-17		thieno[3,2-d]pyrimidin-1-ium		0,2273	50,28 ± 7,88		
Fg-13		thieno[2,3-d]pyrimidin-1-ium		0,1429	99,13 ± 12,01		
Fa-27		thieno[3,2-d]pyrimidin-1-ium		piperidinyl	0,1205	82,64 ± 4,90	
Fg-32		thieno[2,3-d]pyrimidin-1-ium			>0,25	83,85 ± 9,53	
Fa-26		thieno[3,2-d]pyrimidin-1-ium		azepanyl	0,2093	78,02 ± 8,13	
Fg-20		thieno[2,3-d]pyrimidin-1-ium			0,0976	96,62 ± 13,12	
Fg-11		biscationic		thieno[3,2-b]pyridin-1-ium	pyrrolidinyl	0,2321	32,96 ± 1,48
Fg-16				thieno[3,2-d]pyrimidin-1-ium		0,0206	79,74 ± 10,35
Fg-15				thieno[2,3-d]pyrimidin-1-ium		0,0127	97,48 ± 9,22
Fa-25	thieno[3,2-d]pyrimidin-1-ium		piperidinyl	0,0347		nd	
Fg-31	thieno[2,3-d]pyrimidin-1-ium			0,0347		87,41 ± 2,20	
Fa-23	thieno[3,2-d]pyrimidin-1-ium		azepanyl	0,0678		nd	
Fg-19	thieno[2,3-d]pyrimidin-1-ium	0,0715		nd			
Ff-1	biphenethyl	thieno[3,2-b]pyridin-1-ium	pyrrolidinyl	0,0901	47,03 ± 3,53		
Ff-7		thieno[3,2-d]pyrimidin-1-ium		0,0219	55,02 ± 4,07		
Ff-3		thieno[2,3-d]pyrimidin-1-ium		0,0388	nd		
Fa-33		thieno[3,2-d]pyrimidin-1-ium		piperidinyl	0,0439	68,76 ± 6,03	
Ff-6		thieno[2,3-d]pyrimidin-1-ium			0,0279	nd	
Fa-29		thieno[3,2-d]pyrimidin-1-ium		azepanyl	0,0556	46,72 ± 9,49	
Ff-35	thieno[2,3-d]pyrimidin-1-ium	0,0484	57,62 ± 12,29				
Ff-2	diphenoxietthane	thieno[3,2-b]pyridin-1-ium	pyrrolidinyl	0,0740	81,77 ± 3,93		
Ff-8		thieno[3,2-d]pyrimidin-1-ium		0,0352	63,06 ± 6,74		
Ff-4		thieno[2,3-d]pyrimidin-1-ium		0,0121	90,72 ± 4,57		
Fa-28		thieno[3,2-d]pyrimidin-1-ium		piperidinyl	0,0471	84,39 ± 7,53	
Ff-5		thieno[2,3-d]pyrimidin-1-ium			0,0224	91,06 ± 4,27	
Ff-34		thieno[3,2-d]pyrimidin-1-ium		azepanyl	0,0879	52,17 ± 12,43	
Ff-36		thieno[2,3-d]pyrimidin-1-ium			0,1861	46,39 ± 12,44	

IC₅₀ Determination

The most active compounds have been further characterized and their IC₅₀ has been calculated. The IC₅₀ curves are shown in Figure 63 and the results are summarized in Table 10. All the tested compounds displayed a low micromolar potency. Data were taken in triplicates to allow statistical analysis.



Continuing on the following page



Continuing on the following page

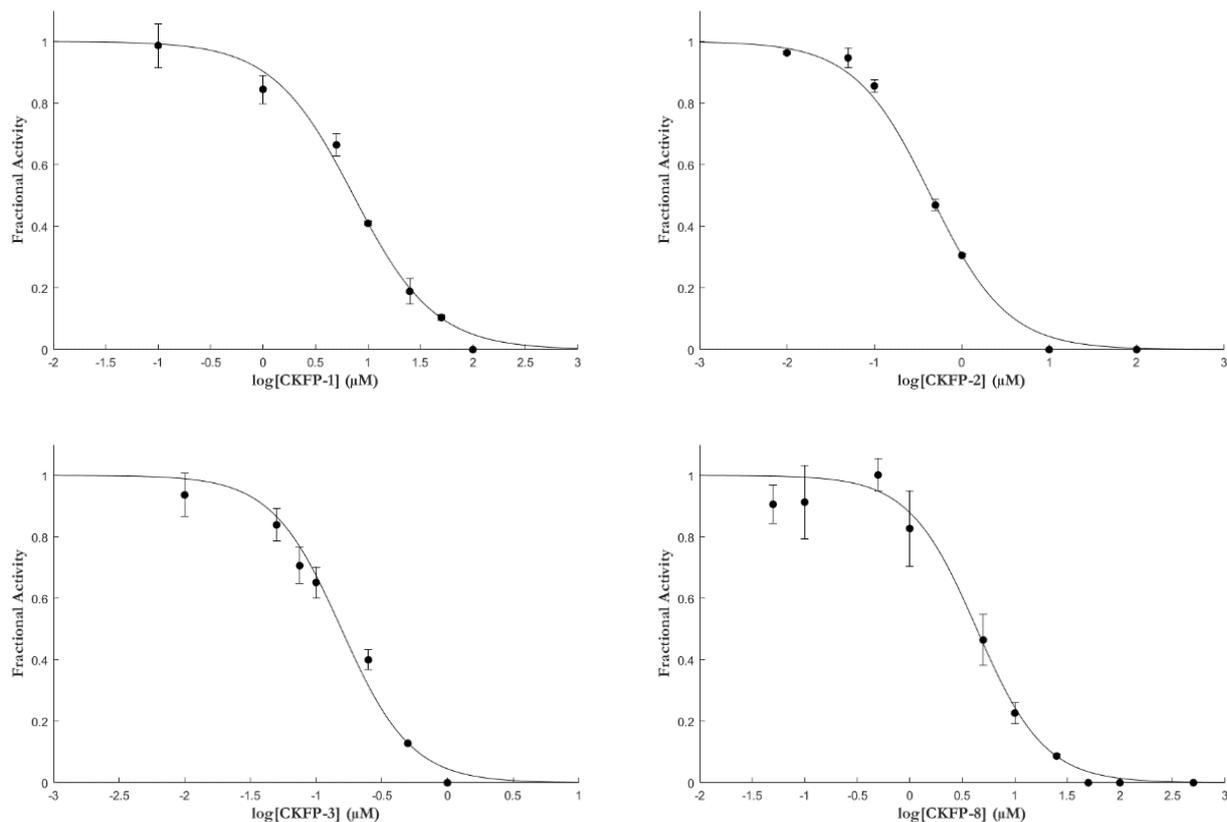


Figure 63: IC_{50} curves of all the tested compounds

The normalized fractional activity is expressed as number comprised from 0 to 1 on the vertical axis, while the logarithm of the concentration of the compound is on the horizontal axis.

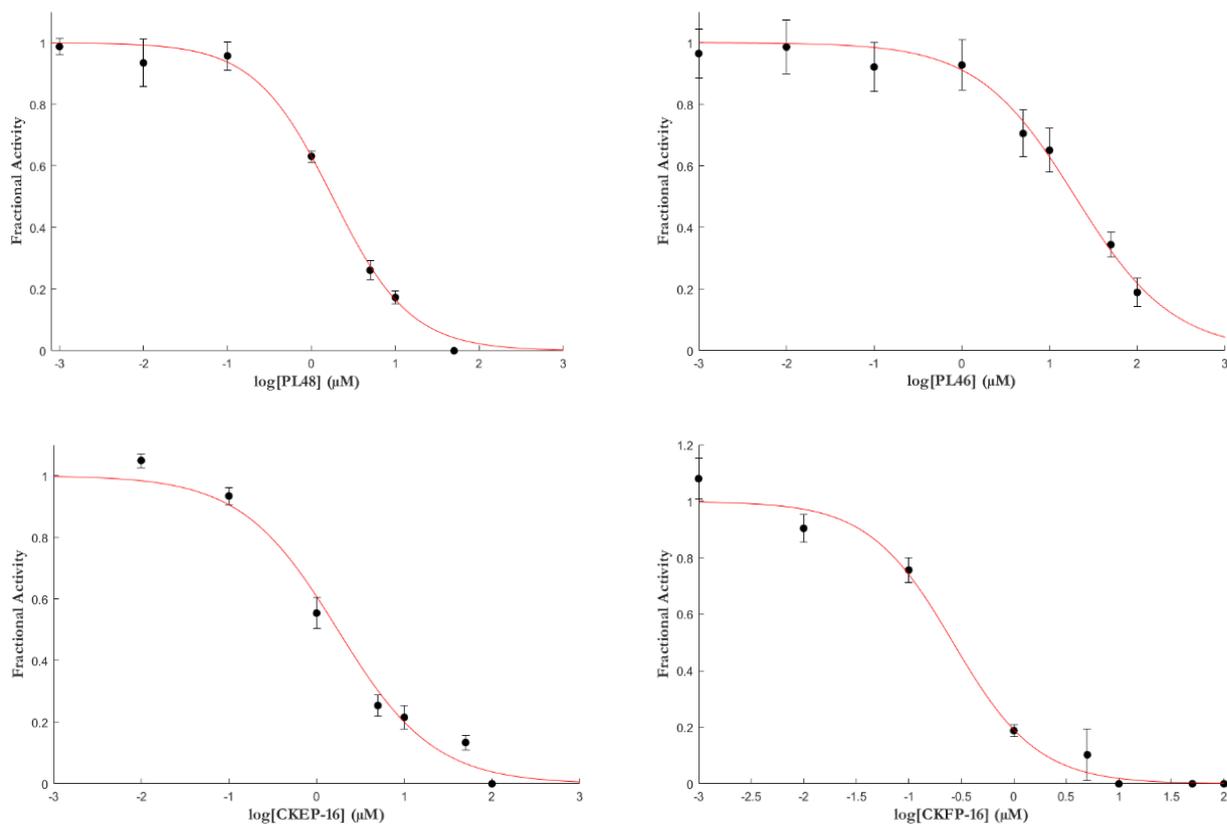
Table 10: Summary of the IC_{50} values calculated from the curves displayed in Figure 47

Compound	IC_{50} (μM)
CKFa-22	11.70 ± 0.49
CKFa-29	4.64 ± 0.20
CKKff-1	10.90 ± 1.33
CKFf-4	37.36 ± 3.24
CKFf-7	13.32 ± 3.82
CKFg-11	16.46 ± 1.43
CKFg-16	26.53 ± 2.46
CKFg-18	11.09 ± 1.00
PL48	1.77 ± 0.32
PL46	22.42 ± 3.30
CKEP-16	1.78 ± 0.39
CKFP-16	0.33 ± 0.05
CKFP-7	0.72 ± 0.12
CKFP-4	0.32 ± 0.02
CKFP-5	0.42 ± 0.01
CKFP-1	7.27 ± 0.23
CKFP-2	0.44 ± 0.03
CKFP-3	0.16 ± 0.01
CKFP-8	4.35 ± 0.63

Determination of the Inhibition Mode

Since the majority of compounds featuring the sulfur-containing linker shows IC_{50} values in the very low micromolar range with most values being less than $1 \mu\text{M}$ (~ 10 times lower than all the IC_{50} values that were calculated for the other compounds that displayed a high *in vitro* antiplasmodial activity). Hence, I set out to evaluate the inhibition modality of this subgroup of compounds in the library. The results of these assays are reported in Figure 64 and Figure 65 and are summarized by Table 11.

As it can be observed, all the IC_{50} values calculated at a concentration of choline equals to its K_M are approximately identical to the previously calculated IC_{50} and a similar trend is observed with the K_i . All values of the α parameter, which represents the index that provides a quantitative measure of the interaction between the inhibitor and the substrate binding site that are calculated for both substrates, are really close to the unitary value. According to the method described by Lai et al. [307], this indicates that the inhibition modality of this class of compounds is the noncompetitive one. In fact, $\alpha = 1$ indicates that there is no interaction between the inhibitor and the substrate binding site (the range of values for synergistic inhibition is $\alpha > 1$, while for antagonistic inhibition is $\alpha < 1$).



Continuing on the following page

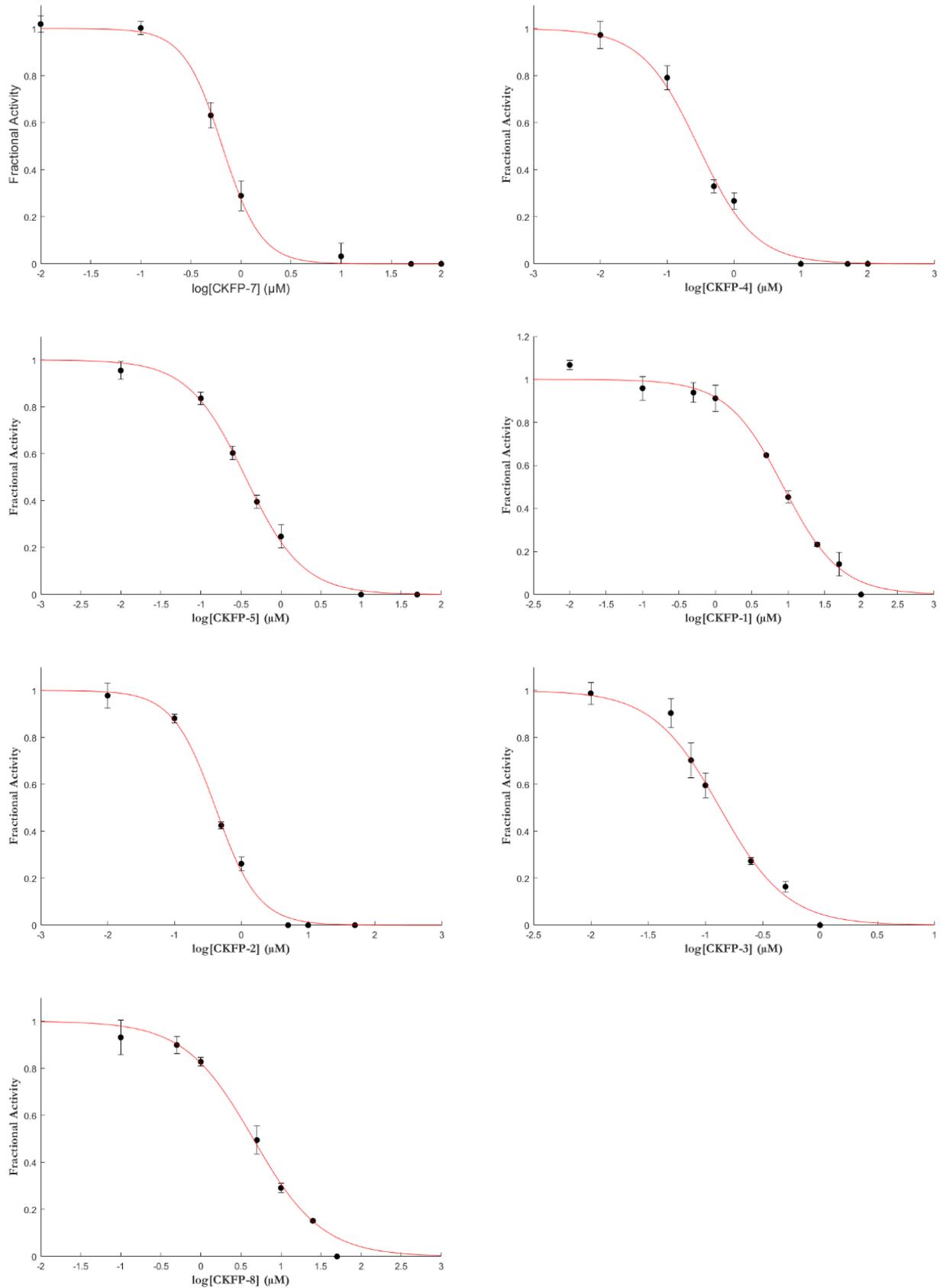
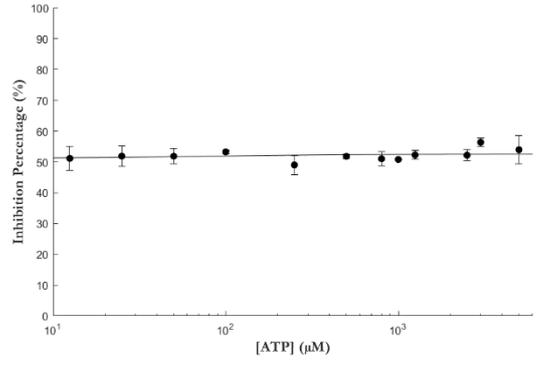
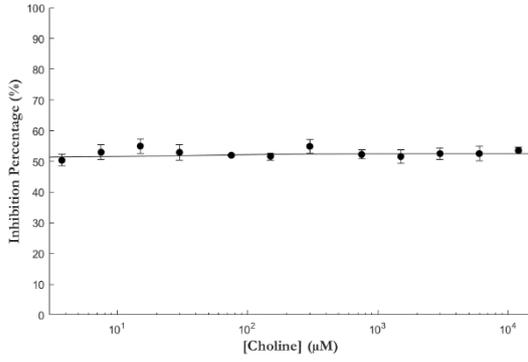
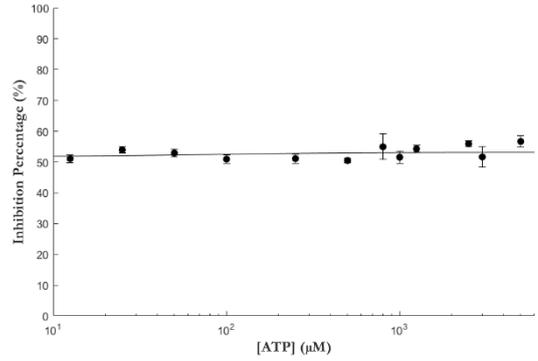
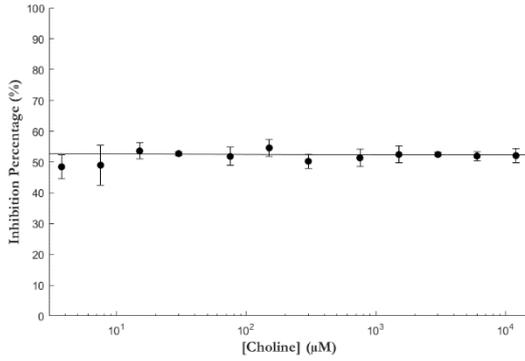


Figure 64: IC_{50} curves of all the tested compounds calculated at a choline concentration equals to its Michaelis-Menten constant. The normalized fractional activity is expressed as number comprised from 0 to 1 on the vertical axis, while the logarithm of the concentration of the compound is on the horizontal axis.

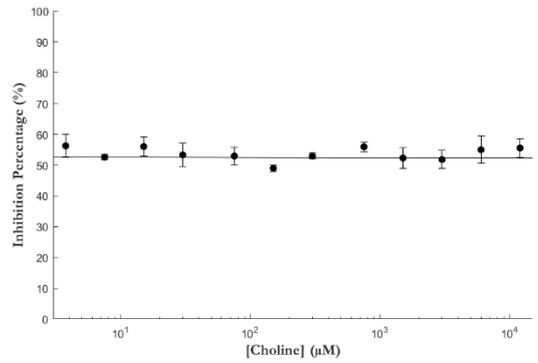
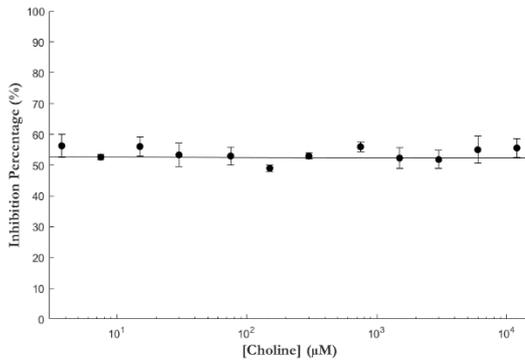
PL48



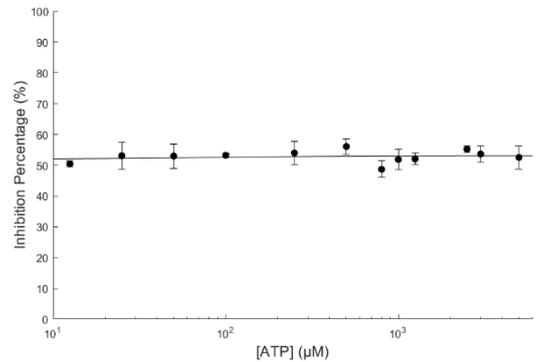
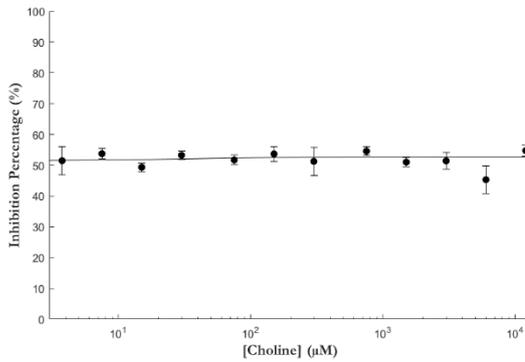
PL46



CKEP-16

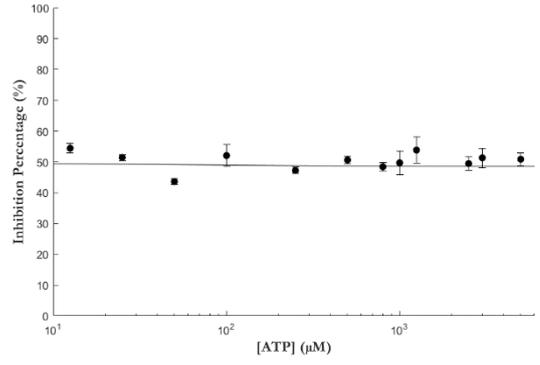
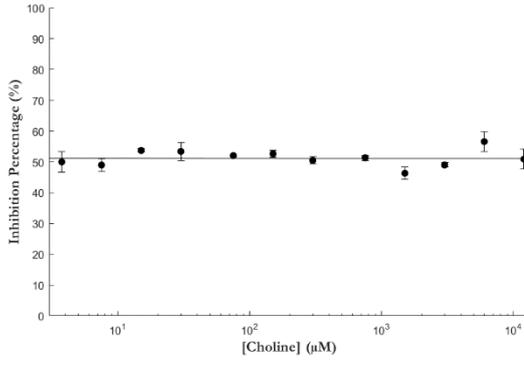


CKFP-16

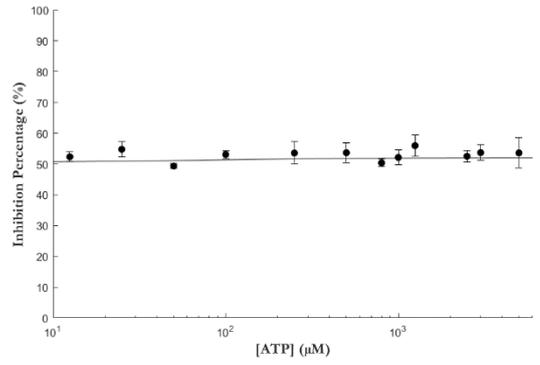
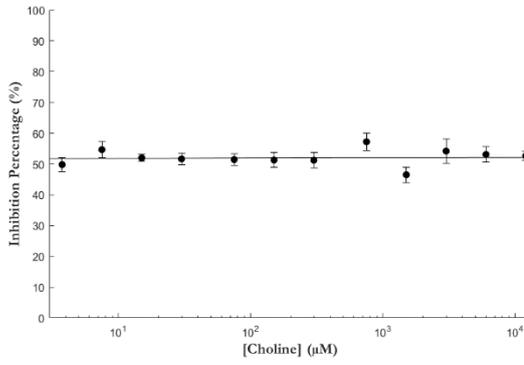


Continuing on the following page

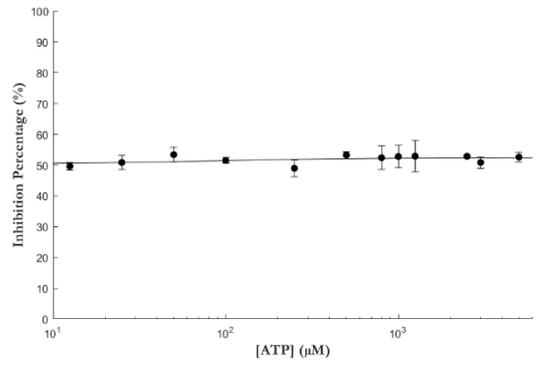
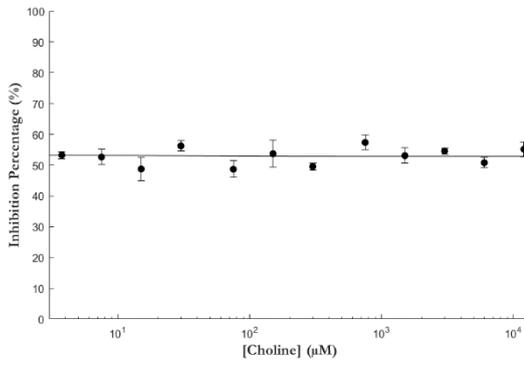
CKFP-7



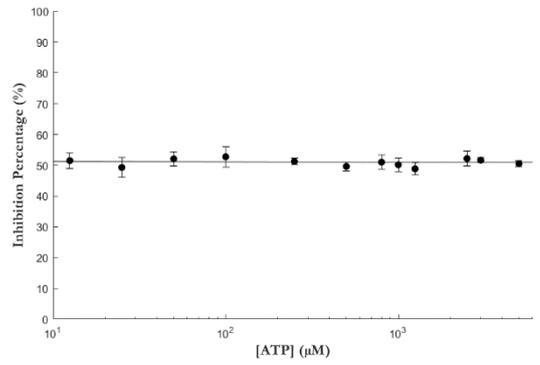
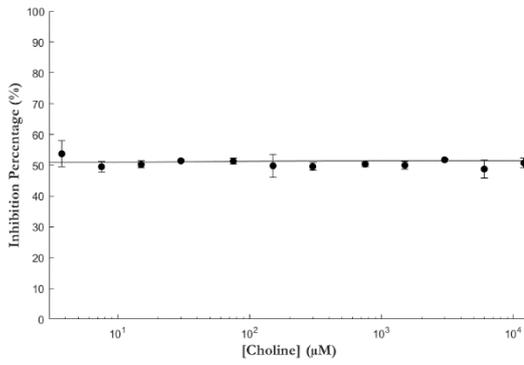
CKFP-4



CKFP-5

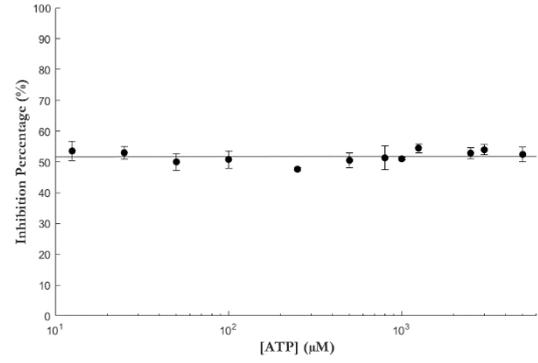
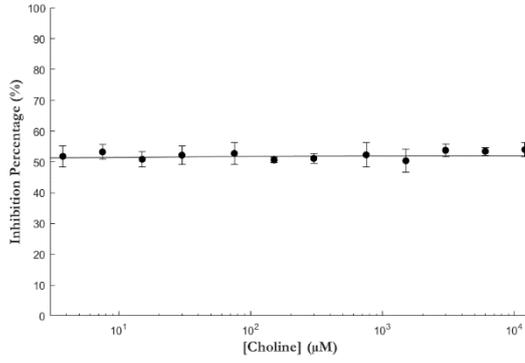


CKFP-1

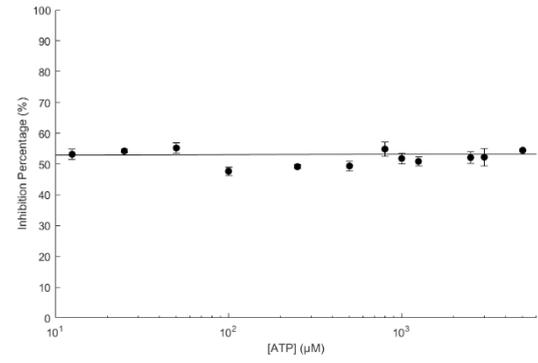
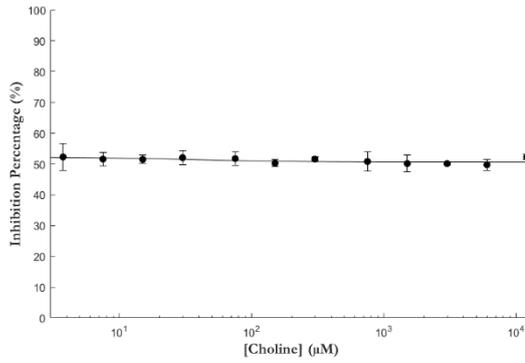


Continuing on the following page

CKFP-2



CKFP-3



CKFP-8

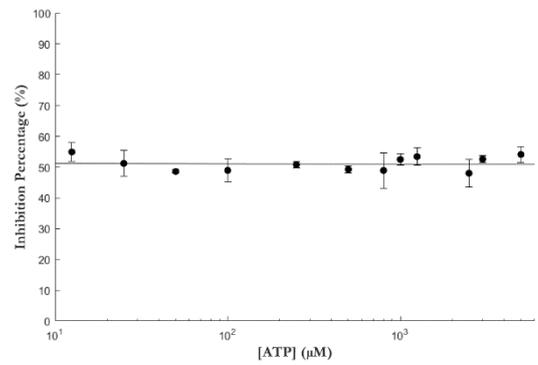
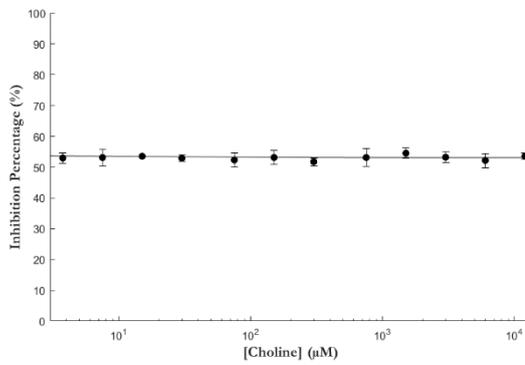


Figure 65: Fractional activity calculated with a constant compound and one substrate concentrations, while varying the concentration of the other
 On the left side of the picture, the compound identifier is shown the logarithms of the varying substrate concentrations used in the experiments are plotted on the horizontal axis. The column of graphs on the left refers to the curves obtained while varying choline, while the other on the right refers to the experiments when ATP was varied.

Table 11: Summary of the K_i and α parameters calculated from the inhibition curves in Figure 64 and Figure 65

Compound	IC ₅₀ (μ M)	IC ₅₀ @ K _M ^{Choline} (μ M)	K _i ^{Choline} (μ M)	α ^{Choline}	K _i ^{ATP} (μ M)	α ^{ATP}
PL48	1.77 \pm 0.32	1.78 \pm 0.07	1.62 \pm 0.07	0.95 \pm 0.05	1.62 \pm 0.12	0.94 \pm 0.09
PL46	22.42 \pm 3.30	19.66 \pm 1.98	17.86 \pm 0.49	1.02 \pm 0.04	18.63 \pm 1.24	0.94 \pm 0.08
CKEP-16	1.78 \pm 0.39	1.74 \pm 0.54	1.52 \pm 0.10	1.02 \pm 0.09	1.71 \pm 0.11	0.91 \pm 0.07
CKFP-16	0.33 \pm 0.05	0.26 \pm 0.04	0.24 \pm 0.02	0.95 \pm 0.08	0.23 \pm 0.01	0.96 \pm 0.08
CKFP-7	0.72 \pm 0.12	0.65 \pm 0.03	0.62 \pm 0.05	1.00 \pm 0.10	0.66 \pm 0.06	1.04 \pm 0.14
CKFP-4	0.32 \pm 0.02	0.29 \pm 0.05	0.26 \pm 0.02	0.98 \pm 0.08	0.27 \pm 0.01	0.94 \pm 0.07
CKFP-5	0.42 \pm 0.01	0.36 \pm 0.01	0.32 \pm 0.02	1.02 \pm 0.09	0.35 \pm 0.01	0.92 \pm 0.05
CKFP-1	7.27 \pm 0.23	8.61 \pm 0.27	8.33 \pm 0.25	0.97 \pm 0.03	8.16 \pm 0.39	1.02 \pm 0.06
CKFP-2	0.44 \pm 0.03	0.41 \pm 0.02	0.39 \pm 0.02	0.97 \pm 0.07	0.38 \pm 0.02	0.99 \pm 0.08
CKFP-3	0.16 \pm 0.01	0.13 \pm 0.01	0.13 \pm 0.01	1.08 \pm 0.06	0.13 \pm 0.01	0.99 \pm 0.01
CKFP-8	4.35 \pm 0.63	4.51 \pm 0.16	3.87 \pm 0.03	1.03 \pm 0.02	4.27 \pm 0.31	1.02 \pm 0.10

Discussion

Structure of the Apo and ADP-bound form of *P. falciparum* Choline Kinase

We produced a 361 amino acid-long truncated form of the enzyme lacking the first 79 N-terminal residues. As shown in previous structural studies on the nematode *Caenorhabditis Elegans* (PDB code: 1NW1) [145] and human (PDB code: 2CKO) [148] variants (ChoKA-2 and human ChoKa1, respectively), this N-terminal domain features a non-conserved region (Figure 66) that plays no role in catalysis.

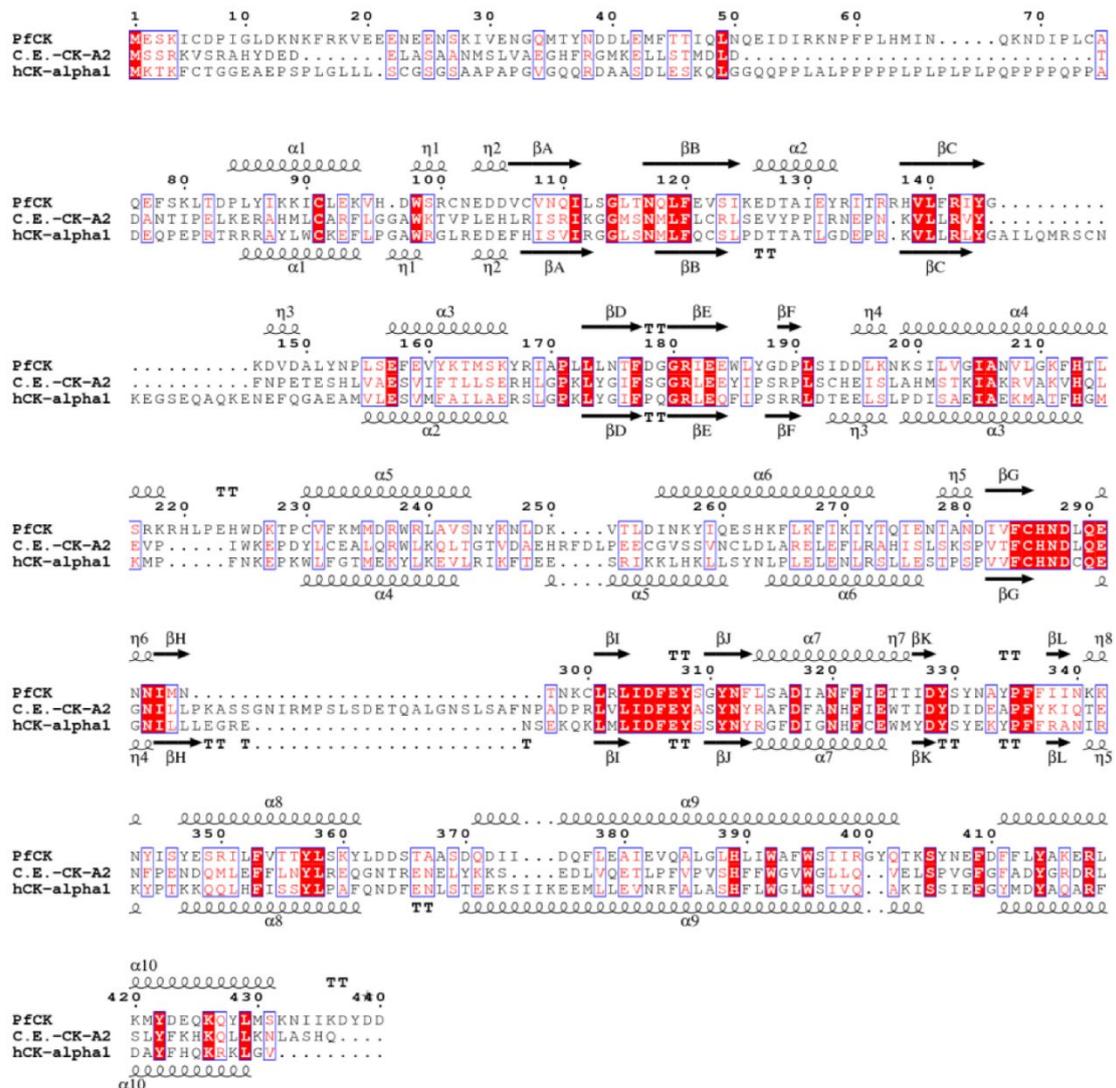


Figure 66: Sequence alignment of the full sequences of *P. falciparum* choline kinase (*PfChoK*), *C. Elegans* *ChoK-A2* and human *ChoKa1*

In the alignment, the secondary structure of the *PfChoK* structure is shown on the top, while the secondary structure of human *ChoKa1* is shown at the bottom. Helices (α) are indicated by numbers, while strands (β) by alphabetical letters. *PfChoK* exhibits a sequence identity of 23.0% with *ChoK-A2* and 25.3% with human *ChoKa1*, respectively. Figure obtained with ESPript-<http://espript.ibcp.fr>.

Like other choline kinase eukaryotic enzymes, the structure of both the PfChoK apoenzyme and its ADP-bound form feature the typical globular and bilobate eukaryotic protein kinase fold, consisting of a 184-residue long N-terminal domain and a 245-residue long C-terminal domain that are connected by a short linker region.

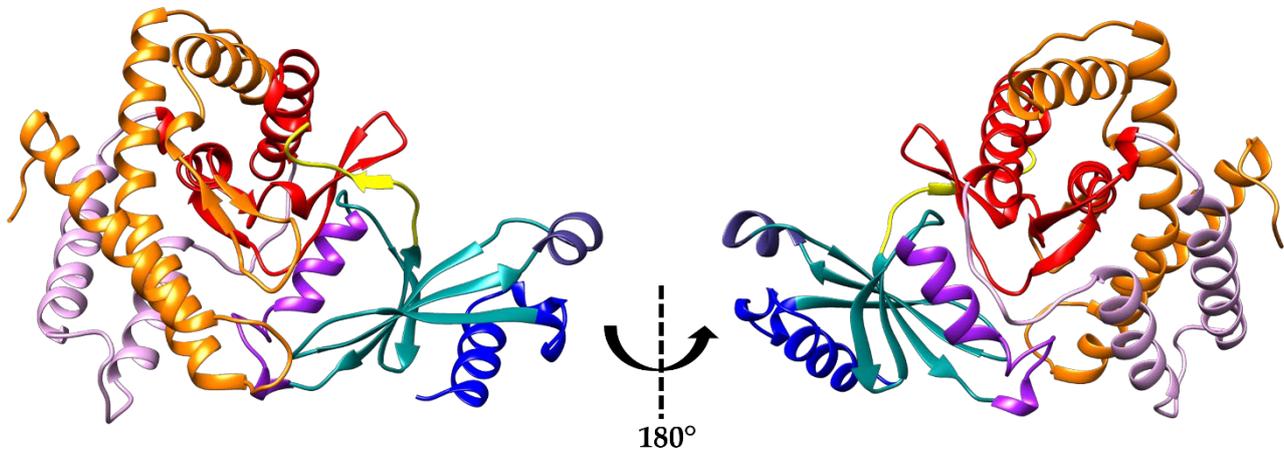


Figure 67: Two 180-degree rotated views of the PfChoK apoenzyme structure (PDB code: 6YXS). The atypical protein kinase fold regions are depicted in color code: helices $\alpha 1$, $\eta 1$, and $\eta 2$ are colored in blue; strands βA , βB , βC , and βD in light sea green; helix $\alpha 2$ in purple, helices $\eta 3$, and $\alpha 3$ in magenta; β strand F in yellow; the central core subdomain in red; the insertion subdomain in pink; and the C-terminal subdomain in orange.

The N-terminal domain begins with one α helix and two 3_{10} helices ($\alpha 1$, $\eta 1$, and $\eta 2$, Figure 67, blue). Helix $\eta 2$ is followed by a five-stranded twisted antiparallel β sheet (Figure 67, light sea green), a highly conserved structural motif among eukaryotic protein kinases, especially strands A and B, which contain the ATP-binding motif. The $\alpha 2$ helix is inserted between strands B and C (Figure 67, purple), while helices $\eta 3$ and $\alpha 3$ are inserted between strands C and D (Figure 67, magenta). Helices $\alpha 2$ and $\eta 3$ are not found in ChoKA-2 and human ChoK $\alpha 1$ and constitute a unique feature of PfChoK, while helix $\alpha 3$ is shorter than the corresponding helices in ChoKA-2 and human ChoK $\alpha 1$. Interestingly, whereas helix $\alpha 3$ mediates dimer formation, both in the ChoKA-2 and in the human ChoK $\alpha 1$ crystal structure, the structure of PfChoK is monomeric. Consistent with this observation, analytical size-exclusion chromatography analysis shows that the oligomerization state of PfChoK in solution is also monomeric (Figure 61).

The N-terminal domain is connected to the C-terminal one by a short linker featuring a short β strand (βF , Figure 67, yellow) within a longer loop portion that makes a 90 degree turn. The PfChoK C-terminal domain is divided in three regions: the central core subdomain (Figure 67, red), the insertion subdomain (Figure 67, pink), and the C-terminal subdomain (Figure 67, orange). The central core subdomain is composed by helices $\eta 4$, $\alpha 4$, and $\alpha 7$ and by a harpin-shaped loop that contains two β sheets (formed by the antiparallel strands βG , βH , and βI , βJ) and $\eta 6$, a 3_{10} helix. These two β sheets

are especially relevant for PfChoK catalytic activity since they house the Brenner's motif (β G, β H, and η 6) and the choline kinase motif (β I, β J).

The insertion subdomain is placed between helix α 4 and strand β G and is composed by two helices, α 5 and α 6, which are connected by a loop and by helix η 5. In the PfChoK structure, this insertion is located behind the central core and the C-terminal subdomains and, unlike the *C. Elegans* and the human enzymes, where a long β hairpin motif and a α helix are present between the helices corresponding to α 5 and α 6, it features a loop.

The C-terminal subdomain is located between the insertion and the core subdomains, and it is connected to the final portion of the central core subdomain by a hairpin-like loop containing a β sheet composed by β strands K and L. Since this hairpin-like loop is positioned in front of the choline binding site, it is believed that it may regulate the accessibility to the active site; hence, the binding of choline, as proposed previously [145].

The harpin-like loop is followed by helices η 8, α 8, α 9, and α 10, which compose the remaining of the C-terminal subdomain. Helix α 8 is situated beneath helix α 4 and is connected to helix α 9 through a reverse turn. Helix α 9 runs across the whole molecule and features a kink in its middle region. Another reverse kink connects helix α 9 to α 10, which extends towards the C-terminus of the enzyme. While structure superimposition between the apo and the ADP-bound form of PfChoK shows no significant conformational variability (rmsd. = 0.38 Å for C α atoms), a comparison between the PfChoK apoenzyme and ChoKA-2 (PDB code: 1NW1) and hCK α 1 (PDB code: 2CKO) shows relatively high rmsd values (3.75 Å and 2.35 Å, respectively). This is despite the overall good sequence alignment and the similar folding topology. However, the rmsd values get smaller when the individual N-terminal domains are considered (1.6 Å for ChoKA-2 and 1.15 Å for human ChoK α 1), while they remain relatively high for the C-terminal domain (2.10 Å for ChoKA-2 and 2.20 Å for human ChoK α 1). These rmsd values for the C-terminal domain seem to be due to the marked differences in the secondary structure of the insertion subdomain.

From a phylogenetic point of view, choline kinases have been classified as members of the atypical kinases (AK) family, a subgroup of eukaryotic protein kinases having a homologous catalytic core but not having all the other usual kinase motifs conserved [147].

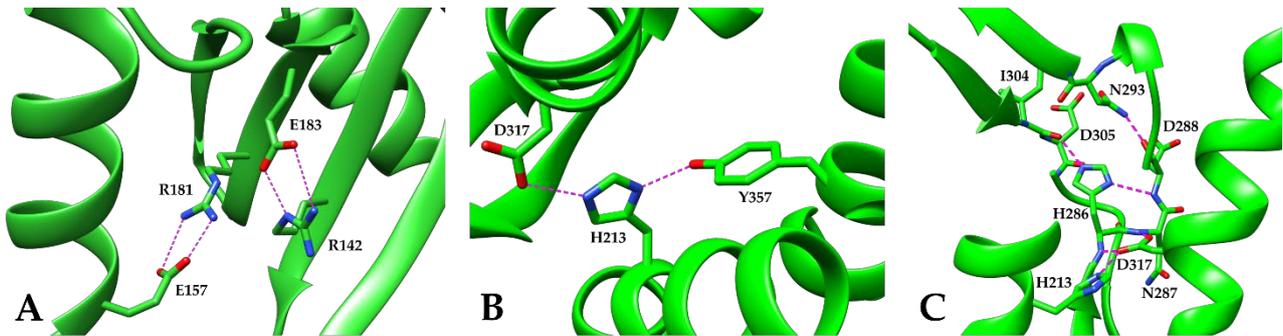


Figure 68: Structural features of PfChoK

(A) The salt bridge interactions between Arg142 and Glu183 and Arg181 and Glu157 in the N-terminal domain. (B) Stabilizing role of His213, which forms H-bonding interactions with Asp317 and Tyr357. (C) hydrogen bonding network stabilizing the central catalytic core domain. The protein is shown as green cartoon, while highlighted residues are depicted in sticks colored with green carbon atoms, blue nitrogen atoms, and red oxygen atoms. Structural details are from the structure of the PfChoK apoenzyme (PDB code: 6YXS).

Aside from the conserved catalytic central core domain, other AK family structural features are also present in the crystal structure of the PfChoK apoenzyme (Figure 68). For instance, two highly conserved residues, Arg142 and Glu183 form a salt bridge that plays an important stabilizing role in the architecture of the N-terminal domain (Figure 68, panel A). Moreover, Arg142 has a role in the binding of ADP since it can interact with both its α and its β phosphate group.

Another salt bridge between conserved residues in the ChoK family is formed between Arg181 and Glu157; thus, stabilizing the interaction between β strand E and helix $\alpha 3$ (Figure 68, panel A).

The residue His213 (Figure 68, panel B), which is situated in helix $\alpha 4$, is also highly conserved and possesses an important structural role. Indeed, by linking together three different helices of the C-terminal domain, it greatly contributes to the overall fold stability. This linkage is provided by the side chain of His213 forming a hydrogen bond with the hydroxyl group of Tyr357 in helix $\alpha 8$ and with the carboxyl group of Asp317 in helix $\alpha 7$.

Another important structural motif that contributes to the stability of the eukaryotic protein kinase fold, and in particular of its catalytic central core, is the hydrogen-bonding network that involves those highly conserved residues that coordinate magnesium. This network (Figure 68, panel C) features a hydrogen bond between His213 and Asp317, while the side chain of the latter contacts the nitrogen atoms of the backbone amide group of His286 and Asn287 situated in the Brenner's motif. The imidazole ring of His286 is sandwiched between the backbone carbonyl groups of Ile304 and Asp305 situated in the choline kinase motif and the backbone nitrogen of Asp288 positioned in the Brenner's motif. The backbone carbonyl group of Asp288 hydrogen bonds the nitrogen in the amide group of Asn293, which is located in helix $\eta 6$ in the Brenner's motif and whose side chain is also involved with the carbonyl group of Asp305 in the coordination of the catalytic magnesium. Four water molecules complete the octahedral coordination of the magnesium ion. All metal coordination distances are within a 2.02–2.15 Å range (Figure 68, panel A).

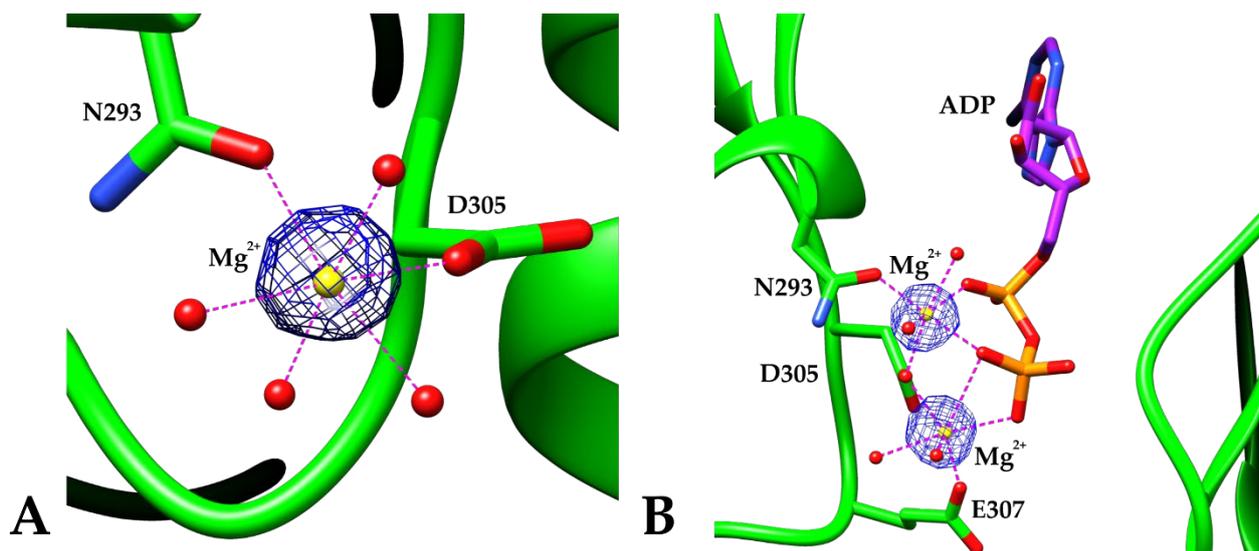


Figure 69: Magnesium ion coordination in the apoenzyme (PDB code: 6YXS) (A) and in the ADP-bound form (PDB code: 6YXT) (B) Magnesium ions are depicted in small yellow spheres, interacting residues are indicated with the corresponding number and one letter code, shown in ball and stick and are colored with carbon atoms in green, nitrogen atoms in blue, and oxygen atoms in red. The electron density from omit maps is contoured at the 4.0σ level. The ADP molecule is drawn in stick with purple carbon atoms, red oxygen atoms and orange phosphorus atoms.

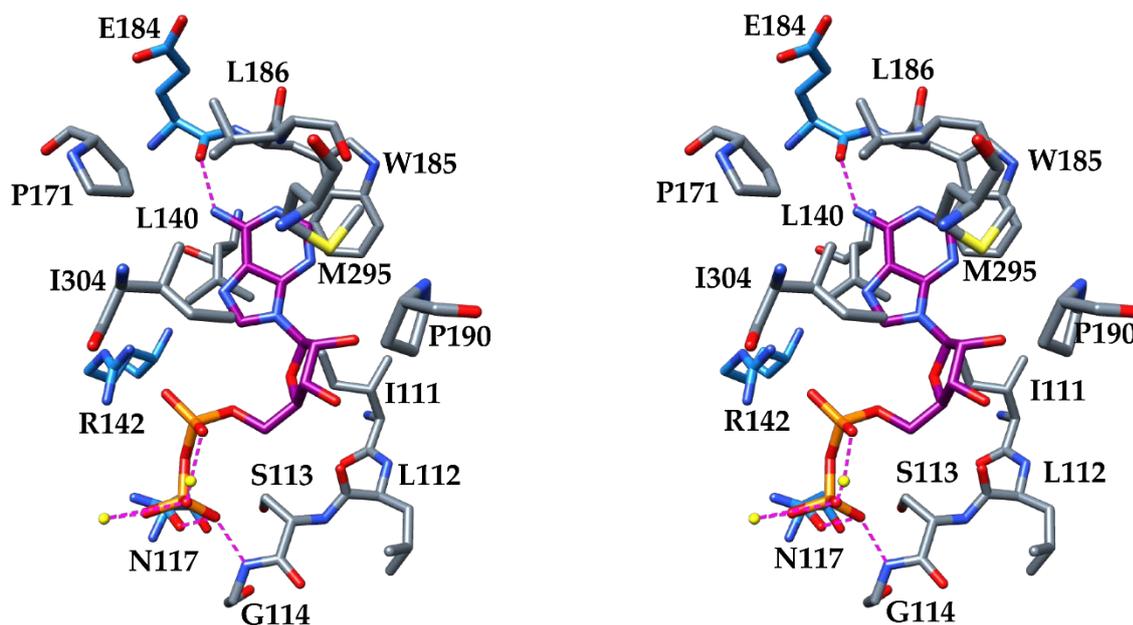


Figure 70: Stereodiagram of the ADP binding mode

The hydrophobic interactions with the non-polar residues that line the pocket in which ADP is bound are shown with grey carbon atoms, while the polar residues that interact with ADP by hydrogen bonding are shown with light blue carbon atoms. All protein residues are drawn in ball and stick, while the ADP molecule is drawn in sticks and shown with purple carbons. Nitrogen atoms are colored in blue, oxygen atoms in red, and phosphorus atoms in orange. Structural details are from the ADP-bound form of PfChoK (PDB code: 6YXT).

The ADP-bound structure reveals the binding mode of the adenosine nucleotide in PfChoK (Figure 70). The ligand is found in a hydrophobic pocket comprised by the N-terminal domain and the C-terminal central core domain and involves many residues of the Brenner's motif. The hydrophobic pocket is mainly lined by non-polar residues, such as Ile111, Leu112, Ser113, Leu140, Pro171, Trp185, Leu186, Pro190, Met295, and Ile304. Moreover, ADP is further stabilized within the enzyme by a number of hydrogen bonds. In particular, the backbone amide group of Gly114 interacts with one of the oxygen atoms in the β phosphate group (2.87 Å), the side chain of Asn117 makes a hydrogen bond with the same oxygen of the β phosphate group (2.98 Å) and the backbone carbonyl group of Glu184 forms a hydrogen bond with the nitrogen of the adenine ring (2.84 Å). Furthermore, here the highly conserved Arg142 is observed in the proximity of one of the oxygen atoms of the α phosphate group of ADP at a distance of 3.42 Å, forming an electrostatic interaction.

The ADP molecule also participates in the coordination of two magnesium ions (Figure 69, panel B). Unlike the regular octahedral coordination of the magnesium ion in the apoenzyme structure, here the octahedral coordination is highly distorted due to the participation of the ADP phosphate groups in the metal coordination sphere. One magnesium ion is coordinated by the side chains of Asp305 and Glu307, by two water molecules and by two oxygen atoms of the β phosphate group, while the second one is coordinated by the side chains of Asn293 and Asp305, two water molecules and two oxygen atoms of the ADP molecule originating from the α and β phosphate groups.

In our PfChoK structure (6YXT), ADP binds to the same ADP-binding motif as in human ChoK α 1 (2CKP) [148], in the incomplete ADP- and phosphocholine-bound form of the *P. falciparum* enzyme (3FI8, unpublished data), and in the co-crystal structure of human ChoK α 1, ADP, and hemicholinium-3 (3G15) [258] (Figure 71), even though, in all these structures, the presence and the number of magnesium ions vary.

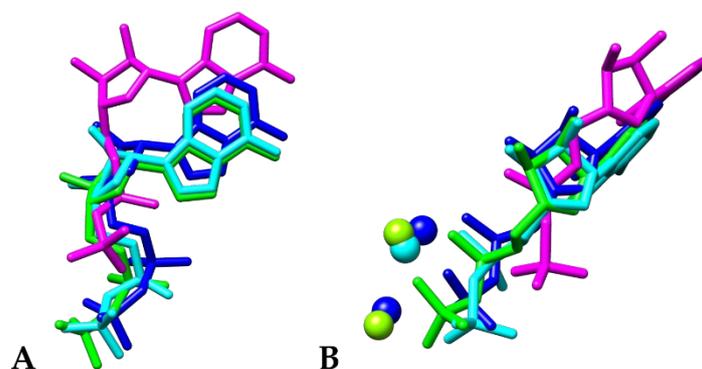


Figure 71: Structural superimposition of ADP molecules and ADP molecules with magnesium ions (A) Structural superimposition of ADP molecules in the ADP-bound form of PfChoK (PDB code: 6YXT, green), the ADP- and phosphocholine-bound form of PfChoK (PDB code: 3FI8, cyan), the ADP-bound form of human ChoK α 1 (PDB code: 2CKP, magenta), and the co-crystal structure of human ChoK α 1, ADP and hemicholinium-3 (PDB code: 3G15, blue). (B) Structural superimposition of ADP and magnesium ions with the same models and with the same color code. ADP molecules are shown in sticks and magnesium ions are drawn as spheres.

By comparing our ADP-bound PfChoK structure (6YXT) with 3FI8, we cannot detect any differences in the enzyme portion and in the residues that bind the ADP molecule. Indeed, the only appreciable difference between the two structures is the presence of only one magnesium ion that is coordinated by Asn293 and Asp305 (cyan sphere in Figure 71, panel B) and that interacts with ADP. The other magnesium ion that is present in 6YXT is not found in 3FI8. The positions of the magnesium ions and of ADP in our PfChoK structure (6YXT) are similar to those observed in the ADP- and hemicholinium-3-bound human ChoK α 1 structure (3G15). Since it is known that two magnesium ions are required to perform the breaking of the ATP phosphodiester bond, the varying number of magnesium ions that are present in the crystal structures may suggest that these are different stages in the catalytic process that leads to the production of phosphocholine.

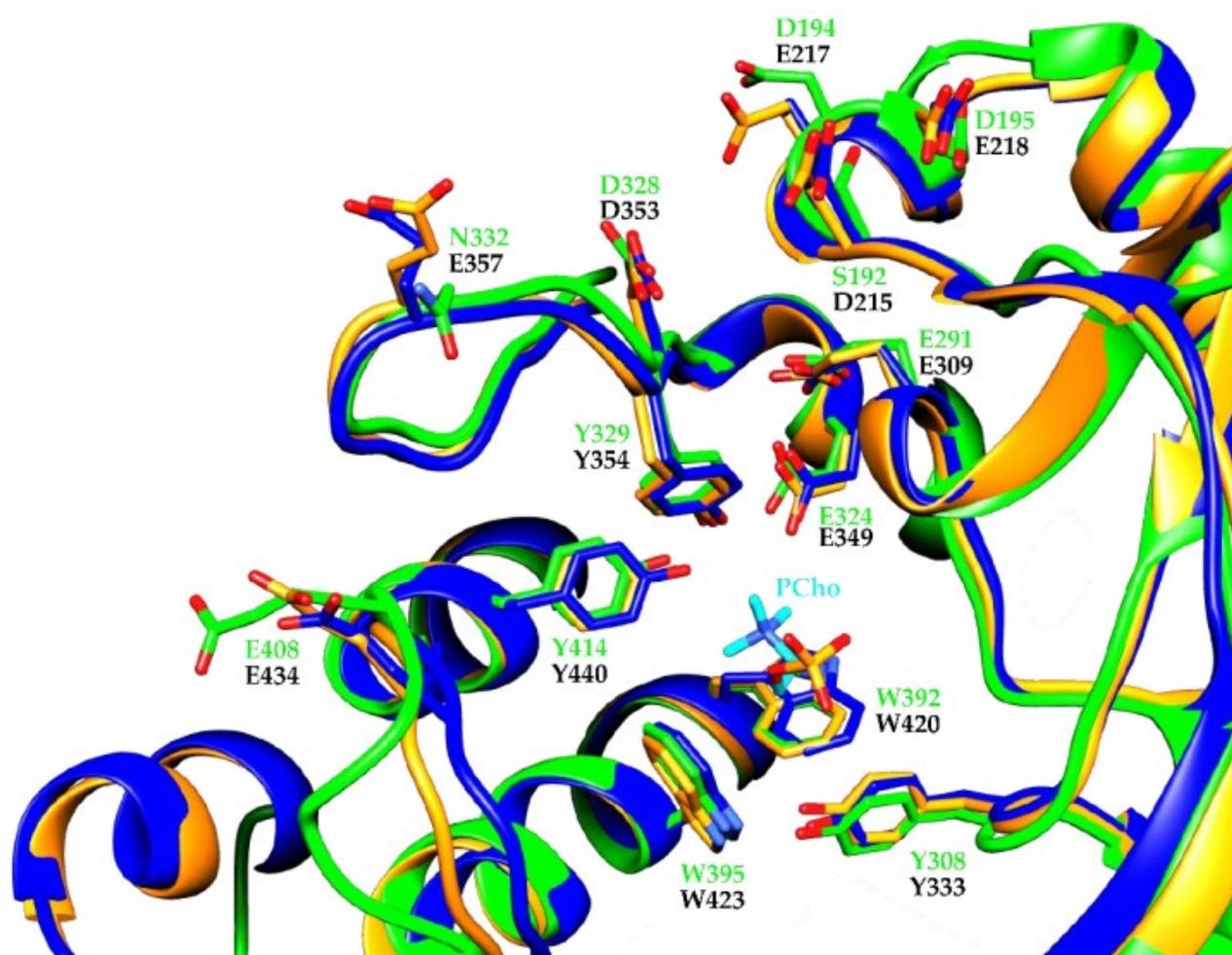


Figure 72: Structure alignment of the PfChoK (green cartoon) and the human ChoKa1 (blue and orange cartoons) choline binding sites

The alignment was obtained by superimposing the human apoenzyme (PDB code: 2CKO, orange cartoon), the phosphocholine-bound form of the human enzyme (PDB code: 2CKQ, blue cartoon) and the apo form of PfChoK (PDB code: 6XYS, green cartoon). Residues involved in choline binding are shown in sticks with orange, blue or green carbon atoms, blue nitrogen atoms, and red oxygen atoms. Residues are labeled with the corresponding one letter code and residue number, which is green for PfChoK and black for 2CKO and 2CKQ. Phosphocholine (labelled PCho) is shown in sticks with carbon atoms in cyan, nitrogen atoms in blue, oxygen atoms in red, and phosphorous atoms in yellow.

As for the choline binding site, the majority of the residues that bind choline in the human enzyme are conserved in PfChoK as well (Figure 72). In the human ChoK α 1, the choline binding site is an essentially hydrophobic pocket that is capped by negatively charged residues [148]. Hence, two different functional regions can be identified: one lined by Asp215, Glu217, Glu218, Glu309, Glu349, Asp353, Glu357, and Glu434 (human ChoK α 1 numbering, Figure 72), which are involved in the electrostatic steering of choline and in the binding of the quaternary amine; the other forming a groove lined by Tyr333, Tyr354, Trp420, Trp423, and Tyr440, which provide direct binding to choline. It should be noted that there are no structural differences between the human apoenzyme (2CKO) and the human phosphocholine-bound form of the enzyme (2CKQ) (Figure 72, orange, and magenta residues).

In PfChoK, while the hydrophobic residues lining the choline-binding pocket are conserved (Tyr308, Tyr329, Trp392, Trp395, and Tyr414 in the parasite enzyme sequence) and superimpose well with the corresponding residues in the choline-binding site of the human enzyme, the majority of the differences are among the negatively charged residues involved in choline binding. Indeed, Glu217, Glu218, and Glu357 in the human enzyme are replaced by Asp194, Asp195, and Asn332 in PfChoK, the latter of the three substitutions altering the electrostatic landscape of the loop that is found in close proximity to the pocket. Finally, Asp215 of human ChoK α 1 is replaced by Ser192 in PfChoK. All other residues (Glu309, Glu349, and Asp353) in human ChoK α 1 are conserved and superimpose well with their PfChoK counterparts (Glu291, Glu324, and Asp328).

Antiplasmodial Activity and *Ex Vivo* Potency

Table 9 and Table 10 and Figure 62 and Figure 63 summarize the effect of the final compounds in infected erythrocytes and *ex vivo*. From here on, for every compound the CK prefix is omitted for brevity. In general, all compounds show nanomolar activity in infected erythrocytes, the most active compound being Fg-14 (11.5 nM). In terms of structure, there seems to be no difference between monocationic and bis-cationic compounds regarding the *in vitro* antimalarial activity, while the *ex vivo* potency of bicationic compounds is higher. Among the monocationic compounds, Fa-M3 (isomer thieno[2,3-d]pyrimidin) stands out, but the rest of the isomers follow up very closely, and the difference in the *in vitro* activity is not considered to be remarkable.

With respect to the bicationic compounds and considering the five types of linker used in *in vitro* experiments, it can be observed that when the spacer is bipyridinyl (Fg-12, Fg-17, Fa-27, Fg-32, Fa-26 and Fg-20) there is a notable decrease in the activity with respect to their biphenylic counterparts (Fg-9, Fa-21, Fg-14, Fa-24, Fg-30, Fg-10, Fa-22, Fg-18, Fp-1 and Fp-8). This difference is more pronounced when the volume of the 4-substituent in the cationic head increases (Fa-22 and Fg-18 vs Fa-26 and Fg-20). Conversely, when the substituent is pyrrolidine there is hardly any difference, except for Fg-13 and Fg-17 vs Fg-14 and Fa-21, where there is more than a 10-fold difference in activity.

The remaining biphenyl and bibenzyl, biphenethyl and 1,2-diphenoxyethane linkers feature very similar values. Starting with the biphenyl spacer, the substituent at 4 has a visible influence, with pyrrolidine again giving the best results (see Table 9 Fa-21 and Fg-14) the increased volume of this cycloalkylamine from piperidine to azepane gradually decreases the activity. However, it can be considered one of the best families, as there is hardly any difference between the less active Fg-18 (GI_{50} = 60 nM) and the more active Fg-14 (GI_{50} = 11 nM). It is worth noting that in this family two thienopyrimidine isomers Fp-1 and Fp-8 substituted at 4 by an *N*-methyl-aniline and *p*-Chloro-*N*-methyl-aniline, were also synthesized and evaluated in order to compare the results with those where the substituent is a cycloalkylamine. However, there was a drastic decrease in activity. Finally, the thienopyrimidin isomers stand out versus the thienopyridin isomers, so the introduction of a second N atom in the cationic head is highly beneficial.

Continuing the SAR analysis, the pattern is the same, the thieno[2,3-d] pyrimidine isomers stand out versus thieno[3,2-d]pyridine in both, when the spacer is bibenzyl (Fg-15 and Fg-31) and when there are 4 carbon atoms between the aromatic rings (Ff-3, Ff-6 and Ff-35). The volume of the pyrrolidine ring is postulated to be the best even if the differences are not too great especially, in the biphenyl derivatives.

The 1,2-diphenoxyethane derivatives behave similarly and the introduction of 2 oxygen atoms between the aromatic rings does not seem to affect them. In comparison to the reference compounds 10a and 10k, previously described in [306], it can be observed how the homologous compounds Ff-4 and Ff-5 provides better values. Both of them belong to the series of bioisosteres derived from thieno[2,3-d]pyrimidin-1-ium, with a pyrrolidine and piperidine group in position 4 respectively. In the consecutive isomers with azepane Ff-34 and Ff-36 (similar 10k with quinoline), the activity decreases and inexplicably in Ff-36 (thieno[2,3-d]pyrimidin-1-ium), it exhibits 6 times less activity than 10k.

All final compounds were analyzed as PfChoK inhibitors in the 25 μM screening, and those with a small residual activity had their IC_{50} calculated. The values obtained show, at first sight, two striking findings; the first is that the monocationic compounds, despite their potent antimalarial activity, do not inhibit choline kinase activity at 25 μM . The second is that the bipyridinic linker derivatives' (with moderate antimalarial values) are likewise no inhibitors of PfChoK at 25 μM .

Starting with the biphenyl linker derivatives, compounds Fp-1 and Fp-8 feature the most remarkable IC_{50} values (7.27 ± 0.23 and 4.17 ± 0.48 μM respectively). Interestingly, these two thienopyrimidine isomers have *N*-methyl-aniline and *p*-Chloro-*N*-methyl aniline as substituents at the 4-position of the heterocycle. Therefore, these compounds can be considered as bioisosteres of compound TCD-717 which had an IC_{50} of 1.75 μM and where the 7-quinoline ring is replaced by the two thienopyrimidine isomers. The antimalarial activity has also decreased up to 10-fold with respect to TCD-717: from 0.0265 μM to 0.1422 μM for Fp-1 and 0.2816 μM for Fp-8. These results point out that the bioisosteric change has favoured neither the inhibitory nor the antimalarial activity. However, it is noticed that there is a direct correlation between both activities and therefore, these compounds would display a similar mechanism of action, i.e. the antimalarial activity is directly related to the inhibition of the enzyme when the substrate is choline and hence shows activity towards choline kinase and not ethanolamine kinase. The differences in enzyme activity are lower than those in antimalarial activity and this could be because of the decreased lipophilicity of these new compounds: Fp-1 ($\text{cLogP} = 3.413$) and Fp-8 ($\text{cLogP} = 4.44$) while for RSM-923A is ($\text{cLogP} = 7.44$). In this same family, however, it is observed that when the substituents in 4 are cycloalkylamines, for example Fa-22 and Fg-18, the IC_{50} of the enzyme activity reaches moderated values of 11 μM , while the antimalarial activity is 0.0514 μM and 0.0604 μM respectively, so it is deduced that an alternative and complementary mechanism to PfChoK inhibition could be involved. Only a moderate increase in enzymatic activity seems to be observed when the isomer is thieno[3,2-d]pyrimidine with respect to the other isomer, which has the sulphur closer to the positive charge of the cationic heads.

The introduction of 2 carbon atoms between the aromatic rings (the bibenzyl family), exerts a negative effect on the enzyme inhibition, being Fg-11 and Fg-16 the most active compounds. In addition, compound Fg-16 was found to be one of the most lethal against the parasite regarding this family and it can be considered that the inhibition of the enzyme is at least partly involved. The addition of 2 more carbon atoms (Ff-1, Ff-7, Ff-3, Fa-33, Ff-6, Fa-29 and Ff-35) produces a significant increase in the enzyme inhibition, which is also partially reflected in the notable improvement of the antimalarial activity.

Among these two families, the most active compound is Fa-29, with an $IC_{50}=4.64 \pm 0.20 \mu\text{M}$. It is a derivative of thieno[3,2-d]pyrimidin-1-ium with an azepanyl group as substituent and biphenethyl as a linker and shows an antimalarial activity of 55.6 nM, which is not the best value but exhibits a direct correlation with the inhibition values.

It is worth noting that the asymmetrical reference compounds BR-23 and BR-25 [305], with excellent antimalarial values, can be considered bioisosteres of the compounds described in this work: Fg-9, Fa-21, Fg-14 for BR-23 and Ff-1, Ff-7 and Ff-3 for BR-25. A 10-fold decrease in antimalarial activity is observed with respect to the reference compounds, but it is nevertheless striking that the synthesized bioisosteres do possess enzymatic activity when the substrate is choline, in contrast to BR-23 and BR-25, which only had inhibitory activity when the enzyme-substrate was ethanolamine.

Analysing the compounds with the linker 1,2-diphenoxyethane, it can be observed that compounds Ff-34 and Ff-36 show the lowest residual enzyme activity with values of $52.17 \% \pm 12.43$ and $46.39 \% \pm 12.44$, which are precisely the homologues of compound 10k [306] with quinoline and azepane as substituent in position 4. However, there is a difference with respect to enzyme activity, since for compound 10a, the mechanism of action must be different as no inhibition of the enzyme in its choline kinase or ethanolamine kinase activity is observed. In contrast, when the head is quinolinic and carries an azepane group in position 4 (10k), there are values similar to those presented here.

Determination of the Inhibition Mode

For low enzyme concentrations (less than 10 nM in our assay) and substrate concentrations around the K_M , the relation between the dissociation constant and the IC_{50} can be described mathematically by the Cheng-Prusoff relationship [308] that has been formulated for each modality of inhibition. For competitive and uncompetitive inhibition, the study of the Cheng-Prusoff equation results in the following relationship:

$$IC_{50} = 2 K_i$$

meaning that the IC_{50} value is always greater than the dissociation constant by a factor of 2.

However, for mixed inhibition the relationship is different and the IC_{50} value is generally greater than the K_i by a factor that can range between 1 and 2, depending on the ratio between the dissociation constant of the inhibitor w.r.t. the enzyme and to the enzyme-substrate complex.

Being a special case of mixed inhibition where the dissociation constants of the inhibitor w.r.t. the enzyme and to the enzyme-substrate complex are equal, noncompetitive inhibition is the only case in which the Cheng-Prusoff relationship:

$$IC_{50} = \left(\frac{[S] + K_M}{\frac{[S]}{K_i^{ES}} + \frac{K_M}{K_i^E}} \right)$$

leads to

$$IC_{50} = K_i$$

since $K_i^{ES} = K_i^E$ and therefore, the IC_{50} value is equal to the dissociation constant.

Thus, a close inspection of the IC_{50} and K_i values reported in Table 11 reveal that in many cases the two measured values are close, indicating that noncompetitive inhibition is the inhibition modality that best fits the experimental data.

Furthermore, the value of the α index also corroborates the noncompetitive inhibition mechanism, since it approaches the unity in all considered cases, it indicates that the binding of the inhibitor does not affect the binding of the substrates.

Conclusions and Future Perspectives

The analysis of Figure 62 and Table 9 reveal that the compounds that feature the sulfur-containing linker perform significantly better as *ex vivo* PfChoK inhibitors and they are characterized by a potency that is almost ten times higher compared to the other compounds and it is always less than 1 μM . This effect seems to be independent from the type of cationic head that is present on the compound and only due to the different chemical nature of the linker. However, since this project is still ongoing, *in vitro* antiplasmodial activity data on this subgroup of compounds is still under investigation and therefore a comparison with the other that are already characterized is not possible yet.

Despite this fact, the higher potency of these compounds prompted us to characterize their mechanism of action, which resulted to be noncompetitive. Therefore, one can speculate based on the obtained biochemical data that these compounds do not bind to the choline binding site as it might be expected from the comparison with their human counterparts, but rather bind the enzyme at another binding site, which still needs to be discovered.

Thus, the future perspectives of this work are the evaluation of the *in vitro* antiplasmodial activity of this class of compounds and the determination of a co-crystal structure, which would allow the elucidation of the novel binding site that can be predicted based on biochemical data.

Bibliography

List of References listed in the order of appearance in the text:

- [1] J. P. Thiery and al., "Epithelial-Mesenchymal Transitions in Development and Disease," *Cell*, vol. 139, no. 5, pp. 871-890, 2009.
 - [2] B. Alberts and al., "Chapter 19: Cell Junctions and the extracellular Matrix," in *Molecular Biology of the Cell 6th Edition*, Garland Science, 2014, pp. 1035-1090.
 - [3] I. Gul and al., "Evolution and diversity of cadherins and catenins," *Experimental Cell Research*, vol. 358, no. 1, pp. 3-9, 2017.
 - [4] P. Hulpiau and al., "Molecular evolution of the cadherin superfamily," *International Journal of Biochemistry and Cell Biology*, vol. 2009, no. 41, pp. 349-369, 2009.
 - [5] K. Sano and al., "Protocadherins: a large family of cadherin-related molecules in central nervous system," *EMBO Journal*, vol. 12, no. 6, pp. 2249-2256, 1193.
 - [6] Q. Wu and al., "A striking organization of a large family of human," *Cell*, vol. 97, pp. 779-790, 1999.
 - [7] M. Frank and al., "Protocadherins," *Current Opinions in Cell Biology*, vol. 14, no. 5, pp. 557-562, 2002.
 - [8] A. Pancho and al., "Protocadherins at the Crossroad of Signaling Pathways," *Frontiers of Molecular Neuroscience*, vol. 13, no. 117, 2020.
 - [9] K. Soo-Young and al., "Non-clustered protocadherin," *Cell Adhesion and Migration*, vol. 5, no. 2, pp. 97-105, 2011.
 - [10] M. W. Wendeler and al., "Unique gene structure and paralogy define the 7D-cadherin family," *Cellular and Molecular Life Sciences*, vol. 63, no. 13, pp. 1564-1573, 2006.
 - [11] B. Kreft and al., "LI-Cadherin-mediated Cell-Cell Adhesion Does Not Require Cytoplasmic Interactions," *Journal of Cell Biology*, vol. 136, no. 5, pp. 1109-1121, 1997.
 - [12] A. M. Goffinet and al., "Seven passcadherins CELSR1-3," *Seminars in Cell & Developmental Biology*, vol. 69, pp. 102-110, 2017.
-

- [13] J. Brash and al., "Homophilic and Heterophilic Interactions of Type II Cadherins Identify Specificity Groups Underlying Cell-Adhesive Behavior," *Cell reports*, vol. 23, no. 6, pp. 1840-1852, 2018.
- [14] O. J. Harrison and al., "The Extracellular Architecture of Adherens Junctions Revealed by Crystal Structures of Type I Cadherins," *Structure*, vol. 19, no. 2, pp. 244-256, 2011.
- [15] Y. Wu and al., "Cooperativity between trans and cis interactions in cadherin-mediated junction formation," *Proceedings of the National Academy of Sciences*, vol. 107, no. 41, pp. 17592-17597, 2010.
- [16] E. Delva and al., "The Desmosome," *Cold Spring Harbor Perspectives in Biology*, vol. 1, no. 2, 2009.
- [17] S. D. Patel and al., "Type II Cadherin Ectodomain Structures: Implications for Classical Cadherin Specificity," *Cell*, vol. 2006, no. 6, pp. 1255-1268, 2006.
- [18] C. Dewitz and al., "Organization of motor pools depends on the combined function of N-cadherin and type II cadherins," *Development*, vol. 146, no. 13, 2019.
- [19] Y. Hirai and al., "Expression and role of E- and P-cadherin adhesion molecules in embryonic histogenesis. II. Skin morphogenesis," *Development*, vol. 105, no. 2, pp. 271-277, 1989.
- [20] M. Fujita and al., "Expression of cadherin cell adhesion molecules during human skin development: morphogenesis of epidermis, hair follicles and eccrine sweat ducts," *Archives of Dermatological Research*, vol. 284, no. 3, pp. 159-166, 1992.
- [21] I. Lung and al., "Expression and role of E- and P-cadherin adhesion molecules in embryonic histogenesis. I. Lung epithelial morphogenesis," *Development*, vol. 105, no. 2, pp. 263-270, 1989.
- [22] Y. Shimoyama and al., "Expression of E- and P-cadherin in gastric carcinomas," *Cancer Research*, vol. 51, no. 8, pp. 2185-2192, 1991.
- [23] Y. Shimoyama and al., "Cadherin intercellular adhesion molecule in hepatocellular carcinomas: loss of E-cadherin expression in an undifferentiated carcinoma," *Cancer Letters*, vol. 57, no. 2, pp. 131-135, 1991.
- [24] D. Shook and al., "Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development," *Mechanisms of Development*, vol. 120, no. 11, pp. 1351-1383, 2003.
- [25] T. J. C. Harris and al., "Adherens junctions: from molecules to morphogenesis," *Nature Reviews Molecular Cell Biology*, vol. 11, no. July 2010 Issue, pp. 502-514, 2010.
-

- [26] M. Takeichi, "Dynamic contacts: rearranging adherens junctions to drive epithelial remodelling," *Nature reviews Molecular Cell Biology*, vol. 15, no. June 2014 Issue, pp. 397-410, 2014.
- [27] T. J. Boggon and al., "C-Cadherin Ectodomain Structure and Implications for Cell Adhesion Mechanisms," *Science*, vol. 296, no. 5571, pp. 1308-1313, 2002.
- [28] A. Al-Amoudi and al., "The molecular architecture of cadherins in native epidermal desmosomes," *Nature*, vol. 450, no. 7171, pp. 832-837, 2007.
- [29] W. He and al., "Untangling desmosomal knots with electron tomography," *Science*, vol. 302, no. 5642, pp. 109-113, 2003.
- [30] O. Pertz and al., "A new crystal structure, Ca²⁺ dependence and mutational analysis reveal molecular details of E-cadherin homoassociation," *EMBO Journal*, vol. 18, no. 7, pp. 1738-1747, 1999.
- [31] K. Kamiya and al., "Cadherin-integrated liposomes with potential application in a drug delivery system," *Biomaterials*, vol. 32, no. 36, pp. 9899-9907, 2011.
- [32] N. Vunnam and al., "Calcium-induced strain in the monomer promotes dimerization in neural cadherin," *Biochemistry*, vol. 50, no. 39, pp. 8437-8444, 2011.
- [33] C. Ciatto and al., "T-cadherin structures reveal a novel adhesive binding mechanism," *Nature Structural & Molecular Biology*, vol. 17, no. 3, pp. 339-347, 2010.
- [34] H. McNeill and al., "Spatial and temporal dissection of immediate and early events following cadherin-mediated epithelial cell adhesion," *Journal of Cell Biology*, vol. 120, no. 5, pp. 1217-1226, 1993.
- [35] J. Brasch and al., "Thinking outside the cell: How cadherins drive adhesion," *Trends in Cell Biology*, vol. 22, no. 6, pp. 299-310, 2012.
- [36] K. Biswas and al., "Early events in the assembly of E-cadherin adhesions," *Experimental Cell Research*, vol. 358, no. 1, pp. 14-19, 2017.
- [37] C. Adams and al., "Mechanisms of Epithelial Cell–Cell Adhesion and Cell Compaction Revealed by High-resolution Tracking of E-Cadherin– Green Fluorescent Protein," *The Journal of Cell Biology*, vol. 142, no. 4, pp. 1105-1119, 1998.
- [38] S. J. Stehbens and al., "Microtubules and cadherins: a neglected partnership," *Frontiers in Bioscience*, vol. 14, no. January 2009 Issue, pp. 3159-3167, 2009.
-

- [39] M. Shahbazi and al., "Microtubules CLASP to adherens junctions in epidermal progenitor cells," *Bioarchitecture*, vol. 4, no. 1, pp. 25-30, 2014.
- [40] B. T. Quang and al., "Principles of E-cadherin supramolecular organization in vivo," *Current Biology*, vol. 23, no. 22, pp. 2197-2207, 2013.
- [41] E. Parisini and al., "The crystal structure of human E-cadherin domains 1 and 2, and comparison with other cadherins in the context of adhesion mechanism," *Journal of Molecular Biology*, vol. 373, no. 2, pp. 401-411, 2007.
- [42] M. J. Bennet and al., "3D domain swapping: a mechanism for oligomer assembly," *Protein Science*, vol. 4, no. 12, pp. 2455-2468, 1995.
- [43] J. Vendome and al., "Molecular design principles underlying β -strand swapping in the adhesive dimerization of cadherins," *Nature Structural and Molecular Biology*, vol. 18, no. 6, pp. 693-700, 2011.
- [44] M. Renaud-Young and al., "In the First Extracellular Domain of E-cadherin, Heterophilic Interactions, but Not the Conserved His-Ala-Val Motif, Are Required for Adhesion," *The Journal of Biological Chemistry*, vol. 277, no. 42, pp. 39609-39616, 2002.
- [45] S. Kudo and al., "Adhesive Dimerization of Human P-Cadherin Catalyzed by a Chaperone-like Mechanism," *Structure*, vol. 24, no. 9, pp. 1523-1536, 2016.
- [46] O. J. Harrison and al., "Two-step adhesive binding by classical cadherins," *Nature Structural and Molecular Biology*, vol. 17, no. 3, pp. 348-357, 2010.
- [47] B. Nagar and al., "Structural basis of calcium-induced E-cadherin rigidification and dimerization," *Nature*, vol. 380, pp. 360-364, 1996.
- [48] S. Kudo and al., "Identification and characterization of the X-dimer of human P-cadherin: implications for homophilic cell adhesion," *Biochemistry*, vol. 53, no. 11, pp. 1742-1752, 2014.
- [49] S. Rakshit and al., "Ideal, catch, and slip bonds in cadherin adhesion," *Proceedings of the National Academy of Sciences of the USA*, vol. 109, no. 46, pp. 18815-18820, 2012.
- [50] V. Nardone and al., "Crystal Structure of Human E-Cadherin-EC1EC2 in Complex with a Peptidomimetic Competitive Inhibitor of Cadherin Homophilic Interaction," *Journal of Medicinal Chemistry*, vol. 59, no. 10, pp. 5089-5094, 2016.
- [51] A. Dalle vedove and al., "The X-ray structure of human P-cadherin EC1-EC2 in a closed conformation provides insight into the type I cadherin dimerization pathway," *Acta Crystallographica Section F: Structural Biology Communications*, vol. 71, pp. 371-380, 2015.
-

- [52] A. V. Priest and al., "Biophysical basis of cadherin mediated cell-cell adhesion," *Experimental cell Research*, vol. 358, no. 1, pp. 10-13, 2017.
- [53] K. Manibog and al., "Molecular determinants of cadherin ideal bond formation: Conformation-dependent unbinding on a multidimensional landscape," *Proceedings of the National Academy of Sciences of the USA*, vol. 113, no. 39, pp. E5711-E5720, 2016.
- [54] J. Thiery and al., "Epithelial-Mesenchymal Transitions in Development and Disease," *Cell*, vol. 139, no. 5, pp. 871-890, 2009.
- [55] J. Thiery and al., "Complex networks orchestrate epithelial-mesenchymal transitions," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 2, pp. 131-142, 2006.
- [56] R. Kalluri and al., "The basics of epithelial-mesenchymal transition," *Journal of Clinical Investigation*, vol. 119, no. 6, pp. 1420-1428, 2009.
- [57] U. N. C. Institute, "Cancer Classification," [Online]. Available: <https://training.seer.cancer.gov/disease/categories/classification.html>.
- [58] M. M. Parker and al., "The Microenvironment in Cancer," in *Tumor Microenvironment*, John Wiley & Sons, 2010, pp. 1-6.
- [59] R. P. Hill and al., "Influence of Hypoxia on Metastatic Spread," in *Tumor Microenvironment*, John Wiley & Sons, 2010, pp. 311-328.
- [60] D. M. Noonan and al., "Invasion and Metastasis," in *The Tumor Microenvironment*, New York, NY: Springer New York, R. G. Bagley, 2010, pp. 213-228.
- [61] J. J. Christiansen and al., "Reassessing epithelial to mesenchymal transition as a prerequisite for carcinoma invasion and metastasis," *Cancer Research*, vol. 66, no. 17, pp. 8319-8326, 2006.
- [62] F. Prall, "Tumour budding in colorectal carcinoma," *Histopathology*, vol. 50, no. 1, pp. 151-162, 2007.
- [63] L. Li and al., "Epithelial-mesenchymal transition in human cancer: Comprehensive reprogramming of metabolism, epigenetics, and differentiation," *Pharmacology and Therapeutics*, vol. 150, pp. 33-46, 2015.
- [64] Y. Wu and al., "Epithelial-to-Mesenchymal Transition in Development and Diseases," in *The Tumor Microenvironment*, New York, NY: Springer New York, R. G. Bagley, 2010, pp. 187-211.
-

- [65] G. Berx and al., "Involvement of members of the cadherin superfamily in cancer," *Cold Spring Harbor Perspectives in Biology*, vol. 1, no. 6, 2009.
- [66] B. Hotz and al., "Epithelial to mesenchymal transition: Expression of the regulators snail, slug, and twist in pancreatic cancer," *Clinical Cancer research*, vol. 13, no. 16, pp. 4769-4776, 2007.
- [67] H. Pinheiro and al., "Allele-specific CDH1 downregulation and hereditary diffuse gastric cancer," *Human Molecular Genetics*, vol. 19, no. 5, pp. 943-952, 2010.
- [68] S. Hansford and al., "Hereditary diffuse gastric cancer syndrome: CDH1 mutations and beyond," *JAMA Oncology*, vol. 1, no. 1, pp. 23-32, 2015.
- [69] E. Lever and al., "Mechanisms and sequelae of E-cadherin silencing in hereditary diffuse gastric cancer," *journal of Pathology*, vol. 220, no. September Issue, pp. 114-125, 2010.
- [70] P. Carneiro and al., "E-cadherin dysfunction in gastric cancer - Cellular consequences, clinical applications and open questions," *FEBS Letter*, vol. 586, no. 18, pp. 2981-2989, 2012.
- [71] G. Li and al., "Downregulation of E-cadherin and Desmoglein 1 by autocrine hepatocyte growth factor during melanoma development," *Oncogene*, vol. 20, no. 56, pp. 8125-8135, 2001.
- [72] Gravdal, "A switch from E-cadherin to N-cadherin expression indicates epithelial to mesenchymal transition and is of strong and independent importance for the progress of prostate cancer," *Clinical Cancer Research*, vol. 13, no. 23, pp. 7003-7011, 2007.
- [73] M. J. G. B. Ussemakers and al., "Complex cadherin expression in human prostate cancer cells," *Cancer*, vol. 450, no. June 1999, pp. 446-450, 2000.
- [74] Y. Shintani and al., "ADH-1 suppresses N-cadherin-dependent pancreatic cancer progression," *International journal of Cancer*, vol. 122, no. 1, pp. 71-77, 2008.
- [75] M. Chaklader and al., "Differential remodeling of cadherins and intermediate cytoskeletal filaments influence microenvironment of solid and ascitic sarcoma," *Molecular Cell Biochemistry*, vol. 382, no. 1-2, pp. 293-306, 2013.
- [76] E. Braungart and al., "Functional loss of E-cadherin and cadherin-11 alleles on chromosome 16q22 in colonic cancer," *journal of Pathology*, vol. 187, no. 5, pp. 530-534, 1999.
- [77] K. Sundfeldt, "Cell-cell adhesion in the normal ovary and ovarian tumors of epithelial origin; an exception to the rule," *Molecular Cell Endocrinology*, vol. 202, no. 1, pp. 89-96, 2003.
-

- [78] P. J. Anson and al., "E-Cadherin Expression in Human Epithelial Ovarian Cancer," *International Journal of Cancer*, vol. 74, no. February Issue, pp. 275-280, 1997.
- [79] Naora, H. and al., "Ovarian cancer metastasis: Integrating insights from disparate model organisms," *Nature Reviews Cancer*, vol. 5, no. 5, pp. 355-366, 2005.
- [80] P. Querzoli and al., "An immunohistochemically positive E-cadherin status is not always predictive for a good prognosis in human breast cancer," *British Journal of Cancer*, vol. 103, no. 12, pp. 1835-1839, 2010.
- [81] C. G. Kleer and al., "Persistent E-cadherin expression in inflammatory breast cancer," *Modern Pathology*, vol. 14, no. 5, pp. 458-464, 2001.
- [82] K. Strumane and al., "Cadherins in cancer," *Handbook of Experimental Pharmacology*, vol. 165, no. 165, pp. 69-103, 2004.
- [83] M. J. Wheelock and al., "Cadherin switching," *Journal of Cell Science*, vol. 121, no. 6, pp. 727-735, 2008.
- [84] A. I. Minchinton and al., "Drug Penetration and Therapeutic Resistance," in *Tumor Microenvironment*, John Wiley & Sons, 2010, pp. 329-352.
- [85] H. M. R. and al., "Impact on Radiotherapy," in *Tumor Microenvironment*, John Wiley & Sons, 2010, pp. 353-376.
- [86] E. Devemy and al., "Identification of a novel N-cadherin antagonist," *Peptides*, vol. 29, no. 11, pp. 1853-1861, 2008.
- [87] H. Yoshioka and al., "In vivo therapeutic effect of CDH3/P-cadherin-Targeting radioimmunotherapy," *Cancer Immunology, Immunotherapy*, vol. 61, no. 8, pp. 1211-1220, 2012.
- [88] O. W. Blaschuk and al., "Identification of a cadherin cell adhesion recognition sequence," *Developmental Biology*, vol. 139, no. 1, pp. 227-229, 1990.
- [89] E. Williams and al., "A novel family of cyclic peptide antagonists suggests that N-cadherin specificity is determined by amino acids that flank the HAV motif," *Journal of Biological Chemistry*, vol. 275, no. 6, pp. 4007-4012, 2000.
- [90] E. Devemy and al., "Identification of a novel dual E- and N-cadherin antagonists," *Peptides*, vol. 30, no. 8, pp. 1539-1547, 2009.
- [91] S. Jothy and al., "Adhesion or anti-adhesion in cancer: what matters more?," *Cancer Metastasis review*, vol. 14, no. 4, pp. 363-376, 1995.
-

- [92] O. W. Blaschuk and al., "N-cadherin antagonists as oncology therapeutics," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, no. 1661, pp. 8-11, 2015.
- [93] K. M. Mrozik and al., "Therapeutic targeting of N-cadherin is an effective treatment for multiple myeloma," *british journal of Haematology*, vol. 171, no. 3, pp. 387-399, 2015.
- [94] O. W. Blaschuk and al., "Discovery and development of N-cadherin antagonists," *Cell and Tissue Research*, vol. 348, no. 2, pp. 309-313, 2012.
- [95] O. W. Blaschuk and al., "Cadherins as novel targets for anti-cancer therapy," *2009*, vol. 625, no. 1-3, pp. 195-198, 195-198.
- [96] A. Perotti and al., "Clinical and pharmacological phase I evaluation of Exherin (ADH-1), a selective anti-N-cadherin peptide in patients with N-cadherin-expressing solid tumours," *Annals of Oncology*, vol. 20, no. 4, pp. 741-745, 2009.
- [97] D. J. Stewart and al., "Final clinical and pharmacokinetic (PK) results from a phase 1 study of the novel N-cadherin (N-cad) antagonist, Exherin (ADH-1), in patients with refractory solid tumors stratified according to N-cad expression," *Journal of Clinical Oncology*, vol. 24, no. 18, p. 3016, 2006.
- [98] N. Yarom and al., "Phase I Clinical Trial of Exherin (ADH-1) in Patients with Advanced Solid Tumors," *Current Clinical Pharmacology*, vol. 8, no. 1, pp. 81-88, 2013.
- [99] G. M. Beasley and al., "Prospective multicenter phase II trial of systemic ADH-1 in combination with melphalan via isolated limb infusion in patients with advanced extremity melanoma," *Jornal of Clinical Oncology*, vol. 29, no. 9, pp. 1210-1215, 2011.
- [100] M. Eslami and al., "Deep analysis of N-cadherin/ADH-1 interaction: a computational survey," *Journal of Biomolecular Structure and Dynamics*, vol. 37, no. 1, pp. 210-228, 2019.
- [101] K. M. Mrozik and al., "LCRF-0006, a small molecule mimetic of the N-cadherin antagonist peptide ADH-1, synergistically increases multiple myeloma response to bortezomib," *federation of American Societies of Experimental Biology*, vol. 2, no. 6, pp. 339-353, 2020.
- [102] A. Senoo and al., "Inhibition of homophilic dimerization and disruption of cell adhesion by P-cadherin-specific small molecules from SPR-based assays," *Chemical Communications*, vol. 54, no. 42, pp. 5350-5353, 2018.
- [103] S. Kudo and al., "Disruption of cell adhesion by an antibody targeting the cell-adhesive intermediate (X-dimer) of human P-cadherin," *Nature Scientific Reports*, vol. 7, no. 39518, 2017.
-

- [104] Q. Sheng and al., "PCA062, a P-cadherin Targeting Antibody–Drug Conjugate, Displays Potent Antitumor Activity Against P-cadherin–expressing Malignancies," *Molecular Cancer Therapeutics*, no. June 3, 2021.
- [105] F. Doro and al., "Computational design of novel peptidomimetic inhibitors of cadherin homophilic interactions," *Organic and Biomolecular Chemistry*, vol. 13, no. 9, pp. 2570-2573, 2015.
- [106] A. Dalle Vedove and al., "Structure-Based Virtual Screening Allows the Identification of Efficient Modulators of E-Cadherin-Mediated Cell–Cell Adhesion," *International Journal of Molecular Sciences*, vol. 20, no. 14, p. Article N° 3404, 2019.
- [107] A. Vulpetti and al., "Design and NMR-Based Screening of LEF, a Library of Chemical Fragments with Different Local Environment of Fluorine," *Journal of American Chemical Society*, vol. 131, pp. 12949-12959, 2009.
- [108] C. Dalvit and al., "NMR methods in fragment screening theory and a comparison with other biophysical techniques," *Drug Discovery Today*, vol. 14, pp. 1051-1057, 2009.
- [109] C. Dalvit and al., "NMR-Based Quality Control Approach for the Identification of False Positives and False Negatives in High Throughput Screening," *Current Drug Discovery Technologies*, vol. 3, pp. 115-124, 2006.
- [110] H. Y. Carr and al., "Effects of diffusion on free precession in nuclear magnetic resonance experiments," *Physical review*, vol. 94, pp. 630-638, 1954.
- [111] T. Battye and al., "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM," *Acta Cryst.*, vol. D67, pp. 271-281, 2011.
- [112] M. Winn and al., "Overview of the CCP4 suite and current developments," *Acta Cryst.*, vol. D67, pp. 235-242, 2011.
- [113] P. Adams and al., "PHENIX: a comprehensive Python-based system for macromolecular structure solution," *Acta Cryst.*, vol. D66, pp. 213-221, 2010.
- [114] P. Emsley and al., "Features and development of Coot," *Acta Cryst.*, vol. D66, pp. 486-501, 2010.
- [115] E. Pettersen and al., "UCSF Chimera - A visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605-1612, 2004.
- [116] X. Robert and al., "Deciphering key features in protein structures with the new ENDscript server," *Nucleic Acids Research*, vol. 42, no. W1, pp. W320-W324, 2014.
-

- [117] C. Siret and al., "Cadherin-1 and cadherin-3 cooperation determines the aggressiveness of pancreatic ductal adenocarcinoma," *British Journal of Cancer*, vol. 118, no. 4, pp. 546-557, 2018.
- [118] C. A. Schneider and al., "NIH Image to ImageJ: 25 years of image analysis," *Nature Methods*, vol. 9, pp. 671-675, 2012.
- [119] S. K. Hanks and al., "Genomic analysis of the eukaryotic protein kinase superfamily: A perspective," *Genome Biology*, vol. 4, no. 111, 2003.
- [120] C. J. Leonard and al., "Novel families of putative protein kinases in bacteria and archaea: Evolution of the "eukaryotic" protein kinase superfamily," *Genome Research*, vol. 8, pp. 1036-1047, 1998.
- [121] P. J. Kennelly, "Protein kinases and protein phosphatases in prokaryotes: A genomic perspective," *Federation of European Microbiological Societies Microbiology Letters*, vol. 206, no. January Issue, pp. 1-8, 2002.
- [122] G. Manning and al., "The protein kinase complement of the human genome.," *Science*, vol. 298, no. 5600, pp. 1912-1934, 2002.
- [123] P. Cohen and al., "Evolution of protein kinase signaling from yeast to man," *Trends in Biochemical Sciences*, vol. 27, no. 10, pp. 514-520, 2002.
- [124] P. J. Kennelly, "Archaeal protein kinases and protein phosphatases: Insights from genomics and biochemistry," *Biochemical Journal*, vol. 370, no. 3, pp. 373-389, 2003.
- [125] S. Hanks and al., "The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification," *The FASEB Journal*, vol. 9, no. 8, pp. 576-596, 1995.
- [126] D. Knighton and al., "Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase," *Science*, vol. 253, p. 407-414, 1991.
- [127] S. S. Taylor and al., "Structural Framework for the Protein Kinase Family," *Annual Reviews of Cell Biology*, vol. 8, pp. 429-462, 1992.
- [128] S. S. Taylor and al., "Three protein kinase structures define a common motif," *Structure*, vol. 2, pp. 345-355, 1994.
- [129] V. Rao and al., "Structure of Type II β Phosphatidylinositol Phosphate Kinase: A Protein Kinase Fold Flattened for Interfacial Phosphorylation," *Cell*, vol. 94, pp. 829-839, 1998.
-

- [130] S. Steinbacher and al., "The crystal structure of the Physarum polycephalum actin – fragmin kinase : an atypical protein kinase with a specialized substrate-binding domain," *The European Molecular Biology Organization Journal*, vol. 18, no. 11, pp. 2923-2929, 1999.
- [131] H. Yamaguchi and al., "Crystal Structure of the Atypical Protein Kinase Domain of a TRP Channel with Phosphotransferase Activity," *Molecular Cell*, vol. 7, pp. 1047-1057, 2001.
- [132] D. Burk and al., "Structural Analyses of Nucleotide Binding to an Aminoglycoside," *Biochemistry*, vol. 40, pp. 8756-8764, 2001.
- [133] al. and B. Kobe, "Principles of Kinase Regulation," in *Handbook of Cell Signaling*, London, Academic Press, 2003, pp. 539-542.
- [134] J. Sowadski and al., "Conformational Diversity of Catalytic Cores of Protein Kinases," *Pharmacology & Therapeutics*, vol. 82, no. 94, pp. 157-164, 1999.
- [135] J. Wittenberg and al., "Choline Phosphokinase," *Journal of Biological Chemistry*, vol. 202, no. 1, pp. 431-444, 1953.
- [136] T. J. Porter and al., "Purification and characterization of choline/ethanolamine kinase from rat liver," *Journal of Biological Chemistry*, vol. 265, no. 1, pp. 414-422, 1990.
- [137] T. Uchida, "Immunologically and enzymatically distinct rat choline kinase isozymes," *Journal of Biochemistry*, vol. 116, no. 6, pp. 1241-1250, 1994.
- [138] K. Ishidate and al., "Complete co-purification of choline kinase and ethanolamine kinase from rat kidney and immunological evidence for both kinase activities residing on the same enzyme protein(s) in rat tissues," *Biochimica and Biophysica Acta*, vol. 836, no. 1, pp. 119-124, 1985.
- [139] S. Yamashita and al., "Choline kinase from yeast," *Biochimica and Biophysica Acta*, vol. 1348, no. 1-2, pp. 63-69, 1997.
- [140] K. H. Kim and al., "Expression, purification, and characterization of choline kinase, product of the CKI gene from *Saccharomyces cerevisiae*," *Journal of Biological Chemistry*, vol. 273, no. 12, pp. 6844-6852, 1998.
- [141] C. Aoyama and al., "Molecular cloning of mouse choline kinase and choline/ethanolamine kinase: their sequence comparison to the respective rat homologs," *Biochimica and Biophysica Acta*, vol. 1393, no. 1, pp. 179-185, 1998.
- [142] C. Aoyama and al., "Structure and characterization of the genes for murine choline/ethanolamine kinase isozymes alpha and beta," *Journal of Lipid Research*, vol. 41, no. 3, pp. 452-464, 2000.
-

- [143] T. Uchida and al., "Molecular cloning, characterization, and expression in *Escherichia coli* of a cDNA encoding mammalian choline kinase," *Journal of Biological Chemistry*, vol. 267, no. 14, pp. 10156-10162, 1992.
- [144] A. Al-Malki and al., "Pea choline kinase: purification, properties and isolation of a cDNA," *Biochemical Society Transactions*, vol. 28, no. 6, pp. 721-723, 2000.
- [145] D. Peisach and al., "The crystal structure of choline kinase reveals a eukaryotic protein kinase fold," *Structure*, vol. 11, no. 6, pp. 703-713, 2003.
- [146] P. Gee and al., "Multiple isoforms of choline kinase from *Caenorhabditis elegans*: Cloning, expression, purification, and characterization," *Biochimica and Biophysica Acta*, vol. 1648, no. 1-2, pp. 33-42, 2003.
- [147] E. D. Scheef and al., "Structural evolution of the protein kinase-like superfamily," *PLoS Computational Biology*, vol. 1, p. e49, 2005.
- [148] E. Malito and al., "Elucidation of human choline kinase crystal structures in complex with the products ADP or phosphocholine," *Journal of Molecular Biology*, vol. 364, no. 2, pp. 136-151, 2007.
- [149] K. Kinoshita and al., "Structural motif of phosphate-binding site common to various protein superfamilies: All-against-all structural comparison of protein-mononucleotide complexes," *Protein Engineering*, vol. 12, no. 1, pp. 11-14, 1999.
- [150] E. D. Scheeff and al., "Structural Evolution of the Protein Kinase-Like Superfamily," *PLoS Computational Biology*, vol. 1, no. 5, pp. 359-381, 2005.
- [151] C. Hudson and al., "Kinetic and mechanistic characterisation of Choline Kinase- α ," *Biochimica and Biophysica Acta*, vol. 1834, no. 6, pp. 1107-1116, 2013.
- [152] S. Whitehouse and al., "Studies on the kinetic mechanism of the catalytic subunit of the cAMP-dependent protein kinase," *Journal of Biological Chemistry*, vol. 258, pp. 3693-3701, 1983.
- [153] C. Erneux and al., "The kinetics of tyrosine phosphorylation by the purified epidermal growth factor receptor kinase of A-431 cells," *Journal of Biological Chemistry*, vol. 258, pp. 4137-4142, 1983.
- [154] I. Posner and al., "Kinetic model of the epidermal growth factor (EGF) receptor tyrosine kinase and a possible mechanism of its activation by EGF," *Journal of Biological Chemistry*, vol. 267, pp. 20638-20647, 1992.
-

- [155] P. V. LoGrasso and al., "Kinetic mechanism for p38 MAP kinase," *Biochemistry*, vol. 36, pp. 10422-10427, 1997.
- [156] T. W. Wong and al., "Goldberg, Kinetics and mechanism of angiotensin phosphorylation by the transforming gene product of Rous sarcoma virus," *Journal of Biological Chemistry*, vol. 259, pp. 3127-3131, 1984.
- [157] D. Pilloff and al., "The kinetic mechanism of phosphomevalonate kinase," *Journal of Biological Chemistry*, vol. 278, pp. 4510-4515, 2003.
- [158] X. Zhang and al., "Kinetic mechanism of AKT/PKB enzyme family," *Journal of Biological Chemistry*, vol. 281, pp. 13949-13956, 2006.
- [159] C. Satishchandran and al., "Adenosine-5'-phosphosulphate kinase from *Escherichia coli* K12. Purification, characterization, and identification of a phosphorylated enzyme intermediate," *Journal of Biochemical Sciences*, vol. 264, pp. 15012-15021, 1989.
- [160] S. N. Kochetkov and al., "Studies on the mechanism of action of histone kinase dependent on adenosine 3':5'-monophosphate. Evidence for involvement of histidine and lysine residues in the phosphotransferase reaction," *European Journal of Biochemistry*, vol. 81, pp. 111-118, 1977.
- [161] S. Morera and al., "Mechanism of phosphate transfer by nucleoside diphosphate kinase: x-ray structures of the phosphohistidine intermediate of the enzymes from *Drosophila* and *Dictyostelium*," *Biochemistry*, vol. 34, pp. 11062-11070, 1995.
- [162] "Symbol report for CHKA," [Online]. Available: https://www.genenames.org/data/gene-symbol-report/#!/hgnc_id/HGNC:1937.
- [163] "Symbol report for CHKB," [Online]. Available: https://www.genenames.org/data/gene-symbol-report/#!/hgnc_id/HGNC:1938.
- [164] "UniProtKB - P35790 (CHKA_HUMAN)," [Online]. Available: <https://www.uniprot.org/uniprot/P35790>.
- [165] "<https://www.uniprot.org/uniprot/Q9Y259>," [Online]. Available: <https://www.uniprot.org/uniprot/Q9Y259>.
- [166] C. Aoyama and al., "Structure and function of choline kinase isoforms in mammalian cells," *Progress in Lipid Research*, vol. 43, no. 3, pp. 266-281, 2004.
-

- [167] D. Gallego-Ortega and al., "Differential role of human choline kinase α and β enzymes in lipid metabolism: Implications in cancer onset and treatment," *PLoS One*, vol. 4, no. 11, p. e7819, 2009.
- [168] "2.7.1.32: choline kinase," [Online]. Available: https://www.brenda-enzymes.org/all_enzymes.php?ecno=2.7.1.32&table=KCat_KM_Value#TAB.
- [169] E. P. Kennedy and al., "The function of cytidine coenzymes in the biosynthesis of phospholipides," *Journal of Biological Chemistry*, vol. 222, no. 1, pp. 193-214, 1956.
- [170] F. Gibellini and al., "The Kennedy pathway-de novo synthesis of phosphatidylethanolamine and phosphatidylcholine," *IUBMB Life*, vol. 62, no. 6, pp. 414-428, 2010.
- [171] S. Pelech and al., "Regulation of Phosphatidylcholine Biosynthesis," *Biochimica et Biophysica Acta*, vol. 779, no. 779, p. 217*251, 1984.
- [172] R. W. Guynn, "Equilibrium constants under physiological conditions for the reactions of choline kinase and the hydrolysis of phosphorylcholine to choline and inorganic phosphate," *Journal of biological Chemistry*, vol. 251, no. 22, pp. 7162-7167, 1976.
- [173] R. Sundler and al., "Regulation of phospholipid biosynthesis in isolated rat hepatocytes," *Journal of Biological Chemistry*, vol. 250, pp. 3369-3387, 1975.
- [174] R. H. Hjelmstad and al., "Chimeric enzymes structure-function analysis of segments of sn-1,2-diacylglycerol choline- and ethanolaminephosphotransferases," *Journal of Biological Chemistry*, vol. 269, pp. 20995-21002, 1994.
- [175] A. Henneberry and al., "Cloning and expression of a human choline/ethanolaminephosphotransferase: synthesis of phosphatidylcholine and phosphatidylethanolamine," *Biochemistry*, vol. 339, pp. 291-298, 1999.
- [176] A. Henneberry and al., "Cloning, genomic organization, and biochemical characterization of a human cholinephosphotransferase," *Journal of Biological Chemistry*, vol. 275, pp. 29808-29815, 2000.
- [177] D. Vance and al., "Hepatic phosphatidylethanolamine N-methyltransferase, unexpected roles in animal biochemistry and physiology," *Journal of Biological Chemistry*, vol. 282, pp. 33237-33241, 2007.
- [178] W. Dowhan and al., "Functional Roles of Lipids in Membranes," in *Biochemistry of Lipids , Lipoproteins and Membranes*, Amsterdam, Oxford, Waltham, Elsevier, 2016, pp. 1-40.
-

- [179] N. D. Ridgway, "Phospholipid Synthesis in Mammalian Cells," in *Biochemistry of Lipids , Lipoproteins and Membranes*, Amsterdam, Oxford, Waltham, Elsevier, 2016, pp. 209-236.
- [180] J. H. Exton, "Phosphatidylcholine breakdown and signal transduction," *Biochimica and Biophysica Acta*, vol. 1212, no. 1, pp. 26-42, 1994.
- [181] M. M. Billah and al., "The regulation and cellular functions of phosphatidylcholine hydrolysis," *The Biochemical Journal*, vol. 269, no. 2, pp. 281-291, 1990.
- [182] Z. Kiss, "Effects of phorbol ester on phospholipid metabolism," *Progress in Lipid Research*, vol. 29, no. 3, pp. 141-166, 1990.
- [183] K. Glunde and al., "Choline metabolism in malignant transformation," *Nature Reviews Cancer*, vol. 11, no. December 2011, pp. 835-848, 2011.
- [184] T. R. DeGrado and al., "Pharmacokinetics and radiation dosimetry of 18F-fluorocholine," *Journal of Nuclear Medicine*, vol. 43, no. 1, pp. 92-96, 2002.
- [185] T. H. Witney and al., "Evaluation of deuterated 18F- and 11C-labeled choline analogs for cancer detection by positron emission tomography," *Clinical Cancer Research*, vol. 18, no. 4, pp. 1063-1072, 2012.
- [186] S. A. Kwee and al., "Cancer imaging with fluorine-18-labeled choline derivatives," *Seminars in Nuclear Medicine*, vol. 37, no. 6, pp. 420-428, 2007.
- [187] M. Li and al., "Value of 11C-choline PET/CT for lung cancer diagnosis and the relation between choline metabolism and proliferation of cancer cells," *Oncology Reports*, vol. 29, no. 1, pp. 205-211, 2013.
- [188] R. Katz-Brull and al., "Kinetics of choline transport and phosphorylation in human breast cancer cells; NMR application of the zero trans method," *Anticancer Research*, vol. 16, no. 3B, pp. 1375-1380, 1996.
- [189] G. Eliyahu and al., "Phosphocholine as a biomarker of breast cancer: molecular and biochemical studies," *International Journal of Cancer*, vol. 120, no. 8, pp. 1721-1730, 2007.
- [190] K. Glunde and al., "Molecular causes of the aberrant choline phospholipid metabolism in breast cancer," *Cancer Research*, vol. 64, no. 12, pp. 4270-4276, 2004.
- [191] E. Iorio and al., "Activation of phosphatidylcholine cycle enzymes in human epithelial ovarian cancer cells," *Cancer Research*, vol. 70, no. 5, pp. 2126-2135, 2010.
- [192] E. Iorio and al., "Alterations of choline phospholipid metabolism in ovarian tumor progression," *Cancer Research*, vol. 70, no. 5, pp. 9369-9376, 2005.
-

- [193] D. Y. Noh and al., "Overexpression of phospholipase D1 in human breast cancer tissues," *Cancer Letters*, vol. 161, no. 2, pp. 207-214, 2000.
- [194] A. Ramirez de Molina and al., "Overexpression of choline," *Biochemical and Biophysical Research Communications*, vol. 296, no. 3, pp. 580-583, 2002.
- [195] K. Glunde and al., "RNA interference-mediated choline kinase suppression in breast cancer cells induces differentiation and reduces proliferation," *Cancer Research*, vol. 65, no. 23, pp. 11034-11043, 2005.
- [196] A. Ramirez de Molina and al., "Expression of choline kinase alpha to predict outcome in patients with early-stage non-small-cell lung cancer: a retrospective study," *The Lancet, Oncology*, vol. 8, no. 10, pp. 889-897, 2007.
- [197] A. Ramirez de Molina and al., "Increased choline kinase activity in human breast carcinomas: clinical evidence for a potential novel antitumor strategy," *Oncogene*, vol. 64, no. 12, pp. 4317-4322, 2002.
- [198] A. Ramirez de Molina and al., "Choline Kinase Is a Novel Oncogene that Potentiates RhoA-Induced Carcinogenesis," *Cancer Research*, vol. 65, no. 13, pp. 5647-5653, 2005.
- [199] M. Noriko and al., "Choline kinase- α protein and phosphatidylcholine but not phosphocholine are required for breast cancer cell survival," *NMR in Biomedicine*, vol. 28, no. 12, pp. 1697-1706, 2015.
- [200] E. Iorio and al., "Key Players in Choline Metabolic Reprogramming in Triple-Negative Breast Cancer," *Frontiers of Oncology*, vol. 6, no. 205, 2016.
- [201] T. Shah and al., "Choline kinase overexpression increases invasiveness and drug resistance of human breast cancer cells," *NMR in Biomedicine*, vol. 23, no. 6, pp. 633-642, 2010.
- [202] S. P. Arluckas and al., "Near infrared fluorescent imaging of choline kinase alpha expression and inhibition in breast tumors," *Oncotarget*, vol. 8, no. 10, pp. 16518-16530, 2017.
- [203] C. Huang and al., "Activation of choline kinase by extracellular Ca^{2+} is Ca^{2+} -sensing receptor, Galph α 12 and Rho-dependent in breast cancer cells," *Cell Signalling*, vol. 21, no. 12, pp. 1894-1900, 2009.
- [204] G. Baio and al., "Correlation between Choline Peak at MR Spectroscopy and Calcium-Sensing Receptor Expression Level in Breast Cancer: A Preliminary Clinical Study," *Molecular Imaging and Biology*, vol. 17, no. 4, pp. 548-556, 2015.
-

- [205] Z. Huang and al., "Use of ^{11}C -Choline positron emission tomography/computed tomography to investigate the mechanism of choline metabolism in lung cancer," *Molecular Medicine Reports*, vol. 11, no. 5, pp. 3285-3290, 2015.
- [206] M. Inazu and al., "Functional expression of choline transporter-like protein 1 (CTL1) in small cell lung carcinoma cells: a target molecule for lung cancer therapy," *Pharmacological Research*, vol. 76, pp. 119-131, 2013.
- [207] K. Nakagami and al., "Increased choline kinase activity in 1,2-dimethylhydrazine-induced rat colon cancer," *Japanese Journal of Cancer Research*, vol. 90, no. 11, pp. 121-1217, 1999.
- [208] K. Nakagami and al., "Increased choline kinase activity and elevated phosphocholine levels in human colon cancer," *Japanese Journal of Cancer Research*, vol. 90, no. 4, pp. 419-424, 1999.
- [209] I. Hu and al., "Overexpression of CHKA contributes to tumor progression and metastasis and predicts poor prognosis in colorectal carcinoma," *Oncotarget*, vol. 7, no. 41, pp. 66660-66678, 2016.
- [210] M. Bagnoli and al., "Choline Metabolism Alteration: A Focus on Ovarian Cancer," *Frontiers of Oncology*, vol. 6, no. 153, 2016.
- [211] A. Rizzo and al., "Choline kinase alpha impairment overcomes TRAIL resistance in ovarian cancer cells," *Journal of Experimental and Clinical Cancer Research*, vol. 40, no. 1, 2021.
- [212] A. Granata and al., "Choline kinase-alpha by regulating cell aggressiveness and drug sensitivity is a potential druggable target for ovarian cancer," *British Journal of Cancer*, vol. 110, no. 2, pp. 330-340, 2014.
- [213] S. Trousil and al., "Alterations of choline phospholipid metabolism in endometrial cancer are caused by choline kinase alpha overexpression and a hyperactivated deacylation pathway," *Cancer Research*, vol. 74, no. 23, pp. 6867-6877, 2014.
- [214] J. M. Mazarico and al., "Choline Kinase Alpha (CHK α) as a Therapeutic Target in Pancreatic Ductal Adenocarcinoma: Expression, Predictive Value, and Sensitivity to Inhibitors," *Molecular Cancer Therapeutics*, vol. 15, no. 2, pp. 323-333, 2016.
- [215] T. Shah and al., "Molecular causes of elevated phosphoethanolamine in breast and pancreatic cancer cells," *NMR in Biomedicine*, vol. 31, no. 8, p. e3936, 2018.
- [216] M. Asim and al., "Choline Kinase Alpha as an Androgen Receptor Chaperone and Prostate Cancer Therapeutic Target," *Journal of the National Cancer Institute*, vol. 108, no. 5, 2015.
-

- [217] A. Challapalli and al., "Exploiting altered patterns of choline kinase-alpha expression on human prostate tissue to prognosticate prostate cancer," *Journal of Clinical Pathology*, vol. 68, no. 9, pp. 703-709, 2015.
- [218] C. Huang and al., "Role of p115RhoGEF in the regulation of extracellular Ca(2+)-induced choline kinase activation and prostate cancer cell proliferation," *International Journal of Cancer*, vol. 128, no. 12, pp. 2833-2842, 2011.
- [219] K. Contractor and al., "Use of [11C]choline PET-CT as a noninvasive method for detecting pelvic lymph node status from prostate cancer and relationship with choline kinase expression," *Clinical Cancer Research*, vol. 17, no. 24, pp. 7673-7683, 2011.
- [220] K. Glunde and al., "Hypoxia regulates choline kinase expression through hypoxia-inducible factor-1 alpha signaling in a human prostate cancer model," *Cancer Research*, vol. 68, no. 1, pp. 172-180, 2008.
- [221] M. Asim and al., "Kinase joins the chaperone club: Androgen-regulated kinome reveals choline kinase alpha as a potential drug target in prostate cancer," *Molecular Cell Oncology*, vol. 3, no. 3, 2016.
- [222] E. Mariotto and al., "Choline Kinase Alpha Inhibition by EB-3D Triggers Cellular Senescence, Reduces Tumor Growth and Metastatic Dissemination in Breast Cancer," *Cancers*, vol. 10, no. 10, p. 391, 2018.
- [223] H. S. Kim and al., "Downregulation of Choline Kinase-Alpha Enhances Autophagy in Tamoxifen-Resistant Breast Cancer Cells," *PLoS One*, vol. 10, no. 10, 2015.
- [224] P. Pftizinger and al., "Indirect cholinergic activation slows down pancreatic cancer growth and tumor-associated inflammation," *Journal of Experimental and Clinical Cancer Research*, vol. 39, no. 1, 2020.
- [225] N. Mori and al., "Choline kinase down-regulation increases the effect of 5-fluorouracil in breast cancer cells," *Cancer Research*, vol. 67, no. 23, pp. 11284-11290, 2007.
- [226] A. de la Cueva and al., "Combined 5-FU and ChoK α inhibitors as a new alternative therapy of colorectal cancer: evidence in human tumor-derived cell lines and mouse xenografts," *PLoS One*, vol. 8, no. 6, 2013.
- [227] D. Gallego-Ortega and al., "Involvement of human choline kinase alpha and beta in carcinogenesis: a different role in lipid metabolism and biological functions," *Advances of Enzyme Regulation*, vol. 51, pp. 183-194, 2011.
-

- [228] S. C. Falcon and al., "A non-catalytic role of choline kinase alpha is important in promoting cancer cell survival," *Oncogenesis*, vol. 3, no. 2, p. e38, 2013.
- [229] X. Chen and al., "Molecular structure and differential function of choline kinases CHKA and CHKB in musculoskeletal system and cancer," *Cytokine & Growth Factor Reviews*, vol. 33, no. February 2017, pp. 65-72, 2017.
- [230] R. Sher and al., "A Rostrocaudal Muscular Dystrophy Caused by a Defect in Choline Kinase Beta, the First Enzyme in Phosphatidylcholine Biosynthesis," *Journal of Biological Chemistry*, vol. 281, no. 8, pp. 4938-4948, 2006.
- [231] G. Wu and al., "Understanding the muscular dystrophy caused by deletion of choline kinase beta in mice," *Biochimica and Biophysica Acta*, vol. 1791, pp. 347-356, 2009.
- [232] P. Gutierrez Rios and al., "Congenital megaconial myopathy due to a novel defect in the choline kinase Beta gene," *Archives of Neurology*, vol. 69, no. 5, pp. 657-661, 2012.
- [233] R. Quinlivan and al., "Muscular dystrophy with large mitochondria associated with mutations in the CHKB gene in three British patients: extending the clinical and pathological phenotype," *Neuromuscular Disorders*, vol. 23, no. 7, pp. 546-556, 2013.
- [234] M. Castro-Gago and al., "Exome sequencing identifies a CHKB mutation in Spanish patient with megaconial congenital muscular dystrophy and mtDNA depletion," *European Journal of Pediatric Neurology*, vol. 18, no. 6, pp. 796-800, 2014.
- [235] G. Haliloglu and al., "Clinical characteristics of megaconial congenital muscular dystrophy due to choline kinase beta gene defects in a series of 15 patients," *Journal of Inherited Metabolic Diseases*, vol. 38, no. 6, pp. 1099-1108, 2015.
- [236] J. Oliveira and al., "New splicing mutation in the choline kinase beta (CHKB) gene causing a muscular dystrophy detected by whole-exome sequencing," *Journal of Human Genetics*, vol. 60, no. 6, pp. 305-312, 2015.
- [237] M. Cabrera-Serrano and al., "New splicing mutation in the choline kinase beta (CHKB) gene causing a muscular dystrophy detected by whole-exome sequencing," *Muscle & Nerve*, vol. 51, no. 1, pp. 140-143, 2015.
- [238] L. Brady and al., "Proximal myopathy with focal depletion of mitochondria and megaconial congenital muscular dystrophy are allelic conditions caused by mutations in CHKB," *Neuromuscular Disorders*, vol. 26, no. 2, pp. 160-164, 2016.
-

- [239] M. Castro-Gago and al., "Congenital neurogenic muscular atrophy in megaconial myopathy due to a mutation in CHKB gene," *Brain Development*, vol. 38, no. 1, pp. 167-172, 2016.
- [240] C. P. De Fuenmayor-Fernández De La Hoz and al., "A milder phenotype of megaconial congenital muscular dystrophy due to a novel CHKB mutation," *Muscle & Nerve*, vol. 54, no. 4, pp. 806-808, 2016.
- [241] S. Marchet and al., "Alteration of mitochondrial membrane inner potential in three Italian patients with megaconial congenital muscular dystrophy carrying new mutations in CHKB gene," *Mitochondrion*, vol. July 2019, no. 47, pp. 24-29, 2019.
- [242] S. H. S. Chan and al., "Megaconial congenital muscular dystrophy: Same novel homozygous mutation in CHKB gene in two unrelated Chinese patients," *Neuromuscular Disorders*, vol. 30, no. 1, pp. 47-53, 2020.
- [243] T. Miyagawa and al., "Variant between CPT1B and CHKB associated with susceptibility to narcolepsy," *Nature Genetics*, vol. 69, pp. 657-661, 2008.
- [244] Z. Li and al., "Choline kinase beta is required for normal endochondral bone formation," *Biochimica and Biophysica Acta*, vol. 1840, pp. 2112-2122, 2014.
- [245] J. Kular and al., "Choline kinase beta mutant mice exhibit reduced phosphocholine, elevated osteoclast activity, and low bone mass," *Journal of Biological Chemistry*, vol. 290, pp. 1729-1742, 2015.
- [246] R. Giménez and al., "Cytidine diphosphate choline administration activates brain cytidine triphosphate: phosphocholine cytidylyltransferase in aged rats," *Neuroscience Letters*, vol. 273, no. 3, pp. 163-166, 1999.
- [247] H. González-Pacheco and al., "Pre-conditioning with CDP-choline attenuates oxidative stress-induced cardiac myocyte death in a hypoxia/reperfusion model," *The Scientific World Journal*, vol. January, no. 21, p. 187071, 2014.
- [248] F. Quartieri and al., "Identification of Unprecedented ATP-competitive Choline Kinase Inhibitors," *Bioorganic & Medicinal Chemistry Letters*, vol. 2021, no. 51, p. e128310, 2021.
- [249] S. Zech and al., "Novel Small Molecule Inhibitors of Choline Kinase Identified by Fragment-Based Drug Discovery," *Journal of Medicinal Chemistry*, vol. 59, no. 2, pp. 671-686, 2016.
- [250] S. Trouzil and al., "The novel choline kinase inhibitor ICL-CCIC-0019 reprograms cellular metabolism and inhibits cancer cell growth," *Oncotarget*, vol. 7, no. 24, pp. 37103-37120, 2016.
-

- [251] S. Spanner and al., "Choline Metabolism in Discrete Areas of the Brain: the Effect of Hemicholinium-3," *Biochemical Society Transactions*, vol. 1, no. 2, pp. 475-477, 1973.
- [252] S. S. G. Ferguson and al., "Stereoselectivity of the Inhibition of [3H]Hemicholinium-3 Binding to the Sodium-Dependent High-Affinity Choline Transporter by the Enantiomers of α - and β -Methylcholine," *Journal of Neurochemistry*, vol. 62, no. 4, pp. 1449-1457, 1994.
- [253] S. O'Regan, "Binding of [3H]Hemicholinium-3 to the High-Affinity Choline Transporter in Electric Organ Synaptosomal Membranes," *Journal of Neurochemistry*, vol. 51, no. 6, pp. 1682-1688, 1988.
- [254] S. Apparsundaram and al., "Molecular cloning and characterization of a murine hemicholinium-3-sensitive choline transporter," *Biochemical Society Transactions*, vol. 29, no. 6, p. 711, 2001.
- [255] J. Cannon, "Structureactivity aspects of hemicholinium3 (HC-3) and its analogs and congeners," *Medicinal Research Reviews*, vol. 14, no. 5, pp. 505-531, 1994.
- [256] J. M. Campos and al., "QSAR of 1,10-(1,2-ethylenebisbenzyl)bis(4-substitutedpyridinium) Dibromides as Choline Kinase Inhibitors: a Different Approach for Antiproliferative Drug Design," *Bioorganic & Medicinal Chemistry Letters*, vol. 10, pp. 767-770, 2000.
- [257] R. Hernández-Alcoceba and al., "In Vivo Antitumor Activity of Choline Kinase Inhibitors: A Novel Target for Anticancer Drug Discovery," *Cancer Research*, vol. 59, no. July 1, pp. 3112-3118, 1999.
- [258] B. Hong and al., "Crystal structures of human choline kinase isoforms in complex with hemicholinium-3: Single amino acid near the active site influences inhibitor sensitivity," *Journal of Biological Chemistry*, vol. 285, no. 21, pp. 16330-16340, 2010.
- [259] R. Hernández-Alcoceba and al., "Choline kinase inhibitors as a novel approach for antiproliferative drug design," *Oncogene*, vol. 15, pp. 2289-2301, 1997.
- [260] J. M. Campos and al., "Quantitative Structure – Activity Relationships for a Series of Symmetrical Bisquaternary Anticancer Compounds," *bioorganic & Medicinal Chemistry*, vol. 10, no. 2, pp. 2215-2231, 2002.
- [261] J. M. Campos and al., "QSAR-Derived Choline Kinase Inhibitors: How Rational can Antiproliferative Drug Design Be?," *Current Medicinal Chemistry*, vol. 10, no. 13, pp. 1095-1112, 2005.
-

- [262] A. Conejo-García and al., "Influence of the Linker in Bispyridium Compounds on the Inhibition of Human Choline Kinase," *Journal of Medicinal Chemistry*, vol. 47, pp. 5433-5440, 2004.
- [263] A. Conejo-García and al., "Choline kinase inhibitory effect and antiproliferative activity of new 1,1',1''-(benzene-1,3,5-triylmethylene)tris[4-[(disubstituted)amino]pyridinium]," *European Journal of Medicinal Chemistry*, vol. 38, no. 1, pp. 109-116, 2003.
- [264] A. Conejo-García and al., "Bispyridinium cyclophanes: novel templates for human choline kinase inhibitors," *Journal of Medicinal Chemistry*, vol. 46, no. 17, pp. 3754-3757, 2003.
- [265] R. Hernández-Alcoceba and al., "In vivo antitumor activity of choline kinase inhibitors: a novel target for anticancer drug discovery," *Cancer Research*, vol. 59, no. 13, pp. 3112-3118, 1999.
- [266] A. Rodríguez-González and al., "Inhibition of choline kinase renders a highly selective cytotoxic effect in tumour cells through a mitochondrial independent mechanism," *International Journal of Oncology*, vol. 26, no. 4, pp. 999-1008, 2005.
- [267] A. Rodríguez-González and al., "Choline kinase inhibition induces the increase in ceramides resulting in a highly specific and selective cytotoxic antitumoral strategy as a potential mechanism of action," *Oncogene*, vol. 23, no. 50, pp. 8247-8259, 2004.
- [268] J. C. Lacal and al., "Preclinical Characterization of RSM-932A, a Novel Anticancer Drug Targeting the Human Choline Kinase Alpha, an Enzyme Involved in Increased Lipid Metabolism of Cancer Cells," *Molecular Cancer Therapeutics*, vol. 14, no. 1, pp. 31-39, 2015.
- [269] "Study of Intravenous TCD-717 in Patients With Advanced Solid Tumors," *ClinicalTrials.gov Identifier: NCT01215864*, p. Available from: <http://clinicaltrials.gov/ct2/show/NCT01215864>.
- [270] S. Kall and al., "Identification of a Unique Inhibitor-Binding Site on Choline Kinase α ," *Biochemistry*, vol. 57, no. 8, pp. 1316-1325, 2018.
- [271] B. Rubio-Ruíz and al., "Design, synthesis, theoretical calculations and biological evaluation of new non-symmetrical choline kinase inhibitors," *European Journal of Medicinal Chemistry*, vol. 50, pp. 154-162, 2012.
- [272] M. Sahún-Roncero and al., "The mechanism of allosteric coupling in choline kinase α 1 revealed by the action of a rationally designed inhibitor," *Angewandte Chemie*, vol. 52, no. 17, pp. 4852-4586, 2013.
-

- [273] M. Sahún-Roncero and al., "Determination of potential scaffolds for human choline kinase $\alpha 1$ by chemical deconvolution studies," *ChemBioChem*, vol. 14, no. 11, pp. 1291-1295, 2013.
- [274] B. Rubio-Ruiz and al., "Discovery of a new binding site on human choline kinase $\alpha 1$: Design, synthesis, crystallographic studies, and biological evaluation of asymmetrical bispyridinium derivatives," *Journal of Medicinal Chemistry*, vol. 57, no. 2, pp. 507-515, 2014.
- [275] S. Schiaffino-Ortega and al., "Design, synthesis, crystallization and biological evaluation of new symmetrical biscationic compounds as selective inhibitors of human Choline Kinase $\alpha 1$ (ChoK $\alpha 1$)," *Scientific Reports*, vol. 6, no. March, pp. 1-17, 2016.
- [276] E. Mariotto and al., "EB-3D a novel choline kinase inhibitor induces deregulation of the AMPK-mTOR pathway and apoptosis in leukemia T-cells," *Biochemical Pharmacology*, vol. 155, no. July, pp. 213-223, 2018.
- [277] WHO, World Malaria Report 2019, Geneva, Switzerland,: World Health Organization, 2019.
- [278] T. Mitamura and al., "Lipid metabolism in Plasmodium falciparum-infected erythrocytes: Possible new targets for malaria chemotherapy," *Microbes Infections*, vol. 5, pp. 545-552, 2003.
- [279] WHO, Guidelines for the Treatment of Malaria, Geneva, Switzerland: World Health Organization: Geneva,, 2015.
- [280] X. Z. Su and al., "Plasmodium genomics and genetics: New insights into malaria pathogenesis, drug resistance, epidemiology, and evolution," *Clinical Microbiology*, vol. 32, pp. 1-29, 2019.
- [281] B. Blasco and al., "Antimalarial drug resistance: Linking Plasmodium falciparum parasite biology to the clinic," *Nature Medicine*, vol. 23, pp. 917-928, 2017.
- [282] S. Krishna, "Science, medicine, and the future: malaria," *Malaria BMJ*, vol. 325, pp. 730-732, 1997.
- [283] A. Trampuz and al., "Clinical review: Severe malaria," *Critical Care*, vol. 7, pp. 315-323, 2003.
- [284] T. Taraschi and al., "Perturbation of red cell membrane structure during intracellular maturation of Plasmodium falciparum," *Science*, vol. 232, pp. 102-104, 1986.
- [285] A. Simões and al., "Selective internalization of choline-phospholipids in Plasmodium falciparum parasitized human erythrocytes," *BBA-Biomembranes*, vol. 1063, pp. 45-50, 1991.
-

- [286] M. Ancelin and al., "Present development concerning antimalarial activity of phospholipid metabolism inhibitors with special reference to in vivo activity," *Memórias do Instituto Oswaldo Cruz*, vol. 89, no. Suppl. 2, pp. 85-90, 1994.
- [287] M. Calas and al., "Antimalarial activity of molecules interfering with plasmodium falciparum phospholipid metabolism. Structure-activity relationship analysis," *Journal of Medicinal Chemistry*, vol. 2623, pp. 3557-3566, 1997.
- [288] M. L. Ancelin and al., "Antimalarial activity of 77 phospholipid polar head analogs: Close correlation between inhibition of phospholipid metabolism and in vitro plasmodium falciparum growth," *Blood*, vol. 91, pp. 1423-1437, 1998.
- [289] M. L. Ancelin and al., "Potent inhibitors of Plasmodium phospholipid metabolism with a broad spectrum of in vitro antimalarial activities," *Antimicrobial Agents and Chemotherapy*, vol. 47, pp. 2590-2597, 2003.
- [290] N. Kilian and al., "Role of phospholipid synthesis in the development and differentiation of malaria parasites in the blood," *Journal of Biological Chemistry*, vol. 293, pp. 17308-17316, 2018.
- [291] H. J. Vial and al., "Malarial Lipids," in *Malaria: Parasite Biology, Pathogenesis, and Protection*, Washington, DC, USA, American Society for Microbiology Press, 1998, pp. 159-175.
- [292] G. Pessi and al., "In vivo evidence for the specificity of Plasmodium falciparum phosphoethanolamine methyltransferase and its coupling to the Kennedy Pathway," *Journal of Biological Chemistry*, vol. 280, pp. 12461-12466, 2005.
- [293] S. Déchamps and al., "Glycerophospholipid acquisition in Plasmodium—A puzzling assembly of biosynthetic pathways," *International Journal of Parasitology*, vol. 40, pp. 1347-1365, 2010.
- [294] S. Wein and al., "Contribution of the precursors and interplay of the pathways in the phospholipid metabolism of the malaria parasite," *Journal of Lipid Research*, vol. 59, pp. 1461-1471, 2018.
- [295] M. L. Ancelin and al., "Several lines of evidence demonstrating that Plasmodium falciparum, a parasitic organism, has distinct enzymes for the phosphorylation of choline and ethanolamine," *FEBS Letters*, vol. 202, pp. 217-223, 1986.
-

- [296] B. Alberge and al., "Comparison of the cellular and biochemical properties of Plasmodium falciparum choline and ethanolamine kinases," *Biochemical Journal*, vol. 425, pp. 149-158, 2010.
- [297] T. Zimmerman and al., "Antiplasmodial activity and mechanism of action of RSM-932A, a promising synergistic inhibitor of Plasmodium falciparum choline kinase," *Antimicrobial Agents and Chemotherapy*, vol. 57, pp. 5878-5888, 2013.
- [298] V. Choubey and al., "Molecular characterization and localization of Plasmodium falciparum choline kinase," *Biochimica et Biophysica Acta*, vol. 1760, pp. 1027-1038, 2006.
- [299] M. Ancelin and al., "Choline kinase activity in Plasmodium-infected erythrocytes: Characterization and utilization as a parasite-specific marker in malarial fractionation studies," *Biochimica et Biophysica Acta*, vol. 875, pp. 52-58, 1986.
- [300] M. L. Ancelin and al., "Quaternary ammonium compounds efficiently inhibit Plasmodium falciparum growth in vitro by impairment of choline transport," *Antimicrobial Agents and Chemotherapy*, vol. 29, pp. 814-820, 1986.
- [301] X. Salom-Roig and al., "Dual molecules as new antimalarials," *Combinatorial Chemistry & High Throughput Screening*, vol. 8, pp. 49-62, 2005.
- [302] S. Peyrottes and al., "Choline analogues in malaria chemotherapy," *Current Pharmaceutical Design*, vol. 18, pp. 3454-3466, 2012.
- [303] V. Choubey and al., "Inhibition of Plasmodium falciparum choline kinase by hexadecyltrimethylammonium bromide: A possible antimalarial mechanism," *Antimicrobial Agents and Chemotherapy*, vol. 51, pp. 696-706, 2007.
- [304] L. Serran-Aguilera and al., "Choline kinase active site provides features for designing versatile inhibitors," *Current topics in Medicinal Chemistry*, vol. 14, pp. 2684-2693, 2014.
- [305] L. Serrán-Aguilera and al., "Plasmodium falciparum Choline Kinase Inhibition Leads to a Major Decrease in Phosphatidylethanolamine causing Parasite Death," *Sci. Rep.*, vol. 6, no. 33189, 2016.
- [306] S. Schiafino-Ortega and al., "1,2-Diphenoxiethane salts as potent antiplasmodial agents," *Bioorganic Medicinal Chemistry Letters*, vol. 28, pp. 2485-2489, 2018.
- [307] C.-J. Lai and al., "A Simple Kinetic Method for Rapid Mechanistic Analysis of Reversible Enzyme Inhibitors," *ASSAY and Drug Development Technologies*, vol. 1, no. 4, pp. 527-535, August 2003.
-

- [308] Y. C. Cheng and al., "Relationship between the inhibition constant (K_1) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction," *Biochemical Pharmacology*, vol. 22, no. 23, pp. 3099-3108, 1973.
- [309] L. Shi and al., "The serine, threonine, and/or tyrosine-specific protein kinases and protein phosphatases of prokaryotic organisms: A family portrait," *Federation of European Microbiological Societies Microbiology Reviews*, vol. 22, no. 4, pp. 229-253, 1998.
-