UNVERSITAT POLITÈCNICA DE VALÈNCIA
Departamento de Biotecnología

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

INSTITUTO DE QUÍMICA-FÍSICA "ROCASOLANO" (IQFR-CSIC)

Departamento de Biología Estructural y Cristalografía

DOCTORAL THESIS / TESIS DOCTORAL:

STRUCTURAL BIOMEDICINE: CHARACTERIZATION OF THE STRUCTURAL BASIS IN PROTEIN-DRUG RECOGNITION IN DIFFERENT HUMAN DISEASES.

Biomedicina Estructural: Caracterización de las Bases Estructurales del Reconocimiento Proteína-Droga en Diferentes Enfermedades Humanas.

Autora:

Alejandra Ángela Carriles Linares

Director:

Dr. Juan A. Hermoso

Valencia, Septiembre 2019

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

El Dr**. Juan A. Hermoso Domínguez**, Profesor de Investigación del Departamento de Cristalografía y Biología Estructural del Instituto de Química-Física "Rocasolano" (Consejo Superior de Investigaciones Científicas, CSIC) certifica que:

Doña Alejandra Ángela Carriles Linares, Licenciada en Biotecnología por la Universidad Politécnica de Valencia, ha realizado bajo mi dirección el trabajo titulado "Biomedicina Estructural: Caracterización de las Bases Estructurales del Reconocimiento Proteína-Droga en Diferentes Enfermedades Humanas", en el Departamento de Cristalografía y Biología Estructural del Instituto de Química-Física "Rocasolano" (Consejo Superior de Investigaciones Científicas, CSIC).

En mi opinión, este trabajo reúne las condiciones exigidas por la legislación vigente y tiene la originalidad, el rigor y la calidad científica necesarios y suficientes para ser presentado por Doña Alejandra Ángela Carriles Linares para optar al Grado de Doctor por la Universidad Politécnica de Valencia.

Para que así conste y con los efectos oportunos, firmo el presente certificado.

Madrid, Septiembre 2019

Fdo: Dr. Juan Antonio Hermoso Domínguez

Profesor de Investigación, CSIC

AGRADECIMIENTOS

Agradecer al Ministerio de Educación, Cultura y Deporte por haberme proporcionado el contrato FPU (FPU14/03190) que me ha permitido desarrollar esta Tesis Doctoral en el Instituto de Química-Física "Rocasolano" del Consejo Superior de Investigaciones Científicas (IQFR-CSIC), así como la financiación otorgada para poder realizar mi estancia predoctoral en el laboratorio del Prof. Hammershmidt, en Greifswald, Alemania (EST17/00751).

A mi supervisor Juan Hermoso, por apostar por mí el primer año de mi Tesis y ofrecerme un contrato desde la Universidad Alcalá de Henares durante el periodo de resolución del contrato FPU previamente mencionado, además de confiar en mí para desarrollar independientemente nuevos proyectos en el departamento, los cuales han resultado ser todo un reto.

A mis colaboradores, en especial a Federico Gago, por dar luz, aportar conocimiento en los proyectos y ayudarme siempre que he necesitado. A Ale y Héctor, gracias. Por todos los lotes de muestras que hemos intercambiado.

A PharmaMar, por darme una nueva visión de la industria científica y aspiraciones profesionales.

A mis compañeros de departamento, que, haya tratado en mayor o en menor medida con ellos, me han enseñado todas las facetas, buenas y malas, del trabajo en grupo y del compañerismo. Aquellos cuya relación ha resultado en amistad, ya os daré las gracias personalmente.

A Martín Martínez-Ripoll, por tener siempre su puerta abierta para sus sabios consejos, apoyo, divulgación de la cristalografía, ayuda y explicaciones de las tablas cristalográficas. Por su cariño y dedicación (y por todas las actualizaciones de software).

A todos los trabajadores del IQFR, desde portería, mantenimiento y otros departamentos, por todos los buenos días con sonrisa incluida que han alegrado mi tiempo aquí.

A los *Local Contact* del Laboratorio de Luz Sincrotrón ALBA. Gracias por vuestra ayuda a cualquier hora de las jornadas interminables de medida, por vuestros consejos, explicaciones, conocimiento y sonrisas al ver a los del "Roca".

Al Prof. Sven Hammershmidt y su grupo. Thank you for letting me be part of your group and making me feel just like home should feel. Thank you for trusting my knowledge and experience enough to "play around" with ÄKTAs and teach new students. I am happy and grateful to have had the opportunity to work at CFunGene and link other departments in new fields of study. Thank you to every single one I met during my stay. A total life changing experience, both personally and professionally.

A las personas que me han ayudado desde la Universidad Politécnica de Valencia, desde mi tutora Belén Picó y Francisco Vera en temas burocráticos, a todos mis responsables de docencia, en especial a Sergi Morais para la asignatura “Experimentación en Biotecnología”. Ha sido una experiencia que he disfrutado muchísimo, y que espero repetir en un futuro.

A mis amigas ya mencionadas en mi Trabajo Final de Carrera, Laura y Sandra, a quien dejé atrás en Valencia y que, a pesar de estar a unos “kilómetrillos” de distancia, han contado conmigo siempre y me han demostrado estar ahí cuando ha hecho falta. Tanto para unas (necesitadas) cervezas como para apoyarme en los momentos difíciles de estos años. A Carol y Silvia, por tanto.

A mis padres. Por todo lo que lleváis aguantado de mí durante todos estos años. Por toda vuestra paciencia, ayuda y apoyo que han hecho que llegue hasta aquí. Por vuestra comprensión, cuidados, esfuerzos, sacrificio y amor… y por fingir que os interesa el maravilloso mundo de la cristalografía.

Y finalmente, a todas las personas que han venido, han estado y se han ido (o no). Gracias por forjarme tal y como soy.

“Progress is made by trial and failure; the failures are generally a hundred times more numerous than the successes; yet they are usually left unchronicled.”

William Ramsay, 1852 to 1916

“If we knew what it was we were doing, it would not be called research, would it?”

Albert Einstein, 1879 to 1955

TABLE OF CONTENTS

SUMMARY

X-ray crystallography is a powerful technique for atomic structure resolution of macromolecules. The information generated impacts different fields involving basic and applied research on biomedicine and drug design and the development of nanotechnology and biotechnological applications. This dissertation focuses on current problematics and the target proteins involved (TryR, eEF1A2 and CBDP35) that are in sight for biotechnological development in the biomedical, pharmaceutical and food industry fields, in which X-ray crystallography plays a crucial role in the elucidation of their atomic structures and functions.

Attaining to biomedical and drug design problematics, we have solved the structure of *Leishmania infantum* TryR in complex with potent oxidoreductase inhibitors prone to further development as anti-trypanosomal drugs, thereby characterizing their binding and mechanism of action. This protein is a long recognized drug target for the treatment of Chagas disease, Human African Trypanosomiasis and leishmaniasis, as it plays a crucial and essential role in the redox-metabolism of the *Trypanosomatidae* parasites. Moreover, the crystallization and diffraction parameters of novel TryR dimerization disruptors have been assayed for inhibitors which have been rationally designed to bind the dimerization interface of TryR.

The "moonlighting" oncoprotein eEF1A2 is known to be highly post-translationally modified and to bind the anticancer drug plitidepsin. X-ray crystallography, combined with mass-spectrometry experiments, have been used as tools to identify novel post-translational modifications and structural features in eEF1A2:GDP. A unique modification, namely the addition of ethanolamine phosphoglycerol (EPG) to conserved glutamic residues (Glu301 and Glu374 in mammals), has been here observed for the first time. Structural analysis of these findings facilitate the understanding of eEF1A2's multiple functions and regulations. The acquirement of a conformationally homogenous eEF1A2:GTP sample, necessary for plitidepsin binding, has been has been assayed for eEF1A2:GTP:plitidepsin complex crystallization.

Regarding the cell wall binding domain of *Listeria monocytogenes* phage-encoded endolysin PlyP35 (CBDP35), we have solved the crystal structure of CBDP35 in complex with natural *Listeria* serovar 1/2a teichoic acid. This structure is the first cell wall binding module in complex with teichoic acids ever elucidated. Structural analysis revealed the main determinants for bacterial cell-wall binding, in particular, the molecular mechanism of N-acetyl-d-glucosamine recognition, a glycosidic moiety in teichoic acids of pathogenic serovars of *L. monocytogenes*. These findings shed light upon the biotechnological development of new tools in the food industry and phage-derived therapies to detect and treat bacterial infections.

RESUMEN

La cristalografía de rayos X es una potente técnica para la resolución de la estructura atómica de macromoléculas. La información generada, tiene gran impacto sobre diferentes campos relacionados con la investigación básica y aplicada, como son la biomedicina y diseño de fármacos, al igual que en el desarrollo de aplicaciones nanotecnológicas y biotecnológicas. Esta Tesis se centra en determinadas problemáticas actuales y en las proteínas involucradas en las mismas (TryR, eEF1A2 y CBDP35), siendo éstas sujeto de desarrollo biotecnológico en los campos de la biomedicina, farmacia y de la industria alimentaria, en el que la cristalografía de rayos X juega un papel crucial para dilucidar sus estructuras atómicas y funciones.

En consideración a la biomedicina y diseño de fármacos, hemos resuelto la estructura de la Tripanotión reductasa (TryR) de *Leishmania infantum* en complejo con potentes inhibidores de su actividad oxidorreductasa, con potencial de desarrollo como fármacos. Así, se ha caracterizado la unión y mecanismo de acción de éstos inhibidores. TryR es una reconocida diana farmacológica para el tratamiento de la enfermedad de Chagas, la Tripanosomiasis Humana Africana y la leishmaniosis, ya que desempeña un papel crucial y esencial en el metabolismo redox de los parásitos de la familia *Trypanosomatidae*. Además, se han analizado los parámetros de cristalización y difracción de novedosos inhibidores de la dimerización de TryR, cuyo diseño racional se basa en la unión a la interfaz de dimerización de la misma.

La oncoproteína eEF1A2, involucrada en múltiples funciones celulares y sujeto de numerosas modificaciones post-traduccionales, se une al fármaco anticancerígeno plitidepsina. La cristalografía de rayos X, combinada con experimentos de espectrometría de masas, se han utilizado como herramientas para identificar nuevas modificaciones post-traduccionales y características estructurales en eEF1A2:GDP. Una modificación única, la adición de etanolamina fosfoglicerol (EPG) a aminoácidos conservados (Glu301 y Glu374 en mamíferos), se ha observado aquí por primera vez. El análisis estructural de estos hallazgos facilita la comprensión de las múltiples funciones y regulaciones de eEF1A2. La adquisición de una muestra conformacionalmente homogénea de eEF1A2:GTP, necesaria para la unión a la plitidepsina, ha sido evaluada en ensayos de cristalización del complejo terciario de eEF1A2: GTP: plitidepsina.

Con respecto al dominio de unión a la pared celular de la endolisina PlyP35 codificada por el fago P35 de *Listeria monocytogenes* (CBDP35), hemos resuelto la estructura cristalina de CBDP35 en un complejo con ácido teicoico natural de *L. monocytogenes* serovar 1/2a. Esta estructura es el primer módulo de unión a la pared celular en complejo con ácidos teicoicos jamás dilucidado. El análisis estructural reveló los principales determinantes para la unión de la pared celular bacteriana, en particular, el mecanismo molecular del reconocimiento de N-acetil-d-glucosamina, una decoración de carácter glicosídico en ácidos teicoicos de serovares patógenos

de *L. monocytogenes*. Estos hallazgos arrojan luz sobre el desarrollo biotecnológico de nuevas herramientas en la industria alimentaria y las terapias derivadas de fagos para detectar y tratar infecciones bacterianas.

RESUM

La cristal·lografia de raig X és una potent tècnica per a la resolució de l'estructura atòmica de macromolècules. La informació generada té gran impacte sobre diferents camps relacionats amb la investigació bàsica i aplicada, com són la biomedicina i disseny de fàrmacs, igual que en el desenvolupament d'aplicacions nanotecnológiques i biotecnològiques. Aquesta Tesi es centra en determinades problemàtiques actuals i en les proteïnes involucrades en les mateixes (TryR, eEF1A2 i CBDP35), sent estes subjecte de desenvolupament biotecnològic en els camps de la biomedicina, farmàcia i de la indústria alimentària, en el que la cristal·lografia de raig X juga un paper crucial per a dilucidar les seues estructures atòmiques i funcions.

En consideració a la biomedicina i disseny de fàrmacs, hem resolt l'estructura de la Tripanotión reductasa (TryR) de Leishmania infantum en complex amb potents inhibidors de la seua activitat oxidorreductasa, amb potencial de desenrotllament com a fàrmacs. Així, s'ha caracteritzat la unió i mecanisme d'acció d'estos inhibidors. TryR és una reconeguda diana farmacològica per al tractament de la malaltia de Chagas, la Tripanosomiasi Humana Africana i la leishmaniosi, ja que exerceix un paper crucial i essencial en el metabolisme redox dels paràsits de la família Trypanosomatidae. A més, s'han analitzat els paràmetres de cristal·lització i difracció de nous inhibidors de la dimerització de TryR, el disseny racional dels quals es basa en la unió a la interfície de dimerización de la mateixa.

L'oncoproteína eEF1A2, involucrada en múltiples funcions cel·lulars i subjecte de nombroses modificacions posttraduccionals, s'unieix al fàrmac anticancerigen plitidepsina. La cristal·lografia de raig X, combinada amb experiments d'espectrometria de masses, s'han utilitzat com a ferramentes per a identificar noves modificacions posttraduccionals i característiques estructurals en eEF1A2:GDP. Una modificació única, l'addició d'etanolamina fosfoglicerol (EPG) a aminoàcids conservats (Glu301 i Glu374 en mamífers), s'ha observat ací per primera vegada. L'anàlisi estructural d'estes troballes facilita la comprensió de les múltiples funcions i regulacions d'eEF1A2. L'adquisició d'una mostra conformacionalmente homogènia d'eEF1A2:GTP, necessària per a la unió a la plitidepsina, ha sigut avaluada en assajos de cristal·lització del complex terciari d'eEF1A2: GTP: plitidepsina.

Respecte al domini d'unió a la paret cel·lular de l'endolisina PlyP35 codificada pel fago P35 de Listeria monocytogenes (CBDP35), hem resolt l'estructura cristal·lina de CBDP35 en un complex amb àcid teicoico natural de L. monocytogenes serovar 1/2a. Esta estructura és el primer mòdul d'unió a la paret cel·lular en complex amb àcids teicoicos mai dilucidat. L'anàlisi estructural va revelar els principals determinants per a la unió de la paret cel·lular bacteriana, en particular, el mecanisme molecular del reconeixement de N-acetil-d-glucosamina, una decoració de caràcter glicosídico en àcids teicoicos de serovares patògens de L. monocytogenes. Estes

troballes fan llum sobre el desenrotllament biotecnològic de noves ferramentes en la indústria alimentària i les teràpies derivades de fagos per a detectar i tractar infeccions bacterianes.

ABBREVIATIONS

aa-tRNA: aminoacylated-transference ribonucleic acid

AdoMetDC: S-adenosylmethionine decarboxylase

ADXV: Area Detector X-ray Viewer

AP: Alkaline Phosphatase

ARG: Arginase

ARS: aminoacyl-tRNA synthetase

CBD: Cell Wall-Binding Domain

CBP: Choline Binding Protein

CC1/2: Correlation Coefficient between two groups of reflections

CCD: Charge-Coupled Device

CCP4: Collaborative Computational Project No. 4

Dap: Aminopimelic acid

DMSO: Dimethyl Sulfoxide

DNA: deoxyribonucleic acid

ESRF: European Synchrotron Radiation Facility

EPG: Ethanolamine phosphoglycerol

F_{hkl}/F(hkl): Structure Factor for reflection with indices *hkl*

FEM: Feature Enhanced Map

GAPDH: glyceraldehyde-3-phosphate dehydrogenase

GEF: guanine nucleotide exchange factor

Gal: Galactose

Glc: Glucose

GlcNAc: N-acetyl-D-glucosamine

GppNHp: 5'-Guanylyl-imidodiphosphate

GR: Glutathione reductase

GroP: Polyglycerol-Phosphate

GSH: Glutathione (reduced)

Gsp: Glutathionylspermidine

GspS: Glutathionylspermidine Synthetase

GSSG: Glutathione (oxidized)

HAT: Human African Trypanosomiasis

I_{hkl}/I(hkl): Intensity for reflection *hkl*

IPTG: Isopropyl β-D-1-thiogalactopyranoside

LB: Luria-Bertani medium

LTA: Lipoteichoic Acid

MAD: Multi-wavelength Anomalous Diffraction

MIR: Multiple Isomorphous Replacement

MS: Mass Spectrometry

MurNAc: N-acetylmuramic acid

NCS: Non-Crystallographic Symmetry

O.D.: Optical Density

ODC: Ornithine Decarboxylase

IP: Isoelectric Point

PA: Polyamine

PCT: Pre-Crystallization Test

PDB: Protein Data Bank

PE: Phosphatidylethanolamine

PEG: Polyethylene Glycol

PKR: double-stranded RNA-activated protein kinase

PPP: Pentose-Phosphate Pathway

PT: Phage Therapy

PTM: Post-Translational Modification

RboP: Polyribitol-Phosphate

Rha (α-L-Rha): α-L-Rhamnose

R_{merge}: measure of the disagreement between the different intensity measurements for a specific reflection

RNA: Ribonucleic Acid

ROS: Reactive Oxygen Species

R_{pim}: improved R_{merge} statistical measurement which accounts for multiplicity when merging data

SAD: Single-wavelength Anomalous Diffraction

SpdS: Spermidine Synthase

TA: Teichoic Acid

TBSV: Tobacco Mosaic Virus

TLS: Translation-Libration-Screw

TRL: Trypanothione Reductase Ligand (synthesized at Medical-Chemistry Institute, CSIC)

TryR: Trypanothione Reductase

Trx: Thioredoxin

TSA: Trypanothione Synthetase-Amidase

$T(SH)_2$: Trypanothione (reduced)

TS_2: Trypanothione (oxidized)

TXN: Tryparedoxin

TXNPx: Tryparedoxin peroxidase

UNGA: United Nations General Assembly

WHO: World Health Organization

α-D-GlcNAc: N-acetyl-α-D-glucosamine

β-D-GlcNAc: N-acetyl-β-D-glucosamine

$|F_{hkl}|$: Module of a structure factor

φ(hkl): Structure factor phase

FIGURES

TABLES

1. INTRODUCTION

The discovery of X-rays by Wilhem Conrad Röntgen by the end of the XIX century, Max von Laue's findings on the interaction of X-rays with crystalline matter in 1912 and the contributions of William H. Bragg and William L. Bragg which allowed the understanding of molecular space configuration within a crystal, led to a progressive development of X-ray crystallography that would later revolutionize the fields of physics, chemistry, biology and biomedicine. Proof of this are the 29 Nobel Prizes awarded involving this methodology.

X-ray crystallography is currently the most favored technique for structure determination of proteins and biological macromolecules, along with NMR and electron microscopy. Thus, by April 2019, there have been a total of 135258 structures solved by X-ray crystallography deposited in the Protein Data Bank (PDB), compared to the 12596 and 3094 entries determined by NMR and electron microscopy, respectively. Increasingly, the biological sciences require structural information to shed light on previously unanswered questions.

A century after the discovery of X-rays, the United Nations General Assembly (UNGA) proclaimed 2014 as the International Year of Crystallography to highlight the continuing importance of this technique and the knowledge of three-dimensional structures of biomolecules. X-ray crystallography has contributed in gaining insight into the structure of matter and understanding critical areas of science. Structural data impacts basic and applied research on health and disease; production of food and energy; material industry and other fields pertaining to global prosperity and environmental sustainability (Burley, *et al*., 2018), as well as on the development of nanotechnology and biotechnological applications. Herein, although crystallography underpins all of the sciences today and has many applications, it remains widely unknown to the general public.

1.1. STRUCTURES, DISEASES AND DRUG DESIGN

From its very beginning, biostructural research has not only provided tremendous breakthroughs in basic biological processes but it has also significantly contributed to the understanding of the molecular mechanisms underlying human disease and the characterization of pharmaceutical and new bioactive compound structures to reveal their mechanism of action in atomic detail.

Although the term *proteopathy* has been recently used to refer to diseases caused by the production of certain structurally abnormal proteins that are misfolded and aggregate in a crystallization-like seeding mechanism (thereby disrupting the function of cells, tissues and organs of the body), it is certain that molecular physiology and structural biochemistry became biomedically relevant since the determination of myoglobin and haemoglobin protein structures in 1959 by John C. Kendrew and Max F. Perutz, respectively. Their discovery allowed the

explanation of the storage and transport of oxygen through the organism and the understanding of sickle cell anemia when an aberrant structure of a hemoglobin variant is produced as a result of gene mutation.

More recently, it is common practice to identify key proteins involved in a disease and subject them to structural characterization. Their atomic distribution enables the understanding of protein function from a mechanistic point of view, portraying the molecular pathophysiology of a certain disease. Beyond single protein approaches, proteomic analysis and structural genomic projects aim to picture at steadily increasing levels of complexity, protein-protein, protein-DNA and protein-RNA complexes in order to elucidate the molecular machinery of the cell at a high-throughput rate and so promoting an integrated understanding of the mechanisms of human disease.

An advanced molecular understanding of disease-relevant factors can open new strategies for both the design and screening procedures of interfering small molecules. Hence, the three-dimensional structures of many human and pathogen proteins guide rational drug design. Successful examples of structure-based drug design are captopril, for hypertension-related diseases (Cushman, et al., 1977), lopinavir for HIV/AIDS (Sham, 1998) and imatinib as an anticancer targeted therapy for chronic myelogenous leukemia (Zimmermann, et al., 1997), among many others. In other cases, the goal of the crystallographic study of a protein-ligand (i.e. target-drug) complex is lead compound optimization, where a molecule's specificity and affinity for its target is improved until adequate safety, efficacy and specificity parameters of a drug are achieved by means of structure-based chemical modifications.

Historically, endogenous bioactive ligands were identified long before their cognate biomolecular receptors could be isolated and before their chemical structures were elucidated. Since the 1970s, drug discovery is driven by biological targets, genetic studies, animal models, molecular biology, gene technology and protein science. Hence, structural biology research is tightly integrated with biophysical, biochemical and cellular studies in attempts to translate the higher-order molecular view into specific options for therapeutic intervention.

Nowadays, major pharmaceutical companies rely on structure-based drug design, along with virtual ligand screening and docking programs (Kitchen, 2004). In this way, chemical libraries are filtered and a richer picture of ligand-receptor interaction is portrayed, based on the physico-chemical descriptors (pharmacophores) inferred from the target's structure. All in all, structure-based drug design, virtual ligand screening and homology modelling, where a reliable structural model is used if the X-ray crystal structure of a target protein is unavailable, have had a tremendous impact on the drug discovery process (Figure 1.1.). Nevertheless, many years of research may be necessary to turn a drug lead into a drug that will be both effective and tolerated

by the human body. Additional years of research and development will bring the drug through clinical trials until it finally reaches the market

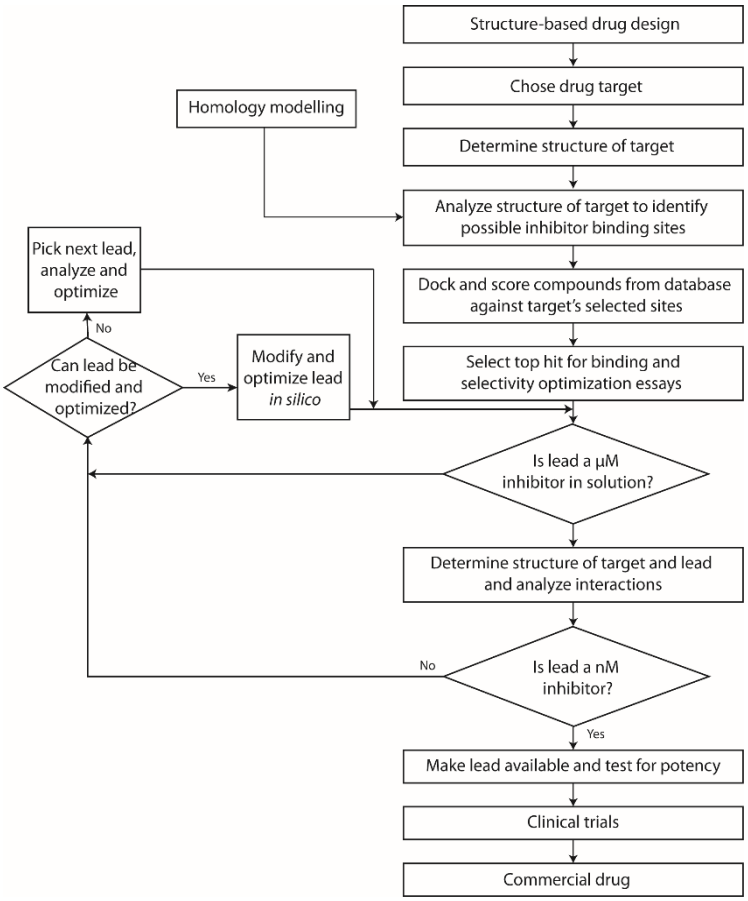


Figure 1.1.: The iterative process of structure-based drug design: The first cycle involves cloning and purifying the target protein to determine its structure. In some occasions homology modelling is used when the structure cannot be obtained experimentally. Using computer algorithms, compounds or fragments of compounds from a database are positioned into a selected region of the structure, which are scored and ranked based on their steric and electrostatic interactions. In the second cycle, structure determination of the target in complex with a promising lead, reveals sites on the compound that can be optimized. Additional cycles include synthesis of the optimized lead, structure determination of the new target:lead complex, and further optimization of the lead compound (Anderson, 2003).

After more than 35 years from the first discussions on structural biology as a tool for understanding disease and for drug discovery purposes (Beddell, *et al*. 1976), structure-based drug design has played a key role in the development of several marketed drugs (Seddon, *et al*., 2012) and is now an integral part of most industrial drug discovery programs (Mountain, 2013). The explosion of genomic, proteomic and structural information, as well as key advancements in computational methods, have aided the identification of new targets involved in disease and assessment of their druggability. For example, consortiums such as RIKEN Structural Genomics/Proteomics Initiative have contributed with the resolution of 2743 macromolecular structures which have been deposited in the PDB. As a result of these high-throughput initiatives, structural biology has been able to tackle more difficult problems than it has ever previously been amenable to structural elucidation.

The following sections describe current problematics and the target proteins there involved that are in sight for biotechnological development in the biomedical, pharmaceutical and food industry fields, in which X-ray crystallography plays a crucial role to shed light onto their atomic structures.

1.2. TARGETING TRYPANOSOMATID'S TRYR

1.2.1. Tropical Neglected Parasitic Diseases

Parasitic diseases caused by *Trypanosomatidae* protozoans include Chagas disease, sleeping sickness -also known as Human African Trypanosomiasis (HAT)- and several forms of leishmaniasis. These are vector borne diseases caused by *Trypanosoma cruzi*, *Trypanosoma brucei* and different species belonging to the genus *Leishmania*, transmitted by the triatominae bugs, tsetse fly and female sandfly, respectively. (Stuart, *et al*., 2018). According to the World Health Organization (WHO) (WHO, 2019), they are considered as the most challenging among the neglected tropical diseases.

Chagas disease is a potentially life-threatening illness that affects 7-8 million people worldwide and causes 10000 deaths per year due to further complications and organ damage because of the persistent presence of parasites within tissues. Infections with *Trypanosoma cruzi* are mainly found in endemic areas of 21 continental Latin American countries, although in the last decades, infection has been increasingly detected in countries where Chagas disease is non-endemic. (WHO, 2019). Chronic Chagas disease slowly attacks the heart and the tissues of the gastrointestinal tract mainly if the patient's immune function is weakened. However, acute-phase symptoms may be resolved spontaneously in a short time period (Vilar, Souza and Lannes, 2015).

Regarding HAT, acute infections are caused by the *Trypanosoma brucei rhodiense* in eastern and southern African countries, while *Typanosoma brucei gambiense* causes chronic infections in west and central African countries (WHO, 2019). According to WHO, the number of new cases has shown a clear decrease during the last ten years, although it does not reflect the lack of control efforts. People who become infected may or may not show signs of illness immediately, but over time the parasite crosses the blood-brain barrier and migrates to the central nervous system. Early diagnosis is therefore difficult due to the lack of specific signs and sensitivity of the parasitological methods available. If untreated, the disease is usually fatal (Fevre, *et al*., 2008).

Leishmaniasis is caused by different parasites from the genus *Leishmania* which translate into different clinical manifestations. These are visceral, mucocutaneous and cutaneous forms. Out of 200 countries and territories reporting to WHO, 97 countries and territories are endemic for leishmaniasis in 2017, distributed in Africa, Asia and Latin America mainly.

Nevertheless, 1713 cases of cutaneous leishmaniasis were reported globally in non-endemic countries (WHO, 2019). According to the WHO, there are currently over 1 billion of people at risk of infection, with 1 million cases of cutaneous leishmaniansis being reported in the last five years and over 20000 yearly deaths caused by visceral leishmaniasis, the most severe form of the disease. Whilst cutaneous leishmaniasis can be caused by almost all species of *Leishmania*, mucocutaneous and visceral forms of the disease is mainly caused by *L. braziliensis*, *L. donovani* and *L. infantum*, respectively (Chappuis, et al., 2007; Reithinger, et al., 2007).

1.2.2. In Need of New Therapies

The chemotherapeutic agents used against the aforementioned protozoa are usually highly toxic and are of limited efficacy. Furthermore, because these kinetoplastid diseases are associated to rural areas of developing countries, pharmaceutical companies remain uninterested in developing new treatments (de Menezes, *et al*., 2015). In the case of Chagas disease, there are only two available medicines that are used in therapy: benznidazole, often the first-line treatment in most countries, and nifurtimox. These are only effective only in the acute phase of the disease, have multiple side effects including kidney, liver failure and other nervous-system disorders, and are prone to resistance development (Bernardes, *et al*., 2013).

For HAT, pentamidine and suramine are used to treat the first stage infection, whilst intravenous melarsoprol, an arsenical derivative, and eflornithine are used once the trypanosomes have crossed the blood-brain barrier. These drugs cause metabolic disorders, are of high costs and only effective against *T. brucei gambiense* (Bernardes, *et al*., 2013). Without prompt diagnosis and treatment, the disease is usually fatal.

Therapy for leishmaniasis mainly relies on injectable pentavalent antimony as a fist-line treatment (administered as sodium stibogluconate and meglumine antimoniate), which has become ineffective in India due to the appearance of resistance of parasites (Jha, *et al*., 1995). Pentamidine, amphotericine B along with their liposomal formulations, paromomycin and miltefosine, the only oral drug available, are used as a second-line treatment for leishmaniasis. These drugs are not specific for *Leishmania*, as they have been previously developed and used for the purpose of treating bacterial and fungal infections, and breast cancer in the case of miltefosine (Croft and Engel, 2006). In addition, cross-treatments can be used for leishmaniasis, such as pentamidine, administered in HAT patients (Monzote, 2009). Major inconveniences are the rapid appearance of resistance, the injection pain, the metabolic disorders, nephrotoxicity and the strong side effects resulting from therapy.

All in all, the chemotherapy used in the treatment of these neglected diseases (Figure 1.2.) present numerous drawbacks and are far from satisfying the current demands of

endemic populations due to their cost, administration, toxicity and resistance issues, resulting in an urgent need to identify and develop new therapeutic alternatives.

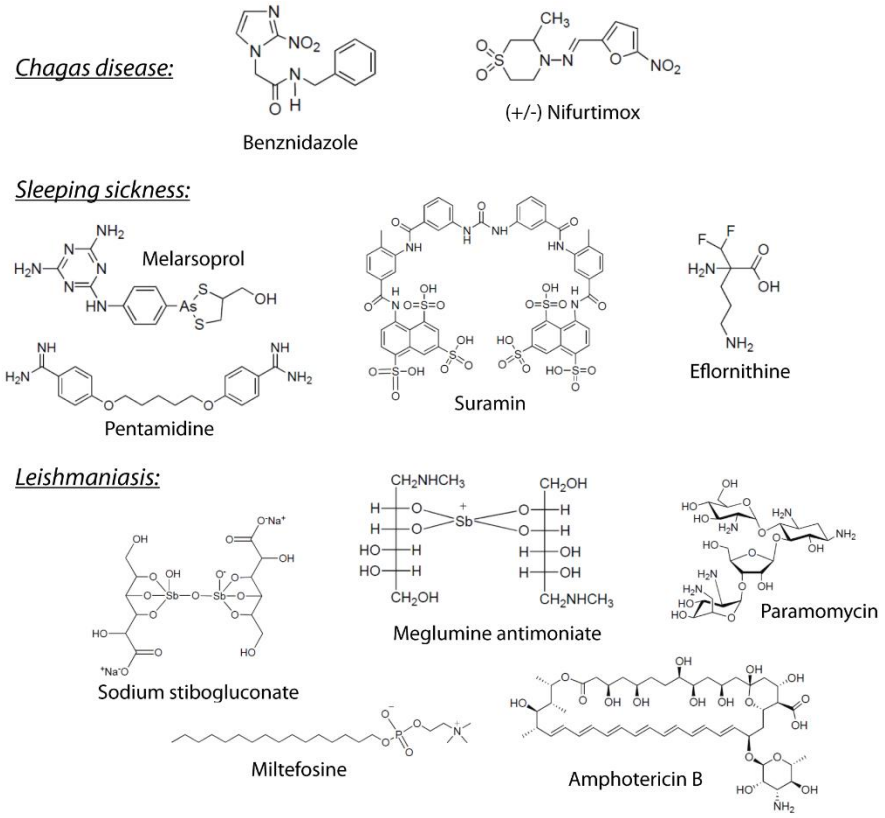


Figure 1.2.: Drugs currently used for the treatment of Chagas disease, sleeping sickness and leishmaniasis. (Figure from Bernardes, *et al*., 2013)

1.2.3. Therapeutic Alternatives: The Trypanothione Metabolism

Comparative studies between *Leishmania spp* and *Trypanosoma* have been conducted aiming to find broad spectrum antiprotozoal agents acting against both parasitic groups. Metabolic pathways that are both essential for parasite survival and absent from the host are considered good possible starting points (El-Sayed, *et al.*, 2005). However, the adaptation of the parasite to diverse environments within the host, and their complex life cycles (Figure 1.3.) hinder the development of broad spectrum drugs for protozoan diseases.

1.2.3.1. Unique Characteristics of *Trypanosoma* and *Leishmania* Genres

The pathogenic trypanosomatids *T. brucei*, *T. cruzi* and *Leishmania spp.* exhibit the most typical eukaryotic organelles. However, several prominent physiological and biochemical features distinguishing trypanosomatids from other eukaryotic cells have guided the development for new drug treatments. Such structures involve the unusual network of interconnected mini and maxi circles of mitochondrial DNA (kDNA), called kinetoplast, and the so-called glycosome, a peroxisome-like organelle in which glycolysis takes place as well as pyrimidine and nucleotide

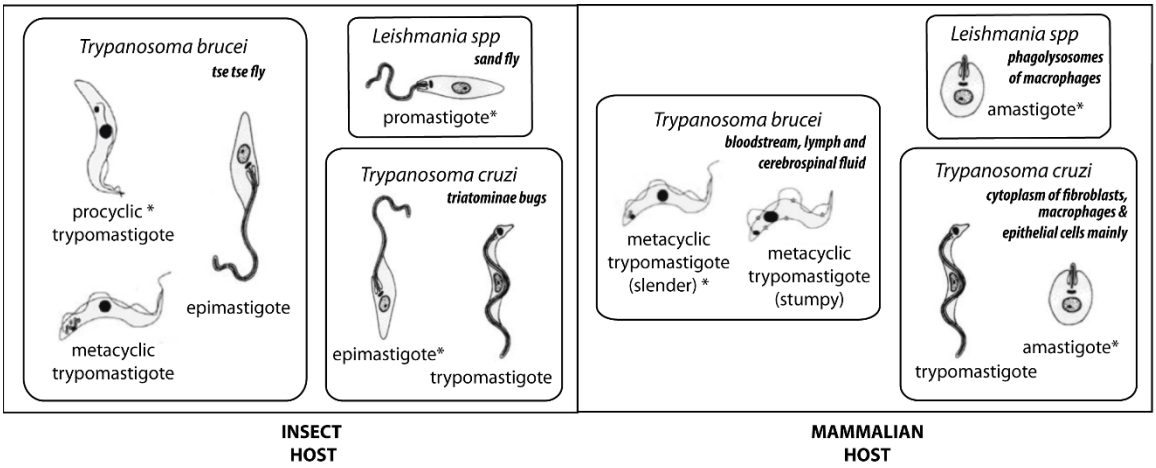


Figure 1.3.: Summary of the developmental stages found in the insect vector and mammalian hosts for the protozoan parasites *Leishmania spp., T. cruzi and T. brucei*: Proliferative stages are indicated by (*). Kinetoplast is represented as a small black circle. Localization of each parasite in vector and host is specified. (Adapted from Rodrigues, Godino and de Sousa, 2014).

biosynthesis, part of the oxidative pentose phosphate pathway (PPP), and among other metabolic routes (Michels, *et al*., 2006). Trypanosomes also contain unusual pathways for lipid metabolism (Lee, Stephens and Englund, 2007) and most noticeably, are equipped with an essential thiol-redox metabolism based on trypanothione (Kruth-Siegel and Comini, 2008). Herein, the polyamine and trypanothione pathways are considered as a whole unique metabolism of trypanosomatids (Ilari *et al*., 2017). Furthermore, the PPP is crucial to these parasites as it is a main source of reducing power in the form of NADPH, used for trypanothione reduction, to detoxify the parasite's cells from reactive oxygen species (ROS) produced by the host (Kovářová and Barrett, 2016).

1.2.3.1.1. The Polyamine-Trypanothione Pathway

Inhibition of the polyamine-trypanothione metabolism enzymes, in particular those involved in the synthesis, use or recycling of trypanothione, are considered the best available options for drug discovery. These enzymes have been validated as drug targets (the enzyme is essential for parasite survival) as well as chemically (enzyme inhibition is possible and selective) (Jager, Koch and Flohe, 2013). A scheme of the polyamine-trypanothione pathway for *Leishmania* is shown in Figure 1.4. The polyamine (PA) biosynthetic pathway starts with the synthesis of the polyamine precursor l-ornithine, catalysed by arginase (ARG), which is then decarboxylated by ornithine decarboxylase (ODC), which produces putrescine, to whom spermidine synthase (SpdS) adds an aminopropyl group donated from the decarboxylated S-adenosylmethionine (dAdoMet), the reaction product of S-adenosylmethionine decarboxylase (AdoMetDC).

The pathways to supply spermidine are not conserved among different trypanosomatids. As shown in Figure 1.4., *Leishmania* transform arginine into spermidine. Trypanosomes lack a

functional ARG and obtain l-ornithine from the mammalian host. The next two steps, namely the transformation of ornithine into spermidine, are identical in *T. brucei* and *Leishmania sp.*, whereas *T. cruzi* is auxotroph for polyamines and scavenges putrescine and spermidine from the host (Krauth-Siegel and Leroux, 2012).

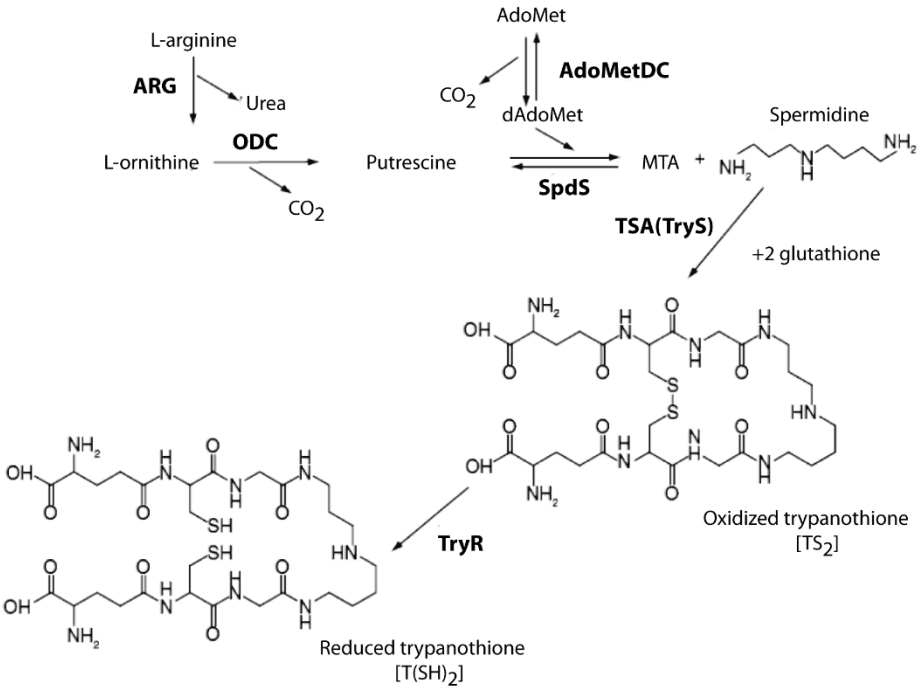


Figure 1.4.: Scheme of the polyamine-trypanothione metabolism in *Leishmania.* Enzymes involved in this metabolism are abbreviated and in bold. (Adapted from Ilari, 2017)

Once spermidine is obtained, trypanothione (N1, N8-bis(glutationyl)spermidine) is synthesized by means of two consecutive reactions in which two molecules of glutathione (GSH) are joined by one molecule of spermidine. As in other organisms, GSH is generated by gamma-glutamylcysteine synthetase and glutathione synthetase, also shown to be essential (Leroux and Krauth-Siegel, 2016). Depending on the parasite's genetic background, either glutathionylspermidine synthetase (GspS) catalyzes the ATP-dependent addition of glutathione to one of the amino groups of spermidine to form glutathionylspermidine (Gsp), and a second GSH molecule is added by trypanothione synthetase-amidase (TSA), or, TSA solely catalyzes both steps of the reaction. Trypanothione is kept reduced $T(SH)_2$ by means of trypanothione reductase (TryR), the only enzyme that connects the NADPH- and thiol-based redox systems in these parasites (Krauth-Siegel and Comini, 2008).

1.2.3.2. Trypanothione and the Thiol-Redox Metabolism

All living organisms are exposed to reactive oxygen and nitrogen species (ROS, RNS) such as superoxide anion ($O_2{\cdot}^-$), hydrogen peroxide (H_2O_2), hydroxyl radical (·OH), nitric oxide (NO·) and peroxynitrite (OONO·) (Olin-Sandoval, Moreno-Sanchez and Saavedra, 2010).

A redox homeostasis is vital for cellular survival, as they participate in cellular signalling and apoptosis (Tomás and Castro, 2013). With regard to RNS, $T(SH)_2$ offers an efficient protection by sequestering NO· to form a stable dinitrosyl iron complex with a higher affinity as compared to GSH (Bocedi, *et al.* 2010).

The generation of ROS mainly begins in the mitochondria (though may be produced elsewhere in the cell and spread to this organelle) with the production of $O_2^{\cdot-}$ as a consequence of an electron being transferred to molecular oxygen (Quijano, *et al.* 2016). The mitochondrial $O_2^{\cdot-}$ radical is disrupted by superoxide dismutases, producing O_2 and H_2O_2. This H_2O_2 can then be reduced to H_2O by peroxidases and, in the presence of reduced transition metals such as Fe_2^+, it can produce ·OH ions.

Mammalian cells/ vertebrates depend upon catalase, glutathione reductase (GR), thioredoxin reductase (TrxR) and glutathione peroxidases (GPx), for the detoxification of hydroperoxides or toxic metabolites and maintenance of cellular redox homeostasis. However, trypanosomatids lack TrxR and GR and so rely on alternative enzymatic systems (Schmidt and Krauth-Siegel, 2002). Most predominantly, they utilize their unique TryR/trypanothione and tryparedoxin/tryparedoxin peroxidase (TXN/TXNPx) systems, ascorbate peroxidases or selenium-free peroxidases similar to GPxs, and other low molecular mass thiols present in low concentrations in the cell, namely ovothiol A, GSH, glutathionyl-spermidine and dihydrolipoic acid (Castro and Tomás, 2008; Krauth-Siegel and Comini, 2008). In *Leishmania* promastigotes, ovothiol A is the most abundant thiol (around 30-60% of low molecular weight intracellular thiols), although it is almost undetectable in the amastigote forms (Ariyanayagam and Fairlamb, 2001), and GSH has been shown to participate in RNS elimination (Romao, et al. 2006). Furthermore, *Leishmania sp.* and *T. cruzi*, reside in the phagolysosome of host macrophages (Figure 1.3.), where they are subjected to an extreme oxidative environment (Horta, *et al.*, 2012), and in which they are capable of adapting and surviving due to their efficient redox metabolism.

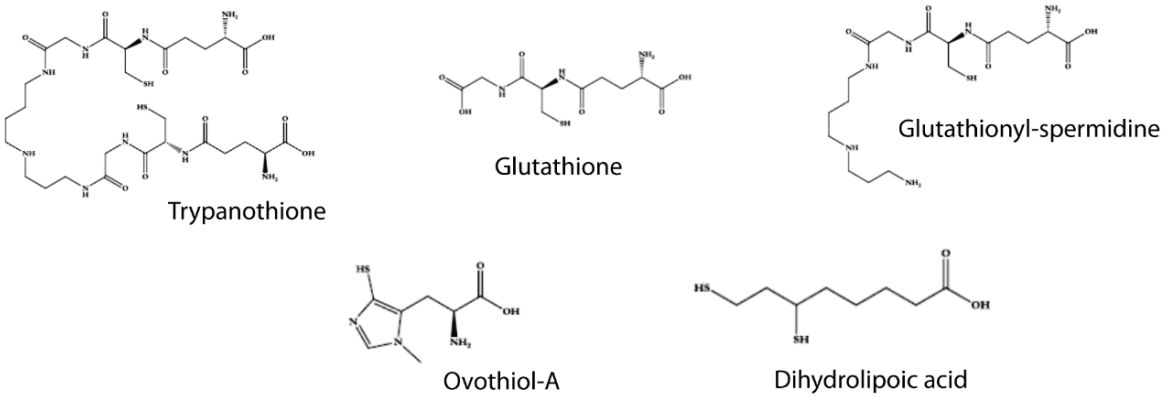


Figure 1.5.: Low molecular weight thiols involved in the thiol-redox metabolism of trypanosomes.

The role of trypanothione in the detoxification of ROS is reflected in the multiplicity of trypanothione-dependent peroxidases. Hence, although $T(SH)_2$ cannot efficiently reduce peroxides itself, it functions as the principal electron donor for the oxidized electron donor substrates of several enzymes of the antioxidant machinery such as the previously mentioned ovothiol A and GSH, as well as the oxidized forms of the dithiol redox proteins, such as TXN and thioredoxin (Trx) (Valko, *et al.*, 2007), which are part of the peroxide detoxification system (Figure 1.6.). Moreover, $T(SH)_2$ fuels many other essential parasite functions, namely heavy metal detoxification, DNA replication and repair, assembly of iron–sulfur clusters, and detoxification of ketoaldehydes and xenobiotics (Comini and Flohé, 2013).

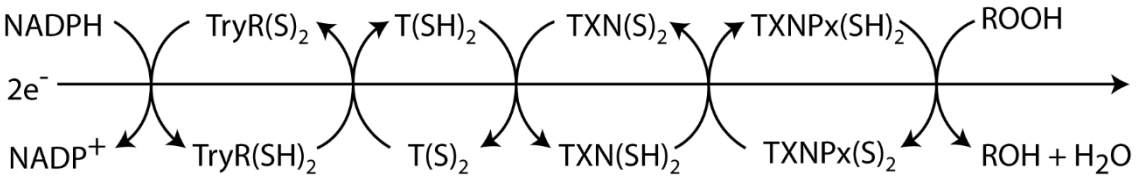


Figure 1.6.: The trypanothione-dependent peroxide detoxification pathway. Trypanothione is used by the tryparedoxin/tryparedoxin peroxidase system (TXN/TXNPx) to reduce hydrogen peroxide and alkyl-hydroperoxide to water and alcohol, respectively. TryR-trypanothione reductase, TXNPx-tryparedoxin peroxidase, TXN-tryparedoxin. (Adapted from Colotti, *et al*., 2013).

Remarkably, several of the currently available anti-trypanosomatid drugs affect the trypanothione metabolism (e.g., nifurtimox, melarsoprol, antimonials, and eflornithine) (Leroux and Krauth-Siegel, 2016). Taking into account the essential nature of the enzymes involved in the synthesis and reduction of trypanothione and Tpx (Castro and Tomás, 2008), virtually all pathway components are potential drug targets. However, most of the drug development approaches have focused on two central enzymes: TryR and TryS.

1.2.3.2.1. Trypanothione vs. Glutathione

In most living cells, the enzymatic antioxidant machinery primarily relies on GSH as a source of electrons to reduce and inactivate ROS and RNS, together with GR, which reduces oxidized glutathione (GSSG) at expense of NADPH oxidation. Just as trypanothione, GSH can non-enzimatically reduce the oxidized forms of cellular antioxidant molecules such as ascorbate and vitamin E due to its higher reducing potential (Valko, *et al.*, 2007). The relative slow rate of these non-enzymatic reactions is insufficient given the severe oxidative stress these parasites are exposed to.

Despite the presence of significant amounts of GSH in trypanosomatid parasites, their antioxidant enzymatic machinery uses its analog $T(SH)_2$. Furthermore, although $T(SH)_2$ and GSH show close redox potential values, the dithiol conjugate of $T(SH)_2$ displays other physico-chemical properties that confer functional advantages. Thus, being a dithiol, the formation of the intramolecular thiol disulphide bond is faster in $T(SH)_2$ than the intermolecular disulphide bond

formation between two GSH molecules (Krauth-Siegel, Comini and Schleker, 2007). Moreover, the positively charged amino group in spermidine of $T(SH)_2$ confers a -SH pK value near 7.4, and so is more reactive than GSH (-SH pK = 8.66) at physiological pH (Krauth-Siegel, Comini and Schleker, 2007). This confers an advantage to trypanosomatids for extreme oxidative stress resistance. Nevertheless, both $T(SH)_2$ and GSH concentrations are similar (0.1 - 2.1 mM and 0.03 - 2.1 mM, respectively), although the concentration ratio can vary among different trypanosomatids, stages or growing phases (Krauth-Siegel and Comini, 2008).

1.2.4. Trypanothione Reductase (TryR)

TryR (EC 1.8.1.12) is a ~50 kDa, FAD-cystine-oxidoreductase that catalyzes the reduction of TS_2 at expenses of NADPH oxidation. Hence, TryR is the main link between the cellular reductive power and the antioxidant system in the unique thiol-redox metabolism of *Trypanosomatidae* parasites (Olin-Sandoval, Moreno-Sanchez and Saavedra, 2010).

TryR is essential for parasite survival. Herein, in *L. donovani*, disruption of the TryR gene decreases the ability of the parasite to survive oxidative stress inside activated macrophages (Dumas, *et al.*, 1997), whilst over-expression of the same gene enhances the ability to regenerate $T(SH)_2$ from oxidized trypanothione (TS_2), although their sensitivity towards exogenous oxidants remain unchanged (Kelly, *et al.*, 1993). Furthermore, conditional knock-out studies in *T. brucei* resulted in growth arrest, enhanced susceptibility towards H_2O_2 and loss of virulence in mice (Krieger, *et al*., 2000). On the other hand, a reduction of TryR expression was observed to decrease levels of $T(SH)_2$ which was associated with an increased sensitivity against arsenic and antimony-derived compounds (Ariyanayagam, *et al.*, 2005). Based on these results, TryR is considered ideal as a potential molecular target for drug design.

Table 1.1.: Protein sequence homology of TryR from different parasite species and its homologue in humans.

TryR sequence homology to *L. infantum* (JPCM5)	**Parasite Species**				**Homologous proteins**
	L. major (taxid:347515)	*L. donovani* (taxid:5661)	*T. brucei gambiense* (taxid:679716)	*T. cruzi* (CL Brener strain)	*Homo sapiens*
	96% (query cover 100%)	99% (query cover 100%)	67% (query cover 100%)	67% (query cover 99%)	Glutathione reductase: 35% (query cover 95%)

TryR is well conserved among all trypanosomatids (Table 1.1.), particularly its active site and overall fold, and so a single inhibitor drug could be developed to treat all three diseases caused by the *Trypanosomatidae* family. Several crystallographic structures of TryR have been deposited in the PDB for *T. brucei*, *T. cruzi* and *L. infantum*, in complex with natural substrates

and inhibitors (Table S1 in Supporting Information). Overall, TryR is a dimeric protein, being both its subunits identical and related by a twofold symmetry axis (Figure 1.7.). Each monomer is arranged in three different domains: the FAD-binding domain, the NADPH-binding domain and the interface domain (Zhang, *et al.,* 1996). In all TryR family members, the trypanothione-binding site is formed by residues of both the FAD-binding domain of one monomer and the interface domain of the second monomer. Therefore, TryR possesses two active sites and must be in a dimeric conformation in order to be functional. Its catalytic residues are two well-known cysteine residues (Cys52 and Cys57 in *L. infantum*). Upon NADPH binding and the consequent electron transfer to FAD, reduction of the Cys52-Cys57 disulfide bridge occurs by formation of a transient charge transfer complex between the flavin and Cys57 thiolate. Upon entry of TS_2 in the active site, Cys52 (previously deprotonated by His461'-Glu466' of the partner monomer), attacks the disulfide bridge of TS_2 nucleophylically, forming a mixed disulfide with it, which is resolved by the attack of Cys57 to Cys52. Finally, the formation of Cys52-Cys57 disulfide bridge and the release of reduced trypanothione takes place.

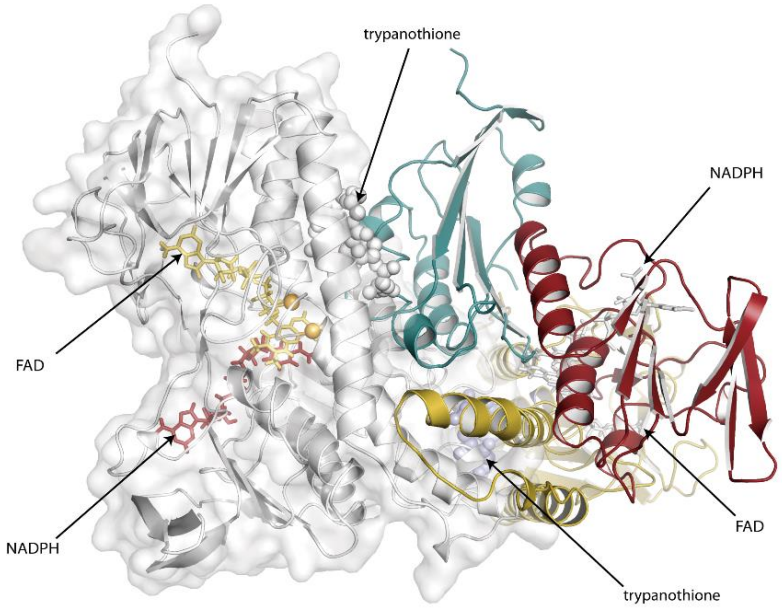


Figure 1.7.: General fold of trypanothione reductase (TryR). Although TryR of *L. infantum* in complex with FAD and NADPH is shown (crystallized in this Thesis), the overall fold is conserved for *T. cruzi* and *T. brucei*. The dimer is shown in cartoon: one monomer is colored in white with its corresponding surface and the second monomer is depicted by its domains: FAD-binding domain is colored in yellow, NADPH binding domain in red and the interface domain in blue. FAD and NADPH is represented as sticks, and only colored in yellow and red respectively in one of the monomers for clarity. Catalytic Cys52 and Cys57 are shown as orange spheres. Trypanothione is also represented in spheres.

TryR is localized in cytoplasm, whereas in *T. cruzi*, TryR is found in the cytoplasm and mitochondria although subsequent studies did not identify TryR in this organelle. Moreover, dual cytosolic/glycosomal localization could be due to the presence of C-terminal extension in trypanosomatid TryR protein which could act as a glycosomal targeting sequence (Singh, Garg and Ali, 2016).

1.2.4.1. Trypanothione Reductase vs. Glutathione Reductase

The TryR/$T(SH)_2$ system substitutes the GR/GSH present in the human host. Although the reduction of TS_2 to $T(SH)_2$ catalyzed by TryR is structurally and mechanistically similar to the reduction of GSH by GR in mammals (Fairlamb and Cerami, 1992) and share three conserved amino acids involved in catalysis (Cys52, Cys57 and His461') (Bond, *et al.*, 1999), TryR has been long considered a target of choice since the substrate binding site of GR displays structural features different to those of TryR. GR's active site is more hydrophilic and smaller than TryR's and it is positively charged due to Arg37, Arg38 and Arg347 that line the cavity to accommodate the negatively charged GSSH. Contrarily, TryR's active site is negatively charged, has a hydrophobic patch (comprised by residues Glu18, Trp21, Met113, Ile106, Ala342 and Ser109 in *L. infantum*), and has a larger substrate-binding pocket to fit its bulkier endogenous ligand (Lee, *et al*., 2005). Mutagenesis studies of TryR suggested the interaction of the hydrophobic region of the spermidine chain of $T(SH)_2$ with Trp21 and the positioning of the extra protonated amino group present in the spermidine component in the negatively charged active site (Cavalli and Bolognesi, 2009). This was later confirmed by the elucidation of the crystal structure of TryR of *T. brucei* and *T. cruzi* in complex with trypanothione (2WOW and 1BZL, respectively). All in all, $T(SH)_2$ promotes significant steric and electrostatic differences within the active site in relation to GR.

The differences above mentioned are exploited when targeting TryR selectively without having an undesired inhibition of GR. Nevertheless, some authors do not consider TryR a good drug target, since many TR inhibitors are not effective in the used infection model (Colotti, *et al.*, 2013) (i.e. the enzyme's and parasite's antiproliferative activities do not correlate directly). In this regard, Krieger *et al*. has explained this phenomenon in *T. brucei* by varying the activity of TryR. The evidence showed that trypanosomes containing less than 10% of wild-type enzyme activity were unable to grow and infect mice, and so it was concluded that more than 90% TryR inhibition is required to prevent parasite growth due to high amounts of trypanothione, which can displace a competitive inhibitor acting at low micromolar range (Krieger, *et al.*, 2000). High affinity and specific inhibitors for TryR are therefore needed. Furthermore, it should be considered that such an extended active site of TryR may require bulky inhibitors based on charge-charge interactions that do not show drug-like properties or, that could interact with the enzyme in many alternative modes, obscuring the binding prediction of diverse core-scaffolds and small drug-like molecules.

1.2.5. Trypanothione Reductase Inhibitors

In the last 30 years, more than 1500 TryR inhibitors have been studied. These can be classified into 6 different categories: i) tricyclic derivatives, ii) diphenylsulphide derivatives, iii) bicyclic and heterocyclic compounds, iv) polyamine derivatives, v) subversive substrates, vi) organometallic compounds, and (vii) irreversible inhibitors (Leroux and Krauth-Siegel, 2016).

1.2.5.1. Tricyclic Derivatives

The antimalarial drug mepacrine was the first tricyclic compound identified as an inhibitor of *T. cruzi* TryR (K_i 19 μM) which, did not affect human GR. As shown by the crystal structure of *T. cruzi* TryR in complex with this drug (PDB 1GXF), the acridine ring of this compound is fixed in the active center of the enzyme near the hydrophobic region formed by Trp21 and Met113, whilst the lateral alkylamine chain is fixed to the side chain of Glu18. Different residues are present in human GR, which explains the specificity of this compound (Jacoby, *et al.*, 1996).

Based on the structure of mepacrine, a series of 9-amino and 9-thioacridines were synthesized (Hammond, Hogg and Gutteridge, 1985; Bonse, *et al.*, 1999) Despite the close similarity, 9-aminoacridines proved to be mild competitive inhibitors (K_i =5-43 μM) with more than one molecule binding simultaneously, whilst 9-thioacridines were able to inhibit TryR by a mixed-type kinetics (Bonse, *et al.*, 1999). An improvement of the mepacrine solubility and TryR activity inhibition was achieved by the introduction of sulfonamides and urea moieties. These derivatives inhibit TryR with 10-fold lower IC_{50} values to mepacrine but are highly toxic due to GR cross-inhibition (Chibale, *et al.*, 2001).

In addition, neuroleptic drug chlorpromazine containing a tricyclic phenothiazine framework showed interesting inhibition of TryR (K_i 10.8 μM) (Chan, *et al.*, 1998). Further investigations using docking fragment methods revealed other tricyclic neuroleptics/antidepressants chlomipramine, amitriptyline and trifluoperazine as specific competitive TryR inhibitors with K_i values in the low micromolar range (i.e. chlomipramine K_i= 6.5 μM) (Benson, *et al.*, 1992).

Rational drug design was applied by Fairlamb *et al.* in order to understand the substrate specificity of GR and TryR and develop new phenothiazine tricyclic compounds of TryR (Benson, *et al.*, 1992). Accordingly, the hydrophobic pocket possibly represented by the Z-site (Phe396', Pro398' and Leu399'), not involved in the substrate binding was identified, with which bulky hydrophobic substituents interact (Benson, *et al.*, 1992). Quaternization of the tertiary

amino side chain of chlorpromazine with suitable hydrophobic groups to access the Z site, such as benzyl derivatives, reduced Ki values of 0.12 μM (Bernardes, *et al.*, 2013).

These compounds have been modelled in the active centre of *L. infantum* TryR: the cyclic core was located in the hydrophobic region of the active centre formed by Trp21 and Met113 and the lateral alkylamine chain towards the residues Glu466' and Glu467' of the other subunit of the enzyme (Khan, *et al.*, 2000). Other models suggest that the alkylamine side chain might interact, as in the case of mepacrine, with the Glu18 of the active site instead (Horvath, 1997). In addition, a different binding in which the cyclic core could be accommodated in the Z-site (Khan, *et al.*, 2000) has been proposed. In any case, these models may not be exclusive since, several inhibitory molecules could bind simultaneously.

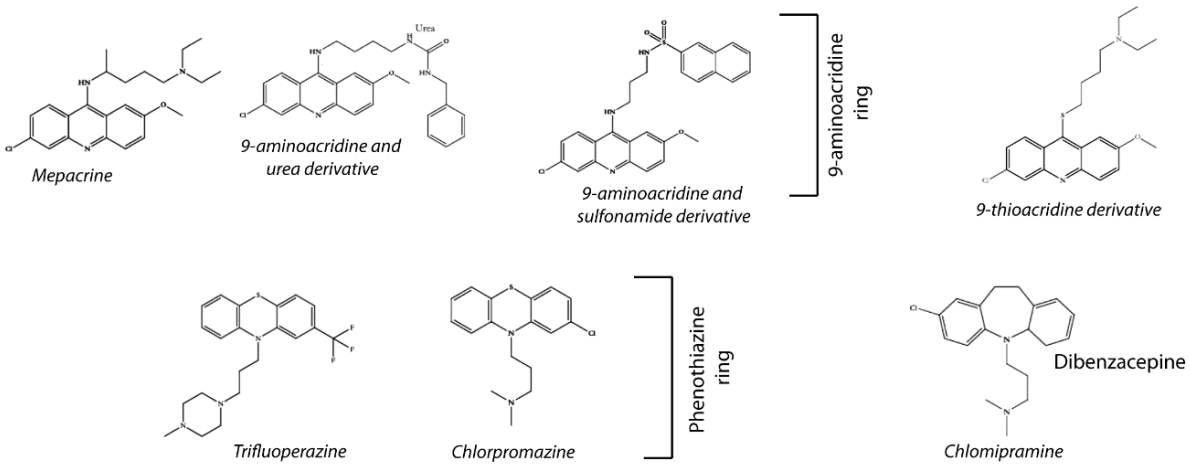


Figure 1.8.: Tricyclic derivative structures.

Novel druglike hits have been also identified from the screening of 1266 compounds, obtained from the Sigma-Aldrich LOPAC1280 library against TryR, (Richardson, *et al.*, 2009). Focusing on the profile of 1-(2-(benzhydryloxy)ethyl)-4-(3-phenylpropyl)piperazine (GBR-12935) (Figure 1.9.) against TryR (IC_{50}=10.9 μM) besides druglikeness, chemical accessibility, and resemblance to tricyclic derivatives, the tricyclic chlorpromazine was mimicked by cyclization of the diphenylmethane moiety. This provided the most active TryR inhibitors (Figure 1.9.) containing a strong binding phenylpropyl group, with IC_{50}=0.75-1.94 μM, respectively (K_i~0.33 μM), with no significant inhibition over human GR (> 100μM). D (Bernardes, *et al.*, 2013).

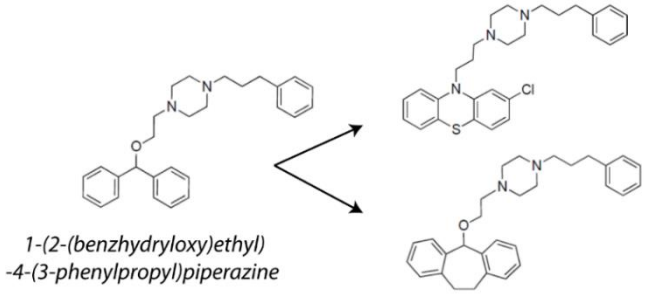


Figure 1.9.: TryR inhibitors containing a 3-phenylpropyl-piperazine moiety.

1.2.5.2. Diphenylsulfide Derivatives

In order to eliminate the primary neuroleptic effect of compounds based on chlorpromazine structure, a rationally designed "open-ring" analogue of chlorpromazine, such as the 2-aminodiphenylsulfide derivative in Figure 1.10. (A) (K_i 25 μM), was designed and shown to interact with the two carboxylate groups of glutamic residues (Glu466' and Glu467') and the hydrophobic pocket (Leu17, Trp21, Tyr110, Met113and Phe114), through its terminal N-methylpiperazine group and the two aromatic rings, respectively, by docking studies (Girault, *et al.*, 1998). These seem to be as powerful as their parent analogues.

The design of symmetrical analogues by introduction of additional polyamine moieties gave rise to competitive-type inhibitors (Figure 1.10. (B) (K_i 0.4 μM; IC_{50} 0.3 μM). Furthermore, the secondary amino group inserted in the polyamine moieties (C), allows different side chains to be introduced, generating multiple compounds with a high inhibitory power. Of these, the most effective is a mixed inhibitor with a IC_{50} value of 200 nM [Figure 1.10. (C)] in the presence of 57 μM of TS_2 (Krauth-Siegel, Bauer and Schirmer, 2005).

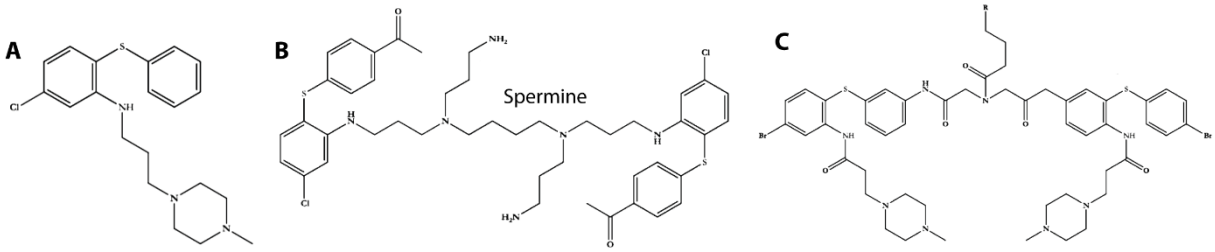


Figure 1.10.: Examples of 2-aminodiphenylsulfide inhibitors.

1.2.5.3. Bicyclic and Heterocyclic Derivatives

High throughput virtual screening of 62,000 compounds led to the identification of novel chemical lead structures related to quinoline or pyrimidopyridazine derivatives. These displayed a potent TryR inhibitory capacity (IC_{50} 1-3 μM) (Spinks, *et al.*, 2009). However, although a structure-function relationship was stablished for these compounds, their inhibitory capacity could not be improved. As drawbacks, quinolines have moderate effects on *T. brucei* parasites and pyrimidopyridazine have a high toxicity in human MRC5 cells, as it shows off-target effects (Spinks, *et al.* 2009).

Similar screening of 100,000 lead-like compounds allowed the identification of novel TryR inhibitor chemotypes with drug-like properties and antiparasitic activity. Five chemical classes were characterized including aryl/alkyl piperidines, basic benzhydryl, 1,2,4-triazine, quinazoline, conjugated indole and iminobenzimidazole. These compounds have IC_{50} ranging from 2 to 50 μM in *T. cruzi* TryR and a higher trypanocidal activity in *T. brucei* than in *T. cruzi*.

However, these compounds present a high toxicity in human MRC5 cells and also suggest off-target effects (Holloway, *et al.*, 2009).

In another screening of the LOPAC1280 library against TryR, 2 new development-prone inhibitors were identified along with the previously mentioned GBR-12935: (i) indatraline, (ii) the benzothiopheno-piperidine derivative BTCP (Richardson, *et al.*, 2009).

Regarding indatraline analogues, a relationship between the structure of the compounds and their activity against *T. cruzi* TryR was demonstrated (TryR IC_{50}=8.84 μM), although off-targets were also observed (Walton, *et al.*, 2011).

BTCP is a competitive inhibitor (K_i=1 μM), with greater activity against *T. cruzi* and *T. brucei* TryR (IC_{50}=3.7 and 3.3 μM respectively) over GR and shows growth inhibition of *T. brucei* ($EC_{50=}$13.6 μM) (Patterson, *et al.*, 2009). A library of 25 analogues of BTCP was synthesized to improve toxicity issues in rat myoblasts. Based on the inhibitory activity of BTCP analogues, the compounds in which the piperidine ring was replaced by pyrrolidine (IC_{50}=0.91 vs 3.3 μM) and the cyclohexyl ring was replaced by piperidine (IC_{50}=0.93 μM), showed limited improvement in the activity against TryR (Patterson, *et al.* 2009), despite unsuccessful trials for obtaining increased potency analogues. The atomic structures of one BTCP derivative (PDB 4NEW and 4NEV) shows to bind to the hydrophobic region of *T. brucei* TryR and *T. cruzi* TryR, just where mepacrine binds, but shows opposite orientations depending on the enzyme (Persch, *et al.*, 2014).

Further small molecule library screening for *L. infantum* TryR allowed to identify a thiomorpholine derivative that competitively inhibits TryR and kills the amastigotes. These compounds show a high toxicity profile in KB human carcinoma cells (Baiocco, *et al.*, 2013). The crystallographic structure of *L. infantum* TryR in its reduced state and in complex with NADPH and thiomorpholine derivative shows two inhibitor molecules in the active site as occurs with mepacrine (PDB entry 4APN).

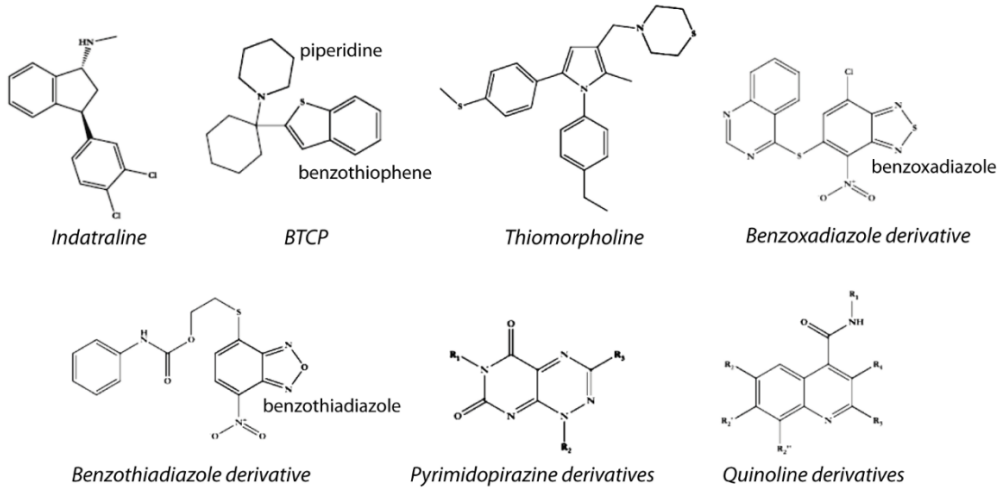


Figure 1.11.: Examples bicyclic and heterocyclic inhibitors. It is worth mentioning that GBR-12935 is also a bicyclic inhibitor (section 1.2.5.2.) although its optimization has led to a tricyclic compound.

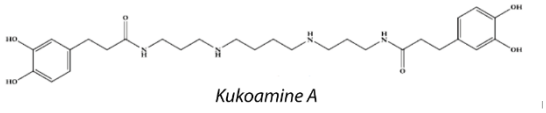
Finally, a combined *in silico/in vitro* approach led to the discovery of 82 *T. cruzi* TryR inhibitors. The most potent compounds present a benzoxadiazole or benzothiadiazole scaffold and K_i values lower than 1 μM with acompetitive inhibition mechanisms. However, the trypanocidal activity for bloodstream forms of *T. brucei* of these compounds is low. (Beig, *et al.* 2015).

1.2.5.4. Polyamine Derivatives

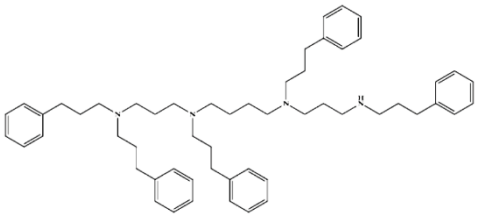
One of the first polyamine derivatives to be discovered was kukoamine A, a spermine derivative obtained from the root bark of solanacea *Lycium chinesea* and behaving as a mixed-type inhibitor of TryR (*Crithidia fasciculate*: K_i=1.8 μM) with no significant inhibition of GR (K_i> mM) (Ponasik, *et al.*, 1995).

The polyamine derivatives take advantage of the major difference between GSH and trypanothione. The approach led to powerful competitive inhibitors like N^1,N^1,N^4,N^8,N^{12}-penta(3-phenylpropyl)spermine and diphenylpropyl-polyaminobiguanides with Ki values for *T. cruzi* TryR of 150 and 950 nM, respectively (Li, *et al.*, 2001; Bi, *et al.*, 2006; Berdarnes, *et al.*, 2013).

Moreover, in order to improve the selectivity index of naphthoquinones and anthraquinones, conjugates with polyamines were synthesized. These were found to be non-competitive inhibitors of *T. cruzi* TryR with K_i 3 - 15 μM, which in some cases, could have a "subversive" behavior (Lizzi, *et al*., 2012). Although a clear structure-activity relationship was not stablished, other studies determine that TryR is most active with naphtoquinone derivatives that contain basic functional groups in side-chain residues. (Henderson, *et al*., 1998).



Kukoamine A



N^1, N^1, N^4, N^8, N^{12} - penta(3-phenylpropyl)-spermine

Figure 1.12.: Structures of polyamine derivatives.

1.2.5.5. Subversive Substrates

Subversive substrates are molecules which are prone to reduction by the transfer of a single electron from another reduced molecule. This electron would then be transferred to an oxygen molecule producing the superoxide radical that initiates ROS and RNS cascades. When the reduction of the subversive substrate is catalyzed by a reductase-disulfide-type antioxidant enzyme, such as TryR, these in turn are converted into a pro-oxidative enzyme. Hence, these substrates generate oxidative stress by (i) diminishing the reducing power within the cell (ii)

increasing ROS and RNS production and (iii) reducing the thiol: disulfide ratios (Krauth-Siegel, Bauer and Schirmer, 2005).

Inhibitors in this category involve 1,4-naphtoquinones such as menadione, plumbagine and lapachol. These compounds interact with TryR and GR. Nevertheless, these compounds show little subversive activity and are mainly reversible inhibitors (Krauth-Siegel, Bauer and Schirmer, 2005). In order to improve the specificity for TryR, derivatives where assayed, the most potent containing two 1, 4 naphtoquinones linked by a polyamine spacer (Figure 1.13.), which showed an increased specificity for *T. cruzi* TryR than human GR (Salmon-Chemin, *et al.*, 2001).

Recently, quinone-coumarin hybrids have been designed as dual inhibitors of *T. brucei* glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and *T. cruzi* TryR. Most of these compounds do not inhibit GAPDH maintaining their activity on TryR. However, the best derivative presents a K_i=2.3 μM for TryR, but shows a high toxicity in rat myoblasts. On the other hand, other less toxic compounds with a higher trypanocidal activity are worse inhibitors of TryR, suggesting off-target effects (Belluti, *et al.*, 2014).

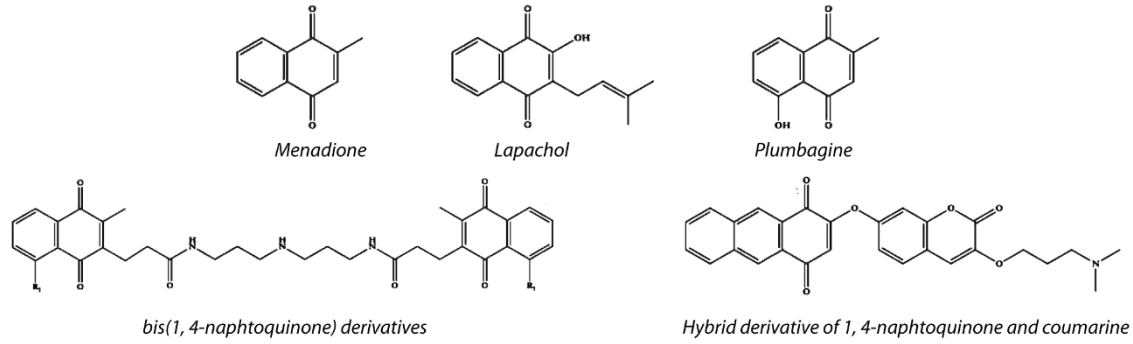


Figure 1.13.: Subversive substrates based on 1, 4-naphtoquinone scaffold.

1.2.5.6. Organometallic Compounds

Derivative complexes of (2,2':6',2´´-terpyridine)Pt(II) complexes are irreversible inhibitors of reduced *T. cruzi* TryR but not human GR. TryR irreversible inhibition is incubation time-dependent: thus, enzyme-inhibitor complexes incubated for short time periods recover the oxidorreductase activity, suggesting that these complexes are initially reversible but become irreversible over time. According to absorption spectra, this irreversible inhibition is probably due to the fixation of these organometallic complexes in TryR active site by Pt(II) coordination of to Cys52 and Cys57 (Bonse, *et al.*, 2000).

Palladium nitrofurylthiosemicarbazone complexes with formulas $PdCl_2(HL)$ and $Pd(L)_2$ might act through multiple mechanisms that combine DNA binding, production of oxidative stress and irreversible inhibition of TryR. These compounds inhibit *T. cruzi* epimastigote growth with EC_{50} values close to 3 μM (Bernardes, *et al*. 2013)).

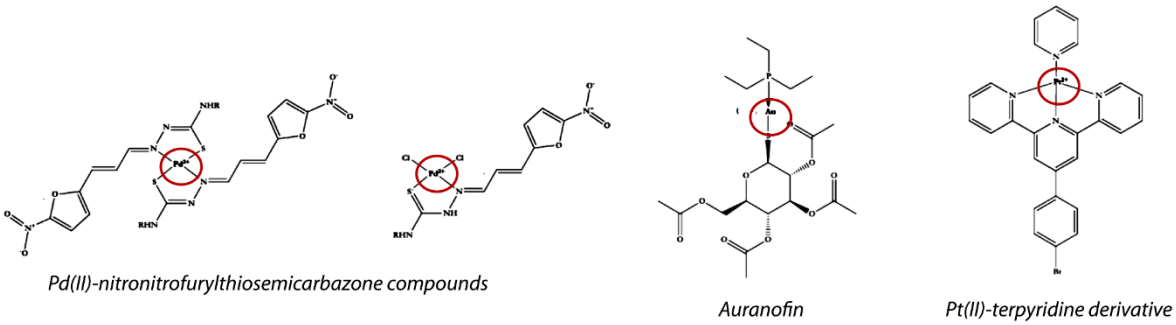


Figure 1.14.: Organometallic derivative examples.

Au(I) complexes such as Auranofin, (1-thio-b-d-glucopyranosato-(triethylphosphine) gold 2,3,4,6-tetraacetate), gold sodium thiomalate, gold thioglucose, and others, have been used for decades against rheumatoid arthritis (Colotti, *et al.*, 2013). This drug is also capable of inhibiting *L. infantum* TryR with an IC_{50} value below μM range. The crystal structure of TryR in complex with Auranofin (PDB entry 2YAU) showed that the Au(I) ion is coordinated by the catalytic cysteines (Cys52 and Cys57). Auranofin has also been shown to have leishmanicidal activity in three murine models of infection with *L. major* and, in some cases, has even been more effective than amphotericin B (Sharlow, *et al.*, 2014).

The mechanism of action of antimonial drugs (mentioned in section 1.2.2.) was determined by the crystal structure of *L. infantum* TryR in complex with Sb(III) ion (PDB entry 2WOH). Sb(III) coordinates Cys52 and Cys57 and His461' of the active site, thereby inhibiting TryR (Ilari, *et al.*, 2012). The ability of silver to inhibit TR is explained at the molecular level by the crystal structure of reduced *L. infantum* TryR in complex with NADPH and silver solved at 3.3 Å resolution (PDB entry 2X50). Similarly, silver inhibits the reduced TR by binding to the catalytic cysteines (Baiocco, *et al.*, 2011).

1.2.5.7. Irreversible Inhibitors

All the covalently-binding inhibitors of TryR that have been described to date modify the catalytic Cys52. The first irreversible inhibitor reported was nitrosourea carmustine, which also inactivates human GR. This inhibitor carbamoiles Cys52 of TryR and the conserved Cys58 of GR (Karplus, *et al.*, 1988).

Other irreversible and natural inhibitors involve ajoene, which also shows subversive properties (Gallwitz, *et al.*, 1999) for both GR and TryR, and an isothiocyanate glycoside isolated from *Moringa peregrine*, which is a weak irreversible inhibitor of *T. cruzi* TryR (Ayyari, *et al.* 2014). Several unsaturated Mannich bases, as well as natural macrocyclic alkaloids that also possess α, β-unsaturated ketones, like lunarine, inhibit TryR in a time-dependent manner (Bernardes, *et al.*, 2013). Lunarine is highly specific for TryR, as it doesn't inhibit GR activity at 500 μM, concentration at which 97% of TryR is inhibited (Hamilton, *et al.*, 2003). Furthermore,

ebsulfide was identified in a high performance screening and was shown to be an irreversible inhibitor of reduced TryR of *T. cruzi* TryR and *T. brucei*. This compound presents a potent trypanocidal activity in different subspecies of *T. brucei* and a low toxicity in mouse cells L929 (Lu, *et al*., 2013).

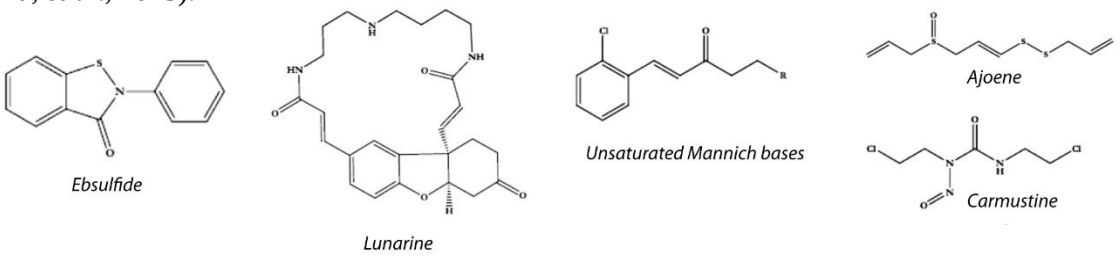


Figure 1.15.: Structures of irreversible inhibitors.

1.2.6. Novel TryR Inhibition Strategy

Most of the research up until today has focused on designing molecules directed at the active site of TryR. In light of the differences between this protein and GR, the dimerization interface of TryR was explored as an alternative target (Toro, *et al.*, 2013). In this way, hampering of TryR dimerization, and thus TryR activity, was proposed as a novel inhibition strategy.

Although the development of protein-protein interaction modulators is considered challenging due to interfaces usually lacking well-defined binding pockets (Wells and McClendon, 2007), significant contribution (>80%) of the binding energy is localized in small

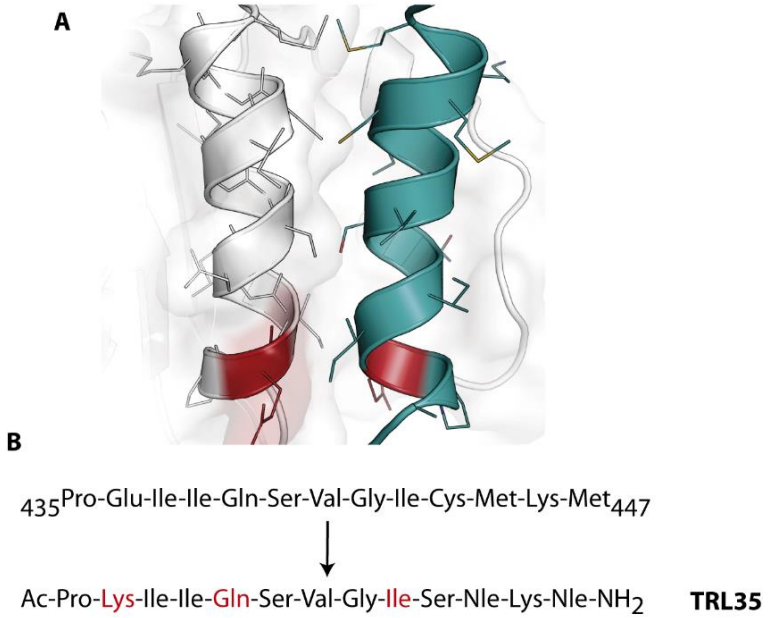


Figure 1.16.: Hot-Spot of *L. infantum* TryR dimerization interface and derived inhibitor peptide. (A) The pair of parallel helices in which Glu436 lies is shown, color coded to Figure 1.7. Glu436 is colored in red. **(B)** TRL35 peptidemimetic sequence is shown, derived from the original protein sequence. Crucial residues for dimerization activity of TRL35 are colored in red. Cys was substituted for Ser to avoid the possibility of redox-dependent disulphide formation with Cys444' in the complementary monomer. Initially, Glu was replaced by Lys in order to assess the importance of a possible interaction with a nearby E436' of the other subunit. Norleucine (Nle) is introduced as a more stable option instead of Met to avoid oxidation issues.

regions or "hot spots" (Thangudu, *et al*., 2012). Hence, validation of interface hot spots by point mutations showed Glu436 of *L. infantum* as a crucial residue for TryR dimer stability and function (Toro, *et al.*, 2013). This residue lies on a pair of parallel helices in the dimerization interface of TryR (Figure 1.16.A.).

Based on this α-helix, our colleagues at the System Biology Department at Alcalá de Henares University designed a 13-*mer* α-helix inhibitory peptide (named TRL35) with the ability to disrupt the dimer (Figure 1.16.B.) (Toro, *et al.*, 2013). This peptide is rapidly degraded by proteases and it is unable to penetrate *Leishmania* parasites. Thus, a battery of shortened peptidemimetics and three different chemical scaffolds were designed and synthetized at the Medical-Chemistry Institute, CSIC, Madrid (section 4.1.1.4.). These show closer drug-like properties and project functional groups that mimic the three essential substituents crucial for dimerization inhibition of TRL35 (Lys2, Gln5 and Ile9 of the peptide) as well as the amphipathic property of the original α-helix. Approximately 200 of the synthesized Trypanothione Reductase Ligands (TRLs) comprising diverse scaffolds have been analyzed for oxidorreductase activity inhibition and dimer disruption of *L. infantum* TryR at the System Biology Department at Alcalá de Henares University. Crystallization attempts regarding *L. infantum* TryR in complex with these inhibitors will be discussed in the following sections (Section 4.1.1.4.).

1.3. THE MANY ROLES OF THE EUKARYOTIC ELONGATION FACTOR 1A2

1.3.1. The "Canonical Function": Protein Synthesis

Protein synthesis is one of the most sophisticated biochemical processes in the cell and comprises the same steps in eukarya, bacteria and archaea: initiation, elongation and termination. Each phase requires the action of not only the ribosome (consisting of a 60S large subunit and a 40S small subunit which come together to form an 80S particle in the case of eukaryotes, as opposed to the 70S prokaryotic ribosome), mRNA and aminoacylated-tRNA (aa-tRNA), but also an entailment of soluble protein factors which strictly regulate each step. By contrast with prokaryotes, this mechanism is of considerable intricacy and complexity in higher eukaryotes.

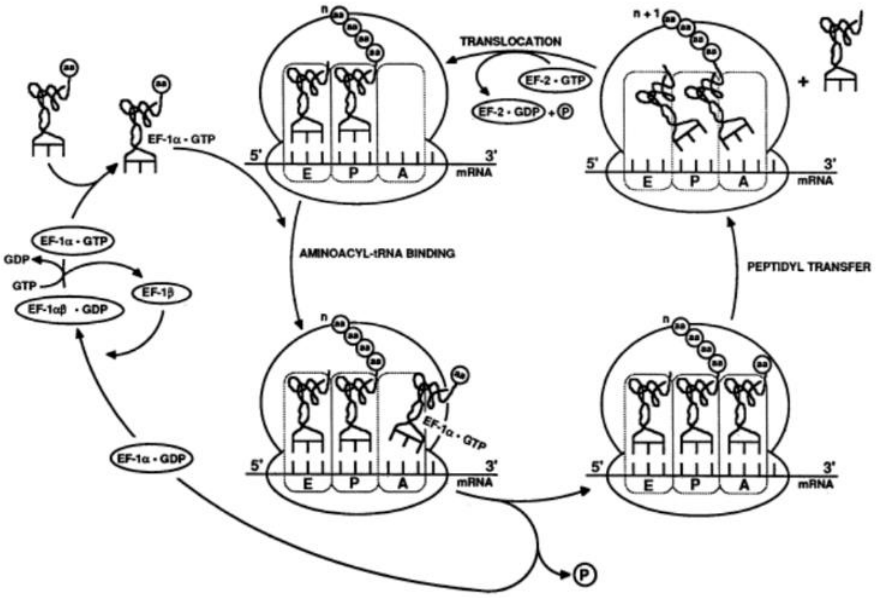


Figure 1.17.: Protein elongation cycle. Role of eEF1A, eEF1B and eEF2 in the first, second and third step of the elongation cycle, respectively. EF-1α, elongation factor 1A; EF-1β, elongation factor 1B; EF-2, elongation factor 2; A, P and E, ribosomal A, P and exit sites, respectively. Adapted from (Merrick, 1992).

Essentially, once translation is initiated at the first codon by the assembled 80S ribosome, it continues onto the elongation phase wherein the peptide chain increases its length cyclically one amino acid at a time (Figure 1.17.). During the translation elongation step, the elongation factor 1A (eEF1A) binds to, and delivers, the aa-tRNA to the A-site of the ribosome upon GTP-GDP hydrolysis. In order to be reactivated, eEF1A later interacts with a guanine nucleotide exchange factor (GEF), eEF1B, which facilitates GDP exchange to GTP. In bacteria, two homologous factors, namely EF-Tu and EF-Ts, are involved in making the appropriate aa-tRNA available to the elongating ribosome and exchange GDP to GTP, respectively. The eukaryotic elongation factor 2 (eEF2, the homolog of bacterial EF-G) acts as a translocase through GTP hydrolysis, allowing the ribosome to advance one codon down the mRNA and move the tRNA

from the A-site to the P-site and the deacylated tRNA from the P-site to the E-site of the ribosome. This translocation permits a new elongation cycle to occur. Peptide elongation continues until the ribosome reaches the stop codon, where point termination is triggered by a single release factor, eRF1, binding to the P-site of the ribosome. In the case of prokaryotes, this is triggered by two factors, RF1 and RF2. Finally, eukaryotic release factor 3 (eRF3), like its prokaryotic counterpart RF3, facilitates the release of both eRF1 and the completed polypeptide. (El'skaya, *et al.,* 1999; Berg, Tymoczko and Stryer, 2002).

In addition to Figure 1.17., another ternary complex (eEF1A·GDP·deacylated tRNA) has never been considered in widely recognized protein elongation schemes. Nevertheless, it had been hypothesized for higher eukaryotes and was later identified (Petrushenko, *et al.,* 2002), demonstrating a high affinity of eEF1A·GDP to deacylated tRNA in comparison to prokaryotic systems. Moreover, a "hand-to-hand" transfer of aa-tRNA from *aminoacyl*-tRNA synthetases (*A*RS) to eEF1A and vice versa (eEF1A transferring deacylated tRNA to ARS) was confirmed by obtaining the quarternary complexes [*Phe*RS·eEF1A·GDP·$tRNA^{Phe}$] and [*Ser*RS·eEF1A·GDP·$tRNA^{Ser}$] (Petrushenko, *et al.,* 2002). This implied that protein biosynthesis in mammalian cells is a channeled pathway in which aa-tRNA is directly transferred from ARS to eEF1A without dissociation into cellular fluid (Negrutskii, Stapulionis and Deutscher, 1994).

1.3.1.1. A Structural Overview on eEF1A and Nucleotide Exchange

The structures of eEF1A and EF-Tu are well known and structurally conserved regardless of their source organism, with over 100 entries in the PDB obtained by X-ray crystallography or electron microscopy when present in the ribosomal complex. Their structure comprises three domains: The N-terminal domain (domain I), responsible for nucleotide binding, and domains II and III, which fold into two β-barrels, all connected by linker sequences (Figure 1.18.). It is well known that this G-protein exhibits two different conformations whether it is bound to GDP or GTP. Herein, domains II and III rotate approximately 90º with respect to domain I, forming new intra- and inter-molecular interaction surfaces.

Eukaryotic EF1A and EF-Tu are classic G-proteins that act as a "molecular switch" for the active and inactive states based on whether GTP or GDP is bound (Bourne, Sanders and McCormick, 1991). Because the intrinsic rate of GDP release from eEF1A is extremely slow, a guanine nucleotide exchange factor (GEF) complex, eEF1B, is required (Pittman, *et al.,* 2006).

The molecular mechanism of this nucleotide exchange and the conformational changes implied have been characterized for EF-Tu (Kawashima, *et al.,* 1996). Nevertheless, the co-crystal structures of yeast eEF1A·GDP and eEF1A·GDPNP in complex with the C-terminus of eEF1B (Andersen, *et al.,* 2000; Andersen, *et al.,* 2001; PDB entries 1F60 and 1G7C, respectively)

revealed a surprising structural divergence from the *E. coli* EF-Tu-EF-Ts (Kawashima, *et al*., 1996, PDB entry 1EFU) and the mammalian mitochondrial EF-Tu-EF-Ts complex (Jeppesen, *et al.*, 2005; PDB entry 1XB2). Despite their similar topology, eEF1A and EF-Tu show different surfaces through which they interact with their GEFs. Hence, the C-terminus of eEF1B interacts with domain I and a distinct pocket of domain II of eEF1A, creating two binding interfaces. In contrast, the bacterial counterpart EF-Ts and mammalian mitochondrial EF-Ts, make extensive contacts with domain I and III of EF-Tu (Pittman, *et al.*, 2009).

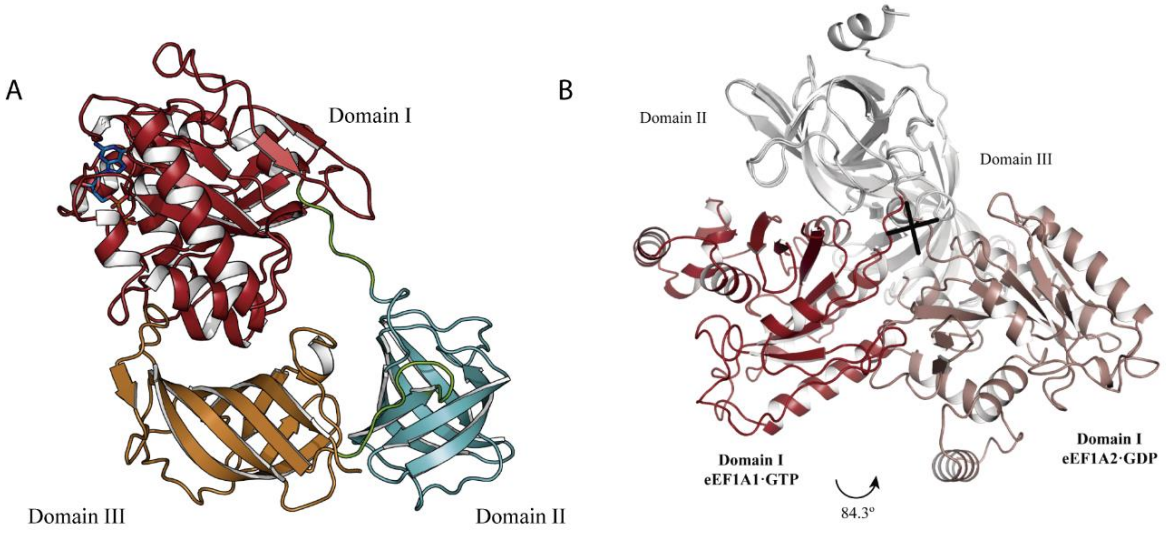


Figure 1.18.: eEF1A general fold and GDP-GTP conformational changes. (A) Conserved general structure of eEF1A·GDP, represented in cartoon. Domain I is colored red, domain II in blue and domain III in yellow. Connecting loops are colored in green. Nucleotide is drawn in stick form. (B) Structural rearrangement of eEF1A upon GTP hydrolysis. Domains II and III are fixed due to clarity issues. Rotation of 84.3º occurs through an axis perpendicular to the sheet of paper, indicated by a black cross. Both A and B structures correspond to eEF1A2·GDP purified from rabbit muscle and crystallized during this Thesis, whilst eEF1A1·GTP from rabbit muscle was extracted from PDB entry 5LZS.

Current models of eEF1A nucleotide exchange suggest that (i) eEF1A has no preference for the type of nucleotide as the equilibrium dissociation constants for GDP ($10x10^{-7}$ M) and GTP ($7x10^{-7}$ M) are very similar (Saha and Chakraburtty, 1986); (ii) the action of eEF1Bα on eEF1A·GTP regeneration is the rate-limiting step in translation elongation (Janssen and Möller, 1988); and (iii) the rate of spontaneous GDP dissociation eEF1A is very slow (shown in yeast) and so it is accelerated 700-fold by eEF1B (Pittman, *et al.*, 2006).

In mammals, a refined molecular nucleotide exchange mechanism has been proposed (Crepin, *et al.*, 2014) based on the reported eEF1A2 structure purified from rabbit muscle which is 100% identical to the human protein (PDB entry 4C0S; see section 1.3.2.), the previously mentioned structures 1F60 and1G7C (see section 1.3.1.1.) and biochemical essays (Pittman, *et al.*, 2006). Herein, switches I and II, the P-loop and the NKXD element are shared within all G-domains of GTP-ases (Ozturk and Kinzy, 2008). Switches I and II are two flexible loop regions (respectively, residues 67-78 and 90-96 in eEF1A2) (Figure 1.19.) which undergo conformational changes triggered by the insertion of Lys205 of eEF1B which break the interactions of the P-loop

with GDP. Among eEF1B binding, Asp91 in switch II of eEF1A moves to form a salt bridge with Lys20 in the P-loop, initially interacting with the β-phosphate of GDP, leading to its destabilization, and so favouring nucleotide exchange.

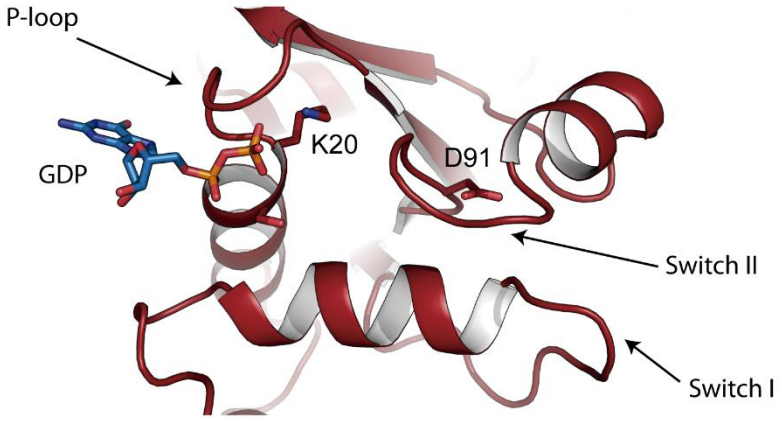


Figure 1.19.: Conserved structural features of G-proteins' G-domain involved in nucleotide exchange. Domain I of eEF1A2 is represented in red cartoon, and GDP is shown in blue sticks. Switches I, II and P-loop are indicated by arrows. Lys20 and D91 are shown in sticks.

In general, Mg^{2+} is important for high affinity binding of nucleotides to proteins (Pittman, *et al.*, 2006). This is true for EF-Tu, in which the interaction of Mg^{2+} with the nucleotide and protein is disrupted by the insertion of the GEF residues (Ozturk and Kinzy, 2008). However, it seems that for eEF1A nucleotide exchange does not involve Mg^{2+}, which explains the similar affinity of eEF1A for GDP and GTP. Although inconsistent with the findings of Mg^{2+} dependent nucleotide exchange, it appears that its binding effect is rather indirect and only results in subtle rearrangements of the GDP-binding site (Schummer, Gromadski and Rodnina, 2007; Crepin, *et al.*, 2014).

1.3.2. eEF1A1 and eEF1A2: Two Isoforms of eEF1A

Mammalian eEF1A occurs in two 98% similar, 92% identical isoforms, namely eEF1A1 and eEF1A2. They are encoded by different genes with different expression patterns which are tissue and development-specific in vertebrates. The expression of the isoforms is mutually exclusive: eEF1A1 is expressed in all tissues throughout the development but is absent in adult muscle and heart. The latter tissues express eEF1A2 instead, as do other cell types, including large motor neurons, islet cells in the pancreas and enteroendocrine cells in the gut (Lee, *et al.*, 1992; Chambers, Peters, and Abbott, 1998). The eukaryotic translation elongation factors are the second most abundant protein (1-3% of total protein content) after actin (Abbas, Kumar and Herbein, 2015). Despite its abundance in the cell, the importance of maintaining the control of eEF1A levels is reinforced by the oncogenic properties of eEF1A2, whose upregulated expression in non-

canonical tissues has been related to breast, ovarian, pancreatic, hepatic and lung cancer (Table 1.2.).

These isoforms show different dissociation rate constants for GDP: it is seven times higher for eEF1A1 than for eEF1A2. In addition, the nucleotide preference ratio (GDP/GTP) for eEF1A1 is 0.82 and for eEF1A2 is 1.50. (Abbas, Kumar and Herbein, 2015).

Table 1.2.: Expression of eEF1A2 in different human cancers. (Adapted from Abbas, Kumar and Herbein, 2015).

Cancer	Relevant Findings	Reference
Breast cancer	eEF1A2 mRNA/protein is high expressed in 50-60% in primary human breast cancer.	Tomlinson, et al., 2005; Kulkarni, et al., 2007
Ovarian cancer	eEF1A2 is highly expressed in 30% of primary ovarian tumors.	Anand, et al., 2002; Pinke, et al., 2008
Lung cancer	Positive Ki-67 expression associated with positive eEF1A2 and KCIP-1.	Li, et al., 2006; Zhu, et al., 2007
Liver cancer	eEF1A2 is highly expressed in half of hepatocellular carcinoma.	Grassi, et al., 2007; Schlaeger, et al. 2008
Pancreatic cancer	eEF1A2 is highly expressed in 83% of pancreatic cancers.	Cao, et al., 2009; Xu, Hu and Zhu, 2013

1.3.2.1. eF1A2 and Oncogenesis

A third pattern of expression belongs to certain tumor cell types and cell lines that express both eEF1A isoforms. The expression of eEF1A2 is strongly upregulated in most of breast tumors: high levels are detected in 60% of primary breast tumors and metastases, but not in normal epithelium (Tomlinson, *et al.*, 2005). The expression of eEF1A2 stimulates the formation of filopodia, cell migration and invasion by Akt and PIK-dependent cytoskeleton remodeling (Amiri, *et al.*, 2007), favoring cellular transformation and oncogenesis.

eEF1A2 gene is also highly expressed in ovarian cancer (Tomlinson, *et al.*, 2007) and it has been suggested that it could favor the development of ovarian primary tumor formation (Lee and Surh, 2009) and poor prognosis. Studies of the oncogenic properties of eEF1A2 in these cancers conclude in the favoring of anchorage-independent growth and increased rate of cellular proliferation (Anand *et al.*, 2002).

Increased pattern of expression has also been observed in lung cancer cells. Comprehensive studies of lung cancer cell lines show that eEF1A2 is a putative oncogene whose expression is correlated with increased DNA copy number and transcript levels, along with KCIP-1 protein. Suppressed cellular proliferation and increased apoptosis rate was observed when the expression of these proteins was modulated by siRNA, suggesting their potentiality as therapeutic targets (Li, *et al.*, 2006).

Upregulation of eEF1A2 gene has also been observed in pancreatic, liver and more recently, gastric cancer. Silencing this gene reduces cell viability, proliferation and increases the apoptosis rates in hepatocellular carcinoma cell lines (Schlaeger, *et al*., 2008). Eighty-three percent of pancreatic cancers display increased expression of eEF1A2, suggesting its important role in pancreatic carcinogenesis (Xu, Hu and Zhu, 2013). This quality has been suggested in order to develop early diagnosis tools (Li, *et al.*, 2004). Regarding gastric cancer specimens, eEF1A2 had significantly increased expression at both mRNA and protein levels (Yang, S. *et al.*, 2015).

1.3.3. Functionality of EF1A: A Moonlighting Protein

A number of processes involving eEF1A, besides the specific function of eEF1A in translation, have also been described (Figure 1.20.). Hence, this factor is acknowledged as a moonlighting protein participating in many important cellular roles throughout prokaryotes and eukaryotes.

1.3.3.1. Macromolecular Partners of eEF1A and Other Non-Canonical Functions

The important aspect for clarifying how the isoform-specific functions of eEF1A1 and eEF1A2 are controlled in mammalian cells is the understanding of whether these isoforms are able to bind different protein targets, or the same target but with different affinity. The potential interaction dissimilarity is suggested by the difference in spatial structures between the isoforms: herein, eEF1A1 was shown to have an open, more relaxed structure, while eEF1A2 has a more compact form (Novosylna, *et al.*, 2007; Timchenko, *et al*., 2013). Furthermore, eEF1A1 has been shown to be more hydrophobic than eEF1A2 (Timchenko, *et al*., 2013) and molecular dynamic simulations described specific regions which could be different in the two isoforms and consequently contribute to the different ability to interact with protein partners (Kanibolotsky, *et al*., 2008).

These differences were shown to correlate with diverse ability to interact with a number of partners. Most predominantly, eEF1A1 is able to form a strong complex with calmodulin in the presence of Ca^{2+} whereas eEF1A2 was not (Novosylna, *et al.*, 2017). It has been shown that tRNA and calmodulin compete for eEF1A1 binding (Novosylna, *et al.*, 2017). Moreover, eEF1B binds to the same surface of eEF1A as tRNA (Andersen, et al., 2000), suggesting a calmodulin-dependent regulation of protein translation in tissues where eEF1A1 is expressed and a mechanism of protection of the protein synthesis process from the sharp and permanently occurring changes in Ca^{2+} concentrations in some specialized eEF1A2-specific cells (Negrutskii, *et al.*, 2018).

It has been over 20 years that actin was shown to be a binding partner of eEF1A1 (Edmonds, et al., 1996), although the interaction of F-actin with eEF1A1 and eEF1A2 was recently proved by the formation of bundles of F-actin (Novosylna, *et al.*, 2017), which was modulated in a Ca^{2+}/calmodulin dependent manner for eEF1A1. As actin cytoskeleton shown significant perturbations in cancer cells), eEF1A2 may contribute to the oncogenic transformation of cells (Stevenson, Veltman and Machesky, 2012. Furthermore, it has been shown that F-actin interacts with domains I and III of *Dictyostelium* EF1α (Liu, *et al.*, 1996) and domain II of human eEF1A from an epidermoid cancer cell line (Lamberti, *et al.*, 2008).

The participation of eEF1A1 and eEF1A2 in the phosphotyrosine-mediated processes has been proposed, and the ability to interact with SH2 and SH3 domains of different signaling molecules *in vitro* has been confirmed (Panasyuk, et al., 2008). Thus, contrary to eEF1A1, phosphotyrosine-containing sites in domain I of eEF1A2 was able to interact with SH2 domains of Grb2, RasGAP, Shc and Shp2, as well as with SH3 domains of Crk, Fgr and Fyn. Moreover, both eEF1A1 and eEF1A2 formed complexes with the SH2 domain of PLCγ via the carboxyl region of eEF1A (Panasyuk, *et al.*, 2008). Many components of signal transduction are mobilized to actin filaments, probably for the efficient relay or cross-talk of various signals (Ejiri, 2002). This data allows the belief that the translation function of eEF1A might be combined with its involvement in various signaling pathways.

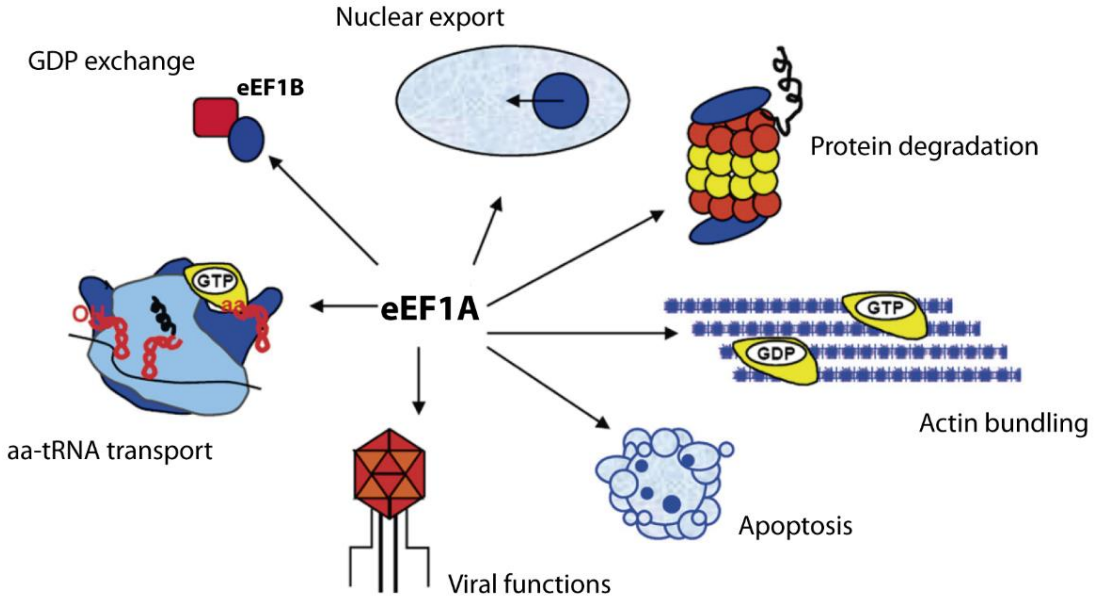


Figure 1.20.: Canonical and some non-canonical functions attributed to eEF1A. (Adapted from Mateyak and Kinzy, 2010)

Other processes in which eEF1A is involved are apoptosis, viral replication, regulation of oxidative stress, proteasome-dependent degradation of ubiquitin-conjugated proteins, regulation of cell cycle progression and nuclear export, among others (Ejiri, 2002). Regarding viral propagation, there are reports on the participation of eEF1A in the control of HIV assembly,

in which eEF1A has been shown to interact with the HIV reverse transcriptase and the HIV-I Gag polyprotein assembled in F-actin filaments (Cimarelli, *et al.,* 1999; Li, *et al.,* 2015). In other viruses, such as the TBSV, the tobacco mosaic virus and the West Nile virus, eEF1A binds both the viral RNA and the viral dependent RNA polymerase, and results suggest that binding of eEF1A to the viral template is important for viral replication (Davis, *et al*., 2007). In general, positive strand viruses may have evolved to utilize eEF1A in order to promote viral propagation through multiple mechanisms (Mateyak and Kinzy, 2010). Moreover, it was shown that eEF1A1 readily forms a complex with Sgt1, a multifunctional protein involved in the anti-viral protection of an organism limiting the viral multiplication. On the other hand, eEF1A2 does not, but shows an increased affinity to viral RNA suggesting a strong pro-viral action of eEF1A2 in a non-dependent Ca^{2+} manner (Novosylna, *et al.*, 2015).

Apoptosis or programmed cell death is a highly regulated series of cellular events that lead to the elimination of damaged or unnecessary cells. It has been shown that eEF1A2 expression correlates with differentiation and has a protective effect against apoptosis, whereas expression of eEF1A1 has the opposite effect (Ruest, Marcotte and Wang, 2002). Moreover, the expression of eEF1A is up-regulated by the pro-apoptotic transcription factor p53, promoting cell death by microtubule-severing in erythroleukemic cell lines (Kato, 1999), and induced in proportion to the concentration of hydrogen peroxide, suggesting its implication in oxidative stress-induced apoptosis (Chen, et al., 2000).

Observations and speculations have been made on "protein synthesis and degradation may be regulated by a common factor, eEF1A". eEF1A has been proposed to be a good candidate for recognizing damaged proteins and shuttling them to the proteasome for degradation. Initially, it was identified as a factor required for the degradation of N-α-acetylated proteins (Gonen, *et al.,* 1994), although later, experiments demonstrated that eEF1A could directly interact with nascent polypeptides while they are being synthesized to help mediate refolding (Hotokezaka, *et al.,* 2002), showing chaperone-like properties.

Several reports have linked eEF1A to nuclear export of aa-tRNA to the cytoplasm (for a review, Sasikumar, Perez and Kinsy, 2012). Although mainly in the cytoplasm, it has been demonstrated that eEF1A could be detected in the nucleus in *S. cerevisiae*, suggesting a role for eEF1A on the nuclear side of the membrane (Murthi, *et al*., 2010). eEF1A was also shown to exhibit an aa-tRNA-dependent binding to the nuclear export machinery in mammalian cells by interacting with VHL and PABP1 proteins through the binding to the transcription-dependent nuclear export motif (TD-NEM) (Khacho, *et al.,* 2008). Because eEF1A binds aa-tRNA in its canonical role, this is really a part of efficient channeling of protein synthesis components (Negrutskii, Stapulionis and Deutscher, 1994).

1.3.3.2. Regulation of eEF1A by Post-Translational Modifications

eEF1A is known to be extensively post-translationally modified. Post-translational modifications (PTMs) of proteins are biochemical modifications that increase the proteome's functional diversity by regulating activity, localization and interaction with other cellular molecules. Such modifications include a wide variety of types, and are mostly catalyzed by enzymes that recognize specific target sequences or folds in specific proteins. eEF1A is subject to different kind of modifications such as phosphorylation, methylation, acetylation, ubiquitination, glucosylation, carbonylation, S-nitrosylation, S-glutathionylation and attachment of ethanolamine phosphoglycerol (EPG) (Soares and Abbott, 2013). These PTMs provide a regulatory switch for other functional roles, enabling eEF1A to interact with other molecular partners in different cellular processes.

The precise roles and locations of these modifications remain unclear. Nevertheless, the linkage of certain PTMs to eEF1A structure and function have been reported, such as the stability of A-A' helices (residues 20-31 and 57-66 in mammalian eEF1A) and impairment of translation by Tyr29 phosphorylation (Negrutskii, *et al.*, 2018), structural rearrangement by phosphorylation of Tyr85 and Tyr86 involving the pro-apoptotic action of eEF1A1 in gastric cancer (Negrutskii, Vlasenko and El'skaya, A., 2012), maintenance of an extended conformation and disruption of the interaction between domain I and the domain II-III body of eEF1A1 via Tyr141 phosphorylation (Negrutskii, Vlasenko and El'skaya, A., 2012), modulation of interaction with mRNA and other binding partners by phosphorylation of Tyr418 (Fan, *et al.*, 2009), cellular apoptosis through C-Raf by phosphorylation of Ser21 (Sanges, *et al.*, 2012), downregulation of mRNA translation and cell proliferation through TGF-β in eEF1A1 due to the phosphorylation of Ser300 (Lin, et al., 2010), and translation regulation through dynamic methylation (Jacobsson, Melcki and Falnes, 2018), among others. Less frequent phosphorylation of Tyr162, Tyr254 and Try418 have been proposed to be likely important for regulation of interaction of eEF1A with some protein partners as they lie on the surface of the protein (Negrutskii, Vlasenko and El'skaya, A., 2012), whilst phosphorylation of Ser21 in eEF1A could also be involved in protein dimerization, crucial for actin bundling (Sanges, *et al.*, 2012).

The possible regulatory relevance of the methylation/acetylation and phosphorylation/acetylation switches have been proposed in cancer signaling networks (Grimes, *et al.*, 2018), as different modifications have been identified for a single residue. Methylation of each site is performed by different methyltransferases (Jakobsson, *et al.*, 2018), being a modulating phenomenon (Wang, *et al.*, 2015) due to its dynamism (i.e. a mixture of non-methylated, mono-, di- and tri-methylated residues may occur simultaneously) (Jakobsson, *et al.*, 2017). In addition, methylation of Lys36, Lys55 and Lys165 of eEF1A have been associated to

changes in ribosome biogenesis and influenced translation rates for some codons and proteins related to tRNA aminoacylation (for a review, Negrutskii, *et al.*, 2018).

On the contrary, EPG attachment to conserved glutamate residues of mammalian and plant eEF1A (described for Glu301 and Glu374 (Rosenberry, *et al.*, 1989)) has not been associated to any specific function, despite the fact it is a unique modification of an essential protein involved in a highly conserved process, i.e. protein translation. This modification is absent in archaea (Greganova, Altmann and Bütikofer, 2011), *E. coli* (Whiteheart, *et al.*, 1989) and *S. cerevisiae* (Cavallius, *et al.*, 1993), representing the only eukaryote lacking this modification. Regarding trypanosome eEF1A, only Glu362 in domain III has been shown to be modified by EPG in *T. brucei* (Signorell, *et al.*, 2008), in which it was proved that this modification is dependent on the three dimensional structure of domain III rather than the amino acid sequence flanking Glu362 (Greganova, Heller, and Bütikofer, 2010). A model for the attachment of EPG to the elongation factor has been proposed (Signorell, *et al.*, 2008; Mittal, *et al.*, 2013) comprising the initial modification of eEF1A by phosphatidylethanolamine (PE) and then the deacylation to EPG (Figure 1.21.).

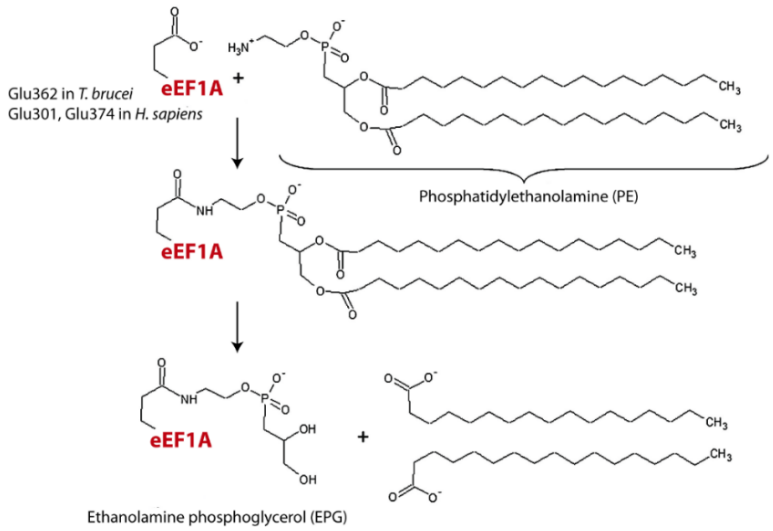


Figure 1.21.: Post-translational modification of eEF1A showing EPG formation. Proposed pathway of EPG attachment involving modification of eEF1A by phosphatidylethanolamine and subsequent deacylation to EPG (adapted from Mittal, *et al.*, 2013).

A previous X-ray crystal structure for rabbit eEF1A2 bound to GDP has been published (PDB code: 4C0S), but only two PTMs were there described (phosphorylation in Thr239 and Ser163). The characterization of PTMs will shed light in the understanding of the molecular basis of the multifunctional behavior and interactions of eEF1A underlying etiological processes. Different PTMs in isoforms eEF1A1 and eEF1A2 may also contribute to their functional and structural inequality, as a more relaxed conformation has been observed for eEF1A1 contrary to eEF1A2 (Novosylna, *et al.*, 2007, section 1.3.3.1.) and different binding partners have been identified for each (Panasyuk, *et al.*, 2008; Negrutskii, Vlasenko and El'skaya, 2012). Hence, a

thorough knowledge of these alternate PTM patterns is crucial for appreciating their diverse biological relevance and how they may lead to eEF1A's "moonlighting" quality.

1.3.4. eEF1A as a Natural Product Drug Target in Cancer Therapy

From long-known antibiotics targeting EF-Tu such as Pulvomycin, Elfamycin, Kirromycin, Enacyloxin and GE2270, to novel antiretrovirals binding eEF1A (Rawle, *et al.*, 2019), translation elongation factors have been proposed as drug targets. As regards of eEF1A2, which reportedly has pro-oncogenic qualities (see section 1.3.2.1.), at least four classes of structurally distinct natural products with potent antiproliferative activities have been shown to target the closely related eEF1A1, namely didemnin B, ternatin, nannocystin A, and ansatrienin B.

Didemnin B, an antitumor agent derived from the marine tunicate *Aplidum albicans* previously demonstrated to "freeze" the protein translation elongation step (Ahuja, *et al.*, 2000), was later shown bound to the rabbit eEF1A1 in the GTP conformation within the elongating ribosomal macromolecular complex by cryo-EM (Shao, *et al.*, 2016, PDB entry: 5LZS), allowing the interpretation the inhibition of eEF1A's canonical function. It has been proposed that all of these natural products share a binding site at a hydrophobic cavity formed between domains I and III in the GTP-conformation of eEF1A1 (Sánchez-Murcia, Cortés-Cabrera and Gago, 2017). Binding of plitidepsin (i.e. dehydrodidemnin B) to this same site in the closely related eEF1A2 is believed to stabilize the "active" conformation of eEF1A2 (Losada, et al., 2016), perhaps by preventing dimer formation. (Sánchez-Murcia, Cortés-Cabrera and Gago, 2017).

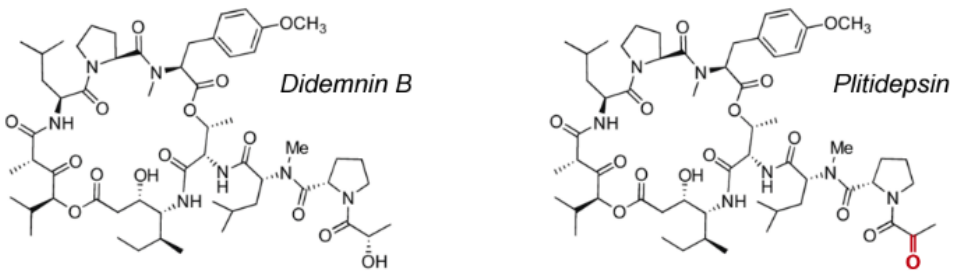


Figure 1.22.: Chemical structure of didemnin B and plitidepsin.

In this regard, the Jun N-terminal kinase and P38 pathways have been shown to be involved in plitidepsin-induced apoptosis in multiple myeloma (Mitsiades, et al., 2008) and double-stranded RNA-activated protein kinase (PKR) has been recently unveiled as a novel eEF1A2-interacting partner whose pro-apoptotic effect is hindered by eEF1A2 sequestering and inhibition of its kinase activity. Targeting eEF1A2 with plitidepsin hampers the formation of the PKR-eEF1A2 complex, facilitating PKR activation and triggering a mitogen-activated protein kinase signaling cascade together with a nuclear factor-κB-dependent activation of the extrinsic apoptotic pathway that leads to tumor cell death (Losada, et al., 2018).

1.4. PHAGE THERAPY AND BIOTECHNOLOGICAL APPLICATIONS OF ENDOLYSINS

1.4.1. Phage Therapy in the Post-Antibiotic Area

Antibiotic resistance is arguably the biggest current threat to global health: over 10 million yearly deaths have been estimated to occur by 2050 due to antimicrobial resistance. This crisis calls for urgent development, standardization and implementation of new therapeutic strategies against infectious diseases, thus, there is a slow but substantial change in the appreciation of phage therapy (PT) and phage-derived proteins. Phage therapy is defined as the administration of virulent phages directly to patients with the purpose of lysing the bacterial pathogen that is causing the infection (Vierte, Ritter and Horz, 2014). Phages (bacteriophages) are viruses that invade bacterial cells. They are ubiquitous, obligate parasites that are highly specific to their host, sometimes at serovar levels (Hermoso, García and García, 2007).

The formal history of phages began with Ernest Hankin, when in 1896 reported the presence of an antibacterial activity against *Vibrio cholerae* (Duckworth, 1976) and with the further work of Hankin, Gamaleya, Twort and d'Herelle (Chanishvili, 2012), who first coined the term "bacteriophage", literally meaning "bacterium eater". d'Herelle used phages to treat dysentery in 1919 and the first reported application of phages to treat infectious diseases in humans came shortly after by Bruynoghe and Maisin, who used bacteriophages to treat a staphylococcal skin disease (Chanishvili, 2012). The use of phage therapy was discontinued after the introduction of antibiotics in the 1940s, and little data was published since. Research regarding the therapeutic use of phages has been somewhat neglected until the past two decades, when the increasing incidence of antibiotic resistant bacteria awoke a renewed interest in phage therapy, and the use of phage proteins to combat pathogenic bacteria (Figure 1.22.).

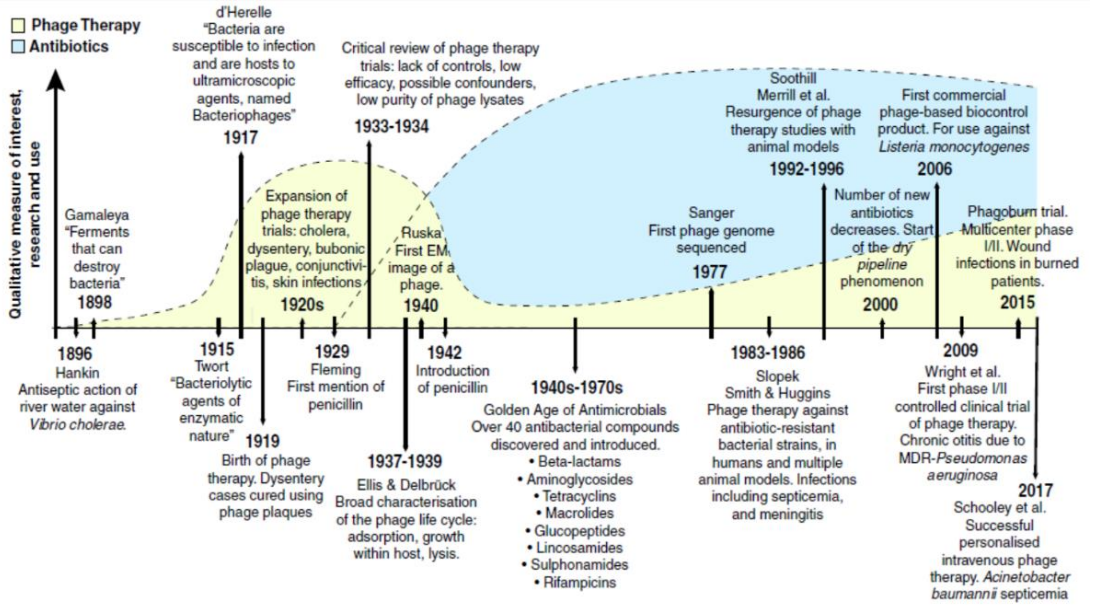


Figure 1.23.: Timeline of major events in phage research, phage therapy and antibiotics. (From Gordillo-Altamirano and Barr, 2019).

1.4.2. *À la Carte* Bioengineered Endolysins

Conventional approaches involve using whole phages as therapeutic agents. Novel tactics involve conjugating biology and engineering principles to produce chimeric phages, thus increasing their therapeutic potential via a range of mechanisms, or the combination of phages and antibiotics (for a review, Gordillo-Altamirano and Barr, 2019). Nevertheless, the following drawbacks should be considered when dealing with phages (López, García and García, 2004; Hermoso, García and García, 2007):

- Phages are high specific for their host, and so the causative bacterium must be identified beforehand, and have a narrow spectrum of action.
- Phages may not always remain lytic under the physiological conditions and bacteria can become resistant to phages after infection.
- Phage preparations should be sterile to meet strict clinical safety requirements, although sterilization treatments may inactivate the phages.
- Phages can be neutralized by antibodies, which can prompt allergic reactions.
- The self-replicating nature of phages complicate the pharmacokinetics of phage treatments.
- Phages might endow bacteria with toxic or antibiotic resistance genes when inserted into the bacterial genome (i.e. lysogenic phages).

In order to avoid some of these disadvantages, different types of purified phage-derived proteins have been evaluated as anti-infective agents.

1.4.2.1. Lysins and the Lytic Cycle of Phages

Lysins (or endolysins) are enzymes encoded by phage genomes, which are translated during the late phase in the lytic cycle to degrade the bacterial cell wall, enabling the release of the progeny (virions) once they have been assembled inside the bacterial host (Loessner, 2005). A schematic representation of the life cycle of phages is shown in figure 1.23.

On the contrary to the late phase, during the initial stages of adsoption and genome injection into the host cell, phages require two groups of proteins: virion-associated peptidoglycan hydrolases and polysaccharide depolymerases (Baca, Coen and Golan, 2012). These are also lytic enzymes, but should be distinguished from endolysins as these are an integral component of the virion base plate tube structure and locally digest the cell wall to facilitate the injection of the phage genome into the bacterial cell. Phage-encoded depolymerases target the polysaccharide components of the bacterial envelope, such as the capsule, lipopolysaccharide of Gram-negative bacteria and extracellular matrix of biofilms, to facilitate access to secondary host receptors located at the cell wall by degrading the structural polysaccharide components (Maciejewska, Olszak and Drulis-Kawa, 2018). A classic example of a well-studied process is the *E. coli* phage T4 where the tube's baseplate protein comprises a lysozyme (Kanamaru, *et al.*, 2005). Phage–

depolymerases may also be used strategically along with antibiotics to facilitate the accessibility of antibiotics through biofilm structures and the action of phagocytes by capsule degradation (Bedi, Verma and Chhibber, 2009; Born, *et al*., 2014).

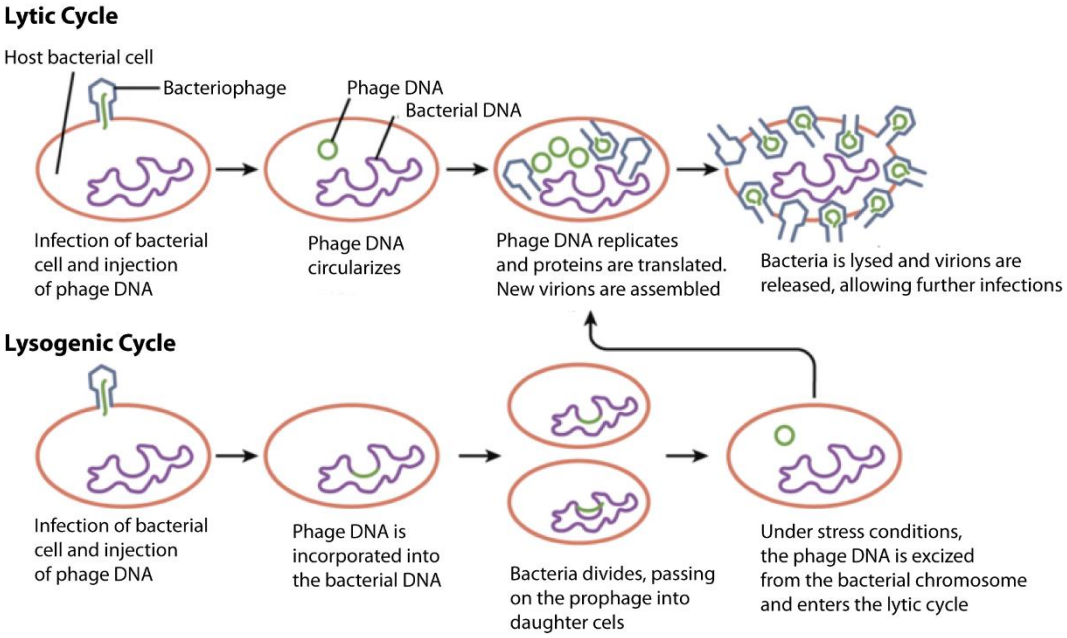


Figure 1.24.: Lytic and lysogenic cycle of bacteriophages. A temperate bacteriophage has both lytic and lysogenic cycles. In the lytic cycle, the phage replicates and lyses the host cell. In the lysogenic cycle, phage DNA is incorporated into the host genome, where it is passed on to subsequent generations. Environmental stressors such as starvation or exposure to toxic chemicals may cause the prophage to excise and enter the lytic cycle**.** (Adapted from https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology).

Although single-strand phage genomes may rely on single-gene lysis systems (Bernhardt, *et al.*, 2002), tailed phages generally encode for a dual protein system of holins and endolysins for bacterial lysis and progeny release during the later phase of their lytic cycle. Holins are small hydrophobic proteins that are expressed in a genetically programmed, allele-specific manner when virions are assembled (Gründling, Bläsi and Young, 2000), fine-tuning the bacterial lysis time. These are passively accumulated in the inner bacterial cell membrane until reaching a specific concentration that triggers their arrangement into holes, allowing the permeabilization of the inner membrane to endolysins (Wang *et al.*, 2000; Young, 1992). However, holins are not capable of independently lysing the cell wall (Roach, *et al.*, 2017).

Once holins rearrange into pores, endolysins translocate from the cytoplasm to the periplasmic space, gaining access to their polymeric substrate outside the cytoplasmic membrane: the peptidoglycan, responsible for the maintenance of cell shape and osmotic stability. Nevertheless, some lysins may contain a signal peptide and use the classic *sec* system to reach the peptidoglycan or rely on a SAR domain that functions as a type II signal anchor. This leaves the secreted protein in the periplasm but is tethered to the membrane in an inactive form. When the membrane is depolarized, the SAR domain exits the bilayer, resulting in activation of the

endolysin and cell lysis. (Hermoso, García and García, 2007). In the case of Gram-negative bacteria, the outer membrane should be disrupted for efficient cell lysis. Spanins achieve this action by catalyzing the fusion of the inner and outer membrane, leading to cell lysis (Young, 2013).

1.4.2.2. Endolysin Classification

Many lysins are produced as a single polypeptide but contain two functional domains: a catalytic domain (EAD) which undergoes the enzymatic activity to degrade the bacterial peptidoglycan, and a cell wall binding domain (CBD) which contributes to the species/strain/genus specificity of the lysin by binding a specific carbohydrate epitope in the cell wall (Yoong, *et al*., 2004; Eugster and Loessner, 2012; Nelson, *et al*., 2006).

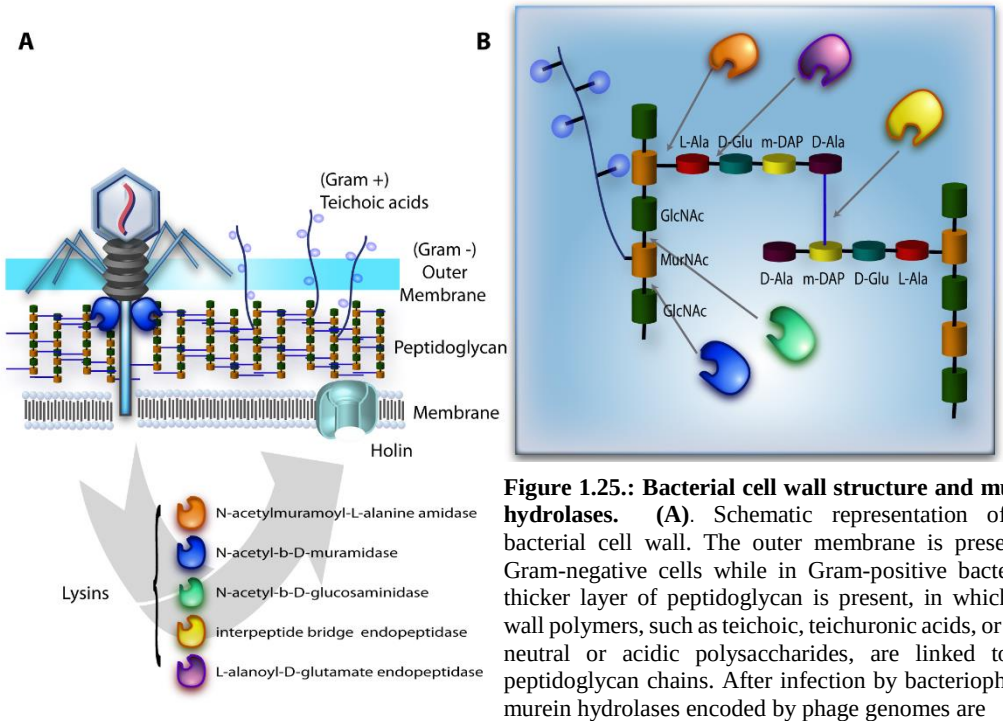


Figure 1.25.: Bacterial cell wall structure and murein hydrolases. (A). Schematic representation of the bacterial cell wall. The outer membrane is present in Gram-negative cells while in Gram-positive bacteria a thicker layer of peptidoglycan is present, in which cell wall polymers, such as teichoic, teichuronic acids, or other neutral or acidic polysaccharides, are linked to the peptidoglycan chains. After infection by bacteriophages, murein hydrolases encoded by phage genomes are produced during the late phase of the lytic cycle. Endolysins gain access to their substrate, through phage-encoded holins. **(B)**. Fine structure of peptidoglycan in *L. monocytogenes* and endolysins targets. Murein glycan strands consist of alternating GlcNAc (N-acetyl-D-glucosamine) and MurNAc (N-acetyl muramic acid) residues. Glycan strands are cross-linked by short peptides. Teichoic acid (TA) units are linked to MurNAc. Bonds cleaved by the different murein hydrolases are indicated by arrows. (Adapted from Hermoso, García and García, 2007).

According to their catalytic activity, endolysins are classified as N-acetylmuraminidases (lysozymes or muramidases), endo-β-N-acetulglucosaminidases (glucosaminidases), N-acetylmuramoyl-L-alanine amidases (NAM-amidases), N-endopeptidases and lytic transglycosylases (Hermoso, García and García, 2007). In general terms, these are glycosidases (glucosaminidases, lysozymes and lytic transglycosilases, which act on the sugar moiety of the peptidoglycan), amidases (which cleave the bond connecting the sugar and peptide constituent) and endopeptidases (in charge of cleaving the peptide cross-bridge) (Roach, *et al*., 2017) (Figure

1.24.). Usually, lysins possess only one type of hydrolytic activity, but enzymes harboring two independent lytic activities have been identified in phages (Hermoso, García and García, 2007).

1.4.2.3. Structural Basis for Cell Wall Recognition

The molecular and structural basis of the catalysis of peptidoglycan hydrolases is generally well known, with hundreds of structures characterized and available in the PDB. Nonetheless, little is known about the molecular and structural basis for cell wall recognition by endolysins.

The first insight on this structural knowledge was provided by the X-ray structure of Cpl-1 lysozyme, a member of the Choline Binding Proteins (CBP) encoded by *Streptococcus* phage Cp-1, in complex with three bacterial cell wall analogues (Pérez-Dorado, *et al.*, 2007). All CBPs have a choline binding module, which is formed by a 20 amino acid repeat, found in tandem copies. The largest peptidoglycan analog bound to Cpl-1 consists of a tetrasaccharide-pentapeptide (PDB entry: 2J8G) which locates along a groove leading to the active site of the glycosidase, thus allowing to hypothesize a processive mechanism for Cpl-1. The cell wall anchoring module is formed by six similar choline-binding repeats, arranged into two different structural regions: a left-handed domain configuring two canonical choline-binding sites and a β-sheet domain, which bring together the whole structure. These allow the specific recognition of the choline-containing teichoic acid of the pneumococcal cell wall.

Other recently reported CBD atomic structures comprise the full length phage-associated cell wall hydrolase PlyPy from *Streptococcus pyogenes* (PDB entry: 5UDN), *Streptococcus* phage phi7917 hydrolase Ly7917 (PDB entry: 5D74), and an endolysin from *Clostridium perfringens* phage phiSM101 (PDB entry: 4KRT), for which no structural analysis has been yet provided. Structure in the PDB of only the CBD of endolysins are the *Bacillus* phage γ endolysin PlyG CBD (PDB entry: 2L48) and the *Streptococcus* phage Cp-7 hydrolysin Cpl-7 CBD (PDB entry: 5I8L). There is only structural analysis for the latter, in which three almost identical CW_7 repeats have been identified. CW_7 repeats are present in many putative cell wall hydrolases encoded by Gram-positive and Gram-negative bacteria as well as by phages of Gram-positive bacteria, allowing a broader spectrum of recognition. These consist of 37–43 amino acids which adopt a three-helix bundle fold with a hydrophobic core (Bustamante, *et al*, 2017). Moreover, there is only a single full-length atomic structure reported and analyzed from the *Listeria* phage PSA endolysin (PlyPSA, PDB entry: 1XOV). The two functional modules, providing enzymatic and cell wall binding activities, are connected via a linker segment of six amino acid residues. The N-terminal catalytic module displays a twisted, six-stranded β-sheet flanked by six helices and shows to be a Zn-dependent peptidase. The CBD of PlyPSA features a novel fold, comprising two copies of a β-barrel-like motif, which are held together by means of swapped β-strand.

Furthermore, PlyPSA requires the CBD to undergo its catalytic activity, indicating that this region is responsible for the efficient binding of the enzyme to its peptidoglycan substrate (Korndörfer, *et al.*, 2006).

The molecular and structural knowledge of the CBD interaction with the bacterial cell wall is the key to understand how the bacterial envelope is degraded by endolysins and their specificity. Moreover, it will also shed light on the function and improvement of enzybiotics and further biotechnological applications.

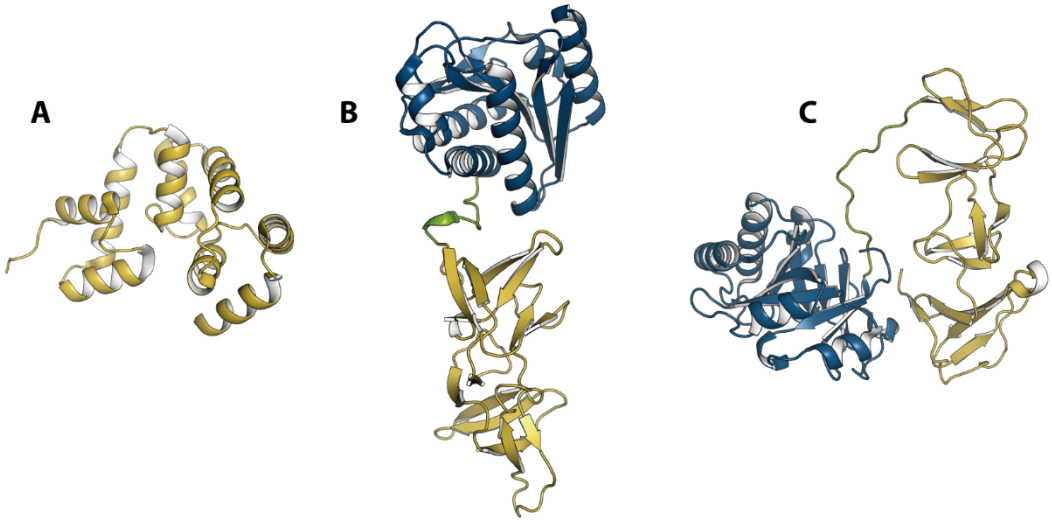


Figure 1.26.: Three-dimensional structures of endolysins. Cartoon representation of the crystal structures of the cell wall-binding domain of Cpl-7 **(A)** and full length PlyPSA **(B)** and Cpl-1 **(C).** Catalytic modules are colored in blue, CBDs in gold and linkers in green.

1.4.2.4. Medical and Biotechnological Applications of Endolysins: Bioengineering *à la Carte*

The natural hydrolytic properties endolysins have given them great potential as antimicrobial agents. Endolysins from Gram-positive bacteria can also carry out their activity from the outside (acting as exolysins) of the bacterial cell, as the peptidoglycan is accessible. Nevertheless, Gram-negative bacteria's outer membrane acts as a barrier to these enzymes, and so the spectrum of endolysins to use as a therapeutic agent is very much reduced (Yang, H. *et al.*, 2015). Different approaches have been devised in order to overcome this difficulty, such as the co-administration of chemicals that permeabilize the outer membrane (chelating agents, EDTA and aminoglycosides) (Briers and Lavigne, 2015) or the administration of artilysins (bioengineered endolysins). These are created by combining the active site of the lysin enzyme with lipopolysaccharide destabilizing peptide which allows for the penetration of the outer membrane. The efficacy of this approach has been proven for *P. aeruginosa* and *A. baumannii* models (Briers, *et al.*, 2014).

Although the capacity of a phage endolysin to kill bacteria was first reported in 1957 (Krause, 1957), it was not until 2001 that purified recombinant endolysins were used as

therapeutic agents to reduce or prevent the colonization of streptococci in mice mucosa (Nelson, Lumis and Fischetti, 2001). Phage lysins have been given plenty of attention as potential antimicrobial agents and have been developed since. These proteins are fast-acting, potent and inactive against eukaryotic cells. A summary of recently published findings on phage lytic enzymes is given in Table 1.3.

Table 1.3.: Recently published findings on phage lytic enzymes. (Adapted from Lin, Koskella and Lin 2017).

	Lysine	Model	Target	Result	Reference
Phage-derived lysins	**ABgp46**	*In vitro*	MDR *A.baumannii*, *P.aeruginosa*, *S.typhimurium*	Cross-inoculation significantly reduced bacterial density	Oliveira, *et al.*, 2016
	PlyF307	Murine	MDR *A. baumannii*	i.p. treatment resulted in mice survival from lethal bacteremia	Lood, *et al.*, 2015
	Cpl-1	Murine	*S. pneumonia*	i.p. treatment resulted in the survival of mice from lethal pneumonia	Witzenrath, *et al.*, 2009
	X6 cocktail	*In vitro* and murine	MRSA	Effective against biofilms *in vitro* and protected mice from lethal sepsis	Schmelcher, *et al.*, 2015
	PlyCD	*In vitro* and *ex vivo*	*C. difficile*	Reduced colonization	Wang, *et al.*, 2015
	PlySs2	Murine	S. pyogenes and MRSA	i.p. treatment resulted in reduced mortality from lethal bacteremia	Gilmer, *et al.*, 2013
	PlyG	*In vitro*	*B. anthracis*	Eliminated spores and vegetative cells	Yang, *et al.*, 2012
	Trx-SA1	*In vivo* (cow)	MRSA	Significant reduction in pathogen levels in milk	Fan, *et al.*, 2016
Bioengineered chimeric lysins	**CHAPK**	*In vitro*	MRSA	Eliminated MRSA and dispersed biofilms	Keary, *et al.*, 2016
	ClyH	Murine	MRSA	Mice survival from bacteremia	Yang, *et al.*, 2014
	Cpl-711	Murine	*S. penumoniae*	Mice survival from bacteremia	Díez-Martínez, *et al.*, 2015
	Ply187	Murine	*S. aureus*	Prevented bacterial endophthalmitis	Singh, *et al.*, 2014
	Artily-sins	Nematode gut, human keratinocyte	*P. aeruginosa*, *A. baumannii*	Decolonized *P. aeruginosa* from gut	Briers, *et al.*, 2014
	HY-133	*In vitro*	MRSA	Significant activity against a collection of strains	Idelevich, *et al.*, 2016
Lysins + antibiotics	**CF-301**	Murine	MRSA	Improved efficiency in combination with vancomycin or daptomycin	Schuch, *et al.*, 2014
	MR-10	Murine	Wound infection	Improved efficiency when in combination with minocycline	Chopra, *et al.*, 2016
	CHAPK	*In vitro*	MRSA	Thermally triggered release of a staphylolytic cocktail from nanoparticles	Hathaway, *et al.*, 2017

MDR: multidrug resistant; i. p.: intraperitoneal injection; MRSA: Mehtillin-resistant *S. aureus.*

Concerns about the therapeutic use of purified endolysins to kill pathogenic bacteria are related to resistance and immune response. Nevertheless, the unlikeliness of bacteria to evolve resistance to lysins it is widely considered: since the phage and the host bacteria are coevolving, endolysins have developed in a way that they bind and cleave highly conserved structures in the cell wall, the modification of which is believed to be detrimental to the bacteria (Borysowski, Weber-Dabrowska and Górski, 2006; Roach and Donovan, 2015). Furthermore, it has been shown

that no resistance was developed from streptococcal bacteria after a repeated exposure to low concentrations of lysins (Fischetti, 2010). Regarding immune responses, although proteins can trigger one when delivered, no anaphylactic response or side effect was observed in the animals tested (Schmelcher, Donovan and Loessner, 2012). Thus, phage-derived lysins are being developed as a therapy to overcome the urge of novel classes of antibacterial agents due to bacterial resistance. Herein, SAL200, an anti-staphylococcal intravenous endolysin therapy is currently in phase I clinical trials to assess its pharmacokinetic profile, establish its ideal dosing schedule, and evaluate its safety (Jun, *et al.,* 2016) and Staphefekt, which is in phase II clinical trials, is used to topically treat skin flora dysbiosis caused by overgrowth of *S. aureus.* (Totté, *et al.,* 2017).

As mentioned, the catalytic activities of lysins are well known and classified according to their limited hydrolytic activities. The modular structure of endolysins provides a unique opportunity for protein engineering in order to modify bacteriolytic activity, specificity, solubility, and other physicochemical properties of these enzymes. Further insight on their specificity given by the CBD may allow to direct their catalytic activity to specific pathogenic bacteria and prevent collateral damage and lysis on yet uninfected host cells (Loessner, *et al.,* 1995). One example of an engineered enzyme with increased lytic activity is Ply187AN-KSH3b, a fusion of the endopeptidase domain of the staphylococcal Ply187 lysin (Ply187AN) with the SH3b CBD of another staphylococcal phage endolysin, LysK. The fusion construct displayed above 10-fold-higher staphylolytic activity than that of Ply187AN and was also more active than LysK in multiple activity assays (Mao, *et al.,* 2013). Nevertheless, application of endolysins as an antimicrobial agent is hampered by the specific requirements of these enzymes, in terms if pH and salt concentration (Hagens and Loessner, 2014). Chimeric recombinant phage lytic proteins are still in the early stages of research, but some of these modified lysins have been proved efficient (Table 1.3.). Further potential of lysin application includes the combination therapy of lysins and antibiotics, which has been shown to be more effective than antibiotics alone (Table 1.3.). A review on biomedical application of endolysins as antimicrobial agents can be found in (Haddad-Kashani, *et al.*, 2018).

Phage-derived endolysins are also used in diagnostics for detection purposes. Sensitive and specific diagnostic protocols to reliably and rapidly identify microbial threats are essential to ensure food safety. Furthermore, endolysins have been applied in animals and crops of agricultural importance as well as for hygiene measures in food production facilities and hospitals. Similar to biomedical applications, CBD-based engineering can be applied for the detection of different pathogenic bacteria, such as *C. perfringens* and *B. cereus*, by fusion with magnetic particles (Kretzer, *et al.,* 2007). Moreover, phage endolysins λSA2 and B30 significantly decrease intramammary concentrations of streptococcal species causing mastitis (Schmelcher, *et al.,* 2015).

Further applications consider the direct addition of endolysins to food products, although only a few studies actually research their efficacy only focusing on cow milk, possibly due to the hampered activity of endolysins in complex food matrices and their high costs when applied beyond the product's surface (Schmelcher and Loessner, 2016; for further information on applications on food safety).

1.4.3. *Listeria monocytogenes* and its Cell Wall

Listeria spp. are small Gram-positive, flagellated, facultative anaerobic rods with a ubiquitous distribution in the environment and a generally saprophytic lifestyle. Currently, the genus *Listeria* is classified into 17 species, based on their genomic and phenotypic characteristics. These species harbor group-specific epitopes, defined by the somatic (O) and flagellar (H) antigens that form the key determinants of serological typing (Orsi, and Wiedmann, 2016). Two of these species, namely *L. monocytogenes* and *L. ivanovii*, are potentially pathogenic. In particular, *L. monocytogenes* is an important foodborne human pathogen and the causative agent of listeriosis, an infectious disease that primarily affects individuals belonging to certain risk groups such as infants, elderly, immunocompromised patients, and pregnant women, in which the bacteria may penetrate the intestinal tract and cause systemic infections including the central nervous system, gastroenteritis, hepatitis and stillbirth or abortion in pregnant woman (Vazquez-Boland *et al.,* 2001). Although rare, listeriosis is associated with serious conditions such as meningitis and septicemia and mortality rates of often more than 30%, which makes it one of the most dangerous foodborne pathogens (Vazquez-Boland *et al.,* 2001) remaining of great public health concern. The disease is caused by the consumption of food contaminated with *L. monocytogenes*, which is a psychrotroph which can grow at refrigeration temperatures and tolerate a wide range of pH and salt concentrations (Gandhi and Chikindas, 2007). To date, at least 12 distinct serovars (i.e. 1/2a, 1/2b, 1/2c, 3a, 3b, 3c, 4a, 4b, 4c, 4d, 4e, and 7) have been identified in *L. monocytogenes*, although at least 95% of the strains isolated from foods and patients are of serovars 1/2a, 1/2b, 1/2c, and 4b (Seeliger and Höhne, 1979; Vazquez-Boland *et al.,* 2001).

L. monocytogenes is a highly successful pathogen that invades eukaryotic host cells, crosses several natural barriers of the host and survives to extreme environments. Hence, its cell wall must necessarily contain molecules making the colonization of these niches possible. While the biochemistry of teichoic acids (TA) and lipoteichoic acids (LTA) of *L. monocytogenes* was inferred three decades ago (Fiedler, 1988), it has only been recently structurally determined in detail by electrospray ionization coupled to mass-spectrometry (Eugster and Loessner, 2011; Shen, *et al.,* 2017). TA polymers are now receiving further attention, as they are known to mediate multiple other biological roles, such as physiology, immunological recognition, antibiotic

resistance, virulence, and interaction with host cells, bacteriophages and endolysins (Pucciarelli, Bierne and Portillo, 2007).

L. monocytogenes has a peptidoglycan formed by glycan chains containing alternating units of the disaccharide N-acetylmuramic acid (MurNAc)-(β-1,4)-N-acetyl-d-glucosamine (GlcNAc). Bound to the MurNAc residue is a stem peptide that in *L. monocytogenes* contains L-alanine-γ-D-glutamic acid-meso-diaminopimelic acid -D-Ala-D-Ala [L-Ala-γ-dD-Glu-m-Dap-D-Ala-D-Ala] (Fiedler, 1988). The glycan chains are crosslinked by 4→3 linkages between the d-Ala residue of one lateral peptide to the m-Dap residue of the other stem peptide (Figure 1.24.B.). This peptidoglycan structure resembles the reported for many gram-negative bacteria as *E. coli* (Schleifer and Kandler, 1972). The TAs are covalently bound to the peptidoglycan by phosphodiester bonds between MurNAc and a special conserved linkage unit (Figure 1.27.), whereas LTAs are amphipathic molecules tethered to the cytoplasmic membrane via a glycolipid moiety (Fischer, 1988).

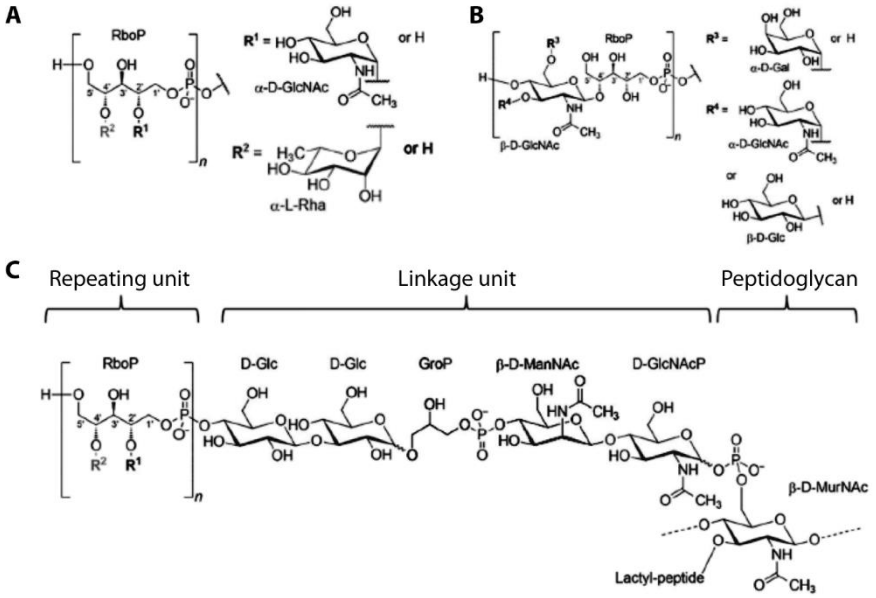


Figure 1.27.: A chemical view of two types of *L. monocytogenes*. (**A)** Structure of the type I repeating unit. (**B)** Structure of the type II repeating unit. **C**. Structure of the repeating unit, linkage unit, and peptidoglycan in the cell wall of *L. monocytogenes* serovar 1/2a. R1: GlcNAc, R2: rhamnose. (Modified from Shen, *et al.*, 2017).

Peptidoglycan-associated TAs are highly variable in structure and often feature species- or even strain-specific variations (Figure 1.27. and 1.28.). *Listeria* TAs are comprised of a polyribitol-phosphate (RboP) chain that consists of 20 to 30 repeating units which can be decorated with a variety of different sugars and (Eugster and Loessner, 2011). Two main structural types exist: type I TAs (serovar groups 1/2, 3 and 7) adopt [5)-Rbo-1-P-(O→]$_n$ chains where the Rbo units are directly interconnected by phosphodiester bonds between C1 and C5, and can be decorated with terminal rhamnose (Rha) or α-D-GlcNAc on the C2 and C4 position of the ribitol molecule. Type II TAs (serovar 4 of *L. monocytogenes*) bear [4)-GlcNAc-(β1→2/4)-Rbo-1-P-(O→]$_n$ as the repeating unit, which incorporate the β-D-GlcNAc moiety within the polymer

chain. The integrated GlcNAc itself may be further decorated with glucose (Glc), galactose (Gal), or with an α-D-GlcNAc residue which may also be O-acetylated (Fujii, *et al.*, 1985; Uchikawa, Sekikawa and Azuma, 1986; Shen, *et al.*, 2017). Herein, variation among the different serovars of *L. monocytogenes* is mostly determined by differences in carbohydrate substitution and components of the RboP subunits of its TAs.

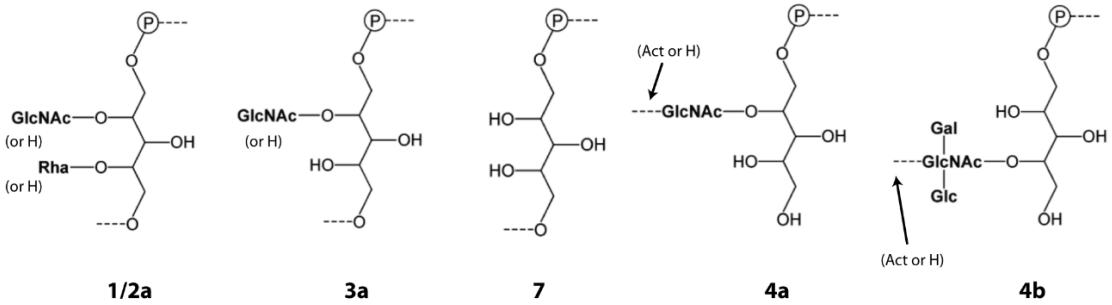


Figure 1.28.: Basic structures of some cell wall TA of *L. monocytogenes* serovars. The serovars are indicated below. The structures represent the linear monomeric ribitol-phosphate repeat units, which are connected by phosphodiester bonds in the TA polymers (dashed lines). (Modified from Eugster, *et al.*, 2011).

Further insight on the specificity and binding mechanism of CBDs from *Listeria* bacteriophage endolysins binding to TAs is crucial, as these molecules can recognize the *Listeria* cell wall in a serovar-dependent manner (Schmelcher, *et al.*, 2010; Eugster, *et al.*, 2011) and, as mentioned in previous sections, have been proposed as novel tools for diagnostics and biocontrol (Schmelcher, Donovan, and Loessner, 2012).

1.4.3.1. *Listeria* Phages: Endolysins and Applications

To date, more than 500 *Listeria* phages have been isolated and characterized. All *Listeria*-specific bacteriophages found are members of the *Caudovirales*, most belonging to the *Siphoviridae* family. Bacteriophages have been found for all major *Listeria* species and serovars, except for *L. grayii*, *L. rocourtii* and *L. marthii*. In general, serovar 3 strains are highly refractory to phage infection, whereas serovar groups 1/2 and 4 strains are particularly susceptible (Klumpp and Loessner, 2013). Herein, absence of phage identification seems to be attributable to differences in teichoic acid composition (Kuenne, *et al.*, 2013).

Over the past decade, *Listeria* phage endolysins have been characterized and analyzed for their binding spectra and catalytic mechanism (Loessner *et al.*, 1995; 1996; 2002; Zimmer *et al.*, 2003; Korndörfer *et al*., 2006; Schmelcher, *et al*., 2010) and several applications designed based on the high binding affinity and specificity of endolysin's CBD. *Listeria* endolysins have been used for removing *Listeria* occurring on food-contact surfaces and biofilms: studies have shown that CBDs immobilized on paramagnetic beads can effectively separate target cells from dilute suspensions, to be then combined with different secondary diagnostic steps (Kretzer, *et al.*, 2007; Walcher, *et al*, 2010). Most recently being, magnetically immobilized CBDs have been coupled with a luciferase-modified A511 bacteriophage for detection purposes (Kretzer,

Schmelcher and Loessner, 2018). Furthermore, CBDs can be used along with electrochemical impedance spectroscopy to measure bacteria captured by the CBD molecules immobilized on a gold screen printed electrode (Tolba, *et al.*, 2012), and differently colored fluorescent proteins fused to CBDs with different recognition and binding spectra allow for a rapid and multiplexed detection and differentiation of *Listeria* strains by fluorescence microscopy (Schmelcher, *et al.*, 2010).

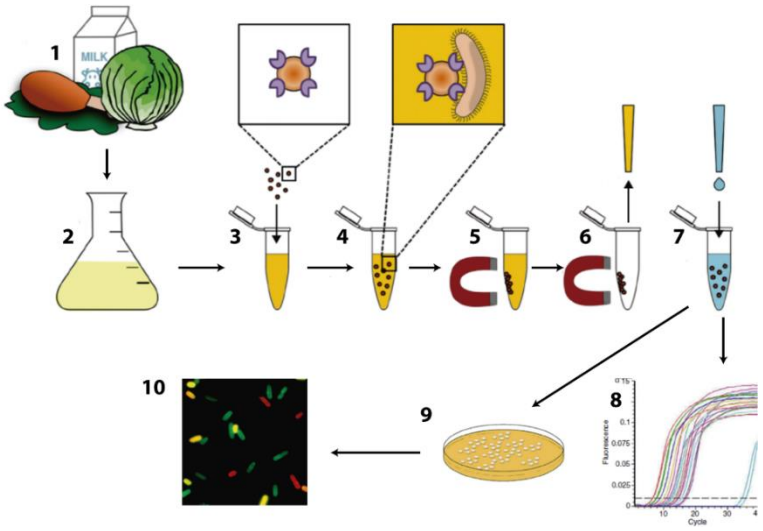


Figure 1.29.: Detection and differentiation of pathogens in food by CBD-based magnetic separation coupled with multiplex cell wall labelling. 1: Take food sample; 2: amplify target bacteria in sample by short selective enrichment; 3: add magnetic beads coated with CBDs specific for target bacteria to aliquot of enrichment culture; 4: allow beads to capture target bacteria; 5: remove bead-bacteria complexes from suspension by applying magnet; 6: remove medium; 7: wash and resuspend in buffer; detect bacteria bound to beads by (8) molecular methods such as quantitative PCR or by (9) direct plating on selective agar; 10: use colony material for multiplex binding assay employing CBDs fused to fluorescent reporters. (Adapted from Schmelcher and Loessner, 2016).

1.4.3.1.1. PlyP35 and *L. monocytogenes* Teichoic Acids

The endolysin of *Listeria* phage P35 (Dorscht *et al.*, 2009), named PlyP35, has been shown to effectively kill viable *Listeria* cells, and it revealed a remarkably high thermostability, which makes it particularly interesting for potential application as an antimicrobial in food products that undergo heat treatment (Schmelcher *et al.*, 2012a). The EAD of this enzyme has been identified by bioinformatics as a putative endopeptidase, which is believed to target the bond between L-alanine and D-glutamate in the stem peptide of the directly cross-linked peptidoglycan of *Listeria* (Schmelcher, *et al.*, 2010, Schleifer and Kandler, 1972). The CBD of PlyP35 (CBDP35) specifically binds to TA-associated GlcNAc moieties in *Listeria* cell wall (Eugster, *et al.*, 2011), resulting in a binding spectrum that covers most strains of serovars 1/2 and 3 and many strains of serovars 4, 5, and 6 (Schmelcher *et al.*, 2010). The interaction between the CBD and the surface of bacterial target cells has been demonstrated to be exceptionally strong, with equilibrium dissociation constants in the nano- to picomolar range (Schmelcher *et al.*, 2010). However, the structural basis and the atomic interactions for this strong and specific binding has not been investigated.

2. OBJECTIVES

Bearing in mind the importance of macromolecular structural insight in biotechnological applications such as rational drug design, biomedicine and the knowledge of the molecular machinery involved in disease, comprehension of biological processes regulation and the development of tools for pathogen detection, three relevant projects have been considered in this dissertation: (i) TryR, an essential and unique kinetoplastid protozoa enzyme key in their redox metabolism, as a drug target for anti-trypanosomal drug design; (ii) eEF1A2, a moonlighting protein subject of many PTMs and with oncogenic properties, as plitidepsin's drug target (PharmaMar S. A.) whose binding is yet uncharacterized, and the structural implications and regulation resulting from PTMs; and (iii) CBDP35, the cell wall binding domain of *Listeria* bacteriophage P35 endolysin PlyP35 which binds to *L. monocytogenes* serovar 1/2 teichoic acids, as a tool in phage therapy and pathogen detection in the food industry.

The specific goals concerning each of above-mentioned proteins are listed below:

· *L. infantum* TryR:

- Crystallization and structural determination of apo-TryR.
- Crystallization and structural determination of TryR in complex with NADPH and trypanothione.
- Exploration of the crystallization and diffraction parameters of TryR dimerization disruptors.
- Crystallization and structural determination of TryR in complex with oxidoreductase activity inhibitors which do not disrupt the dimer formation.
- Structural analysis and characterization of the binding of these inhibitors in order to comprehend the molecular basis of inhibition.

· eEF1A2 from *Oryctolagus cuniculus:*

- Crystallization and structural determination of eEF1A2 in its GDP conformation.
- Discovery and localization of novel PTMs.
- Structural analysis and characterization of PTMs and their possible implications in the regulation of the many functions of eEF1A2.
- Obtainment of a homogenous sample of eEF1A2 in its GTP conformation bound to a non-hydrolyzable GTP analogue, GppNHp.
- Crystallization and structural determination of eEF1A2 in its GTP conformation.
- Crystallization and structural determination of eEF1A2·GppNHp in complex with plitidepsin.

- Structural analysis of eEF1A2·GppNHp·plitidepsin complex.

· PlyP35 CBD:

- Crystallization and structural determination of CBDP35 in complex with *L. monocytogenes* teichoic acids.
- Structural comparison with homologue CBD of PlyPSA.
- Structural analysis and characterization of the binding of GlcNAc to CBDP35.
- Structural analysis and characterization of the binding of teichoic acids with CBDP35 and determination of the molecular basis of α-D-GlcNAc recognition by CBDP35.

3. EXPERIMENTAL PROCEDURES

X-ray Crystallography is the technique used throughout this Thesis for the structural determination of proteins. This technique allows the three-dimensional structure determination of ordered matter, both organic and inorganic, at atomic level. Upon other applications, structural elucidation of proteins and macromolecular complexes by X-ray crystallography provides information on folding and physical-chemical properties of the proteins studied, such as bond distance and angles. This data is crucial for the understanding of the biological function and characterization of the processes in which the protein is involved.

Structural determination by X-ray crystallography consists of several stages (Figure 3.1). Essentially, a protein crystal is needed in order to generate an X-ray diffraction pattern when exposed to this electromagnetic radiation. The diffraction pattern will then be used to determine the final structure of the crystal-forming protein (Als-Nielsen and McMorrow, 2011). The stages composing this technique are described below as well as the different concepts which will allow the understanding and interpretation of the Results and Discussion section of this dissertation.

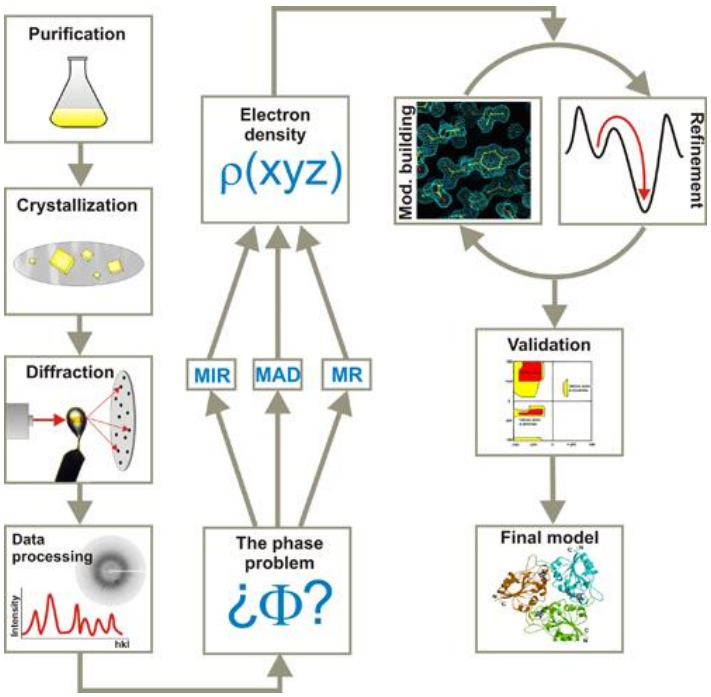


Figure 3.1.: Main steps in the structural resolution of proteins by X-ray crystallography. (Figure: http://www.xtal.iqfr.csic.es/Cristalografia/index-en.html).

3.1. OBTAINING A PURE PROTEIN SAMPLE

The first step towards the structural resolution using X-ray crystallography is obtaining a sufficiently pure protein sample. It is crucial that our sample is of high purity (90-95%), soluble, stable and homogenous, and that sufficient amounts of protein are produced for both preliminary crystallization assays and reproducing final crystals of considerable quality and size (Begfors, 2009).

Therefore, cloning, expression and purification are common steps to be carried out prior to crystallization. Each protein will follow a unique protocol, based on its physical, chemical and biological characteristics, entailing a bottleneck for pure protein acquisition. Moreover, protein behaviour during crystallization trials may be affected by the different decisions and procedures performed throughout these phases, and so going back to initial stages of protein production is typical in X-ray crystallography projects.

3.1.1. Cloning and Protein Expression

DNA cloning is a molecular biology technique which involves the insertion of a protein-of-interest encoding gene into a vector, usually a circular piece of DNA (plasmid). The resultant recombinant plasmid is introduced into the microorganism of choice, which will be suitable for the high-level expression of the encoded protein in order to obtain large yields. The expression system generally involves the use of specific strains of a microorganism (typically *E. coli* bacteria), and therefore cloning and expression pilot trials in different strains and conditions are performed. Variables include cloning of a truncated or mutated protein, restriction sites within the gene construct (i.e. restriction of a N-terminal or C-terminal tag), antibiotic resistance, bacterial strain and microorganism used. Examples of rational directed mutation of specific amino acids of the protein are: enzyme inactivation for co-crystallization with the substrate, modification of the solubility of the protein, hampering the flexibility of mobile regions or domains and formation of new bonds or contacts within protein-protein complexes. According to the stability of the protein and selected strain, temperature, cell culture media and induction of protein expression (such as inducer concentrations, optical density for induction and induction period) are to be assayed.

It is worth mentioning that after many unsuccessful crystallization trials, TryR from *L. infantum* MCAN/ES/89/IPZ229/1/89 strain was cloned at the System Biology Department at Alcalá de Henares University to aid crystallization. Regarding CBDP35, only the cell wall binding domain of endolysin PlyP35 of *L. monocytogenes's* phage P35 was cloned (gene fragment coding for residues 157-293 in PlyP35). This was carried out by our collaborators at the Department of Health Sciences and Technology at the ETH in Zürich.

3.1.1.1. Bacterial Strains

The selection of the bacterial strain is a key component to obtain an enhanced expression and stable protein. Each strain has specific characteristics and its choice for protein expression is closely tied to the properties of the target protein to be expressed (i.e. unstable, insoluble, misfolding, toxic…) and the choice of expression vector. For our work, *E. coli* DH5α has been

used as a plasmid reservoir and for cloning purposes and *E. coli* BL21Gold (DE3) as the expression strain.

DH5α is the most frequently used *E. coli* strain for routine cloning and plasmid storage applications. In addition to supporting blue/white screening by deletion of residues 11-41 of the β-galactosidase encoded by *lacZ* gene, *recA1* and *endA1* mutations in DH5α increase insert stability and improve the quality of plasmid DNA prepared from minipreps, hampering recombination and plasmid degradation by nucleases, respectively. Furthermore, the *hsdR17* mutation eliminates the restriction endonuclease of the EcoKI restriction-modification system, so DNA lacking the EcoKI methylation will not be degraded.

BL21-Gold (DE3) are improved versions of BL21 competent cells. These expression strains are ideal for performing protein expression studies that utilize the T7 RNA polymerase lacUV5 promoter to direct high-level expression via Lac operon. This implies the expression of the cloned gene via IPTG supplementation in the cell culture. Derived from *E. coli* B, these expression strains naturally lack the lon protease, which can degrade recombinant proteins. In addition, these strains are engineered to be deficient for a second protease, ompT, which could also be involved in recombinant protein degradation during purification. BL21-Gold presents the Hte phenotype, which increases the transformation efficiency. In addition, this strain is *endA*$^{-}$. These two improvements allow direct cloning of many protein expression constructs. This strain is used as the expression system for CBDP35.

BL21(DE3)-*Rosetta* strain was used for the expression of TryR. This BL21 derivative is designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli,* as they contain an added plasmid (pRARE) coding for several rare tRNAs. By supplying rare codons, the Rosetta strains provide for "universal" translation, where translation would otherwise be limited by the codon usage of *E. coli*.

3.1.1.2. Expression Vectors

The choice of vector is also key to obtaining a large amount of protein and depends on the length of the gene to insert, copy number, compatibility with other vectors, antibiotic marker, restriction sites, compatibility of restriction sites with the cloned gene's sequence, number of multiple cloning sites, tags to aid protein purification and expression system used.

The vector pET28a and pET28c (Novagen, Germany) have been used for TryR and CBDP35 cloning, respectively. Vectors within the pET family have a replication origin (pBR322) that leads to 20 – 60 copies of the plasmid within a single bacterial cell. These plasmids have a single multiple cloning site whose expression is controlled by means of the T7 promotor and lac operon. Furthermore, these plasmids allow the addition of an 6xhistidine tag in the N-terminal of

the recombinant protein. These pET plasmids provide the cells with kanamycin resistance, and so transformed bacteria can be selected by the presence of this antibiotic.

3.1.1.3. Culture Media

There is a wide variety of cell culture media and its selection highly depends on the organism cultured. Many bacteria are auxotroph for a certain amino acid or have specific nutritional requirements. Culture media may have a defined composition or may be enriched by the use of digested animal or plant extracts. Furthermore, media composition may be designed for bacterial self-induction by controlling diverse sugars concentrations. The culture medium chosen is highly important to maintain an efficient bacterial performance for protein expression. Herein, different additives can be added to the medium, such as sucrose, glucose or ethanol to aid protein solubility.

The liquid culture medium used in this work for TryR and CBDP35 expression was Luria-Bertani medium (LB). This medium contains 10 g/L of tryptone, 5 g/L of yeast extract and 10 g/L of NaCl. The solid culture medium used when transforming bacterial was LB-agar was obtained by supplementing liquid LB with 15 g/L of bacteriological agar. LB medium was supplemented with the required antibiotic for plasmid selection according to vector and strain used.

3.1.1.4. Protein Expression

Protein expression is achieved by growing a cell culture of the transformed bacteria with the gene-containing plasmid in a specific culture media supplemented with the selection antibiotic, during several hours and at an optimum temperature and agitation speed. When a precise optical density of the culture (O.D. at 600 nm, $O.D._{600nm)}$ is reached (usually during the bacterial exponential growth phase), the induction of protein expression is carried out, generally by supplementing the culture with IPTG (concentration of which depends on the vector used) at a specific final concentration, and continuing the culture at precise temperature, agitation and time conditions. Once induction is complete, the culture is centrifuged, and the cell pellet is stored at -80 ºC.

3.1.1.4.1. Expression Conditions for TryR

Although TryR was routinely produced by our collaborators at the System Biology Department at Alcalá de Henares University, expression and purification conditions for crystallization purposes were optimized at the Physical-Chemistry Institute “Rocasolano”.

Transformed BL21(DE3)-*Rosetta E. coli* cells containing pET28a-His-TryR plasmid were grown in LB medium at 37 ºC with a constant 250 rpm agitation until $O.D._{600nm}$ reached 0.5

A.U. IPTG was added at a final concentration of 1 mM and protein expression was induced for 4 hours at 37 ºC.

3.1.1.4.2. Expression Conditions for CBDP35

Transformed BL21-Gold (DE3) *E. coli* cells containing pET28c-His-CBDP35 plasmid were grown in LB medium at 37 ºC and 120 rpm until the $O.D._{600nm}$ reached 0.5 A.U. approximately. Protein expression was induced by adding 0.5 M IPTG and the culture was continued for 4 hours at 30 ºC with agitation.

3.1.2. Protein Purification from Natural Sources (eEF1A2)

In the case of eEF1A2, the protein was purified directly from a natural source (i.e. rabbit muscle). Though laborious, it was of our interest to study the natural behavior and occurrence of the protein and its interaction with plitidepsin in the most similar manner to how the binding should occur in humans. As previously mentioned, this protein is extensively post-translationally modified and is subject to a unique PTM only present in higher eukaryotes, which may regulate its various functions. Though expression systems such as bacteria and insect cells provide a cheaper approach of protein production in which higher yields are obtained, PTMs may be lost. However, eEF1A2 is highly abundant within muscle cells.

Unlike recombinant proteins, naturally isolated proteins do not involve tags, and their molecular behavior, enzymatic activity, and other properties occur under physiological conditions. Isolation and purification of a single protein from native source without containing mixtures of unrelated substances is quite challenging. Regarding the crucial homogeneity of the sample needed for crystallography purposes, naturally obtained proteins entail a huge defiance, as variability between samples may hamper reproducibility in crystallization essays.

eEF1A2 purification was carried out by our collaborators at PharmaMar S. A. Despite the training in the purification of this protein from its natural source, this step was optimized and routinely executed at PharmaMar's facilities. The protocol followed for extraction and purification of eEF1A2 is reported by Losada, A. *et al. (*2016).

3.1.3. Protein Purification

Once the cells have expressed our protein of interest, a purification process is performed in which our protein is extracted from the cells with the highest possible purity criteria. It is crucial that our sample is pure, stable, soluble and conformationally homogenous for crystallization purposes.

Different techniques are available for purification. On one hand, it should be taken into consideration that each step contributes to a higher degree of purity of our sample. On the other hand, large amounts of protein are lost. It is critical to balance purity of sample and amounts (milligrams) of protein purified. This should be monitored throughout every step of the extraction and purification stage. For this Thesis, this was carried out by polyacrylamide gel electrophoresis under denaturing conditions (SDS-PAGE) following the method reported by Laemmli (Laemmli, 1970), using 12 % (m/v) for separating gels and 4 % (m/v) for stacking gels of polyacrylamide.

3.1.3.1. Cell Lysis

Dealing with proteins held in the cytoplasm of the expression system (or natural tissue in the case of eEF1A2), the proteins expressed should be released from inside the cell. Lysis buffer used may affect the efficiency of soluble protein extraction from the cells. It is therefore necessary to consider a compatible pH with both our protein of interest and the first chromatography step. Ionic strength of the buffer by salt addition is also contemplated in order to increase protein solubility and stability. Glycerol is another important component for protein stabilization and to hamper unspecific interaction with other cellular components. According to the protein, chelating agents, protease inhibitors, reducing agents and/or detergents may be added to the lysis buffer to increase protein extraction and solubilization.

Typical lysis techniques include sonication, the use of high pressure by French press, the breakdown of bacterial cell wall by lysozyme and the freeze-thawing technique with liquid nitrogen. Sonication is commonly used, although great amounts of heat and vibrations are produced which could easily cause protein degradation. Heating of our cell extracts is avoided by the use of ice and short sonication cycles.

Following cell lysis, centrifugation of our lysate is necessary for the separation of the soluble fraction from other insoluble cell components.

3.1.3.1.1. Cell Lysis Conditions for TryR

The lysis buffer used for cell resuspension was composed by 50 mM Tris-HCl pH 7, 30 mM NaCl, 25 mM imidazole and was used in a 10 mL / gram of cell culture ratio. Lysozyme was added in a final concentration of 1 mg/mL as well as a protein inhibitor cocktail in a 1:100 ratio. The resuspended cell pellet was then incubated for 30 minutes in ice. Sonication was then used at maximum amplitude for 30 seconds followed by a 30 second incubation in ice for a total time of 10 minutes. Centrifugation of the lysate was performed for 1 hour, 14500 x g at 4 ºC. The supernatant was filtered through a 22 μm Millipore syringe filter and the resultant pellet was discarded.

3.1.3.1.2. Cell Lysis Conditions for CBDP35

The lysis buffer used for cell resuspension contained 50 mM Tris-HCl pH 8, 500 mM NaCl, 0.1 % Tween 20, 5 mM imidazole, 0.5 mM PMSF and was used in a 30 mL / L of cell culture ratio. Sonication was performed at maximum amplitude for 30 seconds followed by a 30 second incubation in ice for a total time of 25 minutes. Centrifugation of the lysate was performed for 45 minutes, 14500 x g at 4 ºC. The supernatant was filtered through a 22 µm Millipore syringe filter and the resultant pellet was discarded.

3.1.3.2. Chromatography Techniques

3.1.3.2.1. Affinity Chromatography

This principle of this purification technique is based on the affinity of a protein to a compound or chemical group bound to a resin, which can be pre-packed into a column format. In the case of TryR and CBDP35, a 6xHistidine tag is fused onto their N-terminal due to the cloning vector used. The interaction between the 6xHistidine tag and the resin is reversible and therefore allows the separation of the protein bound to the column from other proteins in the soluble fraction of the cellular lysate. Ni-NTA resins or columns are commonly used, as they are highly specific for the 6xHistidine tag of the recombinant proteins. This occurs due to the affinity of Ni^{2+} in the resin for the histidine's imidazole group. Elution is carried out by the addition of a chemical compound that competes with resin or protein. For this purpose, imidazole is added to the buffers in different concentrations in order to break the interaction between Ni^{2+} and the 6xHistidine tag. In this way, other contaminant proteins that may bind to the resin are withdrawn and subsequently the protein of interest is eluted.

3.1.3.2.2. Ion Exchange Chromatography

Ionic exchange chromatography separates proteins according to their differences in charge and is based on the interaction between charged polar groups on protein surfaces and the resin (generally prepacked into a column). In order to estimate a protein's charge, the pH of the buffer in which it is solubilized should be monitored. It is of common practice to, at least, alter the buffer's pH 2 units above or below a protein's isoelectric point (IP). If the pH of the buffer is higher than the IP of the protein, a negative surface charge is obtained, and vice versa. Therefore, only the proteins with opposite charge to the column's matrix are retained. There are mainly two types of ionic exchange columns: anionic exchange columns (Q) and cationic exchange columns (S). In order to elute the retained proteins, the charge-charge interactions between the protein and the matrix should be disrupted. This is achieved by a NaCl gradient in the chromatography buffers.

3.1.3.2.3. Size Exclusion Chromatography

This chromatography separates molecules according to their size. The principle of size exclusion chromatography or gel filtration resides in the ability of a sample to flow through the porous matrix of a column composed of long polymer chains. Each matrix is characterized by its pore size. Hence, the larger the molecule, the less able it is to diffuse through the matrix's pores. This leads to a lower retention time inside the column, as the larger molecules travel less distance given a constant flow rate. So, larger molecules will elute from the gel filtration column in earlier fractions than smaller molecules (i.e. less buffer volume is needed for their elution). Given that the size of globular proteins is closely related to their molecular weight, this chromatography separates proteins according to their molecular weight. However, the expected elution volume of non-globular proteins may differ slightly from their actual elution fraction, as their size is not directly proportional to their molecular weight.

This purification technique is handy for buffer exchange purposes and is a common step for crystallization-bound proteins subsequent to any other chromatography technique described above, as it usually provides a good separation between molecules in the sample (even aggregates or oligomers of the same protein) in order to achieve high purity standards needed for crystallization.

3.1.3.3. Tag Cleavage

It is common for affinity tags to be cleaved after purification. These tags are artificial and add a source of flexibility within the protein which sometimes hamper crystallization. Depending on the cloning vector used and after transcription and translation of the protein of interest, a specific protease cleavage site is added between the tag and the protein. The vector will determine which protease should be used, which should not recognize any other cleavage sites within the amino acid protein sequence. Time, temperature, cleavage system (i.e. batch or column-based) and protease: protein ratio should be optimized in every case.

3.1.3.4. TryR Purification

An affinity chromatography was performed using a HisTrap™ Fast Flow Crude 5 mL column (GE Healthcare), running buffer (50 mM Tris-HCl pH 7, 300 mM NaCl, 25 mM imidazole) and elution buffer (50 mM Tris-HCl pH 7, 300 mM NaCl, 200 mM imidazole). The protein was initially loaded into the pre-equilibrated column (running buffer) for 16 hours at 4 ºC using an external pump and eluted by gradually increasing the imidazole concentration from 25 mM to 150 mM.

TryR-containing fractions were pooled and concentrated using an Amicon® 50 000 MWCO (Merck Millipore) prior to a size exclusion chromatography HiPrep™ 26/10 Desalting (GE Healthcare) for buffer exchange purposes (10 mM Tris pH 7.5, 50 mM NaCl).

A final anionic exchange chromatography was performed using a HiTrap™ Q High Performance column (GE Healthcare), using an elution buffer of 10 mM Tris-HCl pH 7.5, 500 mM NaCl).

6xHistidine-tagged TryR was cleaved by adding the Rhinovirus 3c protease in a 1: 30 ratio (protease: TryR, (m/m)) for 16h at 4ºC. After digestion, GSTrap™ Fast Flow (GE Healthacare) and HisTrap™ Fast Flow Crude were used to separate the 3c protease (by GST-affinity chromatography) and 6xHistidine tags and uncleaved TryR from the cleaved versions of the protein. Cleaved TryR was dialyzed into 50 mM Tris-HCl pH 7 and concentrated again to 7-8 mg/mL approximately.

All purification steps of TryR were carried out using an ÄKTA purification system and Rhinovirus 3c protease was produced at the Department of System Biology at Alcalá de Henares University.

3.1.3.5. CBDP35 Purification

The purification of CBDP35 was also based on affinity chromatography. The soluble fraction was loaded into a HisTrap™ Fast Flow Crude 5 mL column previously equilibrated in running buffer (50 mM Tris-HCl pH 8, 500 mM NaCl, 5 mM imidazole, 0.1 % Tween 20) using an external pump. CBDP35 was eluted by using an ÄKTA purification system and an imidazole gradient (elution buffer: 50 mM Tris-HCl pH 8, 500 mM NaCl, 250 mM imidazole, 0.1 % Tween 20).

CBDP35 was dialyzed into a thrombin cleavage buffer (50 mM Tris pH 8, 150 mM NaCl, 10 mM $CaCl_2$), and concentrated to 5 mg/mL concentration using an Amicon® 5 000 MWCO (Merck Millipore). Thrombin was added in 1: 50 (protease: CBDP35 (m/m)) ratio. Cleaved + tag / uncleaved versions of CBDP35 and thrombin were not separated.

3.2. PROTEIN CRYSTALLIZATION

Once a pure, stable and homogenous protein sample is obtained, crystallization assays may be attempted. Achieving high quality crystals is a crucial step towards the atomic, three-dimensional structure resolution of the molecules forming the crystal. In the case of proteins, crystal formation is a major bottleneck, as these macromolecules may have mobile regions which hamper crystallization. Furthermore, the crystallinity degree (order) obtained within a crystal may

not be sufficient to trigger a diffraction pattern when exposed to X-ray radiation. The packing of protein molecules within the crystals usually leaves a large percentage of the crystalline matter filled with disordered solvent molecules (between 30 and 80 % (Matthews, 1968)). Hence, increased solvent content leads to crystal instability, fragility and poor ability to produce strong diffraction patterns. The main goal of crystallization is to obtain appropriately sized and ordered crystals that scatter X-rays in a way that the diffraction pattern registered is interpretable for structural determination of the molecules building the crystals.

3.2.1. Growing Crystals

Once purified, the first step in any crystallization experiment is to coerce the protein gently out of solution, so that crystals can form in the process of phase separation. The solubility of a protein should be reduced in a controlled manner to avoid its massive precipitation into amorphous precipitates, far from the highly ordered crystals that should be yielded. This controlled decrease in protein solubility can lead to the supersaturation of the solution once the solubility limit of the protein is exceeded. The thermodynamically unstable supersaturated solution will eventually, given the necessary kinetic nucleation events, equilibrate and separate into a protein-rich phase (such as crystals or precipitate) and a saturated protein solution.

The practical way to represent the change of protein solubility and how it approaches the supersaturation state we seek is by means of a solubility diagram (Figure 3.2.).

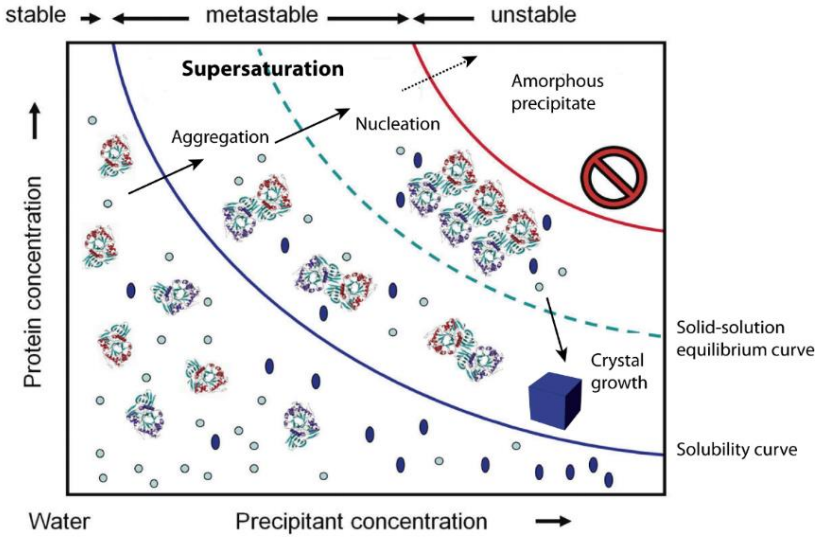


Figure 3.2.: Phase diagram. Arrows indicate el usual behavior of a protein during crystal formation. The metastable area acts as a gradient in which protein molecules gradually aggregate leading to nucleation. (Figure adapted from Rupp, 2010).

In these diagrams, the protein concentration is plotted on the vertical axis, and a second variable in the experiment (usually precipitant concentration), on the horizontal axis. Once the solubility curve is exceeded (either by increasing protein and/or precipitant concentration) and if

this metastable region is not surpassed, the protein solution becomes supersaturated and controlled precipitation occurs through ordered aggregation of protein molecules (nucleation). Moreover, controlled nucleation causes a decrease in protein concentration within the solution, which leads the system to a state in which crystal growth is more energetically favored than the generation of new crystal nuclei. It is here that protein molecules in solution reorder onto the surface of previously formed nuclei. This leads to a gradual increase in size of the crystal. The system will eventually equilibrate when the solubility curved is reached, and so crystal growth ends.

3.2.1.1. Factors Influencing Crystallization Experiments and Crystal Growth

It is fundamental to manipulate certain variables to favor the kinetics and thermodynamics of a crystallization experiment. The main variables are listed below (Begfors, 2009):

· *Protein concentration:* 10 mg/mL is a standard initial protein concentration for crystallization experiments. This concentration will vary according to the solubility and/or stability of our protein. So, the protein concentration should be higher in the case of very soluble proteins in order to reach supersaturation conditions. Nevertheless, a pre-crystallization test is always recommended (PCT, Hampton Research), which shows the trend of a specific protein to precipitate at a precise initial concentration.

· *Precipitating agents:* any compound which facilitates gradual protein aggregation is considered a precipitating agent. Both its concentration and chemical nature are critical and should be assayed in a crystallization experiment. It is common practice to add salts at a low concentration for protein stability (0-300 mM) or different molecular weight polymers (such as polyethylene glycol, PEG), or high salt concentrations. Non-volatile and/or volatile organic compounds, such as alcohols, may also be used.

· *pH:* protein solubility varies according to the pH of the solution it is in. It is recommended to assay a minimum of 2 units below and above the IP of the protein although it has been shown that the correlation between IP and the actual crystallization pH is weak. The strong effect of pH on crystallization success is most likely due to the affectation of the local charge distribution (and therefore creating specific favorable packing interactions) than a net effect on protein solubility. Buffers are generally used at lower concentrations (25-100 mM) and those buffers and/or compounds which tend to crystallize, like phosphate-based buffers, are usually avoided in the crystallization solution.

·*Temperature:* the practical value of temperature selection lies in the control of nucleation a growth kinetics. Generally speaking, lower temperatures mean slower kinetics and decreased thermal agitation, but no prediction of the optimal temperature or a specific crystallization

experiment is possible. Usually, crystallization experiments are set at 18 ºC, but may be performed at 4 ºC depending on the protein's behavior.

· *Additives and cofactors:* low molecular weight compounds may be added at low concentrations to the crystallization solution. These are able to interact with specific areas of a protein, aiding their stabilization and promoting the packing of the protein molecules. Sugars, detergents, ligands, amino acids and lipids are examples of additives used in crystallization and may be used depending on the protein.

3.2.1.2. Protein-Ligand Crystallization

Frequently, it is of our interest to characterize the binding between our protein with another macromolecule, ligand or compound of any sort. Crystallization of protein complexes or proteins bound to a specific ligand can be carried out using the following two approaches (Hassell, *et al.*, 2007):

· *Co-crystallization:* the protein is incubated with the ligand before crystallization assays are sat or the protein and ligand are mixed directly in the crystallization drop. In this way, the complex formed will be the crystallizing unit. Occasionally, incubation of the protein with a ligand prior to crystallization allows the stabilization of mobile regions and adequate packing of the protein.

· *Soaking:* in this case, the protein crystals have been previously formed and are submerged into a ligand-containing solution, which should maintain a similar composition to the one in which the crystals have grown. The ligand will then diffuse through solvent channels within the protein crystal until its binding site is reached. It is of great importance to consider the solubility of the ligand (to facilitate diffusion through the crystal), accessibility of the ligand's binding site (should be exposed) and crystal packing. If the crystal is formed by tightly packed protein molecules, it is likely that the ligand will not be able to diffuse through the solvent channels. This method is commonly used in the case of small-sized ligands.

Each particular scenario is prone to the usage of either one, or both, of these techniques: co-crystallization may change previously optimized crystallization conditions of the protein without the ligand, and ligand binding may cause conformational changes of the protein that can disintegrate or destabilize the previously formed crystal when the soaking technique is used. Incubation or soaking times and ligand concentration and solvent are variables which should be assessed during the experiment.

3.2.1.3. Crystallization Techniques

There are different practical implementations for macromolecular crystallization. Variants apply to each technique and some of them may overlap, albeit they can be classified into four main categories (Chayen and Saridakis, 2008; Rupp, 2010):

3.2.1.3.1. Vapor-Diffusion Techniques

Vapor-diffusion techniques rely on the presence of a reservoir of precipitant that absorbs water from the crystallization drop. In these setups, a droplet of the protein solution is mixed with a droplet of a precipitant solution which are opposed to a reservoir of larger volume and precipitant concentration in a sealed system. In the resulting closed arrangement, water vapor diffuses from the drop into the reservoir, implying a decrease in droplet volume and increase in protein concentration. This will therefore drive the system to supersaturation conditions that will lead to the nucleation and crystal growth phases mentioned above (section 3.2.1.). Once the droplet is in equilibrium with the reservoir, as well as the crystals with the saturated protein droplet, crystal growth terminates, and these will have reached their final size. This method can be easily implemented in two different ways (Figure 3.3):

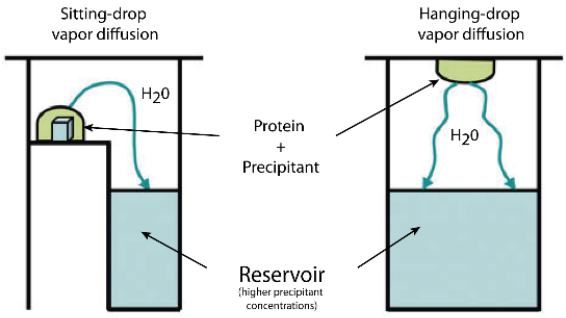


Figure 3.3.: Vapor-diffusion methods. Hanging-drop vapour diffusion is a common method used in small-scale manual setup, while sitting-drop vapor diffusion is preferred with robotic setups. (Figure adapted from Rupp, 2010).

· *Sitting-drop vapor diffusion:* the reservoir solution at higher concentrations is placed in a well surrounding the protein-precipitant mix which rests on a shelf or a post. The system is sealed with a sheet of adhesive tape. This method has been optimized for automation using robotic dispensers.

· *Hanging-drop vapor diffusion:* the protein-precipitant solution drop is placed on a siliconized cover slide. The reservoir or well has a greased rim and is sealed with the flipped-over cover slide.

3.2.1.3.2. Batch Crystallization

In this method, continuous evaporation occurs from the crystallization drop as the system is kept unsealed (Figure 3.4.). In this way, the equilibrium in never reached and the drop may suffer from desiccation. Furthermore, constant evaporation of water and drying of the droplet

causes precipitant precipitation and the loss of crystals (if formed). These experiments should be constantly monitored to determine the optimum time for crystal formation.

Nevertheless, batch crystallization systems are kept isolated by means of water-permeable oils that allow water to diffuse into the environment but reduce the rate of evaporation (Chayen, 1997). These oils are generally a mixture of paraffin and silicon-based oils, and their composition will vary the speed of water evaporation.

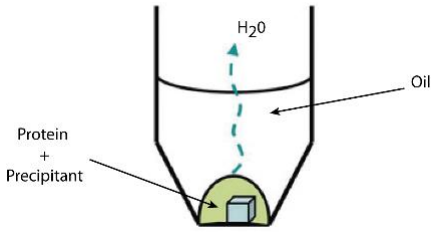


Figure 3.4.: Batch crystallization setup. Additional sealing requirements and the ease of miniaturization factors automated microbatch screening under oil, although crystal harvesting is of higher difficulty. (Figure adapted from Rupp, 2010).

3.2.1.3.3. Dialysis

In this technique, the protein sample is separated from the precipitant by a semi-permeable membrane held by an O-ring. The membrane allows small molecules to pass, but prevents biological macromolecules from crossing the membrane. Crystallization of the sample takes place due to the diffusion of crystallization reagent out of, or into the sample, at constant sample concentration. Salting-in and salting-out are used to reach supersaturation conditions.

3.2.1.3.4. Free Interface Diffusion

The basis of free interface diffusion is to bring the protein solution and precipitant into contact in a narrow vessel without premixing and let the components equilibrate against each other by diffusion only, creating a broad concentration gradient that allows to explore a wide area of the solubility diagram. This technique is used in cases where close-to-optimum crystallization precipitant conditions are known.

3.2.1.3.5. Control of Nucleation by Seeding

As the duration of the crystallization process, from nucleation to the growth of the crystal to its final size, commonly requires several weeks, non-enzymatic hydrolysis of an “unstable” proteins or ligands occurs frequently. Furthermore, in order to achieve the desired balance between nucleation and crystal growth, (i.e. avoiding rapid showers of small crystals, that usually grow irregularly, instead of yielding fewer, larger and better ordered crystals), seeding techniques are implemented (Bergfors, 2003). So, heterogeneous nucleation events are introduced at low

supersaturation, where slow growth is optimal for the formation of diffracting crystals, but spontaneous nucleation is improbable.

· *Microseeding:* few tiny fragments of crystalline matter, of either anisotropic needles, crystal clusters, spherulites of even well-formed crystals, are introduced in a crystallization solution. For this, a dilution series of crushed crystal fragments or other particulate matter is added to the crystallization droplet of somewhat reduced supersaturation, either by pipetting or by streak seeding, in which a whisker or fibre is swiped across a seed crystal solution and streaked through the new drop.

· *Macroseeding*: it implies the transfer of one single, already well-formed small crystal into a new crystallization solution of identical reagents in an attempt to grow the crystal to diffraction size.

3.2.1.4. Identifying Crystallization Conditions and Crystal Optimization

The process of obtaining crystals suitable for diffraction experiments from a protein is cumbersome and not trivial, and so it is the major obstacle in obtaining the detailed 3D structure of a biological macromolecule. Nonetheless, the chemical knowledge of the protein before attempting crystallization trials may increase the chances of success. Due to the many parameters capable of influencing the appearance and growth of a crystal, the process of protein crystallization is not linear, but requires constant re-evaluation of the feedback one gets from many of the steps involved in the procedure in an iterative manner. Although the crystallization conditions vary from protein to protein (and even from versions of the same protein), several procedures that afford the highest probability of success have been proposed (Benvenuti and Mangani, 2007). In general terms, a crystallization experiment can be separated in two different phases (Cudney *et al.*, 1994):

High throughput screening of multiple crystallization conditions is the first step towards the identification of favorable parameters for the crystallization of our protein of interest. Thus, our soluble protein is placed in many different precipitant compositions and the general appearance of the droplets are evaluated regularly. For this purpose, the crystallization robot Innovadine Nanodrop™ ExtY was used during this research for a semi-automated approach in probing initial crystallization conditions identification, using the sitting-drop vapor diffusion method in Innovaplate™ SD-2 plates. The use of robots and small sample volumes (250 nL for the protein solution, 65 µL of reservoir) allows the evaluation of numerous crystallization trials in a cost and sample-effective manner. Along with the Innovadine robot, the commercial crystallization screens Crystal Screen I, Crystal Screen II, INDEX, SaltRX, Detergent Screen and

Additive Screen (Hampton Research); JCSG and PACT Suite (Qiagen); and JBScreen Classic 1-4 (Jena Biosciences), were used.

Usually, crystals obtained in initial screens are not of diffracting quality and need optimization. It is common that reproducibility issues occur in this optimization stage when the crystallization experiments are scaled up to larger volumes of droplet and reservoir solutions aiming to yield larger crystals. Furthermore, subtle deviations from the initial crystallization solution composition are assayed towards examining a broader range of conditions, as well as protein: precipitant ratio (v/v) in the crystallization drop in pursuance of varying the relative protein and precipitant concentration. The use of additive and detergent screens when a favorable crystallization condition has been identified is also a common procedure. It is in this optimization stage when microseeding techniques may be executed.

3.2.1.5. The Real Growth of a Crystal

A single crystal is rarely a true single crystal. The molecules in protein crystals do not arrange themselves flawlessly in each direction across the entire length of the single crystal in a perfect 3-dimensional network during crystal growth. It is typical for crystals to have defects that worsen the ordered positioning of new molecules on the growing crystal surface and hence, the periodicity of the repeating pattern of molecules within the crystal. Different growth islands on the crystal surface may lead to a mosaic crystal of many nearly perfectly aligned domains. The misalignment of the individual domains is described by the mosaicity of the crystal. This parameter is taken into consideration in diffraction data analysis, as each domain will diffract at a slightly different orientation, and the reflections registered in the diffraction pattern will fall at slightly different, overlapping position on the detector. A usable crystal will consist of many domains with a given distribution of domain misalignment of less than 1º or slightly more.

3.2.2. Sample Preparation for X-Ray Diffraction Experiments

After a crystal of adequate size has been grown, it should be cryo-protected, mounted on a system which allows its fixed exposure to X-rays and a 360º rotation, flashed-cooled and transported to the X-ray source used.

The system used comprises a nylon loop (which should be of a similar size to the crystal mounted) joined to a magnetic cap. This cap will then be placed onto a goniometer by magnetism, which permits the centering and rotation of the crystal when immersed in the X-ray beam. To counterbalance the ongoing radiation damage caused during X-ray exposure (Garman, 2010) and to aid the transport of samples to synchrotron radiation facilities, crystals are flash-cooled in liquid nitrogen. Furthermore, a stream of nitrogen at *circa* 100K, permanently showers the crystal

throughout X-ray diffraction experiments, thus, maximizing the "diffractive-life-span" of the crystal.

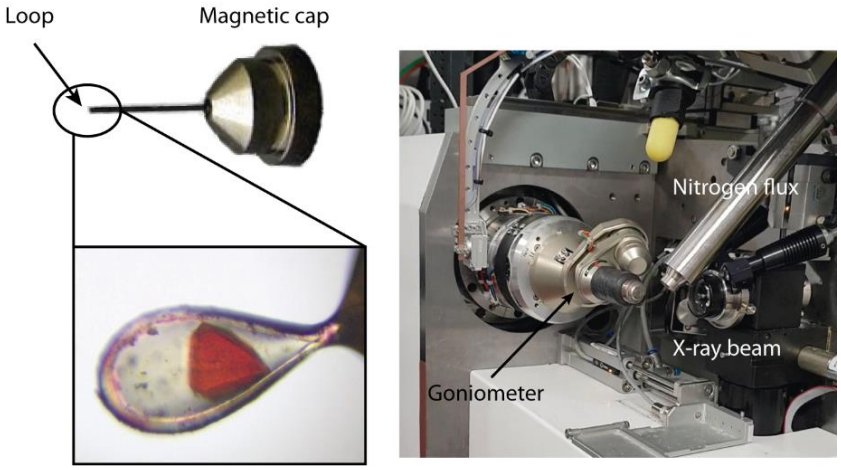


Figure 3.5.: Crystal mounting and placement in an X-ray diffraction system. A safranin-soaked TryR crystal is shown in a vitrified solution and mounted on a nylon loop. The experimental hutch setup at BL13 (Xaloc) at ALBA Synchrotron is shown on the right. The magnetic cap will adhere to the goniometer in order to rotate the crystal when exposed to the X-ray radiation.

If the crystal was to be "fished" onto the nylon loop and directly placed into liquid nitrogen, it is likely that ice crystals would form in the crystallization solution. This usually results in the breakage or deterioration of the crystal and spoiling of the resulting diffraction pattern. It is therefore necessary to soak the crystals in a cryoprotectant solution prior to plunging them into liquid nitrogen. This solution's composition is frequently derived from the crystallization solution and so the crystal is less prone to disintegration due to the changes in its chemical environment. It is therefore interesting to use the minimum cryoprotecting-agent concentration possible that totally prevents ice crystal formation. The cryoprotecting agents most commonly used are glycerol, isopropanol, sugars, MPD, Paratone®, cryosalts and PEGs (Garman and Mitchell, 1996; Rubinson *et al*., 2000; Berejnov et al., 2006).

3.2.3. TryR Crystallization

Diffracting TryR crystals were only obtained from *L. infantum* M/CAN/ES/89/IPZ229/1/89 strain. Prior to the crystallization of this version of TryR, a different *L. infantum* TryR from strain M/CAN/ES/96/BCN150 MON-1, used for dimerization essays at the System Biology Department at Alcalá de Henares University, had been tested. This TryR yielded fragile and non-diffracting crystals despite the efforts in determining its crystallization conditions from the initial stages of commercial crystallization screens.

Optimum crystallization conditions of native TryR comprised a protein concentration of 7-8 mg/mL, a temperature of 18ºC and the vapour-diffusion crystallization method using a

hanging drop layout with a drop volume of 1 µl protein sample, 1 µl crystallization solution and a well volume of 500 µl. This precipitant solution varied from 2-2.4 M $(NH_4)_2SO_4$ and 0.1 M Tris-HCl pH 7.5-8.5. Yellow bi-pyramidal crystals grew within 2-3 days and were then used for soaking experiments.

Although multiple ligands and ligand concentrations were essayed (synthesized at Medical-Chemistry Institute (CSIC), Madrid), both in soaking (in which ligand time exposure was also a variable considered for all inhibitors tested) and co-crystallization methods, the TryR complexes whose structure have been here solved, have been obtained in the following conditions, all by soaking procedures and using a ligand-containing solution of a similar composition to the crystallization condition:

· *TryR·FAD·TRL149*: 1 mM TRL149, 40´ soaking.
· *TryR·FAD·TRL187*: 1.5 mM TRL 187, 30´ soaking.
· *TryR·FAD·TRL190:* 1.5 mM TRL190, overnight (~ 16 hours) soaking.
· *TryR·FAD·TRL156:* 25 mM, overnight (~ 16 hours) soaking.
· *TryR·FAD·TRL187·trypanothione*: 1 mM TRL187, 1 mM TS_2, 1 hour soaking.
· *TryR·FAD·NADPH·trypanothione*: 2 mM TS_2, 5 mM NADPH, 30´ soaking.

It is worth mentioning that FAD was not added to the soaking solution, as FAD is naturally bond to TryR and remains after protein purification, hence the crystals' yellow color. Furthermore, TRLs were dissolved using 100% DMSO due to their insoluble nature.

Crystals were cryoprotected using a modified reservoir solution supplemented with 25 % glycerol prior to mounting and flash-freezing them into liquid nitrogen. If crystals had been previously soaked in a ligand-containing solution, the cryoprotecting mixture also included the corresponding ligand to avoid back-soaking phenomena due to concentration difference.

3.2.4. eEF1A2 Crystallization

Prior to crystallization, the protein was extensively dialyzed against 25 mM MES pH 5.6, 5 mM $MgCl_2$, 2 mM DTT, 5% glycerol, concentrated to around 7 mg/mL and finally supplemented with 100 µM GDP. eEF1A2 concentration was determined spectrophotometrically using ε=28045.380 M^{-1} cm^{-1} as deduced from its amino acid sequence (Swiss-Prot Q71V39). Based on the previously reported crystallization conditions for eEF1A2 (Yaremchuk, *et al.*, 2012), a crystallization grid was carried out across precipitant concentrations of 2 – 2.9 M ammonium sulfate and pH range of 5-7 at 0.1 M and 0.5 intervals respectively. The buffer was changed according to its buffer range (Na citrate or MES buffer used). Crystals were obtained using the hanging-drop vapor diffusion method, mixing 1 µL protein solution with 1 µL reservoir solution

(50 mM Na citrate pH 5/MES pH 5.6, 2.2-2.6 M ammonium sulfate, 10 mM Mg acetate) and equilibrating the drops against 500 µL crystallization solution in the reservoir. The crystals obtained were mounted and flash frozen in liquid nitrogen for data collection without the need for cryoprotection.

It is worth mentioning that numerous attempts were made for eEF1A2·GppNHp and eEF1A2·GppNHp·plitidepsin complex crystallization. Several different protein concentrations with different treatments, GppNHp concentrations, temperatures, buffers and crystallization methods were essayed along with commercial crystallization screens, additive and detergent screens. This will be discussed in the Results and Discussion section.

3.2.5. CBDP35 Crystallization

Crystallization of CBDP35 was attempted by setting up commercial crystallization screens. A hit was obtained in condition C9 of Crystal Screen (4 M sodium formate), thus, the precipitant was buffered by the protein's own solution. Optimized crystals were obtained with the sitting drop vapor-diffusion method, using a protein:precipitant drop volume ratio of 2µl:1µl and 3µl:1µl, and 150 µl reservoir solution. After optimizing the crystallization condition, the best rhombic prism-shaped crystals were yielded in 4.6 M sodium formate.

CBDP35 in complex with *L. monocytogenes* teichoic acid (CBDP35·TA), was finally obtained using co-crystallization methods: CBDP35 was incubated overnight at 4ºC with TAs in a 1:1 molar ratio, followed by crystallization in the conditions previously mentioned. Monomeric *L. monocytogenes* TA was purified by Dr. Shen Yang at the Department of Health Sciences and Technology at the ETH in Zürich.

Whether the crystals were native or product of co-crystallization, they were cryoprotected using saturated lithium sulfate.

3.3. X-RAY DIFFRACTION EQUIPMENT

Once a good quality crystal is obtained, the 3-dimensional structure of its components is calculated by interpreting the diffraction pattern registered when X-ray radiation is applied to the crystal. X-ray diffraction is the physical phenomenon that expresses the fundamental interaction between X-rays and ordered matter, more specifically, when X-rays of specific wavelength interact with the electrons of the atoms in the crystals. Understanding this occurrence and its relationship with symmetry elements within the crystal is essential for the determination of the atomic model of our crystallized molecule. The equipment and elements needed to register the resulting diffraction pattern are commented below.

3.3.1. X-ray Radiation Sources

X-rays are high energy electromagnetic radiation (12.4 keV) with a wavelength of about 10^{-10} m (the equivalent to the unit of length known as one Angstrom). This wavelength magnitude is used as it is close to the interatomic distances found in the crystal-forming molecules. These wavelengths are produced in crystallography laboratories and in large synchrotrons.

3.3.1.1. Rotating Anode Generators

In these generators, a high voltage potential difference is supplied between a negatively charged incandescent filament and a positively charged pure metal (usually copper or molybdenum), which produces an electrical current (of free electrons) between them. Free electrons "jump" from the filament (cathode) to the metal (anode), causing a reorganization of the anode's atomic orbitals. Two different scenarios occur: (i) electrons in the metal return to their initial "relaxed" state emitting energy in the form of X-rays and/or (ii) when the accelerated free electrons come very close to nuclei of atoms in the metal, they are deviated by electromagnetic interactions, losing much energy in the form of electromagnetic radiation (X-rays). In rotating anode generators, the anode takes the form of a cylinder and is maintained in a continuous rotation, so that the incidence of electrons is distributed over its surface obtaining a higher power.

Such X-ray generator was used at the Physical-Chemistry Institute "Rocasolano" (MicroStar, 2.7 kW, Bruker), to discriminate between salt and protein crystals and to identify adequate crystal cryoprotectants.

3.3.1.2. Synchrotron Radiation

A synchrotron facility is composed of an injection system, a booster ring, a storage ring and several tangential beamlines (Figure 3.6.). Within the injection system, electrons are produced in an electron gun, pre-accelerated in a linear injector. These are boosted further in energy (6GeV) and fed into the storage ring (on the order of kilometers), where electrons move at a very high speed in straight channels that occasionally break to match the curvature of the ring. The actual deflection of the electron into a closed path happens in high energy bending magnets. It is at this moment, when electrons change their direction, that they emit a very high energy radiation known as synchrotron radiation tangential to their initial trajectory. This radiation is composed of a continuum of wavelengths ranging from microwaves to the so-called hard X-rays and are collected in the optical hutch in each beamline, where they are filtered, collimated and passed onto the experimental hutch for crystal diffraction.

The X-rays obtained at synchrotrons have two clear advantages for crystallography: the wavelengths can be tuned at will and its brilliance is at least a billion times (10^{21}) higher than those obtained with a conventional X-ray tube.

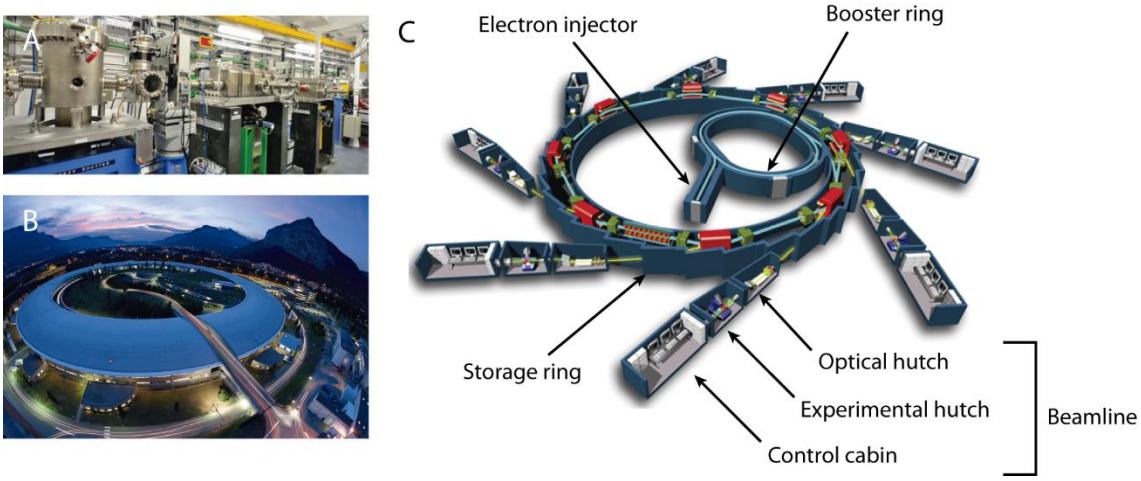


Figure 3.6.: Synchrotron facility and schematics. (A) Optical hutch in BL13 (Xaloc) beamline at ALBA Synchrotron (Spain). **(B)** ESRF (France) synchrotron facility. **(C)** Schematic location of the basic components of a synchrotron.

Diffraction experiments throughout this research have been performed at the beamlines ID29 and ID23-1 at European Synchrotron Radiation Facility (ESRF, Grenoble, France) and beamline BL13 (Xaloc) at the ALBA Synchrotron (Cerdanyola del Vallès, Spain).

3.3.2. X-ray Detectors

3.3.2.1. Imaging Plates

This type of detector was used to register home-lab X-ray diffraction from the rotating anode generator Microstar at the Physical-Chemistry Institute "Rocasolano" to test the salt or protein nature of grown crystals (detector Mar345dtb, MarResearch). This first-generation area detector replaced X-ray film packages and store the X-ray photon energy in a phosphorescent material which coats the surface of the detector, remaining in an excited state for hours. Lasers are used to read out the plates by transforming luminescence centers into a digital image, which are erased and reused, turning out to be slower X-ray detection methods.

3.3.2.2. CCD Detectors

These are predominantly used today and rely in 2-dimensional charge-coupled devices (CCD) semiconductor array chips which directly deliver a digital image of the diffraction pattern when the X-ray photons are absorbed into a fluorescent screen where they are converted to visible light. These eventually generate free electrons in the semiconductor of the CCD in proportion to

the number of photons reached. These detectors exhibit high sensitivity, fast readouts and low noise but are saturated easily. These types of detectors are generally found at synchrotron beamlines and have been used during this research.

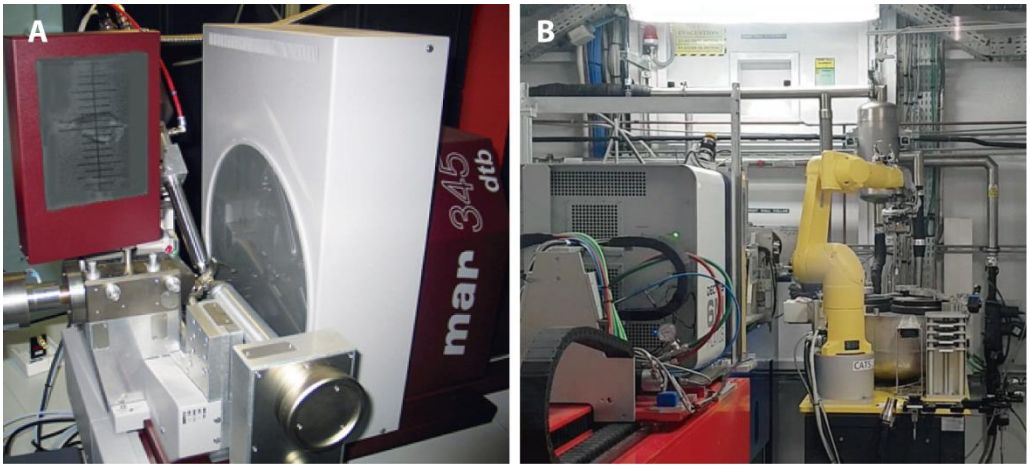


Figure 3.7.: X-ray diffraction detectors. (A). Imaging plate Mar345dtb at the Physical-Chemistry Institute "Rocasolano. (**B)**. Experimental hutch in BL13 (Xaloc) beamline at ALBA Synchrotron (Spain), showing a Pilatus 6M, based on hybrid photon counting (HPC) and single photon counting technology.

3.4. STRUCTURAL DETERMINATION

Once the X-ray diffraction pattern from a single protein crystal is obtained, information should be extracted in order to solve its structure. Diffraction is the interaction of electromagnetic radiation with periodically arranged matter. In order to progress from the scattering of X-rays by periodic arrangements of atoms to solving molecular structures, the fundamentals of diffraction and crystal formation should be described.

3.4.1. Periodic Lattices, Symmetry and Reciprocal Lattices

Symmetry is the consistency or repetition of something in space and/or in time. In our case, protein molecules rearrange in a regular, periodic manner throughout the crystal in order to produce a diffraction pattern. We can therefore describe a full protein crystal by 3D translations of a defined repeating unit or unit cell. The implied translations generate what we call the direct or real lattice, which is basically a network of vectors which describe the atomic positions of every individual atom. The size of the unit cell is determined by the length of three basic vectors (a, b and c) and the three angles between them (α, β and γ). The unit cell contains a motif, usually made up of more than one protein molecule. These protein molecules within the unit cell are related by symmetry elements such as rotations, glide planes and screw axis. It is therefore possible to define the smallest part of the crystal structure, called the asymmetric unit, from which, upon application of space group symmetry operations, the complete unit cell of the crystal is

generated, and so the whole crystal's composition and order within it can be described as a lattice, which is only just a pure mathematical concept.

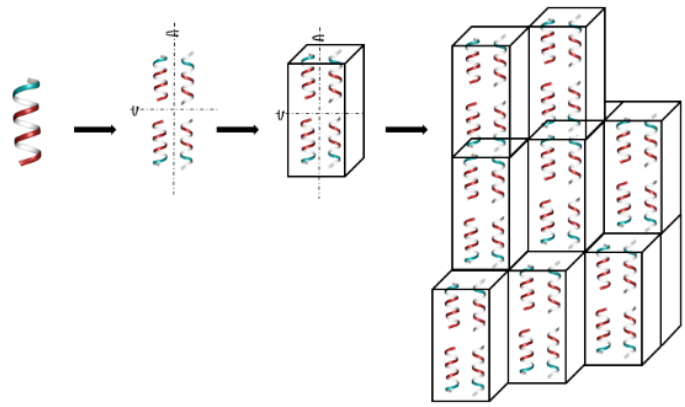


Figure 3.8.: Asymmetric unit, unit cell and crystal representation. The asymmetric unit is repeated within the unit cell and one another can be superposed when the symmetry elements are applied. The unit cell repeats throughout the crystal by translations in the three dimensions.

In periodically repeating systems such as crystals, specific limitations exist as to what operations can be used to create additional copies of the motif in a unit cell, that is, the allowable symmetry operations must be compatible with the translational requirements for the specific lattice and should not generate any changes within the motif. Indeed, combination of symmetry elements and the different 3-dimensional translational lattices possible lead to 230 possible space groups, i.e. different arrangement of motifs in a 3-dimensional periodic crystal structure. The absence of any symmetry in protein molecules as motifs again limits the possible combination of symmetry operations due to the chirality of molecules, resulting in only 65 possible space groups.

A reciprocal lattice may be described from the real lattice depicted above and are related by the reciprocity conditions of its basic vectors. Reciprocal lattice points are defined by Miller Indices (*h, k, l*), which are integers that describe a family of planes in the real lattice. This reciprocal lattice is especially relevant in diffraction experiments as the diffraction patterns obtained as the crystal is rotated in the X-ray beam, is actually the reciprocal lattice obtained from specific planes of the real lattice being exposed to the radiation. So, it is possible to trace back from the reciprocal space shown in the diffraction pattern the associated direct lattice and therefore determine each atom's position.

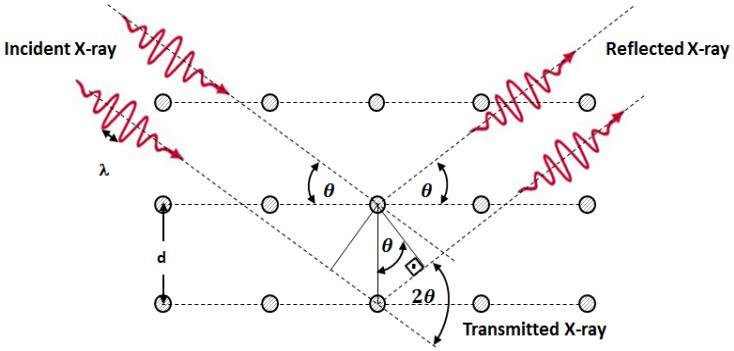
3.4.2. Geometrical and Physical Models in X-Ray Diffraction

X-ray waves interact with matter through the electrons contained in atoms. When the X-rays reach an electron, it becomes a secondary source of electromagnetic radiation that scatters the incident radiation. This scattered radiation waves interfere, causing cancellations and

cooperative scattering of waves, in which their intensity is reinforced. Physical and geometrical models can be used to simplify the diffraction phenomenon:

3.4.2.1. Bragg's Law

Lawrence Bragg stablished in 1913 (Bragg, 1913) that X-ray diffraction by crystals occurs like visual light reflection by imaginary planes described in the crystal lattice. In other words, the hypothesis is to imagine Bragg's diffraction as a reflection of X-rays caused by imaginary "mirrors" formed by atomic planes in the crystal lattice. Due to the repetitive nature of the crystal, these planes would be separated by a constant distance d. So, the diffraction (cooperative scattering) phenomenon will only be produced between parallel planes separated by a constant distant d, for an incident angle θ, when the path difference traveled by the wave fronts an integer number of wavelengths. The planes which satisfy Bragg's law (called crystallographic planes), will produce a characteristic protein diffraction pattern which will be registered. These 3-dimensional lattice planes are indexed accordingly to the reciprocal lattice point they generate (Miller indices). In order to produce all the reflected X-rays possible, the crystal must be rotated in a way that all virtual planes are placed in the right position to obey Bragg's Law.



$$2d \sin \theta = n \lambda \quad (3.1)$$

Figure 3.9: Bragg's Law. When wave fronts reflected are in phase, they interfere constructively and produce a resultant wave which is registered for the set of planes producing the corresponding diffraction. These lattice planes are separated a distance d. Bragg's Law can be formulated by the relationship shown, where θ is the angle of incidence and λ represents the wavelength.

3.4.2.2. Ewald's Sphere

Ewald's sphere (Ewald, 1969) is a representation used to visualize diffraction geometry. According to the wavelength used in each experiment and the tilt of the crystallographic planes to the X-ray beam, this tool allows to describe the spatial disposition of the diffraction pattern's spots and to comprehend the need for rotation of the crystals in the X-ray beam.

In Ewald's model, Bragg's planes are substituted by a geometric node in the reciprocal space. All families of planes described in the crystal will therefore define a net of reciprocal lattice points that is strictly associated with the direct space (or real crystal lattice). It is said that diffraction occurs when each of these reciprocal lattice nodes intersects with the surface of a

theoretical sphere of radius inversely proportional to the wavelength of the X-rays used (Figure 3.10.).

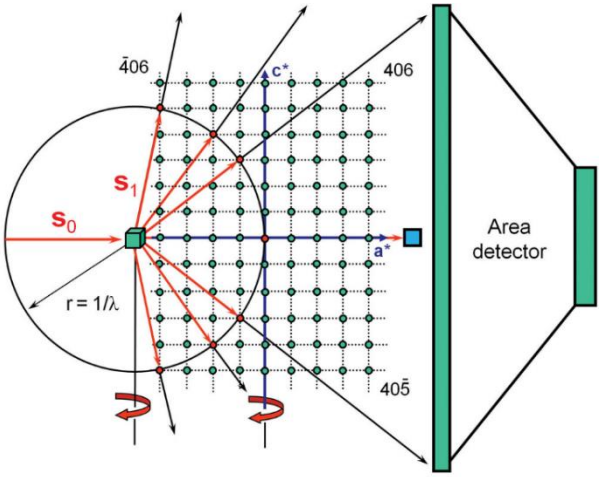


Figure 3.10.: Ewald's sphere. Crystal-centered Ewald's sphere of radius 1/λ allows to define the points observed in the X-ray diffraction pattern registered by the detector, i.e. the interpretation of diffraction in relation to the reciprocal lattice vectors *hkl*. The incoming X-ray beam (S_0) is diffracted as (S_1) when a reciprocal lattice point intersects with the surface of the sphere. a* and c*, being b* perpendicular to the surface of the paper, are the reciprocal lattice axis with origin at (0, 0, 0), corresponding to the red spot on the surface of Ewald's sphere. (Figure reproduced from Rupp, 2010).

The diffracted rays will be registered by an accordingly positioned detector. However, in practice only a few reciprocal points will lie on the Ewald sphere right away, and we will observe only a few diffraction spots at any given orientation of the crystal. The other reciprocal lattice points can be brought to intersection with the Ewald sphere by rotating the crystal, which automatically rotates the crystal's reciprocal lattice. This model allows to visualize the link between real and reciprocal lattice points recorded in a diffraction experiment.

3.4.3. Extracting Information from the Diffraction Pattern

Interpretation of the diffraction pattern is fundamental to recreate the real space lattice inside the crystal. Each spot in the diffraction pattern (described by the Miller indices) gives information on a point in the reciprocal lattice whose position vector is normal to a set of planes in the real lattice (also described by the *hkl* indices). A diffraction image is a 2-dimensional representation of the reciprocal lattice points obeying the laws defined in 3.4.2. It is therefore crucial to determine both the position and intensity of the spots registered in the detector. The position of a spot will be assigned a set of coordinates in the reciprocal lattice and the intensity relates to the shape and size of the spot (i.e. the number of pixels inside a defined area). The distribution of the spots and relation between their Miller indices will reflect the symmetry within the reciprocal lattice which is directly related to the real space crystal point group symmetry.

The symmetry inside the crystal and its resistance to radiation exposure will determine the optimum collection strategy. It is therefore important to align the crystal in a way that all

necessary reciprocal lattice points are brought into diffraction condition and so obtain a complete and redundant diffraction set of data.

Other information extracted visually from a diffraction pattern can be related to the presence of salt, ice, radiation damage, crystal twinning, diffraction resolution, etc.

3.4.4. Diffraction Data Analysis

Diffraction data analysis consists of three individual stages: indexing, integration and scaling of the intensities obtained from the images collected throughout the rotation of the crystal in the X-ray beam. Prior to the description of these stages, it is worth to briefly describe the main goal of diffraction data analysis: to rebuild de scattering units in the real lattice, the electrons, which can be described by the following formula:

$$\rho(x,y,z) = \frac{1}{V}\sum_{-h}^{h}\sum_{-k}^{k}\sum_{-l}^{l}|F_{hkl}|e^{-2\pi(hx+ky+lz-\varphi hkl)} \qquad (3.2.)$$

Where ρ (x, y, z) represents the electron density in each point of the unit cell in the real space. Other variables in the equation are: V, volume of the unit cell; F_{hkl}, representing the resultant diffracted beams of all atoms contained in the unit cell in a given direction. These magnitudes (actually waves), one for each diffracted beam, are known as structure factors and their moduli are directly related to the diffracted intensities; and φ_{hkl}, represent the phases of the structure factors, meaning that, when scattered waves interact with each other producing a single diffracted beam in each direction of space, they can add or subtract according to the difference in distance each individual wave has travelled. At those points where this function takes maximum values (estimated in terms of electrons per cubic Angstrom) is where atoms are located. This means that if we are able to calculate this function, we will "see" the atomic structure of the crystal. In order to calculate the value of the electron density in a single point of coordinates (xyz) it is necessary to use the contributions of all structure factors produced by the crystal diffraction.

The analytic expression of the structure factors, F(hkl), is simple and involves a new magnitude (f), called atomic scattering factor, which takes into account the different scattering powers with which the electrons of the atoms (j) in the unit cell scatter the X-rays:

$$F_{hkl} = \sum_{j} f_j e^{2\pi(hx_j+ky_j+lz_j)} \qquad (3.3.)$$

Experimental-wise, each spot in the diffraction pattern has a structure factor associated to it (a resulting diffracted wave reaching the detector). Thus, the structure factors F(*hkl*) are waves and therefore can be described by their amplitudes, [F(*hkl*)], and phases φ(*hkl*), the unknown

variables in equation 3.1. The magnitude (moduli) of the structure factor associated to a specific reflection in the diffraction pattern is therefore calculated from the intensities of each spot (information extracted directly by analyzing the diffraction pattern just after scaling), and follows the following relationship:

$$I(hkl) \propto |F(hkl)|^2 \tag{3.4.}$$

On the contrary, information on the relative phases of the waves reaching the detector [φ(*hkl*)], is lost during the experiment. This is referred to as "the phase problem" and will be discussed in the following section.

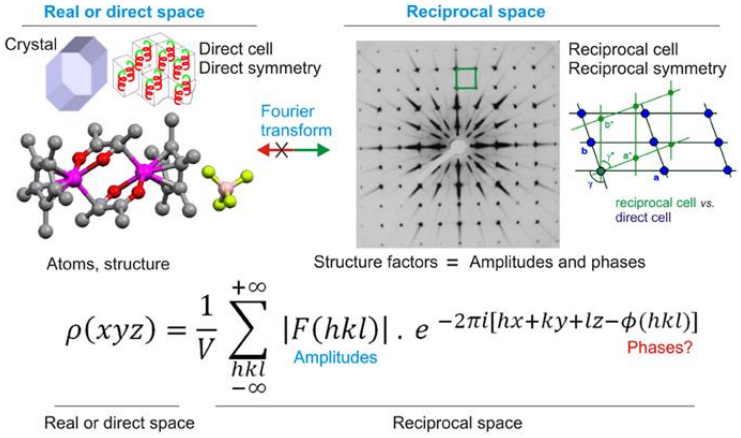


Figure 3.11.: Outline on basic crystallographic concepts, direct and reciprocal space. The issue is to obtain information on the left side (direct space) from the diffraction experiment (reciprocal space). (Figure: http://www.xtal.iqfr.csic.es/Cristalografia/index-en.html).

All in all, between the two mentioned spaces (direct and reciprocal) there is a holistic relationship, which mathematically speaking, is a Fourier transform that cannot directly be solved due to the "phase problem". Therefore, equation 3.2. represents the Fourier transform between the real or direct space (where the atoms are, represented by the function [ρ (x, y, z)] and the reciprocal space (the X-ray pattern) represented by the structure factor amplitudes and their phases.

The following sections describe the stages in diffraction data analysis:

3.4.4.1. Indexing

The first stage of data processing involves the finding and positioning of the diffraction spots in the reciprocal lattice. In other words, some diffraction images obtained are "indexed" by allocating Miller indices to the reflections registered in the diffraction pattern. This information is used to calculate the parameters of the unit cell (a, b, c, α, β, γ) and estimate possible space

groups that could describe the symmetry within the crystal lattice. Crystal mosaicity is also estimated.

Software used for this process in this research were iMOSFLM, of the CCP4 software package (Battye, *et al*., 2011) and XDS (Kabsch, 2010).

3.4.4.2. Integration

It is here that the reading of detector pixels occurs for each reflection and combined into raw reflection intensities. It is worth noting that, as the crystal is rotated and due to its symmetry, the same spot (with the same Miller indices) can be measured in more than one diffraction image obtained throughout the experiment. Up to this point, multiple incidences of the same reflections are treated independently. The integrated raw data are essentially a long list of indices, the reflection intensity of each individual reflection, its standard error, and additional batch information for each frame. The term multiplicity should be here introduced, which reflects the number of times that the same reflection and intensity has been recorded. The more times the same reflection has been recorded, the higher the precision of its intensity estimation is. A second parameter, the completeness of a dataset, refers to the percentage of the reciprocal space measured in the experiment (circa 90% is acceptable).

3.4.4.3. Scaling and Merging: Initial Data Reduction

In addition to the spatial integration, we also must account for the fact that the X-ray reflections were temporally separated during the recording. Thus, the intensity of the radiation flux may vary throughout the experiment and the radiation damage is accumulated towards the end of the experiment. Furthermore, the crystal may be irregular, meaning that, as it is rotated, there may be variations in the crystalline mass exposed to the X-rays. So, after the integration of the reflections, these should be properly scaled and combined to account for these deviations. The initial reduction of data begins with the merging of the multiple measurements of identical reflections according to the initial indexing. The result is a reduced data set that contains all reflections, but yet the further merged symmetry equivalents. Programs such as SCALA (Evans, 2006) and AIMLESS (Evans and Murshdov, 2013), the latter containing the program POINTLESS for symmetry determination, where used for data integrated with iMOSFLM and XSCALA, for XDS integrated data and contained in the package, were used for this step.

3.4.4.4. Statistics

At this point data quality should be analyzed. The intensity of the reflections becomes weaker with increasing resolution. Data collection statistics are thus provided as overall numbers

for the entire data set as well as being binned in resolution shells with an equal number of reflections (Evans, 2011). The following statistics (Rupp, 2010) are therefore scrutinized before data reduction (i.e. resolution cutoff, outliers…).

3.4.4.4.1. Signal-to-Noise Ratio

The consequence of increasing relative error in the measurement in wear reflections is a diminishing signal-to-noise ratio. This provides an estimate for the usefulness of the data as a function of resolution. The expression used to define this parameter is expressed as an average [|I|/σ(I)] summed over all N reflections in a resolution shell:

$$\langle \frac{|I|}{\sigma(I)} \rangle = \frac{1}{N} \sum_{hkl}^{N} \frac{|I_{(hkl)}|}{\sigma(I_{hkl})} \tag{3.5.}$$

A general criterion suggests a resolution cutoff is when [|I|/σ(I)] > 1.5 – 2.

3.4.4.4.2. R_{merge}

The most commonly used quality indicator which describes the disagreement when merging reflections within a resolution range or symmetry-related reflections, is a linear merging R-value,

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^{N} |I_{hkl,i} - \bar{I}_{hkl}|}{\sum_{hkl} \sum_{i=1}^{N} I_{hkl,i}} \tag{3.6.}$$

where the inner summation extends over all N redundant observations for a given flection and $\bar{I}_{hkl}$ is the averaged intensity of each reflection. The outer summation extends over the desired resolution range. Depending on what kind of reflections are actually being merged, this parameter can be termed R_{sym} for merging of symmetry-related reflections, R_{int} for general merging, etc. The lower the value, the more accurate the intensities are throughout the diffraction experiment. This statistic is convenient for high intensity reflections with little error but may drastically increase if the multiplicity is too high, and so should not be considered when determining the resolution cutoff (Evans, 2011).

3.4.4.4.3. R_{pim}

The "precision-indicating merging R-value" accounts for the increase in precision when the multiplicity is high, as the intensities are actually becoming more precise as more and more observations are being merged. It can therefore be considered a "corrected-R_{merge}" statistic in which a normalized N-term is included into the R_{merge} equation.

$$R_{pim} = \frac{\sum_{hkl}\left(\frac{1}{N-1}\right)^{1/2}\sum_{i=1}^{N}\left|I_{hkl,i} - \bar{I}_{hkl}\right|}{\sum_{hkl}\sum_{i=1}^{N} I_{hkl,i}} \quad (3.7.)$$

The precision-indicating merging R-value R_{pim} decreases with redundancy N and seems to be more useful than R_{merge}. Values lower than 0.4 is an acceptable criterion for a resolution cutoff.

3.4.4.4.4. CC^* and $CC_{1/2}$

The indicators previously described report on data precision, so if substantial systematic errors are present the indicators need not reflect the data accuracy. Nevertheless, CC* and $CC_{1/2}$ have been reported to be indicators of first rank utility (Karplus and Diederichs, 2015) for assessing the precision of the merged data. These are Pearson-related correlation coefficients (CC), which is a parameter that could potentially assess both data accuracy and the agreement of model and data on a common scale. $CC_{1/2}$ is calculated between independent sets of observations characterized as a function of resolution and is related to the effective signal to noise of the data. CC values range from 1 to –1 for perfectly correlated versus uncorrelated data, but for properly indexed data these indicators should range from near 1 for highly precise data (lower resolution ranges where intensities are stronger) to near 0 for very imprecise data (higher resolution ranges where signal-to-noise ratios are higher).

CC* provides the potential for a cross-validation independent indication of overfitting and provides an estimate of the CC that would be obtained between the final merged data set and the unknown true intensity values that they are representing. CC* is calculated from $CC_{1/2}$ by the following expression:

$$CC^* = \sqrt{\frac{2CC_{1/2}}{1 + CC_{1/2}}} \quad (3.8.)$$

An advantage of CC-based indicators is that they have well-studied statistical properties so that, for instance, given a CC value and how many observations contributed to it, one can calculate the probability that this value has occurred by chance.

3.4.5. The Phase Problem

As mentioned previously, the phase problem occurs when the structure factor phases (φ_{hkl}) are lost during the diffraction experiment (only the moduli ($|F_{hkl}|$) are obtained which are related to the intensity of each reflection), and so the electron density cannot be calculated (equation 3.2). Phasing techniques need to be used to estimate preliminary phases for each measured reflection which will be later improved to suit experimental values and decrease initial

bias. The available phasing methods can be roughly categorized into four general groups (Rupp, 2010), briefly explained below with special emphasis given to molecular replacement, which has been the method used in this research.

3.4.5.1. Direct Methods

The methodology generally exploits constraints or statistical correlations between the phases of different Fourier components, in other words, between different sets of structure factors. This method is established in terms of probability, thus, the direct methods use equations that relate the phase of a reflection with the phases of other neighbor reflections, assuming that these relationships are "probably true" by limitations in the distribution of phases associated with the structure factors (inferred from atomicity of molecules). Direct methods usually require high resolution data (below 2 Å) and small sized-proteins with a predicted high percentage of α-helices for *ab initio* determination of a protein structure, although great developments have been made to date (Millán, Sammito and Usón, 2015).

3.4.5.2. Molecular Replacement

This method requires the availability of a structurally similar model as a molecular search probe, thus based on the idea that proteins with similar peptide sequences show a very similar folding. Initial phases are taken from the ones obtained for the homologous protein and are later optimized. The 3-dimensional model of the known homologous protein (search model) is combined with the experimental data obtained from the protein crystal whose structure is pending. Transferring the molecular structure of the known protein from its own crystal structure to a new crystal packing may be challenging. The positioning of the known molecule into the unit cell of the unknown protein requires determining its correct orientation and position within the unit cell and is reliant on Patterson-based rotation and translation functions (Figure 3.12.).

As the number of structures present in the PDB increases, the possibility that an unknown structure has some features in common with one or more already characterized protein becomes greater. Molecular replacement is therefore the main method for solving the phase problem. Nevertheless, in more difficult cases, it is possible that an adequate probe molecule is lacking (i.e. homology above 30% does not apply to the full length of the target protein – general criteria to carry out molecular replacement procedures). In these cases, a probe model should be prepared or a “domain search” performed, in which different conserved regions (usually domains) of a different probe structure are translated and rotated into the target structure, hoping to localize a specific domain folding.

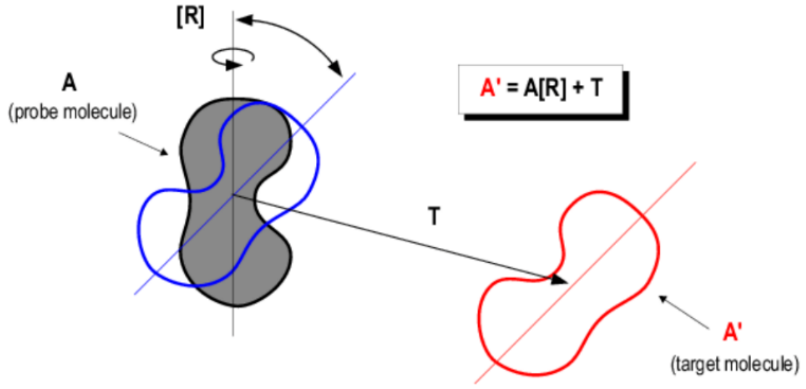


Figure 3.12.: Principle of molecular replacement. A series of translation (T) and rotation (R) operations are carried out on a probe molecule (A) with similar sequence to the target molecule whose structure is unknown (A') in order to supply an initial set of phases to solve the crystal structure. If the probe molecule is similar to the target molecule, the bias of the resulting electron density will be minor, and so the solution of molecular replacement will be similar to the real space of the crystal. (Figure adapted from http://www.xray.bioc.cam.ac.uk.)

A scoring system, as well as visual analysis, is needed to determine whether the molecular replacement has been, or not, successful. The Z-scores (RFZ – rotation function Z-score, TFZ – translation function Z-score) are calculated when the rotation and translation functions are applied to the probe model. Z-scores > 7 criteria are used to define a plausible molecular replacement solution. Another parameter, the contrast score, gives a ratio between the top score of an applied combination of rotation and translation functions and the average. Visual analysis of density fit and clashes between symmetry generated molecules is also performed.

Once the probe is correctly positioned, phases can be calculated from this model and combined with the observed amplitudes to give an approximate Fourier synthesis of the structure of interest. This model will then have to be carefully analyzed, and through several refinement and model rebuilding steps to account for the differences between the probe and the target molecule, will provide a model that maximizes the agreement with the experimental data.

Software used for structural resolution by molecular replacement in this Thesis where MOLREP (Vagin and Teplyakov, 1997) and Phaser (McCoy, *et al.*, 2007) of the CCP4 software package.

3.4.5.3. Marker Atom Substructure Methods

If the homology of the unknown protein is below a certain threshold (~30%) to any protein of known structure, other techniques should be used to obtain the real phases from diffraction experiments. The following methods share the same principle and do not depend on prior structural information, but on the determination of a marker atom substructure from diffraction data, generally by Patterson search methods or direct methods, which is taken as a

reference to calculate the remaining protein structure (Giacovazzo, 2013). Two main strategies can be described:

· *Multiple Isomorphous Replacement (MIR):* this method is applied after introducing "heavy" atoms (large scatterers) in the crystal structure. However, the difficulty of this methodology lies in the fact that the heavy atoms should not affect the crystal formation or unit cell dimensions in comparison to its native form. The heavy atoms will alter the X-ray scattering in a way that they can be localized and estimate initial phases.

· *Single/Multi-wavelength Anomalous Diffraction (SAD/MAD):* changes can be caused in the intensity of diffraction by modifying the physical properties of atoms (i.e. introducing heavy atoms). Thus, if the incident X-ray radiation has a frequency close to the natural vibration frequency (resonance) of the electrons in a given atom, the atom behaves as an "anomalous scatterer". The intensity changes generated in the diffraction pattern can be used to estimate initial phases. Depending on the number of wavelengths used in the diffraction experiment, we refer to single (SAD) or multiple (MAD) anomalous diffraction. Compared to MIR, the MAD technique uses data collected from a single crystal, the problems derived from lack of isomorphism, common in the MIR method, do not apply.

For these techniques, heavy atom derivatives need to be prepared by introducing heavy atoms (Hg, Se, Sm, Mo…) in the crystal (Garman and Murray, 2003), which should be ordered within it. This can be done by either:

·*Soaking techniques:* as mentioned in section 3.2.1.2., a pre-formed crystal of the native protein is submerged into a solution containing the heavy atom. The heavy atom should then diffuse through the solvent channels and bind to the side chains of specific regions which will stabilize the anomalous scatterer in ordered regions within the crystal. Just as for ligand binding, different heavy atom concentration and soaking time should be tested. Furthermore, co-crystallization techniques can also be used for the same purpose.

·*Substitution:* native atoms of the protein are substituted by heavy atoms. It is of general practice that minimum culture media supplemented with modified amino acids (i.e. Se-methionine or Se-cysteine, in which the sulfur atom is substituted for selenium) are used for protein expression, and so these heavy-atom amino acids are incorporated into the primary structure of the protein. Other options are, in the case of metalloproteins, one can replace their endogenous metals by heavier ones (i.e. Zn by Hg, Ca by Sm, etc.).

3.4.5.4. Density Modification Techniques

The phase improvement procedure is usually referred to as density modification because new phase estimates are obtained by modification of the electron density to conform prior expectations. These techniques are used after the substructure phasing methods in practically all de novo structure determinations. They are used to improve already somehow-calculated phases and maps (Cowtan, 2010). They do not require an atomic model and can be considered as an extension of experimental phasing. The result of this procedure should be a better set of phases, that can be used to make a map for interpretation (i.e. model building).

3.4.6. Electron Density Calculation and Structural Model Building

Once a set of initial phases and the structure factors are obtained from the diffraction experiment, a first electron density function can now be calculated. The analysis and interpretation of the electron density function leads to an initial distribution of atomic positions within the unit cell which can be represented by points or small spheres to build an initial structural model. According to the quality of the electron density map (i.e. resolution), the structural model can be built automatically, using for example ARP/wARP (Lazmin, *et al*., 2012), Autobuild (Terwilliger, 2008) or Buccaneer (Cowtan, 2006) software for high-quality automated model building and refinement, which majorly contribute towards automation and high-throughput structure resolution. This is also the case of the server Auto-Rickshaw (Panjikar, *et al*., 2005), which combines both phase determination and automatic structure determination. Nevertheless, inspection of the model fitting is always needed, as well as (sometimes challenging) manual reconstruction, especially in cases where initial phases from which the electron density is reconstructed are inaccurate or resolution is low. For this purpose, Coot software (Emsley, *et al.*, 2010) has been used for the model building of the structures here solved.

Two types of maps are calculated: basic difference maps $(F_o\text{-}F_c)\cdot\exp(i\varphi_c)$ and combined $(2F_o\text{-}F_c)\cdot\exp(i\varphi_c)$ maps (Rupp, 2010), where F_o represents the observed structure factor amplitudes and F_c and φ_c come from the starting phasing model. Difference maps show red and green contours, which represent negative density values (regions where the model has electron density calculated for but where there should not be density according to experimental data) and positive density values (regions where the model does not have density calculated for but where there should actually be density), respectively. Combined maps can be interpreted as a combination of the $(F_o\text{-}F_c)\cdot\exp(i\varphi_c)$ map and a $(F_o)\cdot\exp(i\varphi_c)$ map, and is contoured to amplify positive density and is well suited to early model building stages. These maps allow the correct tracing of the model to suit the observed experimental intensities.

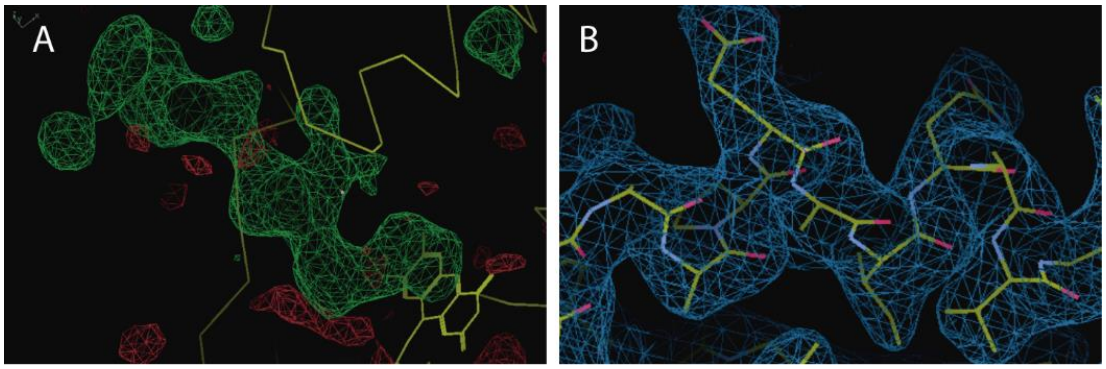


Figure 3.13.: Electron density maps. (A) Difference map (Fo-Fc) used for TryR·NADPH·FAD·trypanothione complex. It shows positive density (coloured in green) indicating the presence of the NADPH molecule stacking onto the FAD molecule, shown in yellow sticks. **(B)** 2Fo-Fc map, of the same structure, but of a well-traced region of the dimerization interface, more specific, E436, crucial for TryR dimerization.

3.4.6.1. Structural Refinement

Refinement is the word given to iterative adjustment of variable parameters of a model so that the fit between model and observed data is optimized (Rupp, 2010). The atomic positions of the first solution obtained after an initial approximation of phases are rarely correct, so

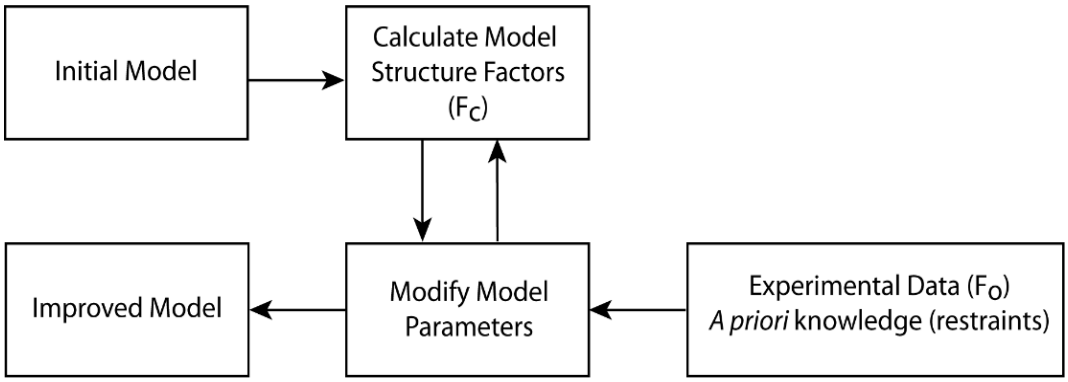


Figure 3.14.: "The circle of refinement". Having constructed an initial model from the first obtained electron density, structure factors are calculated (|Fc|, φc). These new phases should be closer to the real phases. Nevertheless, the moduli of the calculated structure factors should be discarded in order to use the measured experimental amplitudes (|Fo|). Thus, a new electron density is calculated using |Fo| and φc, closer to reality and allowing further construction and modifications of the model according to human interpretation in an iterative cycle of improved phase calculation.

interpretation of the electron density is usually necessary (i.e. errors in side chains or main chain disposition, sequence and/or ligand presence should be accounted for). In general terms, the refinement process involves the modification of the atomic positions of the first model obtained, calculating the resultant phases and structure factors of this newly generated model and contrasting it with experimental structure factors in order to re-determine a better quality electron density map from the more precise newly curated phases (Figure 3.14). From the new electron density map, more accurate atomic positions can be derived, which lead to even better phases, and so forth, thus iteratively eliminating errors introduced during initial model building.

REFMAC5 (Murshdov, *et al.*, 2011) from the CCP4 software package and Phenix.refine (Afonine, *et al.*, 2012) from the Phenix software package have been used for this purpose.

The goal of macromolecule refinement involves the minimization of a target function (or maximization of its likelihood), which is usually a maximum likelihood function. The maximum likelihood principle states that "the best estimate for any parameter is the value that maximizes a likelihood function, that is, the joint probability distribution of experimental outcomes that can accommodate incompleteness and various errors (as well as prior information) of the model" (Rupp, 2010). How the target function is minimized depends to a degree on the choice of target function, although the aim is to search for a global minimum for a real space scenario.

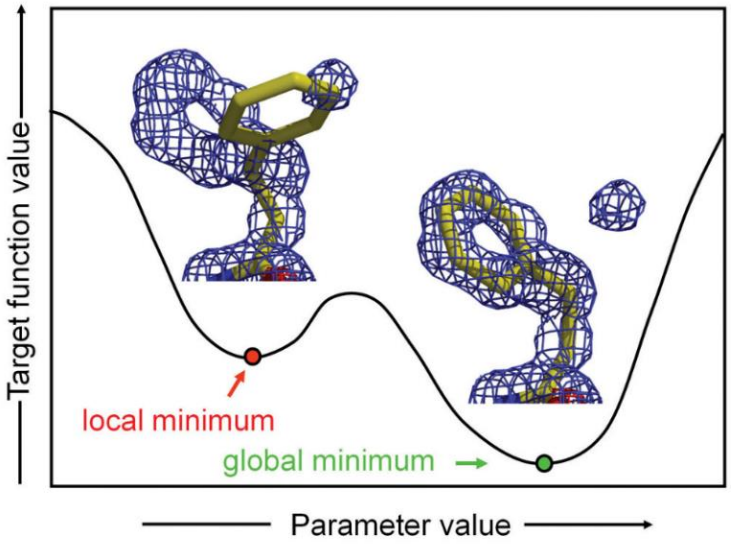


Figure 3.15.: Aiming for a global minimum in refinement. Sometimes, a refinement program may not be able to proceed upwards over the activation barrier, (or may allow only limited positional parameter shifts), which refers to an increase in the target function, thus preventing the large movement of the entire side chain out of the partial density until it snaps into the correct electron density. Either manual modelling or parameters optimization is needed to obtain a global minimum. Figure adapted from Rupp, 2010.

3.4.6.1.1. Molecule Parametrization

Certain restraints are considered during refinement, referring to the assumptions used to introduce chemical or physical information into a refinement as additional experimental observations. These restraints include bond angles, bond distances, chirality, torsion angles, etc. (Evans, 2007). Other experimentally extracted parameters which are considered and "tuned" (parametrization of macromolecular structures) according to the quality of experimental data, of current model and the data-to-parameters ratio during refinement stages involve:

· *Atomic coordinates:* exact positions of atoms, described by (x, y, z) coordinates, in the real space.

· *B-factor (atomic displacement parameter):* term associating each atom and its thermal vibrational state, defined individually or as group B-factors, and is related to the atomic positional

deviations from its mean position due to agitation of the molecule. Furthermore, B-factors describe relative positional uncertainty. In a first approach, their thermal vibration is considered as isotropic (spherical) around its position of equilibrium. Thus, its value refers to the radius of the sphere of vibration. However, when dealing with high resolution data, the state of vibration can be considered anisotropic, i.e. distinguishing between different directions of vibration in the form of an ellipsoid, representing a more realistic model of agitation due to bond restrictions (Merritt, 2012).

· *Occupancy:* fraction of atoms or molecules that actually occupy a crystallographic position. In other words, in the case of solvent molecules, ligands and side chains or loops with multiple conformations, the specific position given to an atom may not be fully occupied by it, and so a fraction $0 \leq n \geq 1$ is given to describe its presence throughout the crystal.

· *Non-Crystallographic Symmetry (NCS):* Nearly half of all structures in the PDB are oligomers. NCS involves the presence of more than one copy of a motif in the asymmetric unit. Molecules in the asymmetric unit may be related through a general combination of rotation and translation, nonetheless, NCS accounts for local symmetry and is not limited to crystallographic symmetry. Refinements can be restricted to this new generated symmetry.

· *Translation-Libration-Screw (TLS) parametrization:* this parameter describes more complex, anisotropic motions within regions of the molecule that move as rigid bodies. Which atoms form a group of correlated movement can either be defined directly from knowledge about the structure (hinged domains of a multi-domain structure) or even from analysis of the isotropic B-factors. The interesting part from a biological point of view is the anisotropic TLS movement of groups or domains.

· *Rigid body refinement:* interatomic distances are kept constant, that is, the whole molecule moves as a rigid unit. Useful at low resolutions, for example after molecular replacement and lower resolution data refinement.

· *Restrained refinement:* the model is refined to give the best fit to the experimental density while maintaining good geometry.

· *Unrestrained refinement:* this mode of refinement does not restrain the model geometry. It is only useful at high resolution.

3.4.6.1.2. Refinement Optimization Methods

Refinement optimization methods are available to improve the calculated maps at lower resolutions in order to obtain reliable models. The following have been used for the structural determination of proteins here presented:

· *Jelly-body refinement:* REFMAC5 includes this tool which involves a regularization function in interatomic distance space, allowing flexibility of the molecule in order to facilitate the placing of domains or regions of the protein with higher mobility. Useful for lower resolution data refinement.

· *Simulated annealing:* refers to a general technique of energy refinement, suited by the multiple energy minima characteristic of crystallographic refinement, by successively minimizing the potential energy of a perturbed model that slowly returns to equilibrium. Alternatively, the model is heated and progressively cooled into a different state minimum energy which may differ from the initial state. It is useful in normalization of an initial structural solution after molecular replacement (Adams, et al., 2010) to reduce phase bias.

· *Map Sharpening:* implemented by REFMAC5, map sharpening attempts to remove overall B-values and automatically produces a map with more structural features (optimization of details) by enhancing differences between solvent and macromolecular electron density. This sort of density modification procedures will be commented below. Useful for lower resolution data refinement (Nicholls, et al., 2012).

3.4.6.1.3. Monitoring the Refinement

In order to monitor the state of the refinement stage, a statistic is calculated which determines the disagreement between the model and the experimental diffraction data (Rupp, 2010), so that we obtain a best fit between the observed structure factor amplitudes and the computed model structure factor amplitudes. This overall fit is numerically quantified by the R-value:

$$R_{work} = \frac{\sum_{hkl}\left||F_o| - |F_c|\right|}{\sum_{hkl}|F_o|} \quad (3.9.)$$

Acceptable values of R_{work} depend on the data resolution. In other words, the model should agree to a higher degree with the experimental data at higher resolution values. Tolerated R_{work} values comprise lower percentages than 25% for medium to low resolution data, whilst circa 10% error is expected for high resolution data.

A random sample of reflections, usually around 5-15% of the data, is set aside for the purpose of R_{free} calculation and never included in the modelling or refinement stages. R_{free} value is calculated to asses over the modelling of the structure and depicts the agreement between observed and computed structure factor amplitudes for this "test" set of data. (Brünger, 1992). In this way, R_{free} reflects if the model can predict a set of reflexions which have not been used to calculate the density map used for constructing the model. Furthermore, there is a high correlation

between R_{free} and the accuracy of the atomic model phases. This is useful because experimental phase information is usually inaccurate, incomplete or unavailable.

The repeated alternation between local real space rebuilding and refinement is in general a successful combination. Nevertheless, it is possible that the model structure falls into a false local refinement minimum that the refinement itself cannot recover from. R-values may not reflect this situation as they are a rough global quality indicator that reflects nothing about which parts of the model are good or not. Moreover, on the contrary to when a global minimum is approached in which R_{work} and R_{free} values converge, R-values disband to over 10% of R_{work}'s value if over-refinement occurs.

3.4.6.2. Other Generated Maps

There are at least three different factors that routinely affect the quality of crystallographic maps used during the structure-solution workflow, which make the interpretation of crystallographic maps ambiguous, nontrivial or non-unique at typical macromolecular resolutions of approximately 1–4 Å. (Afonine, et al., 2015): errors in and incompleteness of the data, signal weakness and model bias. Along with choosing a correct weighting term (Read, 1986), density modification techniques (Cowtan, 2010) and obtaining maps derived from ensembles of perturbed models or structure factors (Lang, et al., 2014), such as with simulated annealing, OMIT maps are a common tool to overcome these errors and to verify the presence of atoms in the model, specially ligands. They can be defined as the electron density reconstruction in whose computation, a questionable part of the model has been omitted (Rupp, 2010). The following maps have been generated when the presence of ligands has been doubtful.

· *Polder maps*: a polder map is an omit map which excludes the bulk solvent around the omitted region. This way, weak densities, which can be obscured by bulk solvent, may become visible (Liebschner, *et al*., 2017). So, a residual map is calculated by updating the structure factors when the region of interest (i.e. ligand) is omitted, and so, it is then expected that if these atoms are present in the crystal structure, the electron density for the omitted atoms will be seen as positive features in this map. Furthermore, the presence of a ligand is confirmed by CC calculations which relates the positive density observed to bulk solvent density or to the atomic features of the polder OMIT selection. These types of maps apply a "Polder Mask", generated by a modified "Bulk-Solvent-Mask" which prevents the bulk-solvent mask from penetrating the region in question. Polder OMIT maps can be particularly useful for displaying weak densities of ligands, solvent molecules, side chains, alternative conformations and residues both in terminal regions and in loops. Polder maps have been generated by phenix.polder in the PHENIX software package (Adams, *et al.* 2010).

· *Feature Enhanced Maps (FEM):* these types of maps are generated by a procedure that improves a weighted $2F_o-F_c$ map (Afonine, et al., 2015) and are calculated for the whole protein. The result is a new map that possesses a reduced level of noise and model bias and that also shows enhancement of weak features, often bringing them onto the same scale as the strong features, to make them more easily interpretable for both human and model-building software. Unlike density modification, the FEM is not an iterative phase-improvement procedure and it is generally not expected to reveal features that are not already present in the original map. FEM maps have been generated by phenix.fem in the PHENIX software package (Adams, *et al.* 2010).

3.4.7. Structure Validation and Deposition

Even though the previously described R_{work} and R_{free} values are acceptable, the structure requires of further validation prior to deposition in the Protein Data Bank (PDB). Parameters that should be checked are mainly related to the molecule's geometry, such as bond angle and distances, Ramachandran outliers, side chain rotamer outliers and clashes, among others. These parameters are checked by tools present in Coot during manual construction and modification of the model, by other programs such as MolProbity (Chen et al., 2010) or by the PDB Validation service executed prior to coordinate deposition. All of these have been used for structural validation of the structures here solved.

Furthermore, macromolecular structures which have not been obtained using the latest experimental and computational methods sometimes need double-checking: platforms, such as PDB-REDO (Joosten, *et al.*, 2009) re-refines already existing structure models in the PDB automatically for further improvement and validation. In the case of the previously reported eEF1A2 structure (PDB code 4C0S), although not re-refined by PDB-REDO, we have manually reprocessed the already deposited coordinates again to improve the model for molecular replacement of our diffraction data.

3.4.8. Other Software Used

· *DALI:* The Dali program is widely used for carrying out automatic comparisons of protein structures determined by X-ray crystallography or NMR. Used for the identification of other "structurally-homologous" proteins that may share particular folding, but sequence similarity is low. (Holm, *et al*., 2006).

· *BLAST:* used for structural alignment of protein sequences. (Boratyn, *et al*., 2013).

· *PYMOL:* used for graphical representation and visualization of structures. (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.).

· *ADXV:* used for visualizing diffraction images.

· *PhosphositePlus:* online server to identify potential post-translational modification sites according to sequence (Hornbeck, 2015).

· *PISA server:* online server to identify interactions between given molecules (Krissinel and Henrick, 2007).

· *RCD2+:* RCD+ (v2) server is a fast loop-closure modeling tool based on an improved version of the RCD method (Chys and Chacón, 2013).

· e*LBOW:* automated generation of geometry restraint information for refinement of novel ligands and improved geometry restraint information for standard ligands.

· *PDB-REDO:* a procedure to optimize crystallographic structure models, providing algorithms that make a fully automated decision making system for refinement, rebuilding and validation, used for CBDP35:TA structure validation (section 3.4.7.).

4. RESULTS AND DISCUSSION

4.1. DEVELOPING NEW TryR INHIBITORS

4.1.1. Structural Determination of *L. infantum* TryR and Try Complexes

4.1.1.1. TryR Crystallization

As mentioned in section 3.2.3. of Experimental Procedures, initial crystallization trials were carried out using TryR from *L. infantum* M/CAN/ES/89/IPZ229/1/89 strain. *L. infantum* TryR structure had already been reported along with its crystallization conditions (Baiocco, *et al.*, 2009; PDB 2JK6), comprising 2.2 M ammonium sulfate, 0.1 M Tris pH 7.5 using a hanging drop technique. Following the reported methodology, crystallization assays were set up around this condition, yielding extremely fragile, rod-like yellow crystals which in occasions grew in a twinned manner (Figure 4.1.). It is worth mentioning that the yellow color is due to the presence of the FAD cofactor bound to TryR. Due to crystal-bundle growth, the initial condition was optimized by adding additives, detergents and low molecular weight PEGs at low concentrations. Crystals were cryoprotected with the same crystallization solution supplemented with 25% glycerol and subjected to X-ray diffraction experiments. These crystals diffracted X-rays at circa 8 Å resolution.

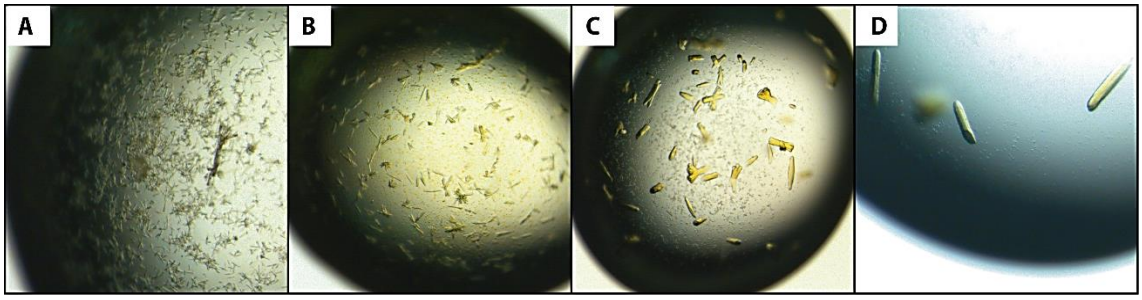


Figure 4.1.: *L. infantum* strain CAN/ES/89/IPZ229/1/89 TryR crystals. (A-C) Crystals obtained in 2-2.4 M ammonium sulfate, 0.1 M Tris pH 7-8.5. Those shown in **(C)** had a supplementation of 5% PEG 400. **(D)** Optimized crystallization was shown for 0.1 M Tris pH 8.5, 2 M ammonium sulfate, 5% PEG 400, using the hanging-drop vapour diffusion method. Here, individual rod-shaped crystals were grown.

Several approaches were carried out in order to enhance the X-ray diffraction properties of these crystals (section 3.2.) but no considerable improvement was achieved. Sequence analysis revealed a total of 18 mutations with respect to the previously crystallized *L. infantum* TryR (Baiocco, *et al.*, 2009) (Figure S2 in Supporting Information). When these mutations were pin-pointed in the three dimensional structure of the protein (PDB 2JK6), it was observed that these are generally distributed on its surface. Non-conservative mutations (12 out of the 18) contribute in the modification of the surface electrostatic potential of the protein, and so affecting its ability to tightly pack into regular, well-diffracting crystals.

A new construction of *L. infantum* M/CAN/ES/96/BCN150/MON-1 strain TryR was cloned and purified at the System Biology Department at Alcalá de Henares University to crystallization standards. This protein was crystallized as mentioned in section 3.2.3., mimicking

reported conditions (Baiocco, *et al.*, 2009). Bi-pyramidal crystals (Figure 4.2.) grew within two-three days in a broad range of pHs and ammonium sulfate concentrations.

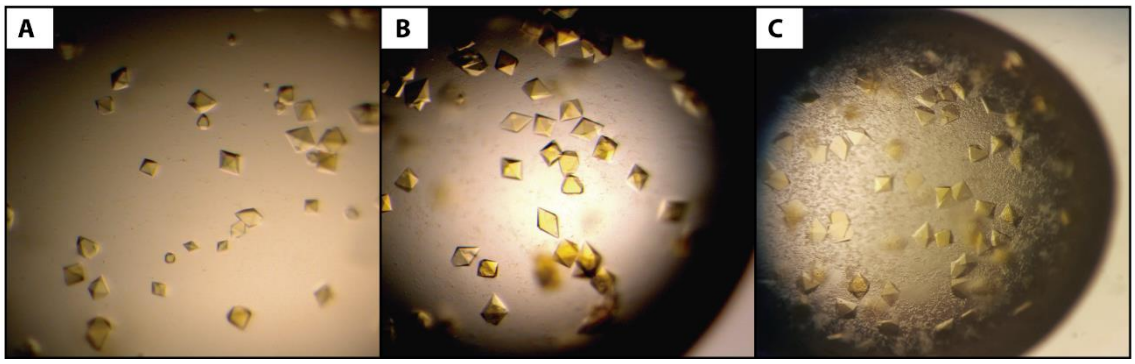


Figure 4.2.: *L. infantum* strain M/CAN/ES/96/BCN150/MON-1 TryR crystals. Yellow bi-pyramidal crystals were yielded 3 days after crystallization assays. Figure shows increasing ammonium sulfate concentrations (1.9M –A-; 2.1 M –B-; 2.3M –C-) in 0.1 M Tris- HCl pH 8. Protein concentration: 8 mg/mL.

These crystals were used in soaking experiments (section 3.2.3.) to obtain the complex with its natural substrate and NADPH cofactor, following reported conditions (Baiocco, *et al.*, 2013) and with those inhibitors synthetized at the Medical-Chemistry Institute (CSIC, Madrid).

4.1.1.2. Structural Determination of TryR and TryR:NADPH:trypanothione Complex

Crystals were taken to beamlines BL13 at the ALBA Synchrotron (Cerdanyola del Vallès, Spain). Diffraction patterns collected were indexed, integrated, scaled and merged using XDS and Aimless. The best crystals diffracted up to 1.9 Å, but resolution diminished to 2.4 Å in the case of crystals that had been soaked in trypanothione and NADPH. Both apo and complexed crystals belong to the P 4_1 2_1 2 space group (unit cell parameters shown in Table 4.1.) Diffraction patterns revealed anisotropic diffraction, and crystals were highly sensitive to radiation. High resolution intensities were therefore excluded (final resolutions of 2.4 Å and 2.6 Å for apo and complexed crystals, respectively). Matthews coefficient is ~2.6 A^3/Dalton, with one molecule per asymmetric unit and ~53 % of solvent content. Because the same protein structure had been solved already, (PDB 2JK6) it was used in MOLREP as a search model for molecular replacement.

Molecular replacement solution was refined using Phenix.refine (section 3.4.6.). Final refinement statistics and model geometry show values within accepted criteria (Table 4.1.). RMS values for bond length and angles and Ramachandran outliers indicate an optimum geometry for the model obtained. Along with water molecules, sulfate ions and glycerol molecules present in the cryoprotectant solution were also identified.

In comparison to PDB 2JK6, our crystals show a lower amount of solvent, hence the higher resolution (2.4 Å vs 2.95 Å for TryR; 2.6 Å vs 3.6 Å for TryR:NADPH:trypanothione complex). Moreover, our crystals present a higher grade of symmetry, and so the asymmetric unit

comprises only one molecule (TryR monomer), as opposed to PDB 2JK6, which belongs to $P4_1$ space group and has two molecules per asymmetric unit (TryR dimer). Nevertheless, when symmetry operators are applied to our TryR monomer coordinates, the dimer is formed along with a symmetry-related molecule (Figure 4.3.A.).

Table 4.1.: Crystallographic data collection and refinement statistics for apo-TryR and TryR:NADPH:trypanothione complex. Value for the highest resolution shell is shown in parenthesis.

Parameters	**TryR**	**TryR:NADPH:trypanothione**
Data collection		
Space group	P 4_1 2_1 2	P 4_1 2_1 2
Cell dimensions		
a, *b*, *c* (Å)	103.82 103.82 192.21	103.128 103.128 193.342
α, β, γ (º)	90, 90, 90	90, 90, 90
Wavelength (Å)	0.979260	0.979260
Resolution (Å)	48.27-2.40 (2.49)	48.34 -2.6 (2.69)
Total reflections	83762 (8224)	65128 (6355)
No. unique reflections	41881 (4113)	32867 (3208)
R_{pim}	0.032 (0.376)	0.019 (0.296)
$CC_{1/2}$	0.998 (0.689)	1.0 (0.937)
$I/\sigma(I)$	14.42 (1.56)	11.79 (1.62)
Completeness (%)	100 (100)	99.84 (99.63)
Multiplicity	18.6 (2.0)	6.4 (6.0)
Refinement		
Resolution range (Å)	48.27-2.40	48.34 -2.6
R_{work}/R_{free}	0.179/0.221	0.189/0.242
No. atoms		
Protein	3713	3713
Water	317	81
Ligand	113	189
R.m.s. deviations		
Bond length (Å)	0.013	0.009
Bond angles (º)	1.100	1.090
Ramachandran favored/outliers (%)	99.59/0.41	99.38/0.62
Residues in AU	491	491
Average B value overall (Å^2)	60.58	69.23
PDB code	**----**	**----**

4.1.1.3. Structural Characterization of TryR and TryR:NADPH:trypanothione Complex

Although *L. infantum* TryR and TryR:NADPH:trypanothione complex structures were known, structural analysis of our complexes was carried out.

4.1.1.3.1. Structure Overview

Our structure reveals one FAD molecule bound to one monomer of TryR, or one FAD, NADPH and trypanothione molecule bound to the TryR monomer if the TryR crystal had been previously soaked with the cofactor and natural substrate. Overall, our TryR structure shows small differences with PDB 2JK6 (rmsd = 0.191Å for chain A and 0.192 Å for chain B). Hence, the

overall fold resembles that of TryR of other species, revealing the three different domains: FAD binding domain (residues 1-160 and 289-360) which adopts a Rossman fold typical of GR family members, the NADPH binding domain (residues 161-289) and the interface domain (361-488).

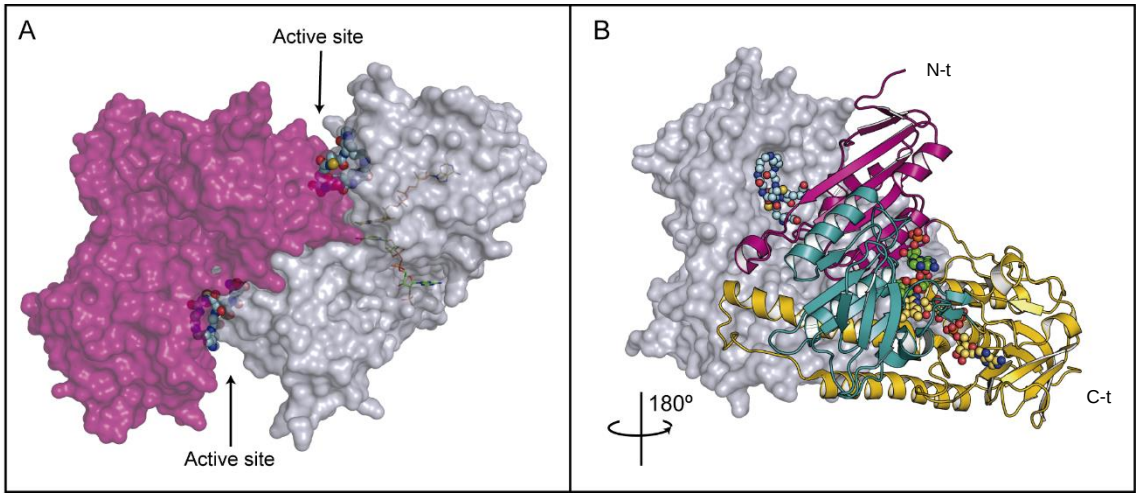


Figure 4.3.: Overview of TryR structure. (A) Surface representation of TryR. If symmetry operations are applied to the monomer in the asymmetric unit (purple), the dimer is generated (grey), related by a two-fold axis. Two catalytic clefts are formed, in which trypanothione is present when TryR crystals are exposed to its natural substrate (sphere representation). NADPH and FAD are represented as green and yellow sticks respectively. **(B)** General domain fold of TryR. The monomer in the asymmetric unit is shown as cartoon. FAD binding domain is colored yellow, NADPH binding domain in blue and interface domain in purple. Trypanothione is represented as blue spheres and NADPH and FAD in green and yellow spheres, respectively. Structure determined during this Thesis.

4.1.1.3.2. Crystal Packing

TryR has crystallized in space group P 4_1 2_1 2. As mentioned previously, a single TryR monomer is observed in the asymmetric unit. Herein, a resultant crystallographic binary axis diagonal to axis *a* and *b* of the cell, generates the true TryR dimer forming the crystal, which naturally occurs with each monomer related by a twofold axis. Large solvent channels are observed (Figure 4.4.), allowing smaller molecules to diffuse through the crystal during soaking experiments. TryR's active sites are exposed to these solvent channels.

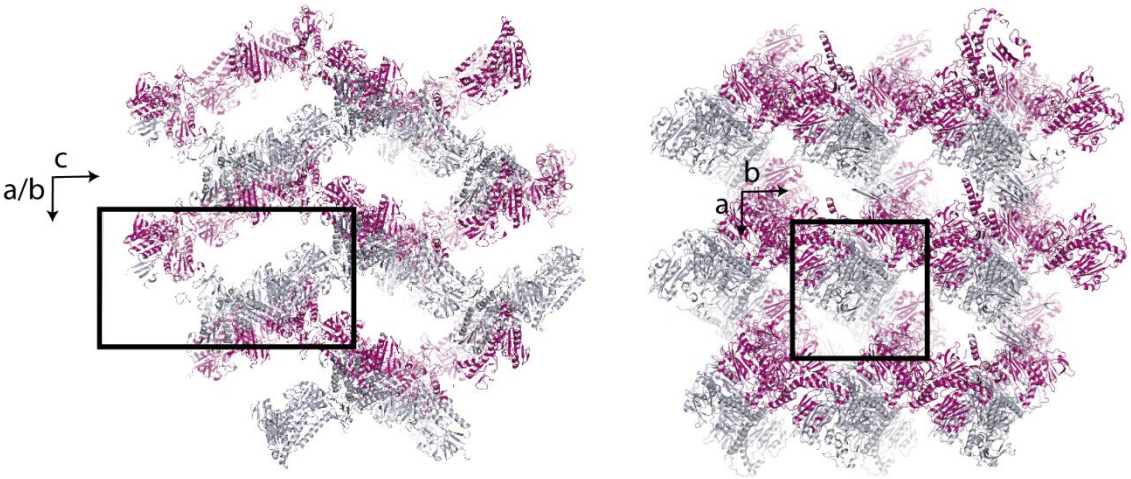


Figure 4.4: Symmetry and packing TryR crystals: The protein chains are represented as cartoon. The monomer is colored in purple. For clarity, the other chain forming the dimer is colored grey.

4.1.1.3.3. Active Site and Trypanothione Binding

Our TryR:NADPH:trypanothione structure shows a different conformation of trypanothione to the previously published structure complex (PDB 4ADW). TryR's active site is a 15 Å wide and 20 Å deep cleft, formed by residues from both the FAD binding site and the interface domain. As previously described (Lantwin, *et al.*, 1994; section 1.2.4.), the catalytic residues comprise two cysteines, namely Cys52 and Cys57, and His461' from the other monomer. Although it is expected TryR to be oxidized, we can observe in our TryR apo-structure that the disulfide bridge has been reduced. Nevertheless, in the presence of trypanothione, a double conformation of the catalytic cysteines is observed (Figure 4.5.), both oxidized and reduced. Furthermore, the distance between cysteines is larger when the crystals were soaked with trypanothione (3.4 Å vs. 2.7 Å in the apo structure). It is worth mentioning that disulfide bridge formation seemed independent whether NADPH was present or not in the soaking solution.

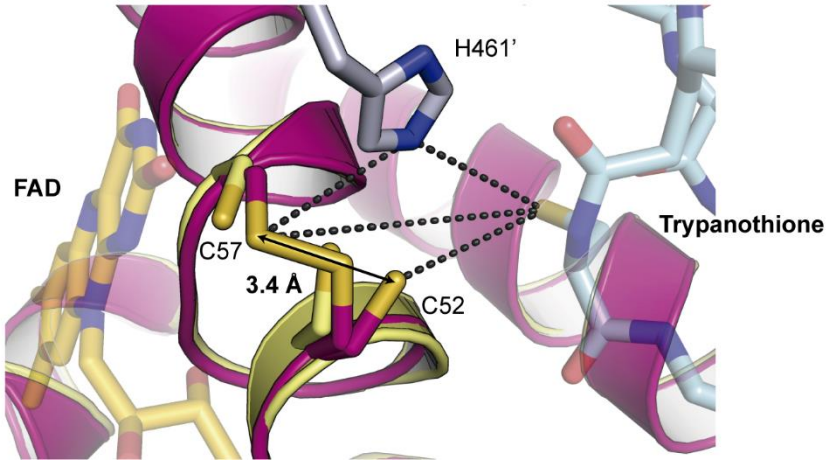


Figure 4.5.: Catalytic residues in TryR's binding site. TryR is shown as cartoon. TryR:NADP:trypanothione complex is colored purple and regular crystallized TryR is colored in light yellow. FAD and trypanothione are shown as yellow and blue sticks. Catalytic residues are shown as sticks and labelled accordingly. His461' from the other monomer is colored grey. Interactions are represented by dashes.

Due to the large cavity that comprises the active site of TryR, trypanothione positions differently within it. Multiple interactions are made with residues belonging to both of the TryR monomers. Herein, in the previous trypanothione-bound *L. infantum* TryR structure (PDB 4ADW), trypanothione displays in a linear conformation, spread across the active site. Oppositely, in our structure, trypanothione is curled deep in the cavity close to the catalytic cysteines. *T. brucei* TryR with NADPH and trypanothione shows a similar network of interactions as in our structure. These interactions are represented in Figure 4.6. and summarized in Table 4.2. It is worth mentioning that, other than the residues shown in Figure 4.5., key residues which are consistent in the different trypanothione conformations comprise Trp21 and Tyr110 from one monomer, and Glu466', Glu467' and Thr463' from the other. Furthermore, the spermidine bridge between both glutathionyl tails of trypanothione is involved in the interaction with Trp21, due to

charge-charge interactions with –NH groups of the spermidine moiety and those π electrons of Trp21's rings.

B factors of trypanothione suggest high mobility, as they are twice as high as the average protein B factor. Moreover, it is worth mentioning that residual electron densities present within the active site suggested possible alternative conformations (mobility) of trypanothione.

Table 4.2.: Interactions observed in our trypanothione bound TryR complex.

Residue	Atom	TS_2	Distance(Å)
Cys52	Sγ	Sγ2	4
His461'	Nε2	Sγ2	3.5
Glu18	Oε1	N1S	3.2
Tyr110	OH	O3	2.8
Thr463'	O	Oδ7	4
Glu466'	Oε1/Oε2	N1/O27	3/2.8
Glu467'	Oε2	O17	3.1
Ser470'	Oγ	O21	3

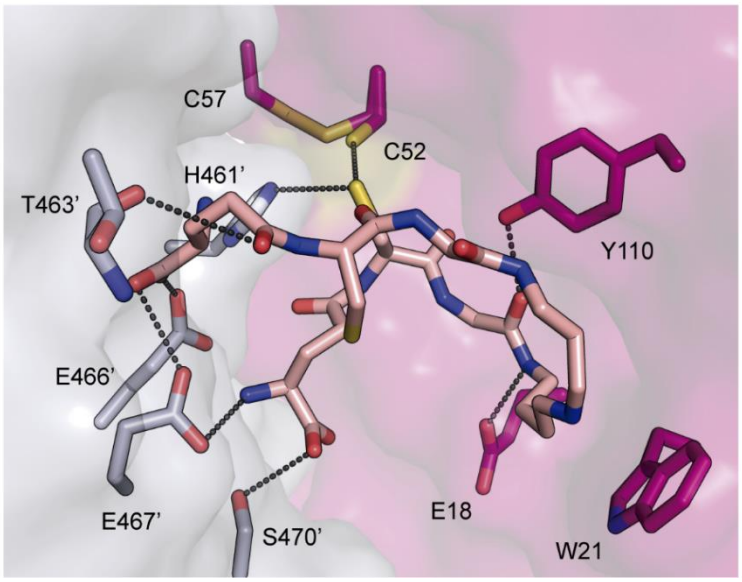


Figure 4.6.: Residues involved in trypanothione binding. TryR's surface is shown. Different colors represent different monomers. Key residues involved in trypanothione binding (represented as pink sticks) are shown in stick form, labelled and colored according to their monomer. Non-interacting main chain atoms are not shown for clarity purposes.

4.1.1.3.4. FAD and NADPH Binding Sites

The FAD binding domain is formed by three-stranded antiparallel β-sheets (residues 126-153), a five stranded parallel β-sheets (comprising residues 7-10, 31-35, 120-124 and 155-158), and four α-helices (14-27, 24-92, 69-161 and 335-351). Strong conservation of FAD-binding residues is observed in all species, although subtle differences may occur in the residues interacting with FAD through their main chains. The FAD binding site and residues involved in its stabilization are portrayed in Figure 4.7. In general, N of Gly127 is hydrogen bonded to N1 of

the adenine ring. This bond is much longer in our NADPH complex structure than in the apo-structure (4 vs 2.95 Å), suggesting a displacement of this loop upon NADPH binding. Although it has been reported that the NA7 of the adenine ring interacts with Arg290, we observe Arg290 interacting with O2B of the ribose sugar. The ribose ring strongly interacts with Asp35 (distance OD2-O3B =2.7 Å). The pyrophosphate group is hydrogen bonded to the N and O atoms of Thr51 (distances OA1-N and OA2-OG1, 2.7 and 3.1 Å respectively) and to the main chains of Ser14 (N-OP1 = 3 Å) and Asp327 (N-O2P = 3 Å) by means of weak hydrogen bonds. Furthermore, the ribosyl chain also hydrogen bonds to Asp327 (distances ~3Å) through its side chain. Finally, the flavin ring interacts with the protein moiety through Lys60's amine side chain group (NZ-O4 = 2.6Å).

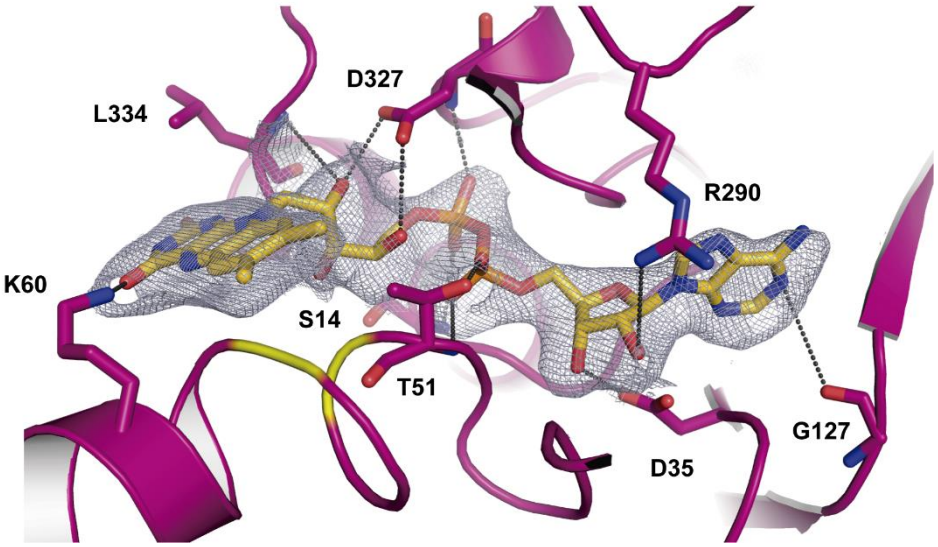


Figure 4.7.: FAD binding site. TryR is shown as cartoon and colored purple Cys52 and Cys57 side chains are not shown but their Cα are colored yellow. Residues involved in FAD binding (shown in yellow sticks) are shown in stick form and labelled. Main chains not involved in FAD binding are hidden for clarity purposes. Dashes represent interactions between atoms. 2Fo-Fc map is shown and colored grey, contoured at 0.9σ.

Regarding NADPH binding, the adenine moiety does not interact directly with the protein (Figure 4.8.A.). Nevertheless, the 2′-phosphate group is bound to Arg222, Arg228, and Tyr221 through its oxygen atoms by means of hydrogen bonds and electrostatic interactions (Arg222: O2 χ –Nε = 3.3 Å, O3χ-NH2 = 3.5 Å; Tyr221: O2χ-OH = 3.5 Å and Arg228: O1χ-NH1 = 2.9Å). The pyrophosphate interacts with the main chain of Tyr198 (N-O1α = 3.5 Å), which acts as a gate by changing conformation to allow the NADPH entrance (Figure 4.8.B.), and Gly286 (O2η-N = 2.9 Å). The ribose sugar is only stabilized by means of hydrogen bonds with the side chain of Met333 (O2δ-O = 2.8 Å). Finally, the nicotinamide moiety interacts with Glu202 (N7-Oε2 = 3Å) and the side chain of Ala365 (N7η-O = 3 Å). These interactions are all illustrated in Figure 4.8.A.

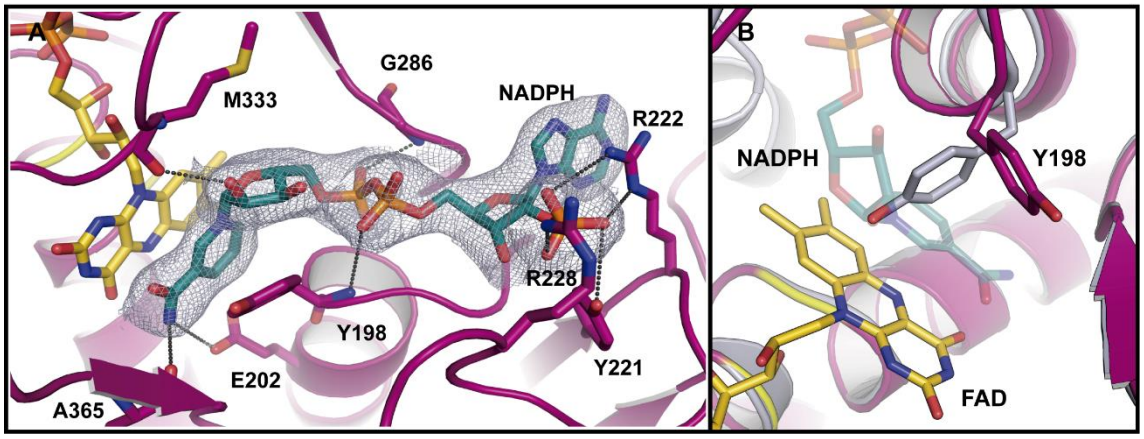


Figure 4.8.: NADPH binding site. (A) TryR is shown as cartoon and colored purple Cys52 and Cys57 side chains are not shown but their Cα are colored yellow for reference. Residues involved in NADPH binding (shown in green sticks) are shown in stick form and labelled. Main chains not involved in the binding are hidden for clarity purposes. Dashes represent interactions between atoms. 2Fo-Fc map is shown and colored grey, contoured at 1σ. **(B)** Tyr198 detail. When NADPH binds to TryR, Try198 acts as a gate, changing conformation in order to allow NADPH entrance.

4.1.1.3.5. Dimerization Interface

PISA server was used to analyze the monomer-monomer interactions between TryR dimer. TryR monomers interact through a 3187.7 Å, solvent accessible, non-polar surface, composed by 85 residues. Nevertheless, those residues involved in hydrogen bonds are Lys61, Gly66, Tyr69, Glu75, Phe79, Gly80, Trp81, Glu82, Asn91, Tyr210, Ser433, Ile437, Gln439, Ser440, Cys444, Ile458, Val460, Pro462, Ser464 and Ala465. The atoms involved in hydrogen bond formation and corresponding distances are listed in Table S2 in Supporting Information. These interactions are also conserved in human GR, except for Phe79, Gly80, Trp81, Asn91, Tyr210 and Ser440.

A central cavity is embedded within the dimerization interface of both GR and TryR. Despite the amino acid conservation between these proteins, the hydrophobic cavities of GR and TryR shows only 35% sequence identity between them. Regarding TryR, this cavity displays a 1755 $Å^2$ surface and a 2280 $Å^3$ volume, and is conserved between TryRs of different *Tripanosomatidae* species. This cavity has been explored in the case of GR. Crystal structures of GR in complex with xanthene-derived compounds (i.e. PDB 3SQP, 1XAN) show that the residues involved in the binding of these compounds (Trp70, Asn71, Val74, His75, Phe78, Met79, His82 and Tyr407 in GR) are not conserved in TryR.

This cavity connects both active sites where trypanothione binds (Figure 4.9.). In particular, in the case of *L. infantum* TryR, hydrophobic residues lining the central cavity include Val64, Leu72, Phe367, Pro371 and Pro435. The previously identified dimerization hot-spot (Glu436) is also exposed to this cavity. Furthermore, Glu466 and Glu467 connect the active site to the central cavity (Figure 4.9.). These amino acids lie only ~9 Å away from the catalytic cysteines.

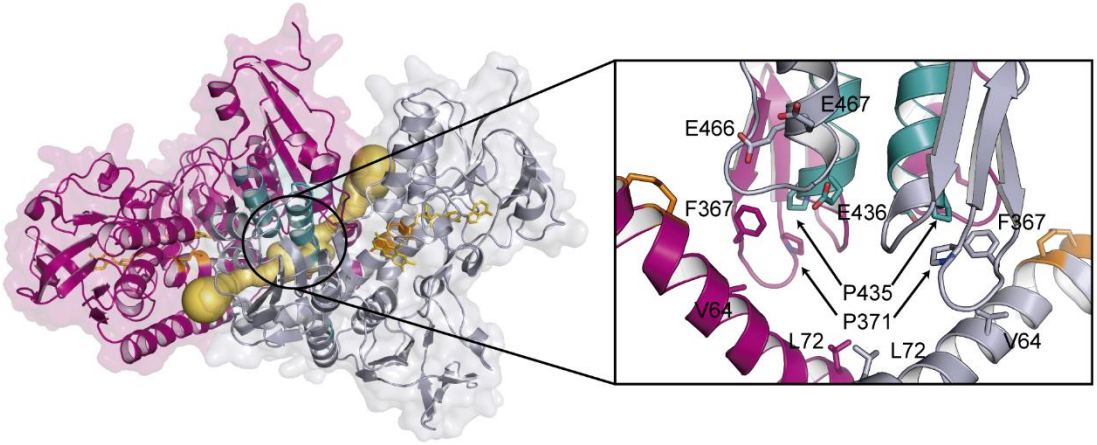


Figure 4.9.: Central cavity in *L. infantum* TryR. TryR is represented as cartoon, each monomer colored purple and grey. FAD molecules are shown as yellow sticks. Catalytic cysteine residues are colored orange. The α-helix in the dimerization interface, crucial for TryR oligomerization, is colored green. The tunnel connecting both active sites, calculated by CAVER (pymol plugin) is shown as yellow spheres. The panel on the right shows this cavity in detail. Hydrophobic residues lining the cavity are shown as sticks and labelled.

4.1.1.4. Trypanothione Reductase Ligands (TRLs) and Background Knowledge

As mentioned (section 1.2.6.), TryR dimerization hot-spot identification led to the design of short peptides mimicking the α-helix involved in the interaction with Glu436 (Figure 4.10.A) and with probed dimer inhibition capacity (TRL35) (Toro, *et al.*, 2013). TRL35 behaves as a non-competitive inhibitor (K_i = 0.5 μM) and is capable of irreversibly inactivate TryR by causing TryR precipitation. Nevertheless, this peptidemimetic suffered from protease degradation and was unable to cross the membrane barrier to enter parasitic cells (Toro, 2017).

In order to optimize TRL35, Grossmann's peptidemimetic inhibitor classification (Pelay-Gimeno, *et al.*, 2015) was considered. This classification involves four different groups based on the inhibitor's resemblance to the initial peptide (Table 4.2.). Although multiple peptidemimetics from classes A-C have been developed by our colleagues at Medical-Chemistry Institute (CSIC), and their activity assayed at the System Biology Department at Alcalá de Henares University, crystallization experiments were only performed for class C compounds.

The specific contribution of each residue and length to the inhibitory capacity of TRL35 (Figure 1.16.) was evaluated by our colleagues at the System Biology Department at Alcalá de Henares University, leading to a shorter and active version, namely TRL38 (Figure 4.10.A). Along with docking analysis by Prof. Federico Gago, a pharmacophore was proposed (Figure 4.10.C), used for the design of small molecule structural mimetics. These studies shed light on key residues of TRL38: Lys2, Gln5 and Ile9 (Figure 4.10.C). Hence, Lys2 was proposed to interact with Asp423' and Glu436' of TryR, Gln5 would hydrogen bond the main chain atoms of Ile458' and Val460' and Ile9 of the inhibitor peptide would make hydrophobic interactions with Phe454' of TryR (Figure 4.10.C).

Table 4.3.: Classification of peptidemimetics by Grossmann and colleagues. Adapted from (Pelay-Gimeno, *et al.*, 2015).

Peptides		Natural peptide sequences derived from proteins and ribosomal peptides
Peptidic character	Class A: Modified peptides	Peptides mainly formed by α-amino acids with minor side chain or backbone alterations
	Class B: Modified peptides/foldamers	Peptides with various backbone and side chain alterations also including foldamers
Small molecules	Class C: Structural mimetics	Small molecule-like scaffolds that project substituents in analogy to peptide side chains
	Class D: Mechanistic mimetics	Molecules that mimic the mode of action of a peptide without a direct link to the side chains

Herein, class C peptidemimetics were designed based on pyrrolopyrimidine, 5-6-5-imidazole-phenyl-thiazole and 5-6-5-triazole-phenyl-thiazole scaffolds, developed by Lee, Cummings and our colleagues at the Medical-Chemistry Institute (CSIC), respectively, who patented this new molecular architecture of peptidemimetics (Cummings, *et al.*, 2009; Lee, *et al.*, 2011). These scaffolds allow the projection of substituents that specifically mimic lateral chains at positions i, i+3 (or i+4 in the case of pyrrolopyrimidines) and i+7 of the designed peptide. The substituents would face the same side of the model α-helix. Thus, substituents in position R1 of the scaffold mimic the interaction of Lys2; R2 that of Gln5; and R3 mimics Ile9 in the TRL38 peptide (Figure 4.11). Regarding 5-6-5-triazole-phenyl-thiazole scaffolds, these can be classified in types I and II whether a fourth substituent is introduced (Figure 4.11.).

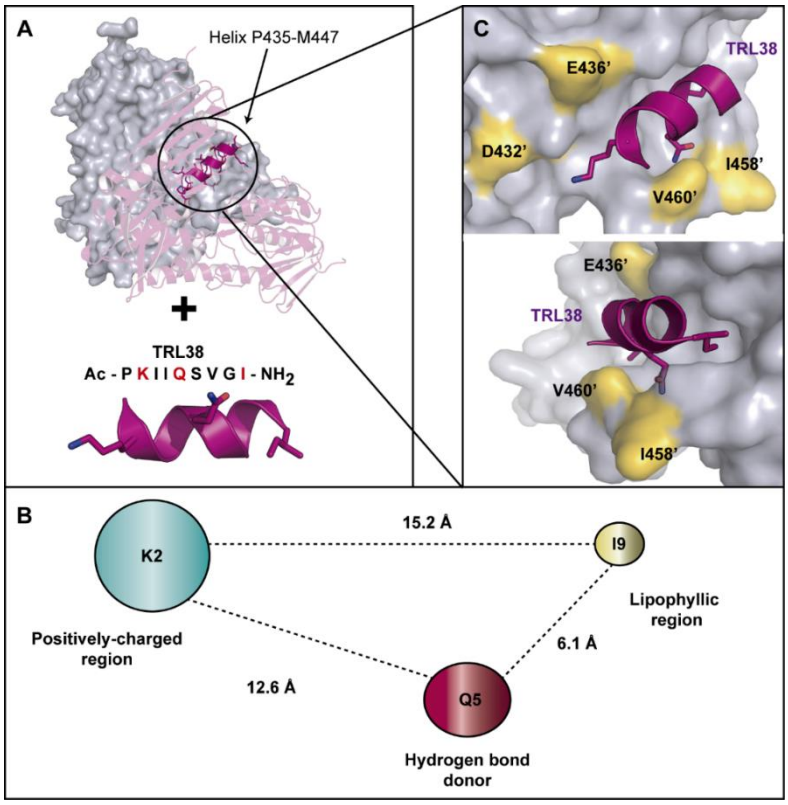


Figure 4.10.: Background knowledge of TRLs (A) Based on the original protein sequence of helix P435-M447, and inhibitor optimization, TRL38 is designed. **(B)** Pharmacophore designed on TRL38 properties. **(C)** Docking of TRL38.

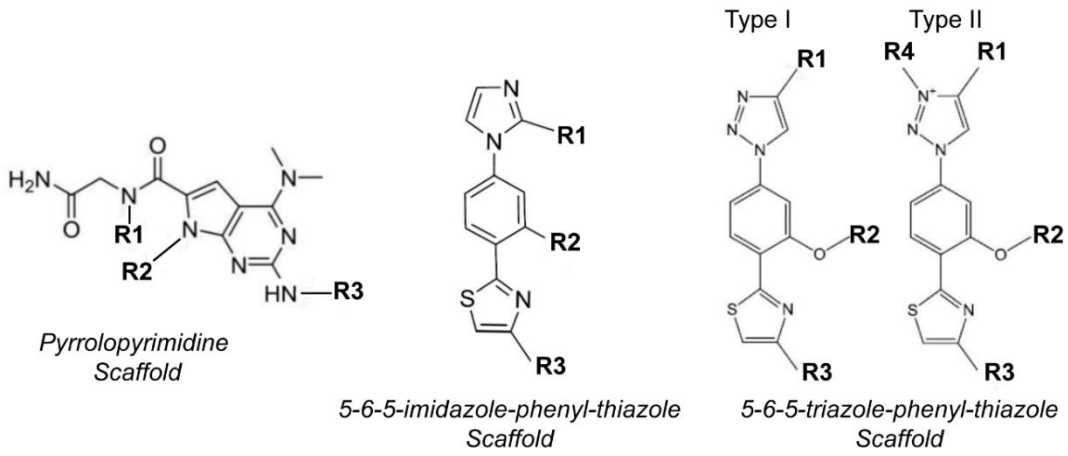


Figure 4.11.: Chemical scaffolds used as peptidemimetics of the P435-M447/TRL38 helix.

Herein, although many TryR inhibitors were essayed, only the atomic complex structures with TRL149, TRL156, TRL187 and TRL190 were obtained. TRL156 has a pyrrolopyrimidine scaffold, and biological assays show no effect in dimerization inhibition and high IC_{50} values (IC_{50}=52.2 ± 1.8). TRL190, and TRL149 and TRL187 are type I and type II 5-6-5-triazole-phenyl-thiazole scaffolds, respectively. These show an improved TryR oxidoreductase inhibition compared to pyrrolopyrimidine scaffolds, being IC_{50}=3.1± 0.2 for TRL149, IC_{50}=3.6 ± 0.3 for TRL187 and IC_{50}=5.3 ± 0.9 for TRL190.

4.1.1.5. TryR:TRL Complex Crystallization and Diffraction Experiments

Section 1.2.6. briefly described a novel TryR inhibition strategy based on rationally designed peptidemimetics which target the dimerization interface of TryR. These can be categorized according to their chemical scaffold or to their activity on TryR. Hence, regarding crystallization, it is worth differentiating between (i) dimer disruptors (which inactivate TryR) and (ii) oxidoreductase inhibitors (which initially do not affect the equilibrium between TryR monomeric and dimeric forms). The general observations here presented will be discussed below.

4.1.1.5.1. Co-Crystallization Experiments

Co-crystallization was attempted by incubating TryR with selected TRLs at different concentrations and time periods (initially, x10 ligand to protein molar ratio and decreased to x0.5 ligand:protein), either prior to the crystallization set-up or "in-drop" co-crystallization (addition of ligand as an additive to the drop). Observations can be differentiated according to ligand activity:

· *Dimer disruptors:* incubation of TryR with key peptidemimetics caused a sudden and quasi-total precipitation of the enzyme. Herein, our colleagues at the System Biology Department had proved that, initially, a small percentage of the total protein remained soluble but in a dimeric

conformation, and, after a 16h, room-temperature incubation, TryR precipitated totally due to the instability of TryR in the monomeric form. This observation proved X-ray crystallography not to be a useful tool for the characterization of the binding of these peptidemimetics. Co-crystallization in the presence of detergents such as β-octyl glucoside, which could help mask exposed hydrophobic patches when in the monomeric form, did not avoid enzyme precipitation.

· *Oxidoreductase inhibitors:* two scenarios were observed:

a) Crystal formation: crystals formed in the presence of oxidoreductase inhibitors with the same morphology as in Figure 4.2., either in the same crystallization condition as with the apo-protein or in new conditions identified in commercial crystal screenings. This occurred when TryR was exposed to the ligands directly in the crystallization drop or when the crystallization assays were set up after short incubation periods. When exposed to X-rays, these crystals either (i) generated a poor protein diffraction pattern characteristic of crystal damage; (ii) were unable to diffract X-rays (iii) no electron density was observed for the ligand when a poor dataset was collected.
b) Protein precipitation/no crystals: this was observed when TryR was incubated with oxidoreductase inhibiting TRLs for longer periods of time (overnight incubations) prior to crystallization or in crystallization drops in which crystals did not form.

4.1.1.5.2. Soaking Experiments

Due to unsuccessful co-crystallization experiments, crystal-soaking techniques were attempted using previously obtained TryR crystals in its apo-form. Soaking outcomes for oxidoreductase inhibitors and dimer disruptors should be again differentiated. It is worth commenting on the wide variety of time and TRL concentrations analyzed for each TRL.

· *Dimer disruptors:* soaking experiments led to crystal cracking and disintegration after TRL exposure beyond 30 minutes and 1 mM TRL concentration. If a crystal was plunged into liquid nitrogen prior to its disintegration and subjected to X-ray diffraction experiments, a poor diffraction pattern was produced. Furthermore, these were extremely sensitive to radiation damage impeding the collection of quality-sufficient diffraction data or a complete diffraction pattern given the crystal symmetry. Soaking conditions involving lower time and concentration values led to X-ray diffracting crystals which had not bound the TRL (i.e. no electron density corresponding to the TRL was observed after processing diffraction data).

· *Oxidoreductase inhibitors:* soaking apo-TryR crystals using relatively high TRL concentration (~ 5 mM) for short time periods (< 30 min) or low concentrations (0.5-1 mM) for longer periods (days) resulted in non-diffracting or damaged crystals. Optimization for each individual ligand

led to the soaking conditions mentioned in section 3.2.3., in which acceptable X-ray diffraction data and TRL binding came to an agreement.

4.1.1.6. Structural Determination of TryR:TRL Complexes

Crystals were taken to beamlines BL13 at the ALBA Synchrotron (Cerdanyola del Vallès, Spain). Diffraction patterns collected were indexed, integrated, scaled and merged using XDS, iMosflm and Aimless. Resolution significantly diminished to ~3 Å when exposed the ligands. It is worth mentioning that ligands were dissolved in 100% DMSO and 100 mM stocks were prepared in order to avoid high DMSO concentrations in soaking solutions. Unit cell and refinement parameters of all complexes are shown in Table 4.4. Diffraction images collected revealed an even more pronounced anisotropic diffraction and sensitivity to radiation than native crystals. Intensities from high resolution shells were therefore excluded (final resolutions ranging from 2.5-3.3 Å). This implied low multiplicity values, as images collected beyond 90-100º rotation angle showed extreme radiation damage (in general terms, diffraction ended after 150º crystal rotation). Because this protein structure was already known, (PDB 2JK6) it was used as a search model for molecular replacement.

Molecular replacement solution was refined using Refmac (section 3.4.6.), employing tight geometric and non-crystallographic symmetry restraints, jelly body refinement, and map sharpening, which gave improved statistical results. Model building was carried out using the program COOT. Final refinement statistics and model geometry show values within accepted criteria. RMS values for bond length and angles and Ramachandran outliers indicate an optimum geometry for the model obtained, all. Along with water molecules, sulfate ions, glycerol and DMSO molecules present in the soaking and cryoprotectant solution were identified

Just like native crystals, most of these complexes belonged to the $P4_12_12$ space group. Nevertheless, TryR:TRL156 crystals belong to space group $P2_12_12_1$. As mentioned in section 4.1.1.4., pyrrolopyrimidine scaffolds show decreased efficacy in TryR inhibition, with IC_{50} values ten times larger than 5-6-5-triazole-phenyl-thiazole scaffolds. Hence, despite the increased concentration of TRL156 in the soaking solution, only one molecule of TRL156 was shown bound to one of the two active sites of dimeric TryR. Therefore, because of the dissimilarity between active sites upon TRL156 binding, the space group symmetry is decreased and two molecules (TryR dimer) are depicted in the asymmetric unit. However, crystal packing is maintained.

Table 4.4.: Crystallographic data collection and refinement statistics for TryR in complex with TRL156, TRL149, TRL187, TRL190 and TRL187+trypanothione. Value for the highest resolution shell is shown in parenthesis.

Parameters	TryR:TRL156	TryR:TRL149	TryR:TRL187	TryR:TRL190	TryR:TRL187 :trypanothione
Data collection					
Space group	P 2_1 2_1 2_1	P 4_1 2_1 2	P 4_1 2_1 2	P 4_1 2_1 2	P 4_1 2_1 2
Cell dimensions					
a, *b*, *c* (Å)	94.18, 106.39, 186.55	103.5, 103.5, 192.22	103.43, 103.43, 192.79	103.11,103.11 191.68	103.5, 103.5, 192.39
α, β, γ (º)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	0.979260	0.979260	0.979260	0.979260	0.979260
Resolution (Å)	46.64-3.30 (3.56)	73.19-2.5 (2.59)	68.38-3.0 (3.11)	48.05-2.8 (2.9)	73.19-2.65 (2.75)
Total reflections	78272 (11212)	72557 (6991)	41206 (4122)	52357 (5104)	62129 (6062)
No. unique	41881 (4113)	36702 (3606)	21384 (2111)	26209 (2549)	31166 (3049)
R_{pim}	0.07 (0.39)	0.034 (0.30)	0.067 (0.294)	0.025 (0.263)	0.026 (0.219)
$CC_{1/2}$	0.99 (0.79)	0.998 (0.915)	0.998 (0.863)	0.999 (0.939)	0.999 (0.948)
$I/\sigma(I)$	5.2 (1.3)	10.04 (2.02)	5.90 (1.62)	11.48 (2.02)	11.28 (2.25)
Completeness(%)	81.7 (74)	99.23 (99.20)	98.42 (99.48)	99.89 (99.80)	99.89 (99.93)
Multiplicity	3.3 (2.6)	2.7 (1.9)	1.9 (2.0)	3.6 (2.1)	2.0 (2.0)
Refinement					
Resolution range (Å)	46.64-3.30	73.19-2.5	68.38-3.0	48.05-2.8	73.19-2.65
R_{work}/R_{free}	0.243/0.271	0.232/0.242	0.247/0.282	0.238/0.273	0.212/0.255
No. atoms					
Protein	7403	3712	3710	3704	3707
Water	195	271	150	246	218
Ligand	170	161	138	133	258
R.m.s. deviations					
Bond length (Å)	0.0075	0.014	0.013	0.028	0.014
Bond angles (º)	1.61	1.90	1.89	2.09	1.88
Ramachandran favored/outliers (%)	99.5/0.5	99.59/0.41	99.38/0.62	99.18/0.82	98.77/1.23
Residues in AU	977	489	489	489	489
Average B value overall (Å^2)	61.54	33.86	52.38	54.73	66.74
PDB code	**6I7N**	**----**	**----**	**----**	**----**

4.1.1.7. Structural Characterization of TryR:TRL Complexes

4.1.1.7.1. Pyrrolopyrimidine Scaffold (TRL156)

The final model contains a TryR functional dimer consisting of two subunits related by a non-crystallographic 2-fold axis. Interestingly, only chain B presents one TRL156 molecule bound to its active site (Figure 4.3.). This inhibitor is placed close to the catalytic residues (Cys52 and Cys57) and to the FAD cofactor (Figure 4.12).

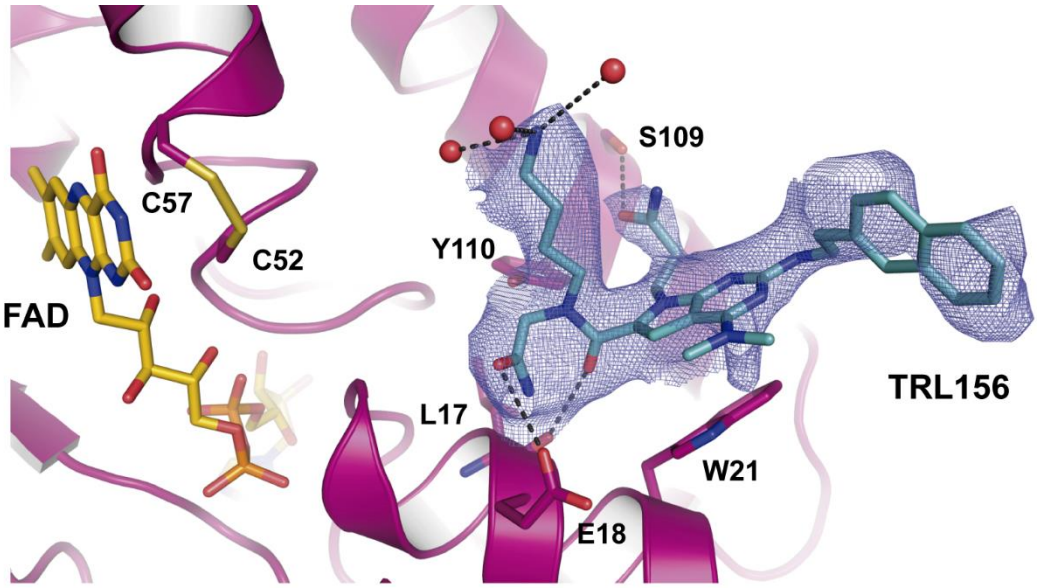


Figure 4.12.: Close up view of TRL156 binding site. The ligand prevents trypanothione from fitting into the cavity. Polar interactions are represented by dashed lines. Relevant amino acids are depicted as capped sticks and labelled. Water molecules are represented as red spheres. TRL156 is represented as blue capped sticks and its corresponding electron density (2Fo-Fc map) is contoured in blue at 0.9σ.

The central pyrrolopyrimidine core makes a stacking interaction with Trp21 (Figure 4.12.) at the so-called polyamine-binding site, where the spermidine cross-bridge of trypanothione usually binds. Substituents R1 and R2 also play an important role in the stabilization of TRL156 in TryR's active site by establishing different polar interactions with the protein. Thus, the Glu18 side chain and backbone atoms of Leu17 residues establish hydrogen bond interactions with the amide moiety and keto groups in R1, respectively. Another residue involved is Ser109, which interacts with the amide group of the glutamine mimetic in R2. Moreover, the lysine mimetic substituent forms ion-dipole interactions with water molecules positioned in the active site of TryR. The naphthalene ring in R3 protrudes out of the cavity, and its electron density is poorly defined, indicating high mobility in this region.

4.1.1.7.2. 5-6-5-Triazole-Phenyl-Thiazole Scaffolds (TRL149/187/190)

These models contain one TryR molecule, and so the functional dimer is formed by symmetry-related molecules. Hence, both active sites of TryR are occupied by the correspondent TRL. Just like in TRL156, these inhibitors are placed close to the catalytic residues (Cys52 and Cys57) and to the FAD cofactor, and bind to the same polyamine-binding site that trypanothione binds to (Figure 4.13).

All of these 5-6-5-triazole-phenyl-thiazole-derived inhibitors are positioned similarly in the active site of TryR, independent on whether they belong to type I or type II peptidemimetics (section 4.1.1.4.). R1 and R3 substituents are the same for these TRLs. The amino-butyl substituent in R1 faces towards the outside of the active site. On the other hand, R3 faces a

hydrophobic region inside of the cavity where Ile106 lies, also formed by Val53, Val102 and Val103. Both of these substituents show a weaker defined electron density (Fo-Fc, 1σ), suggesting mobility of this region. R2 presents a imidazolidinone (TRL149) or a piperidine (TRL187 and TRL190) substituent depending on the ligand. In the case of TRL149, the imidazolidinone hydrogen bonds Glu18 and Trp21 (Figure 4.13.A.). In the case of TRL187 and TRL190, the –NH group in the piperidine ring makes a single hydrogen bond with Glu18 (Figure 4.13.B. and C.). However, the ring stacks with Trp21 (3.9 Å distance). The thiazole ring in the scaffold of these TRLs also stacks with the side chain of Tyr110 at an angle, forming a so-called herringbone interaction, and hydrogen bonds Ser109 (Figure 4.13.).

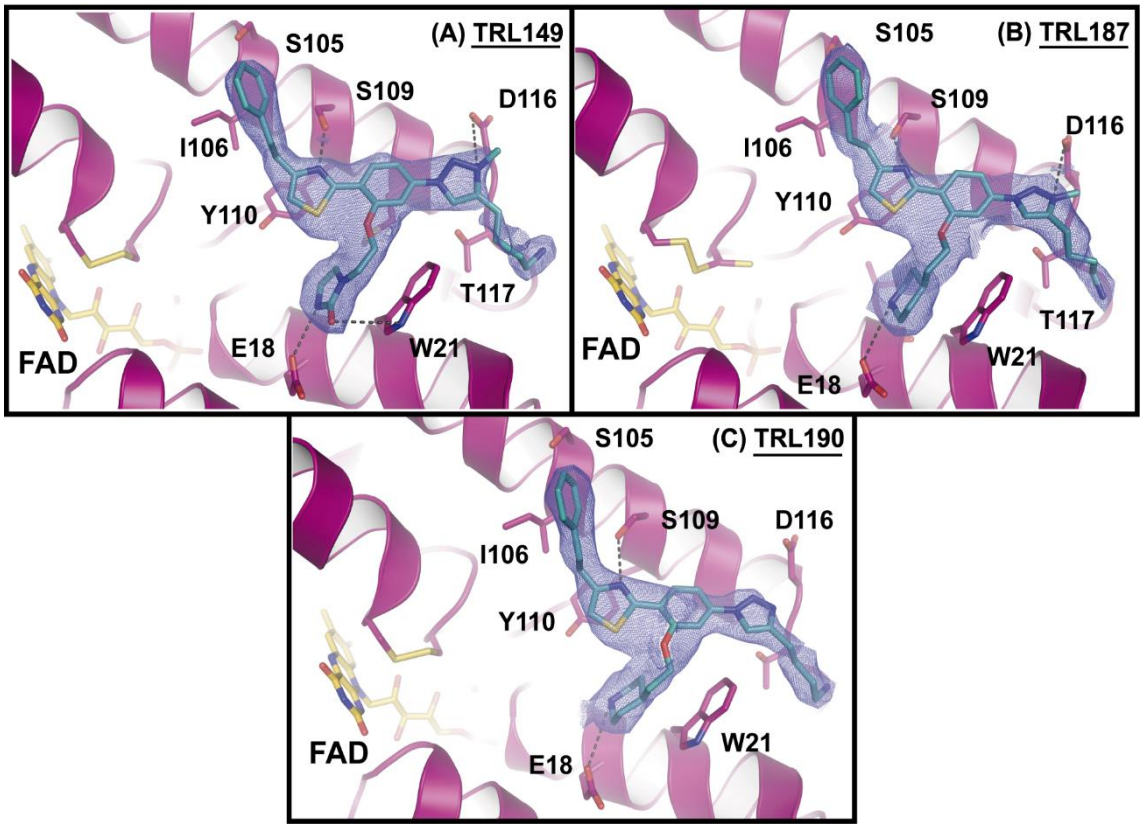


Figure 4.13.: Close up view of TRL149, TRL187 and TRL190 binding site. Interactions are represented by dashed lines. Relevant amino acids are depicted as capped sticks and labelled. TRLs are represented as blue capped sticks and its corresponding electron density (2Fo-Fc map) is contoured in blue at 1σ for TRL149 and TRL187 and 0.8σ for TRL190.

The triazole ring also varies its charge depending on the TRL. Thus, TRL149 and TRL187 have a formal extra positive charge (type II 5-6-5-thiazole-phenyl-triazole scaffold) which is involved in an electrostatic interaction with Asp116 (distance ~3.8 Å). Moreover, it is worth mentioning that the electron density for the triazole ring is less defined for TRL190, as well as the R1 substituent it holds, suggesting mobility of the triazole ring (i.e. rotation around the bond with which it binds to the phenyl ring) and allowing further interactions of R1 with nearby polar residues (Thr117).

4.1.1.7.3. TRL187 and Trypanothione

Soaking experiments of TRL187 in the presence of oxidized trypanothione were carried out in order to check the effects of TRL187 on trypanothione binding. Soaking at equal concentrations of TRL187 and trypanothione were executed for different time periods (15 minutes – 2 hours). Most datasets showed the same binding of TRL187 as in Figure 4.13.B. However, the crystal obtained after 1 hour of soaking revealed a different scenario.

Just as mepacrine (PDB 1GXF), TRL187 is capable of staking within the active site of TryR by means of its 5-6-5-thiazole-phenyl-triazole scaffold (Figure 4.14.A). Hence, two nearly perpendicularly-oriented molecules of TRL187 were observed at 4.2 Å distance from each other. TRL187_1 (Figure 4.14.B) maintains the same orientation and interactions as previously seen in section 4.1.1.7.2. TRL187_2 stacks above TRL187_1, without making any interactions with TryR.

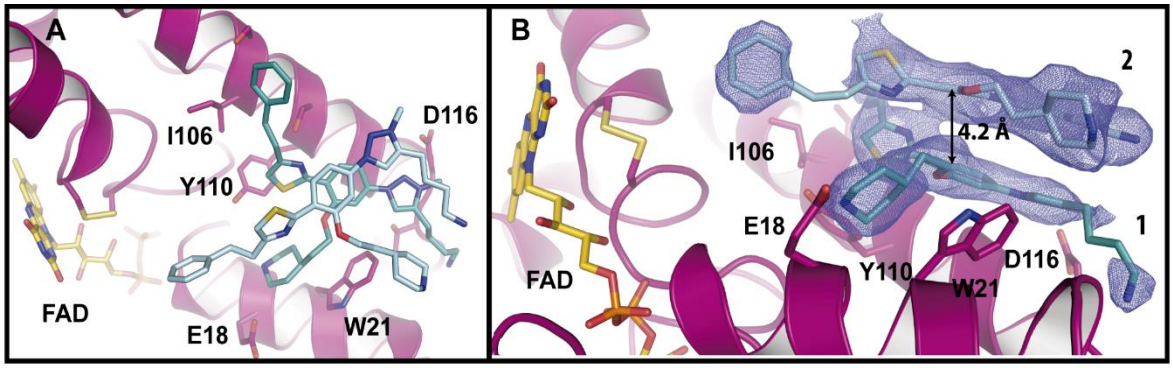


Figure 4.14.: Close up view of TRL187 stacking in the TryR active site. Interactions are represented by dashed lines. Relevant amino acids are depicted as capped sticks and labelled. TRLs are represented as blue capped sticks and its corresponding electron density (2Fo-Fc map) is contoured in blue at 1σ for TRL149 and TRL187 and 0.8σ for TRL190.

Continuous electron density was observed for the first glutathionyl- molecule of trypanothione (Figure 5.15.). This section interacts by means of hydrogen bonds with the thiol group of Cys52 (Sγ2-Sγ = 3 Å) and His463' and the carboxyl group of the main chain of Thr463' (N7-O = 3.5 Å) (Figure 4.15.). However, the spermidine bridge is displaced from its binding site above Trp21 of one TryR monomer towards a polar region comprising Glu410' and Glu467' of the other monomer. These residues interact by means of weak hydrogen bonds with the second glutathionyl- molecule of trypanothione and the spermidine cross-link, respectively (Glu410: Sγ2-Oε1 = 3.5 Å; Glu467: N6σ-Oε1 = 3.8 Å). Nevertheless, poor electron density was observed beyond the first glutathionyl- molecule likely due to scarce interactions needed for stabilization and therefore implying mobility of the solvent-oriented glutathionyl- moiety of trypanothione.

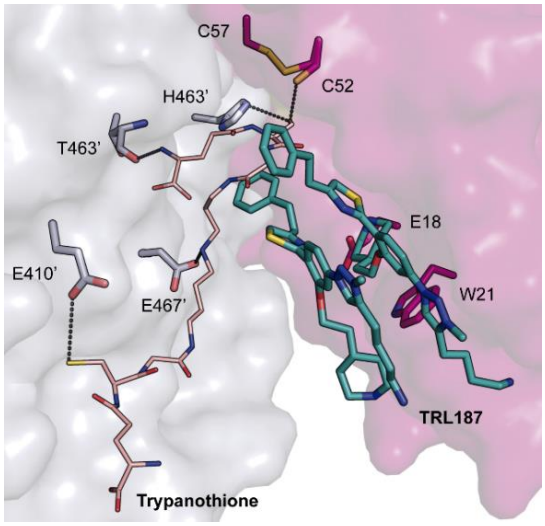


Figure 4.15.: Displacement of trypanothione in the active site upon TRL187 stacking. Interactions are represented by dashed lines. Relevant amino acids are depicted as capped sticks and labelled. TRL187 are represented as blue capped sticks and trypanothione as pink lines. TRL187 binds in the polyamine binding site (Glu18, Trp21), displacing trypanothione from its natural conformation (Figure 4.6.).

4.1.1.8. Consequences Upon Ligand Binding

The ligand−protein interaction does not seem to dramatically alter the structure of the TryR dimer. All TryR structures are highly similar (average RMSD between all TryR chains circa 0.2 Å). Changes observed between the backbone of both chains (Figure 4.16.A.) are concentrated in different loops and at the N-terminal tail. Ligand binding at the active site provokes side chain reorientation in those residues directly involved in their stabilization (Figure 4.16.B.). These side chain accommodations vary according to the ligand scaffold bound. For example, Tyr110 side chain reorients when the pyrrolopyrimidine scaffold (TRL156) binds to the active site or TryR, approaching the nitrogen atom R1. Furthermore, Trp21 shows a different rotamer conformation whenever an inhibitor binds to that of apo-TryR, favoring the stacking interaction with R2 substituent (Figure 4.16.B.).

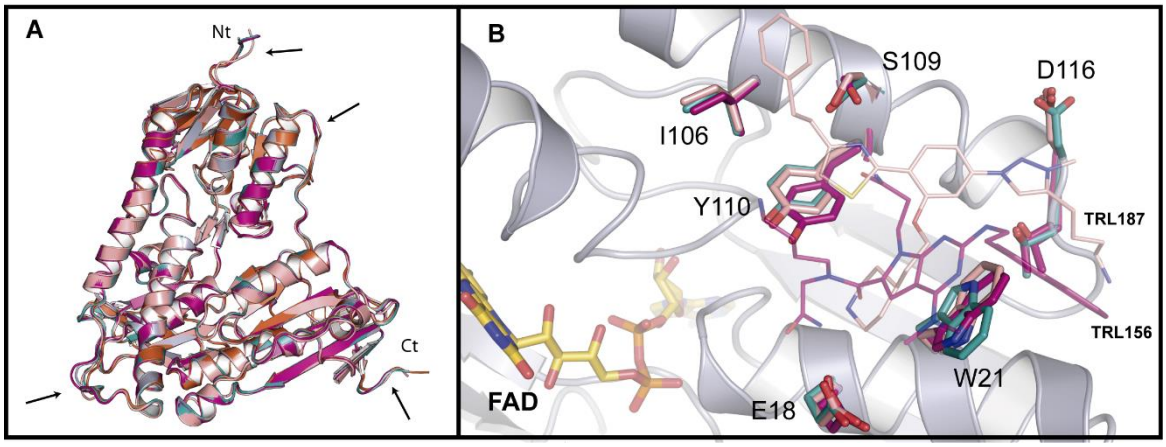


Figure 4.16.: Conformational changes upon ligand binding. (A) TryR structure is conserved in all structures, only minor changes in loops are observed, indicated by arrows. **(B)** Reorientations of side changes upon ligand binding. Only the structures of apo-TryR and in complex with TRL156 and TRL187 are shown. Side chains of relevant residues are shown as capped sticks and colored magenta for TryR:TRL156, pink for TryR:TRL187 and green for apo-TryR. Tyr110 reorients towards the inhibitor in the case of TRL156 binding and Trp21 switches conformer upon ligand interaction.

4.1.2. Structure-Inhibition Correlation for TRLs and Future Work

4.1.2.2. Structure-Activity Relation of TRLs

The fact that all crystallized TRLs occupy the active site strongly suggests that they all prevent the binding/exit of Try. However, the TryR:TRL187:trypanothione ternary complex shows that trypanothione is still capable of entering the active site of TryR in the presence of TRL187. The results obtained in the TryR:TRL187:trypanothione complex could be therefore explained as an intermediate state between functional and inhibited TryR. Hence, TRL187 could bind to the polyamine binding site, displacing a previously reacting trypanothione molecule from its regularly adopted conformation (i.e. spermidine bridge interacting with Trp21, Figure 4.6.) and inhibiting further substrate molecules entering the active site. Furthermore, the similar chemical structure of TRL149, 187 and 190 suggest that this ligand stacking phenomena could apply for all cases.

The TryR structures obtained in complex with different inhibitors only involve TRLs which have been proven inactive in displacing the equilibrium from a dimeric to a monomeric state of TryR but have shown oxidoreductase capacity in the micromolar range (Lucio, 2018). Ligand structure and activity was analyzed for the battery of compounds by our colleagues at the System Biology Department at Alcalá de Henares University. In general terms, it was concluded that pyrrolopyrimidine scaffolds were poor dimer disruptors and oxidoreductase inhibitors, hence, a larger TRL156 concentration was required in soaking experiments in order to obtain the TRL:TRL156 complex. Regarding the 5-6-5-thiazole-phenyl-triazole scaffold, it has been shown that the substitution of the imidazolidinone R2 (TRL149) substituent for a piperidine (TRL187 and TRL190) completely eliminates the dimerization inhibiting capacity of TRLs. However, type II 5-6-5-thiazole-phenyl-triazole inhibitors (TRL149 and TRL187) containing a positive charge in the triazole ring, also hampers TryR dimerization. This explains why TRL149, which incorporates an imidazolidinone as its R2 substituent, does not show dimer-inhibition capacity.

In structural terms, the positive charge in the triazole ring fixes the type II 5-6-5-thiazole-phenyl-triazole scaffold by interacting with Asp116 by means of electrostatic interactions (~3.8 Å), retaining the inhibitor in the polyamine binding site and preventing the TRL to diffuse into the interface cavity (section 4.1.1.3.5.). Another interpretation involves the electron density for the triazole ring and the R1 substituent of TRL190 holds (no formal positive charge). In comparison to the electron density of type II 5-6-5-thiazole-phenyl-triazole scaffolds, TRL190 shows a weaker electron density for this region (less continuity and definition), suggesting mobility along the phenyl-triazole bond due to the absence of the electrostatic interaction with Asp116 required for stabilization. In the case of piperidine-containing substituents (TRL187 and TRL190), although a hydrogen bond is lost in comparison to the imidazolidinone ring, the piperidine ring stacks Trp21, holding the TRL in the position observed (Figure 4.13.).

It has been shown that bulkier, longer and more hydrophobic substituents for R3 result in a stronger oxidoreductase- and dimerization-inhibiting capacity of TRLs (Lucio, 2018). As observed in our results, -CH_2-phenyl group of R3 is oriented towards a hydrophobic environment where Ile106 lies, right at the entrance of the target interface cavity. Based on molecular biology results, it has been suggested that bulkier and longer R3 substituents promote stronger hydrophobic interactions with Val53, Val102, Val103 and Ile106, allowing the displacement of the TRL from the polyamine binding site towards the interface cavity and so promoting their dimerization-inhibiting activity.

All in all, these inhibitors have shown a competitive inhibition behavior towards *L. infantum* TryR (Lucio, 2018), coherent to the binding observed in our crystal structures. Moreover, except for TRL156, IC_{50} values of TRL149, TRL187 and TRL190 are an order of magnitude below mepacrine (IC_{50} = 12 μM, Section 1.2.5.1.), showing efficient TryR inhibition by novel chemical scaffolds prone to development as anti-trypanosome drugs. However, as mentioned in section 1.2.4.1., more than 90% TryR inhibition is required to prevent parasite growth (Krieger, *et al.*, 2000). This implies that competitive inhibitors with K_i values in the micromolar range are insufficient for trypanosomiasis and leishmaniosis treatment, hence the development and need of mixed-type and irreversible inhibitors (i.e. dimer disrupting TRLs) or competitive inhibitors with K_i values in the nanomolar range.

4.1.2.2. Other Structurally Characterized TryR Inhibitors

Eleven crystal structures of TryR (either from *T. brucei*, *T. cruzi* or *L. infantum)* in complex with other polyamine site-binding inhibitors have been published (4APN, 2WP5, 2WP6, 2WPC, 2WPE, 2WPF, 4NEV, 6BTL, 6BU7, 4NEW, 1GXF, Table S1 Supporting Information). Other TryR inhibitors whose mechanism of action have been structurally revealed (Table S1 Supporting Information) involve the fixation of the catalytic cysteine residues by means of heavy atoms. When the crystal structure of TryR in complex with TRLs 149, 156, 187 and 190 is best-fit superimposed onto these structures, the only common interaction observed to all inhibitors involves the side chain of Trp21 in *L. infantum* TryR.

Phenyl-quinazoline derivatives (PDB: 2WP5, 2WP6, 2WPC, 2WPE and 2WPF, Figure 4.17.A.) engage Trp21 and Tyr110. The phenyl group creates T-stacking interactions with Trp21 and Tyr110 is shown to change conformer upon ligand binding. Furthermore, Glu18 is also involved by means of hydrogen bonds with the quinazoline moiety.

Cyclohexylpyrrolidine derivatives (PDB: 4NEV, 6BTL and 6BU7, Figure 4.17.B. and C.) also bind in the same region and involve Asp116 for its stabilization, which pivots towards the pyrrolidine moiety. However, the same inhibitor as in PDB 4NEV (*T. brucei* TryR) was

crystallized in complex with *T. cruzi* TryR (PDB 4NEW, Figure 4.17.B.). In this case, the cyclohexylpyrrolidine moiety turns away from Asp116 but maintains the interaction with Trp21.

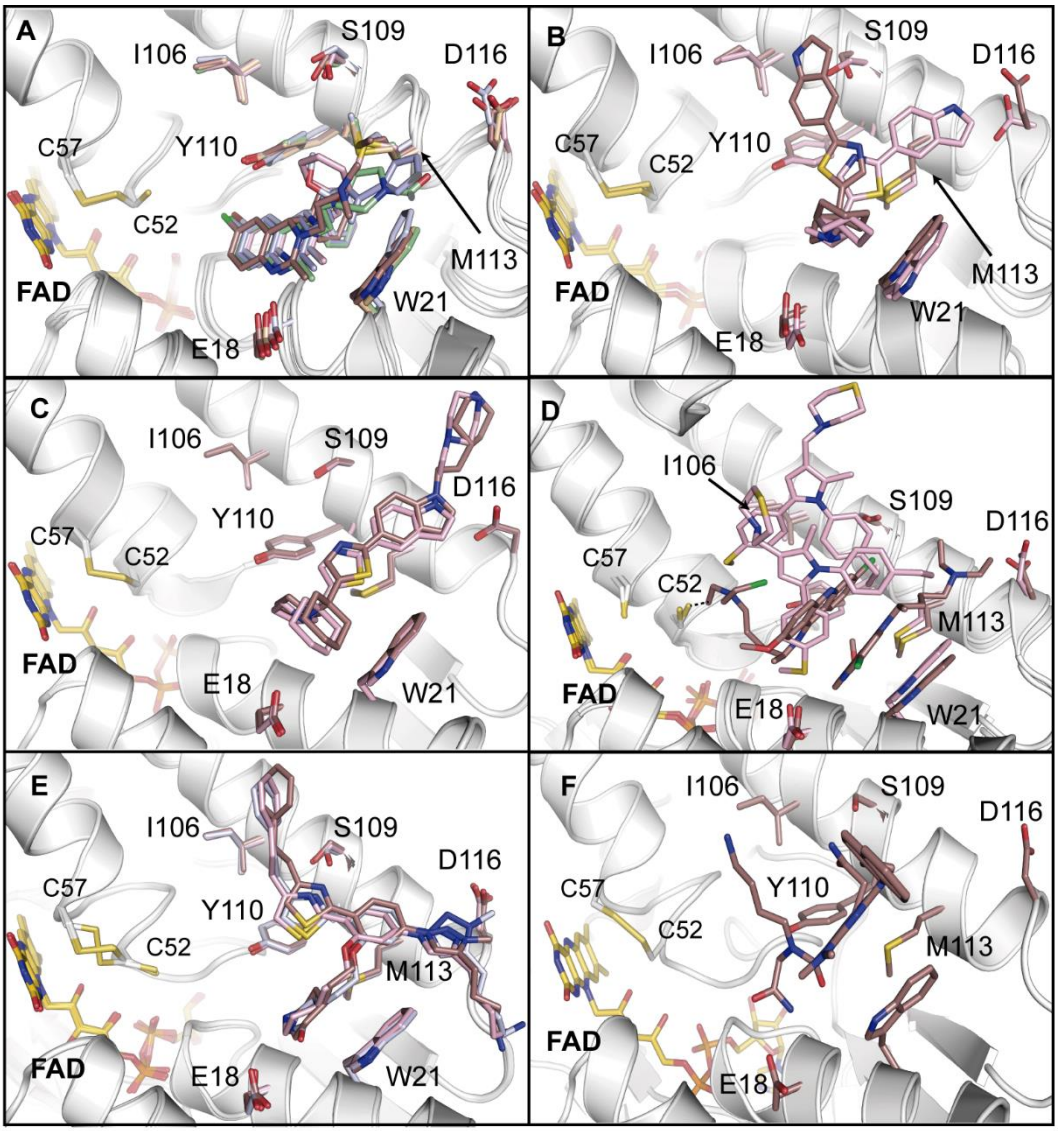


Figure 4.17.: Polyamine binding site is common to different TryR inhibitors. Figures show different complex structures of TryR with inhibitors bound to its polyamine binding site (either *L. infantum*, *T.brucei* or *T.cruzi*). TryR is shown in white cartoon. FAD and catalytic cysteines are shown as sticks and shown for reference. Residues with which the ligands here determined interact are shown as sticks, labelled and colored accordingly. *L. infantum* TryR numbering is considered. In all, the interaction with Trp21 is maintained despite the chemical structure of the inhibitor. **(A)** Phenyl-quinazoline derivatives (PDB: 2WP5-blue, 2WP6-pink, 2WPC-green, 2WPE-yellow, 2WPF-brown). **(B)** Cyclohexylpyrrolidine derivatives. PDB 4NEV-pink, and 4NEW-brown. **(C)** Cyclohexylpyrrolidine derivatives. PDB 6BTL-pink, 6BU7-brown. **(D)** Thiomorpholine derivative (PDB 4APN) and quinacrine mustard (mepacrine) (PDB 1GXF). Both stack two molecules by means of Trp21 to the active site. **(E)** TRL149, TRL187 and TRL190, revealed in this research. **(F)** TRL156, determined in this research.

The other two TryR structures in complex with a thiomorpholine derivative (PDB 4APN, Figure 4.17.D.) and quinacrine mustard (mepacrine) (PDB 1GXF, Figure 4.17.D.) involve inhibitors which stack above Trp21. Hence, two molecules are observed inside the active site of TryR hampering trypanothione binding. Mepacrine spans the whole active site, one molecule

stacking above Trp21, on a perpendicular plane to the 5-6-5-thiazole-phenyl-triazole scaffold, and the second molecule hydrogen bonding Cys52. Ser109, Ser112 and Asp116 are also involved. However, the thiomorpholine derivative binds with a different orientation by forming a herringbone interaction with Trp21 and stacking a second molecule in close proximity to the Ile106-Ser109 helix. Again this inhibitor's positioning is also based on Trp21 stacking and hydrophobic interactions.

In conclusion, oxidoreductase inhibitor binding to the so-called polyamine or mepacrine binding site do not seem to cause significant conformational changes on the TryR structure besides side chain reorientations. Nevertheless, these inhibitors have been shown to interact with conserved residues in TryR of different species, mainly being Glu18, Trp21, Tyr110 and Asp116. Other than Asp116, these residues are involved in trypanothione positioning within the active site by binding its spermidine crosslink. Hence, interrupting the interactions of Glu18, Trp21 and Try110 with trypanothione seems to be crucial for TryR inhibition. Moreover, analogue residues and the extra hydrophobic patch to which these inhibitors bind are absent on GR (Section 1.2.4.1.). This difference is therefore exploited in the design of specific competitive inhibitors in order to avoid cross-inhibition and toxicity issues.

4.1.2.3. TRLs and Crystallization

The TRLs here crystallized were not effective in disrupting the dimerization of *L. infantum* TryR (Lucio, 2018). As explained in Section 4.1.1.5., co-crystallization experiments were thwarted regarding TRLs that are capable of efficiently displacing the dimer to monomer equilibrium (IC_{50} < 10 μM) due to enzyme precipitation. In this sense, it is noteworthy that the same precipitation phenomenon was also observed by our colleagues at the System Biology Department at Alcalá de Henares University (Toro, 2017).

Low concentrations of dimerization disruptors also lead to enzyme precipitation in a time-dependent manner. Herein, although TryR precipitated partially soon after TRL addition, nearly total precipitation of TryR was observed after 1-2-hour exposure to the ligand or during on-going crystallization.

Regarding soaking experiments, dimer-inhibiting TRLs lead to crystal cracking and disintegration, therefore limiting soaking time periods, or resulting in non-diffracting crystals. As observed for the apo-TryR, crystals are formed by the dimeric protein that crystallizes in a $P4_1\ 2_1\ 2$ space group (Figure 4.4.). Therefore, the effects observed upon crystallization with dimer-disrupting TRLs suggest that, among other possible effects, inhibitors binding to the interface cavity of TryR may distort the two-fold related dimer and alter the crystal packing, leading to fragile crystals and poor or no diffracting crystals. Moreover, TryR crystals soaked in a solution

containing weak dimer disruptors (TRLs with IC_{50} beyond 20 μM) did not result in diffraction patterns that revealed the TRL. This could imply that insufficient TRL molecules had regularly bound to TryR dimers in the crystal, thus preserving the crystal packing and allowing diffraction, and so were not observed in the average calculated electron density.

Crystal solvent content was ~70% for all crystals, regardless of its space group. As shown in Figure 4.4., large solvent channels are present in TryR crystals, allowing TRLs to diffuse throughout. However, both oxidoreductase and dimer inhibitors assayed are highly insoluble in water. For that reason, DMSO was used as a solvent. The link between DMSO concentration, exposure time and crystal damage was studied, showing a positive correlation. Concentrated TRL stocks in 100% DMSO (100 mM stocks) allowed low percentages of the organic solvent in the final soaking solution (1-5%). Hence, diffusion of the TRLs through the crystal channels could be hampered due to ligand insolubility in aqueous solutions and high ionic concentration. These considerations allow the understanding of our negative outcomes and difficulties in obtaining TryR in complex with TRLs.

4.1.2.4. Time-Dependent TryR Inactivation

Time dependent TryR inactivation by dimer disruption has been observed for many TRLs analyzed, despite their initial oxidoreductase activity, as a consequence of long incubation periods of *L. infantum* TryR and TRL compounds (Lucio, 2018). In particular, incubation of TryR with stoichiometric concentrations of TRL149 for over 16 h eventually caused enzyme precipitation, likely due to the disruption of the dimeric conformation of TryR. Thus, we hypothesized that the crystal complex structures here determined could represent an initial phase (or transient state) prior to the occupancy of the hydrophobic cavity at the dimerization interface, that we consider the putative target site of the TRLs (section 4.1.1.3.5.). This line of reasoning is supported by results from ongoing steered molecular dynamics simulations carried out in Prof. Gago's lab at Alcalá de Henares University. Due to the difficulty in crystallizing dimer disrupting TRLs, docking analysis are currently on-going using our coordinates as a starting point for simulations. Herein, these simulations enforce the shift of selected ligands from the active site pocket to the intermonomeric cavity where an automated docking program finds binding poses that are compatible with our hypothesis.

4.2. NOVEL STRUCTURAL FEATURES AND PTMs IN eEF1A2

4.2.1. Structural Characterization of eEF1A2:GDP

4.2.1.1. Crystallization of eEF1A2:GDP

As mentioned in section 3.1.2. and 3.2 of Experimental Procedures, eEF1A2 protein was purified by our collaborators at PharmaMar S. A. and crystals of eEF1A2:GDP were obtained using the hanging-drop vapor diffusion method, mixing 1 µL protein solution with 1 µL reservoir solution (50 mM Na citrate pH 5/MES pH 5.6, 2.2-2.6 M ammonium sulfate, 10 mM Mg acetate) and equilibrating the drops against 500 µL crystallization solution in the reservoir.

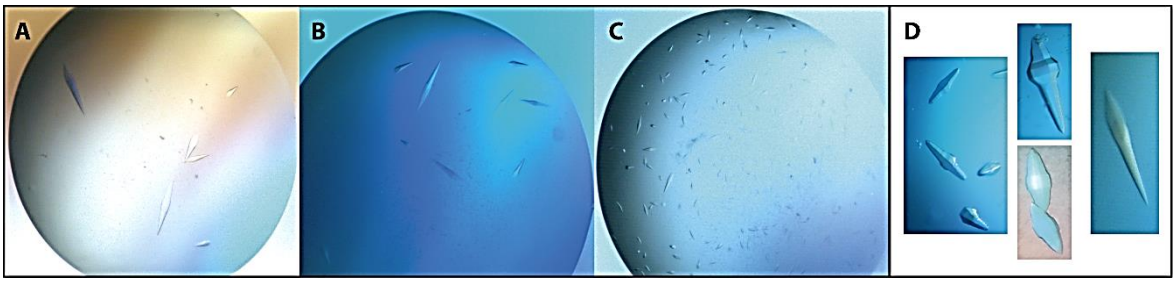


Figure 4.18: eEF1A:GDP crystals: Pictures were taken in the presence of polarized light, showing the slight ability of this crystals to change the direction of the electromagnetic wave. **(A)** Crystals obtained at 2.6 M ammonium sulfate, 50 mM Na citrate pH 5, 10 mM Mg acetate. **(B)** eEF1A2 crystals in 2.7 M ammonium sulfate, 50 mM Na citrate pH 5, 10 mM Mg acetate. **(C)** Crystals grown in 2.8 M ammonium sulfate, 50 mM Na citrate pH 5, 10 mM Mg acetate. **(D)** Close up of crystals obtained throughout crystallization trials.

As shown in Figure 4.18. (A-C), nucleation events increase as precipitant concentration increases. Furthermore, Figure 4.18.D shows crystals of different morphology, attaining to the variability of the sample. Herein, the sample was purified from natural source, leading to less reproducible results between batches and so the crystals yielded showed either arrow- or scepter-like shapes.

4.2.1.2. Structural Determination of eEF1A2:GDP

Crystals obtained throughout crystallization trials were taken to beamlines BL13 and ID29, at the ALBA Synchrotron (Cerdanyola del Vallès, Spain) and European Synchrotron Radiation Facility (ESRF, Grenoble, France), respectively, where diffraction patterns of sufficient quality for atomic structural resolution were obtained (Table 4.5.).

Diffraction patterns collected were indexed, integrated and scaled using XDS and Aimless. The best crystals diffracted to 2.3-2.5 Å and shown to belong to the P 6_1 2 2 space group with unit cell parameters a=b=133.52 Å, c=305.44 Å; α=β=90º, γ=120º. Nevertheless, the diffraction pattern demonstrated radiation damage, limiting the resolution to 2.7 Å and below were processed. Matthews coefficient is 3.89 A^3/Dalton, indicating two molecules per asymmetric unit and a 68.4 % of solvent. Because the same protein structure was already known, PDB 4C0S was used as a search model for molecular replacement in MOLREP.

Table 4.5.: Crystallographic data collection and refinement statistics for eEF1A2. Value for the highest resolution shell is shown in parenthesis.

Parameters	eEF1A2:GDP Complex
Data collection	
Space group	P 6_1 2 2
Cell dimensions	
a, *b*, *c* (Å)	133.52, 133.52, 305.44
α, β, γ (º)	90, 90, 120
Wavelength (Å)	0.979260
Resolution (Å)	61.17-2.70 (2.79)
Total reflections	585551 (48781)
No. unique reflections	45146 (4339)
R_{pim}	0.034 (0.320)
$CC_{1/2}$	0.995 (0.840)
$I/\sigma(I)$	12.3 (1.9)
Completeness (%)	100 (100)
Multiplicity	13 (11.2)
Refinement	
Resolution range (Å)	61.17-2.70
R_{work}/R_{free}	0.195/0.256
No. atoms	
Protein	7053
Water	183
Ligand	115
R.m.s. deviations	
Bond length (Å)	0.006
Bond angles (º)	1.485
Ramachandran favored/outliers (%)	98.8/1.2
Residues in AU	911
Average B value overall (Å^2)	80
PDB code	**6RA9**

Molecular replacement solution was refined using REFMAC (section 3.4.6.), with 10 cycles of restrained refinement. Manual adjustments of solvent-exposed mobile loops and Ramachandran/rotamer outliers, and addition of PTMs was carried out using COOT. Final refinement statistics were R_{work}=0.1995 and R_{free}=0.2557, shown in Table 4.5. RMS values for bond length and angles and Ramachandran outliers indicate an optimum geometry for the model obtained. On the other hand, the average B value is slightly above to what is usually expected, which may relate to scarce electron densities for multiple side chains of exposed residues. Despite the resolution and B value, 183 water molecules were observed, along with sulfate ions and acetate molecules present in the crystallization solution.

4.2.1.3. Structural Analysis of eEF1A2:GDP

Electron density maps allowed the tracing of the eEF1A2 protein structure from residues 4-462 (see below), a longer version to what had been already published (PDB 4C0S), revealing a novel fold in the C-terminal region. Furthermore, as opposed to PDB 4C0S, all residues had sufficient density for their Cα, and so all loops were pictured clearly in our structure.

4.2.1.3.1. Structure Overview

The final model contains two copies of the eEF1A2 (chains A and B) in a dimeric conformation, each containing a bound GDP molecule (451 residues for monomer A and 460 for monomer B). The three well-known domains are observed (Figure 4.19.; Figure 1.18.): domain I (residues 4-234, composed of 8 main α-helices and 8 β-laminas), responsible for nucleotide binding, and domains II (241-328, 6 β-laminas) and III (337-454 in A and 337-462 in B, 7 β-laminas), which adopt β-barrel structures. Superposition of individual domains shows high grade of resemblance between chains (Domain I: rmsd=0.394 Å; Domain II: rmsd=0.273 Å; Domain III: rmsd=0.236 Å). Nevertheless, the overall rmsd for all the Cα atoms backbone is 1.163 Å, thus indicating that changes between chains A and B are mainly due to slightly different rearrangements between domains.

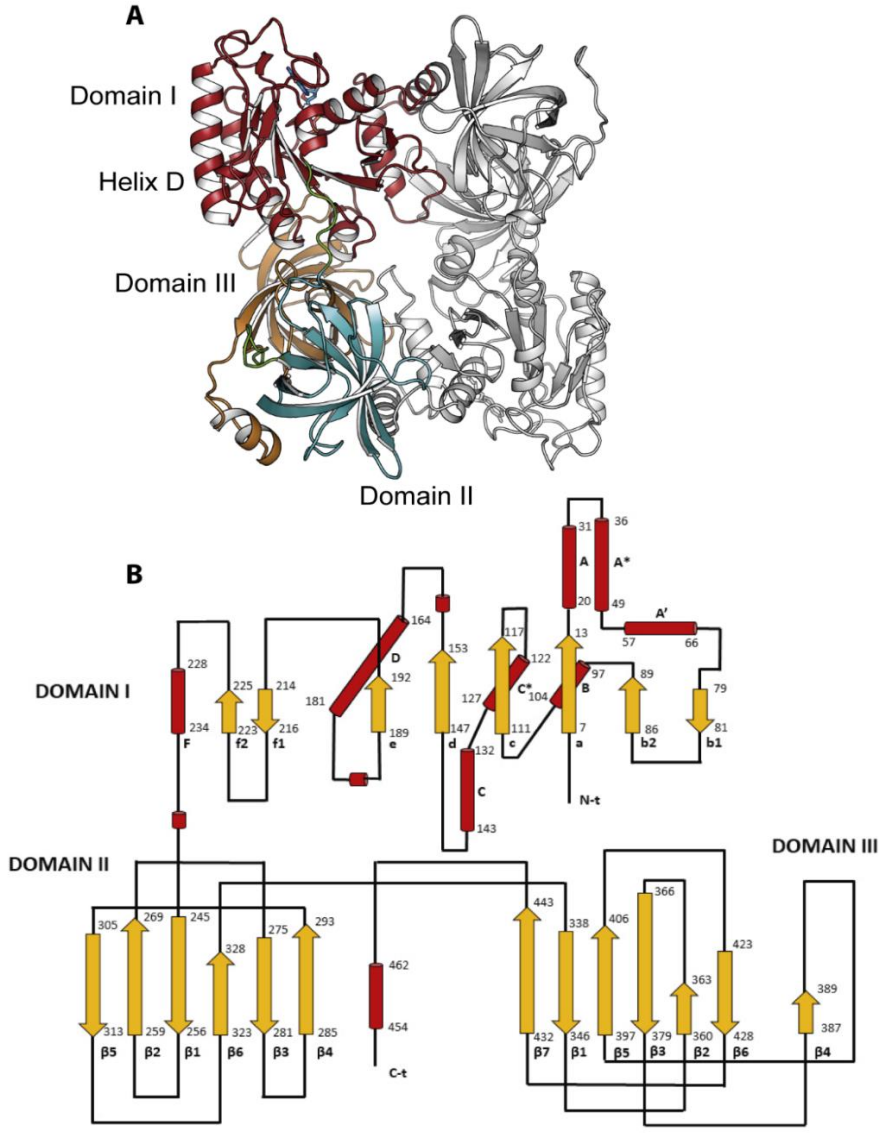


Figure 4.19: Structure overview of eEF1A2: (A) eEF1A2 crystallographic dimer. The elongation factor is represented in cartoons. Chain A is shows labelled domains and colored accordingly (domain I-red, domain II-light blue and domain III-copper) Chain B is coloured in grey. Helix D is pinpointed for reference. **(B)** Topology diagram of eEF1A. Helices are represented in red cylinders and β-laminas as yellow arrows. These are proportional to the number of residues composing the secondary structure. Connecting loops are represented as black lines and are not proportional to their length.

4.2.1.3.2. Crystal Packing

eEF1A2 crystallized in space group P 6_1 2 2, hence, a single hexagonal axis and two binary axes are present which relate the molecules within the crystal lattice. When symmetry operations are applied, 12 molecules may be fitted into the unit cell. Herein, the eEF1A2 dimer is stacked in a way that domain I of one monomer interacts with β-barrels (domains II and III) of the other. This dimeric conformation repeats throughout the crystal (Figure 4.20.). Conformational changes and structural differences, discussed below, are explained due to crystalline contacts with neighboring dimers that may stabilize interactions or sterically hamper the positioning of certain regions.

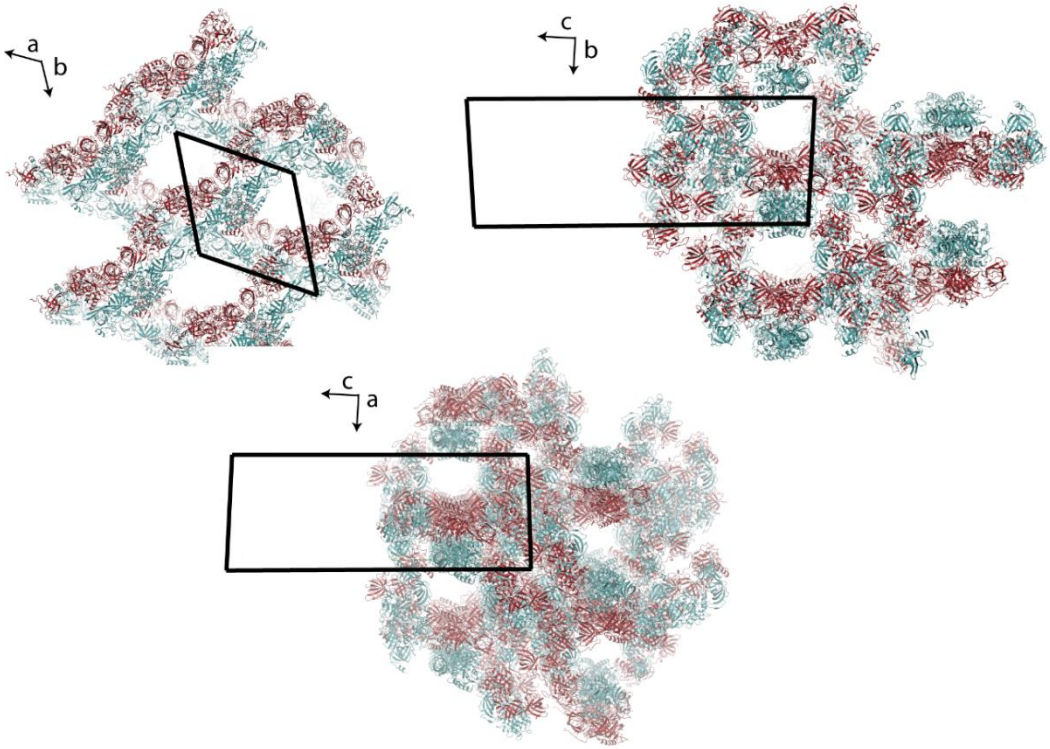


Figure 4.20: Symmetry and packing of eEF1A2:GDP crystals: The protein chains are represented as cartoon, chain A colored red and chain B in blue. Black tetrahedrons represent the unit cell and the orientation is indicated for each orientation.

4.2.1.3.3. Nucleotide Binding Site

The nucleotide-binding residues are conserved throughout species (Figure S1). In eEF1A2, recognition of GDP (Figure 4.21.) involves Asp17, Gly19, Lys20 and Thr22, which directly interact with the α- and β-phosphates of GDP, respectively; the side chains of Asn153 and Asp156 hydrogen bond to the guanine ring; and Lys154 interacts with the O1' ribose sugar. Other non-conserved residues engaged in the binding with GDP through hydrogen bonds with the main chain are Ser194 and Trp196. Moreover, Ser21 (Thr26 in E. *coli*) establishes hydrogen bonds with the β-phosphate of the nucleotide. In the case of chain A, a water molecule mediates

the interaction of Asp17 with the β-phosphate of GDP. This water molecule is positioned as the Mg^{2+} ion in the previously reported structure of eEF1A2 (PDB 4C0S). Nevertheless, our density is insufficient for a strongly bound Mg^{2+} ion. It is worth mentioning that this position is different to the Mg^{2+} ion found in *E. coli* EF-Tu, whose nucleotide exchange has been proved dependent on Mg^{2+} (Kawashima, *et al.*, 1996). Due to the absence of Mg^{2+} in our structure, the independence from Mg^{2+} for GDP binding is therefore demonstrated and supports the disengagement between the presence of Mg^{2+} and the structural rearrangements involved in nucleotide exchange, as opposed to EF-Tu (Vitagliano, *et al.*, 2004).

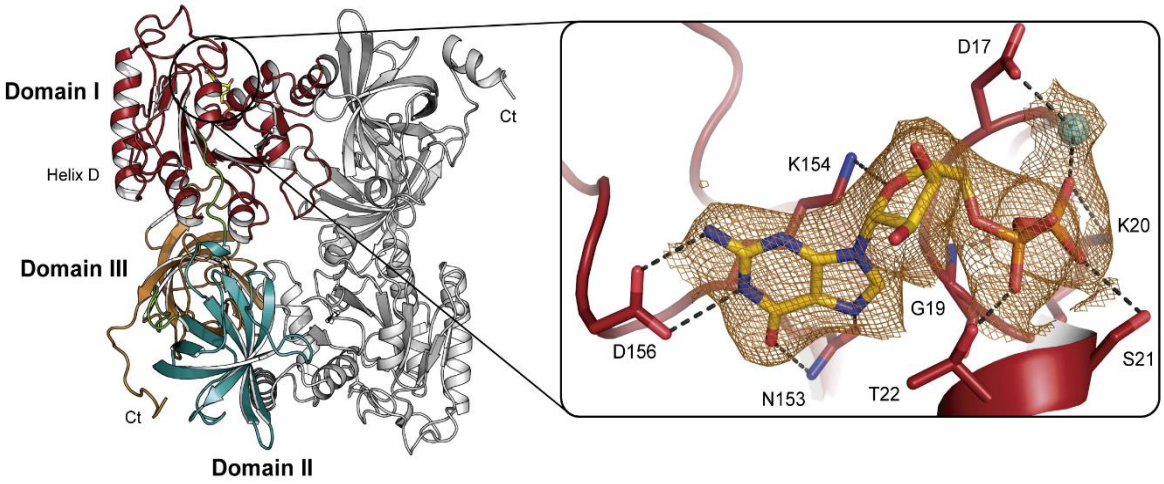


Figure 4.21: GDP binding site of eEF1A2. GDP is drawn in capped yellow sticks and its binding site is zoomed to the right. The eEF1A dimer is shown again as in Figure 4.19. as a reference for the positioning of the binding site. The 2Fo-Fc electron density map (contoured at 1 σ) is depicted for GDP. Residues involved in GDP binding are represented as sticks and main interactions are displayed. The water molecule, present in chain A, is shown as a blue sphere.

4.2.1.3.4. Structural Differences Between Chains A and B

Besides the differences within the nucleotide binding sites, other structural differences are observed between chains A and B. These differences mainly concern displacements of specific loops and the unfolding of helix C* (residues 122-127) (Figure 4.22.). The 122-127 region, unfolded in chain A, folds into an α-helix in chain B. This is due to the crystal packing mentioned in section 4.2.1.3.2., hence, distinct crystal environments are observed for this region in both chains that could affect the final structure. C* helix in chain B interacts with the same helix of a symmetrical chain, stabilizing this conformation. This same region from chain A is exposed to solvent channels, unfolding into a loop. These observed differences in secondary structure for this region point to an intrinsic feature of eEF1A2. It is worth to note that unfolding of helix C* causes a maximum lateral displacement of 1.7 Å of proximal helix D (residues 164-181), which propagates throughout β strands e, f1 and f2 of domain I (Figure 4.22.).

Minor loop displacements ranging from 1.5 to 2 Å throughout the structure lead to subtle translations of secondary structure elements. This can be observed, for example, in the loop

connecting β3 to β4 in domain III (residues 380-386), which shows a reorientation of 1.8 Å and causes a maximum lateral displacement of 1.6 Å of the β4 strand. Furthermore, switch I is a mobile region involved in nucleotide exchange (section 1.3.1.1.) and, in our case, presents two different conformations 3.75 Å apart in chains A and B (Figure 4.22.).

The C-terminal region also varies between chains (Figure 4.22.). Chain B shows 8 more residues, which fold into a helix (residues 454-462). This is a previously uncharacterized region of eEF1A2. Again, because of crystal packing, the C-terminus of chain A is closely situated to a symmetry-related molecule, forcing this region into a different conformation due to steric clashes. This results in the C-terminus being exposed into solvent channels, not showing up in the electron density maps, most likely due to mobility.

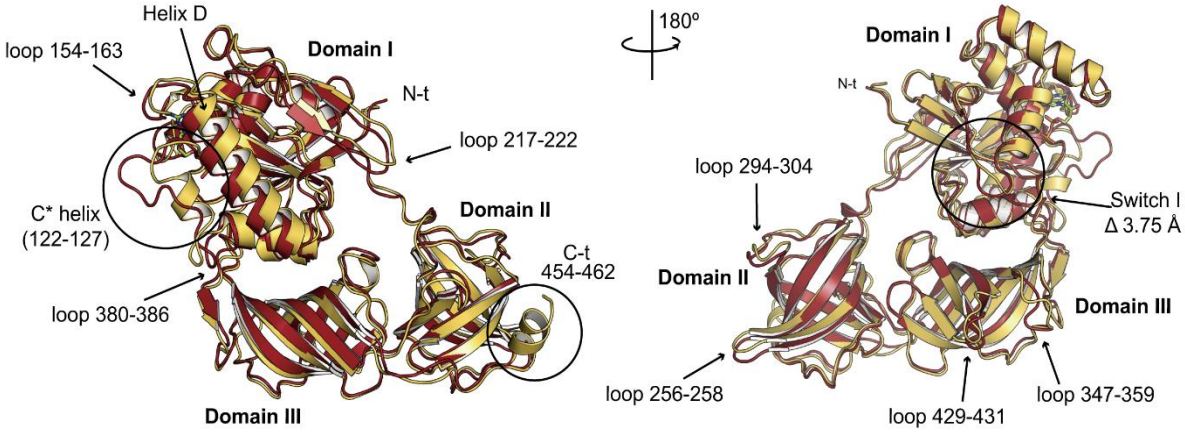


Figure 4.22.: Structural comparison of chains A and B of eEF1A2. Chain A is represented in red cartoon and chain B in yellow cartoon. Relevant regions are indicated by arrows or encircled and labelled.

4.2.1.3.5. Switching from GDP to GTP Bound Forms

As mentioned in section 1.3.1.1., strong structural rearrangements are known to occur for eEF1A when it switches from the GTP to the GDP conformation. Nevertheless, the GTP bound structure of the mammalian eEF1A was not known until end of 2016 (Shao, *et al.*, 2016). Considering the high sequence identity between eEF1A1 and eEF1A2, structural comparison with eEF1A1 in the GTP (active) conformation, as present in the ribosomal complex (PDB code 5LZS), allows to precisely predict the conformational changes in eEF1A2 from its GDP to GTP conformations (Figure 4.23.). When eEF1A is bound to GDP, the interfaces formed between domains comprise mainly helices C and B in domain I and strands β1, β5 and β7 in domain III. On the other hand, a new interface is created between the β-strands of domains I and II when GTP is bound. This implies an 84.3° rotation and 1.4 Å translation of both domains II and III with respect to domain I, as calculated with the DYNDOM server (Girdlestone and Hayward, 2016). Thus, domains II and III act as a whole structural unit with respect to domain I, thanks to the flexibility of loop 234-235, regardless of whether eEF1A binds to GDP or GTP.

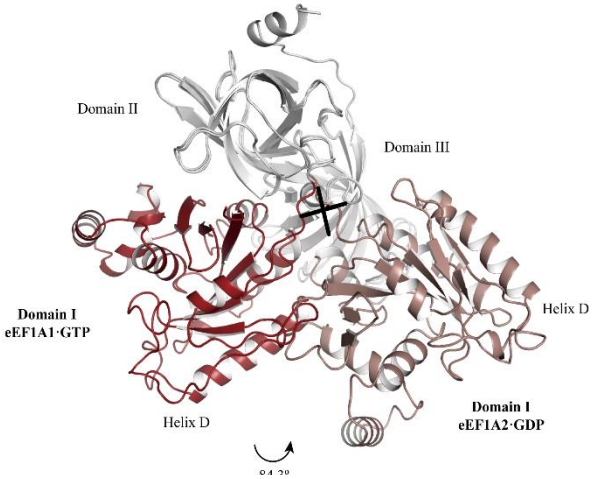


Figure 4.23.: Conformational change of eEF1A upon nucleotide hydrolysis. Structural superposition of eEF1A domains II and III (colored grey, which remain invariable whether the GDP or GTP is bound) allows to observe the changes in the relative position of domain I upon nucleotide hydrolysis. Changes from the GTP to the GDP conformation involve the rotation of domain I of 84.3º around an axis perpendicular to the plane of the paper (black cross) as calculated by DynDom server.

Domain rearrangement among nucleotide binding therefore provides a source of interaction surface variability. Moreover, the three domains of eEF1A2 present a diverse interactome (Panasyuk, *et al.*, 2008; Lambertim *et al.*, 2011). In particular, the interaction with eEF1B to reactivate eEF1A takes place through domains I and II when eEF1A is in the active conformation, and tRNA interacts with all domains, as shown in several complex structures deposited in the PDB (i.e. PDB codes 5O8W - yeast eEF1A:eEF1B-;1IJE - nucleotide exchange intermediates of yeast eEF1A:eEF1Ba complex-; and 1TTT - Phe-tRNA:EF-TU:GDPNP ternary complex of *Thermus aquaticus*-).

4.2.1.4. Novel Structural Features of eEF1A2

As mentioned throughout this Dissertation, a previous eEF1A2 structure had been published in 2014 (PDB 4C0S). Nevertheless, our structure presents novel features and dissimilarities (Figure 4.24.):

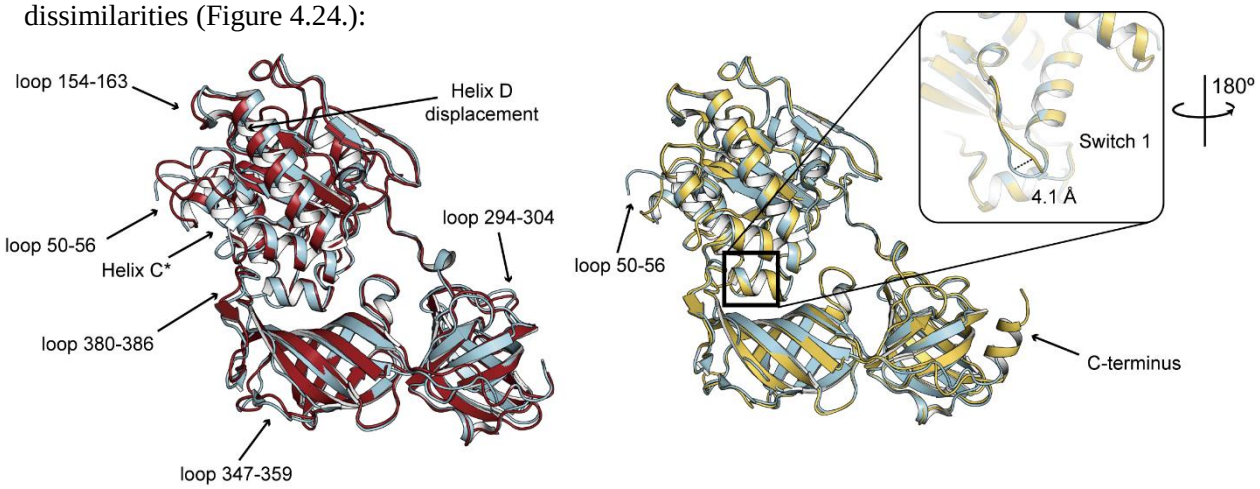


Figure 4.24.: Superposition of chains A and B of the present eEF1A2 with PDB 4C0S. Our chains A and B are represented in red and yellow cartoon, respectively. PDB 4C0S is shown in blue cartoon. Major differences are pointed by arrows**.**

While the overall backbone conformations are comparable (average rmsd of 0.65 Å between chains for Cα atoms), relevant differences are observed in loops 50-56 and 154-163 of domain I (Figure 4.24.), as well as other subtle loop reorientations. The C-terminal region (residues 454-462) was left untraced in the previous structure of eEF1A2, but is now observed for the first time in a non-mitochondrial eEF1A. Minor changes between structures also involve the relative side-chain orientation of many lysines and arginines. This is likely due to the stabilization of amino and guanidinium groups by electrostatic interactions with SO_4^{2-} ions present in the crystallization solution, the distribution of which appears to mimic that of the phosphates in the backbone of a single-stranded RNA molecule (Figure 4.25.).

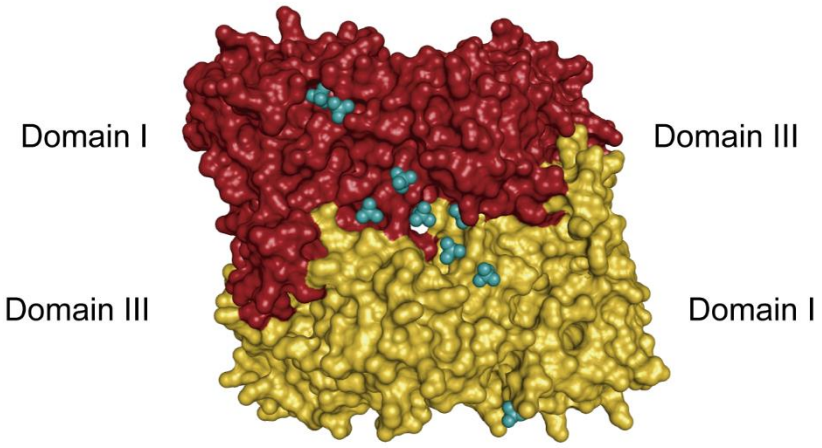


Figure 4.25.: Sulfate ions disposition in the eEF1A2 dimer. Chains A and B are represented in red and yellow cartoon, respectively. Sulfate ions are represented as blue speres. SO_4^{2-} ions are assembled in the surface comprised by domains I and III.

The previous rabbit eEF1A2 structure displayed an incomplete loop (residues 50-56) in chain A. Tracing of this loop connecting helices A* and A' (Figure 4.24.) in domain I has been possible in both A and B chains of the present structure, despite the increased B-factors in both loops and their different conformations, which attest to the mobility of this region. The same scenario is observed in switch I, which presents different conformations in all chains between PDB 4C0S chains and ours. This loop shows a maximum displacement of 4.1 Å in chain B with respect to the 4C0S structure (Figure 4.24.).

Different loop arrangements can also be witnessed in domains II and III, the most noticeable being those comprising residues 294-304 between β4-β5 in domain II and those joining β1-β2 (residues 347-359) and β3-β4 (residues 380-386) in domain III. Furthermore, the previously mentioned unfolding of helix C* in domain I is only observed in chain A, thus causing the displacement of helix D, as discussed above.

Regarding the last 8 residues (residues 454-462), which comprise a lysine-rich region, these fold into an α-helix in chain B (Figure 4.24.). Despite the fact that this region presents higher B-factors overall as well as an exposed location and minimal contacts suggesting flexibility, the

three residues (Gly449-Lys450-Val451) preceding the C-terminal α-helix establish hydrogen bonds with β4 of domain II in a quasi-β-sheet manner (Figure 4.26.) keeping the C-terminal α-helix close to the protein core. This folding of the C-terminus has also been observed in bovine mitochondrial EF-Tu (PDB 1D2E), which has been alleged to possibly interact with tRNA (Andersen, *et al.*, 2000).

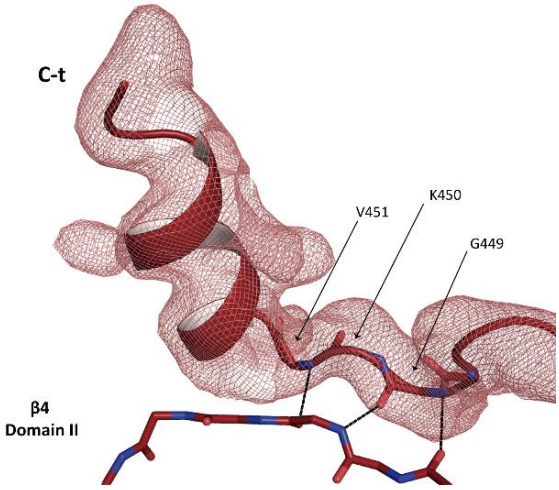


Figure 4.26: C-terminal region of eEF1A2. The C-terminal region of chain B of the present structure is shown as cartoon. The main chain of the residues involved in stabilizing this region are shown and labelled. The main chain of β4 of domain II is shown as sticks**.** Side chains are not represented for clarity reasons. The network of hydrogen bonds between the loop connecting the C-terminal helix and domain II are shown by dashes, which resemble a β-sheet.

4.2.1.5. Structural Characterization of PTMs in eEF1A2 by X-ray Crystallography and Mass Spectrometry

PTMs in mammalian eEF1A play an important role in differentiation between highly homologous eEF1A1 and eEF1A2 forms (Soares and Abbott, 2013) and also as a potential source of moonlighting activities and regulation for this promiscuous protein. Direct isolation of eEF1A2 protein from rabbit muscle allowed us to study the nature of such PTMs. Hence, mass-spectrometry (MS) experiments were carried out as a complementary tool for identification of other PTMs in our sample. MS assays were carried out at Proteomics Unit in the Faculty of Pharmacy at the Complutense University of Madrid. We performed data mining and interpretation of results.

Briefly, the MS/MS data acquired were analyzed with different software to identify peptides and PTMs in eEF1A2. Proteome Discoverer software v.2.2 (Thermo Scientific) with search engine MASCOT 2.6 (MatrixScience, London, UK) and Sequest HT were used to identify the peptides against a home-made database with a FASTA sequence (Q71V39) and contaminant database of 247 sequences. Furthermore, Peaks Studio v. 8.5 Software (Bioinformatics solution Inc., trial version), was also used for two LC-MS/MS data analysis to identify peptide/protein *de novo* sequencing and PTM analysis.

The PTMs observed in our crystal structure and/or determined by MS, some of which have been identified for the first time in eEF1A, are shown in Figure 4.27. together with their distribution over the eEF1A2 structure. Full lists of PTMs characterized by MS surpassing the acceptance criteria (Protein Discoverer: FDR<1%, where at least one unique peptide was identified with high confidence (CI>95%, p<0.05). Only modified peptides with a site probability ≥75 % were considered. Further limitations assumed trypsin digestion with up to 2 miss-cleavage allowance and a fragment ion mass tolerance of 0.02 Da. Peaks Studio: database searches, PSM-FDR< 0.5; peptide -1OlogP ≥ 23.7; protein -1OlogP ≥ 20 with at least one unique peptide. De novo sequences or tags were filtered by ALC score ≥ 80, meaning p< 0.01), are given in Tables S3-S8 in Supporting Information.

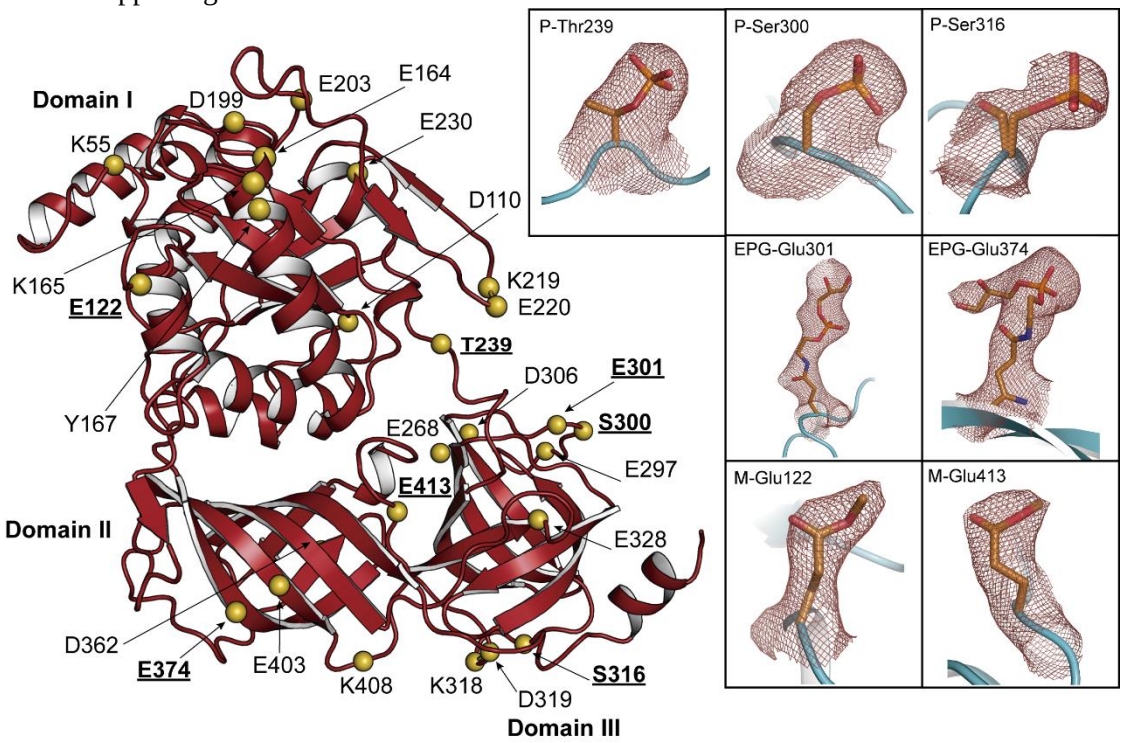


Figure 4.27.: PTMs on eEF1A2 observed by MS and X-ray crystallography analyses. The eEF1A2 backbone is represented as a red cartoon. The Cα of each residue displaying a PTM is pinpointed as a yellow sphere and labelled accordingly. Only those PTMs which have been identified by both Protein Discoverer and PEAKS Studio software are shown. Amino acids whose PTMs are observed in the crystal structure are underlined. The associated electron density (2Fo-Fc map contoured at 1σ) for each modified residue is depicted in the boxes on the right. S316 displays a double conformation. The GDP molecule is removed for clarity purposes. P-Thr: phosphorylated threonine; P-Ser: phosphorylated serine; EPG-Glu: EPG addition to glutamate residue; M-Glu: methylated glutamic acid.

In total, seven PTMs were found consistent with the electron density map after a complete refinement of the whole structure (Figure 4.27.). These PTMs fall into three different categories: (i) attachment of EPG to Glu301 and Glu374 of chain A and B respectively; (ii) phosphorylation of Ser300 and Ser316 of chain A and in Thr239 of chain B; and (iii) methylation of Glu122 and Glu413 in domains I and III, respectively, of chain B. Of these, only Thr239 appears phosphorylated in publicly available 3D structures of both eEF1A1 in the ribosomal complex (PDB code 5LZS) and eEF1A2 (PDB code 4C0S). In stark contrast, although the unique

EPG modification of eEF1A is of general knowledge, it has not been observed or reported in any crystallographic structure published to date.

EPG modification of Glu301 in domain II was detected by MS using both Protein Discoverer and PEAKS software; the latter identified this PTM in a scarce 7.28% of the peptides analyzed (Table S8 in Supporting Information). On the other hand, EPG on Glu374 in domain III is accounted for only by the PEAKS software. This is likely due to the fact that, because this amino acid is adjacent to lysine and arginine residues in the protein sequence, a larger number of missed trypsin cleavage sites than those allowed in the Protein Discoverer peptide search would be necessary. Nonetheless, the PTM search performed by PEAKS, in which the spectra with high confident de novo scores that are not assigned by database search are mapped against the identified proteins, allowed us to confirm the modification of Glu374 by EPG in 33.7% of the peptides analyzed.

Table 4.6.: Summary of post-translational modifications (PTMs) identified by mass-spectrometry.

Detection Software	*Proteome Discoverer and PEAKs Studio*	**K55 (A)**, D110 (M), E122 (M), E164 (M), **K165 (A)**, **Y167 (P)**, D199 (M), E203 (M), K219 (A), E220 (M), E230 (M), E268 (M), E297 (M), **S300 (P)**, E301 (EPG,M), D306 (M), **K318 (A)**, D319 (M), D328 (M), D362 (M), E403 (M), **K408 (A)**, E413 (M)
	Proteome Discoverer	**T23 (P)**, E124 (M), S224 (P), **K244 (A)**, E293 (M), **K395 (A)**, D398 (M)
	PEAKs Studio	D35 (M), **K36 (A)**, **K146 (A)**, D156 (M), **E374 (EPG)**

Letters in parenthesis indicate the PTM observed: A-acetylation, EPG-ethanolamine phosphoglycerolamination, M-methylation, P-phosphorylation. PTMs have been grouped according to the software of detection, all having a p-value < 0.05. PTMs in bold have been previously reported (Soares and Abbott, 2013).

In agreement with previous reports for PDB entries 5LZS and 4C0S, electron density consistent with phosphorylation of Thr239 was also visible in our maps (Figure 4.27.) albeit this PTM was not identified by the MS experiments. Furthermore, the high-resolution electron density map reveals the phosphorylation of Ser316 despite the lack of evidence from the MS experiments. In this regard, it is noteworthy that the phosphorylation prediction server NetPhos 3.1 (Blom, *et al.,* 2004) does identify this modification with a 99.7% trust score. On the other hand, both Proteome Discoverer and PEAKs identified Ser300 phosphorylation, surpassing both acceptance criteria (Table 4.6.) and we do observe this PTM in our electron density maps (Figure 4.27.).

While most of the PTMs described in literature for eEF1A2 seem to be involved in specific recognition by other protein partners, our work reveals that some of the identified PTMs have also a direct impact on structure and/or catalysis. P-Ser316 and EPG-Glu374 are fully exposed modifications that alter the intramolecular interactions pattern on eEF1A2's surface by creating (P-Ser316) or altering (EPG-Glu374) salt bridge interactions (Figure 4.28.).

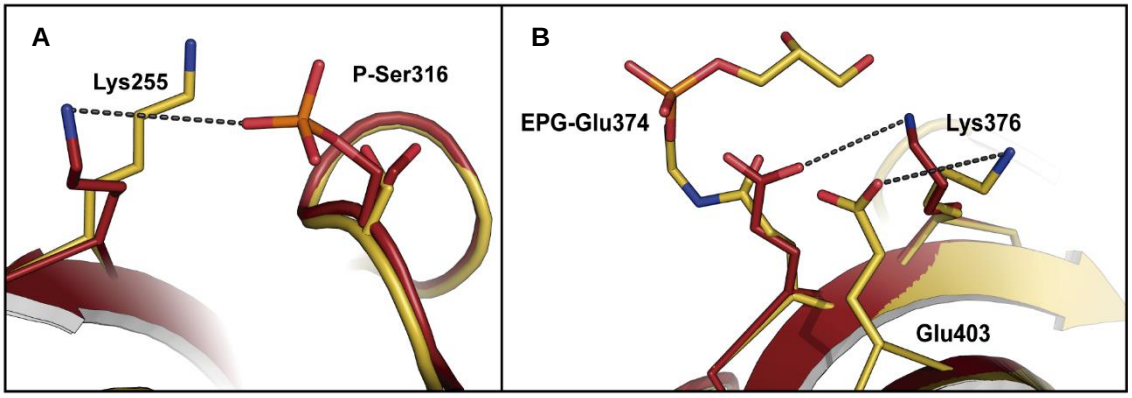


Figure 4.28.: Modifications of the interactions of P-Ser316 and EPG-Glu374 within eEF1A2: Chain A and B of eEF1A are shown in cartoon and colored red and yellow respectively. Those residues involved in the modification of interactions within the molecule are shown in stick form and colored accordingly. In the case of phosphorylated Ser316 (P-Ser316), potential new electrostatic interactions are made with Lys 255. In the case of the addition of EPG to Glu374, the electrostatic bond formed with Lys376 is broken, which leads to Lys376 to electrostatically interact with nearby Glu403 in the neighboring β-lamina.

Other modifications such as P-Thr239 or M-Glu413 directly affect interactions between domains (Figure 4.29.A and Figure 4.29.B, respectively). Moreover, the phosphorylation of Thr239 (observed in our structure and in PDB 4C0S), located in the linker between domains I and II, allows creation of salt bridge interactions with Lys244 (domain II) or with Lys146 (domain I in PDB 4C0S) (Figure 4.29.A) strengthening the connection between both domains. Modification of Glu413 has also implications in interdomain interactions and likely in the stabilization of the C-terminal region of eEF1A2. Native Glu413 is involved in a large network of interactions (Figure 4.29.B.) connecting domains II and III in which Arg247 plays a central role. Glu413 residue establishes a double salt bridge linking Arg247 (domain II) with Lys443 (domain III). When Glu413 is methylated, electrostatic interaction with Lys443 is broken, resulting in a change in its conformation. This conformational change seems to propagate along the C-terminal region

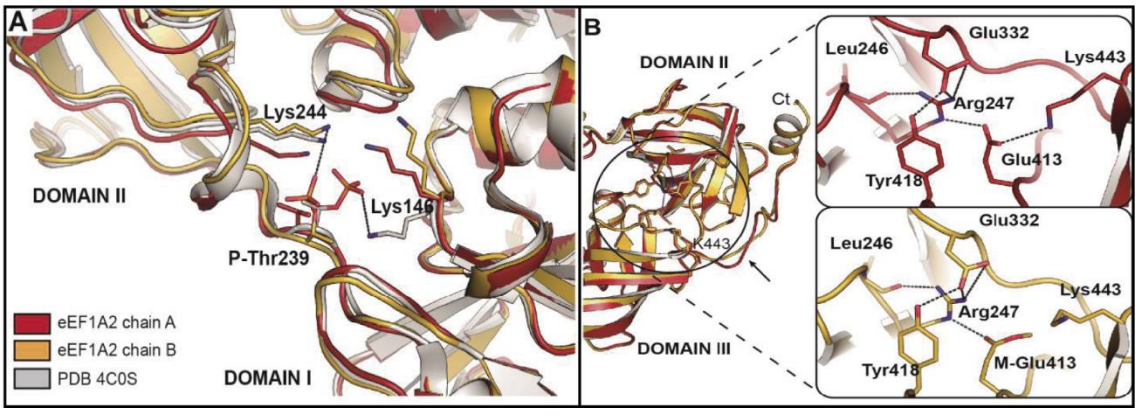


Figure 4.29.: Structural implications of P-Thr239 and M-Glu413 in eEF1A2. (A) Superposition of Chains A (red) and B (yellow) of our eEF1A2 structure and PDB 4C0S (grey). Residues involved in the new interactions are portrayed in sticks. Phosphorylation of Thr239 (P-Thr239) allows new interactions between domains I and II to be formed. **(B)** Structural superposition of chains A (red) and B (yellow) of eEF1A2 (present work). Native Glu413 (chain A) is involved in a network of interactions connecting domain II and III. Methylation of Glu413 (chain B) abolish interaction with Lys443. Differences in the C-terminal region in both chains is highlighted with an arrow**.**

that shows a different structure in the presence or absence of PTM at Glu413. Interestingly, at the end of this region the C-terminal helix exists, only observed for the chain containing this PTM.

PTMs such as M-Glu122, also affect the shape and properties of the essential GDP binding site (Figure 4.30.). In the absence of this PTM, as shown in PDB 4C0S, Glu122 establishes a double salt-bridge interaction with Lys154, keeping the ε-amino group at 2.7Å of the O' of the ribose moiety of GDP. When Glu122 is methylated (as observed in chain B of our structure), the electrostatic interaction with Lys154 is weakened and this leads to a conformational shift of Asp17 towards Lys154 that allows formation of a new electrostatic interaction with this residue. This is reflected in a longer distance between Lys154 and ribose (3.0 Å). Another structural configuration is also possible in the GDP-binding site. Unfolding of α-helix C* (observed in chain A of our structure) orientates Glu122 towards the solvent and towards the outside the GDP binding site, resulting in a longer distance between Lys154 and ribose (3.5Å). In this case, Asp17 is found involved in stabilization of phosphate groups in GDP via a water molecule.

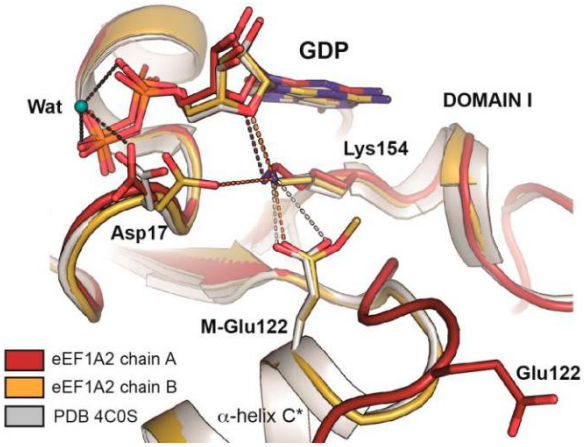


Figure 4.30.: Conformational changes involving the methylation of Glu122. In the absence of this PTM (PDB 4C0S), Glu122 establishes a double salt-bridge interaction with Lys154, involved in the binding of the ribose in GDP (2.75Å). When Glu122 is methylated (chain B) the salt-bridge interaction with Lys154 is weakened, leading to an increase distance with ribose (3.0Å). Unfolding of α-helix C* (chain A) orients Glu122 towards the solvent weakening the interaction of Lys154 with ribose in GDP (3.5 Å). Protein is represented in cartoon and colored as indicated in the key. Residues involved in the interaction are shown as sticks. The interactions are represented by dashed lines and colored as protein structures.

4.2.2. Plausible Implications of PTMs and the C-Terminal Helix of eEF1A2

4.2.2.1. The C-Terminal Region as a Source of Regulation in Protein Translation

The structure of the C-terminal region (residues 454-462) has been identified in eEF1A2. While the physiological role for this amino acid stretch has not been described for eEF1A yet, our structural analysis reveals potential in vivo implications. The structure of the mammalian ribosomal elongation complex (PDB code 5LZS) obtained by cryo-electron microscopy (Shao, *et al.*, 2016) shows eEF1A1 in its GTP conformation interacting with the protein synthesis inhibitor didemnin B (Figure 4.31.). This complex allows a direct comparison between the GDP

conformation and the GTP conformation of eEF1A when interacting with the ribosome. Interestingly, both conformations keep the C-terminal-binding surface intact, moving as a whole together with domains II and III. Thus, the hydrogen-bond interaction network connecting the C-terminal main chain with domain II can be preserved in the ribosome-bound state. Furthermore, the eEF1A C-terminus would lay between domain II and the rRNA of the small ribosomal subunit, providing a positively charged interface due to amino acid composition that could directly interact with the ribosome's nucleic acids (Figure 4.31.).

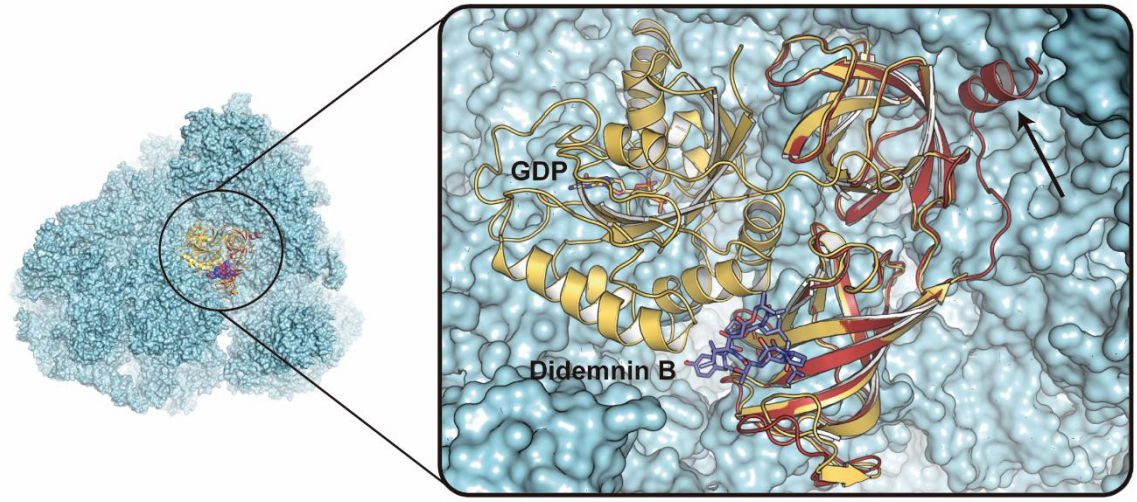


Figure 4.31.: C-terminal helix interacts with the ribosome. Superposition of domains II and III of the eEF1A2:GDP complex here solved onto mammalian ribosomal elongation complex with aminoacyl-tRNA, eEF1A, and didemnin B (PDB code 5LZS). The new region at the C-terminal region of eEF1A2 (highlighted by an arrow) nicely accommodates into the ribosomal surface and may interact with the negatively charged phosphate backbone of the 40S rRNA (ribonucleoproteins of the 80S ribosome are represented in cyan surface mode). Domains II and III of eEF1A2:GDP (red) have been superposed onto eEF1A1 in its GTP conformation (yellow). GDP and didemnin B from PDB 5LZS have been represented in yellow and purple sticks respectively.

The same folding of the C-terminal region has also been identified in the bovine mitochondrial EF-Tu:GDP complex (Andersen, *et al.*, 2000a) and ascribed the role of interacting with tRNA. Structural superposition of domain III of the our eEF1A2:GDP complex onto the bovine mitochondrial EF-Tu:GDP complex (PDB code 1D2E) indicates that the C-terminal regions are oriented differently in both cases (Figure 4.32.A.). In order to assess a potential interaction of this region with tRNA we superimposed both structures onto the Phe-tRNA:EF-Tu:GDPNP ternary complex of *Thermus aquaticus* (Nissen, *et al.*, 1995) (PDB code 1TTT). As observed in Figure 4.32.B., both C-terminal helices are far away from the tRNA. Thus, major conformational changes of the C-terminal region would be required for interaction with tRNA.

Considering the regulation of the interaction of this C-terminal region with tRNA, lysine methylation in eEF1A has been related to efficient mRNA translation, as it influences the ability of eEF1A to interact with aminoacyl-tRNA and/or mediate its interaction with the ribosome (Jakobsson, Malecki and Falnes, 2018). Lysines are predominant in the amino acid composition of the C-terminus, but our MS experiments failed to identify PTMs in this region. Although

methylesterification is a specific C-terminal modification in yeast (Zobel-Thropp, *et al.*, 2000), the only evidence to date of C-terminal PTMs in eEF1A are the previously reported acetylation events predicted for Lys453, Lys457 and Lys460 (Soares and Abbott, 2013; Hornbeck, *et al.*, 2015).

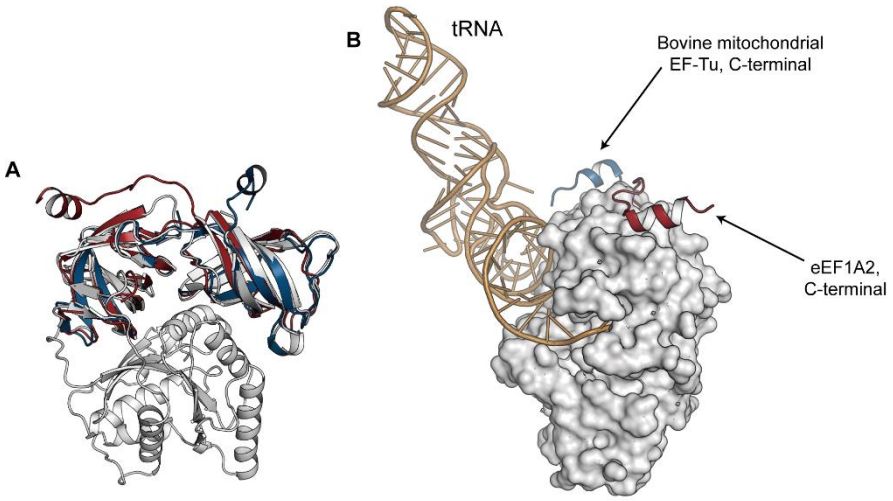


Figure 4.32.: Conformational changes of the C-terminal region are needed for tRNA interaction. (A) Structural superposition of domains II and III of the present eEF1A2 structure (red cartoon) and of bovine mitochondrial EF-Tu (PDB: 1D2E; blue cartoon) onto the same domains of *Thermus aquaticus* EF-Tu in complex with Phe-tRNA (PDB: 1TTT; grey cartoon). **(B)** *T. aquaticus* EF-Tu is represented as a surface, the tRNA molecule is represented as cartoon and coloured in yellow. Terminal regions of bovine mitochondrial EF-Tu is represented as a blue cartoon and chain B of the crystal structure of eEF1A2 is represented as a red cartoon. As stated by Andersen, G. L. *et al.* (Andersen, *et al.*, 2000a), the C-terminal region of mitochondrial EF-Tu may interact with tRNA as it lies in a close location to it. In the case of eEF1A2, the C-terminal helix is repositioned and folded onto the β4 stand of domain II, facing away from the tRNA molecule. Due to this different orientation, the C-terminal region of eEF1A2 lies at a further distance from the tRNA, and so extreme repositioning would be necessary in order to interact with the tRNA molecule. Nevertheless, the liker between the β7 lamina of domain III and the C-terminal helix is longer in the case of eEF1A2.

4.2.2.2. PTMs Lie on the Surface of the eEF1A Dimer

The previous eEF1A2 structure crystallized in the same dimeric configuration. Crepin, *et al.* suggested that the dimer observed in the crystal structure of eEF1A2 would represent a possible in vivo arrangement for eEF1A (Crepin, *et al.*, 2014). Dimerization of eEF1A would be required for actin bundling (Bunai, *et al.*, 2006), its stability and function via Raf kinases (Sanges, *et al.*, 2012), and its interaction with tumor suppressor p16INK4a (Lee, *et al.*, 2013). Recently, heterodimer formation between eEF1A1 and eEF1A2 in the cellular environment has been demonstrated by using a FRET approach, both in the cytoplasm and in the plasma membrane (Migliaccio, *et al.*, 2018). Moreover, this localization has also been observed in HeLa cells for eEF1A2-GTP-plitidepsin complexes, which are assumed to contain at least one subunit of this protein (García, *et al.*, 2018). Since both eEF1A1 and eEF1A2 display such a high sequence identity, it has been considered that quaternary structures of eEF1A can be present as homodimers or heterodimers (Sanges, *et al.*, 2012; Lee, *et al.*, 2013).

Structural analysis reveals that, when mapped onto the crystallographic dimer, both the PTMs reported in this work and those reported by Abbott et al. (Soares and Abbott, 2013) are mainly exposed and lie outside the dimerization interface (Figure 4.33.). In other words, the PTMs are mostly located in the "sequence-wise variable face", in other words, those areas of the protein which vary in primary structure between isoforms eEF1A1 and eEF1A2). Thus, the dimerization interface is conserved and barely any residues are subjected to PTMs that could interfere with monomer-monomer interactions. Exceptions are (i) Asp35; (ii) Ser76 and Thr432; and (iii) Tyr254 and Lys318, which are target to methylation, phosphorylation and acetylation events, respectively. PTMs on residues responsible for dimerization could affect the efficiency of eEF1A oligomerization. This suggests that these residues are involved in directing eEF1A's functions by regulating its oligomeric form. Nonetheless, a conformational change triggered by heterodimerization has been proposed (Sanges, *et al.*, 2012), leading to the exposure of residues usually involved in nucleotide binding, as in the case of Ser21. This would facilitate the access to previously unavailable residues, which would undergo PTMs in this region, as described by (Soares and Abbott, 2013) and in the present research. Furthermore, the intrinsic properties of the unfolding of α-helix C* would also allow modifying enzymes to access these unreachable residues.

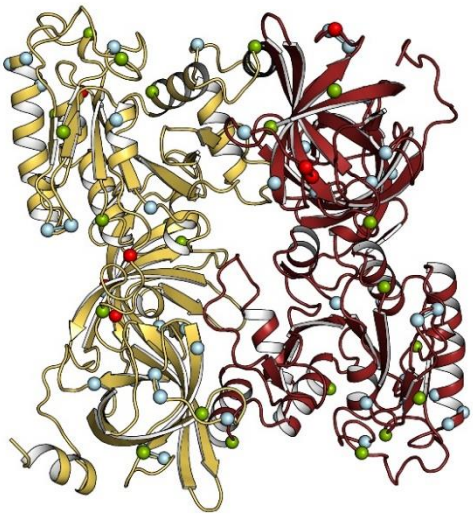


Figure 4.33.: Dimer formation in the crystal structure and distribution of PTMs. Chain A and B are represented in cartoon, in red and yellow respectively. Cα of residues with PTMs identified in this study are shown as spheres. Those identified by both Proteome Discoverer and PEAKs Studio software are coloured in light blue and those identified by either one or the other in green. Those PTMs whose electron density is found in the crystal structure are coloured in red. It is observed that the dimerization interface present in the crystal structure scarcely contains any PTMs, but are present in the outer surface of the dimer.

4.2.2.3. PTMs as a Source of Regulation to a "Moonlighting" Protein

PTMs entail a source of diversity that relates to conformational changes, new functions and variable interaction partners. In eEF1A, numerous PTMs have been reported with strong relevance in regulation and function. Among the different kinds of PTMs observed in our crystal structure, the addition of EPG to Glu301 and Glu374 most noteworthy. This widely known

modification is unique and strictly specific for these glutamate residues of mammalian eEF1A but has never before been observed in any crystal structure, to the best of our knowledge. A model for the attachment of EPG to the elongation factor has been proposed (section 1.3.3.2., Figure 1.21.) in which the eEF1A is first modified by phosphatidylethanolamine (PE) and then deacylated to EPG. Considering that eEF1A can localize close to membranes (Migliaccio, *et al.*, 2018), the presence of PE in eEF1A would allow the protein to bind to membranes and return to the cytoplasm after deacylation. This could correlate with the observed non-canonical eEF1A functions in proximity to the membrane, such as its participation in the nuclear export of TD-NEM-containing proteins by interacting with von Hippel-Lindau (VHL) tumor suppressor and poly(A)-binding protein (PABP1) (Khacho, *et al.,* 2008*;* section 1.3.3.1.). Moreover, the rearrangements of Lys376 observed in our crystal structure caused by the addition of EPG reveals possible new interaction surfaces and networks for domain III (Figure 4.28.).

Phosphorylation is known to play a critical role in the regulation of cellular processes, being the most common means of protein function regulation and signal transmission throughout the cell. The phosphorylated residues observed in our eEF1A2 crystal structure are Thr239, Ser300, and Ser316. The function of some phosphorylated residues in eEF1A is known (Negrutskii, Vlasenko, and El'skaya, 2012) but the vast majority of phosphorylation repercussions remain unclear. Herein, P-Thr239 establishes salt-bridge interactions with domains I and II, and so could be involved in the stabilization and flexibility of these domains. On the other hand, Ser300 and Ser316 are present in mobile and exposed loops and their modification would allow the creation of a new pattern of both intra- (Ser316) and inter-molecular (Ser300) interactions.

With respect to the consequences of other PTMs, methylation of Glu413 may regulate the rearrangement between domains II and III, and stability of the C-terminal region, whereas methylation of Glu122 directly impacts on GDP stabilization via Lys154. Hence, due to the conformational differences observed between chains A and B in our structure and those in PDB entry 4C0S, we can infer that the involvement of Asp17 in GDP binding and the strength of the interaction with Lys154 are modulated by methylated Glu122 and the conformation of the C*-helix (Figures 4.29. and 4.30.).

Considering the strong structural conservation between species for this elongation factor, we can infer which residues prone to PTMs may interfere in tRNA or eEF1B binding. Alignment of eEF1A2 with EF-Tu in the Phe-tRNA:EF-Tu:GDPNP ternary complex (PDB code 1TTT) shows Lys318 and Lys408, which can be acetylated, interacting directly with bases in the aminoacyl tRNA region and with the phosphate backbone of tRNA, respectively. Other modified residues present in this interaction surface include Ser76, Tyr254 and Tyr257 (susceptible to

phosphorylation) and Lys255 (prone to acetylation), previously described in (Soares and Abbott, 2013), if conformational changes of its side chains are considered. When domains of eEF1A2 are structurally aligned to equivalent domains of homologous *S. cerevisiae* eEF1A1 in complex with eEF1Bα in nucleotide exchange intermediates (PDB codes 1IJE, 1IJF, 1G7C) it can be observed that Ser21, Ser76, Tyr254, Glu293 and Glu297, all of them subject to PTMs, lie at the interface between eEF1A and eEF1Bα. Except for the methylation of Glu293 and Glu297 identified here by MS, phosphorylation events have been previously reported for the former residues (Soares and Abbott, 2013). It is tempting to suggest that these modifications may play a role in the modulation of eEF1B and tRNA binding and hence, efficiency of nucleotide exchange and translation, which could provoke the switching of eEF1A towards non-translation-related functions.

Nevertheless, the interface between eEF1A and these structurally characterized binding partners comprises a region that is scarce in PTMs. This observation underscores the preservation and invariability of this binding site, as most of the PTMs identified lie on the outer surface of the protein surrounding the interface between eEF1A and eEF1Bα and tRNA.

Deciphering the binding partners of eEF1A is crucial to understanding the network of molecular interactions in which this protein is involved and how its structural domains and PTMs may modulate eEF1A's numerous functions and connections.

Several studies have been carried out in which different molecules have been identified to interact with eEF1A, either involving the entire protein or just localized regions within specific domains, some reviewed in section 1.3.3.1. Thus, it has been shown that (i) F-actin interacts with domains I and III of *Dictyostelium* EF1α (Liu, *et al.,* 1996) and domain II of human eEF1A from an epidermoid cancer cell line (Lamberti, *et al.,* 2008); (ii) activation-induced deaminase (AID) binds to Domain III of eEF1A leading to cytoplasmic retention (Hälser, Rada and Neuberger, 2012); (iii) SH3 domain-containing adaptor protein SORBS2, involved in the assembly of signaling complexes, interacts with domain II of eEF1A1 near the membrane (Lamberti, *et al.,* 2011); and (iv) SH2 and SH3 domains of diverse signaling molecules, such as Crk, Fgr, Fyn, Grb2, RasGAP, Shc and Shp2, possess different abilities to bind to phosphotyrosine-containing sites in domain I of both eEF1A1 and eEF1A2 (Panasyuk, *et al.,* 2008). Further studies directed at characterizing eEF1A complexes and the PTMs involved are required to clarify, at the molecular level, the plethora of events in which this elongation factor is involved.

4.2.3. Mass Spectrometry and X-ray Crystallography in PTM Identification

In our study, we used a combined approach by X-ray crystallography and MS proteomics to characterize PTMs in eEF1A2 isolated from rabbit muscle. While basic agreement is found

between both techniques, crystallography fails in the identification of many PTMs suggested by proteomics (likely due to flexibility and/or not complete modification in most of the proteins composing the crystal) and MS also misses the characterization of some of the PTMs identified by crystallography. It is worth noting that the lack of PTM detection in eEF1A2 by MS does not necessarily mean that those residues were not modified in that particular sample. Coverage of the full protein sequence is highly challenging and some modifications may be lost during the preparation and analysis of the sample, as exemplified by labile phosphate groups that can be released during fragmentation. Furthermore, heterogeneity of the protein sample should also be considered, as it is purified from a natural source and this leads to reproducibility issues. Low occupancy of PTMs is also a common aspect: only a small fraction of equal-mass peptides appears as modified. This idea is reinforced by the X-ray crystallography results, as the electron density is representative of a fraction of the molecule population in the sample being crystallized. Moreover, both chains in the asymmetric unit have a different PTM pattern and Ser316, which only seems to be phosphorylated in one chain, shows a dual conformation and phosphorylation is observed only in chain A. In addition, exposed residues are highly mobile (especially lysine residues) and this results in lack of electron density for their side chains that hampers the observation of previously described PTMs by X-ray crystallography

4.2.4. Crystallization and Structural Determination of eEF1A2:GTP and eEF1A2:GTP:Plitidepsin

4.2.4.1. Initial Approach

Due to the oncogenic properties of eEF1A2 and the previous reports on eEF1A2 binding plitidepsin in its GTP conformation (Losada, et al., 2016), we aimed for the structural determination of eEF1A2:GTP:plitidepsin ternary complex in order to characterize the binding of this didemnin B derivative.

Initially, crystallization trials were carried out by adding a non-hydrolysable GTP analogue (GppNHp) in x100 molar ratio to the protein concentration and launching commercial crystallization screens in a semi-automated procedure using robots. After scaling up several promising conditions and despite the difficulty in reproducing previously observed outcomes, the most reproducible was 0.1 M HEPES Na pH 7, 0.1M Na acetate, 24% PEG 4000. This condition yielded twinned crystalline plates, and in the best scenarios, single crystalline plates (Figure 4.34.). A blue dye (IZIT) was used to prove their protein nature. Nevertheless, these plates were extremely thin and fragile. In order to optimize these crystals, additives and detergents were assayed (Figure 4.34.), as well as lower temperatures, protein concentrations, GppNHp

concentration, seeding techniques and different crystallization set-ups (under-oil crystallization). None of the approaches led to improvement of the crystals.

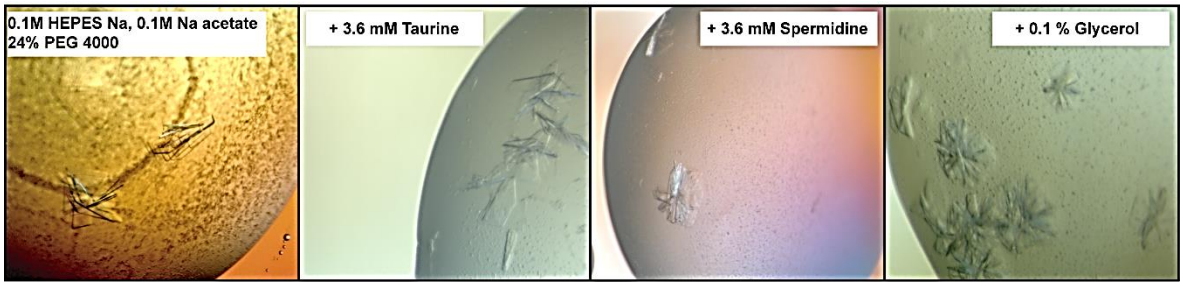


Figure 4.34.: eEF1A:GppNHp crystallization trials: Pictures were taken under polarized light. Each picture shows the outcome after optimization of the initial crystallization "hit" (left picture: 0.1M HEPES Na, 0.1M Na acetate, 24% PEG 4000) with additives. Protein concentration: 5.5 mg/mL.

Regarding structure determination, these plates were cryoprotected by supplementing the crystallization condition with 25% glycerol or increasing PEG concentration to 30%. In either cases, poor diffraction patterns were obtained (~10 Å). Indexing of these diffraction patterns was unsuccessful.

4.2.4.2. A Second Approach Based on Alkaline Phosphatase

A second approach was considered, in which the conformational variability between the possible GDP/GTP domain rearrangements, leading to sample heterogeneity which could hamper the crystallization of a single conformation, was accounted for despite the inconsistency of the protein sample (natural source, section 3.1.2). Nucleotide binding affinity was assessed at PharmaMar S.A. (binding of 3H-GDP to eEF1A2), revealing that, although GDP and GTP affinity was similar (K_D = 1 and 2.4 μM, respectively), the affinity for the non-hydrolysable GTP analogue was reduced (K_D = 23.3 μM). Moreover, HPLC analysis showed that the elongation factor isolated from rabbit muscle was exclusively in the GDP-bound conformation, which was detected at an equimolecular ratio to the eEF1A2.

Hence, to obtain the GTP-bound conformation of eEF1A2, the bound GDP was exchanged with GppNHp using calf intestinal alkaline phosphatase (AP) immobilized on agarose in the presence of high concentrations of GppNHp, and following the procedure described in (Smith and Rittinger, 2002). This method is based on the cleavage abilities of AP. AP is capable of cleaving the phosphoester bonds of GTP and GDP but is unable to hydrolyze imido- type bonds (such as GppNHp) or others that involve non-phosphate groups, such as thiophosphate (as in GTPgammaS). Thus, AP hydrolyses GDP and the resulting nucleoside (guanosine) is displaced from the binding site by GppNHp. In this way the GppNHP-bound eEF1A2 population was enriched in the sample to almost 100%.

The new sample had the buffer changed to 20 mM Tris-Hcl pH 7.5, 2 mM $MgCl_2$ and 2 mM DTT due to scarce GppNHp stability at lower pH values. The new 97% GTP-bound eEF1A2 sample was used for semi-automated preliminary crystallization trials, both in the native form and in co-crystallization assays with plitidepsin. In this case, plitidepsin (previously dissolved in 100% DMSO) was added in a final concentration x10 greater than the protein concentration evaluated for a specific experiment, and was either incubated with the protein overnight (4ºC) or directly added to the crystallization drops. Lens-like crystals were yielded in 0.1 M Na citrate pH 5.5, 0.2 M Na/K tartrate, 2 M ammonium sulfate, both in the presence and absence of plitidepsin, within 3 days (Figure 4.35.A.) Moreover, similar-looking crystals were obtained in 0.1 M HEPES Na pH 7.5, 0.2 M $(Li)_2SO_4$, 35% PEG 3350 within 2 weeks (Figure 4.35.B). Both results were obtained using a sitting-drop setup at 18ºC and were not always reproducible.

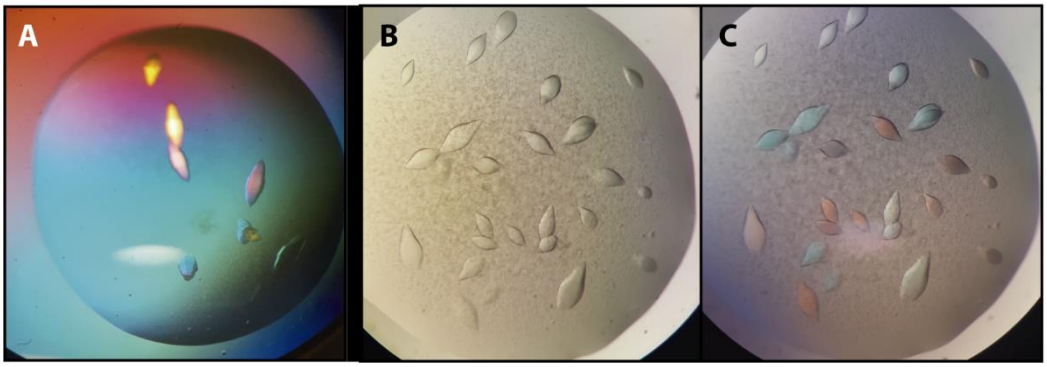


Figure 4.35.: 97% conformationally homogenous eEF1A2:GppNHp sample crystallization. (A) Crystals obtained in 0.1 M HEPES Na pH 7.5, 0.2 M (Li)2SO4, 35% PEG 3350 under polarized light. These resembled eEF1A2:GDP crystals. **(B)** Lens-like crystals obtained in 0.1 M Na citrate pH 5.5, 0.2 M Na/K tartrate, 2 M ammonium sulfate. **(C)** Drop B under polarized light. Protein concentration 6 mg/mL.

These crystals were exposed to synchrotron X-rays, having been previously cryoprotected either with saturated Li_2SO_4 or 25% glycerol according to the crystallization condition. Those crystals obtained in 35% PEG 3350 conditions diffracted to ~ 4Å, despite their light-polarizing ability. Nevertheless, indexing of the reflections revealed the same unit cell parameters and symmetry as eEF1A2:GDP crystals (a=b=134.3 Å, c= 304.5 Å; α=β= 90º, γ=120º). Lens-like crystals grown in 0.1 M Na citrate pH 5.5, 0.2 M Na/K tartrate, 2 M ammonium sulfate diffracted to ~2.7 Å. Datasets obtained were processed, showing an identical crystal composition to eEF1A2:GDP. However, electron density maps were inspected for the presence of plitidepsin or conformational changes derived from plitidepsin addition.

Thorough analysis and refinement of electron density maps revealed the presence of an extra phosphate molecule in the nucleotide-binding site of domain I of eEF1A2 (Figure 4.36.), belonging to the γ-phosphate of GppNHp. These results suggest that, although GppNHp is a non-hydrolysable GTP analogue which has been used for crystallization of other GTPases, its stability criteria are insufficient for eEF1A2 crystallization. Herein, the manufacturer strongly suggests

that "for reasons of stability, […] the pH value of a solution of this product should never drop below 7.0" and that "short term exposure (up to 1 week cumulative) to ambient temperature is possible" (JenaBioSciences). Therefore, despite the treatment with AP to obtain a conformationally homogenous sample, GppNHp hydrolyzed due to the acidic pH in the crystallization condition and/or long crystallization periods at mild temperatures, causing the "closed" GTP conformation to revert back to the known "open" GDP rearrangement.

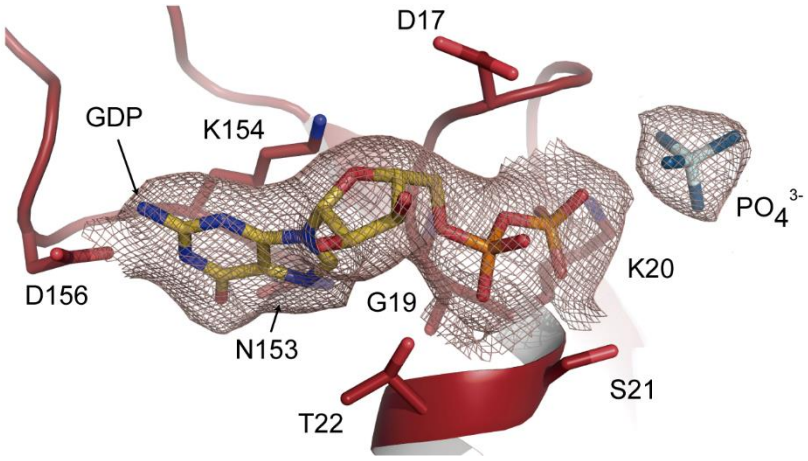


Figure 4.36.: GppNHp hydrolysis in the nucleotide binding site. Processing of diffraction data obtained from the crystallized eEF1A2 sample treated with alkaline phosphatase reveals the hydrolysis of the GppNHp nucleotide analogue within the binding site in domain I. eEF1A2 is represented as cartoon and colored red.Those residues involved in nucleotide binding are shown as sticks. GDP is shown in sticks and colored yellow. The extra phosphate molecule (absent in GDP-bound eEF1A2 crystals) is shown in stick form and colored blue. The electron density (2Fo-Fc map contoured at 0.9σ) is shown as a brown mesh.

Finally, in 2016, the structure of the mammalian ribosomal elongation complex with aminoacyl-tRNA, eEF1A, and didemnin B (PDB 5LZS) determined by electron microscopy, was published. This unveiled the binding of didemnin B to eEF1A1 at the molecular level (resolution 3.3 Å) (Figure 4.31.). Due to sequence similarity and structural conservation between eEF1A2 and eEF1A1 and the extremely close chemical structure of didemnin B and plitidepsin (Figure 1.22.), it is assumed that the binding of this cyclic depsipeptide is identical.

4.3. DECIPHERING THE MOLECULAR BASIS OF *Listeria* TEICHOIC ACID RECOGNITION BY PlyP35

4.3.1. Structural Characterization of CBDP35 in complex with *L. monocytogenes* cell-wall teichoic acid.

4.3.1.1. CBDP35 Crystallization in Complex with Teichoic Acids

CBDP35 was crystallized using the sitting drop vapor diffusion method, under the conditions mentioned in section 3.2.5. Protein and precipitant volume variations led to deviations in the nucleation events given a protein concentration of 5 mg/mL. Hence, optimization led to a final protein:precipitant drop volume ratio of 2µl:1µl and 3µl:1µl (Figure 4.37.).

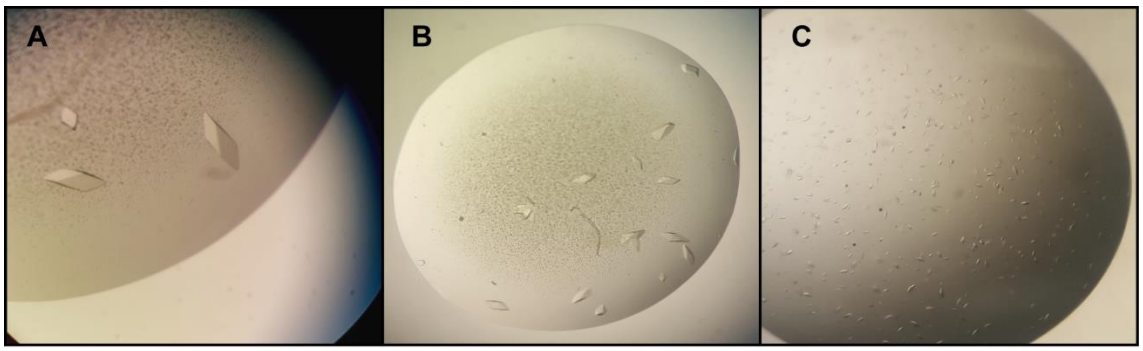


Figure 4.37.: CBDP35:TA crystals. (A and B) CBDP35 crystals grown in 4.6M sodium formate. Protein:precipitant ratio 2µl:1µl (A) and 3µl:1µl (B). **(C)** Increased nucleation events occurred when protein:precipitant ratio varied to 4µl:1µl. Protein concentration: 5mg/mL.

Prior to CBDP35:TA complex crystallization, the Se-Met derivative of the apo-CBDP35 structure had been solved in our laboratory and the CBDP35:GlcNAc complex determined (β-D-GlcNAc). These samples crystallized under similar conditions to CBDP35:TA complex and yielded crystals of similar morphology.

4.3.1.2. Structural Determination of CBDP35:TA

Diffraction datasets for these crystals were collected at beamline BL13 in the ALBA Synchrotron (Cerdanyola del Vallès, Spain). Diffraction patterns of sufficient quality for atomic structural resolution were obtained (Table 4.7.).

Diffraction patterns collected were indexed, integrated, scaled and merged using iMosflm and Aimless. The best crystals diffracted to 2.3 Å and shown to belong to the C 2 space group with unit cell parameters *a*=336.378 Å, *b*=96.288 Å, *c*= 85.056 Å; α= γ=90º, β =90.056º. However, reflections from higher resolution shells than 2.83 Å were excluded. Matthews coefficient is 3.54 A^3/Dalton, indicating twelve molecules per asymmetric unit and a 65.25 % of solvent in the crystal. Because the apo-protein had been solved in our laboratory, it used as a search model for molecular replacement using MOLREP.

Table 4.7.: Crystallographic data collection and refinement statistics for CBDP35:TA. Value for the highest resolution shell is shown in parenthesis.

Parameters	**CBDP35:TA Complex**
Data collection	
Space group	C 1 2 1
Cell dimensions	
a, *b*, *c* (Å)	336.378 96.288 85.056
α, β , γ (º)	90.000 90.056 90.000
Wavelength (Å)	0.97918
Resolution (Å)	46.83 -2.83
Total reflections	110726 (10641)
No. unique reflections	62245 (6210)
R_{pim}	0.08392 (0.4541)
$CC_{1/2}$	0.992 (0.753)
$I/\sigma(I)$	6.72 (1.72)
Completeness (%)	95.56 (96.55)
Multiplicity	1.8 (1.7)
Refinement	
Resolution range (Å)	46.83-2.83
R_{work}/R_{free}	0.222/0.256
No. atoms	
Protein	13620
Water	718
Ligand	461
R.m.s. deviations	
Bond length (Å)	0.013
Bond angles (º)	1.770
Ramachandran favored/ouliers (%)	99.94/0.06
Residues in AU	1668
Average B value overall (Å^2)	43.24
PDB code	**----**

Molecular replacement solution was refined using REFMAC (section 3.4.6.), using strong geometric restraints for refinement. Manual adjustments of TA molecules, α-D-GlcNAc moieties and solvent was carried out using COOT. Final refinement statistics were R_{work}=0.211 and R_{free}=0.270, shown in Table 4.7., after applying NCS restraints. RMS values for bond length and angles and Ramachandran outliers indicate an optimum geometry for the model obtained. Despite diffraction resolution, numerous water molecules (886) and sulfate ions were observed.

4.3.1.3. Structural Characterization of CBDP35:TA

4.3.1.3.1. Structure Overview

The CBD of PlyP35 is composed by two β-barrels, namely a proximal and distal subdomain, relative to the N and C-terminal (Figure 4.38.A.). These subdomains display a certain degree of homology, connoting gene duplication or insertion of equivalent coding sequences (Figure S3 in Supporting Information). The proximal and distal subdomains are formed by two pairs and a triplet of antiparallel β-strands framing the previously mentioned β-barrel structures.

Moreover, single 3_{10} helixes are integrated within each β-barrel (Figure 4.38.B.). However, the distal subdomain consists of a narrower barrel.

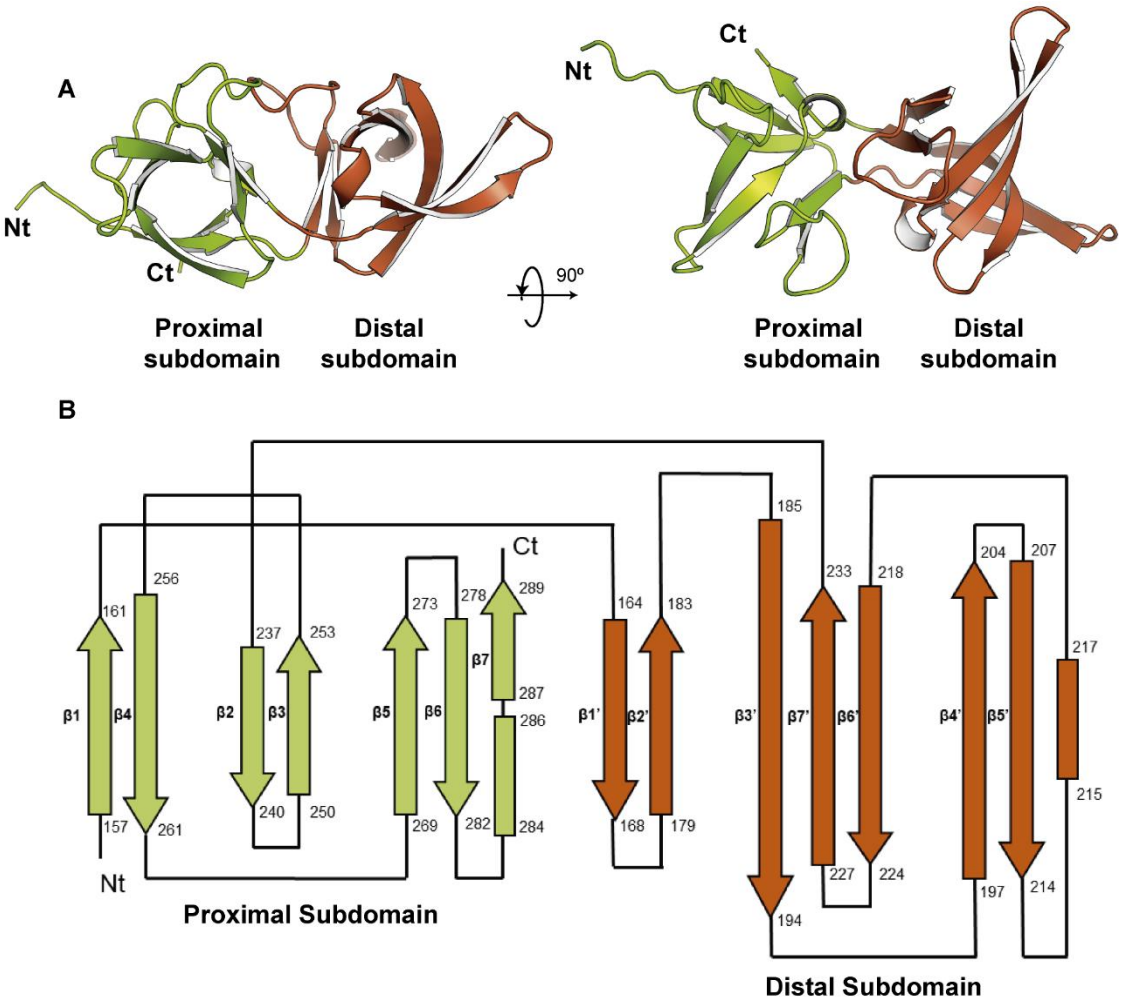


Figure 4.38.: Overall fold of CBDP35 (A) Overview of the fold of CBDP35, which is shown as cartoon and its two subdomains colored green and brown for the proximal and distal regions, respectively. **(B)** Topology diagram of CBDP35. B-laminas are represented as arrows and colored green when forming part of the proximal subdomain and brown when in the distal subdomain. 3_{10} helices are represented as a rectangle. Both β-laminas and helixes are in proportion to the amino acid length. In opposition to PlyPSA, numbering of β-sheets

Notably, CBDP35 features a high percentage of positively charged residues (17.4% of Lys and Arg) evenly distributed throughout the surface of both subdomains, and a considerable number of aromatic amino acids (total number of Tyr, Phe, Trp residues: 20 -14.5% of CBDP35). The majority of these aromatic residues lie in the proximal subdomain, although interestingly, most of them face towards a cleft formed between both proximal and distal subdomains. In this way, hydrophobic interactions predominate in the maintenance of the double β-barrel fold, with only Lys162-Gln181 and Gln168-Tyr285 are involved in polar interactions.

This exact conformation of CBDP35 is observed in all Se-Met-CBDP35, CBDP35:GlcNAc and CBDP35:TA complex structures.

4.3.1.3.2. Crystal Packing

CBDP35:TA complex has crystallized in the monoclinic space group C 2. As mentioned previously, twelve molecules are identified in the asymmetric unit. Large solvent channels are observed (Figure 4.39.), allowing smaller molecules to diffuse through the protein crystal during soaking experiments. The crystal contacts vary from one chain to another as each chain has a different orientation with respect to its neighboring molecule. Hence, chains may interact trough their proximal domains, through the C-terminal of one molecule with β6'-β7' of another, or through their β3' laminas of the distal subdomain. This will limit the exposure of each TA binding site to the solvent or enhance the stability of the TA molecules within the pockets to which they bind.

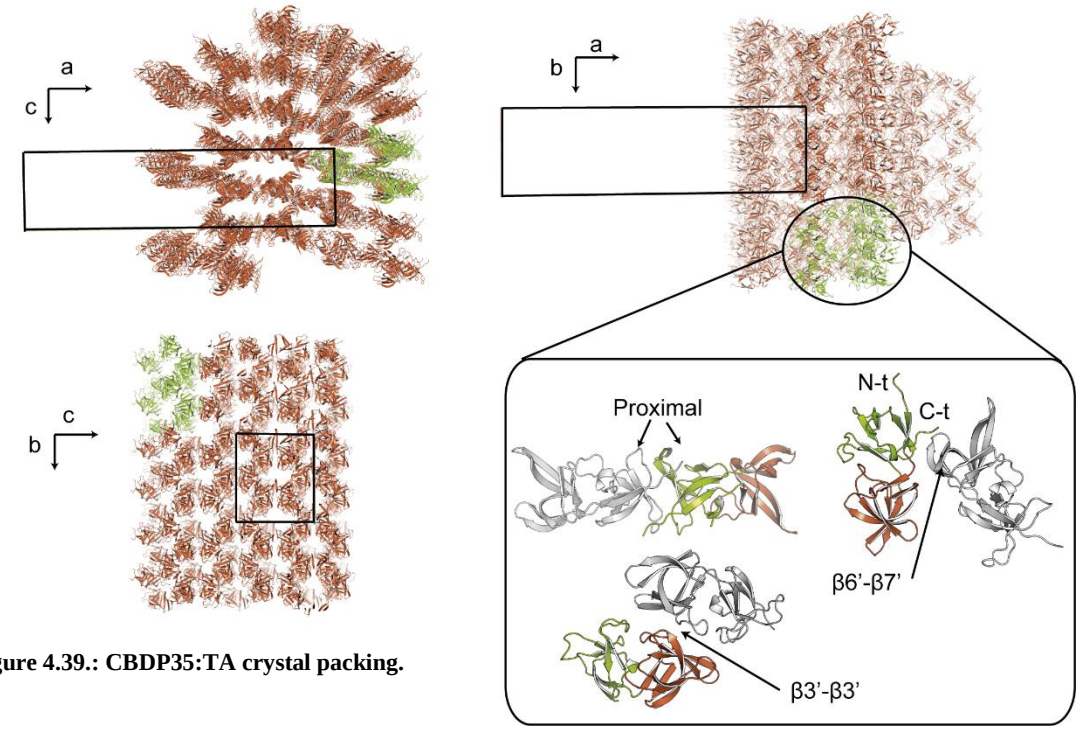


Figure 4.39.: CBDP35:TA crystal packing.

4.3.1.3.3. Apo-CBDP35 vs. PlyPSA

The overall fold of CBDP35 (described in section 4.3.1.3.1.) has been only previously observed in the CBD of PlyPSA, an N-acetyl muramoyl-L-alanine amidase encoded by the *Listeria* phage PSA (PDB 1XOV, Figure 4.40.). Herein, despite the low sequence identity between PlyPSA's CBD and PlyP35's CBD (20%), both domains can be superimposed with an RMSD of 1.46Å. In PlyPSA's CBD, β1 and β7' are structurally swapped (Korndörfer, *et al.*, 2006). On the contrary, when proximal and distal subdomains within CBDP35 are sequence-wise superimposed, this is not observed, as these β-strands do not show sequence homology.

Although the proximal and distal subdomain arrangement is conserved in the CBD of P35 and PSA endolysins, certain dissimilarities exist (Figure 4.40.) Different loop conformations are

observed, such as in residues 169-178 in CBDP35, which connect β1' and β2' of the distal subdomain, and the linker loop connecting the distal to the proximal subdomain (residues 234-236 in CBDP35), the latter region being 4 residues longer in PlyPSA than in CBDP35.

Variations can also be observed in the secondary structure of both CBDs. Overall, β laminas comprising the distal subdomain are shorter in PlyPSA than in CBDP35. In particular, β6' and β7' consist of 7 residues in CBDP35 and are connected by a 4 residue loop (residues 224-227), whereas the equivalent in PlyPSA are comprised of 3 and 5 amino acids respectively. Thus, this region is weakly stabilized in PlyPSA as it unfolds into a loop. Same scenario, is observed for β4' and β5', which are two residues shorter in PlyPSA than CBDP35.

Regarding the proximal subdomain, the antiparallel β1-β4 couple is stabilized at a different angle in CBDP35 than in PlyPSA, causing variation on the loop connecting β4 with β5 (residues 262-268 in CBDP35). This change is propagated through to the β5-β6 pair (Figure 4.40.).

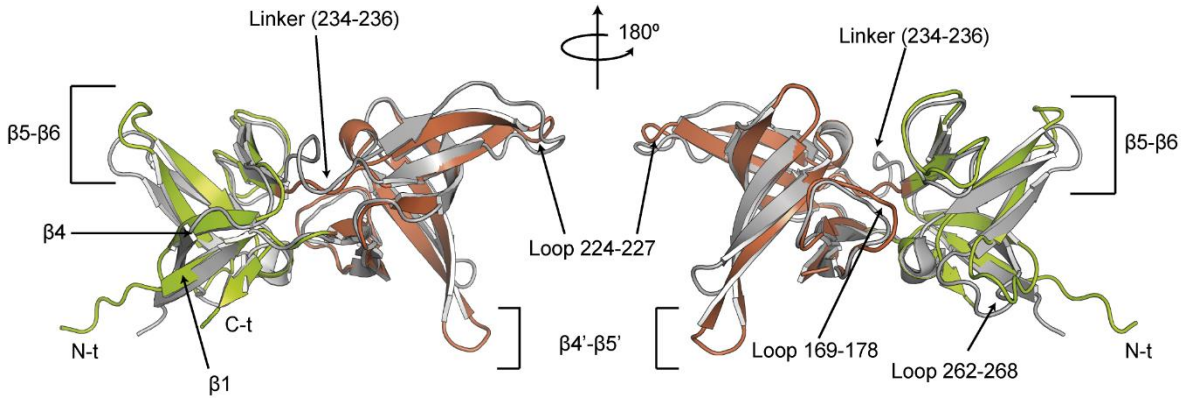


Figure 4.40.: Structural differences of CBDP35 with the CBD of PlyPSA. Both CBDs are shown as cartoon. PlyPSA is colored in grey and CBDP35 green and brown, respectively to the proximal and distal subdomains. Regions showing structural differences are labelled and numbered according to CBDP35 numbering.

In addition, a search with DALI server identified structural similarities with SH3 and SH3b-like domains in phage and non-phage cell wall-associated proteins (PDBs: 2KT8, 5UDM, 4LXC, 5BND, 3H41) in the case of the proximal subdomain.

4.3.1.3.4. N-acetyl-β-D-glucosamine Binding Sites

The crystal structure of apo-CBDP35 and in complex with β-D-GlcNAc had been previously obtained in our laboratory. Although the CBDP35:GlcNAc complex structure showed no conformational changes with respect to the apo-CBDP35 structure, it revealed two GlcNAc binding sites, one situated in the terminal surface of the distal subdomain and the other in a pocket localized at the interface between the proximal and distal subdomains (Figure 4.41.). Both binding sites have a certain hydrophobic character, although they are strongly defined by their positive

charges framing the pockets (Figure 4.41.). For each site, a single molecule of β-D-GlcNAc was bound in all 12 molecules of the asymmetric unit.

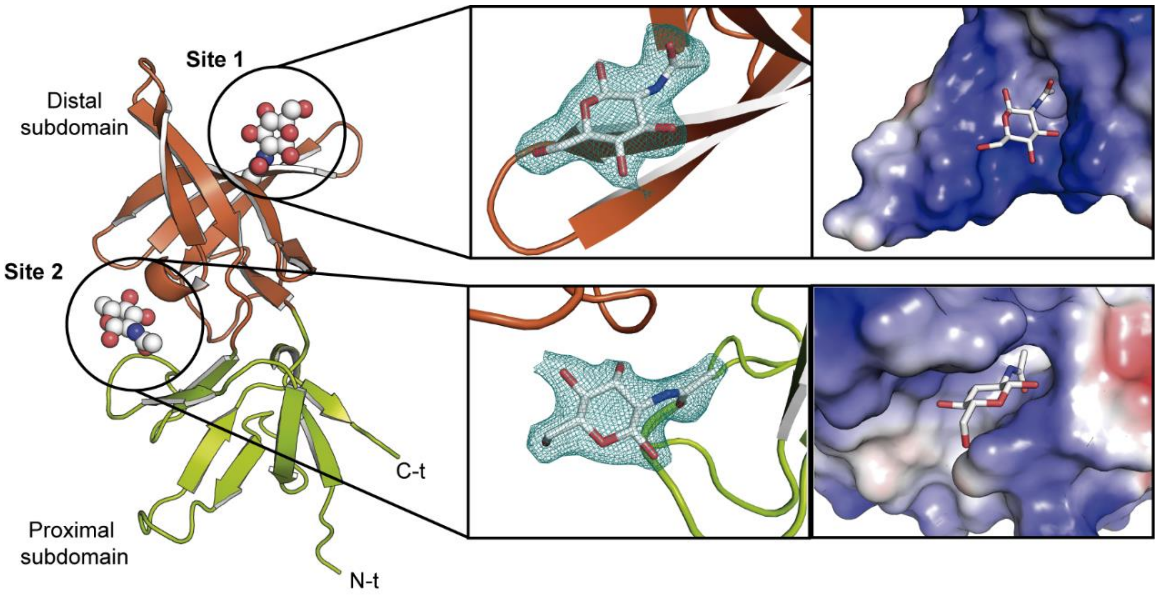


Figure 4.41.: β-D-GlcNAc binding sites in CBDP35. The panel on the left shows CBDP35 as cartoon and colored green and brown according to its subdomains. β-D-GlcNAc is shown in spheres. The right panel shows the electron density (2Fo-Fc) observed for β-D-GlcNAc contoured at 1σ and the electrostatic potential surface for both pockets to which GlcNAc binds (white sticks).

Site 1 is comprised of an exposed surface that involves residues of only the distal subdomain. These residues lie on strands β2' (Trp182), β3' (Lys189) and β4' (Val201) (Figure 4.42.A.). Herein, the main chain atoms of Lys189 and Val201 hydrogen bond β-D-GlcNAc in all CBDP35 molecules through its -$NHCOCH_3$ substituent at C5. However, the side chain of Lys189 is eventually engaged in binding of O4 of β-D-GlcNAc. Water molecules mediate the stabilization of this sugar, involving other residues such as Trp182. These observations denote a weak stabilization of β-D-GlcNAc, as no side chains are implicated in the binding of this sugar.

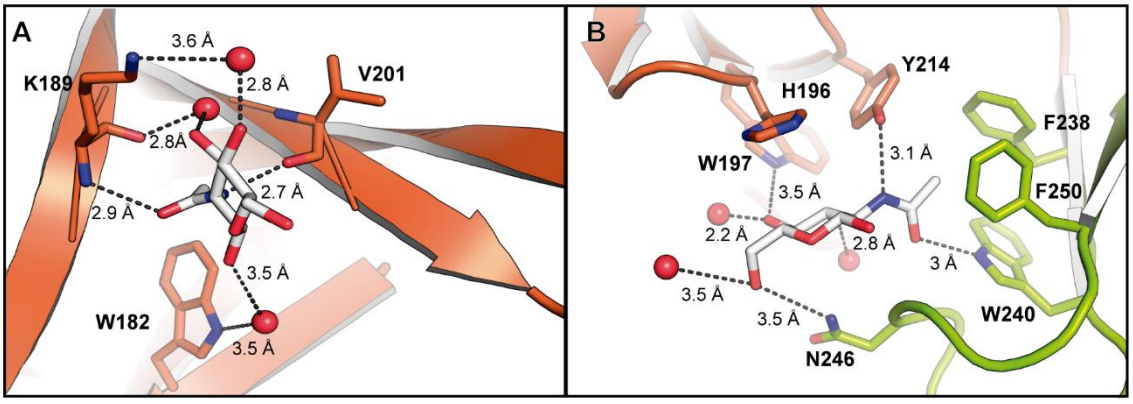


Figure 4.42.: Interactions of β-D-GlcNAc in binding site (A) 1 and 2 (B). Distal and proximal subdomains are colored brown and green respectively. Residues involved in β-D-GlcNAc binding are portrayed as sticks. Hydrogen bonds are represented as black dashes. Distances and residues are labelled accordingly. Water molecules are shown as red spheres.

Site 2 involves an enclosed cavity in which different types of interactions with β-D-GlcNAc are depicted (Figure 4.42.B.). Side chains of residues from both the distal and proximal subdomains are implicated by means of electrostatic/polar interactions (Trp197 and Tyr214 in β4' and β5' of the distal subdomain, respectively; Trp240 in β2 and Asn246 in the loop connecting β2 and β3 of the proximal subdomain). Moreover, hydrophobic interactions are observed between the $–CH_3$ of the $-NHCOCH_3$ substituent at C5 of the β-D-GlcNAc ring and Phe238 and Phe247, as well as stacking interactions with His196. Just as in binding site 1, water molecules also contribute in the stabilization of the sugar.

The residues above mentioned are equally positioned in both the apo and complex structures, showing no reorientation upon β-D-GlcNAc binding. Furthermore, three water molecules mediate the stabilization of β-D-GlcNAc in each binding site. Details and distances are given in Figure 4.42.

4.3.1.3.5. Teichoic Acid Binding Sites

Our colleagues at the Department of Health Sciences and Technology at the ETH in Zürich purified listerial peptidoglycan to obtain monomeric teichoic acids (one repeating unit, section 1.4.3.) from *L. monocytogenes* serovar 1/2a. The monomeric form consists of one RboP molecule decorated with one α-D-GlcNAc and one α-L-Rha molecule in its C2 and C4 positions, respectively (Figure 4.43.).

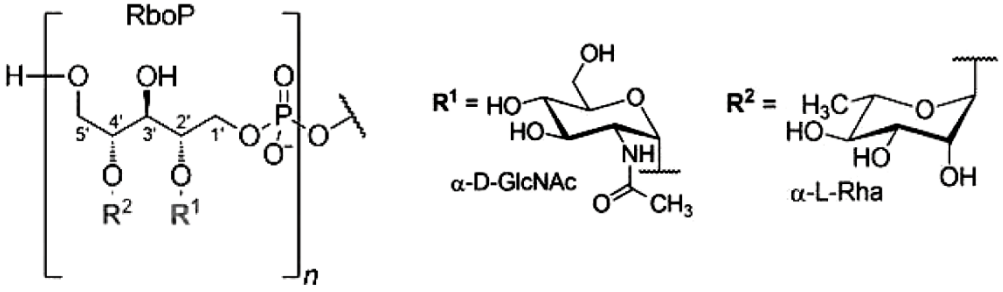


Figure 4.43.: Teichoic acid scheme from *L. monocytogenes* serovar 1/2a.

As mentioned in section 4.3.1.3.2., twelve molecules compose the asymmetric unit of CBDP35 crystals. In the case of CBDP35:TA complex and contrary to CBDP:GlcNAc crystals, not all the twelve molecules displayed equal densities in their binding sites. Three different scenarios were considered attending to density continuity and morphology: either (i) a full TA molecule (for which occupancy was refined) was present bound to the previously determined pockets (site 1); (ii) only the α-D-GlcNAc moiety was observed, implying mobility of the TA monomer or (iii) a SO_4^{2-} ion from the cryoprotectant solution was found in the binding site (Figure 4.44.).

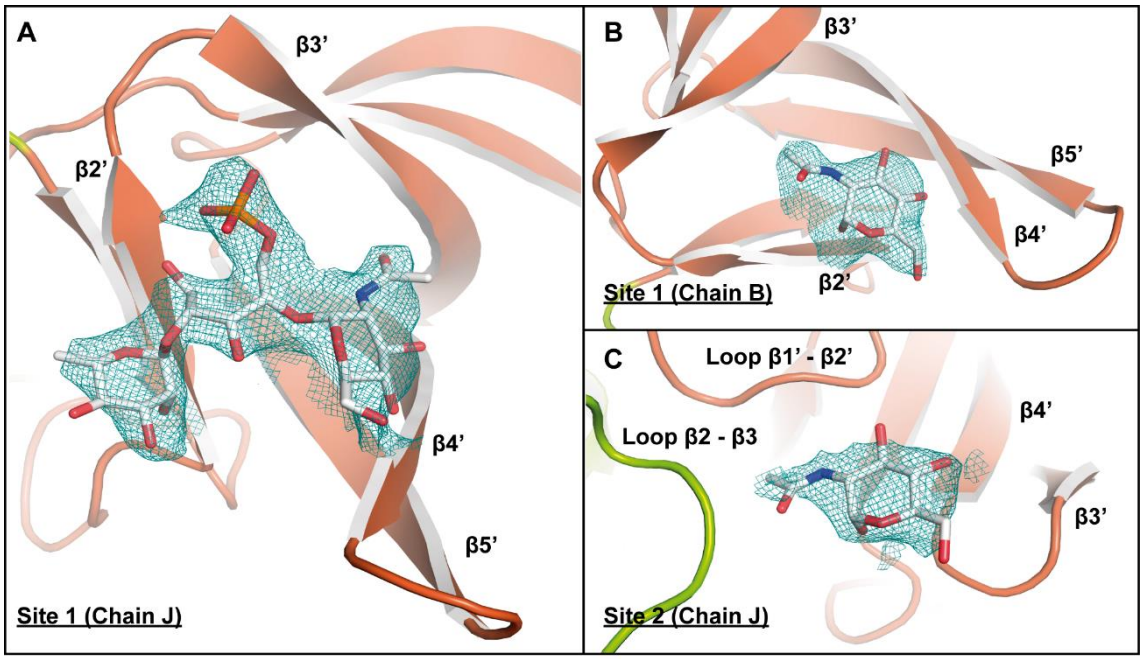


Figure 4.44.: Electron density of the teichoic acid monomer and α-D-GlcNAc moiety for binding sites 1 and 2. CBDP35 is shown as cartoon, proximal subdomain colored in green and distal subdomain in brown. Relevant residues (discussed below) are depicted as sticks. TA monomer and α-D-GlcNAc are shown as white sticks. Green grid represents the 2Fo.Fc map. Both site 1 and site 2 show differences in electron density depending on the chain, and so the TA or α-D-GlcNAc molecules were positioned according to density continuity or morphology. **(A)** Chain J, 2Fo-Fc map contoured at 0.8σ. **(B)** Chain B, 2Fo-Fc map contoured at 0.8σ. **(C)** Chain J, 2Fo-Fc, contoured at 0.75σ.

· Binding Site 1

Binding site 1 is a wide and solvent-exposed cavity comprised by strands β2', β3' and β4' of the distal subdomain. TAs bound to site 1 do not appear tightly anchored to the pocket, and so variations in the pose of the TA (or α-D-GlcNAc moiety) attached are witnessed. However, although certain flexibility of the monomeric form of the TA is allowed, some residues framing the pocket seem to be crucial for its binding (Figure 4.45.). Continuous electron density for the full TA molecule was only observed in binding site 1 of chains D and J (Figures 4.44 and 4.45.).

The phosphate moiety of the monomeric TA pivots between β2' and β3' of the distal subdomain, showing a displacement of ~ 2.3 Å between the TA of chains D and J. This phosphate group engages the main chain of Lys183 and side chain of Thr186 (chain D) for its stabilization by means of hydrogen bonds (Figure 4.45).

The α-D-GlcNAc moiety of the monomeric TA faces the β4' strand. Along with chains D and J, the α-D-GlcNAc sugar alone can be observed for chains A, B, C, G, H and I (Figure 4.46.). In chain D and J, this moiety of the monomeric TA is stabilized by polar interactions with the main chain of Lys203 (Figure 4.45. and 4.46.) through its O4 and O5 atom. This α-D-GlcNAc substituent also engages Arg180 (chains B, C, H and I), which forms a network of interactions with Asp204 and Val179 in order to maintain the β-strand disposition and the pocket

conformation. Arg180 also interacts with O5, the oxygen atom of α-D-GlcNAc ring (chains A, C, Figure 4.46.), stabilizing the disposition of this sugar in binding site 1. Moreover, the -$NHCOCH_3$ substituent at C2 of the α-D-GlcNAc ring is also shown oriented towards β3' of the distal subdomain. Hence, hydrogen-bonding with residues in this strand for its stabilization involve the main chain atoms Lys187.

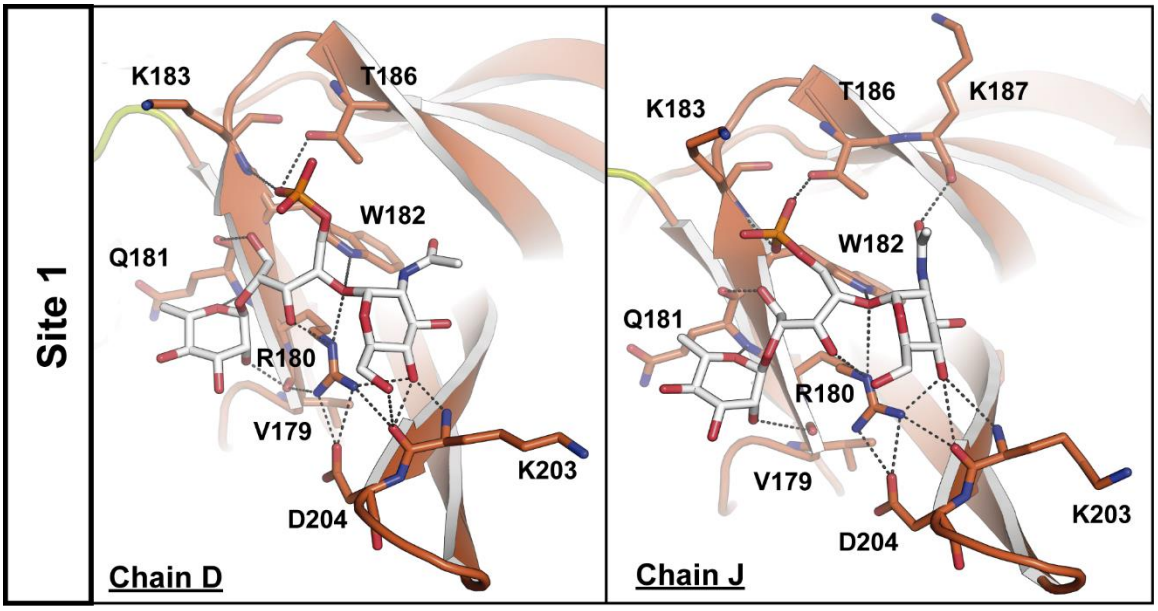


Figure 4.45.: Residues involved in the binding of the monomeric TA molecule in binding site 1. CBDP35 is shown as cartoon and colored brown (all residues belonging to the distal subdomain). Residues involved in TA binding are shown as sticks and labelled accordingly. TA molecule is shown as white sticks. Polar interactions/hydrogen bonds/salt bridges are shown as dashed lines. Those residues also shown as sticks, such as Val202, with no depicted interactions may be involved in Van der Waals interactions. Water molecules are represented by red spheres. Range of distances considered: 2.4-3.5 Å.

Arg180 and Trp182 are crucial for the recognition of the RboP repeating unit of the TA monomer. Hence, the side chain of the previously mentioned residues (Nε of Arg180 and the –NH group of Trp182) interact with O1 of α-D-GlcNAc, resultant of the glycosidic bond linking α-D-GlcNAc to the RboP molecule, and so ensure the positioning and stabilization of the TA within the cavity. Moreover, the RboP sugar makes polar interactions through O3 and O5 with Arg180 and Gln181, respectively, leading to further mobility due to the lack of strong and defined interactions.

The Rha sugar is exposed to the solvent (Figure 4.45.). In our complex, O of the Rha moiety hydrogen bonds the main chain of Gln181 in both chains D and J. Furthermore, the main chain of Val179 is engaged in the weak stabilization of the sugar by means of hydrogen bonding O2 of the Rha moiety. However, just as the RboP backbone, this sugar is weakly stabilized, suggesting mobility.

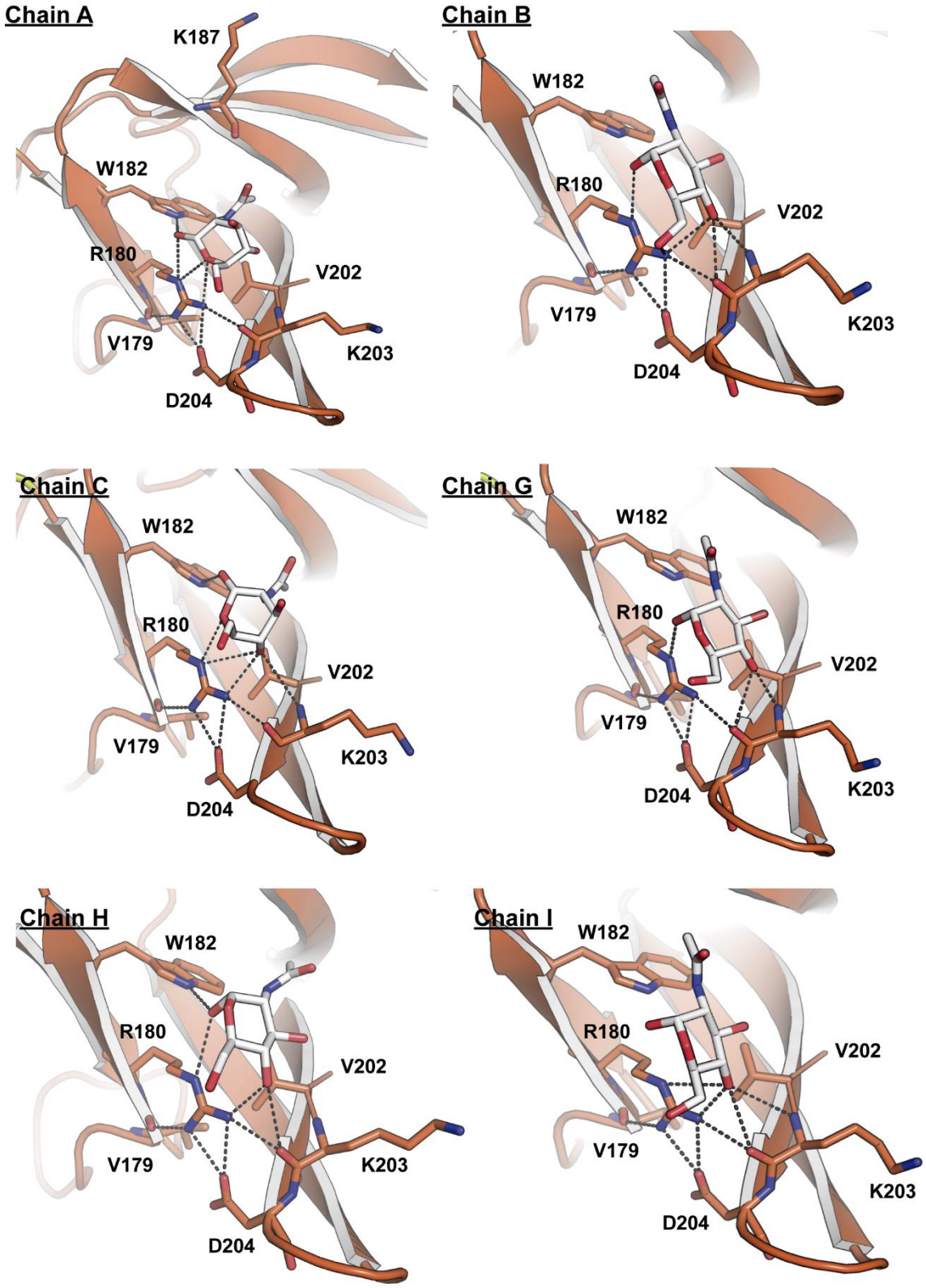


Figure 4.46.: α-D-GlcNAc binding in site 1. CBDP is shown as cartoon and colored brown (all residues belonging to the distal subdomain). Residues involved in α-D-GlcNAc binding are shown as sticks and labelled accordingly. α-D-GlcNAc molecule is shown as white sticks. Interactions are shown as dashed lines. Chains E, F, K and L have a SO_4^{2-} ion in their binding site and are not shown. Range of distances considered: 2.2-3.5 Å. Those residues which may participate in hydrogen-bond formation but lie slightly beyond 3.5Å (i.e. Lys187 in chain A and Trp182 in chain I) are also shown.

It is worth mentioning that, although the orientation of the α-D-GlcNAc moiety is conserved, the chair conformation is flipped to properly fit the electron density observed in each cavity. Furthermore, the protein structure does not change upon TA binding with respect to apo-CBDP35, other than for side chain readjustments.

· Binding Site 2

No chains showed sufficient electron density definition and continuity to allow the placement of the complete monomeric TA molecule in binding site 2. However, adequate electron density was observed for the α-D-GlcNAc moiety in chains B, H, J and L (Figure 4.44. and Figure 4.47.), implying the exposure to the solvent and mobility of the RboP and α-Rha sugars. Each of these chains was analyzed regarding α-D-GlcNAc moiety interactions in binding site 2.

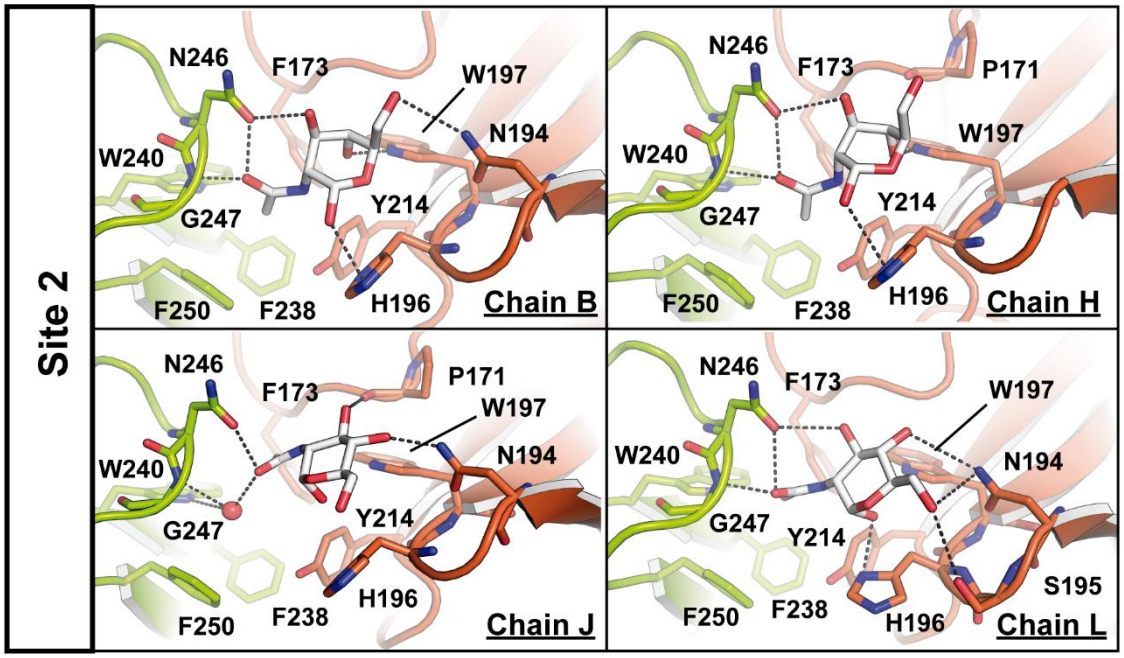


Figure 4.47.: α-D-GlcNAc binding in site 2. CBDP is shown as cartoon and colored brown (all residues belonging to the distal subdomain). Residues involved in α-D-GlcNAc binding are shown as sticks and labelled accordingly. α-D-GlcNAc molecule is shown as white sticks. Interactions are shown as dashed lines. Chains E, F, K and L have a SO_4^{2-} ion in their binding site and are not shown. Range of distances considered: 2.2-3.5 Å.

Just as in binding site 1, α-D-GlcNAc showed a certain degree of variability in its positioning depending on the CBDP35 chain (ring inversion). However, in binding site 2, some residues with which the α-D-GlcNAc sugar interacts remain consistent throughout all protein molecules. In general terms, the $–NHCOCH_3$ substituent at C2 is oriented towards the loop connecting β2 and β3 of the proximal subdomain, whilst the C3 and C4 substituents face the β1'-β2' loop and β3 and β4 strands of the distal subdomain (Figure 4.44 and Figure 4.47.), therefore engaging residues from both the proximal and distal subdomains.

Herein, α-D-GlcNAc moiety interacts with Asn246 and Gly247 through its acetyl group at C2 by means of hydrogen bonds, either directly or by a bridging water molecule (chain J, Figure 4.47.). Moreover, Asn246 may also hydrogen bond the hydroxyl group at C3 of the α-D-GlcNAc sugar, which may also interact with Pro171 depending on how the ring is tilted. In the case of O4, it may engage either Asn194 or Trp197, the former also participating in hydrogen bonding the hydroxyl group at C5 along with Ser195 (chain L, Figure 4.47.).

Regarding O1 (which should result in the glycosidic bond between RboP and α-D-GlcNAc), it is oriented away from the binding cavity and interacts with His196. Furthermore, the distance between the sugar moiety (ring) and His196 (~4 Å) reveals possible stacking interactions with this residue. Hence, His196 plays a similar role as Arg180 of binding site 1 in determining the disposition of α-D-GlcNAc within the cavity. Other residues lining this cavity that may also contribute to the stabilization of the sugar and confer its hydrophobic character involve Phe173, Tyr214, Phe238 and Phe248 (Figure 4.47.).

As in binding site 1, ligand binding does not alter the general fold of CBDP35, although side chain reorientations can be observed among different CBDP35 chains of the same crystal structure, as well as between apo-CBDP35.

4.3.2. Further Considerations

4.3.2.1. TA Recognition and Specificity is Determined by the Distal Subdomain

As mentioned in section 4.3.1.3.1., a DALI search identified structural homology only for the proximal subdomain of CBDP35 with SH3 and SH3b-like domains in phage and non-phage cell wall-associated proteins. Hence, TA recognition and specificity of P35 binding to *L. monocytogenes* TA is given by the distal subdomain.

This is coherent with our results, in which two TA binding sites are determined: site 1, which is only comprised by residues of the distal subdomain, and site 2, which involves residues of both the proximal and distal subdomains.

4.3.2.1.1. TA Binding Sites in CBDP35 and PlyPSA

PlyPSA, the only DALI hit which shows the same double barrel fold as CBDP35 (Figure 4.48.A.), has been shown to bind to an evenly distributed ligand on the surface of *L. monocytogenes* serovar 4 cells by epifluorescence microscopy (Korndöfer, *et al.* 2006). Here, a putative binding region for an oligosaccharide ligand was identified, which was confirmed by a prediction of energetically favorable binding sites. In fact, this region corresponds to our binding site 2 of CBDP35.

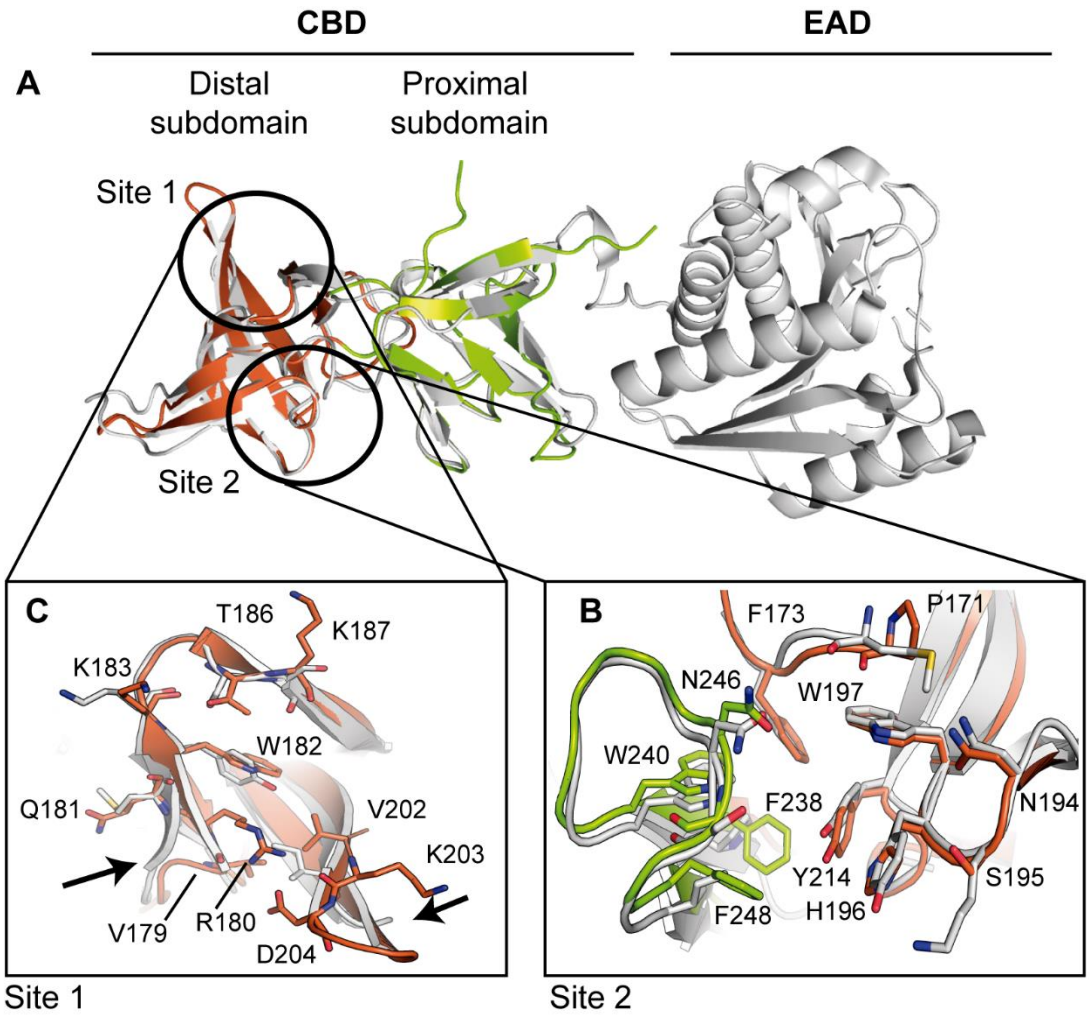


Figure 4.48.: TA binding sites of PSA and P35 CBDs. (A) CBDP35 and PlyPSA are represented as cartoon and structurally superposed. CBDP35 is colored by subdomains and PlyPSA is colored grey. TA binding sites identified in CBDP35 are encircled and zoomed in. Residues involved in TA binding are shown as sticks, as well as the equivalent in PlyPSA, and labelled according to CBDP35 numbering. **(B)** Binding site 1 identified in CBDP35. β2' and β4' are longer and stabilized by a salt bridge in CBDP35 (R180-D204), determining the binding site conformation and key region for TA recognition. Conformational changes between CBDP35 and PlyPSA are indicated by arrows. **(C)** The fold and most residues in binding site 2 are conserved in PlyPSA.

The putative PlyPSA residues reported prone to TA interaction were Phe200, Try222, Trp223, Tyr236, Phe239 and Trp279 (Korndöfer, *et al.* 2006), now associated to our binding site 2. The equivalent residues in CBDP35 are Phe173, His196, Trp197, Tyr214, Met217 and Asn252. From these CBDP35 residues, our results confirm the implication of His196, Trp197 and Tyr214 in the binding of the α-D-GlcNAc decoration of *L. monocytogenes* TA. Moreover, key CBDP35 residues that have been here determined to be involved in TA binding were analyzed for PlyPSA (Figure 4.48.B.).

Residues in binding site 2 which have been shown to hydrogen-bond GlcNAc by means of their side chains (Figure 4.48.B.) are either (i) conserved in PlyPSA (CBDP35: Asn194, Trp240, Asn246; PlyPSA: Asn220, Trp267, Asn273, respectively), or (ii) have side chains which

can also form hydrogen bonds with GlcNAc (Ser195, His196 in CBDP35; Lys221, Tyr222 in PlyPSA). Moreover, non-conserved residues such as Pro171 (Met198 in PlyPSA) have been shown to bind α-D-GlcNAc by means of their main chain. Furthermore, it is worth mentioning that not only the residues are highly conserved, but the overall fold of binding site 2 is identical for both CBDP35 and PlyPSA.

A different scenario is observed in the case of binding site 1. This binding region was not predicted for PlyPSA. Structural superposition and analysis of TA interacting residues of site 1 show pronounced diversity in comparison to binding site 2 (Figure 4.48.C.). Although the stabilization of TA in site 1 has been shown to be partly mediated by main chain atoms of CBDP35 (Figure 4.45. and 4.46., section 4.3.1.3.5.), those residues which interact by means of their side chains (Arg180, Trp182 and Thr186 in CBDP35; Lys206, Tyr208 and Ser212 in PlyPSA) are different in PlyPSA, but still capable of forming polar interactions with the TA molecule. However, a conformational change can be observed in the binding sites 1 (Figure 4.48.C.) between CBDP35 and PlyPSA.

Binding site 1 is framed by β2', β3' and β4' of the distal subdomain. In CBDP35, β4' and β5' are two residues longer each than in PlyPSA (Figure 4.40.). One of them is Asp204, which forms a salt bridge with Arg180 of β2' (Figure 4.48.C.). This interaction fixes the conformation between β2' and β4' in order to maintain a tight packing between these β-strands to build a "platform" to which the TA molecule clasps (Figure 4.49.). PlyPSA lacks this interaction, resulting in a larger separation between its β2' and β4' strands (Figure 4.48.C.), thus widening and deforming the protrusion forming binding site 1. Furthermore, Arg180 and Asp204 are involved in RboP recognition and positioning of α-D-GlcNAc within binding site 1 of CBDP35.

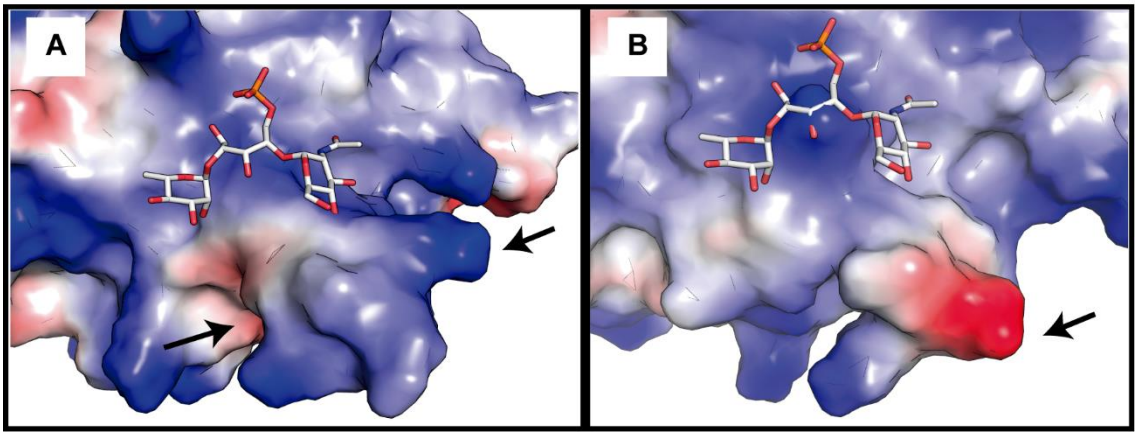


Figure 4.49.: TA binding sites 1 in CBDP35 and PlyPSA. The surface electrostatic potential of CBDP35 binding site 1 **(A)** and the corresponding region of PlyPSA **(B)** is shown. CBDP35:TA and PlyPSA structures have been superposed in order to localize the TA molecule (white sticks) in the PlyPSA structure.The CBDP35 has a bulkier and positively-charged protrusion corresponding to the longer β4' and β5' laminas of the distal subdomain (marked by arrows) which the TA molecule clasps. PlyPSA does not show this bulge and presents a negatively charged region which is missing in CBDP35.

These findings suggest that PlyPSA could also recognize GlcNAc moieties due to the similarity of its putative binding site to binding site 2 of CBDP35. Herein, *L. monocytogenes* serovar 4 cells have type II TAs (Figure 1.28, section 1.4.3.). These serovars incorporate the β-D-GlcNAc moiety within the polymer chain, which may be further decorated with other sugars (Glc, Gal, α-D-GlcNAc). Hence, the ability of the interface region between the proximal and distal subdomains to bind α-D-GlcNAc would explain the observations made by Korndöfer, *et al.*, in the above mentioned epifluorescence microscopy experiments.

4.3.2.2. Crystal Packing May Affect TA Stability Within the Binding Sites

TA binding site 1 showed the largest differences in TA binding among CBDP35 chains in the asymmetric unit. While binding site 1 is exposed in all 12 chains, the environment is not the same for each as a consequence of the crystal packing (Figure 4.39.). We thus analyzed differences in crystal environment in order to see its potential effect on the observed TA binding. Binding site 1 of most CBDP35 chains in the asymmetric unit is oriented towards the inter-subdomain region and the C-terminal region of neighboring protein molecules (Figure 4.39.). In particular, Lys219' (in β6') and Lys291' (C-terminal), both belonging to nearby CBDP35 chains, may be involved in the stabilization of the TA molecule. Two different scenarios are illustrated in Figure 4.50., in which either (i) O5 of RboP or O4 of the Rha moiety interact with Lys219' (Figure 4.50.A.); (ii) the TA molecule interacts with Lys219' through the α-D-GlcNAc moiety (Figure 4.50.B.).

Interaction with Lys219' may be considered a source of variation in the RboP conformation shown in our crystal (Figure 4.50.). In chain D, Lys219' interacts with RboP (Figure 4.50.A.), repositioning O5 of the TA. Furthermore, as shown in chain I (Figure 4.46., 4.50.B.), Lys291' of neighboring molecule E interacts with α-D-GlcNAc through its C2 substituent. This sugar is shown pivoted in comparison to other α-D-GlcNAc molecules observed in other chains Hence, electrostatic interactions between Lys291' and RboP and α-D-GlcNAc may also affect the placement of the phosphate and sugar moiety.

All in all, the interactions here shown are only bi-products of crystallization due to the crystal packing for CBDP35 molecules, therefore implying different environments in binding site 1 that may explain the observed variability and lack of electron density in this site.

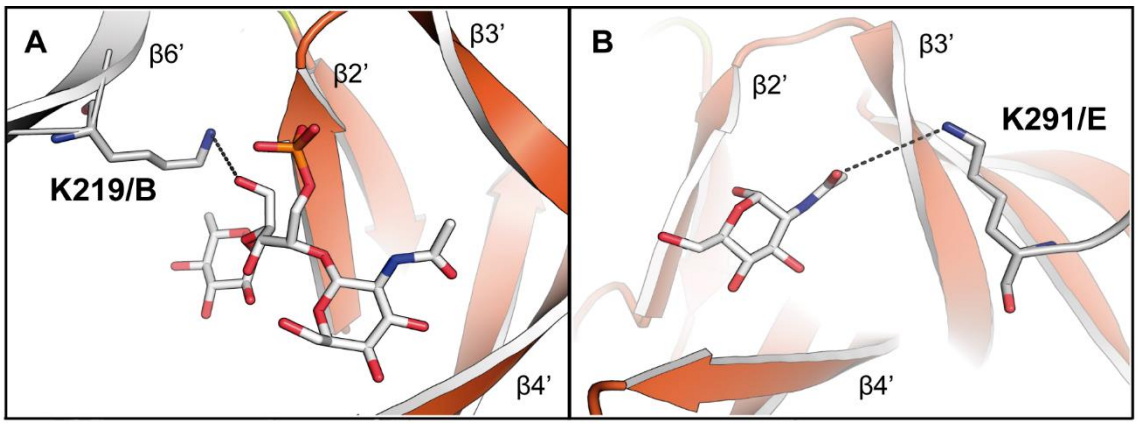


Figure 4.50.: TA in binding site 1 interacts with neighboring chains in the crystal. CBDP35 is shown as cartoon and the distal subdomain colored brown. TA molecule is shown as sticks. Interactions with other molecules within the asymmetric unit (colored grey and shown as carton) are shown by dashed lines and the interacting residues are labelled. **(A)** TA in binding site 1 of chain D is hydrogen bound to Lys219 of chain B. **(B)** α-D-GlcNAc in binding site 1 of chain I contacts Lys291 of chain E. **(C)** TA in binding site 1 of chain J can potentially interact with the C-terminal of chain F and Lys219 of chain H. **(D)** Diverse disposition of the TA's Rha and RboP moieties in binding sites 1 of chains D (white sticks) I (blue sticks) and J (grey sticks) may be influenced by interactions with the neighboring CBDP35 molecules within the crystal.

Regarding binding site 2, crystal packing of CBDP35 molecules, results in one binding site 2 of a molecule facing the binding site 2 of another (Figure 4.51.). This occurs for every pair of chains in the crystal. Although there is sufficient density for the α-D-GlcNAc moiety in binding site 2, no continuous density was observed allowing the positioning of the RboP and Rha sugars. This implies internal disorder within the crystal. Furthermore, a symmetry axis relates these confronted CBDP35 chains, hampering the placement of the TA molecule.

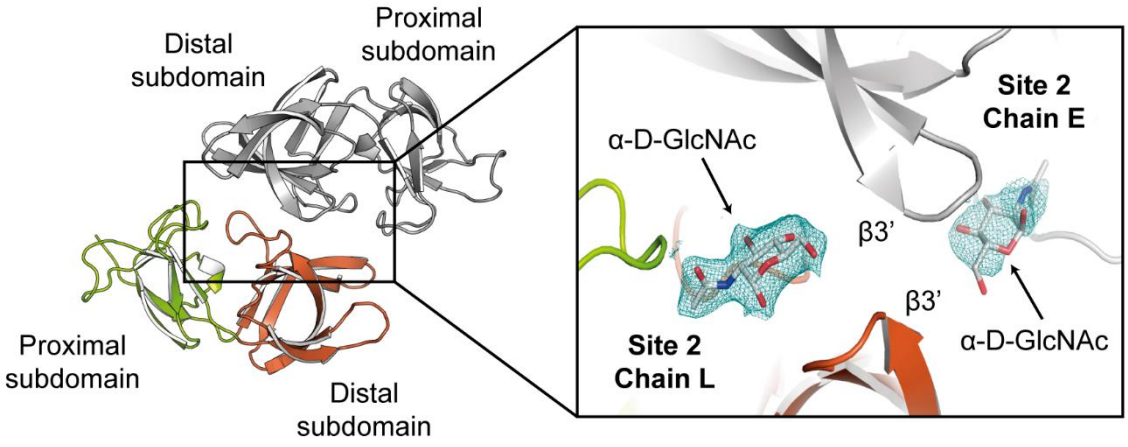


Figure 4.51.: Binding site 2 layout within the crystal. CBDP35 molecules (shown in cartoon) are packed in a way that β3' laminas of different CBDP35 chains are confronted, relating binding sites 2 by internal symmetry. Mobility of the TA molecule, low resolution and the presence of symmetry axis hamper the visualization of the complete TA molecule in binding site 2. CBDP35 is colored by subdomains (proximal subdomain-green; distal subdomain-brown) and a symmetry related molecule is colored grey. α-D-GlcNAc is shown in sticks and the 2Fo-Fc map is shown as a green mesh, contoured at 0.7σ.

4.3.2.3. Monomeric TA vs. β-D-GlcNAc: Different Molecules, Different Interactions

Regarding CBDP35:GlcNAc and CBDP35:TA complexes, differences are observed in the positioning of GlcNAc in binding in sites 1 and 2. In site 1, the α-D-GlcNAc sugar of the TA monomer is displaced an average of 6.5 Å away from the β3'-β4' pocket β-D-GlcNAc binds (Figure 4.52.) and tilted 10-20º with respect to β-D-GlcNAc alone. The CBDP35:GlcNAc complex reveals residues Trp182, Lys189 and Val201 to be involved in β-D-GlcNAc binding in site 1 (Figure 4.41.). Except for Lys189 and Val201, these residues have also been shown to interact with the monomeric TA (Figure 4.45. and 4.46.), although engaging different oxygen atoms to β-D-GlcNAc only. Hence, Trp182 is involved in the stabilization of β-D-GlcNAc by means of water molecules in CBDP35:GlcNAc (Figure 4.42.), whilst in the CBDP35:TA complex, it interacts with O1 resultant from the glycosidic linkage between the RboP and α-D-GlcNAc (Figure 4.45 and 4.46.), determining the disposition of the TA molecule within the binding site along with Arg180.

This displacement of the TA monomer with respect to the β-D-GlcNAc molecule is given by interactions between the RhoP and CBDP35. Hence, the TA molecule forms new interactions with residues in the β2', β3' and β4' strands, which are absent in the CBDP35:GlcNAc complex stabilization, fixing the TA molecule (and therefore the α-D-GlcNAc moiety) in a new position.

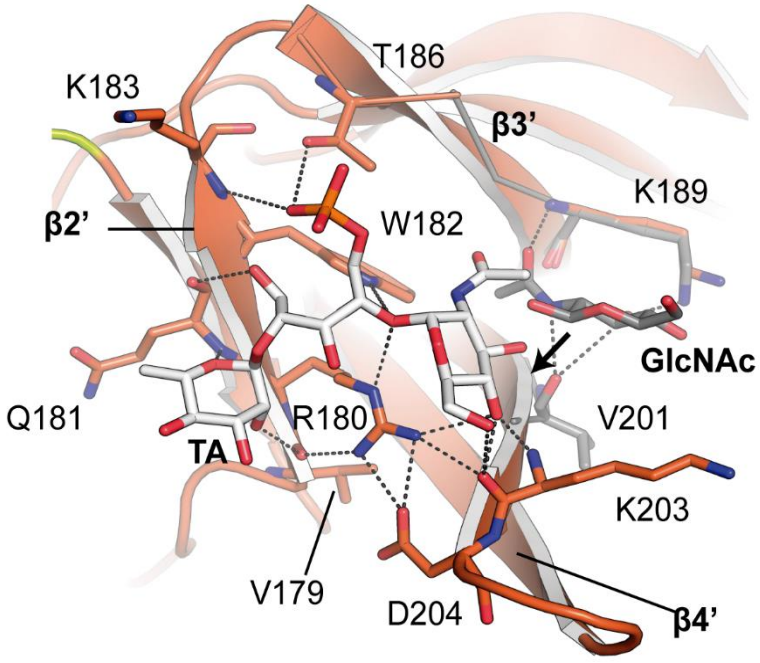


Figure 4.52.: β-D-GlcNAc vs TA in binding site 1. CBDP35:GlcNAc (grey) and CBDP35:TA (chain J, brown) complexes have been superposed and shown in cartoon. Interacting residues are shown as sticks and labelled accordingly. TA and β-D-GlcNAc molecules are shown in white and grey sticks, respectively. It can be observed that the α-D-GlcNAc decoration of the TA monomer is displaced from the previously determined binding site due to newly formed interactions with the Arg180 network.

Regarding binding site 2, a narrower range of displacement distances between the TA's α-D-GlcNAc moiety in CBDP35:TA and β-D-GlcNAc molecules in CBDP35:GlcNAc are observed, spanning ~3 Å (Figure 4.53.). Furthermore, β-D-GlcNAc and the α-D-GlcNAc of the TA molecule are equally oriented and a slight tilt between the sugar molecules, different for all chains, is perceived. Hence, the residues which are relevant for β-D-GlcNAc binding in the CBDP35:GlcNAc complex have also been determined in the binding of the TA molecule (Figures 4.42. and 4.47.), although variations in the positioning of the TA's α-D-GlcNAc decoration involve other residues such as Pro171, Asn194, Ser195 and Gly247 for its stabilization which had not identified in the CBDP35:GlcNAc complex. Moreover, binding site 2 comprises a smaller, less exposed surface than binding site 1, and so the RboP and Rha sugars would be completely solvent-exposed. The lack of interactions in binding site 2 other than for the α-D-GlcNAc moiety, can be related to the TA's mobility in this pocket leading to scarce electron densities which make the determination of the complete TA molecule difficult.

L. monocytogenes cell wall TAs are comprised of a poly-RboP chain that consists of 20 to 30 repeating units (section 1.4.3.) All in all, TA monomers (one repeating unit) were used for CBDP35:TA elucidation and binding characterization. It is expected that, upon CBDP35-GlcNAc recognition in a poly-RboP TA, further subtle rearrangements and interaction variations could occur from what we have here elucidated, although still conserving the residues involved in α-D-GlcNAc recognition for full TA binding.

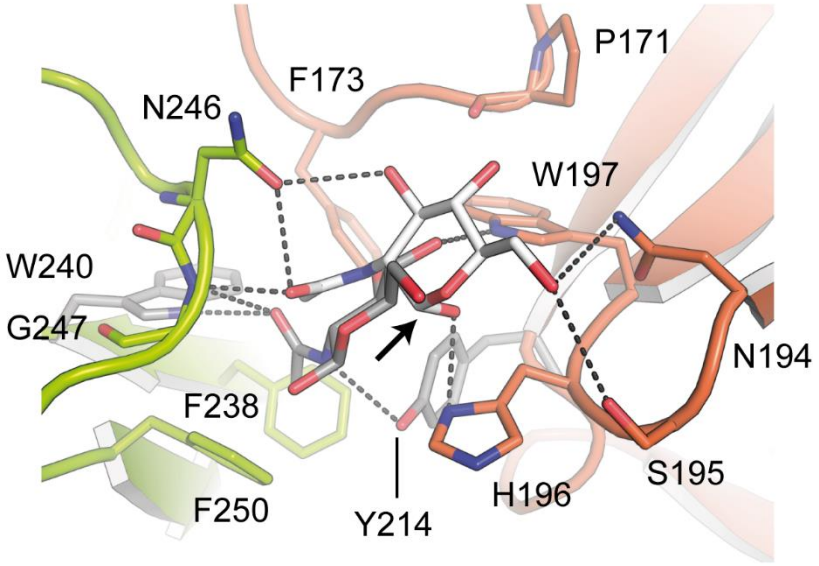


Figure 4.53.: β-D-GlcNAc vs α-D-GlcNAc of TA in binding site 2. CBDP35:GlcNAc (grey) and CBDP35:TA (chain L, green and brown for proximal and distal subdomains, respectively; white for α-D-GlcNAc molecule) complexes have been superposed and shown in cartoon. Interacting residues are shown as sticks and colored and labelled accordingly, showing in grey the residue with which only the β-D-GlcNAc molecule interacts. GlcNAc molecules are shown in sticks. α-D-GlcNAc is displaced by ~3 Å and tilted with respect to the CBDP35:GlcNAc complex.

4.3.2.4. Final Remarks

The previous sections describe the binding of CBDP35 with *L. monocytogenes* serovar 1/2a TA. Given the previous knowledge of CBDP35 specificity for GlcNAc (Eugster, *et al.*, 2011) and our results from CBDP35:GlcNAc complex structure analysis, we identified two binding sites for β-D-GlcNAc, namely site 1 and 2 which comprise residues from only the distal subdomain or both proximal and distal subdomains, respectively.

These binding sites differ in nature. Binding site 1 is exposed and does not show a defined cavity for only α-D-GlcNAc binding, but allowing the TA repeating unit molecule to also interact with CBDP35 through its RboP and Rha moieties. On the other hand, binding site 2 is a tightly enclosed pocket that only recognizes the α-D-GlcNAc decoration of the RboP backbone, leaving the RboP and Rha moieties exposed to the solvent.

Remarkably, the amino-acidic composition of CBDP35 shows a high percentage of positively-charged residues, leading to an expected basic electrostatic potential of CBDP35 surface for teichoic acid binding. Despite TA flexibility and exposure to solvent, we have determined key residues involved in TA stabilization and α-D-GlcNAc recognition. In binding site 1, residues from β2', β3' and β4' of the distal subdomain contribute to TA binding. Herein, Val202 and Lys203 are generally involved in α-D-GlcNAc stabilization. Moreover, Arg180 and Trp182 contribute to the positioning of this sugar within binding site 1, as they interact with the resultant O1 from the glycosidic link between RboP and α-D-GlcNAc. The phosphate group is oriented towards β2'-β3', engaging Lys183 and Thr186 directly through the main chain atoms or side chain, respectively. In the case of the Rha decoration of the TA, it has been observed that residues from β2' strand of the distal subdomain, more specifically Val179 and Gln181, may be involved in its stabilization. In binding site 2, Pro171, Asn194, Ser195, His196, Trp197, Asn246 and Gly247 participate in α-D-GlcNAc recognition. However, other neighboring residues such as Tyr214, Phe238, Trp240 and Phe250 may be also involved by Van der Waals interactions and contribute to the hydrophobic environment of binding site 2.

Both CBDP35:GlcNAc and CBDP35:TA complexes suggest that sites 1 and 2 bind GlcNAc regardless of strong and defined interactions that position the TA in a specific conformation. This may be related to the activity of endolysins, as they do not require a precise and regulated peptidoglycan hydrolysis, but the rapid and efficient degradation of the bacterial cell wall allowing a prompt release of its progeny.

5. CONCLUSIONS

The main conclusions obtained throughout this Dissertation are listed for each of the three proteins studied:

· ***Leishmania infantum*** **Trypanothione reductase (TryR):**

1. TryR crystals soaked with trypanothione and NADPH yielded a structure at better resolution than previously reported by others (PDB 4ADW). This showed a novel conformation of trypanothione bound to the protein active site, in which the spermidine that bridges both glutathionyl moieties of trypanothione is situated in the so called polyamine binding site and interacts with Trp21 of TryR.

2. Structural determination of TryR complexes obtained by soaking with four oxidoreductase inhibitors was achieved; three of them (TRL149, TRL187 and TRL190) having the 5-6-5-triazole-phenyl-thiazole scaffold; and one (TRL156) being a pyrrolopyrimidine. These ligands inhibit TryR by occupying its active site, more precisely the polyamine-binding site, engaging three residues for its stabilization that are essential for trypanothione binding (Glu18, Trp21 and Tyr110). The inhibitors did not trigger significant conformational changes on the structure of TryR.

3. In the case of TRL187, two inhibitor molecules have been found within TryR's active site by means of stacking interactions in a mepacrine-like fashion, resulting in the relocation of trypanothione which exhibited a non-canonical conformation and mode of binding to TryR.

4. Despite their IC_{50} values in the low micromolar range, insight derived from these crystallographic studies may guide further attempts to design optimized oxidoreductase inhibitors with enhanced affinity for TryR, increasing their potential for clinical use.

· **Rabbit muscle Elongation Factor 1A2 (eEF1A2)*:***

1. Natural (non-recombinant) eEF1A2 in complex with GDP was crystallized and its structure determined, revealing conformational changes with the existing structure of eEF1A2·GDP (PDB 4C0S). Most noticeably, these included the unfolding of helix C* in domain I; the structural delineation as a helix of the previously unobserved C-terminal residues; and the lack of Mg^{2+} in the GDP site, a finding that strongly suggests that Mg^{2+} is unessential for nucleotide binding and supports the disengagement between the presence of Mg^{2+} and the structural rearrangements occurring with GDP-GTP exchange.

2. A combined X-ray crystallography and mass-spectrometry (MS) approach revealed 36 post-translational modifications (PTMs) in eEF1A2, 23 of these had been previously unreported. In particular, two unique PTMs, namely the addition of an ethanolamine-phosphoglycerol moiety to Glu301 and Glu374, visualized in Glu301 of chain A and Glu374 of chain B, were determined by X-ray crystallography for the first time. Other PTMs found in the electron density map of eEF1A2 were phosphorylations of Ser300 and Ser316 of chain A and of Thr239 of chain B; and the glutamate side-chain sterification by a methanol (glutamate methylation) at Glu122 of domain I and Glu413 of domain III, both in chain B.

3. Although PTMs mostly map on a variable surface of eEF1A2 and do not affect the conserved surface through which the subunits of the dimer interact, some PTMs alter intramolecular salt bridges and interdomain interactions, regulating domain flexibility and stabilization. Methylation of Glu122, which sits at the nucleotide binding site, has been shown to possibly regulate nucleotide binding affinity. These changes could have regulatory potential, as well as our newly described C-terminal helix through plausible interactions with the rRNA of the small ribosomal subunit.

· Structural complex of the cell wall binding domain of endolysin PlyP35 (CBDP35) with teichoic acids from *Listeria monocytogenes* serovar 1/2a:

1. The crystal structure of CBDP35 revealed a double barrel fold (namely distal and proximal subdomains each formed by 7 β-strands) that had only been observed in the endolysin PlyPSA from another *L. monocytogenes* phage which recognizes *L. monocytogenes* serovar 4 cells. The electrostatic potential shows a major basic surface, coherent for TA binding.

2. Two N-acetyl-β-D-glucosamine (β-D-GlcNAc) binding sites were identified in the CBDP35·GlcNAc complex. Site 1 is formed by residues in the distal subdomain and is framed by β2', β3' and β4', whilst site 2 is a small cavity formed in the interface between the proximal and distal subdomains.

3. In the CBDP35:TA complex, single TA molecules (or only the N-acetyl-α-D-glucosamine (α-D-GlcNAc) moiety) were determined bound to site 1 and site 2. Site 1 comprises a wide and exposed region in which the TA molecule is bound. In site 2, the

polyribitol-phosphate (RboP) and α-L-Rhamnose (α-L-Rha) sugars remain solvent-exposed. In both cases GlcNAc moiety of the TA molecules are recognized.

4. Our crystal structure of the CBDP35:TA complex reveals, Val179, Arg180, Gln181, Trp182, Lys183, Thr186, Val202 and Lys 203 from site 1, and Pro171, Asn194, Ser195, His196, Trp197, Phe238, Trp240, Asn246, Gly247 and Phe250 from site 2 as crucial for GlcNAc recognition and TA anchoring. Furthermore, the longer β4'-β5' pair of the distal subdomain and their anchoring to β2' determine the structure for binding site 1, conforming a β-strand architecture absent in PlyPSA endolysin.

5. Previously reported PlyPSA hypothetical TA binding site shows high grade of similarity to CBDP35 binding site 2, suggesting that PlyPSA may also recognize α-D-GlcNAc moieties also decorating *L. monocytogenes* serovar 4 cells. Hence, it is inferred that α-D-GlcNAc recognition is given by the distal subdomain, as no other structural homologue is identified with this fold.

6. BIBLIOGRAPHY

Abbas, W., Kumar, A. amd Herbein, G. (2015) Th eEF1A proteins: at the crossroads of oncogenesis, apoptosis and viral infections. *Front. In Oncol*. 5:1-10.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. y Zwart, P. H. (2010) PHENIX: a comprehensive Pythonbased system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, 66, 213–221.

Afonine, P. V., Grosse-Kuntsleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Musttakimov, M., Terwilliger, T. C., Urzhumstsev, A., Zwart, P. H., Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68, 352-67.

Afonine, P. V., Moriarty, N. W., Mustyakimov, M., Sobolev, O. V., Terwilliger, T. C., Turk, D., Urzhumtsev, A., Adams, P. D. (2015) FEM: feature-enhanced map. *Acta Cryst*. D71, 646-666.

Ahuja, D., Vera, M. D., SirDeshpande, B. V., Morimoto, H., Williams, P. G., Joullié, M. M., Toogood, P. L. (2000) Inhibition of protein synthesis by didemnin B: how EF-1alpha mediates inhibition of translocation. *Biochemistry*. 39(15):4339-4346.

Als-Nielsen, J. y McMorrow, D. (2011) Elements of modern X-ray physics. John Wiley & Sons.

Amiri, A., Noei, F., Jeganathan, S., Kulkarni, G., Pinke, D. E., Lee, J. M. (2007) eEF1A2 activates Akt and stimulates Akt-dependent actin remodelling, invasion and migration. *Oncogene*. 26(21):3027-3040.

Anand, N, Murthy, S., Amann, G., Wernick, M., Porter, L. A., Cukier, I. H., Collins, C., Gray, J. W., Diebold, J., Demetrick, D. J., Lee, J. M. (2002) Protein elongation factor eEF1A2 is a putative oncogene in ovarian cancer. *Nat. Genet*. 31(3):301-305.

Andersen, G. R., Thirup, S., Spremulli, L. L., Nyborg, J. (2000a) High resolution crystal structure of bovine mitochondrial EF-tu in complex with GDP. *J. Mol. Biol.* 297(2), 421-436.

Andersen, G. R., Pedersen, L., Valentem L., Chatterjee, I., Kinzy, T. G., Kjeldgaard, M., Nyborg, J. (2000) Structural basis for nucleotide exchange and competition with tRNA in the yeast elongation factor complex eEF1A:eEF1Balpha. *Mol Cell.* 6(5):1261-6.

Andersen, G. R., Valente, L., Pedersen, L., Kinzy, T. G., Nyborg, J. (2001) Crystal structures of nucleotide exchange intermediates in the eEF1A-eEF1Balpha complex. *Nat. Struct. Biol.* 8(6):531-4.

Anderson, A. C. (2003) The Process of Structure-Based Drug Design. *Chem.&Biol*. 10: 787–797.

Ariyanayagam M.R. and Fairlamb A.H. (2001) Ovothiol and trypanothione as antioxidants in trypanosomatids. *Mol Biochem Parasitol.* 115(2):189-198.

Ariyanayagam, M. R., Oza, S. L., Guther, L. S., Fairlamb, A. H. (2005) Phenotypic analysis of trypanothione synthetase knockdown in the African trypanosome. *Biochem. J.* 391(pt.2):425-432.

Ayyari, M., Salehim P., Ebrahimi, S. N., Zimmermann, S., Portmann, L., Krauth-Siegel, R. L., Kaiser, M., Brun, R., Rezadoost, H., Rezazadeh, S., Hamburger, M. (2014)

Antitrypanosomal isothiocyanate and thiocarbamate glycosides from *Moringa peregrina.* 80:86-89.

Baca, Q. J., Coen, D. M., and Golan, D. E. (2012) Principles of combination chemotherapy. In Golan, D. E., Tashjian, A. H., Jr. Amstrong, E. J., Armstrong, A. W. (ed), Principles of pharmacology: the pathophysiologic basis of drug therapy, 3rd ed. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, PA. Pg 716-727.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. and Leslie, A. G. W. (2011) IMosflm: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst. D.* 67, 271-281.

Baiocco, P., Colotti, G., Franceschini, S., Ilari, A. (2009) Molecular Basis of Antimony Treatment in Leishmaniasis. *J. Med. Chem*. 52(8): 2603-2612.

Baiocco, P., Ilari, A., Ceci P., Orsini, S., Gramiccia, M., Di Muccio, T., Colotti, G. (2011) Inhibitory effect of silver nanoparticles on trypanothione reductase activity and Leishmania infantum proliferation. *ACS Med. Chem. Lett.* 2, 230–233.

Baiocco, P., Poce, G., Alfonso, S., Cocozza, M., Porretta, G.C., Colotti, G., Biava, M., Moraca, F., Botta, M., Yardley, V., Fiorillo, A., Lantella, A., Malatesta, F., Ilari, A. (2013) Inhibition of Leishmania infantum trypanothione reductase by azole-based compounds: a comparative analysis with its physiological substrate by X-ray crystallography. *ChemMedChem.* 8(7):1175-1183.

Beddell, C. R., Goodford, P. J., Norrington, F. E., Wilkinson, S., Wootton, R. (1976) Compounds designed to fit a site of known structure in human haemoglobin. *J. Pharmacol*. 57(2):201-209.

Bedi, M. S., Verma, V., Chhibber, S. (2009) Amoxicillin and specific bacteriophage can be used together for eradication of biofilm of Klebsiella pneumoniae B5055. *World J. Microbiol. Biotechnol.* 25:1145

Beig, M., Oellien, F., Garoff, L., Noack, S., Krauth-Siegel, R. L., Selzer, P. M. (2015) Trypanothione reductase: a target protein for a combined in vitro and in silico screening approach. *PLoS Negl Trop Dis.* 9(6):e0003773.

Belluti, F., Uliassi, E., Veronesi, G., Bergamini, C., Kaiser, M., Brun, R., Viola, A., Fato, R., Michels, P. A. M., Krauth-Siegel, R. L., Cavalli, A., Bolognesi, M. L. (2014) Towards the development of dual-targeted glyceraldehyde-3-phosphate dehydrogenase /trypanothione reductase inhibitors against Trypanosoma brucei and Trypanosoma cruzi. *ChemMedChem.* 9(2):371-82.

Benson, T. J., McKie, J. H., Garforth, J., Borges, A. Fairlamb, A. H., Douglas, K. T. (1992) Rationally designed selective inhibitors of trypanothione reductase. Phenothiazines and related tricyclics as lead structures. *Biochem. J.* 286: 9-11.

Benvenuti, M. y Mangani, S. (2007) Crystallization of soluble proteins in vapor diffusion for X-ray crystallography. *Nat Protoc*, 2, 1633–1651.

Berejnov, V., Husseini, N. S., Alsaied, O. A. y Thorne, R. E. (2006) Effects of cryoprotectant concentration and cooling rate on vitrification of aqueous solutions. *J Appl Cryst*, 39, 244–251.

Berg, J. M., Tymoczko, J. L., Stryer, L. (2002). Biochemistry. 5th Edition. Section 29.5. W. H. Freeman.

Bergfors, T. (2009) Protein crystallization. International University Line.

Bernardes, L. S. C., Zani, C. L. and Carvalho, I. (2013) Trypanosomatidae Diseases: From the Current Therapy to the Efficacious Role of Trypanothione Reductase in Drug Discovery. *Curr. Med. Chem*. 20, 2673-2696.

Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., Brunak, S. (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics.* 4(6), 1633-49.

Bocedi, A., Dawood, K. F., Fabrini, R., Federici, G., Gradoni, L., Pedersen, J. Z., Ricci, G. (2010) Trypanothione efficiently intercepts nitric oxide as a harmless iron complex in trypanosomatid parasites. *Faseb J*. 24(4):1035-1042.

Bond, C. S., Zhang, Y., Berriman, M., Cunningham, M. L., Fairlamb, A.H., Hunter, W. N. (1999) Crystal structure of Trypanosoma cruzi trypanothione reductase in complex with trypanothione and the structure-based discovery of new natural product inhibitors. *Structure.* 7:81-89

Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezhuk, Y., Raytselis, Y., Sayers, E.W., Tao, T., Ye, J., Zaretskaya, I. (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res*. 41: W29-W33.

Borysowski, J., Weber-Dabrowska, B., Górski, A. (2006) Bacteriophage endolysins as a novel class of antibacterial agents. *Exp. Biol. Med.* 231(4):366-377.

Bonse, S., Santelli-Rouvier, C., Barbe, J., Krauth-Siegel, R. L. (1999) Inhibition of Trypanosoma cruzi Trypanothione Reductase by acridines: kinetic studies and structure-activity relationships. *J. Med. Chem.* 42, 5448-5454.

Bonse, S., Richards, J. M., Ross, S. A., Lowe, G., Krauth-Siegel, R.L. (2000) (2,2':6',2''-Terpyridine) platinum(II) complexes are irreversible inhibitors of Trypanosoma cruzi trypanothione reductase but not of human glutathione reductase. *J. Med. Chem*. 43(25):4812-4821.

Born, Y., Fieseler, L., Klumpp, J., Eugster, M. R., Zurfluh, K., Duffy, B., Loessner, M. J. (2014) The tail-associated depolymerase of Erwinia amylovora phage L1 mediates host cell adsorption and enzymatic capsule removal, which can enhance infection by other phage. *Environ. Microbiol.* 16:2168–2180.

Bourne, H. R., Sanders, D. A., and McCormick, F. (1991) The GTPase superfamily: conserved structure and molecular mechanism. *Nature.* 349 117-127.

Bragg, W. L. (1913) The diffraction of short electromagnetic waves by a crystal. *Proc Cambridge Philos Soc*, 17, 43–57.

Briers, Y., Walmagh, M., Van Puyenbroeck, V., Cornelissen, A., Cenens, W., Aertsen, A., Oliveira, H., Azeredo, J., Verween, G., Pirnay, J. P., Miller, S., Volckaert, G., Lavigne, R. (2014) Engineered endolysin-based "artilysins" to combat multidrug-resistant gram-negative pathogens. *mBio* 5:e01379-14.

Brünger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355, 472–475.

Bunai, F., Ando, K., Ueno, H., Numata, O. (2006) Tetrahymena eukaryotic translation elongation factor 1A (eEF1A) bundles filamentous actin through dimer formation. *J. Biochem.* 140, 393-399.

Burley, S.K., Berman, H.M., Christie, C., Duarte, J., Feng, Z., Westbrook, J., Young, J. and Zardecki, C. (2018) RCSB Protein Data Bank: sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci*., 27, 316–330.

Bustamante, N., Iglesias-Bexiga, M., Bernardo-García, N., Silva-Martín, N., García, G., Campanero-Rhodes, M. A., García, E., Usón, I., Buey, R., García, P., Hermoso, J. A., Bruix, M., Menéndez, M. (2017) Deciphering how Cpl-7 cell wall-binding repeats recognize the bacterial peptidoglycan. *Sci. Rep.* 7(16494).

Cao, H., Zhu, Q., Huang, J., Li, B., Zhang, S., Yao, W., Zhang, Y. (2009) Regulation and functional role of eEF1A2 in pancreatic carcinoma. *Biochem. Biophys. Res. Commun.* 24(1):11-16.

Castro H. and Tomás A.M. (2008) Peroxidases of trypanosomatids. *Antioxid Redox Signal.* 10(9):1593-606.

Cavalli, A. and Bolognesi, M. L. (2009) Neglected tropical diseases: multi-target-directed ligands in the search for novel lead candidates against Trypanosoma and Leishmania. *J. Med. Chem.* 52:7339-7359.

Cavallius, J., Zoll, W., Chakraburtty, K. and Merrick, W. C. (1993) Characterization of yeast EF-1alpha: non-conservation of post-translational modifications. *Biochim. Bio-phys. Acta.* 1163, 75-80.

Chan, C., Yin, H., Garforth, J., McKie, J. H., Jaouhari, R., Speers, P., Douglas, K. T., Rock, P. J., Yardley, V., Croft, S. L., Fairlamb, A. H. (1998) Phenothiazine inhibitors of trypanothione reductase as potential antitrypanosomal and antileishmanial drugs. *J. Med. Chem.* 41:148-156.

Chanishvili N. (2012) Phage therapy—history from Twort and d'Herelle through Soviet experience to current approaches. *Adv. Virus. Res*. 83:3–40.

Chappuis, F., Sundar, S., Hailu, A., Ghalib, H., Rijal, S., Peeling, R. W., Alvar, J., Boelaert, M. (2007) Visceral leishmaniasis: what are the needs for diagnosis, treatment and control*? Nat. Rev. Microbiol.* 5(11):873-882.

Chayen, N. E. (1997) The role of oil in macromolecular crystallization. *Structure*, 5, 1269–1274.

Chayen, N. E. y Saridakis, E. (2008) Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Methods*, 5, 147–153.

Chen, E., Proestou, G., Bourbeau, D., Wang, E. (2000) Rapid up-regulation of peptide elongation factor EF-1A protein levels is an immediate early event during oxidative stress-induced apoptosis. *Exp. Cell Res*. 259: 140-148.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. y Richardson, D. C. (2010) MolProbity: all-atom structure

validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, 66, 12–21.

Chopra, S., Harjai, K., Chhibber, S. (2016) Potential of combination therapy of endolysin MR-10 and minocycline in treating MRSA induced systemic and localized burn wound infections in mice. *Int. J. Med. Microbiol.* 306: 707-716.

Chys, P. and Chacón, P. (2013) Random Coordinate Descent with Spinor-matrices and Geometric Filters for Efficient Loop Closure. *JCTC*. 9(3), 1821–1829.

Cimarelli, A. and Luban, J. (1999) Translation elongation factor 1A interacts specifically with the human immunodeficiency virus type 1 nucleocapsid with actin. *J. Virol*. 73: 2901-2908.

Colotti, G., Baiocco, P., Fiorillo, A., Boffi, A., Poser, E., Di Chiaro, F., Ilari, A. (2013). Structural insights into the enzymes of the trypanothione pathway: targets for antileishmaniasis drugs. *Future Med. Chem.* 5(15):1861-1875.

Comini, M. A. and Flohé, L. (2013). Trypanothione-based redox metabolism of trypanosomatids. In Trypanosomatid Diseases: Molecular Routes to Drug Discovery, chapter 9 (eds P. M. Selzer, T. Jäger, O. Koch and L. Flohé). doi:10.1002/9783527670383.ch9

Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr*, 62, 1002–1011.

Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr*, 66, 470–478.

Cummings C. G., Ross, N. T., Katt, W. P., Hamilton, A. D. (2009) Synthesis and biological evaluation of a 5-6-5 imidazole-phenyl-thiazole based alpha-helix mimetic. *Org. Lett*. 11(1):25-8.

Crepin, T., Vyacheslav, F S., Yaremchuk, A. D., Vlasenko, D. O., McCarthy, A., Negrutskii, B. S., Tukalo, M. A., El'skaya, A. (2014) Mammalian translation elongation factor eEF1A2: X-ray structure mechanism in higher eukaryotes. *Nucl. Acid Res.* 20(42):12939-12948.

Croft, S. L. and Engel, J. (2006) Miltefosine discovery of the antileishmanial activity of phospholipid derivatives. *Trans. R. Soc. Trop. Med. Hyg*. 100 Suppl 1: S4-8.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. y McPherson, A. (1994) Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr D Biol Crystallogr*, 50, 414–423.

Cushman, D. W., Cheung, H. S., Sabo, E. F., Ondetti, M. A. (1977) Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids. *Biochemistry*. 16(25):5484-5491.

Davis, W. G., Blackwell, J. L., Shi, P. Y., Brinton, M. A. (2007) Interaction between the cellular protein eEF1A and the 3′-terminal stem-loop of West Nile virus genomic RNA facilitates viral minus-strand RNA synthesis. *J. Virol*. 81,10172–1018.

Díez-Martínez, R., De Paz, H. D., García-Fernández, E., Bustamante, N., Euler, C. W., Fischetti, V. A., Menendez, M., García, P. (2015) A novel chimeric phage lysin with high in vitro and in vivo bactericidal activity against Streptococcus pneumoniae. *J. Antimicrob. Chemother.* 70: 1763-1773.

Dorscht, J., J. Klumpp, R. Bielmann, M. Schmelcher, Y. Born, M. Zimmer, R. Calendar, Loessner, M. J. (2009) Comparative genome analysis of Listeria bacteriophages reveals extensive mosaicism, programmed translational frameshifting, and a novel prophage insertion site. *J.Bacteriol.* 191: 7206-7215.

Duckworth, D. H. (1976) Who discovered bacteriophage? *Bacteriol. Rev.* 40:793-802.

Dumas, C., Ouellette, M., Tovar, J., Cunningham, M. L., Fairlamb, A. H., Tamar, S., Olivier, M., Papadopoulou, B. (1997) Disruption of the trypanothione reductase gene of Leishmania decreases its ability to survive oxidative stress in macrophages. *EMBO J.* 16(10): 2590-2598.

Edmonds, B. T., Wyckoff, J., Yeung, Y. G., Wang, Y., Stanley, E. R., Jones, J., Segall, J., Condeelis, J. (1996) Elongation factor-1 alpha is an overexpressed actin binding protein in metastatic rat mammary adenocarcinoma. *J. Cell. Sci.* 109(11) 2705-2714.

Ejiri, S. (2002) Moonlighting functions of polypeptide elongation factor 1: from actin bundling to zinc finger protein R1-associated nuclear localization. *Biosc. Biotech. Bio-chem.* 66(1), 1-21.

El-Sayed N.M., Myler P.J., Blandin G., Berriman M., Crabtree J., Aggarwal G., Caler E., Renauld H., Worthey E.A., Hertz-Fowler C., Ghedin E., Peacock C., Bartholomeu D.C., Haas B.J., Tran A.N., Wortman J.R., Alsmark U.C., Angiuoli S., Anupama A., Badger J., Bringaud F., Cadag E., Carlton J.M., Cerqueira G.C., Creasy T., Delcher A.L., Djikeng A., Embley T.M., Hauser C., Ivens A.C., Kummerfeld S.K., Pereira-Leal J.B., Nilsson D., Peterson J., Salzberg S.L., Shallom J., Silva J.C., Sundaram J., Westenberger S., White O., Melville S.E., Donelson J.E., Andersson B., Stuart K.D., Hall N. (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science.* 309(5733)404-409.

El'skaya, A. V., Turkovskaya, G. V., Rachkov, A. E., Starodub, N. F. (1999) Regulation of protein synthesis in higher eukaryotes: facts and hypotheses. *Mol. Biol*. 33:(N6), 1043-1053.

Eugster, M. R., Haug, M. C., Huwiler, S. G., Loessner, M. J. (2011) The cell wall binding domain of Listeria bacteriophage endolysin PlyP35 recognizes terminal GlcNAc residues in cell wall teichoic acid. *Mol. Microb*. 81(6):1419-1432.

Eugster, M. R. and Loessner, M. J. (2011) Rapid analysis of Listeria monocytogenes cell wall teichoic acid carbohydrates by ESI-MS/MS. *PloS ONE.* 6(6): e21500.

Eugster, M. R. and Loessner, M. J. (2012) Wall teichoic acids restrict access of bacteriophage endolysin Ply118, Ply511, and PlyP40 cell wall binding domains to the Listeria monocytogenes peptidoglycan. *J. Bacteriol.* 194:6498–6506.

Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, 62, 72–82.

Evans, P. R. (2007) An introduction to stereochemical restraints. *Acta Crystallogr D Biol Crystallogr*, 63, 58–61.

Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr*, 67, 282–292.

Evans, P. R. y Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr*, 69, 1204–1214.

Ewald, P. P. (1969) Introduction to the dynamical theory of X-ray diffraction. *Acta Cryst*., A25, 103–108.

Fan, K., Chrzanowska-Lightowlers, Z. M. A. and Hesketha, J. E. (2009) Fine mapping of interactions between eEF1α protein and 3′UTR of metallothionein-1 mRNA. *Biochem Biophys Res Commun.* 386(1), 82–88.

Fan, J., Zeng, Z., Mai, K., Yang, Y., Feng, J., Bai, Y., Sun, B., Xie, Q., Tong, Y., Ma, J. (2016) Preliminary treatment of bovine mastitis caused by Staphylococcus aureus, with trx-SA1, recombinant endolysin of S. aureus bacteriophage IME-SA1. *Vet. Microbiol.* 191:65–71.

Fairlamb, A. H. and Cerami, A. (1992) Metabolism and functions of trypanothione in the kinetoplastida. *Annu. Rev. Microbial*. 46, 695-729.

Fevre, E. M., Wissmann, B.V., Welburn S. C., Lutumba, P. (2008) The burden of human African trypanosomiasis. *PLoS Negl. Dis*. 2, e333.

Fiedler, F. (1988) Biochemistry of the cell surface of Listeria strains: a locating general view. *Infection.* 16(Suppl. 2):S92-S97.

Fischer, W. (1988) Physiology of lipoteichoic acids in bacteria. *Adv. Microb. Physiol.* 29: 233–302.

Fischetti, A. V. (2010) Bacteriophage endolysins: a novel anti-infective to control Gram-positive pathogens. *Int. J. Med. Microbiol.* 300:357-362.

Fujii, H., Kamisango, K., Nagaoka, M., Uchikawa, K., Sekikawa, I., Yamamoto, K., Azuma, I. (1985) Structural study on teichoic acids of Listeria monocytogenes types 4a and 4d. *J. Biochem.* 97, 883–891.

Gallwitz H., Bonse, S., Martinez-Cruz, A., Schlichting, I., Schumacher, K., Krauth-Siegel, R. L. (1999) Ajoene is an inhibitor and subversive substrate of human glutathione reductase and Trypanosoma cruzi trypanothione reductase: crystallographic, kinetic, and spectroscopic studies. *J. Med. Chem.* 42(3):364-72.

Gandhi, M. and Chikindas, M. L. (2007) Listeria: a foodborne pathogen that knows how to survive. *Int. J. Food Microb.* 113: 1-15.

Garman, E. F. (2010) Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr D Biol Crystallogr*, 66, 339–351.

Garman, E. F. y Mitchell, E. P. (1996) Glycerol concentrations required for cryoprotection of 50 typical protein crystallization solutions. *J. Appl. Cryst*., 29, 584–587.

Giacovazzo, C. (2013) Phasing in crystallography: a modern perspective. Oxford Science Publications.

Gilmer, D. B., Schmitz, J. E., Euler, C. W., Fischetti, V. A. (2013) Novel bacteriophage lysin with broad lytic activity protects against mixed infection by Streptococcus pyogenes and methicillin-resistant Staphylococcus aureus. *Antimicrob. Agents. Chemother.* 57: 2743-2750.

Girault, S., Davioud-Charvet, E., Salmon, L., Berecibar, A., Debreu, M.A., Sergheraert, C. (1998) Structure-activity relationships in 2- aminodiphenylsulfides against trypanothione reductase from Trypanosoma cruzi. *Bioorg. Med. Chem*. Lett. 8: 1175-1180.

Girdlestone, C. and Hayward, S. (2016) The DynDom3D webserver for the analysis of domain movements in multimeric proteins. J. Comp. Bio. 23(1), 21-26.

Gonen, H., Smith, C. E., Siegel, N. R., Kahana, C., Merrick, W. C., Chakraburtty, K., Schwarz, A. L., Ciechanover, A. (1994) Protein synthesis elongation factor EF-1 alpha is essential for ubiquitin-dependent degradation of certain N alpha-acetylated proteins and may substituted by a bacterial elongation factor EF-Tu. *Proc. Natl. Acad. Sci. USA*. 91:7648-7652.

Gordillo-Altamirano, F. L. and Barr, J. J. (2019) Phage therapy in the postantibiotic era. *Clin. Microbiol. Rev.* 32: e00066-18.

Grassi, G., Scaggiante, B., Farra, R., Dapas, B., Agostini, F., Baiz, D., Rosso, N., Tiribelli, C. (2007) The expression levels of the translational factors eEF1A1/2 correlate with cell growth but not apoptosis in hepatocellular carcinoma fell lines with different differentiation grade. *Biochimie*. 89(12):1544-1552.

Greganova, E., Altmann, M., Bütikofer, P. (2011) Unique modifications of translation elongation factors. *FEBS J.* 2613-2624.

Greganova, E., Heller, M. and Bütikofer, P. (2010) A structural domain mediates attachment of ethanolamine phosphoglycerol to eukaryotic elongation factor 1A in *Trypanosoma brucei*. *PLoS ONE*. 5, e9486.

Grimes, M., Hall, B., Foltz, L., Levy, T., Rikova, K., Gaiser, J., Cook, W., Smirnova, E., Wheeler, T., Clark, N. R., Lachmann, A., Zhang, B., Hornbeck, P., Ma'ayan, A., Comb, M. (2018) Integration of protein phosphorylation, acetylation and methylation data sets to outline lung cancer signalling networks. *Sci. Signal.* 11(531):314-319.

Gründling, A., Bläsi, U., Young, R. (2000) Genetic and biochemical analysis of dimer and oligomer interactions of the λ S holin. *J. Bacteriol.* 182:6082–6090.

Haddad-Kashani, H., Schmelcher, M., Sabzalipoor, H., Hosseini, E. S., Moniri, R. (2018) Recombinant endolysins as potential therapeutics against antibiotic-resistant Staphylococcus aureus: current status of research and novel delivery strategies. *Clin. Microbiol. Rev.* 31(1): e00071-17.

Hagens, S. and Loessner, M. J. (2014) Phages of Listeria offer novel tools for diagnostics and biocontrol. *Fron. Microb*. 5,159.

Hälser, J., Rada, C., Neuberger, M. S. (2012) The cytoplasmic AID complex. *Seminars in Immun.* 24, 273-280.

Hamilton, C. J., Saravanamuthu, A., Eggleston, I. M., Fairlamb, A. H. (2003) Ellman's-reagent-mediated regeneration of trypanothione in situ: substrate-economical microplate and time-dependent inhibition assays for trypanothione reductase. *Biochem J*. 369(Pt 3):529-537.

Hammond, D. J., Hogg, J., Gutteridge, W. E. (1985) Trypanosoma cruzi: possible control of parasite transmission by blood transfusion using amphiphilic cationic drugs. *Exp. Parasitol*. 60: 32-42.

Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, 3rd, H. L., Deng, S.-J. J., Gampe, R. T., Grisard, T. E., Madauss, K. P., Nolte, R. T., Rocque, W. J., Wang, L., Weaver, K. L., Williams, S. P., Wisely, G. B., Xu, R. and Shewchuk, L. M. (2007) Crystallization of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr*, 63, 72–79.

Hathaway, H., Ajuebor, J., Stephens, L., Coffey, A., Potter, U., Sutton, J. M., Jenkins, A. T. A. (2017) Thermally triggered release of the bacteriophage endolysin CHAP K and the bacteriocin lysostaphin for the control of methicillin resistant Staphylococcus aureus (MRSA). *J. Control Release*. 245:108 –115.

Henderson, G. B., Ulrich, P., Fairlamb, A. H., Rosenberg, I., Pereira, M., Sela, M., Cerami, A. (1988) "Subversive" substrates for the enzyme trypanothione disulfide reductase: alternative approach to chemotherapy of Chagas disease. *Proc Natl Acad Sci USA*. 85(15): 5374–5378.

Hermoso, J. A., García, J. L. and García, P. (2007) Taking aim on bacterial pathogens: from phage therapy to enzybiotics. *Curr. Op. Microbiol.* 10:461-472.

Holm L, Kääriäinen S, Wilton C, Plewczynski D. (2006) Using Dali for structural comparison of proteins. *Curr Protoc Bioinformatics*. 5:5.5.

Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., Skrzypek, E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res*. 43, 512-520.

Horta, M.F., Mendes, B.P., Roma, E.H., Noronha, F.S., Macêdo, J.P., Oliveira, L.S., Duarte, M. M., Vieira, L. Q. (2012) Reactive oxygen species and nitric oxide in cutaneous leishmaniasis. *J. Parasitol. Res*. 2012:203818.

Horvath, D. (1997) A virtual screening approach applied to the search for trypanothione reductase inhibitors. *J. Med. Chem*. 40(15):2412-23.

Hotokezaka, Y., Tobben, U., Hotokezaka, H., Van Leyen, K., Beatrix, B., Smith, D. H., Nakamura, T., Wiedmann, M. (2002) Interaction of the eukaryotic elongation factor 1A with newly synthesized polypeptides. *J. Biol. Chem*. 277: 18545-18551.

Idelevich, E. A., Schaumburg, F., Knaack, D., Scherzinger, A. S., Mutter, W., Peters, G., Peschel, A., Becker, K. (2016) The recombinant bacteriophage endolysin HY-133 exhibits in vitro activity against different African clonal lineages of the Staphylococcus aureus complex, including Staphylococcus schweitzeri. *Antimicrob. Agents Chemother.* 60:2551–2553.

Ilari, A., Baiocco, P., Messori, L., Fiorillo, A., Boffi, A., Gramiccia, M., Di Muccio, T., Colotti, G. (2012) A gold containing drug against parasitic polyamine metabolism: the x-ray structure of trypanothione reductase from *Leishmania infantum* in complex with auranofin reveals a dual mechanism of enzyme inhibition. *Amino Acids*. 42(2–3), 803–811.

Ilari, A., Fiorillo, A., Genovese, I., Colotti, G. (2017) An update on structural insights into the enzymes of the polyamine-trypanothione pathway: targets for new drugs against leishmaniasis. *Future Med. Chem.* 9(1):61-77.

Jakobsson, M. E., Malecki, J. and Falnes, P. Ø. (2018) Regulation of eukaryotic elonga-tion factor 1 alpha (eEF1A) by dynamic lysine methylation. *RNA Biol*. 15(3), 314-319.

Jakobsson, M. E., Malecki, J., Nilges, B. S., Moen, A., Leidel, S A., Falnes, P. O. (2017) Methylation of human eukaryotic elongation factor alpha (eEF1A) by a member of a novel

protein lysine methyltransferase family modulates mRNA translation. *Nucl. Acids Res.* 45(14):8239-8254.

Jager, T., Koch, O., Flohe, L. (2013) Trypanosomatid diseases: molecular routes to drug discovery. Drug Discovery in Infectious Diseases. Wiley-Blackwell Ed., Weinheim, Germany.

Janssen, G. M. C. and Moller, W. (1988) Kinetic studies on the role of elongation factors 1 beta and 1 gamma in protein synthesis. *J. Biol. Chem.* 263, 1773–1778

Jeppesen, M. G., Navratil, T., Spremulli, L. L., Nyborg, J. (2005) Crystal structure of the bovine mitochondrial elongation factor Tu.Ts complex. *J. Biol. Chem.* 280(6):5071-81.

Jha, T. K., Giri, Y. N., Singh, T. K., Jha, S. (1995) Use of amphotericin B in drug-resistant cases of visceral leishmaniasis in north Bihar, India. *Am. J. Trop. Med. Hyg*. 52(6):536-538.

Joosten, R. P., Womack, T., Vriend, G. y Bricogne, G. (2009) Re-refinement from deposited X-ray data can deliver improved models for most PDB entries. *Acta Crystallogr D Biol Crystallogr*, 65, 176–185.

Jun, S. Y., Jang, I. J., Yoon, S., Jang, K., Yu, K-S., Cho, J. Y., Seong, M. W., Jung, G. M., Yoon, S.J., Kang, S. H. (2017) Pharmacokinetics and tolerance of the phage endolysin-based candidate drug SAL200 after a single intravenous administration among healthy volunteers. *Antimicrob. Agents Chemother.* 61:e02629–e02616.

Kabsch, W. (2010). XDS. *Acta Cryst*. D66, 125-132.

Kanamaru, S., Ishiwata, Y., Suzuki, T., Rossmann, M. G., Arisaka, F. (2005) Control of bacteriophage T4 tail lysozyme activity during the infection process. *J. Mol. Biol.* 1013-1020

Kanibolotsky, D. S., Novosyl'na, O. V., Abbott, C. M., Negrutskii, B. S., El'skaya, A.V. Multiple molecular dynamics simulation of the isoforms of human translation elongation factor 1A reveals reversible fluctuations between "open" and "closed" conformations and suggests specific for eEF1A1 affinity for Ca2+-calmodulin. *BMC Struct. Bio*. 8:4.

Karplus, P. A. and Diederichs, K. (2015) Assessing and maximizing data quality in macromolecular crystallography. *Curr Opin Struct Biol*. 34: 60–68.

Karplus, P. A., Krauth-Siegel, R. L., Schirmer, R. H., Schulz, G.E. (1988) Inhibition of human glutathione reductase by the nitrosourea drugs 1,3-bis(2-chloroethyl)-1-nitrosourea and 1-(2-chloroethyl)-3-(2-hydroxyethyl)-1-nitrosourea. A crystallographic analysis. *Eur J Biochem.* 171(1-2):193-198.

Kato, M. V. (1999) The mechanisms of death of an erythroleukemic cell line by p53: involvement of the microtubule and mitochondria. *Leuk. Lymphoma.* 33: 181-186.

Kawashima, T., Berthet-Colominas, C., Wulff, M., Cusack, S., Leberman, R. (1996) The structure of the Escherichia coli EF-Tu.EF-Ts complex at 2.5 A resolution. *Nature.* 379(6565):511-8.

Keary, R., Sanz-Gaitero, M., van Raaij, M. J., O'Mahony, J., Fenton, M., McAuliffe, O., Hill, C., Ross, R. P., Coffey, A. (2016) Characterization of a bacteriophage-derived murein peptidase for elimination of antibiotic-resistant Staphylococcus aureus. *Curr. Protein Pept. Sci.* 17: 183-190.

Kelly, J. M., Taylor, M. C., Smith, K., Hunter, K. J., Fairlamb, A. H. (1993) Phenotype of recombinant Leishmania donovani and Trypanosoma cruzi which over-express trypanothione reductase. Sensitivity towards agents that are thought to induce oxidative stress. *Eur. J. Biochem.* 218(1):29-37.

Khacho, M., Mekhail, K., Pilon-Larose, K., Pause, A., Côte, J., Lee, S. (2008) eEF1A2 is a novel component of the mammalian nuclear protein export machinery. *Mol. Biol. Cell*. 19, 5296-5308.

Khan, M.O., Austin, S.E., Chan, C., Yin, H., Marks, D., Vaghjiani, S.N. (2000) Use of an additional hydrophobic binding site, the Z site, in the rational drug design of a new class of stronger trypanothione reductase inhibitor, quaternary alkylammonium phenothiazines. *J Med Chem*. 43(16):3148-3156.

Kitchen, D. B., Decornez, H., Furr, J. R., Bajorath, J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov*. 3(11):935-949.

Korndörfer, I. P., Danzer, J., Schmelcher, M., Zimmer, M., Skerra, A., Loessner, M. J. (2006) The crystal structure of the bacteriophage PSA endolysin reveals a unique fold responsible for specific recognition of Listeria cell walls. *J. Mol. Biol.* 364:678-689.

Kovářová J. and Barrett M. P. (2016) The Pentose Phosphate Pathway in Parasitic Trypanosomatids. *Trends Parasitol.* 32(8):622-634.

Krause, R. M. (1957) Studies on bacteriophages of haemolytic streptococci. I. Factors influencing the interaction of phage and susceptible host cell. *J. Exp. Med.* 106:365-384.

Krauth-Siegel, R.L., Bauer, H., Schirmer, R.H. (2005) Dithiol proteins as guardians of the intracellular redox milieu in parasites: old and new drug targets in trypanosomes and malaria-causing plasmodia. *Angew. Chem. Int. Ed. Engl.* 44(5):690-715.

Krauth-Siegel, R. L. and Comini, M. A. (2008) Redox control in trypanosomatids, parasititc protozoa with trypanothione-based thiol metabolism. *Biochim. Biophys. Acta*. 1780(11), 1236-1248.

Krauth-Siegel, R. L., Comini, M. A. and Schlecker, T. (2007) The trypanothione system. *Subcell Biochem.* 44: 231-251

Krauth-Siegel, R. L., Leroux, A. E. (2012) Low-molecular-mass antioxidants in parasites. *Antioxid. Redox Signal.* 17, 583–607.

Kretzer, J. W., Lehmann, R., Schmelcher, M., Banz, M., Kim, K. P., Korn, C., Loessner, M. J. (2007) Use of high-affinity cell wall-binding domains of bacteriophage endolysins for immobilization and separation of bacterial cells. *Appl. Environ. Microbiol*. 73, 1992–2000.

Kretzer, J. W., Schmelcher, M. and Loessner, M. J. (2018) Ultrasensitive and fast diagnostics of viable Listeria cells by CBD magnetic separation combined with A511::luxAB detection. *Viruses.* 10(11): 626.

Krieger, S., Schwarz, W., Ariyanayagam, M. R., Fairlamb, A. H., Krauth-Siegel, R. L., Clayton, C. (2000) Trypanosomes lacking trypanothione reductase are avirulent and show increased sensitivity to oxidative stress. *Mol. Microbiol.* 35(3): 542-552.

Krissinel, E. and Henrick, K. (2007). 'Inference of macromolecular assemblies from crystalline state.'. *J. Mol. Biol*. 372, 774--797.

Klumpp, J. and Loessner, M. J. (2013) Listeria phages: genomes, evolution and applications. *Bacteriophage*. 3:3, e26861.

Kuenne, C., Billion, A., Mraheil, M. A., Strittmatter, A., Daniel, R., Goesmann, A., Barbuddhe, S., Hain, T., Chakraborty, T. (2013) Reassessment of the Listeria monocytogenes pan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. *BMC Gen.* 14:47.

Kulkarni, G., Turbin, D. A., Amiri, A., Jeganathan, S., Andrade-Navarro, M. A., Wu, T. D., Huntsman, D. G., Lee, J. M. (2007) Expression of protein elongation factor eEF1A2 predicts favourable outcome in breast cancer. *Breast Cancer Res. Treat.* 102(1):31-41-

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227, 680–685.

Lamberti, A., Sanges, C., Longo, O., Chambery, A., Di Maro, A., Parente, A., Masullo, M., Arcari, P. (2008) Analysis of nickel-binding peptides in a human epidermoid cancer cell line by Ni-NTA affinity chromatography and mass spectrometry. Prot. & Peptide Letters. 15(10), 1126-1131.

Lamberti, A., Sanges, C., Chambery, A., Migliaccio, N., Rosso, F., Di Maro, A., Papale, F., Marra, M., Parente, A., Caraglia, M., Abbruzzese, A., Acari, P. (2011) Analysis of interaction partners for eukaryotic translation elongation factor 1A M-domain by functional proteomics. *Biochimie.* 93(10), 1738-1746.

Lang, P. T., Holton, J. M., Fraser, J. S. & Alber, T. (2014) Protein structural ensembles are revealed by redefining X-ray electron density noise. *Proc. Natl Acad. Sci*. USA, 111, 237–242.

Lantwin, C. B., Schlichting, I., Kabsch, W., Pai, E. F., Krauth-Siegel, R. L. (1994) The structure of Trypanosoma cruzi trypanothione reductase in the oxidized and NADPH reduced state. *Proteins*. 18(2):161-73.

Lee, B., Bauer, H., Melchers, J., Ruppert, T., Rattray, L., Yardley, V.; Davioud-Charvet, E., Krauth-Siegel, R. L. (2005) Irreversible inactivation of trypanothione reductase by unsaturated Mannich bases: a divinyl ketone as key intermediate. *J. Med. Chem*. 48:7400-7410.

Lee, S., Francoeur, A. M., Liu, S., Wang, E. (1992) Tissue-specific expression in mammalian brain, heart and muscle of S1, a member of the elongation factor-1 alpha gene family. *J. Biol. Chem*. 267, 24064-24068.

Lee, S. H., Stephens, J. L. and Englund, P. T. (2007) A fatty-acid synthesis mechanism specialized for parasitism. *Nat. Rev. Microbiol*. 5(4), 287-297.

Lee, M. H. and Surh, Y. J. (2009) eEF1A2 as a putative oncogene. *Ann. N. Y. Acad. Sci.* 1171:87-93.

Lee, J. H., Zhang, Q., Jo, S., Chai, S. C., Oh, M., Im, W. (2011) Novel pyrrolopyrimidine-based α-helix mimetics: cell-permeable inhibitors of protein−protein interactions. *J. Am. Chem. Soc*. 133(4):676-9.

Lee, M. H., Choi, B. Y., Cho, Y. Y., Lee, S. Y., Huang, Z., Kundu, J. K., Kim, M. O., Kim, D.J., Bode, A. M., Surh, Y. J., Dong, Z. (2013) Tumour suppressor p16INK4a inhibits cancer cell growth by downregulating eEF1A2 through a direct interaction. *J.Cell Science.* 126, 3796.

Leroux, A. E. and Krauth-Siegel, R. L. (2016) Thiol redox biology of trypanosomatids and potential targets for chemotherapy. *Mol. & Biochem. Parasitology*. 206:67-74.

Li, Z., Fennie, M. W., Ganem, B., Hancock, M. T., Kobaslija, M., Rattendi, D., Bacchi, C. J., O'Sullivan, M. C. (2001) Polyamines with N-(3-phenylpropyl) substituents are effective competitive inhibitors of trypanothione reductase and trypanocidal agents. *Bioorg. Med. Chem. Lett.* 11: 251-254

Li, D., Xie, K., Wolff, R., Abbruzzese, J. L. (2004) Pancreatic cancer. *Lancet.* 363(9414):1049-1057.

Li, R., Wang, H., Bekele, B. N., Yin, Z., Caraway, N. P., Katz, R. L., Stass, S. A., Jiang, F. (2006) Identification of putative oncogenes in lung adenocarcinoma by a comprehensive functional genomic approach. *Oncogene.* 25(18):2628-2635.

Li, D., Wei, T., Rawle, D. J., Qin, F., Wang, R., Soares, D. C., Jin, H., Sivakumaran, H., Lin, M. H., Spann, K., Abbott, C. M., Harrich, D. (2015) Specific interaction between eEF1A and HIV RT is critical for HIV-1 reverse transcription and a potential anti-HIV target. *Plos. Pathog*. 11(12):e1005289.

Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., Adams, P. D. (2017) Polder maps: Improving OMIT maps by excluding bulk-solvent. *Acta Cryst Struct Biol*. D73, 148-157

Lin, K. W., Yakymovych, I., Jia, M., Yakymovych, M., Soulchelnytskyi, S. (2010) Phosphorylation of eEF1A1 at Ser300 by TβR-I results in inhibition of mRNA translation. *Curr. Biol*. 20, 1615-1625.

Liu, G., Tang, J., Edmonds, B. T., Murray, J., Levin, S., Condeelis, J. (1996) F-actin sequesters elongation factor 1α from interaction with aminoacyl-tRNA in a pH-dependent reaction. J. Cell Biol. 135(4), 953-963.

Lizzi, F., Veronesi, G., Belluti, F., Bergamini, C., López-Sánchez, A., Kaiser, M., Reto, B., Krauth-Siegel, R. L., Hall, D. G., Rivas, L., Bolognesi, M. L. (2012) Conjugation of quinones with natural polyamines: toward an expanded antitrypanosomatid profile. *J. Med. Chem.* 55(23):10490-10500.

Loessner, M. J., Wendlinger, G. and Scherer, S. (1995) Heterogeneous endolysins in Listeria monocytogenes bacteriophages: a new class of enzymes and evidence for conserved holing genes within the siphoviral lysis cassettes. *Mol. Microbiol.* 16, 1231-1241.

Loessner, M.J., Schneider, A., and Scherer, S. (1996) Modified Listeria bacteriophage lysin genes (ply) allow efficient overexpression and one-step purification of biochemically active fusion proteins. *Appl. Environ. Microbiol*. 62: 3057–3060.

Loessner, M.J., Kramer, K., Ebel, F., and Scherer, S. (2002) C-terminal domains of Listeria monocytogenes bacteriophage murein hydrolases determine specific recognition and high-affinity binding to bacterial cell wall carbohydrates. *Mol. Microbiol*. 44: 335–349.

Loessner, M. J. (2005) Bacteriophage endolysins—current state of research and applications. *Curr. Opin. Microbiol.* 8:480-487.

Lood, R., Winer, B.Y., Pelzek, A.J., Diez-Martinez, R., Thandar, M., Euler, C. W., Schuch, R., Fischetti, V. A. (2015) Novel phage lysin capable of killing the multidrug-resistant gram-negative bacterium Acinetobacter baumannii in a mouse bacteremia model. *Antimicrob. Agents Chemother*. 59: 1983-1991.

López R, García E and García P (2004) Enzymes for anti-infective therapy: phage lysins. *Drug Discov. Today Ther. Strat.* 1:469-474.

Losada, A., Muñoz-Alonso, M. J., García, C., Sánchez-Murcia, P. A., Martínez-Leal, J. F., Domínguez J. M., Lillo, M. P., Gago, F., Galmarini, C. M. (2016) Translation elongation factor eEF1A2 is a novel anticancer target for the marine natural product plitidepsin. *Sci. Reports*. 6(35100), 1-15.

Losada, A., Muñoz-Alonso, M. J., Martínez-Díez, M., Gago, F., Domínguez, J. M., Martínez-Leal, J. F., Galmarini, C. M. (2018) Binding of eEF1A2 to the RNA-dependent protein kinase PKR modulates its activity and promotes tumour cell survival. *British J. Cancer*. 119, 1410–1420.

Lu, J., Vodnala, S. K., Gustavsson, A. L., Gustafsson, T. N., Sjöberg, B., Johansson, H. A., Kumar, S., Tjernberg, A., Engman, L., Rottenberg, M. E., Holmgren, A. (2013) Ebsulfur is a benzisothiazolone cytocidal inhibitor targeting the trypanothione reductase of Trypanosoma brucei. *J. Biol. Chem*. 288(38):27456-27468.

Lucio, H. 2018. Inhibución de la Tripanotión Reductasa de Leishmania infantum Mediada por Peptidomiméticos de la Hélice α2 de la Interfaz de Dimerización. –Universidad Alcalá de Henares.

Maciejewska, B., Olszak, T., Drulis-Kawa, Z. (2018) Applications of bacteriophages versus phage enzymes to combat and cure bacterial infections: an ambitious and also a realistic application? *Appl. Microbiol. Biotechnol*. 102:2563–2581.

Mao, J., Schmelcher, M., Harty, W. J., Foster-Frey, J., Donovan, D. M. (2013) Chimeric Ply187 endolysin kills Staphylococcus aureus more effectively than the parental enzyme. *FEMS Microbiol. Lett.* 342:30–36.

Mateyak, M. K. and Kinzy, T. G. (2010) eEF1A: Thinking Outside theRibosome. *J. Biol. Chem.* 285(28): 21209-21213.

Matthews, B. W. (1968) Solvent content of protein crystals. J Mol Biol, 33, 491–497.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Cryst*. 40, 658-674.

Merrick, W. (1992) Mechanism and Regulation of Eukaryotic Protein Synthesis. Microb. Reviews, 56(2)291-315.

Merritt, E. A. (2012) To B or not to B: a question of resolution? *Acta Crystallogr D Biol Crystallogr*, 68, 468–477.

de Menezes, H. P., Guedes, C. E., Petersen, A. L., Fraga, D. B., Veras, P. S. (2015) Advances in the Development of New Treatments for Leishmaniasis. *Biomed. Res. Int.* 2015:1-11.

Michels, P. M., Bringaud, F., Herman, M. and Hannaert, V. (2006) Metabolic functions of glycosomes in trypanosomatids. *Biochim. Biophys. Acta*, 1763(12):1463-1477.

Migliaccio, N., Sanità, G., Ruggiero, I., Martucci, N. M., Sanges, C., Rippa, E., Quagl-iariello, V., Papale, F., Arcari, P. and Lamberti, A. (2018) Cellular interaction of human eukaryotic elongation factor 1A isoforms. *Prot-Prot. Int. Essays*. DOI: 10.5772/intechopen.74733

Millan, C., Sammito, M. and Usón, I. (2015) Macromolecular ab initio phasing enforcing secondary and tertiary structure. *IUCrJ*. 2:95-105.

Mitsiades, C. S., Ocio, E. M., Pandiella, A., Maiso, P., Gajate, C., Garayoa, M., Vilanova, D., Montero, J. C., Mitsiades, N., McMullan, C. J., Munshi, N. C., Hideshima, T., Chauhan, D., Aviles, P., Otero, G., Faircloth, G., Mateos, M. V., Richardson, P. G., Mollinedo, F., San-Miguel, J. F., Anderson, K. C. (2008) Aplidin, a marine organism-derived compound with potent antimyelma activity in vitro and in vivo. *Cancer Res*. 68:5216-5125.

Mittal, N., Subramanian, G., Bütikofer, P., Madhubala, R. (2013) Unique posttranslational modifications in eukaryotic translation factors and their roles in protozoan parasite viability and pathogenesis. *Mol&Biochem Parasit.* 187, 21-31.

Monzote, L. (2009) Current treatment of leishmaniasis: a review. *Open Antimicrob. Agents J*. 1:9-19.

Mountain, V. (2013) Astrex, Structural Genomix, and Syrrx. *Chem. Biol*. 10, 95-98.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. y Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*, 67, 355–367.

Murthi, A., Shaheen, H. H., Huang, H. Y., Preston, M. A., Lai, T. P., Phizicky, E. M., Hopper, A. K. (2010) Regulation of tRNA bidirectional nuclear-cytoplasmic trafficking in Saccharomyces cerevisiae. *Mol. Biol. Cell.* 21,639–649.

Negrutskii, B. S., Novosylna, O. V., Porubleva, L. V., Vislovukh, A. A. (2018) Control of the amount and funtionality of the eEF1A1 and eEF1A2 isoforms in mammalian cells. *Biopol. and Cells.* 34(6):411-425.

Negrutskii, B. S., Stapulionis, R. and Deutscher, M. P. (1994) Channelling of aminoacyl-tRNA for protein synthesis *in vivo*. *Proc. Natl. Acad. Sci. USA.* 91(3): 964-968.

Negrutskii, B., Vlasenko, D., El'skaya, A. (2012) From global phosphoproteomics to individual proteins: the case of translation elongation factor eEF1A. *Expert Rev. Proteomics.* 9(1); 71-83.

Nelson, D., Loomis, L. and Fischetti, V. A. (2001) Prevention and elimination colonization of mice by group A streptococcus by using a bacteriophage lytic enzyme. *Proc. Natl. Acad. Sci. USA*. 98:4107-4112.

Nelson, D., Schuch, R., Chahales P., Zhu, S., Fischetti, V. A. (2006) PlyC: a multimeric bacteriophage lysin. *Proc. Natl. Acad. Sci. USA.* 103:10765–10770.

Nicholls, R. A., Long, F. and Murshdov, G. N. (2012) Low-resolution refinement tools in REFMAC5. *Acta Crystallogr D Biol Crystallogr*. D68, 404–417.

Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B. F., Nyborg, J. (1995) Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science*. 270(5241):1464-72.

Novosylna, O., Doyle, A., Vlasenko, D., Murphy, M., Negrutskii, B., El'skaya, A. (2017) Comparison of the avility of mammalian eEF1A1 and its oncogenic variant eEF1A2 to interact with actin and calmodulin. *Biol. Chem,*.398(1):113-124.

Novosylna, O., Jurewicz, E., Pydiura, N., Goral, A., Filipek, A., Negrutskii, B., El'skaya, A. (2015) Translation elongation factor eEF1A1 is a novel partner of a multifunctional protein Sgt1. *Biochimie.* 119:137-145.

Novosylna, O. V., Timchenko, A. A., Tiktopulo, E. I., Serdyuk, I. N., Negrutskii, B. S., El'skaya, A. V. (2007) Characterization of physical properties of two isoforms of translation elongation factor eEF1A. *Biopolym. Cell.* 23(5), 386-390.

Olin-Sandoval, V., Moreno-Sánchez, R. and Saavedra, E. (2010) Targeting trypanothione metabolism in trypanosomatid human parasites. *Curr. Drug Targ.* 11, 1614-1630.

Oliveira, H., Vilas Boas, D., Mesnage, S., Kluskens, L. D., Lavigne, R., Sillankorva, S., Secundo, F., Azeredo, J. (2016) Structural and enzymatic characterization of ABgp46, a novel phage endolysin with broad anti-Gram-negative bacterial activity. *Front. Microbiol*. 7: 208.

Orsi, R. H. and Wiedmann, M. (2016) Characteristics and distribution of Listeria spp., including Listeria species newly described because 2009. *Appl. Microbiol. Biotechnol*. 100, 5273–5287.

Ozturk, S. B. and Kinzy, T. G. (2008) Guanine nucleotide exchange factor independence of the G-protein eEF1A through novel mutant forms and biochemical properties. *J. Biol. Chem.* 283:23244-23253.

Panasyuk, G., Nemazanyy, I., Filonenko, V., Negrutskii, B., El'skaya, A. (2008) A2 isoform of mammalian translation factor eEF1A displays increased tyrosine phosphorylation and ability to interact with different signalling molecules. *Int. J. Biochem. Cell Biol*. 40(1), 63-71.

Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). Auto-Rickshaw - An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Cryst*. D61, 449-457.

Patterson, S., Jones, D. C., Shanks, E. J., Frearson, J. A., Gilbert, I.H., Wyatt, P.G. (2009) Synthesis and evaluation of 1-(1-(Benzo[b]thiophen-2-yl)cyclohexyl)piperidine (BTCP) analogues as inhibitors of trypanothione reductase. *ChemMedChem*. 4(8):1341-53.

Pelay-Gimeno, M., Glas, A., Koch, O., Grossmann, T. N. (2015) Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes. *Angew. Chem. Int. Ed. Engl*. 54(31):8896-927.

Pérez-Dorado, I., Campillo, N. E., Monterroso, B., Hesek, D., Lee, M., Páez, J. A., García, P., Martínez-Ripoll, M., García, J. L., Mobashery, S., Menéndez, M., Hermoso, J. A. (2007) Elucidation of the molecular recognition of bacterial cell wall by modular pneumococcal phage endolysin CPL-1. *J. Biol. Chem.* 282:24990-24999.

Petrushenko, Z. M., Budkevich, T. V., Shalak, V. F., Negruskii, B. S., El'skaya,A. V. (2002) Novel complexes of mammalian translation elongarion dator eE1A·GDP with uncharged tRNA and amino-acyl-tRNA synthetase. Implications for tRNA channeling. *Eur. J. Biochem.* 269(19):4811-4818.

Pinke, D. E., Kalloger, S. E., Francetic, T., Huntsman, D. G., Lee, J. M. (2008) The prognostic significance of elongation factor eEF1A2 in ovarian cancer. *Gynecol. Oncol.* 108(3):561-568.

Pittman, Y: R., Kandl, K., Lewis, M., Valente, L., Kinzy, T. G. (2009) Coordination of Eukaryotic Translation Elongation Factor 1A (eEF1A) Function in Actin Organization and Translation Elongation by the Guanine Nucleotide Exchange Factor eEF1Bα*. *J. Biol. Chem.* 284(7): 4739–4747.

Pittman, Y. R., Valente, L., Jeppesen, M. G., Andersen, G. R., Patel, S., Kinzy, T. G. (2006) Mg2+ and a key lysine modulate exchange activity of eukaryotic translation elongation factor 1B alpha. *J Biol Chem.* 281(28):19457-68.

Ponasik, J. A., Strickland, C., Faerman, C., Savvides, S., Karplus, P. A., Ganem, B. (1995) Kukoaminea and other hydrophobic acylpolyamines-potent and selective inhibitors of *Crithidia-fasciculata* trypanothione reductase. *Biochem. J.*, 311: 371-375.

Pucciarelli, M. G., Bierne, H. and Portillo F.G. (2007) The Cell Wall of Listeria monocytogenes and its Role in Pathogenicity. In: Goldfine H., Shen H. (eds) Listeria monocytogenes: Pathogenesis and Host Response. Springer, Boston, MA

Quijano, C., Trujillo, M., Castro, L., Trostchansky, A. (2016) Interplay between oxidant species and energy metabolism. *Redox Biol.* 8:28-42.

Rawle, D. J., Li, D., Wu, Z., Wang, L., Choong, M., Lor, M., Reid, R. C., Fairlie, D. P., Harris, J., Tachedjian, G., Poulsen, S. A., Harrich, D. (2019) Oxazole-benzenesulfonamide derivatives inhibit HIV-1 reverse transcriptase interaction with cellular eEF1A and reduce viral replication. *J. Virol.* doi: 10.1128/JVI.00239-19

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst*. A42, 140–149.

Reithinger, R., Dujuardin, J. C., Louzie, H., Pirmez, C., Alexander, B., Brooker, S. (2007) Cutaneous leishmaniasis. *Lancet Infect. Dis.* 7(9):581-596.

Richardson, J. L., Nett, I. R. E., Jones, D. C., Abdille, M. H., Gilbert, I. H., Fairlamb, A. H. (2009) Improved tricyclic inhibitors of trypanothione reductase by screening and chemical synthesis. *ChemMedChem*, 4: 1333-1340.

Roach, D. R. and Donovan, D. M. (2015) Antimicrobial bacteriophage-derived proteins and therapeutic applications. *Bacteriophage.* 5:e1062590.

Roach, D. R., Leung, C. Y., Henry, M., Morello, E., Singh, D., Di Santo, J. P., Weitz, J. S., Debarbieux, L. (2017) Synergy between the host immune system and bacteriophage is essential for successful phage therapy against an acute respiratory pathogen. *Cell Host Microbe* 22:38–47.

Rodrigues J.C.F., Godinho J.L.P., de Souza W. (2014) Biology of human pathogenic trypanosomatids: epidemiology, lifecycle and ultrastructure. Chapter 1 in: Santos A., Branquinha M., d'Avila-Levy C., Kneipp L., Sodré C. (eds) Proteins and Proteomics of

Leishmania and Trypanosoma. Subcellular Biochemistry, vol 74. Springer, Dordrecht. DOI 10.1007/978-94-007-7305-9_1

Romão, P.R., Tovar, J., Fonseca, S.G., Moraes, R.H., Cruz, A.K., Hothersall, J.S. (2006) Glutathione and the redox control system trypanothione/trypanothione reductase are involved in the protection of Leishmania spp. against nitrosothiol-induced cytotoxicity. *Braz J Med Biol Res.* 39(3):355-63.

Rosenberry, T. L., Krall, J. A., Dever, T. E., Haas, R., Louvard, D., and Merrick, W. C. (1989) Biosynthetic incorporation of [3H] ethanolamine into protein synthesis elongation factor 1 alpha reveals a new post-translational protein modification. J. Biol. Chem. 264, 7096-7099.

Rubinson, K. A., Ladner, J. E., Tordova, M. y Gilliland, G. L. (2000) Cryosalts: suppression of ice formation in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, 56, 996–1001.

Ruest, L. B., Marcotte, R., and Wang, E. (2002) Peptide elongation factor eEF1A2/S1 expression in cultured differenciated myotubes and its protective effect against caspase-3-mediated apoptosis. *J. Biol. Chem.* 277,5418–5425.

Rupp, B. (2010) Biomolecular crystallography: principles, practice and application to structural biology. Garland Science.

Saha, S. K., and Chakraburtty, K. Protein synthesis in yeast. Isolation of variant forms of elongation factor 1 from the yeast Saccharomyces cerevisiae. (1986) *J. Biol. Chem*. 261, 12599–12603.

Salmon-Chemin, L., Buisine, E., Yardley, V., Kohler, S., Debreu, M. A., Landry, V., Sergheraert, C., Croft, S., Krauth-Siegel, R. L., Davioud-Charvet, E. (2001) 2- and 3-substituted 1,4-naphthoquinone derivatives as subversive substrates of trypanothione reductase and lipoamide dehydrogenase from Trypanosoma cruzi: synthesis and correlation between redox cycling activities and in vitro cytotoxicity. *J. Med. Chem.* 44(4):548-565.

Sánchez-Murcia, P. A., Cortés-Cabrera, A. and Gago, F (2017) Structural rationale for the cross-resistance of tumor cells bearing the A399V variant of elongation factor eEF1A1 to the structurally unrelated didemnin B, ternatin, nannocystin A and ansatrienin B. *J. Comp. Aided Mol. Design*. 31(10) 915–928.

Sanges, C., Scheuermann, C., Zahedi, R. P., Sickmann, A., Lamberti, A., Migliaccio, N., Baljuls, A., Marra, M., Zappavigna, S., Rapp, U., Abbruzzese, A., Caraglia, M., Arcari, P. (2012) Raf kinases mediate the phosphorylation of eukaryotic translation elongation factor 1A and regulate its stability in eukaryotic cells. Cell Death and Disease. 3, e276.

Sasikumar, A. N., Perez, W. B., Kinzy, T. G. (2012) The many roles of the Eukaryotic Elongation Factor 1 Complex. *Wiley Interdiscip Rev RNA.* 3(4):543-555.

Schlaeger, C., Longerich, T., Schiller, C., Bewerunge, P., Mehrabi, A., Toedt, G., Kleeff, J., Ehemann, V., Eils, R. Lichter, P., Schirmacher, P., Radlwimmer, B. (2008) Etiology-dependent molecular mechanisms in human hepatocarcinogenesis. *Hepatology.* 47(2):511-520.

Schleifer, K. H. and Kandler, O. (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.* 36(4):407-477.

Schmelcher, M., Donovan, D. M., and Loessner, M. J. (2012) Bacteriophage endolysins as novel antimicrobials. *Future Microbiol*. 7, 1147–1171.

Schmelcher, M., Waldherr, F. and Loessner, M. J. (2012a) Listeria bacteriophage peptidoglycan hydrolases feature high thermoresistance and reveal increased activity after divalent metal cation substitution. *Appl. Microbiol. Biotechnol.* 93: 633-643.

Schmelcher, M. and Loessner, M. J. (2016) Bacteriophage endolysins: applications for food safety. *Curr. Op. Biotech*. 37:76-87.

Schmelcher, M., Powell, A. M., Camp, M. J., Pohl, C. S., Donovan, D. M. (2015) Synergistic streptococcal phage lambdaSA2 and B30 endolysins kill streptococci in cow milk and in a mouse model of mastitis. *Appl. Microbiol. Biotechnol.* 99:8475-8486.

Schmelcher, M., Shabarova, T., Eugster, M. R., Eichenseher, F., Tchang, V. S., Banz, M., Loessner, M. J. (2010) Rapid multiplex detection and differentiation of Listeria cells by use of fluorescent phage endolysin cell wall binding domains. *Appl. Environ. Microbiol*. 76, 5745–5756.

Schmelcher, M., Shen, Y., Nelson, D. C., Eugster, M. R., Eichenseher, F., Hanke, D. C., Loessner, M. J., Dong, S., Pritchard, D. G., Lee, J. C., Becker, S. C., Foster-Frey, J., Donovan, D. M. (2015) Evolutionarily distinct bacteriophage endolysins featuring conserved peptidoglycan cleavage sites protect mice from MRSA infection. *J. Antimicrob. Chemother.* 70: 1453-1465.

Schmidt, A. and Krauth-Siegel, R. L. (2002) Enzyms of the trypanothione metabolism as targets for antitrypanosomal drug development. *Curr. Top Med. Chem.* 2:1239-1259.

Schuch, R., Lee, H. M., Schneider, B. C., Sauve, K. L., Law, C., Khan, B. K., Rotolo, J. A., Horiuchi, Y., Couto, D. E., Raz, A., Fischetti, V. A., Huang, D. B., Nowinski, R. C., Wittekind, M. (2014) Combination therapy with lysin CF-301 and antibiotic is superior to antibiotic alone for treating methicillin-resistant Staphylococcus aureus-induced murine bacteremia. *J. Infect. Dis*. 209: 1469-1478.

Schummer, T., Gromadski, K. B. and Rodnina, M. V. (2007) Mechanism of RF-Ts-catalyzed guanine nucleotide exchange in EF-Tu: contribution of interactions mediated by helix B of EF-Tu. *Biochemistry*. 46, 4977-4984.

Seddon, G., Lounnas, V., McGuire, R., van den Bergh, T., Bywater, R. P., Oliveira, L., Vriend, G. (2012) Drug design for ever, from hype to hope. J. Comput. *Aided Mol. Des*. 26(1):137-150.

Seeliger, H. P. R. and Höhne, K. (1979) Serotyping of Listeria monocytogenes and related species. *Methods Microbiol.* 13:31-49.

Sham, H. L., Kempf, D. J., Molla, A., Marsh, K. C., Kumar, G. N., Chen, C. M., Kati, W., Stewart, K., Lal, R., Hsu, A., Betebenner, D., Korneyeva, M., Vasavanonda, S., McDonald, E., Saldivar, A., Wideburg, N., Chen, X., Niu, P., Park, C., Jayanti, V., Grabowski, B.,

Granneman, G. R., Sun, E., Japour, A. J., Leonard, J. M., Plattner, J. J., Norbeck, D. W. (1998) *Antimicrob. Agents Chemother*. 42(12):3218-3224.

Shao, S., Murray, J., Brown, A., Taunton, J., Ramakrishnan, V., Hegde, R. S. (2016) Decoding Mammalian Ribosome-mRNA States by Translational GTPase Complexes. *Cell.* 167(5), 1229-1240.

Sharlow, E. R., Leimgruber, S., Murray, S., Lira, A., Sciotti, R. J., Hickman, M., Hudson, T., Leed, S., Caridha, D., Barrios, A. M., Close, D., Grögl, M., Lazo, J. S. (2014) Auranofin is an apoptosis-simulating agent with in vitro and in vivo anti-leishmanial activity. *ACS Chem Biol*. 9(3):663-672.

Shen, Y., Boulos, S., Sumrall, E., Gerber, B., Julian-Rodero, A., Eugster, M. R., Fieseler, L., Nyström, L., Ebert, M. O., Loessner, M. J. (2017) Structural and functional diversity in Listeria cell wall teichoic acids. *J. Biol. Chem.* 292(43): 17832–17844.

Signorell, A., Jelk, J., Rauch, M. and Bütikofer, P. (2008) Phosphatidylethanolamine is the precursor of the ethanolamine phosphoglycerol moiety bound to eukaryotic elongation factor 1A. *J. Biol. Chem.* 283, 20320-20329.

Singh, P. K., Donovan, D. M., Kumar, A. (2014) Intravitreal injection of the chimeric phage endolysin Ply187 protects mice from Staphylococcus aureus endophthalmitis. *Antimicrob. Agents Chemother*. 58: 4621-4629.

Singh, K., Garg, G., Ali, V. (2016) Current therapeutics, their problems and thiol metabolism as potential drug targets in Leishmaniasis. *Curr. Drug Metab*. 17:1-23.

Smith, S. J. and Rittinger, K. (2012) Preparation of GTPases for structural and biophysical analysis. *Methods Mol. Biol*. 189:13-24.

Soares, D. C. and Abbott, C. M. (2013) Highly homologous eEF1A1 and eEF1A2 exhibit differential post-translational modification with significant enrichment around localised sites of sequence variation. *Biol. Direct.* 8, 29.

Spinks, D., Shanks, E. J., Cleghorn, L. A. T., McElroy, S., Jones, D., James, D., Fairlamb, A. H., Frearson, J. A., Wyatt, P. G., Gilbert, I. H. (2009) Investigation of trypanothione reductase as a drug target in Trypanosoma brucei. *ChemMedChem*. 4: 2060-2069.

Stevenson, R. P., Veltman, D., Machesky, L. M. (2012) Actin-bundling proteins in cancer progression at a glance. *J. Cell. Sci.* 125(5) 1073-1079.

Stuart, K., Brun, R., Croft, S., Fairlamb, A. H., Gürtler, R. E., McKerrow, J., Reed,S., Tarleton,R., (2008) Kinetoplastids: related protozoan pathogens, different diseases. *J. Clin. Invest.* 118, 1301-1310.

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P. V., Moriarty, N.W., Zwart, P.H., Hung, L., Read, R.J., Adams, P.D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Cryst*. D64, 61-69.

Timchenko, A. A., Novosylna, O. V., Prituzhalov, E. A., Kihara, H., El'skaya, A. V., Negrutskii, B. S., Serdyuk, I. N. (2013) Different oligomeric properties and stability of highly

homologous A1 and proto-oncogenic A2 variants of mammalian translation elongation factor eEF1. *Biochem.* 52(32):5345-5353.

Thangudu, R. R., Bryant, S. H., Panchenko, A. R., Madej, T. (2012) Modulating protein-protein interactions with small molecules: the importance of binding hotspots. *J. Mol. Biol.* 415:443-453.

Tolba, M., Ahmed, M. U., Tlili, C., Eichenseher, F., Loessner, M. J., Zourob, M. (2012) A bacteriophage endolysin-based electrochemical impedance biosensor for the rapid detection of Listeria cells. *Analyst.* 137, 5749–5756.

Tomás A.M. and Castro H. (2013) Redox metabolism in mitochondria of trypanosomatids. *Antioxid Redox Signal*.19(7):696-707.

Tomlinson, V. A., Newbery, H. J., Bergmann, J. H., Boyd, J., Scott, D., Wray, W. R., Sellar, G. C., Gabra, H., Graham, A., Williams, A. R., Abbott, C. M. (2007) Expression of eEF1A2 is associated with clear cell histology in ovarian carcinomas: overexpression of the gene is not dependent on modifications at the eEF1A2 locus. *Br. J. Cancer*. 96(10):1613-1620.

Tomlinson, V. A., Newbery, H. J., Wray, N. R., Jackson, J., Larionov, A., Miller, W. R., Dixon, J. M., Abbott, C. M. (2005) Translation elongation factor eEF1A2 is a potential oncoprotein that is overexpressed in two-thirds of breast tumours. *BMC Cancer.* 5:113.

Toro, M. A., Sánchez- Murcia, P. A., Moreno, D., Ruiz-Santaquiteria, M., Fernando Alzate, J., Negri, A., Camarasa, M. J., Gago, F., Velázquez, S., Jiménez-Ruiz, A. (2013) Probing the dimerization interface of *Leishmania infantum* trypanothione reductase with site directed mutagenesis and short peptides. *Chem. Bio. Chem.* 14:1212-1217.

Toro, M. A. 2017. Diseño y evaluación de inhibidores peptídicos dirigidos a la interfaz de dimerización de la tripanotión reductasa. – Universidad de Alcalá.

Totté, J., de Wit, J., Pardo, L., Schuren, F., van Doorn, M., Pasmans, S. (2017) Targeted anti-staphylococcal therapy with endolysins in atopic dermatitis and the effect on steroid use, disease severity and the microbiome: study protocol for a randomized controlled trial (MAAS trial). *Trials.* 18:404.

Uchikawa, K., Sekikawa, I. and Azuma, I. (1986) Structural studies on teichoic acids in cell walls of several serotypes of Listeria monocytogenes. *J. Biochem.* 99: 315–327.

Vagin, A. A. and Teplyakov, A. (1997) MOLREP: an automated program for Molecular Replacement. *J. Appl. Cryst*. 30, 1022 – 1025.

Vázquez-Boland, J. A., Kuhn, M., Berche, P., Chakraborty, T., Domínguez-Bernal, G., Goebel, W., González-Zorn, T., Wehland, J., Kreft, J. (2001) Listeria pathogenesis and molecular virulence determinants. *Clin. Microb. Rev*. 584-640.

Valko, M., Leibfritz, D., Moneo, J., Cronin, M.T., Mazur, M., Telser, J. (2007) Free radicals and antioxidants in normal physiologica1 functions and human disease. *Int. J. Biochem. Cell Biol.* 39: 44-84.

Viertel, T. M., Titter, K. and Horz, H-P. (2014) Viruses versus bacteria—novel approaches to phage therapy as a tool against multidrug-resistant pathogens. *J Antimicrob. Chemother*. 69:2326 –2336.

Vilar-Pereira, G., de Souza-Ruivo, L. A., and Lannes-Vieira, J. (2015) Behavioural alterations are independent of sickness behaviour in chronic experimental Chagas disease. *Mem. Inst. Oswaldo Cruz.* 110(8):1042-1050.

Vitagliano, L., Ruggiero, A., Masullo, M., Cantiello, P., Arcari, P., Zagari, A. (2004) The cristal structure of Sulfolobus solfataricus elongation factor 1 alpha in complex with magnesium and GDP. *Biochem.* 43:6630-6636.

Walcher, G., Stessl, B., Wagner, M., Eichenseher, F., Loessner, M. J., Hein, I. (2010) Evaluation of paramagnetic beads coated with recombinant Listeria phage endolysin-derived cell-wall-binding domain proteins for separation of Listeria monocytogenes from raw milk in combination with culture-based and real-time polymerase chain reaction-based quantification. *Foodborne. Pathog. Dis*. 7, 1019–1024.

Walton, J. G. A., Jones, D. C., Kiuru, P., Durie, A. J., Westwood, N. J., Fairlamb, A. H. (2011) Synthesis and evaluation of indatraline-based inhibitors for trypanothione reductase. *ChemMedChem.* 6: 321-328.

Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., Lander, E. S., Sabatini, D. M. (2015) Identification and characterization of essential genes in the human genome. *Science.* 350(6264):1096-1101.

Wang, Q., Euler, C. W., Delaune, A., Fischetti, V. A. (2015) Using a novel lysin to help control Clostridium difficile infections. *Antimicrob. Agents Chemother.* 59: 7447-7457.

Wang, I., Smith, D. L. and Young, R. (2000) HOLINS: The Protein Clocks of Bacteriophage. *Ann. Rev. Microbiol.* 54: 799–825.

Wells, J. A. and McClendon, C. L. (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature.* 450:1001-1009.

Whiteheart, S. W., Shenbagamurthi, P., Chen, L., Cotter, R. J. and Hart, G. W. (1989) Murine elongation factor 1alpha is posttranslationally modified by novel amide-linked ethanolamine-phosphoglycerol moieties. *J. Biol. Chem.* 264, 14334-14341.

WHO. World Health Organization, 2019. Available at https://www.who.int/ (Accessed April 2019).

Witzenrath, M., Schmeck, B., Doehn, J. M., Tschernig, T., Zahlten, J., Loeffler, J. M., Zemlin, M., Müller, H., Gutbier, B., Schütte, H., Hippenstiel, S., Fischetti, V. A., Suttorp, N., Rosseau, S. (2009) Systemic use of the endolysin Cpl-1 rescues mice with fatal pneumococcal pneumonia. *Crit. Care Med.* 37: 642-649.

Xu, C., Hu, D. M. amd Zhu, Q. (2013) eEF1A2 promotes cell migration, invasion and metastasis in pancreatic cancer by upregulating MMP-9 expression through Akt activation. Clin. Exp. *Metastasis*. 30(7):933-944.

Yang, S., Lu, M., Chen, Y., Meng, D., Sun, R., Yun, D., Zhao, Z., Lu, D., Li, Y. (2015) Overexpression of eukaryotic elongation factor 1 alpha-2 is associated with poorer prognosis in patients with gastric cancer. *J. Cancer Res. Clin. Oncol.* 141:1265–1275

Yang, H., Wang, M., Yu, J., Wei, H. (2015) Antibacterial activity of a novel peptide-modified lysin against Acinetobacter baumannii and Pseudomonas aeruginosa. *Front. Microbiol.* 6:1471.

Yang, H., Wang, D. B., Dong, Q., Zhang, Z., Cui, Z., Deng, J., Yu, J., Zhang, X. E., Wei, H. (2012) Existence of separate domains in lysin PlyG for recognizing Bacillus anthracis spores and vegetative cells. *Antimicrob. Agents Chemother.*56: 5031-5039.

Yang, H., Zhang, Y., Yu, J., Huang, Y., Zhang, X. E., Wei, H. (2014) Novel chimeric lysin with high-level antimicrobial activity against methicillin-resistant Staphylococcus aureus in vitro and in vivo. *Antimicrob. Agents Chemother*. 58: 536-542

Yaremchuk, A., Shalak, V. F., Novosylna, O. V., Negrutskii, B. S., Crepin, T., El'skaya, A. V., Tukalo, M. (2012) Purification, crystallization and preliminary X-ray crystallographic analysis of mammalian translation elongation factor eEF1A2. *Acta Cryst. F Struct. Biol. Cryst. Commun*. 68(3), 295-297.

Young, R. (1992) Bacteriophage lysis: mechanism and regulation. *Microbiol. Rev.* 56: 430–481.

Young, R. (2013) Phage lysis: do we have the whole story yet? *Curr. Opin. Microbiol.* 16:790 – 797.

Zhang, T., Bond, C. S., Bailey, S., Cummingham, M. L., Failamb, A. H., Hunter, W. N. (1996) The crystal structure of trypanothione reductase from the human pathogen *Trypanosoma cruzi* at 2.3 Å resolution. *Prot. Sci.* 5, 52-61.

Zhu, H., Lam, D. C., Han, K. C., Tin, V. P., Suen, W. S., Wang, E., Lam, W. K., Cai, W. W., Chung, L.P., Wonga, M. P. (2007) High resolution analysis of genomic aberrations by metaphase and array comparative genomic hybridization identifies candidate tumour genes in lung cancer cell lines. *Cancer Lett.* 245(1-2):303-314.

Zimmer, M., Sattelberger, E., Inman, R.B., Calendar, R., and Loessner, M.J. (2003) Genome and proteome of Listeria monocytogenes phage PSA: an unusual case for programmed + 1 translational frameshifting in structural protein synthesis. *Mol. Microbiol.* 50: 303–317.

Zimmermann, J., Buchdunger, E., Mett, H., Meyer, T., Lydon, N. (1997) Potent and selective inhibitors of the Abl-kinase phenylamino-pyrimidine (PAP) derivatives. *Bioorg. Medicinal Chem.* 187-192.

Zobel-Thropp, P., Yang, M. C., Machado, L. and Clarke, S. (2000) A novel post-translational modification of yeast elongation factor 1A. Methylesterification at the C terminus. J. Bio. Chem. 275(47), 37150-37158.

SUPPORTING INFORMATION

```
E.coli        MSKEKFERTKPHVNVGTIGHVDHGKTTLTAAITTVLAKT--------------YGGA-AR  45
Archaea       ---GSHMAEKPHMNLVVIGHVDHGKSTLVGHLLYRLGYIEEKKLKELEEQAKSRGKESFK  57
Rabbit        -----MGKEKTHINIVVIGHVDSGKSTTTGHLIYKCGGIDKRTIEKFEKEAAEMGKGSFK  55
Yeast         -----MGKEKSHINVVVIGHVDSGKSTTTGHLIYKCGGIDKRTIEKFEKEAAELGKGSFK  55
                        * *:*: .***** **:* .. :    .                 *    :

E.coli        AFDQIDNAPEEKARGITINTSHVEYDTPTRHYAHVDCPGHADYVKNMITGAAQMDGAILV  105
Archaea       FAWILDKMKEERERGITIDLTFMKFETKKYVFTIIDAPGHRDFVKNMITGASQADAAILV  117
Rabbit        YAWVLDKLKAERERGITIDISLWKFETTKYYITIIDAPGHRDFIKNMITGTSQADCAVLI  115
Yeast         YAWVLDKLKAERERGITIDIALWKFETPKYQVTVIDAPGHRDFIKNMITGTSQADCAILI  115
                  :*:   *: *****: :  :::* .   : :*.*** *::******::* * *:*:

E.coli        VAATDGPM-------PQTREHILLGRQVGVPYIIVFLNKCDMVDD---EELLELVEMEVR  155
Archaea       VSARKGEFEAGMSTEGQTREHLLLARTMGIEQIIVAVNKMDAPDVNYDQKRYEFVVSVLK  177
Rabbit        VAAGVGEFEAGISKNGQTREHALLAYTLGVKQLIVGVNKMDSTEPAYSEKRYDEIVKEVS  175
Yeast         IAGGVGEFEAGISKDGQTREHALLAFTLGVRQLIVAVNKMDSV--KWDESRFQEIVKETS  173
              ::.  * :        ***** **.  :*:  :** :** *       :.  : :

E.coli        ELLSQYDFPGDDTPIVRGSALK-----ALEGDAEWEAK----I-------LELAGFLDSY  199
Archaeal      KFMKGLGYQVDKIPFIPVSAWKGDNLIERSPNMPWYNG-----------PTLVEALD-Q  224
Rabbit        AYIKKIGYNPATVPFVPISGWHGDNMLEPSPNMPWFKGWKVERKEGNASGVSLLEALD-T  234
Yeast         NFIKKVGYNPKTVPFVPISGWNGDNMIEATTNAPWYKGWEKETKAGVVKGKTLLEAID-A  232
                :.  .:     *::  *. :         :  *                *   :*
```

Figure S1: Protein sequence alignment of N-terminal residues of EF-Tu and eEF1A from different organisms. Those residues involved in the binding of GDP are highlighted in yellow. Conserved residues are marked with (*), similar residues are depicted as (.) or (:).

```
TryR_A      MSRAYDLVVLGAGSGGLEAGWNAAVTYKKKVAVIDAQATHGPPFFAALGGTCVNVGCVPK  60
TryR_B      MSRAYDLVVLGAGSGGLEAGWNAAVTHKKKVAVVDVQATHGPPLFAALGGTCVNVGCVPK  60

TryR_A      KLMVTGAQYMDLIRESGGFGWEMNRESLCPNWKTLIAAKNKVVNDINESYKSMFADTEGL  120
TryR_B      KLMVTGAQYMDLIRESGGFGWEMDRESLCPNWKTLIAAKNKVVNSINESYKSMFADTEGL  120

TryR_A      SFHMGFGALQDAHTVLVRKSEDPNSDVLETLDTEYILIATGSWPTRLGVPGDEFCITSNE  180
TryR_B      SFHMGFGALQDAHTVVVRKSEDPHSDVLETLDTEYILIATGSWPTRLGVPGDEFCITSNE  180

TryR_A      AFYLEDAPKRMLCVGGGYIAVEFAGIFNGYKPRGGYVDLCYPGDLILRGFDTEVRKSLTK  240
TryR_B      AFYLEDAPKRMLCVGGGYIAVEFAGIFNGYKPCGGYVDLCYRGDLILRGFDTEVRKSLTK  240

TryR_A      QLGANGIRVRTNLNPTKITKNEDGSNHVHFNDGTEEDYDQVMLAIGRVPRSQALQLDKVG  300
TryR_B      QLGANGIRVRTNLNPTKITKNEDGSNHVHFNDGTEEDYDQVMLAIGRVPRSQALQLDKAG  300

TryR_A      VQTGKNGAVQVDAYSKTSVDNIYAIGDVTNRVMLTPVAINEGAAFVETVFGGKPRATDHT  360
TryR_B      VRTGKNGAVQVDAYSKTSVDNIYAIGDVTNRVMLTPVAINEGAAFVETVFGGKPRATDHT  360

TryR_A      KVACAVFSIPPIGTCGMTEEAAAKNHETVAVYESCFTPLMHNISGSKHKEFMIRIITNQP  420
TryR_B      KVACAVFSIPPIGTCGMTEEEAAKNYETVAVYASSFTPLMHNISGSKHKEFMIRIITNES  420

TryR_A      TGEVLGVHMLGDSAPEIIQSVGICMKMGAKISDFHSTIGVHPTSAEELCSMHTPAYFYES  480
TryR_B      NGEVLGVHMLGDSAPEIIQSVGICMKMGAKISDFHSTIGVHPTSAEELCSMRTPAYFYES  480

TryR_A      GKRVEKLSSNL  491
TryR_B      GKRVEKLSSNL  491
```

Figure S2: Protein sequence alignment of *L. infantum* TryR of different strains. TryR_A: M/CAN/ES/89/IPZ229/1/89 strain, used in initial crystallization trials. TryR_B: M/CAN/ES/96/BCN150/MON-1 strain, used for crystallization purposes.

```
DISTAL          VGYFQDKPQFLNSKSVRQWKHGTKVKLTKHNSH---WYTGVVKDGNKSVR
                                  ......: :.|.|:|   |       .||.:.
PROXIMAL        ------------------MMENINI-VIKDNAHAFCW-------DNKKLN

DISTAL          GYIYHSMAK----VTSKNSDGSVNATI------------------
                   |..:.::.:    :|...|||......
PROXIMAL        GGDFINLKRGFKGITHPASDGFYPLYFASRKKTFYIPRYMFDIKK
```

Figure S3: Alignment between proximal and distal subdomains of CBDP35. Secondary structure elements are highlighted in red.

Table S1: Trypanothione reductase (TryR) PDB entries available for *T. brucei, T. cruzi* and *L. infantum*

	Parasite	PDB Code	Ligands
PDB accession codes for TryR	*L. infantum*	**2JK6**	FAD
		2WOH	FAD, NADPH, Antimony (III) ion
		2X50	FAD, NADPH, Ag^{+}
		2YAU	FAD, NADPH, Auranofin
		4ADW	FAD, NADPH, trypanothione
		4APN	FAD, 4-((1-(4-ethylphenyl)-2-methyl-5-(4-(methylthio)phenyl)-1H-pyrrol-3-yl)methyl)thiomorpholine
		5EBK	FAD, 6-(sec-butoxy)-2-((3-chlorophenyl)thio)pyrimidin-4-amine
		6ER5	FAD, 2-(diethylamino)ethyl4-((3-(4-nitrophenyl)-3oxopropyl)amino)benzoate
		6I7N	FAD, TRL156
	T. brucei	**2WBA**	FAD, NADPH
		2WOI	FAD
		2WOV	FAD, NADPH
		2WOW	FAD, NADPH, trypanothione
		2WP5	FAD, methyl [(4s)-6-bromo-2-methyl-4-phenylquinazolin-3(4h)-yl]acetate
		2WP6	FAD, (4s)-3-benzyl-6-chloro-2-methyl-4-phenyl-3,4-dihydroquinazoline
		2WPC	FAD, (4s)-6-chloro-3-{2-[4-(furan-2-ylcarbonyl)piperazin-1-yl]ethyl}-2-methyl-4-phenyl-3,4-dihydroquinazoline
		2WPE	FAD, n-{2-[(4s)-6-chloro-2-methyl-4-phenylquinazolin-3(4h)-yl]ethyl}furan-2-carboxamide
		2WPF	FAD, 3-[(4s)-6-chloro-2-methyl-4-(4-methylphenyl)quinazolin-3(4h)-yl]-n,n-dimethylpropan-1-amine
		4NEV	FAD, 5-{5-[1-(pyrrolidin-1-yl)cyclohexyl]-1,3-thiazol-2-yl}-1H-indole
		6BTL	FAD, 1-[2-(piperazin-1-yl)ethyl]-5-{5-[1-(pyrrolidin-1-yl)cyclohexyl]-1,3-thiazol-2-yl}-1H-indole
		6BU7	FAD, 1-[2-(piperidin-4-yl)ethyl]-5-{5-[1-(pyrrolidin-1-yl)cyclohexyl]-1,3-thiazol-2-yl}-1H-indole
	T. cruzi	**1AOG**	FAD
		1BZL	FAD, trypanothione
		1GXF	FAD, quinacrine mustard
		1NDA	FAD
		4NEW	FAD, 5-{5-[1-(pyrrolidin-1-yl)cyclohexyl]-1,3-thiazol-2-yl}-1H-indole

Table S2: Residues involved in hydrogen bond formation in TryR dimerization interface.

Residue+Atom Name	Distance (Å)	Residue+Atom Name
B:LYS 61[NZ]	2.85	A:PRO 462[O]
B:TYR 69[OH]	3.57	A:GLU 75[OE1]
B:TRP 81[NE1]	2.97	A:GLY 66[O]
B:TRP 81[NE1]	3.68	A:TYR 210[OH]
B:ASN 91[N]	2.71	A:GLY 80[O]
B:ASN 91[ND2]	3.30	A:GLU 82[OE2]
B:TYR 210[OH]	2.72	A:PHE 79[O]
B:SER 433[OG]	3.84	A:SER 433[OG]
B:GLN 439[NE2]	3.65	A:ILE 458[O]
B:CYS 444[N]	3.87	A:CYS 444[SG]
B:CYS 444[SG]	3.85	A:SER 440[O]
B:CYS 444[SG]	3.33	A:CYS 444[SG]
B:VAL 460[N]	2.57	A:GLN 439[OE1]
B:SER 464[N]	3.18	A:GLU 436[OE1]
B:SER 464[N]	3.07	A:GLU 436[OE2]
B:ALA 465[N]	2.98	A:GLU 436[OE1]
B:PRO 462[O]	2.88	A:LYS 61[NZ]
B:GLU 75[OE1]	3.55	A:TYR 69[OH]
B:TYR 210[H]	3.79	A:TRP 81[NE1]
B:GLY 66[O]	2.99	A:TRP 81[NE1]
B:GLY 80[O]	2.71	A:ASN 91[N]
B:GLU 82[OE2]	3.14	A:ASN 91[ND2]
B:PHE 79[O]	2.59	A:TYR 210[OH]
B:ILE 458[O]	3.63	A:GLN 439[NE2]
B:ILE 437[O]	3.23	A:SER 440[OG]
B:CYS 444[SG]	3.82	A:CYS 444[N]
B:SER 440[O]	3.85	A:CYS 444[SG]
B:GLN 439[OE1]	2.67	A:VAL 460[N]
B:GLU 436[OE1]	3.19	A:SER 464[N]
B:GLU 436[OE2]	2.94	A:SER 464[N]
B:GLU 436[OE1]	2.92	A:ALA 465[N]

(Generated by PISA server).

Table S3: Protein Discoverer mass spectrometry data analysis (Mascot 18-83)

Position	Annotated Sequence	Theoretical . MH+ [Da]	Missed Cleavages	Modifications	PSMs	Found in Sample	Ions Score	Confidence	Charge	m/z [Da]	q-Value
21-30	[K].STTTGHLIYK.[C]	1200.57	0	1xPhospho [T23(100)]	2	High	43	High	2	600.78	0.001427
101-129	[K].NMITGTSQADCAVLIVAAGVGEFEAGISK.[N]	2939.45	0	1xMethyl [D110(100)]	73	High	38	High	4	735.62	0.001427
267-273	[R].VETGILR.[P]	801.48	0	1xMethyl [E268(100)]	5	High	38	High	2	401.25	0
155-166	[K].MDSTEPAYSEKR.[Y]	1443.64	1	1xMethyl [E164(100)]	10	High	32	High	3	481.89	0
155-165	[K].MDSTEPAYSEK.[R]	1271.55	0	1xMethyl [E164(100)]	6	High	28	High	2	636.28	0.001427
155-165	[K].MDSTEPAYSEK.[R]	1287.54	0	1xMethyl [E164(100)]	6	High	26	High	2	644.28	0.001723
409-423	[K].PMCVESFSQYPPLGR.[F]	1781.83	0	1xMethyl [E413(100)]	9	High	21	High	2	891.42	0.001723
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2693.23	0	1xGlycerylPE [E301(100)]	29	High	18	High	3	898.42	0.002441
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2677.23	0	1xGlycerylPE [E301(99.5)]	63	High	18	High	3	893.08	0.001427
155-166	[K].MDSTEPAYSEKR.[Y]	1427.65	1	1xMethyl [E164(100)]	2	High	18	High	3	476.55	0.005189
409-423	[K].PMCVESFSQYPPLGR.[F]	1797.83	0	1xMethyl [E413(100)]	5	High	17	High	3	599.95	0.001723
323-330	[R].GNVCGDSK.[S]	850.37	0	1xMethyl [D328(100)]	19	High	16	High	2	425.69	0.004251
396-408	[K].SGDAAIVEMVPGK.[P]	1303.66	0	1xMethyl [E403(100)]	1	High	14	High	2	652.33	0.001427
396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3066.47	0	2xMethyl [E403(100); E413(100)]	4	High	13	High	3	1022.83	0.005189
101-129	[K].NMITGTSQADCAVLIVAAGVGEFEAGISK.[N]	2967.48	0	3xMethyl [D110(100); E122(100); E124(100)]	1	High	11	Medium	3	989.83	0.0122

322-330	[R].RGNVCGDSK.[S]	1006.47	1	1xMethyl [D328(100)]	1	High	9	High	2	503.74	0.008798
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2773.19	0	1xPhospho [S300(97.9)]	4	High	9	High	4	694.05	0.004892
396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3096.48	1	3xMethyl [D398(100); E403(100); E413(100)]	2	High	8	High	4	774.88	0.003945
220-244	[K].EGNASGVSLLEALDTILPPTRPTDK.[P]	2674.33	1	1xPhospho [S224(97.8)]	1	High	5	High	3	892.12	0.004615
331-371	[K].SDPPQEAAQFTSQVIILNHPGQISAGYSPVIDCHTAHIACK.[F]	4472.18	0	1xMethyl [D362(100)]	5	High	4	High	5	895.24	0.001427
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2524.21	0	2xMethyl [E297(99.3); E/D]	4	High	4	Low	3	842.08	0.05323
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2719.28	0	3xMethyl [D306(100); E]	1	High	4	Low	3	907.10	0.05868
181-212	[K].IGYNPATVPFVPISGWHGDNMLEPSPNMPWFK.[G]	3706.73	0	2xMethyl [D199(100); E203(100)]	2	High	2	High	4	927.44	0.002067

* Carbamidomethylation- and oxidation- only modifications are not shown as these modifications are artefacts of mass spectrometry experiments. Peptides are sorted according to their Ions score.

Table S4: Protein Discoverer mass spectrometry data analysis (Mascot 18-83 Semitarget)

Positions	Annotated Sequence	Theoretical MH+ [Da]	Missed Cleavages	Modifications	PSM	Found in Sample	Ions Score	Confidence	Charge	m/z [Da]
101-129	[K].NMITGTSQADCAVLIVAAGVGEFEAGISK.[N]	2939.45	0	1xMethyl [D110(100)]	73	High	44	High	3	980.49
267-273	[R].VETGILR.[P]	801.48	0	1xMethyl [E268(100)]	5	High	24	High	2	401.24
155-166	[K].MDSTEPAYSEKR.[Y]	1443.64	1	1xMethyl [E164(100)]	10	High	22	High	3	481.89
323-330	[R].GNVCGDSK.[S]	850.37	0	1xMethyl [D328(100)]	19	High	21	High	2	425.69
409-423	[K].PMCVESFSQYPPLGR.[F]	1797.83	0	1xMethyl [E413(100)]	5	High	19	High	3	599.95
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2693.23	0	1xGlycerylPE [E301(100)]	29	High	14	Low	4	674.06

* Carbamidomethylation- and oxidation- only modifications are not shown as these modifications are artefacts of mass spectrometry experiments. Peptides are sorted according to their Ions score.

Table S5: Protein Discoverer mass spectrometry data analysis (Sequest HT 18-83)

Position	Annotated Sequence	Theoretical MH+ [Da]	Missed Cleavages	Modifications	PSM	Found in Sample	XCorr	Confidence	Charge	m/z [Da]	q-Value
331-371	[K].SDPPQEAAQFTSQVIILNHPGQISAGYSPVIDCHTAHIACK.[F]	4472.18	0	1xMethyl [D362(100)]	5	High	3.2	High	5	895.24	0.0005681
155-172	[K].MDSTEPAYSEKRYDEIVK.[E]	2203.02	2	1xAcetyl [K165(100)]	1	High	2.33	High	4	551.51	0.002381
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2693.23	0	1xGlycerylPE [E301(100)]	29	High	2.32	High	3	898.42	0.0005681
101-129	[K].NMITGTSQADCAVLIVAAGVGEFEAGISK.[N]	2939.45	0	1xMethyl [D110(100)]	73	High	2.22	High	4	735.62	0.0005681
155-165	[K].MDSTEPAYSEK.[R]	1287.54	0	1xMethyl [E164(100)]	6	High	2.21	High	2	644.27	0.0007298
155-166	[K].MDSTEPAYSEKR.[Y]	1443.64	1	1xMethyl [E164(100)]	10	High	2.12	High	3	481.89	0
21-30	[K].STTTGHLIYK.[C]	1200.57	0	1xPhospho [T23(100)]	2	High	2.07	High	2	600.78	0.0007298
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2677.23	0	1xGlycerylPE [E301(99.5)]	63	High	1.96	High	3	893.08	0
396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3080.45	0	1xAcetyl [K408(100)]	13	High	1.94	High	4	770.87	0.0005681
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2524.21	0	2xMethyl [E297(99.3); E/D]	4	High	1.94	High	3	842.08	0.0007298
396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3078.47	0	1xAcetyl [K408(100)]	5	High	1.93	High	4	770.37	0.0005681
155-166	[K].MDSTEPAYSEKR.[Y]	1427.65	1	1xMethyl [E164(100)]	2	High	1.82	High	3	476.55	0.0007298
101-129	[K].NMITGTSQADCAVLIVAAGVGEFEAGISK.[N]	2967.48	0	3xMethyl [D110(100); E122(100); E124(100)]	1	High	1.74	High	3	989.83	0.0007298
52-62	[K].GSFKYAWVLDK.[L]	1355.70	1	1xAcetyl [K55(100)]	1	High	1.66	High	3	452.57	0.005586
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2757.20	0	1xPhospho [S300(100)]	2	High	1.65	High	3	919.74	0.0007298

396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3174.43	0	1xMethyl [E413(98.5)]; 1xAcetyl [K408(100)]	10	High	1.6	High	3	1058.83	0.0005681
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2719.28	0	3xMethyl [D306(100); E]	1	High	1.57	High	3	907.10	0.0007298
267-290	[R].VETGILRPGMVVTFAPVNITTEVK.[S]	2782.48	0	1xMethyl [E268(100)]; 1xGlycerylPE [E288(100)]	2	High	1.55	High	4	696.37	0.0007298
155-165	[K].MDSTEPAYSEK.[R]	1271.55	0	1xMethyl [E164(100)]	6	High	1.46	High	2	636.28	0.0007298
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2522.23	0	3xMethyl [E293(99.3); E297(100); E301(100)]	1	High	1.43	High	3	841.41	0.001063
314-321	[K].NVSVKDIR.[R]	986.56	1	1xMethyl [D319(100)]; 1xAcetyl [K318(100)]	9	High	1.41	High	3	329.53	0.0007298
396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3094.47	0	1xMethyl [E413(99.4)]; 1xAcetyl [K408(100)]	40	High	1.39	High	3	1032.16	0.0005681
396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3096.45	0	1xAcetyl [K408(100)]	5	High	1.32	High	3	1032.82	0.0005681
396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3066.47	0	2xMethyl [E403(100); E413(100)]	4	High	1.28	High	3	1022.83	0.0007298
396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3110.46	0	1xAcetyl [K408(100)]	12	High	1.19	High	4	778.37	0.0005681
219-247	[R].KEGNASGVSLLEALDTILPPTRPTDKPLR.[L]	3116.73	1	2xMethyl [E220(100); E230(100)]	1	High	1.18	High	4	779.93	0.0007298
155-166	[K].MDSTEPAYSEKR.[Y]	1471.64	1	1xAcetyl [K165(100)]	1	Not Found	1.11	Medium	3	491.22	0.04889
396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3092.49	0	1xAcetyl [K408(100)]	2	High	1.11	High	3	1031.49	0.004965
323-330	[R].GNVCGDSK.[S]	850.37	0	1xMethyl [D328(100)]	19	High	1.09	High	2	425.69	0.0007298
322-330	[R].RGNVCGDSK.[S]	1006.47	1	1xMethyl [D328(100)]	1	High	1.08	Medium	2	503.74	0.04504
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2773.19	0	1xPhospho [S300(97.9)]	4	High	1.05	High	3	925.07	0.0007298
396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3106.50	0	3xMethyl [D398(100); E403(100); E413(100)]; 1xAcetyl [K408(100)]	1	High	1.05	High	3	1036.16	0.0007298

393-423	[K].SLKSGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3500.67	1	1xAcetyl [K395(98.5)]	1	High	0.98	High	4	875.93	0.003027
219-247	[R].KEGNASGVSLLEALDTILPPTRPTDKPLR.[L]	3130.71	1	1xAcetyl [K219(100)]	1	High	0.94	High	4	783.44	0.0007298
167-172	[R].YDEIVK.[E]	846.36	0	1xPhospho [Y167(100)]	1	High	0.94	High	1	846.37	0.0007298
181-212	[K].IGYNPATVPFVPISGWHGDNMLEPSPNMPWFK.[G]	3706.73	0	2xMethyl [D199(100); E203(100)]	2	High	0.63	Medium	3	1236.25	0.01476
220-247	[K].EGNASGVSLLEALDTILPPTRPTDKPLR.[L]	3002.62	0	1xAcetyl [K244(100)]	1	High	0.62	High	3	1001.54	0.006208

* Carbamidomethylation- and oxidation- only modifications are not shown as these modifications are artefacts of mass spectrometry experiments. Peptides are sorted according to their XCorr value.

Table S6: Protein Discoverer mass spectrometry data analysis (Sequest HT 18-83 Semitarget)

Position	Annotated Sequence	Theo. MH+ [Da]	Missed Cleavages	Modifications	PSMs	Found in Sample	XCorr	Confidence	Charge	m/z [Da]
155-166	[K].MDSTEPAYSEKR.[Y]	1443.64	1	1xMethyl [E164(100)]	10	High	3.06	High	3	481.89
291-313	[K].SVEMHHEALSEALPGDNVGFNVK.[N]	2693.23	0	1xGlycerylPE [E301(100)]	29	High	2.98	High	3	898.42
101-129	[K].NMITGTSQADCAVLIVAAGVGEFEAGISK.[N]	2939.45	0	1xMethyl [D110(100)]	73	High	2.67	High	3	980.49
155-166	[K].MDSTEPAYSEKR.[Y]	1471.64	1	1xAcetyl [K165(100)]	1	High	2.36	High	3	491.22
396-423	[K].SGDAAIVEMVPGKPMCVESFSQYPPLGR.[F]	3110.46	0	1xAcetyl [K408(100)]	12	Not Found	1.93	Medium	4	778.37
314-321	[K].NVSVKDIR.[R]	986.56	1	1xMethyl [D319(100)]; 1xAcetyl [K318(100)]	9	Not Found	1.91	Medium	3	329.53
323-330	[R].GNVCGDSK.[S]	850.37	0	1xMethyl [D328(100)]	19	High	1.25	Medium	2	425.69
52-62	[K].GSFKYAWVLDK.[L]	1355.70	1	1xAcetyl [K55(100)]	1	Not Found	1.15	Medium	3	452.57
155-165	[K].MDSTEPAYSEK.[R]	1287.54	0	1xMethyl [E164(100)]	6	Not Found	0.92	Medium	2	644.27

* Carbamidomethylation- and oxidation- only modifications are not shown as these modifications are artefacts of mass spectrometry experiments. Peptides are sorted according to their XCorr value.

Table S7: PEAKS-Database mass spectrometry data analysis

Position	Peptide	-10logP	Mass [Da]	z	m/z	ppm error	Intensity PTMS	PSM	AScore
135-146	R.EHALLAY(+79.97)TLGVK.Q	85.17	1393.70	2	697.86	-0.1	3.44E+07	4	Y7:Phosphorylation (STY):14.02
155-166	K.MDSTEPAYSE(+14.02)KR.Y	69.45	1426.64	2	714.33	1.6	3.00E+09	3	E10:Methylation(D E):55.60
155-166	K.M(+15.99)DSTEPAYSE(+14.02)KR.Y	69.14	1442.63	3	481.89	0.4	1.65E+10	11	M1:Oxidation (M):1000.00; E10:Methylation(D E):51.00
101-129	K.NM(+15.99)ITGTSQAD(+14.02)C(+57.02)AVLIVAAGVGEFEAGISK.N	67.7	2938.44	3	980.49	3.5	9.47E+08	44	M2:Oxidation (M):1000.00; D10:Methylation(D,E):78.56; C11:Carbamidomethylation:1000.00
396-423	K.SGDAAIVE(+14.02)MVPGK(+42.01)PM(+15.99)C(+57.02)VESFSQYPPLGR.F	65.57	3093.46	4	774.37	-1	2.30E+08	5	E8:Methylation(D E):0.00; K13:Acetylation (K):1000.00; M15:Oxidation (M):0.00; C16:Carbamidomethylation:1000.00
396-423	K.SGDAAIVEM(+15.99)VPGK(+42.01)PM(+15.99)C(+57.02)VESFSQYPPLGR.F	60.42	3095.44	3	1032.82	0.2	1.30E+08	8	M9:Oxidation (M):1000.00; K13:Acetylation (K):1000.00; M15:Oxidation (M):1000.00; C16:Carbamidomethylation:1000.00
155-165	K.M(+15.99)DSTEPAYSE(+14.02)K.R	59.54	1286.53	2	644.27	0.8	1.21E+09	5	M1:Oxidation (M):1000.00; E10:Methylation(D E):39.44
156-166	M.DSTEPAYSE(+14.02)KR.Y	58.82	1295.60	2	648.81	0.2	1.94E+08	2	E9:Methylation(D E):68.60
396-423	K.SGDAAIVE(+14.02)MVPGK(+42.01)PM(+15.99)C(+57.02)VE(+14.02)SFSQYPPLGR.F	57.66	3107.48	4	777.87	-3.1	6.02E+07	2	E8:Methylation(D E):52.68; K13:Acetylation (K):1000.00; M15:Oxidation (M):0.00; C16:Carbamidomethylation:1000.00; E18:Methylation(D E):27.33

355-371	S.AGYSPVID(+14.02)C(+57.02)HTAHIAC(+57.02)K.F	57.57	1912.89	4	479.23	0.4	5.16E+07	2	D8:Methylation(D E):1000.00; C9:Carbamidomethylation:1000.00; C16:Carbamidomethylation:1000.00
396-423	K.SGDAAIVEM(+15.99)VPGK(+42.01)PMC(+57.02)VESFSQYPPLGR.F	57	3079.45	4	770.87	-1	2.80E+08	3	M9:Oxidation (M):0.00; K13:Acetylation (K):1000.00; C16:Carbamidomethylation:1000.00
155-166	K.M(+15.99)DSTEPAYSEK(+42.01)R.Y	56.1	1470.63	3	491.22	-0.3	5.67E+07	4	M1:Oxidation (M):1000.00; K11:Acetylation (K):1000.00
155-165	K.MDSTEPAYSE(+14.02)K.R	55.52	1270.54	2	636.28	-1.9	1.10E+08	2	E10:Methylation(D E):45.16
323-330	R.GNVC(+57.02)GD(+14.02)SK.S	54.08	849.37	2	425.69	0	4.33E+07	19	C4:Carbamidomethylation:1000.00; D6:Methylation(D E):1000.00
396-423	K.SGDAAIVEM(+15.99)VPGK(+42.01)PM(+15.99)C(+57.02)VE(+14.02)SFSQYPPLGR.F	53.94	3109.46	4	778.37	2.4	4.45E+08	10	M9:Oxidation (M):1000.00; K13:Acetylation (K):1000.00; M15:Oxidation (M):1000.00; C16:Carbamidomethylation:1000.00; E18:Methylation(D E):0.00
396-423	K.SGDAAIVEMVPGK(+42.01)PM(+15.99)C(+57.02)VE(+14.02)SFSQYPPLGR.F	53.45	3093.46	4	774.37	1.9	4.20E+08	5	K13:Acetylation (K):1000.00; M15:Oxidation (M):0.00; C16:Carbamidomethylation:1000.00; E18:Methylation(D E):0.00
409-423	K.PMC(+57.02)VE(+14.02)SFSQYPPLGR.F	52.76	1780.83	3	594.62	0.4	8.15E+08	9	C3:Carbamidomethylation:1000.00; E5:Methylation(D E):1000.00
409-423	K.PM(+15.99)C(+57.02)VE(+14.02)SFSQYPPLGR.F	52.7	1796.82	3	599.95	0.2	1.08E+09	10	M2:Oxidation (M):1000.00; C3:Carbamidomethylation:1000.00; E5:Methylation(D E):1000.00
396-423	K.SGDAAIVEMVPGK(+42.01)PM(+15.99)C(+57.02)VESFSQYPPLGR.F	52.37	3079.45	3	1027.49	-3.1	1.28E+08	5	K13:Acetylation (K):1000.00; M15:Oxidation (M):0.00; C16:Carbamidomethylation:1000.00

396-423	K.SGDAAIVEMVPGK(+42.01)PMC(+57.02)VE(+14.02)SFSQYPPLGR.F	52.03	3077.47	4	770.37	-4.9	4.73E+07	2	K13:Acetylation (K):1000.00; C16:Carbamidomethylation:1000.00; E18:Methylation(D E):4.78
85-96	K.YYIT(+79.97)IIDAPGHR.D	51.88	1497.70	2	749.85	-5.8	2.99E+07	1	T4:Phosphorylation (STY):22.45
155-166	K.M(+15.99)DSTEPAYSE(+14.02)K(+42.01)R.Y	51.16	1484.65	3	495.89	-0.3	1.81E+06	1	M1:Oxidation (M):1000.00; E10:Methylation(D E):10.19; K11:Acetylation (K):1000.00
411-423	M.C(+57.02)VE(+14.02)SFSQYPPLGR.F	50.03	1552.73	2	777.38	0.8	6.82E+09	18	C1:Carbamidomethylation:1000.00; E3:Methylation(D E):1000.00
396-423	K.SGDAAIVE(+14.02)M(+15.99)VPGK(+42.01)PMC(+57.02)VE(+14.02)SFSQYPPLGR.F	49.18	3107.48	4	777.88	3.8	6.02E+07	1	E8:Methylation(D E):45.16; M9:Oxidation (M):0.00; K13:Acetylation (K):1000.00; C16:Carbamidomethylation:1000.00; E18:Methylation(D E):25.39
296-313	H.HEALSE(+197.05)ALPGDNVGFNVK.N	48.33	2092.98	3	698.67	0.2	1.53E+08	4	E6:Glycerylphosphorylethanolamine:23.44
296-313	H.HE(+14.02)ALSE(+197.05)ALPGDNVGFNVK.N	47.28	2107.00	3	703.34	0	4.07E+07	2	E2:Methylation(D E):20.17; E6:Glycerylphosphorylethanolamine:20.17
142-154	Y.TLGVK(+42.01)QLIVGVNK.M	46.39	1409.86	2	705.94	0.5	1.58E+07	1	K5:Acetylation (K):63.64
299-313	A.LSE(+197.05)ALPGDNVGFNVK.N	46.14	1755.84	2	878.93	0.8	2.44E+07	2	E3:Glycerylphosphorylethanolamine:1000.00
396-410	K.SGDAAIVEM(+15.99)VPGK(+42.01)PM.C	44.66	1558.74	2	780.38	1.2	1.94E+08	4	M9:Oxidation (M):9.45; K13:Acetylation (K):1000.00
295-313	M.HHEALSE(+197.05)ALPGDNVGFNVK.N	43.28	2230.04	3	744.35	-2.2	1.55E+08	2	E7:Glycerylphosphorylethanolamine:25.18
396-423	K.SGDAAIVEM(+15.99)VPGK(+42.01)PMC(+57.02)VE(+14.02)SFSQYPPLGR.F	41.97	3093.46	4	774.37	-6.7	0	1	M9:Oxidation (M):0.00; K13:Acetylation (K):1000.00; C16:Carbamidomethylation:1000.00; E18:Methylation(D E):0.00

300-312	L.SE(+197.05)ALPGDNVGFNVK.N	41.09	1642.76	2	822.39	1.2	5.37E+08	4	E2:Glycerylphosphorylethanolamine:1000.00
314-321	K.NVSVK(+42.01)DIR.R	40.86	971.54	3	324.85	0.4	0	1	K5:Acetylation (K):1000.00
357-371	G.YSPVID(+14.02)C(+57.02)HTAHIAC(+57.02)K.F	40.54	1784.83	3	595.95	-0.9	1.50E+07	1	D6:Methylation(D E):1000.00; C7:Carbamidomethylation:1000.00; C14:Carbamidomethylation:1000.00
348-371	L.NHPGQISAGYSPVID(+14.02)C(+57.02)HTAHIAC(+57.02)K.F	39.58	2646.24	4	662.57	-1.7	6.19E+08	4	D15:Methylation(D E):1000.00; C16:Carbamidomethylation:1000.00; C23:Carbamidomethylation:1000.00
322-330	R.RGNVC(+57.02)GD(+14.02)SK.S	38.42	1005.47	2	503.74	-0.1	2.76E+07	16	C5:Carbamidomethylation:1000.00; D7:Methylation(D E):1000.00
173-179	K.EVSAY(+79.97)IK.K	38.34	888.40	1	889.41	2.3	4.01E+07	1	Y5:Phosphorylation (STY):33.18
396-408	K.SGDAAIVE(+14.02)MVPGK.P	38.23	1286.65	2	644.33	0.1	5.80E+07	1	E8:Methylation(D E):25.58
155-166	K.MD(+14.02)STEPAYSEK(+42.01)R.Y	37.87	1468.65	2	735.33	-0.1	6.25E+07	1	D2:Methylation(D E):28.36; K11:Acetylation (K):1000.00
52-62	K.GSFK(+42.01)YAWVLDK.L	37.85	1354.69	2	678.35	-0.8	2.27E+08	8	K4:Acetylation (K):63.15
396-410	K.SGDAAIVEMVPGK(+42.01)PM(+15.99).C	37.49	1558.74	2	780.38	0.6	1.29E+08	3	K13:Acetylation (K):1000.00
331-367	K.SDPPQEAAQFTSQVIILNHPGQISAGYSPVID(+14.02)C(+57.02)HTAH.I	37.36	3998.92	4	1000.74	1.3	3.89E+09	8	D32:Methylation(D E):30.48; C33:Carbamidomethylation:1000.00
155-166	K.MDSTEPAYSEK(+42.01)R.Y	37.08	1454.63	3	485.88	-6.6	0	1	K11:Acetylation (K):1000.00
298-313	E.ALSE(+197.05)ALPGDNVGFNVK.N	36.86	1826.88	2	914.45	2.3	1.38E+07	1	E4:Glycerylphosphorylethanolamine:1000.00
155-166	K.MDSTE(+14.02)PAYSE(+14.02)KR.Y	36.21	1440.66	2	721.34	0.9	3.19E+09	5	E5:Methylation(D E):44.44; E10:Methylation(D E):25.00

358-371	Y.SPVID(+14.02)C(+57.02)HTAHIAC(+57.02)K.F	35.89	1621.77	3	541.60	1	7.36E+07	1	D5:Methylation(D E):1000.00; C6:Carbamidomethylation:1000.00;C13:Carbamidomethylation:1000.00
396-423	K.SGDAAIVE(+14.02)M(+15.99)VPGK(+42.01)PM(+15.99)C(+57.02)VESFSQYPPLGR.F	35.78	3109.46	4	778.38	5.7	0	1	E8:Methylation(D E):0.00; M9:Oxidation (M):1000.00; K13:Acetylation (K):1000.00; M15:Oxidation (M):1000.00; C16:Carbamidomethylation:1000.00
297-313	H.EALSE(+197.05)ALPGDNVGFNVK.N	35.72	1955.92	2	978.97	1	1.00E+07	1	E5:Glycerylphosphorylethanolamine:28.70
396-423	K.SGDAAIVE(+14.02)M(+15.99)VPGKPM(+15.99)C(+57.02)VE(+14.02)SFSQYPPLGR.F	35.61	3081.46	3	1028.15	-8.1	5.85E+07	2	E8:Methylation(D E):45.16; M9:Oxidation (M):1000.00; M15:Oxidation (M):1000.00; C16:Carbamidomethylation:1000.00; E18:Methylation(D E):25.39
396-410	K.SGDAAIVEMVPGK(+42.01)PM.C	35.2	1542.74	2	772.38	1.5	2.30E+08	8	K13:Acetylation (K):1000.00
155-166	K.M(+15.99)DSTE(+14.02)PAYSE(+14.02)KR.Y	34.6	1456.65	3	486.56	0.1	6.75E+08	6	M1:Oxidation (M):1000.00; E5:Methylation(D E):68.47; E10:Methylation(D E):40.94
396-403	K.SGDAAIVE(+14.02).M	34.57	774.38	1	775.38	1.2	4.50E+07	2	E8:Methylation(D E):67.85
256-266	K.IGGIGT(+79.97)VPVGR.V	34.16	1104.57	2	553.29	-0.8	1.19E+07	1	T6:Phosphorylation (STY):1000.00
356-371	A.GYSPVID(+14.02)C(+57.02)HTAHIAC(+57.02)K.F	33.91	1841.86	4	461.47	0.2	2.91E+08	3	D7:Methylation(D E):1000.00; C8:Carbamidomethylation:1000.00; C15:Carbamidomethylation:1000.00
101-114	K.NM(+15.99)ITGTSQAD(+14.02)C(+57.02)AVL.I	33.82	1509.68	2	755.85	0.8	1.36E+08	3	M2:Oxidation (M):1000.00; D10:Methylation(D E):1000.00; C11:Carbamidomethylation:1000.00
160-166	E.PAYSE(+14.02)KR.Y	33.72	863.45	2	432.73	2.4	1.11E+07	1	E5:Methylation(D E):1000.00

396-408	K.SGDAAIVEM(+15.99)VPGK(+42.01).P	32.97	1330.64	2	666.33	0.7	7.82E+07	3	M9:Oxidation (M):1000.00; K13:Acetylation (K):1000.00
157-166	D.STEPAYSE(+14.02)KR.Y	32.56	1180.57	2	591.29	1.7	7.65E+07	3	E8:Methylation(D E):70.86
101-129	K.NM(+15.99)ITGTSQAD(+14.02)C(+57.02)AVLIVAAGVGE(+14.02)FEAGISK.N	32.33	2952.46	3	985.16	-1.3	6.78E+05	1	M2:Oxidation (M):1000.00; D10:Methylation(D E):6.89; C11:Carbamidomethylation:1000.00; E22:Methylation(D E):22.85
291-313	K.SVEM(+15.99)HHEALS(+79.97)E(+14.02)ALPGD(+14.02)NVGFNVK.N	32.31	2603.17	3	868.74	5.8	1.86E+08	1	M4:Oxidation (M):1000.00; S10:Phosphorylation (STY):22.83; E11:Methylation(D E):25.70; D16:Methylation(D E):18.61
155-166	K.M(+15.99)DSTE(+14.02)PAYSEK(+42.01)R.Y	32.16	1484.65	2	743.33	-2.6	0	1	M1:Oxidation (M):1000.00; E5:Methylation(D E):0.00; K11:Acetylation (K):1000.00
401-423	A.IVEM(+15.99)VPGKPM(+15.99)C(+57.02)VE(+14.02)SFSQYPPLGR.F	31.4	2666.29	3	889.77	1.5	2.56E+07	2	M4:Oxidation (M):1000.00; M10:Oxidation (M):1000.00; C11:Carbamidomethylation:1000.00; E13:Methylation(D E):21.51
291-301	K.SVEM(+15.99)HHEALSE(+197.05).A	30.91	1480.59	2	741.30	-0.3	3.41E+08	1	M4:Oxidation (M):1000.00; E11:Glycerylphosphorylethanolamine:42.89
396-410	K.SGDAAIVE(+14.02)MVPGK(+42.01)PM.C	30.51	1556.76	2	779.38	-3.9	9.37E+06	2	E8:Methylation(D E):52.68; K13:Acetylation (K):1000.00
267-279	R.VE(+14.02)TGILRPGMVVT.F	29.56	1384.77	2	693.40	1.2	2.44E+07	1	E2:Methylation(D E):1000.00
291-313	K.SVE(+197.05)MHHE(+14.02)ALSE(+14.02)ALPGD(+14.02)NVGFNVK.N	29.27	2718.27	3	907.10	1.3	9.48E+07	1	E3:Glycerylphosphorylethanolamine:4.73; E7:Methylation(D E):4.73; E11:Methylation(D E):2.52; D16:Methylation(D E):1000.00

341-371	F.TSQVIILNHPGQISAGYSPVID(+14.02)C(+57.02)HTAHIAC(+57.02)K.F	29.17	3400.70	5	681.15	1.1	4.56E+08	2	D22:Methylation(D E):1000.00; C23:Carbamidomethylation:1000.00; C30:Carbamidomethylation:1000.00
219-242	R.K(+42.01)E(+14.02)GNASGVSLLEALDTILPPTRPT.D	28.71	2534.36	3	845.79	0.6	0	1	K1:Acetylation (K):1000.00
314-321	K.NVSVK(+42.01)D(+14.02)IR.R	28.29	985.56	3	329.53	0.4	1.57E+09	2	K5:Acetylation (K):1000.00; D6:Methylation(D E):1000.00
396-423	K.SGDAAIVEMVPGK(+42.01)PM(+15.99)C(+57.02)VE(+14.02)S(+79.97)FSQYPPLGR.F	28.03	3173.43	3	1058.83	8.2	4.69E+07	1	K13:Acetylation (K):1000.00; M15:Oxidation (M):0.00; C16:Carbamidomethylation:1000.00; E18:Methylation(D E):0.00; S19:Phosphorylation (STY):0.00
101-118	K.NM(+15.99)ITGTSQAD(+14.02)C(+57.02)AVLIVAA.G	27.41	1863.91	2	932.96	0.2	2.26E+07	1	M2:Oxidation (M):1000.00; D10:Methylation(D E):1000.00; C11:Carbamidomethylation:1000.00
396-423	K.SGDAAIVEMVPGK(+42.01)PMC(+57.02)VESFSQYPPLGR.F	27.19	3063.45	3	1022.15	-7.8	7.88E+05	1	K13:Acetylation (K):1000.00; C16:Carbamidomethylation:1000.00
301-313	S.E(+197.05)ALPGDNVGFNVK.N	27.05	1555.73	2	778.87	-0.7	8.70E+06	1	E1:Glycerylphosphorylethanolamine:1000.00
396-410	K.SGDAAIVE(+14.02)M(+15.99)VPGK(+42.01)PM(+15.99).C	26.52	1588.75	2	795.38	-2.2	0	1	E8:Methylation(D E):52.68; M9:Oxidation (M):1000.00; K13:Acetylation (K):1000.00; M15:Oxidation (M):1000.00
267-279	R.VE(+14.02)TGILRPGM(+15.99)VVT.F	26.34	1400.77	2	701.39	-1.1	2.58E+07	1	E2:Methylation(D E):1000.00; M10:Oxidation (M):1000.00
155-172	K.MDSTEPAYSEK(+42.01)RYDEIVK.E	26.17	2202.01	4	551.51	-0.2	1.13E+08	1	K11:Acetylation (K):67.12
396-423	K.SGDAAIVE(+14.02)MVPGK(+42.01)PMC(+57.02)VESFSQYPPLGR.F	26.09	3077.47	4	770.37	-7.2	1.12E+08	2	E8:Methylation(D E):0.00; K13:Acetylation (K):1000.00; C16:Carbamidomethylation:1000.00
52-57	K.GSFK(+42.01)YA.W	25.68	713.34	2	357.68	-0.8	1.11E+08	1	K4:Acetylation (K):1000.00

181-212	K.IGYNPATVPFVPISGWHGD(+14.02)NMLE(+14.02)PSPNMPWFK.G	25.67	3625.75	4	907.45	1	1.24E+05	1	D19:Methylation(D E):1000.00; E23:Methylation(D E):1000.00
372-376	K.FAE(+197.05)LK.E	25.04	803.38	2	402.70	0.7	1.34E+08	1	E3:Glycerylphosphorylethanolamine:1000.00
396-430	K.SGDAAIVEMVPGK(+42.01)PM(+15.99)C(+57.02)VESFSQYPPLGRFAVRD(+14.02)M(+15.99)R.Q	24.32	3984.90	4	997.24	3.5	1.91E+08	1	K13:Acetylation (K):1000.00; M15:Oxidation (M):0.00; C16:Carbamidomethylation:1000.00; D33:Methylation(D E):0.00; M34:Oxidation (M):0.00
267-273	R.VE(+14.02)TGILR.P	24.21	800.48	2	401.25	1.3	6.79E+07	2	E2:Methylation(D E):1000.00
396-410	K.SGDAAIVE(+14.02)M(+15.99)VPGK(+42.01)PM.C	23.82	1572.75	2	787.38	-4	7.18E+05	1	E8:Methylation(D E):52.68; M9:Oxidation (M):0.00; K13:Acetylation (K):1000.00
52-60	K.GSFK(+42.01)YAWVL.D	23.74	1111.57	2	556.79	-0.1	1.17E+07	1	K4:Acetylation (K):1000.00

* Carbamidomethylation- and oxidation- only modifications are not shown as these modifications are artefacts of mass spectrometry experiments. Peptides are sorted according to their -10logP value.

Table S8: PEAKS-PTM mass spectrometry data analysis results

Position	Modified Peptide Sequence	Modifications	Best -10logP	Best AScore_PTMS	Peptides modified	Peptides unmodified
E135	AGISKNGQTReHALLAYTLGV	Methylation(D E)	72.61	1000	4.34E+09	2.94E+10
E164	KMDSTEPAYSeKRYDEIVKEV	Methylation(D E)	69.45	1000	2.11E+10	3.58E+10
K55	EAAEMGKGSFkYAWVLDKLKA	Dimethylation(KR)	68.52	1000	1.42E+10	4.09E+09
D110	KNMITGTSQAdCAVLIVAAGV	Methylation(D E)	67.7	1000	2.24E+08	8.87E+09
K408	DAAIVEMVPGkPMCVESFSQY	Acetylation (K)	65.57	1000	2.48E+09	6.16E+10
E173	SEKRYDEIVKeVSAYIKKIGY	Methylation(D E)	61.89	1000	2.20E+09	4.35E+10
D362	QISAGYSPVIdCHTAHIACKF	Methylation(D E)	57.57	1000	5.28E+09	1.21E+10
K165	MDSTEPAYSEkRYDEIVKEVS	Acetylation (K)	56.1	1000	1.21E+08	2.69E+10
D328	KDIRRGNVCGdSKSDPPQEAA	Methylation(D E)	54.08	1000	7.10E+07	7.19E+07
D252	TDKPLRLPLQdVYKIGGIGTV	Methylation(D E)	53.63	1000	2.72E+08	1.74E+10
E413	EMVPGKPMCVeSFSQYPPLGR	Methylation(D E)	52.76	1000	2.94E+09	5.07E+10
T88	WKFETTKYYItIIDAPGHRDF	Phosphorylation (STY)	51.88	22.45	2.99E+07	4.43E+10
K79	RGITIDISLWkFETTKYYITI	Dimethylation(KR)	51.01	29.87	1.15E+09	6.48E+07
E81	ITIDISLWKFeTTKYYITIID	Methylation(D E)	51.01	40.82	6.38E+08	7.27E+07
E442	TVAVGVIKNVeKKSGGAGKVT	Methylation(D E)	50.12	1000	4.71E+09	2.90E+08
K55	EAAEMGKGSFkYAWVLDKLKA	Methylation(KR)	49.49	104.79	1.32E+08	4.10E+09
E403	SLKSGDAAIVeMVPGKPMCVE	Methylation(D E)	49.18	67.85	2.32E+08	5.98E+10
E301	SVEMHHEALSeALPGDNVGFN	Glycerylphosphorylethanolamine	48.33	1000	1.28E+09	1.63E+10
E297	TEVKSVEMHHeALSEALPGDN	Methylation(D E)	47.28	22.82	4.07E+07	2.20E+10
K36	HLIYKCGGIDkRTIEKFEKEA	Dimethylation(KR)	46.44	27.95	7.64E+06	0.00E+00
K146	HALLAYTLGVkQLIVGVNKMD	Acetylation (K)	46.39	63.64	1.58E+07	3.20E+10

D17	HINIVVIGHVdSGKSTTTGHL	Methylation(D E)	42.25	1000	3.85E+08	3.48E+09
K318	VGFNVKNVSVkDIRRGNVCGD	Dimethylation(KR)	41.23	68.47	3.07E+10	0.00E+00
K55	EAAEMGKGSFkYAWVLDKLKA	Acetylation (K)	39.34	1000	3.86E+08	4.10E+09
D91	ETTKYYITIIdAPGHRDFIKN	Methylation(D E)	39.14	1000	1.29E+09	5.13E+10
K165	MDSTEPAYSEkRYDEIVKEVS	Dimethylation(KR)	38.6	26.02	4.19E+07	2.37E+10
Y177	YDEIVKEVSAyIKKIGYNPAT	Phosphorylation (STY)	38.34	33.17	4.01E+07	4.09E+10
D156	KQLIVGVNKMdSTEPAYSEKR	Methylation(D E)	37.87	28.35	6.25E+07	5.57E+10
K371	IDCHTAHIACkFAELKEKIDR	Methylation(KR)	37.4	1000	2.89E+08	5.45E+09
E159	IVGVNKMDSTePAYSEKRYDE	Methylation(D E)	34.6	68.47	4.33E+07	5.54E+10
R273	PVGRVETGILrPGMVVTFAPV	Methylation(KR)	34.23	40.47	6.57E+07	5.27E+10
T261	QDVYKIGGIGtVPVGRVETGI	Phosphorylation (STY)	34.16	1000	1.19E+07	6.30E+10
E122	AVLIVAAGVGeFEAGISKNGQ	Methylation(D E)	32.33	22.85	6.78E+05	1.38E+10
E301	SVEMHHEALSeALPGDNVGFN	Methylation(D E)	32.31	25.7	1.86E+08	2.07E+10
S300	KSVEMHHEALsEALPGDNVGF	Phosphorylation (STY)	32.31	22.82	1.86E+08	2.26E+10
K386	KEKIDRRSGKkLEDNPKSLKS	Methylation(KR)	32.08	104.47	9.48E+07	3.80E+09
D319	GFNVKNVSVKdIRRGNVCGDS	Methylation(D E)	31.68	1000	2.93E+10	3.01E+09
D168	TEPAYSEKRYdEIVKEVSAYI	Methylation(D E)	31.43	33.97	2.01E+09	4.72E+10
D389	IDRRSGKKLEdNPKSLKSGDA	Methylation(D E)	30.74	33.97	4.41E+08	3.65E+09
E268	GIGTVPVGRVeTGILRPGMVV	Methylation(D E)	29.91	1000	5.27E+09	4.74E+10
H367	YSPVIDCHTAhIACKFAELKE	Methylation(others)	29.86	22.45	1.35E+08	1.83E+10
D306	HEALSEALPGdNVGFNVKNVS	Methylation(D E)	29.27	1000	9.48E+07	2.39E+10
R166	DSTEPAYSEKrYDEIVKEVSA	Methylation(KR)	28.91	113.99	2.13E+07	6.33E+10
K318	VGFNVKNVSVkDIRRGNVCGD	Acetylation (K)	28.29	1000	1.59E+09	0.00E+00
D199	PFVPISGWHGdNMLEPSPNMP	Methylation(D E)	25.67	1000	1.24E+05	8.71E+09

E203	ISGWHGDNMLePSPNMPWFKG	Methylation(D E)	25.67	1000	1.24E+05	9.40E+09
E374	HTAHIACKFAeLKEKIDRRSG	Glycerylphosphorylethanolamine	25.04	1000	1.34E+08	2.61E+08
K36	HLIYKCGGIDkRTIEKFEKEA	Acetylation (K)	23.63	1000	1.03E+07	0.00E+00
D35	GHLIYKCGGIdKRTIEKFEKE	Methylation(D E)	23.63	1000	1.03E+07	3.13E+07

* Modifications other than acetylations, methylations, phosphorylations and glycerylphosphorylethanolamination were neglected Modifications are sorted according to their -10logP value.

Table S9: PEAKS *De novo* mass spectrometry data analysis

Peptide	Mass	m/z	z	ppm	Intensity	ALC (%)	PTM	Tag (>=80%)
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-1	1,18E+07	96	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-0,5	1,58E+07	96	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
K(+42.01)PLQDVYK	1031,6	516,79	2	0,4	2,24E+08	95	Acetylation (K)	K(+42.01)PLQDVYK
K(+42.01)PLQDVYK	1031,6	516,79	2	0,1	2,52E+08	94	Acetylation (K)	K(+42.01)PLQDVYK
STTTK(+42.01)HLYK	1119,6	560,81	2	2,8		94	Acetylation (K)	STTTK(+42.01)HLYK
D(+14.02)HALLAYTLRK	1313,7	657,87	2	-8,5	1,56E+07	94	Methylation(D E)	D(+14.02)HALLAYTLRK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-1,2	1,18E+07	94	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-1	1,18E+07	94	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
Y(+79.97)PRYTK	906,4	454,21	2	-0,3		94	Phosphorylation (STY)	Y(+79.97)PRYTK
K(+42.01)PLQDVYK	1031,6	516,79	2	0,8	2,99E+08	93	Acetylation (K)	K(+42.01)PLQDVYK
TSTTK(+42.01)HLYK	1119,6	560,8	2	0		93	Acetylation (K)	TSTTK(+42.01)HLYK
K(+42.01)PLQDVYK	1031,6	516,79	2	0,7	2,24E+08	93	Acetylation (K)	K(+42.01)PLQDVYK
K(+42.01)PLQDVYK	1031,6	516,79	2	0,1	5,47E+07	93	Acetylation (K)	K(+42.01)PLQDVYK
VETK(+42.01)LR	786,46	394,23	2	-7,4	3,33E+06	93	Acetylation (K)	VETK(+42.01)LR
APVNLTE(+14.02)SVK	1070,6	536,31	2	-0,3	7,51E+07	93	Methylation(D E)	APVNLTE(+14.02)SVK
D(+14.02)HALLAYTLRK	1313,7	657,87	2	-8,4	1,56E+07	93	Methylation(D E)	D(+14.02)HALLAYTLRK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-0,9	9,54E+06	93	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-1,1	1,18E+07	93	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-0,8	1,18E+07	93	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-1,2	1,18E+07	93	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-0,7		93	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-1,3	1,18E+07	93	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-0,8	9,54E+06	93	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-0,6	1,18E+07	93	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-1,1	1,18E+07	93	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
RWM(+15.99)EAK(+42.01)SK	1092,5	547,28	2	6,3	2,92E+08	93	Oxidation (M); Acetylation (K)	RWM(+15.99)EAK(+42.01)SK

VETK(+42.01)LR	786,46	394,24	2	-0,8		92	Acetylation (K)	VETK(+42.01)LR
STTTK(+42.01)HLYK	1119,6	560,8	2	1,4	3,14E+06	92	Acetylation (K)	[87.0]TTTK(+42.01)HLYK
STTTK(+42.01)HLYK	1119,6	560,8	2	0,1	6,25E+06	92	Acetylation (K)	STTTK(+42.01)HLYK
K(+42.01)PLQDVYK	1031,6	516,79	2	-1,3	7,67E+09	92	Acetylation (K)	K(+42.01)PLQDVYK
TLD(+14.02)KFEK	893,49	447,75	2	-0,1	4,96E+07	92	Methylation(D E)	TLD(+14.02)KFEK
VE(+14.02)TK(+42.01)LR	800,48	401,25	2	1,2	6,79E+07	92	Methylation(D E); Acetylation (K)	VE(+14.02)TK(+42.01)LR
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-0,8	1,18E+07	92	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)GK
RWM(+15.99)EAK(+42.01)SK	1092,5	547,28	2	6,6	1,51E+06	92	Oxidation (M); Acetylation (K)	RWM(+15.99)EAK(+42.01)SK
K(+42.01)PLQDVYK	1031,6	516,79	2	0,6	7,67E+09	91	Acetylation (K)	K(+42.01)PLQDVYK
K(+42.01)PLQDVYK	1031,6	516,79	2	0,3	7,67E+09	91	Acetylation (K)	K(+42.01)PLQDVYK
K(+42.01)PLQDVYK	1031,6	516,79	2	-0,1		91	Acetylation (K)	K(+42.01)PLQDVYK
AK(+42.01)LVGVNK	869,53	435,77	2	0,6	5,32E+08	91	Acetylation (K)	[71.0]K(+42.01)LVGVNK
AK(+42.01)LVGVNK	869,53	435,77	2	-0,2	5,32E+08	91	Acetylation (K)	[71.0]K(+42.01)LVGVNK
VETK(+42.01)LR	786,46	394,23	2	-7,4	3,33E+06	91	Acetylation (K)	VETK(+42.01)LR
AK(+42.01)LVGVNK	869,53	435,77	2	-0,8	3,05E+10	91	Acetylation (K)	[71.0]K(+42.01)LVGVNK
VAVTK(+42.01)PR	811,49	406,75	2	-0,1	6,31E+07	91	Acetylation (K)	VAVTK(+42.01)PR
VETK(+42.01)LR	786,46	394,24	2	0,1	4,45E+05	91	Acetylation (K)	VETK(+42.01)LR
LGK(+42.01)NVFC(+57.02)K	1006,5	504,27	2	8,3	5,56E+06	91	Acetylation (K);	LGK(+42.01)NVFC(+57.02)K
LELSD(+14.02)LNR	972,52	487,27	2	1,4	1,06E+09	91	Methylation(D E)	LELSD(+14.02)LNR
TQALVD(+14.02)LVK	999,6	500,81	2	0,1	4,51E+07	91	Methylation(D E)	[229.1]ALVD(+14.02)LVK
D(+14.02)HALLAYTLRK	1313,7	657,87	2	-8,8	1,56E+07	91	Methylation(D E)	D(+14.02)HALLAYTLRK
KGLD(+14.02)DNPK	899,47	450,74	2	-0,8		91	Methylation(D E)	KGLD(+14.02)DNPK
RWM(+15.99)EAK(+42.01)SK	1092,5	547,28	2	5,6	1,56E+06	91	Oxidation (M); Acetylation (K)	RWM(+15.99)EAK(+42.01)SK
RWM(+15.99)EAK(+42.01)SK	1092,5	547,28	2	7,1	2,72E+06	91	Oxidation (M); Acetylation (K)	RWM(+15.99)EAK(+42.01)SK
RWM(+15.99)EAK(+42.01)SK	1092,5	547,28	2	5,2		91	Oxidation (M); Acetylation (K)	RWM(+15.99)EAK(+42.01)SK
RWM(+15.99)D(+14.02)AK(+42.01)SK	1092,5	547,28	2	6	4,72E+06	91	Oxidation (M); Methylation(D E); Acetylation (K)	RWM(+15.99)D(+14.02)A[170.1]SK
TLY(+79.97)LK	716,35	359,18	2	-8,7	1,18E+07	91	Phosphorylation (STY)	TLY(+79.97)LK
S(+79.97)PLQDVYK	1028,5	515,24	2	2,3	4,90E+07	91	Phosphorylation (STY)	S(+79.97)PLQDVYK
T(+79.97)T(+79.97)EDLAQK	1064,4	533,2	2	-3,9		91	Phosphorylation (STY)	[181.0]T(+79.97)EDLAQK

VSVDLVK(+42.01)	800,46	401,24	2	0,5	5,69E+07	90	Acetylation (K)	VSVDLVK(+42.01)
VETK(+42.01)LR	786,46	394,24	2	1,8	6,84E+05	90	Acetylation (K)	VETK(+42.01)LR
VETK(+42.01)LR	786,46	394,24	2	1,6	7,53E+05	90	Acetylation (K)	VETK(+42.01)LR
VETK(+42.01)LR	786,46	394,24	2	-3,8	5,94E+05	90	Acetylation (K)	VETK(+42.01)LR
KYE(+14.02)DVVK	893,49	447,75	2	0,2	1,33E+07	90	Methylation(D E)	KYE(+14.02)DVVK
KLE(+14.02)DNPK	856,47	429,24	2	-0,5	5,32E+06	90	Methylation(D E)	KLE(+14.02)DNP[128.1]
VE(+14.02)TK(+42.01)LR	800,48	401,25	2	1,3	6,79E+07	90	Methylation(D E); Acetylation (K)	VE(+14.02)TK(+42.01)LR
RWM(+15.99)EAK(+42.01)SK	1092,5	547,28	2	6,3	1,56E+06	90	Oxidation (M); Acetylation (K)	RWM(+15.99)EAK(+42.01)SK
RWM(+15.99)D(+14.02)AK(+42.01)SK	1092,5	547,28	2	6		90	Oxidation (M); Methylation(D E); Acetylation (K)	RWM(+15.99)D(+14.02)A[170.1]SK
M(+15.99)D(+14.02)AK(+42.01)SK	750,36	376,19	2	1,9		90	Oxidation (M); Methylation(D E); Acetylation (K)	M(+15.99)D(+14.02)AK(+42.01)SK
VAVTK(+42.01)PR	811,49	406,75	2	-0,4	6,31E+07	89	Acetylation (K)	VAVTK(+42.01)[97.1]R
STTTK(+42.01)HLYK	1119,6	560,8	2	0,8	2,53E+06	89	Acetylation (K)	STTT[307.2]LYK
VETK(+42.01)LR	786,46	394,24	2	2,5	4,45E+05	89	Acetylation (K)	VETK(+42.01)LR
K(+42.01)PLQDVYK	1031,6	516,79	2	-0,7	5,47E+07	89	Acetylation (K)	K(+42.01)PL[128.1]DVYK
AK(+42.01)LVGVNK	869,53	435,77	2	-0,5		89	Acetylation (K)	[71.0]K(+42.01)LVGVNK
VETK(+42.01)LR	786,46	394,24	2	-1	5,40E+05	89	Acetylation (K)	VETK(+42.01)LR
EPGD(+14.02)FK	705,33	353,67	2	-3,7	2,51E+08	89	Methylation(D E)	EP[57.0]D(+14.02)FK
TLLD(+14.02)GEESR	1032,5	517,26	2	0,8	3,73E+08	89	Methylation(D E)	TLLD(+14.02)[57.0]EESR
LD(+14.02)LSENLR	972,52	487,27	2	-0,4	1,59E+07	89	Methylation(D E)	LD(+14.02)LSENL[156.1]
VE(+14.02)TK(+42.01)LR	800,48	401,24	2	-0,5	2,98E+06	89	Methylation(D E); Acetylation (K)	VE(+14.02)TK(+42.01)LR
VE(+14.02)TK(+42.01)LR	800,48	401,24	2	-1	2,98E+06	89	Methylation(D E); Acetylation (K)	VE(+14.02)TK(+42.01)LR
FD(+14.02)AK(+42.01)SK	750,39	376,2	2	-1,6	3,60E+06	89	Methylation(D E); Acetylation (K)	FD(+14.02)A[170.1]SK
DSTD(+14.02)PAYSEK(+42.01)K	1295,6	648,81	2	8,8	1,05E+08	89	Methylation(D E); Acetylation (K)	DSTD(+14.02)PAY[87.0]EK(+42.01)[128.1]
RWM(+15.99)D(+14.02)AK(+42.01)SK	1092,5	547,28	2	5,2		89	Oxidation (M); Methylation(D E); Acetylation (K)	RWM(+15.99)D(+14.02)A[170.1]SK
NPLDQVHK(+42.01)	991,51	496,76	2	-9,8	2,23E+08	88	Acetylation (K)	NPLDQVHK(+42.01)
AK(+42.01)LVGVNK	869,53	435,77	2	-0,6	3,18E+07	88	Acetylation (K)	[71.0]K(+42.01)LV[57.0]VNK
GSFFK(+42.01)	626,31	314,16	2	-0,8	5,04E+06	88	Acetylation (K)	[57.0]SFFK(+42.01)
VETK(+42.01)LR	786,46	394,24	2	1,4		88	Acetylation (K)	VETK(+42.01)LR

KGLD(+14.02)DNPK	899,47	450,74	2	-1,1	6,23E+07	88	Methylation(D E)	KGLD(+14.02)DNP[128.1]
NEQVD(+14.02)LR	886,45	444,23	2	-0,1	2,43E+08	88	Methylation(D E)	[114.0]EQVD(+14.02)LR
LPLE(+14.02)TDYK	991,52	496,77	2	-0,4	8,58E+07	88	Methylation(D E)	[113.1]PLE(+14.02)T[115.0]YK
YNDD(+14.02)LVK	879,43	440,72	2	-0,2	5,35E+05	88	Methylation(D E)	[163.1]NDD(+14.02)LVK
M(+15.99)E(+14.02)ELVK	777,39	389,7	2	-4	8,94E+07	88	Oxidation (M); Methylation(D E)	[147.0]E(+14.02)ELVK
RWM(+15.99)D(+14.02)AK(+42.01)SK	1092,5	547,28	2	6,1		88	Oxidation (M); Methylation(D E); Acetylation (K)	RWM(+15.99)D(+14.02)A[170.1]SK
LPLDY(+79.97)YK	990,45	496,23	2	1,3	3,78E+06	88	Phosphorylation (STY)	[113.1]PLD[243.0]YK
LPLY(+79.97)DYK	990,45	496,23	2	4	1,60E+07	88	Phosphorylation (STY)	[113.1]PLY(+79.97)DYK
VETK(+42.01)LR	786,46	394,24	2	2,6	1,67E+06	87	Acetylation (K)	VETK(+42.01)L[156.1]
VETK(+42.01)LR	786,46	394,24	2	1,1	7,42E+05	87	Acetylation (K)	VETK(+42.01)LR
TESLAK(+42.01)R	845,46	423,74	2	-0,8	4,36E+06	87	Acetylation (K)	TESLAK(+42.01)[156.1]
WK(+42.01)GTVALPK	1040,6	521,3	2	-8,9	1,47E+07	87	Acetylation (K)	[186.1]K(+42.01)[57.0]TVALPK
VAK(+42.01)SLPR	811,49	406,75	2	0,9	6,31E+07	87	Acetylation (K)	VAK(+42.01)[87.0]LPR
STTTK(+42.01)HLYK	1119,6	560,8	2	-1,1		87	Acetylation (K)	[87.0]TTT[170.1]HLYK
K(+42.01)GTVPVGR	854,5	428,26	2	1,1	5,59E+08	87	Acetylation (K)	K(+42.01)GTVPVG[156.1]
VEDLD(+14.02)SVGK	974,49	488,25	2	-0,9	1,52E+08	87	Methylation(D E)	[99.1]EDLD(+14.02)SVG[128.1]
SD(+14.02)LTERLR	1002,5	335,19	3	-1	2,85E+07	87	Methylation(D E)	SD(+14.02)L[101.0]ERL[156.1]
TAHNFD(+14.02)K	845,4	423,71	2	4,4	6,72E+06	87	Methylation(D E)	[172.1]HNFD(+14.02)K
KGLD(+14.02)DNPK	899,47	450,74	2	-1,2		87	Methylation(D E)	KGLD(+14.02)D[114.0]P[128.1]
KGLD(+14.02)DNPK	899,47	450,74	2	0	3,24E+07	87	Methylation(D E)	KGLD(+14.02)DNP[128.1]
KGLD(+14.02)DNPK	899,47	450,74	2	-0,4	6,23E+07	87	Methylation(D E)	KGLD(+14.02)DNP[128.1]
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-1	3,11E+06	87	Methylation(D E); Oxidation (M)	[200.1]AD(+14.02)M(+15.99)GK
RWM(+15.99)D(+14.02)AK(+42.01)SK	1092,5	547,28	2	6,3	4,72E+06	87	Oxidation (M); Methylation(D E); Acetylation (K)	RWM(+15.99)D(+14.02)[241.1]SK
S(+79.97)LAYSEK	876,36	439,19	2	0,4		87	Phosphorylation (STY)	S(+79.97)LAYSEK
Y(+79.97)HLRK	795,38	398,7	2	8,9	1,48E+07	87	Phosphorylation (STY)	[243.0]HLRK
T(+79.97)ALLASFK	929,46	465,74	2	-1,8	1,55E+06	87	Phosphorylation (STY)	T(+79.97)ALLASF[128.1]
LPLY(+79.97)FDK	974,45	488,23	2	-4,1	5,34E+06	87	Phosphorylation (STY)	LPLY(+79.97)FD[128.1]
KLT(+79.97)LK	681,38	341,7	2	-2,6	1,29E+08	87	Phosphorylation (STY)	KL[181.0]LK

RGLTEVK(+42.01)	843,48	422,75	2	0,9	6,06E+08	86	Acetylation (K)	R[57.0]LTEVK(+42.01)
VAK(+42.01)SLPR	811,49	406,75	2	-1,5	6,31E+07	86	Acetylation (K)	VAK(+42.01)[87.0]LPR
K(+42.01)SSPATLNSR	1101,6	551,8	2	0,4	1,22E+07	86	Acetylation (K)	[170.1]SSPATLNSR
VETK(+42.01)LR	786,46	394,24	2	4,9	1,09E+06	86	Acetylation (K)	VET[170.1]L[156.1]
NVSVTK(+42.01)LR	957,56	320,19	3	-0,3	5,60E+06	86	Acetylation (K)	[114.0]VSVTK(+42.01)LR
DVSTVK(+42.01)LR	958,54	480,28	2	0	7,82E+07	86	Acetylation (K)	DVSTV[170.1]LR
DVSTVK(+42.01)LR	958,54	320,52	3	-3,2	1,39E+07	86	Acetylation (K)	[115.0]VSTVK(+42.01)L[156.1]
EK(+42.01)PRYTK	962,52	482,27	2	-0,2	6,64E+06	86	Acetylation (K)	[299.1]PRYTK
STTTK(+42.01)HLYK	1119,6	560,8	2	0,4		86	Acetylation (K)	[87.0]TTT[307.2]LYK
STTTK(+42.01)HLYK	1119,6	560,8	2	1,4		86	Acetylation (K)	[87.0]TTT[170.1]HLYK
NVSVTK(+42.01)LR	957,56	320,19	3	-0,7	5,60E+06	86	Acetylation (K)	[114.0]VSVTK(+42.01)LR
QK(+42.01)LVGVNK	926,55	464,28	2	-0,5	8,54E+07	86	Acetylation (K)	[298.2]LVGVNK
LGK(+42.01)GTVPVGR	1024,6	513,31	2	0,9		86	Acetylation (K)	L[57.0]K(+42.01)GTVPVG[156.1]
K(+42.01)ALVGVNK	869,53	435,77	2	0,4		86	Acetylation (K)	[241.1]LVGVNK
AHAYSEK(+42.01)K	974,48	488,25	2	0,7	6,04E+07	86	Acetylation (K)	[71.0]HAYSEK(+42.01)K
SK(+42.01)ELVLAD(+14.02)ALLLLR	1609	805,5	2	1,9	5,54E+06	86	Acetylation (K); Methylation(D E)	[257.1]E[113.1]VLAD(+14.02)ALLLL[156.1
RE(+197.05)LAYTLRK	1345,7	673,87	2	8,7	3,29E+07	86	Glycerylphosphorylethanolamine	[482.2]LAYTLRK
RAGD(+14.02)FD(+14.02)AGLSK	1163,6	582,8	2	-9,6	4,34E+07	86	Methylation(D E)	[156.1]AGD(+14.02)FD(+14.02)A[57.0]LSK
NVSLE(+14.02)LK	815,48	408,74	2	-1,5	1,27E+08	86	Methylation(D E)	NVSL[143.1]LK
EKVD(+14.02)R	659,36	330,69	2	-0,4	1,01E+07	86	Methylation(D E)	EKVD(+14.02)[156.1]
TLLD(+14.02)GEESR	1032,5	517,26	2	0,3		86	Methylation(D E)	TLLD(+14.02)[57.0]EES[156.1]
KLE(+14.02)DNPK	856,47	429,24	2	-0,9	5,32E+06	86	Methylation(D E)	KLE(+14.02)DNP[128.1]
VEE(+14.02)RLK	786,46	394,24	2	1,4	1,77E+06	86	Methylation(D E)	VEE(+14.02)[156.1]LK
E(+14.02)EE(+14.02)TNPK	873,41	437,71	2	-1,1	3,43E+06	86	Methylation(D E)	[143.1]EE(+14.02)TNPK
VE(+14.02)QD(+14.02)LAEYK	1121,6	561,79	2	9,8	1,79E+07	86	Methylation(D E)	[242.1]QD(+14.02)LAEYK
KGLD(+14.02)DNPK	899,47	450,74	2	-1,1		86	Methylation(D E)	[185.1]LD(+14.02)DNP[128.1]
MDSTD(+14.02)PAYSEKK(+42.01)	1426,6	714,33	2	9,4	2,19E+09	86	Methylation(D E); Acetylation (K)	MDSTD(+14.02)P[71.0]YSEK[170.1]
VD(+14.02)TK(+42.01)LR	786,46	394,24	2	-0,3	1,38E+08	86	Methylation(D E); Acetylation (K)	VD(+14.02)T[170.1]LR
LPLY(+79.97)FDK	974,45	488,23	2	-1,2	5,34E+06	86	Phosphorylation (STY)	LPLY(+79.97)FD[128.1]

ST(+79.97)DELVK	870,37	436,2	2	5,2	1,01E+06	86	Phosphorylation (STY)	[268.0]DELVK
Y(+79.97)LQEDK	874,35	438,18	2	-9,2	3,41E+05	86	Phosphorylation (STY)	[243.0]L[128.1]EDK
APVLSDSAS(+79.97)K	1053,5	527,74	2	-9,1	3,05E+06	86	Phosphorylation (STY)	APVLSDS[366.1]
LY(+79.97)DPK	714,3	358,16	2	4,2		86	Phosphorylation (STY)	LY(+79.97)DPK
T(+79.97)LLVYK	815,42	408,72	2	-2,6		86	Phosphorylation (STY)	T(+79.97)LLVY[128.1]
LPLY(+79.97)DYK	990,45	496,23	2	-1,6	3,78E+06	86	Phosphorylation (STY)	[113.1]PLY(+79.97)[278.1]K
S(+79.97)LQDVYK	931,41	466,71	2	2,7		86	Phosphorylation (STY)	[167.0]L[128.1]DVYK
MKT(+79.97)TFTM(+15.99)K	1082,5	542,24	2	8,9	1,09E+07	86	Phosphorylation (STY); Oxidation (M)	MKT(+79.97)TF[101.0]M(+15.99)[128.1]
K(+42.01)GGLGTVPVGR	1081,6	541,82	2	1	3,87E+09	85	Acetylation (K)	[170.1]GGLGTVPVG[156.1]
EK(+42.01)PRYTK	962,52	482,27	2	-0,3	2,06E+06	85	Acetylation (K)	[299.1]PRYTK
VAAEK(+42.01)PR	811,46	406,73	2	0,1	1,43E+08	85	Acetylation (K)	VA[71.0]EK(+42.01)PR
K(+42.01)GVLK	585,38	586,39	1	-0,2	8,11E+06	85	Acetylation (K)	[227.1]VLK
NVSVTK(+42.01)LR	957,56	320,19	3	-0,3		85	Acetylation (K)	[114.0]VSVTK(+42.01)LR
AK(+42.01)LVGVNK	869,53	435,77	2	0,1		85	Acetylation (K)	[71.0]K(+42.01)LVG[99.1]NK
K(+42.01)VGVLK	684,45	685,46	1	1,4	1,91E+07	85	Acetylation (K)	[170.1]V[57.0]VLK
STTTK(+42.01)HLYK	1119,6	560,8	2	0,7	1,19E+06	85	Acetylation (K)	STTT[420.2]YK
K(+42.01)PLY(+79.97)DYK	1047,5	524,74	2	-2,9		85	Acetylation (K); Phosphorylation (STY)	[170.1]PLY(+79.97)DY[128.1]
E(+14.02)QLVGVNK	899,51	450,76	2	-2	8,33E+08	85	Methylation(D E)	[271.1]LVGVNK
E(+14.02)D(+14.02)D(+14.02)QTPK	873,41	437,71	2	0,4	3,43E+06	85	Methylation(D E)	[143.1]D(+14.02)D(+14.02)QTPK
VD(+14.02)TGLLGHAK	1023,6	512,79	2	-4	8,76E+06	85	Methylation(D E)	VD(+14.02)TGLL[57.0]H[199.1]
QLE(+14.02)DNPK	856,43	429,22	2	0,4	2,57E+07	85	Methylation(D E)	[128.1]LE(+14.02)DNPK
EAAD(+14.02)SLYQSK	1124,5	563,28	2	0,8	3,02E+08	85	Methylation(D E)	[200.1]AD(+14.02)SLYQS[128.1]
TNYD(+14.02)QLAEQNR	1364,6	683,32	2	-0,2	4,03E+08	85	Methylation(D E)	[378.2]D(+14.02)QLAEQNR
VLDD(+14.02)LATLR	1028,6	515,3	2	-0,1	6,26E+07	85	Methylation(D E)	[99.1]LDD(+14.02)L[172.1]LR
QLE(+14.02)DNPK	856,43	429,22	2	-0,8	2,57E+07	85	Methylation(D E)	QLE(+14.02)[229.1]PK
TTM(+15.99)C(+57.02)LYK	931,41	466,71	2	-6,3	1,09E+06	85	Oxidation (M); Carbamidomethylation	TT[307.1]LYK
KT(+79.97)VRLPLPK	1130,7	566,33	2	-9,8	8,39E+06	85	Phosphorylation (STY)	[309.1]VRLPLP[128.1]
T(+79.97)ALLTFK	872,44	437,23	2	-1,4	2,76E+08	85	Phosphorylation (STY)	T(+79.97)ALLTF[128.1]

FS(+79.97)LGVK	729,35	365,68	2	1,5	2,39E+08	85	Phosphorylation (STY)	[147.1]S(+79.97)L[57.0]VK
TVS(+79.97)LLAYTLGVK	1343,7	672,86	2	0,6	3,73E+07	85	Phosphorylation (STY)	[367.1]LLAYTLGVK
T(+79.97)ALLTFK	872,44	437,23	2	-1,1	2,76E+08	85	Phosphorylation (STY)	T(+79.97)ALLTF[128.1]
APVLSDWS(+79.97)K	1081,5	541,75	2	-5,2	2,18E+07	85	Phosphorylation (STY)	[71.0]PVLSD[186.1]S(+79.97)K
T(+79.97)LLVYK	815,42	408,72	2	-3,4		85	Phosphorylation (STY)	T(+79.97)LLVY[128.1]
K(+42.01)LYK	592,36	593,37	1	1,5	1,42E+08	84	Acetylation (K)	[170.1]LYK
K(+42.01)GVLK	585,38	586,39	1	0,4	1,98E+07	84	Acetylation (K)	[227.1]VLK
VAK(+42.01)SLPR	811,49	406,75	2	-1,2	6,31E+07	84	Acetylation (K)	VAK(+42.01)[87.0]LPR
STTTK(+42.01)HLYK	1119,6	560,8	2	0,3		84	Acetylation (K)	[87.0]TTT[307.2]LYK
SVPVK(+42.01)DLR	954,55	319,19	3	-1	1,87E+06	84	Acetylation (K)	SVPV[285.1]L[156.1]
K(+42.01)VGVLK	684,45	685,46	1	-0,5	1,91E+07	84	Acetylation (K)	[170.1]V[57.0]VLK
QK(+42.01)LVGVNK	926,55	464,28	2	-1,5		84	Acetylation (K)	[298.2]LVGVNK
K(+42.01)PLQDVYGK	1088,6	545,3	2	-0,1	7,18E+06	84	Acetylation (K)	[170.1]PLQD[99.1]YGK
TLRK(+42.01)E(+14.02)LLVGVNK	1424,9	713,44	2	-1	4,45E+08	84	Acetylation (K); Methylation(D E)	T[582.4]LLVGVNK
KVK(+42.01)GSLS(+79.97)NK	1081,6	541,79	2	8,4	3,54E+06	84	Acetylation (K); Phosphorylation (STY)	[227.2]K(+42.01)GSLS(+79.97)NK
C(+57.02)QWLE(+14.02)K	876,42	439,21	2	-9,2		84	Carbamidomethylation; Methylation(D E)	[160.0]QWLE(+14.02)K
APVNHC(+57.02)M(+15.99)VK	1070,5	536,25	2	-6,8		84	Carbamidomethylation; Oxidation (M)	[71.0]PVNH[307.1]VK
D(+14.02)YD(+14.02)LQVTVGR	1192,6	597,31	2	-0,7	5,46E+07	84	Methylation(D E)	[292.1]D(+14.02)LQVTVGR
ALELD(+14.02)D(+14.02)ALQK	1142,6	572,32	2	1,7	2,69E+07	84	Methylation(D E)	[184.1]ELD(+14.02)D(+14.02)ALQK
TQALVD(+14.02)LVK	999,6	500,81	2	-0,3	4,51E+07	84	Methylation(D E)	[229.1]ALVD(+14.02)LVK
D(+14.02)YD(+14.02)LQLTAGR	1178,6	590,3	2	-2,7	5,06E+08	84	Methylation(D E)	[292.1]D(+14.02)LQLTAGR
GYDD(+14.02)LVK	822,41	412,21	2	1,8	3,30E+07	84	Methylation(D E)	[220.1]DD(+14.02)LVK
SMDD(+14.02)LVK	820,4	411,21	2	-4,9	1,94E+07	84	Methylation(D E)	[218.1]DD(+14.02)LVK
VATE(+14.02)GPLR	855,48	428,75	2	1,2	9,00E+08	84	Methylation(D E)	VATE(+14.02)[154.1]LR
GYDD(+14.02)LVK	822,41	412,21	2	1,8	3,30E+07	84	Methylation(D E)	[220.1]DD(+14.02)LVK
LE(+14.02)E(+197.05)K	728,34	365,17	2	-5,4	5,06E+05	84	Methylation(D E); Glycerylphosphorylethanolamine	L[143.1]E(+197.05)K
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	0,1		84	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)[57.0]K

EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-2,5		84	Methylation(D E); Oxidation (M)	EAAD(+14.02)M(+15.99)[57.0]K
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	-1,3	3,11E+06	84	Methylation(D E); Oxidation (M)	[200.1]AD(+14.02)M(+15.99)GK
AEAD(+14.02)M(+15.99)GK	750,32	376,17	2	-1	3,11E+06	84	Methylation(D E); Oxidation (M)	[200.1]AD(+14.02)M(+15.99)GK
RAWM(+15.99)EAK(+42.01)SK	1163,6	582,8	2	5,5	4,34E+07	84	Oxidation (M); Acetylation (K)	[413.2]M(+15.99)EAK(+42.01)SK
APVLSS(+79.97)SYK	1030,5	516,24	2	1,4	3,45E+08	84	Phosphorylation (STY)	APVLS[254.0]Y[128.1]
TY(+79.97)ELVK	831,38	416,7	2	6,8	4,87E+08	84	Phosphorylation (STY)	[344.1]ELVK
S(+79.97)PLAGDVYK	1028,5	515,24	2	0,5	4,90E+07	84	Phosphorylation (STY)	S(+79.97)PLA[172.0]VYK
LPLY(+79.97)DFK	974,45	488,24	2	4,6		84	Phosphorylation (STY)	[113.1]PLY(+79.97)[390.2]
T(+79.97)ELVGVNK	938,45	470,23	2	-9,9	2,52E+07	84	Phosphorylation (STY)	[181.0]ELVGVNK
T(+79.97)GLVGVGGGK	923,45	462,73	2	2	9,94E+07	84	Phosphorylation (STY)	[238.0]LVGVGGGK
LPLY(+79.97)FDK	974,45	488,23	2	-0,6	2,41E+07	84	Phosphorylation (STY)	[113.1]PLY(+79.97)[390.2]
LPLY(+79.97)DFK	974,45	488,24	2	7,5		84	Phosphorylation (STY)	LPLY(+79.97)[390.2]
T(+79.97)GLVGVGGGK	923,45	462,73	2	1,7	9,94E+07	84	Phosphorylation (STY)	[238.0]LVGVGGGK
T(+79.97)ALLTFK	872,44	437,23	2	-2,3	2,76E+08	84	Phosphorylation (STY)	T(+79.97)ALLTF[128.1]
Y(+79.97)GTTLLK	874,42	438,22	2	9,8	9,42E+05	84	Phosphorylation (STY)	Y(+79.97)[57.0]TTLL[128.1]
T(+79.97)ELVGVNK	938,45	470,23	2	-7,9	2,65E+07	84	Phosphorylation (STY)	[181.0]ELVGVNK
AT(+79.97)VLLAYTLGVK	1327,7	664,86	2	-1,5	3,28E+07	84	Phosphorylation (STY)	[351.1]LLAYTL[57.0]VK
QY(+79.97)SK(+42.01)R	802,34	402,18	2	7,7		84	Phosphorylation (STY); Acetylation (K)	[128.1]Y(+79.97)SK(+42.01)[156.1]
NVSVTK(+42.01)LR	957,56	320,19	3	0,2	1,39E+09	83	Acetylation (K)	[213.1]SV[101.0]K(+42.01)LR
K(+42.01)GVLK	585,38	586,39	1	-3,4	1,98E+07	83	Acetylation (K)	[227.1]VLK
RGLTEVK(+42.01)	843,48	422,75	2	0,9	6,06E+08	83	Acetylation (K)	[213.1]LTEVK(+42.01)
VYAQPK(+42.01)	746,4	374,21	2	-1,2		83	Acetylation (K)	VYA[225.1]K(+42.01)
QK(+42.01)LVGVNK	926,55	464,28	2	-0,3		83	Acetylation (K)	[128.1]K(+42.01)LV[57.0]VNK
QK(+42.01)AVGVLK	883,55	442,78	2	-0,4	1,63E+08	83	Acetylation (K)	[128.1]K(+42.01)AV[156.1]LK
VEAK(+42.01)LR	756,45	379,23	2	-0,4	7,18E+07	83	Acetylation (K)	[99.1]EA[170.1]LR
TPSFEAK(+42.01)SK	1035,5	518,77	2	2,1	8,86E+07	83	Acetylation (K)	[198.1]SFEA[170.1]SK
TNVEADLDK(+42.01)R	1201,6	601,8	2	-1,7	3,83E+07	83	Acetylation (K)	[215.1]VEADLDK(+42.01)[156.1]
K(+42.01)EDNPK	771,38	386,69	2	-1,5	3,79E+06	83	Acetylation (K)	[170.1]EDN[225.1]
NVSVTK(+42.01)LR	957,56	320,19	3	-1,9	1,39E+09	83	Acetylation (K)	[213.1]SV[101.0]K(+42.01)LR

QK(+42.01)LVGVNK	926,55	464,28	2	-0,4	8,54E+07	83	Acetylation (K)	[298.2]LVGVNK
RE(+197.05)LAYTLRK	1345,7	673,87	2	8,6	7,06E+07	83	Glycerylphosphorylethanolamine	[482.2]LAYTLRK
E(+197.05)LTLLDAYK	1261,6	631,82	2	9,9	1,72E+06	83	Glycerylphosphorylethanolamine	[540.2]LLDAYK
E(+197.05)VNGS(+79.97)K	909,32	455,67	2	-3,5	1,07E+07	83	Glycerylphosphorylethanolamine; Phosphorylation (STY)	[425.2]NGS(+79.97)K
EAVLD(+14.02)NAR	900,47	451,24	2	0,2	3,90E+06	83	Methylation(D E)	[200.1]VLD(+14.02)NA[156.1]
RE(+14.02)PATLNDK	1056,6	529,29	2	-0,3	3,64E+07	83	Methylation(D E)	[299.2]PATLNDK
TNYD(+14.02)QLAEQNR	1364,6	683,32	2	-0,5	4,03E+08	83	Methylation(D E)	[378.2]D(+14.02)QLAEQN[156.1]
D(+14.02)PGEFK	705,33	353,67	2	-5,1	2,51E+08	83	Methylation(D E)	D(+14.02)P[57.0]EFK
KGLD(+14.02)DNPK	899,47	450,74	2	-1,1		83	Methylation(D E)	[185.1]LD(+14.02)DNP[128.1]
E(+14.02)E(+14.02)VAVRLK	970,58	486,3	2	-0,4	3,21E+09	83	Methylation(D E)	[286.1]VAVRLK
E(+14.02)E(+14.02)VAVRLK	970,58	486,3	2	-0,2	2,95E+08	83	Methylation(D E)	[286.1]VAVRLK
D(+14.02)YD(+14.02)LQVTVGR	1192,6	597,31	2	2	5,46E+07	83	Methylation(D E)	[292.1]D(+14.02)LQVTVG[156.1]
SSE(+14.02)QNK	705,33	353,67	2	1,6	4,28E+06	83	Methylation(D E)	[174.1]E(+14.02)QNK
REE(+14.02)GLVK	843,48	422,75	2	0,2	6,06E+08	83	Methylation(D E)	[156.1]EE(+14.02)[170.1]VK
AE(+14.02)PLNEVYK	1075,6	538,78	2	0,3	1,35E+06	83	Methylation(D E)	[214.1]PLNE[99.1]YK
RAAWM(+15.99)D(+14.02)AK(+42.01)SK	1234,6	618,32	2	4,5	1,29E+07	83	Oxidation (M); Methylation(D E); Acetylation (K)	[298.2]WM(+15.99)D(+14.02)AK(+42.01)SK
QNLSY(+79.97)LR	972,44	487,23	2	-7,6		83	Phosphorylation (STY)	[242.1]LSY(+79.97)LR
T(+79.97)ALLTFK	872,44	437,23	2	-1,7	2,76E+08	83	Phosphorylation (STY)	T(+79.97)ALL[376.2]
NS(+79.97)LLYR	844,38	423,2	2	-0,8	2,73E+07	83	Phosphorylation (STY)	[281.0]LLYR
SYS(+79.97)ELVK	904,39	453,21	2	7,7		83	Phosphorylation (STY)	[417.1]ELVK
LPLY(+79.97)DFK	974,45	488,24	2	8	3,42E+07	83	Phosphorylation (STY)	[113.1]PLY(+79.97)D[275.2]
T(+79.97)LLVYK	815,42	408,72	2	-3,3		83	Phosphorylation (STY)	T(+79.97)LLVY[128.1]
MS(+79.97)LLYR	861,38	431,7	2	1,7	2,84E+07	83	Phosphorylation (STY)	[298.0]LLYR
T(+79.97)ALLTFK	872,44	437,23	2	-2,8	2,76E+08	83	Phosphorylation (STY)	T(+79.97)ALLT[275.2]
T(+79.97)ALLTFK	872,44	437,23	2	-1,8	3,75E+07	83	Phosphorylation (STY)	T(+79.97)ALL[376.2]
LPLY(+79.97)FDK	974,45	488,23	2	-0,3	2,41E+07	83	Phosphorylation (STY)	[113.1]PLY(+79.97)[390.2]
LPLY(+79.97)DYK	990,45	496,23	2	-3	1,83E+07	83	Phosphorylation (STY)	[113.1]PLY(+79.97)[278.1]K
VEAK(+42.01)LR	756,45	379,23	2	-1,4	7,18E+07	82	Acetylation (K)	[99.1]E[241.1]LR

THLNLVVHK(+42.01)	1101,6	551,82	2	0,3	4,08E+08	82	Acetylation (K)	THLNLVVH[170.1]
VYAQPK(+42.01)	746,4	374,2	2	-1,6		82	Acetylation (K)	[99.1]YA[128.1]PK(+42.01)
VETK(+42.01)LR	786,46	394,24	2	-5,3		82	Acetylation (K)	VET[170.1]L[156.1]
K(+42.01)GVLK	585,38	586,39	1	0,6	1,98E+07	82	Acetylation (K)	[227.1]VLK
GSFFGK(+42.01)	683,33	342,67	2	-0,5	3,07E+06	82	Acetylation (K)	[57.0]SFF[57.0]K(+42.01)
DVTSVK(+42.01)LR	958,54	480,28	2	0,2	7,82E+07	82	Acetylation (K)	[115.0]V[188.1]VK(+42.01)L[156.1]
THLNLVVHK(+42.01)	1101,6	551,82	2	-0,1	1,53E+08	82	Acetylation (K)	THLNLVVH[170.1]
TSTTK(+42.01)HLYK	1119,6	560,8	2	0,2	1,56E+07	82	Acetylation (K)	[101.0]ST[408.2]LYK
NVSVTK(+42.01)LR	957,56	320,19	3	-1,2	1,39E+09	82	Acetylation (K)	[213.1]SV[271.2]LR
AK(+42.01)LVGVNK	869,53	435,77	2	-0,4		82	Acetylation (K)	[71.0]K(+42.01)LV[156.1]NK
VPSHAVK(+42.01)R	934,53	312,52	3	-0,3	4,35E+06	82	Acetylation (K)	VPSH[170.1]K(+42.01)[156.1]
EFEAK(+42.01)SK	879,43	440,72	2	1,8	3,68E+08	82	Acetylation (K)	[276.1]EA[170.1]SK
GSFFK(+42.01)	626,31	314,16	2	-0,6		82	Acetylation (K)	[57.0]SFFK(+42.01)
NVSVTK(+42.01)LR	957,56	479,79	2	-0,6	1,51E+09	82	Acetylation (K)	[114.0]VS[370.2]LR
E(+197.05)FTLNSR	1062,5	532,24	2	-10		82	Glycerylphosphorylethanolamine	[473.2]TLNSR
E(+197.05)RLAYTLRK	1345,7	673,87	2	9,5	1,12E+08	82	Glycerylphosphorylethanolamine	[482.2]LAYTLRK
E(+197.05)PLVEM(+15.99)VPGK	1310,6	656,32	2	-0,1	1,86E+07	82	Glycerylphosphorylethanolamine; Oxidation (M)	[423.1]LVEM(+15.99)VPGK
E(+14.02)NTTEVK	833,41	417,71	2	-0,1	1,11E+08	82	Methylation(D E)	[257.1]TTEVK
LVND(+14.02)LTEFAK	1162,6	582,32	2	-0,2	7,23E+06	82	Methylation(D E)	[212.2]ND(+14.02)LTEFAK
RAGD(+14.02)FD(+14.02)AGLSK	1163,6	582,8	2	-9,3	4,34E+07	82	Methylation(D E)	[227.1]G[129.0]FD(+14.02)A[57.0]LSK
SSE(+14.02)KNK	705,37	353,69	2	4	5,58E+06	82	Methylation(D E)	[317.1]KNK
E(+14.02)QLVGVNK	899,51	450,76	2	0,2	8,33E+08	82	Methylation(D E)	[271.1]LVGVNK
GPLTE(+14.02)NPK	868,47	435,24	2	0,7	3,63E+07	82	Methylation(D E)	[57.0]PL[244.1]NPK
LGPLE(+14.02)TDFK	1032,5	517,28	2	0,3	4,58E+07	82	Methylation(D E)	[267.2]LE(+14.02)TDFK
VAE(+14.02)GTLPR	855,48	428,75	2	0,6	9,00E+08	82	Methylation(D E)	VA[301.1]LPR
AE(+14.02)DQYMTK	998,44	500,22	2	-3,2	5,37E+07	82	Methylation(D E)	[214.1]D[128.1]YMTK
AEAD(+14.02)M(+15.99)GK	750,32	376,17	2	-0,9	3,11E+06	82	Methylation(D E); Oxidation (M)	[200.1]AD(+14.02)M(+15.99)GK
EAAD(+14.02)M(+15.99)GK	750,32	376,17	2	0		82	Methylation(D E); Oxidation (M)	[200.1]AD(+14.02)M(+15.99)GK
RWM(+15.99)EAK(+42.01)SK	1092,5	547,28	2	5,9		82	Oxidation (M); Acetylation (K)	[342.2]M(+15.99)EA[170.1]SK

RQWM(+15.99)EAK(+42.01)SK	1220,6	611,31	2	6	3,03E+06	82	Oxidation (M); Acetylation (K)	[284.2]WM(+15.99)EA[170.1]SK
RAAWM(+15.99)EAK(+42.01)SK	1234,6	618,32	2	3,9	1,29E+07	82	Oxidation (M); Acetylation (K)	[298.2]WM(+15.99)EAK(+42.01)SK
VESATM(+15.99)K(+42.01)	822,38	412,2	2	2		82	Oxidation (M); Acetylation (K)	[99.1]ESATM(+15.99)[170.1]
TM(+15.99)LDE(+14.02)M(+15.99)K	912,39	457,21	2	8,1	2,45E+07	82	Oxidation (M); Methylation(D E)	[248.1]LDE(+14.02)M(+15.99)K
WT(+79.97)ELVK	854,39	428,21	2	8,9	1,21E+08	82	Phosphorylation (STY)	[367.1]ELVK
T(+79.97)HAPGLK	802,37	402,19	2	-0,9		82	Phosphorylation (STY)	[318.1]AP[57.0]LK
T(+79.97)RPVLDK	907,45	454,73	2	-9,4	5,07E+06	82	Phosphorylation (STY)	[337.1]PVLDK
YSAAS(+79.97)YK	868,34	435,17	2	-9	7,99E+07	82	Phosphorylation (STY)	[250.1]AAS(+79.97)YK
KT(+79.97)FLAYTLGVK	1319,7	660,86	2	6,2	1,90E+07	82	Phosphorylation (STY)	[456.2]LAYTLGVK
LPLY(+79.97)DFK	974,45	488,24	2	6,4	3,42E+07	82	Phosphorylation (STY)	[113.1]PLY(+79.97)[390.2]
YSAAS(+79.97)YK	868,34	435,17	2	-9,4	5,74E+07	82	Phosphorylation (STY)	[250.1]AAS(+79.97)YK
T(+79.97)FELVK	815,38	408,7	2	9,2	1,09E+09	82	Phosphorylation (STY)	[328.1]ELVK
LS(+79.97)ASVGAHK(+42.01)	990,45	496,23	2	-8,7	5,49E+06	82	Phosphorylation (STY); Acetylation (K)	LS(+79.97)ASV[57.0]AH[170.1]
NRS(+79.97)HLAC(+57.02)K	1064,5	355,83	3	5	1,34E+07	82	Phosphorylation (STY); Carbamidomethylation	[270.1]S(+79.97)HLAC(+57.02)K
S(+79.97)YAQE(+14.02)LK	931,41	466,71	2	0,6	2,13E+06	82	Phosphorylation (STY); Methylation(D E)	[167.0]YA[271.1]LK
NVSVTK(+42.01)LR	957,56	320,19	3	-1,1		81	Acetylation (K)	[213.1]S[370.2]LR
NVSVTK(+42.01)LR	957,56	479,79	2	-0,2	1,04E+07	81	Acetylation (K)	[114.0]VS[370.2]LR
LGK(+42.01)GTVSGHK	1024,6	513,29	2	2,7		81	Acetylation (K)	LGK(+42.01)GTVS[322.2]
YDELGK(+42.01)	765,35	383,68	2	-0,2	2,52E+07	81	Acetylation (K)	YDEL[57.0]K(+42.01)
YDELGK(+42.01)	765,35	383,68	2	0,4	2,52E+07	81	Acetylation (K)	YDEL[227.1]
K(+42.01)VGVLK	684,45	685,46	1	3,8	1,91E+07	81	Acetylation (K)	[170.1]V[57.0]VLK
SSPAAAK(+42.01)MK	931,48	466,74	2	-9,8	1,81E+06	81	Acetylation (K)	SSPA[312.2]MK
NVSVTK(+42.01)LR	957,56	320,19	3	-2,1	1,39E+09	81	Acetylation (K)	[213.1]SV[101.0]K(+42.01)LR
YEELGK(+42.01)	779,37	390,69	2	-4,6	1,57E+09	81	Acetylation (K)	[163.1]EEL[227.1]
TNVEADLNK(+42.01)R	1200,6	601,31	2	1,5	3,04E+06	81	Acetylation (K)	[314.2]EADLNK(+42.01)[156.1]
K(+42.01)C(+57.02)LVRNK	958,54	480,28	2	4,9	2,01E+06	81	Acetylation (K); Carbamidomethylation	[330.1]LVRNK
K(+42.01)DDDPAYSD(+14.02)K	1208,5	605,27	2	-1,2	5,25E+06	81	Acetylation (K); Methylation(D E)	[515.2]PAYSD(+14.02)K

LPPDVC(+57.02)M(+15.99)K	974,46	488,23	2	-5	2,41E+07	81	Carbamidomethylation; Oxidation (M)	[113.1]PPD[406.1]K
RE(+197.05)TLLDARYK	1460,7	731,38	2	3,9		81	Glycerylphosphorylethanolamine	[482.2]T[113.1]LDARY[128.1]
E(+197.05)RLAYTLRK	1345,7	673,87	2	9,5	1,12E+08	81	Glycerylphosphorylethanolamine	[482.2]LAYTLRK
SE(+197.05)NLTTEVK	1216,6	609,29	2	-1,8		81	Glycerylphosphorylethanolamine	[413.1]NLTTEVK
APE(+14.02)AVGVLGK	953,55	477,78	2	-2,8	1,27E+07	81	Methylation(D E)	[311.2]AVGVLGK
AE(+14.02)PELQNVK	1040,6	521,29	2	5,7	4,46E+07	81	Methylation(D E)	[71.0]E(+14.02)PELQ[114.0]VK
FNE(+14.02)LK	663,36	332,69	2	-2,5	8,50E+06	81	Methylation(D E)	F[257.1]LK
QLE(+14.02)TNPK	842,45	422,23	2	-1	1,16E+06	81	Methylation(D E)	[128.1]LE(+14.02)T[211.1]K
VEE(+14.02)GVLK	786,45	394,23	2	-9,8		81	Methylation(D E)	VEE(+14.02)[156.1]LK
NVSLE(+14.02)LK	815,48	408,74	2	-1,5	1,27E+08	81	Methylation(D E)	NV[456.3]K
D(+14.02)HALLAYTLRK	1313,7	657,87	2	-7,8	1,25E+09	81	Methylation(D E)	[337.1]LLAYTLRK
VD(+14.02)TGLTAHK	954,51	478,26	2	-0,2	3,98E+08	81	Methylation(D E)	VD(+14.02)TGL[309.1]K
YE(+14.02)NYK	729,33	365,68	2	6,2	8,34E+06	81	Methylation(D E)	[163.1]E(+14.02)[114.0]YK
D(+14.02)YD(+14.02)LQQWAK	1193,6	597,79	2	-3	3,17E+07	81	Methylation(D E)	[292.1]D(+14.02)LQQWAK
AFAYGSD(+14.02)R	899,41	450,72	2	3,4	8,45E+01	81	Methylation(D E)	[71.0]FAYG[87.0]D(+14.02)[156.1]
E(+14.02)GVMPK	673,35	337,68	2	-3,9	1,46E+07	81	Methylation(D E)	[200.1]VM[97.1]K
STTTPPLD(+14.02)YK	1135,6	568,8	2	8,9	3,78E+06	81	Methylation(D E)	[87.0]TTT[194.1]LD(+14.02)YK
RE(+14.02)K(+42.01)LLVRGGK	1210,8	606,38	2	-8,5	1,22E+08	81	Methylation(D E); Acetylation (K)	[469.3]LLV[156.1]GGK
E(+14.02)S(+79.97)YK	619,23	310,62	2	5,9		81	Methylation(D E); Phosphorylation (STY)	[310.1]YK
EPWM(+15.99)D(+14.02)AGLSK	1162,5	582,27	2	-6,4	1,92E+07	81	Oxidation (M); Methylation(D E)	E[283.1]M(+15.99)D(+14.02)A[57.0]LSK
M(+15.99)LY(+79.97)VVK	847,39	424,7	2	-9,5	4,05E+06	81	Oxidation (M); Phosphorylation (STY)	[260.1]Y(+79.97)VVK
T(+79.97)FGAVVLGK	970,49	486,26	2	9,5		81	Phosphorylation (STY)	[328.1]GA[99.1]VLGK
KS(+79.97)GLLAYTLGVK	1328,7	665,36	2	-7,9	5,35E+07	81	Phosphorylation (STY)	[352.1]LLAYTLGVK
T(+79.97)FELVK	815,38	408,7	2	8,7	1,09E+09	81	Phosphorylation (STY)	[328.1]ELVK
LPLY(+79.97)DFK	974,45	488,23	2	-1,5	5,34E+06	81	Phosphorylation (STY)	[113.1]PLY(+79.97)[390.2]
Y(+79.97)RFEVLASPGK	1345,6	673,84	2	9,4	1,10E+07	81	Phosphorylation (STY)	[546.2]EVLASPGK
LPLY(+79.97)VS(+79.97)S(+79.97)K	1145,4	573,72	2	3,6	3,90E+06	81	Phosphorylation (STY)	LPLY(+79.97)V[462.1]

T(+79.97)SDELVK	870,37	436,2	2	4,7	4,18E+05	81	Phosphorylation (STY)	[268.0]DELVK
AT(+79.97)EAHLASSK	1093,5	547,75	2	1,4	4,94E+06	81	Phosphorylation (STY)	[252.1]EAHLASS[128.1]
LPLY(+79.97)FDK	974,45	488,23	2	0,6	2,41E+07	81	Phosphorylation (STY)	[113.1]PLY(+79.97)[390.2]
T(+79.97)GLVGVGGGK	923,45	462,73	2	2,2	9,94E+07	81	Phosphorylation (STY)	[238.0]LVGVGG[57.0]K
T(+79.97)ALLVMK	854,43	428,22	2	-6,7	7,61E+06	81	Phosphorylation (STY)	[181.0]ALLVM[128.1]
Y(+79.97)GNVVGNK	929,4	465,7	2	-7,9	3,93E+06	81	Phosphorylation (STY)	[513.2]V[57.0]NK
RVKY(+79.97)ADQM(+15.99)TSK	1421,6	711,83	2	4,1	1,48E+07	81	Phosphorylation (STY); Oxidation (M)	[255.2]KY(+79.97)ADQM(+15.99)TSK
LGK(+42.01)GTVPVGR	1024,6	513,31	2	-0,3		80	Acetylation (K)	L[57.0]K(+42.01)GTV[97.1]V[213.1]
SGYFK(+42.01)	642,3	322,16	2	-1,1	9,65E+07	80	Acetylation (K)	[144.1]YFK(+42.01)
STTK(+42.01)HLK	855,48	428,75	2	-1,4	3,17E+06	80	Acetylation (K)	STT[307.2]LK
K(+42.01)PLQDVYGK	1088,6	545,3	2	-0,9	9,20E+07	80	Acetylation (K)	[267.2]LQDVYGK
NVSVTK(+42.01)LR	957,56	320,19	3	-0,3		80	Acetylation (K)	[213.1]SVT[170.1]LR
K(+42.01)GGLGTVPVGR	1081,6	541,82	2	0,2	1,46E+09	80	Acetylation (K)	[170.1]GGLGTV[409.2]
K(+42.01)PLSPASYGK	1088,6	545,3	2	0,1	7,18E+06	80	Acetylation (K)	[170.1]PL[255.1]SYG[128.1]
RADYEAK(+42.01)SK	1108,6	555,28	2	-8,1	8,92E+06	80	Acetylation (K)	[505.2]EA[170.1]SK
EK(+42.01)PRYTK	962,52	321,85	3	0		80	Acetylation (K)	[396.2]RYTK
YDELGK(+42.01)	765,35	383,69	2	2,9	2,52E+07	80	Acetylation (K)	YDEL[227.1]
AK(+42.01)LVGVNK	869,53	435,77	2	-0,3		80	Acetylation (K)	[71.0]K(+42.01)LV[156.1]NK
AK(+42.01)LVGVNK	869,53	435,77	2	-0,4	5,20E+06	80	Acetylation (K)	[71.0]K(+42.01)LV[156.1]NK
TPSFEAK(+42.01)SK	1035,5	518,77	2	3	8,86E+07	80	Acetylation (K)	[285.1]FEA[170.1]SK
VESATYK(+42.01)	838,41	420,21	2	-1,8	5,05E+07	80	Acetylation (K)	[99.1]ESATY[170.1]
SVPVK(+42.01)DLR	954,55	319,19	3	-0,1	6,93E+08	80	Acetylation (K)	SV[481.3]LR
RK(+42.01)M(+15.99)VLDK	946,53	474,27	2	6	2,61E+07	80	Acetylation (K); Oxidation (M)	[326.2]M(+15.99)VLDK
K(+42.01)QLPLY(+79.97)VYK	1272,7	637,33	2	-8,3	2,23E+07	80	Acetylation (K); Phosphorylation (STY)	[298.2]LPL[342.1]YK
EPPPC(+57.02)S(+79.97)K	893,34	447,67	2	-1,5	1,84E+07	80	Carbamidomethylation; Phosphorylation (STY)	EPP[424.1]K
NPVE(+197.05)LK	895,44	448,73	2	-1,7	4,77E+07	80	Glycerylphosphorylethanolamine	[211.1]VE(+197.05)LK
RE(+197.05)LAYTLRK	1345,7	673,87	2	9,3	9,04E+07	80	Glycerylphosphorylethanolamine	[482.2]LAYTLRK

E(+197.05)VVLNYK(+42.01)	1102,5	552,28	2	8,8		80	Glycerylphosphorylethanolamine; Acetylation (K)	[425.2]VL[447.2]
TLEGKFD(+14.02)K	950,51	317,84	3	-0,4	1,92E+09	80	Methylation(D E)	TLE[332.2]D(+14.02)K
TDAYGTE(+14.02)K	897,41	449,71	2	-4,5	3,80E+06	80	Methylation(D E)	TDAY[158.1]E(+14.02)K
TSD(+14.02)PAYSKE(+14.02)R	1180,6	591,29	2	1,8	2,34E+07	80	Methylation(D E)	TSD(+14.02)PAY[514.3]
D(+14.02)HALLAYTLRK	1313,7	657,87	2	-7,9		80	Methylation(D E)	[337.1]LLAYTLRK
D(+14.02)GKFD(+14.02)K	736,38	369,19	2	-0,2	8,95E+07	80	Methylation(D E)	[314.2]FD(+14.02)K
TNYD(+14.02)QLAEQELR	1492,7	498,58	3	7	1,99E+07	80	Methylation(D E)	[215.1]YD(+14.02)[128.1]LAEQEL[156.1]
DGNDLVD(+14.02)YVPGK	1304,6	653,32	2	3		80	Methylation(D E)	[401.1]LVD(+14.02)YVPGK
LPLENYD(+14.02)K	1004,5	503,27	2	0,8	1,02E+07	80	Methylation(D E)	[113.1]PLE[277.1]D(+14.02)K
ATE(+14.02)LDSLLAD(+14.02)VK	1301,7	651,86	2	1,4	2,04E+07	80	Methylation(D E)	[315.1]LDSLLAD(+14.02)VK
VND(+14.02)LDPELQK	1183,6	592,81	2	0,3	3,31E+08	80	Methylation(D E)	[455.2]D[97.1]ELQK
THDKANM(+15.99)QM(+15.99)K(+42.01)	1276,6	639,29	2	3,3	8,42E+07	80	Oxidation (M); Acetylation (K)	[238.1]DKA[261.1]QM(+15.99)[170.1]
T(+79.97)STTGHLLYK	1199,6	600,78	2	-7,7	1,90E+07	80	Phosphorylation (STY)	[181.0]STT[57.0]HLLY[128.1]
EHALLAY(+79.97)TLGVK	1393,7	697,86	2	-0,1	3,44E+07	80	Phosphorylation (STY)	E[208.1]LLA[344.1]L[57.0]VK
DY(+79.97)DVPLPK	1025,4	513,74	2	9,9		80	Phosphorylation (STY)	[115.0]Y(+79.97)DVPLP[128.1]
KS(+79.97)GLLAYTLGVK	1328,7	665,36	2	-8,9	5,35E+07	80	Phosphorylation (STY)	K[224.0]LLAYTL[57.0]VK
LPLY(+79.97)FDK	974,45	488,23	2	-2,2	8,23E+07	80	Phosphorylation (STY)	[113.1]PLY(+79.97)[390.2]
YSPVLDAEPT(+79.97)K	1298,6	650,29	2	-6,4	2,32E+07	80	Phosphorylation (STY)	YSPVL[412.2]T(+79.97)K
TY(+79.97)ELVK	831,38	416,7	2	8,6	4,87E+08	80	Phosphorylation (STY)	[344.1]ELVK
LPLY(+79.97)FDK	974,45	488,23	2	-2,2	8,23E+07	80	Phosphorylation (STY)	[113.1]PLY(+79.97)[390.2]
T(+79.97)FAVRLK	913,48	457,75	2	7,5	7,16E+07	80	Phosphorylation (STY)	[328.1]AVRLK
LPLY(+79.97)DYK	990,45	496,23	2	0,5	2,26E+06	80	Phosphorylation (STY)	[113.1]PLY(+79.97)[278.1]K
S(+79.97)LLLAYTLGVK	1256,7	629,35	2	-0,5	1,68E+07	80	Phosphorylation (STY)	[280.1]LLAYTLGVK
T(+79.97)FELVK	815,38	408,7	2	6,9	1,09E+09	80	Phosphorylation (STY)	[328.1]ELVK
S(+79.97)PLADVYGK	1028,5	515,24	2	1	4,90E+07	80	Phosphorylation (STY)	S(+79.97)PLADVY[185.1]
SPVLMS(+79.97)K	840,38	421,2	2	-2,4	3,81E+07	80	Phosphorylation (STY)	SPVL[426.1]
LPLY(+79.97)DYK	990,45	496,23	2	-2,6	1,15E+06	80	Phosphorylation (STY)	[113.1]PLY(+79.97)[278.1]K
TY(+79.97)ELVK	831,38	416,7	2	9,8	6,97E+05	80	Phosphorylation (STY)	[344.1]ELVK
S(+79.97)SAAVGVLGK	967,47	484,74	2	-0,2	2,55E+07	80	Phosphorylation (STY)	[325.1]AVGVLGK