

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Modeling and exploring human IRE1 as a strategy to design novel inhibitors: a computational approach

ANTONIO CARLESSO



UNIVERSITY OF GOTHENBURG

Department of Chemistry and Molecular Biology

UNIVERSITY OF GOTHENBURG

Göteborg, Sweden 2020

Modeling and exploring human IRE1 as a strategy to design novel inhibitors: a computational approach

Antonio Carlesso

Department of Chemistry and Molecular Biology University of Gothenburg
SE-412 96 Göteborg
Sweden

© Antonio Carlesso, 2019

ISBN: 978-91-7833-754-5 (PRINT)

ISBN: 978-91-7833-755-2 (PDF)

<http://hdl.handle.net/2077/62404>

Printed by BrandFactory, Kållerød, Sweden, 2019

Cover design: © Filippo Lessio

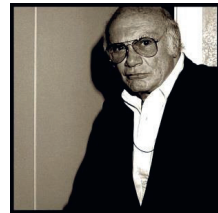
Author photograph: Antonio Carlesso by © Matteo Tamini

«If we say 'all animals', that does not pass for zoology; for the same reason we see at once that the words absolute, divine, eternal, and so on do not express what is implied in them. »

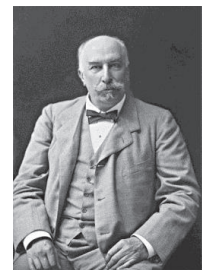
Preface to the Phenomenology of Spirit (1807).



«La vita di un regista sono i suoi film. Non tutta la sua vita certo, ma quella parte di essa attraverso la quale ha espresso la sua relazione con il mondo, con le idee e con gli uomini »



« ... le leggi devono tener conto anche dei difetti e delle manchevolezze di un paese... Il sarto che ha da vestire un gobbo, se non tiene conto della gobba, non riesce. »



Modeling and exploring human IRE1 as a strategy to design novel inhibitors: a computational approach

ANTONIO CARLESSO

Department of Chemistry and Molecular Biology
University of Gothenburg

ABSTRACT: Inositol Requiring Enzyme 1 (IRE1) is a bifunctional serine/threonine kinase and endoribonuclease that is the major mediator of the Unfolded Protein Response (UPR) during endoplasmic reticulum (ER) stress. The association of IRE1 dysregulation with a wide range of human diseases, stimulated research towards the discovery of small organic molecules able to modulate IRE1 signalling, and to potentially be used as novel therapeutics. In this thesis we performed *in silico* three-dimensional (3D) molecular modeling analysis encompassing: (i) the selection of suitable protocols for docking and virtual screening in the IRE1 serine/threonine kinase and endoribonuclease domains studies, (ii) the exploration of IRE1 and PERK ligand interaction networks, (iii) the study of IRE1-ligand recognition phenomena in order to understand the mechanism of action of IRE1 small organic modulators and (iv) offers important insights relevant to hit-discovery and lead optimization of novel IRE1 modulators.

Our structure-based drug design approach provides useful information for designing improved IRE1 ligands, as confirmed by one soon-to-be-filed patents on new inhibitors targeting IRE1, developed during the PhD period.

KEYWORDS: ER stress, unfolded protein response, cancer, inflammation, neurodegeneration, therapeutic targets, molecular docking, molecular dynamics.

LIST OF PUBLICATIONS AND CONTRIBUTION REPORT

This thesis is based on the work presented in the following papers:

Paper I

Carlesso, A.; Chintha, C.; Gorman, A. M.; Samali, A.; Eriksson, L. A. Binding Analysis of the Inositol-Requiring Enzyme 1 Kinase Domain. *ACS Omega* 2018, 3 (10), 13313–13322.

A.C. performed the calculations and initial analyses, and wrote the manuscript.

Paper II

Carlesso, A.; Chintha, C.; Gorman, A. M.; Samali, A.; Eriksson, L. A. Merits and Pitfalls of Conventional and Covalent Docking in Identifying New Hydroxyl Aryl Aldehyde like Compounds as Human IRE1 Inhibitors. *Sci. Rep.* 2019, 9 (1), 3407.

A.C. performed the calculations, wrote the first draft, and revised the text into final form.

Paper III

Carlesso, A.; Eriksson, L. A. Selective Inhibition of IRE1 Signalling Mediated by MKC9989: New Insights from Molecular Docking and Molecular Dynamics Simulations. *ChemistrySelect* 2019, 4 (11), 3199–3203.

A.C.: Planning and conceiving the study; simulations and analyses; writing of manuscript.

Paper IV

Carlesso A.; Chintha C.; Gorman A. M.; Samali A.; Eriksson L. A. Effect of Kinase Inhibiting RNase Attenuator (KIRA) Compounds on the Formation of Face-to-Face Dimers of Inositol-Requiring Enzyme 1: Insights from Computational Modeling. *Int J Mol Sci.* 2019;20(22). doi: 10.3390/ijms20225538.

A.C. performed the computations, analyzed the data, wrote the first draft, and revised the text.

Paper V

Chintha, C.; **Carlesso, A.**; Gorman, A. M.; Samali, A.; Eriksson, L. A. Molecular modeling provides structural basis for PERK inhibitor selectivity towards RIPK1. Status: *under revision*

A.C. analysed the data and finalized the submitted manuscript.

Paper VI

Doultinos, D.; **Carlesso, A.**; Chintha, C.; Rainot, A.; Paton, J.C.; Paton, A.W.; Samali, A.; Chevet, E.; Eriksson, L. A. Peptidomimetic-based identification of FDA approved compounds inhibiting IRE1 activity
Status: *Manuscript in preparation*

A.C. analysed the Molecular Dynamics (MD) simulation data and wrote the parts about MD simulation of the paper.

Other works not presented in this thesis:

Review I

Almanza, A.; **Carlesso, A.**; Chintha, C.; Creedican, S.; Doultinos, D.; Leuzzi, B.; Luís, A.; McCarthy, N.; Montibeller, L.; More, S.; et al. Endoplasmic Reticulum Stress Signalling – from Basic Mechanisms to Clinical Applications. FEBS J. 2018, 241-278.

Review II

Pelizzari R.D., Doultinos D., **Carlesso A.**, Eriksson L.A., Chevet E. Pharmacological targeting of IRE1 in health and disease: current status and future challenges.
Status: *Submitted*

Table of Contents

1. Unfolded Protein Response (UPR)	1
1.1 General overview.....	1
1.2 IRE1	1
1.3 PERK	2
1.4 ATF6	3
2. The Kinome World.....	3
3. Targeting IRE1 signaling	3
3.1. Pharmacological modulators of IRE1.....	3
3.2. Ligands that interact with IRE1 RNase domain	3
3.3. Ligands that interact with the IRE1 kinase domain	4
4. Methodology	5
4.1 Molecular mechanics.....	5
4.1.1 Force Fields.....	5
4.2 Energy Minimization.....	9
4.3.1 Scoring functions.....	12
4.4 Classical Molecular Dynamics (MD).....	14
4.5. Experimental Methods used for IRE1 drug design.....	17
4.5.1. Micro-scale thermophoresis (MST)	17
4.5.2. Fluorescence resonance energy transfer (FRET)	18
4.5.3. Protein kinases assay using radiolabeled ATP	18
5. RESULTS	19
PAPER I. Binding analysis of the Inositol-requiring enzyme 1 (IRE1) kinase domain	20
PAPER II. Merits and pitfalls of conventional and covalent docking in identifying new hydroxyl aryl aldehyde like compounds as human IRE1 inhibitors.....	21
PAPER III. Selective Inhibition of IRE1 Signalling mediated by MKC9989: New Insights from Molecular Docking and Molecular Dynamics Simulations	22
PAPER IV. Effect of Kinase Inhibiting RNase Attenuator (KIRA) Compounds on the Formation of Face-to-Face Dimers of Inositol-Requiring Enzyme 1: Insights from Computational Modeling.....	24
PAPER V. Molecular modeling provides structural basis for PERK inhibitor selectivity towards RIPK1.....	26
Paper VI. Peptidomimetic-based identification of FDA approved compounds inhibiting IRE1 activity	28
6. Concluding remarks.....	29
7. Author`s Acknowledgements	34
Bibliography.....	35

1. Unfolded Protein Response (UPR)

1.1 General overview

The endoplasmic reticulum (ER) is a fundamental cellular compartment in protein folding¹. The ER is involved in the synthesis of one third of the entire proteome². A cellular condition known as ER stress could alter the functionality of this organelle, leading to accumulation of unfolded or misfolded proteins inside the ER¹.

The unfolded protein response (UPR) is a cellular response related to the endoplasmic reticulum³. It is triggered by the accumulation of proteins in the luminal domain of the ER. In this context, the UPR response has two purposes: restoring normal cell function by readjusting protein synthesis, and increasing the production of molecular chaperones involved in protein folding. If these goals are not achieved within a given time frame, the UPR programs for cell death (apoptosis)⁴.

The UPR is involved in numerous physiological processes, ranging from cellular homeostasis, cellular differentiation, inflammation, lipid and cholesterol metabolism^{5,6}. This wide range of activities suggests its important role in the progression of several diseases (*i.e.* cancer, neurodegenerative disorders and diabetes)⁷. On the basis of these pharmacological observations, several academic laboratories and pharmaceutical companies have made efforts in order to identify UPR modulators⁷. Promising and attractive indications highlight the possibility of modulating ER stress levels using small organic molecules⁸.

In mammals, the major ER stress-sensing molecular machines are three ER transmembrane proteins²: PKR-like ER kinase (PERK), Inositol-requiring enzyme 1 (IRE1), and activating transcription factor 6 (ATF6) (Figure 1).

1.2 IRE1

IRE1 is a type I transmembrane ER-resident protein that contains a N-terminal luminal domain, a transmembrane domain, and a cytoplasmic C-terminal kinase and endoribonuclease (RNase) effector domain¹ (Figure 1). Mammalian IRE1 is present in two isoforms, α and β . IRE1 α (hereafter referred to as IRE1) is ubiquitously expressed whereas IRE1 β is sparsely expressed⁹. IRE1 activation is triggered by the accumulation of unfolded or misfolded proteins within the ER¹.

With an imbalance in ER homeostasis, IRE1 dimerizes, trans autophosphorylates and activates its own endoribonuclease domain on the cytosolic side¹. RNase domain activation and oligomerisation (dimer of dimers) results in X-box-binding protein 1 (XBP1) mRNA splicing, which generates the transcription factor XBP1 with a length of 376 amino-acids².

XBP1s (s stands for the spliced form) translocates to the nucleus promoting the expression of genes that enhance protein degradation and UPR response² (Figure 1).

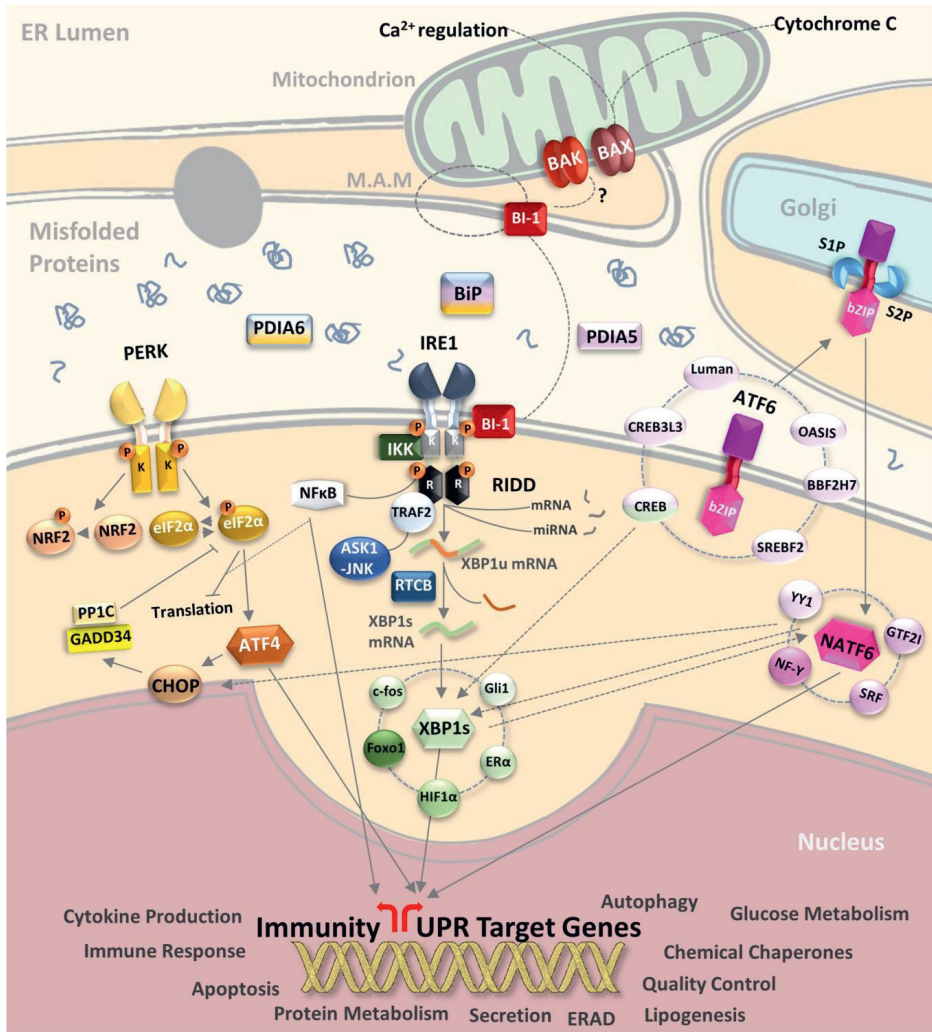


Figure 1. The complexity of the UPR signalling and downstream pathways².

1.3 PERK

Pancreatic ER kinase (PERK) is a type 1 transmembrane protein with a luminal domain and a cytosolic kinase domain¹ (Figure 1). Upon ER stress PERK undergoes oligomerization and transautophosphorylation¹. PERK activation prompts phosphorylation of the eukaryotic translation initiation factor-2 (eIF2 α), that leads to translation reduction and inhibition of mRNA translation³.

However, the short open reading frame of mRNA could translate, encoding for the transcription factor ATF4¹.

C/EBP homologous protein (CHOP), and growth arrest and DNA damage-inducible 34 (GADD34), are the two critical target genes induced by ATF4¹. CHOP is a pro-apoptotic

transcription factor, while GADD34, genes encoding the protein phosphatase PP1C, balances PERK activity by dephosphorylating eIF2 α ¹.

Hence, PERK has a dual behaviour, from protective to cell death response, played out at different signalling levels².

1.4 ATF6

ATF6 is a type 2 transmembrane transcription factor constituted by a luminal and a cytosolic domain¹. Under ER stress ATF6 is released, translocated to the Golgi organelle, and cleaved by site-1 and site-2 proteases removing the luminal domain and the transmembrane domain¹⁰. The cytosolic domain of ATF6 (ATF6c) is translocated into the cell nucleus, where it activates the transcription of UPR target genes involved in the transcription of ER chaperones, of folding enzymes, and of transcription factors such as XBP1¹.

2. The Kinome World

Protein kinases such as IRE1 and PERK are enzymes that phosphorylate specific amino acid residues (*i.e.* serine, threonine, and tyrosine) in substrate proteins¹¹. Dysfunctional signalling by overexpressed or mutated protein kinases have been observed in many types of cancers¹². Protein kinases can be divided into two main classes: tyrosine kinases and serine-threonine kinases. Tyrosine kinases phosphorylate the phenolic group of tyrosine residues, while serine-threonine kinases phosphorylate the alcohol group of serine and threonine residues. All the kinases use adenosine triphosphate (ATP) as phosphorylating agent. The crystallographic structures of the protein kinases complexed with adenosine triphosphate (ATP) were studied, and the acquired knowledge was used for the design of selective kinase inhibitors¹³. Kinase inhibitors are classified as type I or type II inhibitors¹⁴. Type I inhibitors act on the active conformation of the enzyme by blocking substrate access¹⁴. Type II inhibitors bind to the enzyme in an inactive conformation, stabilizing this¹⁴.

3. Targeting IRE1 signaling

3.1. Pharmacological modulators of IRE1

As reported in literature, UPR can be modulated by several small organic molecules². Currently, several IRE1 structures co-crystallized with endogenous or exogenous ligands, are available¹⁵. This piece of information will allow structure-based drug design in order to identify new classes of IRE1 modulators¹⁵.

3.2. Ligands that interact with IRE1 RNase domain

Different chemical scaffolds have been classified as IRE1 RNase inhibitors⁸. The salicylaldehyde inhibitor (Figure 2) co-crystallized in the murine IRE1 highlights Lysine 907 as an important residue in establishing a Schiff Base with these series of compounds (PDB code: 4PL3¹⁶). This crystallographic structure has profoundly increased our interest in covalent

drug discovery, and in the understanding of these series of covalently bound hydroxyl aryl aldehyde (HAA) inhibitors.

3.3. Ligands that interact with the IRE1 kinase domain

Two chemical classes are known to inhibit the IRE1 Kinase active site:

-ATP-competitive inhibitors that inhibit the kinase domain and activate the RNase^{17,18},

-ATP-competitive inhibitors that inhibit the kinase and inactivate RNase domain, also known as “kinase inhibiting RNase attenuators” (hereafter referred to as KIRA) (Figure 2)^{19,20}.

The progress made in the field of IRE1 small organic modulators of the kinase domain prompted our scientific interest in the IRE1 kinase domain as well, with several different questions addressed, understood and clarified during the PhD period.

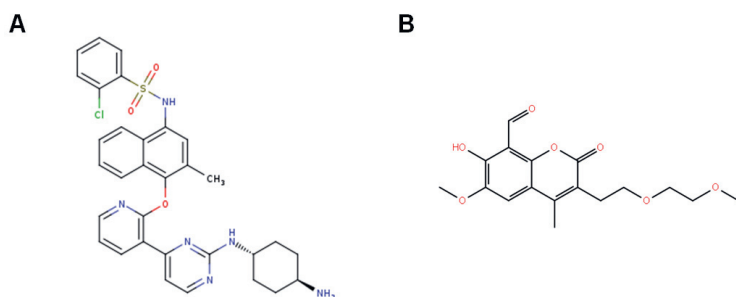


Figure 2. Ligands co-crystallized in the IRE1 cytosolic domain investigated during the PhD period: (A) KIRA in the kinase active site and (B) MKC9989 in the RNase domain.

4. Methodology

4.1 Molecular mechanics

The description of the energy state of a molecular system, as a function of its atomic coordinates, requires the resolution of the Schrödinger equation²¹:

$$\hat{H}\Psi(\mathbf{r}, t) = E\Psi(\mathbf{r}, t) \quad (1)$$

Where \hat{H} represents the Hamiltonian operator of the system, Ψ the wave function, E the energy, \mathbf{r} the position of the vector and t the time, respectively.

Despite the equation having general validity, its practical application is excessively complex to investigate biological molecules²².

In molecular mechanics, the quantum-mechanical effects are therefore ignored²³. The atoms of biological systems are treated, from a physical point of view, like macroscopic bodies described by potential functions.

In computational chemistry, a mathematical function, given the coordinates and the nature of the atoms of a molecular system, is able to provide a numerical value that quantifies the energy of the molecular system of interest. The energy of a system is characterized by two components: potential and kinetic energy (fundamental aspects in both Newtonian deterministic physics and quantum physics). Potential energy defines the ability of an object to carry out work; such contribution differs from the energy acquired by the molecular object during its motion (*i.e.* kinetic energy). Molecular mechanics calculates the molecular system's potential energy U , while classical molecular dynamics²⁴ uses molecular mechanics to study the physical movements of atoms and molecules. Thanks to molecular mechanics and the application of force fields, it is possible to calculate the potential energy of molecules containing several hundred thousand atoms (*i.e.* proteins and protein complexes, membrane models, DNA molecules).

4.1.1 Force Fields

The potential energy U of a molecule is described as the sum of several variables (*i.e.* terms for bonded and non-bonded interactions) and constants, parameterized as a function of the atom-types considered in the calculation.

If we need to apply energy to break a chemical bond, it is according to the law of conservation of energy possible to consider the chemical bond as the main descriptor of the potential energy of a molecule. Through this deduction, defining the fundamental equation of molecular mechanics (*i.e.* force field) is relatively straightforward:

$$U_{TOT} = U_{bond\ stretching} + U_{angle\ bending} + U_{dihedral} + U_{non-bonded\ interactions} \quad (2)$$

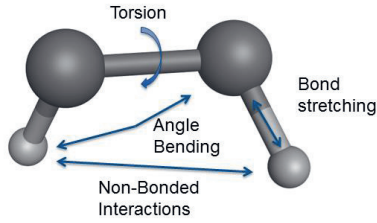


Figure 3. Representation of energy contributions to the force field of a molecule.

In order to calculate the position and motion of a molecular system, it is necessary to develop the functional form of the potential energy U (*i.e.* force fields²⁵⁻²⁷). This functional form depends on the number and type of atoms, and on the forces that each atom exerts on all other atoms (*i.e.* the set of possible atomic interactions^{25,26}).

These interactions can be divided into two main classes: the first, regarding covalent bond interaction between the atoms, and the second describing non-covalent interactions, such as electrostatic²⁴ and Van der Waals interactions²⁴. Usually, these energy terms are added together to derive the total potential energy of the system:

$$U_{TOT} = U_{bonded} + U_{non-bonded} \quad (3)$$

where each term is subdivided in covalent and non-covalent terms, as shown in (4) and (5):

$$U_{bonded} = U_{bond\ stretching} + U_{angle\ bending} + U_{dihedral} \quad (4)$$

$$U_{non-bonded} = U_{Van\ der\ Waals} + U_{electrostatic} \quad (5)$$

The analytical form of a force field is represented below:

$$U_{TOT} = \sum_{bonds} k_{str} (r - r_e)^2 + \sum_{angles} k_{ben} (\tau - \tau_e)^2 + \sum_{dihedral} A [1 + \cos(n\tau - \Theta)] + \sum_i \sum_j -\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \sum_i \sum_j \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad (6)$$

$i = 1 \dots N$ (number of atoms in the systems)

$i, j = 1 \dots N$ (numbers of atoms in the systems)

r , k , and Θ represents the stretching of bonds, angles and dihedral angles with r_e and τ_e the equilibrium value for bond length and angle

k_{str}, k_{ben} constitute the two force constants for deviate the bonding, and angle from the equilibrium one

A is the amplitude of each cosinusoid, n is the periodicity and Θ the phase

A_{ij} and B_{ij} are constants specific to atom i and j

r_{ij} is the distance between atom i and atom j
 q_i, q_j are the partial atomic charge of atom i and j
 ϵ_0 is the vacuum permittivity and ϵ_r the relative permittivity

The first term of the force field expressed in equation (6) is defined by the harmonic potential, which quantifies the energy associated with vibration of the equilibrium bond length of a chemical bond (equation 7). It differs in every molecular system, depending on its minimum, corresponding to the distance in which the diatomic system reaches the minimum of energy.

By summing up this function for all bonds, the contribution from bond stretching to the potential energy is obtained for a given molecule.

Using both experimental and quantum-mechanics computational techniques, it is possible to calculate the value of the force constants (k_{str}), and the equilibrium length of the bond r_e ²⁷.

The greater the force constant, the greater the energy needed to displace the bonding distance from the equilibrium.

$$U_{bond\ stretching} = \sum_{bonds} k_{str} (r - r_e)^2 \quad (7)$$

For molecules with a number of atoms greater than two, the bending energy (equation 8) estimates the energy associated with the bending of the equilibrium bond angle. By varying the angle formed by three atoms, the potential energy of the system is varied. Again, the function is approximated as quadratic.

$$U_{angle\ bending} = \sum_{angles} k_{ben} (\tau - \tau_e)^2 \quad (8)$$

If the molecular system presents torsional energy, it is necessary to define a trigonometric mathematical function, such as an expansion of periodic functions (equation 9), that quantifies the energy associated with each of the dihedral angles. Since the torsion angle is periodic, a typical form used in force fields is:

$$U_{dihedral} = \sum_{dihedral} A [1 + \cos(n\tau - \theta)] \quad (9)$$

A is the amplitude of each cosinusoid, n is the periodicity and θ the phase

The last component defining the force field is the energy associated with non-bonded interactions (equation 10). Non-bonded interactions can be classified as electrostatic interactions and dipolar interactions (*i.e.* Van der Waals interactions). For the calculation of electrostatic interactions, the Coulomb equation is applied, whereas to date, Lennard Jones's potential is a simple mathematical model that approximates the dipolar contributions.

Starting from a pair of neutral atoms, the Lennard–Jones 12-6 potential describes dipole behaviour in the gaseous state. When the two atoms are in close proximity, the dipole orientation takes place reaching the most stable geometric conformation (*i.e.* the lower one). The functional form of this potential derives from the combination of two terms: the first describes the attractive contribution of the Van der Waals forces, which varies with the sixth power of the distance between the centres of two atoms²⁴; the second describes the contribution of the repulsive forces that are established at short distances between nuclei (<4Å). The repulsive force varies with the twelfth power of the distance between the centres of the two atoms, with a sign opposite to the first term of the Lennard-Jones 12-6 potential. Electrostatic interactions are instead described by the Coulomb potential, with the application of a relative dielectric constant value²⁴ (*i.e.* ϵ_r), depending on the surrounding medium. This will take into account the presence of a solvent in the simulation (*i.e.* the solvent shields the charges by interacting with ions and polar molecules, damping the intensity of the attractive or repulsive force).

$$U_{non-bonded} = \left(\sum_i \sum_j -\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} \right) + \left(\sum_i \sum_j \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \right) \quad (10)$$

$i = 1, \dots, N$
 $i, j = 1, \dots, N$ (numbers of atoms in the systems)

The presence of expressions representing the non-bonding energies places heavy limitations on the computational capacity for molecular systems consisting of many atoms (such as biological macromolecules). In these cases, to make the calculation method more efficient, the interactions between atoms placed at a distance greater than a certain threshold (cut-off) are neglected. Completely neglecting the interactions that occur at a greater cut-off distance is a simplistic approach that can generate artefacts in the simulation.

To overcome this problem, alternative methods have been developed over the years. Among them, two of the most used are the Shifted force method, and the exchange of potential at the cut-off²⁴.

In the case of Shifted force method, the expression describing the interaction is translated (*i.e.* it is null at the cut-off distance). This approach has the disadvantage of generating smaller equilibrium distances between the atoms compared with observed experimental value.

In the potential exchange, the function that describes the potential is modified only in a small interval, through the introduction of a switch function. This function has two important parameters: (a) the cut-off value at which the potential approaches zero and (b) the distance interval in which the switch function influences the potential. The application of a cut-off is partly justified by the observation that the Van der Waals interactions decrease with the sixth power of the distance between the centres of two atoms when considering the Lennard-Jones 12-6 potential. In this case, 8-10 Å cut-off may be sufficient.

A separate discussion would be needed to calculate the electrostatic interactions, which manifest their effects even at long range, decreasing in a linear manner with distance. In these cases, more complex computational approaches are adopted^{28,29}, such as the Ewald summation technique²⁴.

Finally, it is worth noting the term force field has a dual meaning: (a) a library of parameters applied to characterize the energy function of a systems of atoms or (b) the specifics of the potential energy function to describes a molecular system²⁴.

4.2 Energy Minimization

Force field represents an empirical approximation of the potential energy surface of a molecular system. By using a function that describes the potential energy of the system as a function of its atomic coordinates, it would be possible, theoretically, to calculate the potential energy at each point in this hypersurface of $3N-6$ dimensions (where N represents the number of atoms present in the system). For example, the energetic potential associated with Formaldehyde (H_2CO) would be described by a 6-dimensional hyperspace. This hypersurface is obviously not representable. Nevertheless, this representation is extremely useful in the description of several molecular properties. In fact, molecular modelling, including molecular mechanics and molecular dynamics, is often interested in the analysis of equilibrium (minimum) points of the potential energy surface, corresponding to energetically stable molecular conformations. Energy minimization is the process to reach the closest energy minimum, starting from a molecular conformation³⁰. Energy minimization occurs in two steps: (a) identifying a set of atomic coordinates x_k and a function f that describes the potential energy associated with each conformation k and (b) adjusting the conformation k to lower the value of the potential energy. These two conditions are satisfied when:

(a) the first partial derivative of f with respect to each variable x_k is equal to zero, (b) the second partial derivative is positive and (c) the determinant and the trace of the Hessian matrix is positive :

$$(a) \frac{\partial f}{\partial x_k} = 0 \quad (11)$$

$$(b) \frac{\partial^2 f}{\partial x_k^2} > 0 \quad (12)$$

$$H_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j} \quad (13)$$

where x_i and x_j are any of the atomic coordinates of the molecule

Two approaches used in geometry minimization in molecular mechanics are Steepest Descents and Conjugate Gradients³⁰.

4.3 Molecular Docking

Molecular Docking is one of the most important tools in the computational-pharmaceutical field, from which it is possible to obtain valid information for the discovery and optimization of drug candidates.

With the availability of the three-dimensional structure of the molecular target, molecular docking is a method to predict the interactive profile of ligand(s) in a receptor(s) binding site. The prerequisite for the usage and exploitation of this powerful tool is the availability of the three-dimensional structure of the target.

The exponential growth of resolved three-dimensional structures of biological macromolecules³¹ has revolutionized Computer-aided drug design (CADD). Many of the three-dimensional structures available belong to proteins that play a primary role in cellular physiology and offer opportunities in the field of structure based drug design.

If the three-dimensional structure of the receptor is known, *in silico* exploration of the energetically favourable conformations of the ligand within the target could provide insights on the placement, orientation, and conformation of the ligand in the receptor. This computational methodology is called molecular docking³². Molecular docking performs both roto-translation and variation of the internal coordinates of the molecular objects, generating conformations through various conformational analysis methods. By converting these operations into algorithms, docking analysis provide useful information to predict the orientation of ligand(s) to target(s).

From the geometric arrangement of the molecular object into the recognition pocket, the concept of "pose" is defined as the position and conformation adopted in space.

A historical approach, within the molecular docking methodologies, is rigid docking³². Nowadays, few docking programs work with this method because the conformation of the ligand subjected to docking analysis is determined *a priori*. From the conformational analysis of the ligand, the lowest potential energy conformation will be chosen, without being able to determine whether the selected conformer is the one that maximizes complementarity with the target. When computing powers were limited, and conformational analysis algorithms were not so powerful, the first docking programs worked as follows: they selected the more stable conformer, docked into the cavity (where the amino acid residues were fixed), and performed roto-translation operations. A mathematical function, called the scoring function, determines which of the final predictions is the most stable.

The problem with this obsolete methodology comes from the ligand conformation used. It is not possible to establish *a priori* the lower ligand energy conformation as the one that maximizes the minimum potential energy in the ligand-protein binding mode. In the simplest case where the ligand is kept rigid (*i.e.* we do not consider the torsion angles of the rotatable bonds in the ligand), solving a docking-ligand receptor problem is equivalent to identifying an appropriate translation vector and rotation matrix to be applied to the initial coordinates of the ligand. To understand this concept, we consider the elementary case of an atom projected into a two-dimensional surface (*i.e.* a point in a two-dimensional space). The position of the point is described by a vector of dimensions (x, y), which starts from the

origin of the Cartesian axes. To translate the point into translated coordinates (x' , y') it will be necessary to apply to the first vector, through vector addition, a second vector of components (a, b):

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x + a \\ y + b \end{pmatrix} \quad (14)$$

In three dimensions, the translation vector will consist of three values. Also in this case, therefore, the rotation is given by specifying three values, corresponding to the rotation with respect to the X, Y and Z axis. In two dimensions, to rotate an atom, it will be necessary to make the product of the vector for the rotation matrix:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (15)$$

For example, in the case of an atom of coordinates (1, 0) and a rotation of 90° , we will have:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} (1 * 0) + (-1 * 0) \\ (1 * 1) + (0 * 0) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (16)$$

In summary, if the ligand is considered rigid, transforming the initial coordinates equals to defining six values: three for translation (translation along the X, Y and Z axis) and three for rotation (rotation along the X, Y and Z axis).

These six variables constitute the degrees of freedom (*i.e.* the number of independent variables that define the position and orientation of the ligand).

The ligand can also be considered (completely or partially) flexible: in this case, the number of degrees of freedom increases with the number of torsional angles. When the ligand is considered flexible, the exponential increase in the number of possible ligand conformations makes the problem rapidly intractable. To circumvent the exponential increase, any docking approach must implement an efficient algorithm. Based on the algorithm implemented, the approaches are divided into systematic and stochastic.

In systematic conformational research, molecular docking explores the largest number of conformations that can be adopted by the ligand in the target-binding site. The advantage of this approach is the exhaustiveness of the investigation (*i.e.* all or almost all the possible solutions are considered) and the reproducibility (*i.e.* we always obtain the same results by repeating the molecular docking with the same settings parameters). In order to deal with very flexible ligands, docking programs that make use of a systematic approach (among the most important, we can mention Glide³³), use *divide et impera* philosophy: the ligand is initially divided into fragments; among the generated fragments, one is considered to be the core fragment; conformational search within the target-binding site is exhaustively performed for the core fragment. At the end of this phase, the flexible fragments are

progressively added to the core fragment, one at the time, and the rotatable torsional angles are systematically explored.

On the other hand, the algorithms used in stochastic docking apply random geometric transformations on the ligand conformations.

In genetic algorithms (for example used in GOLD³⁴ docking program), the conformational analysis are performed based on the principles of biological evolution.

In addition to genetic algorithms, other stochastic approaches have been adopted in docking. PRO_LEADS³⁵, for example, adopts a method for ligand conformational space analysis called tabu search. At each iteration the conformations discarded (*i.e.* tabu conformations) become part of a tabu list. In each new iteration, ligand conformations are compared with the tabu conformations, and discarded if too similar. In this way, the algorithm avoids re-exploring the same conformation region.

The methodologies described, both systematic and stochastic, can be accompanied by one or more steps of energy minimization of the final conformations.

4.3.1 Scoring functions

An ideal scoring function should be able to accurately determine the binding-free energy given by the Gibbs-Helmholtz equation^{36,37}.

One might think that even docking studies, like any experimental procedure, could give a binding value constant. This is absolutely false. In fact, the final state of a system cannot be compared with the initial one.

When performing binding experiments, or determining the association constant, the obtained numerical value is an average value, not an absolute one. Each experimental measure is the result of an Avogadro number of events (*i.e.* macroscopic system consisting of 10^{23} particles or more²⁴). Once clarified this aspect, it is trivial to understand how a docking result cannot be a thermodynamic measure: the ligand-target complex describes a single event. Hence, results obtained from a molecular docking experiment are not the binding constants of the ligand-receptor complex, but rather the differences in stability between different compounds/conformers.

Molecular docking takes into account the non-covalent interactions that play a crucial role between the receptor and the ligand, in particular the electrostatic interactions and Van der Waals interactions. Moreover, to take into account the entropic and desolvation effects, empirical terms could be introduced in the force field. These effects are relevant in the receptor-ligand interaction, but they are usually neglected in classical force fields.

In addition to the molecular mechanics force field, empirical energy functions^{38,39} are often used, based on the analysis of Gibbs free energy of interaction ($\Delta G_{\text{binding}}$) which is experimentally measured in numerous receptor-ligand complexes.

These functions approximate the $\Delta G_{\text{binding}}$ value as a sum of uncorrelated energy terms:

$$\begin{aligned}\Delta G &= \Delta G_0 + \Delta G_{idr} + \Delta G_{met} + \Delta G_{hyd} + \Delta G_{rot} \quad (17) \\ \Delta G_0 &= k_0 \\ \Delta G_{idr} &= k_1 X_{idr} \\ \Delta G_{met} &= k_2 X_{met} \\ \Delta G_{hyd} &= k_3 X_{hyd} \\ \Delta G_{rot} &= k_4 X_{rot}\end{aligned}$$

where k_0 , k_1 , k_2 , k_3 and k_4 are coefficients derived from a multiple linear regression analysis on a training set of protein–ligand complexes and X_{idr} , X_{met} , X_{hyd} and X_{rot} are scores for hydrogen bonding, acceptor-metal, lipophilic interactions and loss of conformational entropy of the ligand upon binding to the protein.

In equation 17, the ChemScore energy function^{40,41} implemented in some docking programs (including Gold) is shown. The construction of these functions involves a linear regression analysis of the $\Delta G_{\text{binding}}$ values as a function of a large number of ligands related to their structure (X_{idr} , X_{met} , X_{hyd} , etc etc) by using the linear relationship between two variables and derive the coefficients k_0 , k_1 ... k_n that appear in the various terms of the energy function. Despite the progress made, the binding affinity predictions are far from being perfect. The reasons of the discrepancy with the experimental values can be several. Firstly, the scoring functions present many simplifications, necessary to increase the efficiency of the computational calculation. In particular, entropic effects estimations are one of the main limiting factors. A second molecular docking limitation is the solvent treatment: the solvation/desolvation effects of the receptor-ligand complex are often neglected or considerably simplified in the scoring function; moreover, water molecules often mediate the interaction between receptor and ligand. Finally, the representation of the receptor in molecular docking could be unrealistic. Even without considering the errors present in a structure^{42–46}, a 3D structure obtained by X-ray crystallography represents an average in space and time of a set of conformational states of the receptor. The most populated among these states, may not correspond with the most populated state in the presence of a ligand. If this happens, considering the rigid receptor during docking constitutes an excessive simplification of the system.

In addition, the complexity of computational calculations required for the evaluation of the interaction energies of the ligand atoms with each single atom of the receptor, imposes a limitation on the dimensions of the space explored from the ligand. In many docking programs (like Glide), the energy contribution of the receptor in the interaction with the ligand is approximated by means of a grid map³³.

In conclusion, although molecular docking has not yet reached full maturity, the increasing number of promising ligands developed using docking methods (a datum confirmed in my doctoral work as well), fully justifies its use and the efforts being made in refining it.

4.4 Classical Molecular Dynamics (MD)

Molecular Dynamics (MD) is a computational approach useful in investigating the dynamic properties of molecules³². MD simulates the evolution of a system in time, based on the force field used. MD uses equations 18 and 19, which can be reformulated as equation 20:

$$\mathbf{F}_i(t) = m_i \mathbf{a}_i(t) \quad (18)$$

$$-\frac{\partial U}{\partial \mathbf{r}_i} = m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} \quad (19)$$

$$-\frac{\partial U}{\partial \mathbf{r}_i} = m_i \frac{\partial \mathbf{v}_i}{\partial t} \quad (20)$$

where U is the potential energy with respect to the coordinates of the atom i at time t (described by the vector \mathbf{r}_i), m is the mass of the atom i , and \mathbf{v}_i and \mathbf{a}_i represents the velocity and acceleration of the atom i at time t . This equation is deterministic: once the initial coordinates and velocities of the system are known, it is possible to study the evolution of coordinates and velocities in time. The initial coordinates of the system, especially for biological macromolecules, are represented by experimentally resolved three-dimensional structures (X-ray crystallography or NMR), usually deposited in the Protein Data Bank database, or obtained through molecular modelling techniques (*i.e.* homology modelling⁴⁷). In both cases, to ensure appropriate geometry in the initial configurations, energy minimization is performed. The initial velocities of the atoms are randomly selected from the Maxwell-Boltzmann distribution, which describes the distribution of velocities as a function of temperature:

$$p(v_{ix}) = \sqrt{\left(\frac{m_i}{2\pi k_b T}\right)} \exp\left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_b T}\right] \quad (21)$$

where p represents the probability distribution function of the velocity along the x axis for the particle i , k_b the Boltzmann constant, m_i the mass of the particle i and T the temperature of the system. Once the initial coordinates and velocities of the system are known, and a force field has been defined, it is possible to solve equation (20) through numerical methods because the equations of motion in molecular dynamics are too complex to be solved analytically.

The analytical solution of the motion equation is possible in very simple cases (*i.e.* the harmonic oscillator). Consider for example a diatomic molecule: based on molecular mechanics, the potential energy of the system can be approximated to a single term (*i.e.* the stretching of the bond):

$$U = \frac{1}{2}K_{ij}(r_{ij} - r_{0ij})^2 \quad (22)$$

where r_{0ij} represents the equilibrium position. By placing this value at the origin of the coordinates, and by considering the one-dimensional case, it is possible to simplify (22) as follows:

$$U = \frac{1}{2}K_{ij}x^2 \quad (23)$$

The gradient (first derivative) of (23) can be solved analytically, obtaining the force acting on the system (anti-gradient):

$$-\frac{\partial U}{\partial x} = -k_{ij}x \quad (24)$$

equation (24) describes the force acting on the bond stretching in the same manner as an elastic body, for example a spring (Hooke's law). Starting from an initial velocity and an initial time t_0 , by replacing the (24) in (20) and integrating the resulting equation, we get:

$$\int_{v_0}^{v_i} dv = \int_{t_0}^{t_i} -\frac{kx}{m} dt \quad (25)$$

The solution of (25) gives the value of the velocity v_i when v_0 is known:

$$v_i = v_0 - \frac{kx}{m}(t_i - t_0) \quad (26)$$

In the same way, the x_i position of the system can be obtained from the initial position x_0 by integrating the equation with the velocity v_i :

$$x_i = x_0 + v_i(t_i - t_0) \quad (27)$$

As previously mentioned, the systems investigated by molecular dynamics are far too complex to be solved in an analytical form. By calculating the evolution of positions and velocities on small discrete intervals of Δt time (timestep) we replace the derivatives with the respective incremental ratios (*i.e.* the limit for Δt tending to zero is the derivative itself):

$$\lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt} \quad (28)$$

The substitution of the derivatives with the incremental ratios allows to divide the whole MD process in small intervals Δt , and to reiterate the integration of equations of the motion. If

the positions and the initial velocities of the system (time $t = 0$), and the force acting on each atom (given by the force field) are known, the acceleration of each atom can be obtained, and then (through the algorithms illustrated below) the positions and the velocities in the instant $t + \Delta t$ can be derived, by assuming the force as constant during Δt .

There are several algorithms for integrating the equations of motion using finite difference methods. One of the most used in molecular dynamics is called the Störmer-Verlet method²⁴. After dividing the time interval into smaller Δt , Taylor series is used to expand $x(t)$ or, more generally, $r(t)$ (*i.e.* the vector based on the Cartesian coordinates of the system) in the interval $t+\Delta t$ and $t-\Delta t$:

$$x(t - \Delta t) = x(t) - x'(t)\Delta t + \frac{1}{2}x''(t)(\Delta t)^2 - \frac{1}{6}x'''(t)(\Delta t)^3 \quad (29)$$

$$x(t + \Delta t) = x(t) + x'(t)\Delta t + \frac{1}{2}x''(t)(\Delta t)^2 + \frac{1}{6}x'''(t)(\Delta t)^3 \quad (30)$$

For simplicity, the Taylor series expansion is truncated at the third derivative, excluding position uncertainty of higher order derivatives.

Next, by adding (29) to (30) we get:

$$x(t + \Delta t) = -x(t - \Delta t) + 2x(t) + x''(t)(\Delta t)^2 \quad (31)$$

and by using Newton's second law we get :

$$x(t + \Delta t) = -x(t - \Delta t) + 2x(t) - \frac{1}{m} \frac{\partial U}{\partial x} (\Delta t)^2 \quad (32)$$

Therefore, with this equation, we obtain the position of the system at time $t + \Delta t$ once the system positions at the time $t-\Delta t$ and the gradient of the field is known.

A more correct and more used method is a variation of the Verlet algorithm⁴⁷, called leap-frog⁴⁷. This method calculates the speed at $\Delta t / 2$ as described below:

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) - \frac{1}{m} \frac{\partial U}{\partial x} \Delta t \quad (33)$$

Starting from this value, the position $t + \Delta t$ is calculated using the relation:

$$x(t + \Delta t) = x(t) + v\left(t + \frac{\Delta t}{2}\right) \Delta t \quad (34)$$

One of the basic assumptions of the Störmer-Verlet algorithm, as well as of alternative integration algorithms, is that the velocities and accelerations of each atom in the considered system remain constant in the Δt time interval (timestep) considered. Otherwise, the integration process is inaccurate, resulting in system artefacts (*i.e.* instability during simulation). Since molecular vibrations have times scale between 10^{-13} - 10^{-14} s, an efficient

simulation requires that each movement is sampled at least 10 times, which implies values of Δt between 0.1 and 1 fs (femtoseconds, 10^{-15} s). To enable larger integrating time steps constraint on the fastest modes of bonds vibrations are normally applied²⁴. The duration of a molecular dynamics simulation is therefore strongly limited from the high number of frames. Finally, in molecular dynamics the ensemble needs to be chosen. The simplest ensemble is the NVE, for which number of atoms, volume, and energy are kept constant²⁴. An ensemble that is usually more coherent with chemical processes is the NPT, in which number of atoms, pressure, and temperature are kept constant²⁴.

4.5. Experimental Methods used for IRE1 drug design

Although my PhD work follows a predominantly computer-driven drug discovery path, focusing primarily on developing inhibitors towards IRE1 and PERK, respectively, our computational efforts have been extended by assay development and testing of the identified compounds, and by further compound optimization.

These subsequent phases of the work were carried out within Afshin Samali's group (NUI Galway), Eric Chevet's team (INSERM, France), 2bind (Regensburg), and Reaction Biology Corporation.

4.5.1. Micro-scale thermophoresis (MST)

Experimental methods have been developed to aid studies on protein-ligand interactions. Each sort of approach has its own strengths and limitations in terms of sensitivity and specificity.

Microscale thermophoresis (MST) is a technique useful for quantitative analysis of protein-ligand/protein-protein interactions in free solutions: it has been applied extensively in academia and in industry⁴⁸. This technique detects the motion of molecules (*i.e.* relative distribution of molecules) using an infrared (IR) laser inside a capillary, in a microscopic temperature gradient, given by the following formula:

$$\frac{C_{hot}}{C_{cold}} = \exp(-S_T \Delta T) \quad (35)$$

C_{hot} : molecule concentration in the hot area of the capillary; C_{cold} : molecule concentration in the cold area of the capillary; S_T : Soret coefficient.

This relative distribution of molecules in a microscopic temperature gradient is a phenomenon called thermophoresis, which can be analysed by fluorescence using the intrinsic fluorescence of tryptophans in proteins, or by applying an extrinsic fluorophore coupled to one of the molecular partners. This effect is highly sensitive to binding-induced changes in various molecular properties, such as size, charge, and conformation or hydration

shell, hence offering a powerful technique for quantitative measurement of interactions between protein-ligands.

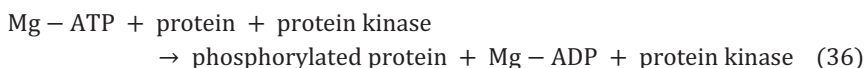
4.5.2. Fluorescence resonance energy transfer (FRET)

FRET is a phenomenon of energy transfer between fluorophores. The mechanism exploits the presence of two fluorescent molecules, called donor and acceptor. The donor can be excited at a specific wavelength. This molecule transfers energy to the acceptor, which can consequently emit fluorescence. This process occurs only if the two molecules are at a reasonable distance (*i.e.* 1–10 nm).

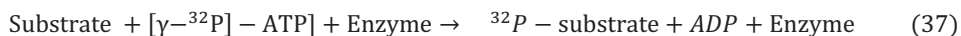
The fluorescence signal generated could therefore be useful in investigating changes in molecular complex associations and/or dissociations²⁰.

4.5.3. Protein kinases assay using radiolabeled ATP

Protein kinases catalyse the phosphorylation of serine, threonine, tyrosine residues on their substrate proteins, according to the following equation:



Based on this molecular evidence, protein kinase activity could be evaluated using radiolabeled [γ -³²P] ATP with an appropriate protein substrate⁴⁹:



To capture the resulting phosphorylated proteins phosphocellulose paper immersed in phosphoric acid is used. This will remove the radiolabeled [γ -³²P] ATP and leave radiolabeled product bound to the paper, respectively. This method is the 'gold standard' to investigate the phosphorylation level of all protein being positively charged at a pH of 1.8⁴⁹.

5. RESULTS

In this chapter, the main results and conclusions achieved in Paper I-VI are presented.

PAPER I. Binding analysis of the Inositol-requiring enzyme 1 (IRE1) kinase domain

IRE1 is a key protein in UPR signalling, and its kinase domain has been targeted with small organic molecules. The availability of IRE1 entries in the Protein Data Bank (PDB), co-crystallized with small organic molecules in the kinase active site (exogenous or endogenous compounds) have triggered the usage of Structure-based drug design (SBDD) approach for the identification of novel, potent and selective IRE1 ligands.

In this study we investigated the plethora of IRE1 structures available, and identified the most suitable one for a Virtual Screening campaign in the IRE1 kinase domain, as an important piece of information for the development of novel KIRA-like compounds capable of allosterically inhibit RNase activity.

Firstly, we selected the docking protocols able to reproduce the ligand-protein complexes in the self-docking procedure and, secondly, we ranked the series of co-crystallized ligands towards the full pool of IRE1 crystal structures by using MASC score⁵⁰ (*i.e.* the Normalized values of docking score). Thirdly, we tested the IRE1 structures by fitting in them the available KIRA allosteric inactivators, 25 KIRA compounds⁵¹ for which IRE1 kinase and RNase activity were available⁵¹. By using molecular docking, coupled with MASC score evaluation (Figure 4), we were able to identify the holo IRE1 conformation (PDB code: 4U6R) as highly suitable for docking-based VS, and hence also for identifying novel KIRA-like compounds.

This study has facilitated the process of selection of novel potential IRE1 ligands from the application of Virtual Screening. This work resulted in one patent on new inhibitors targeting IRE1, currently about to be filed.

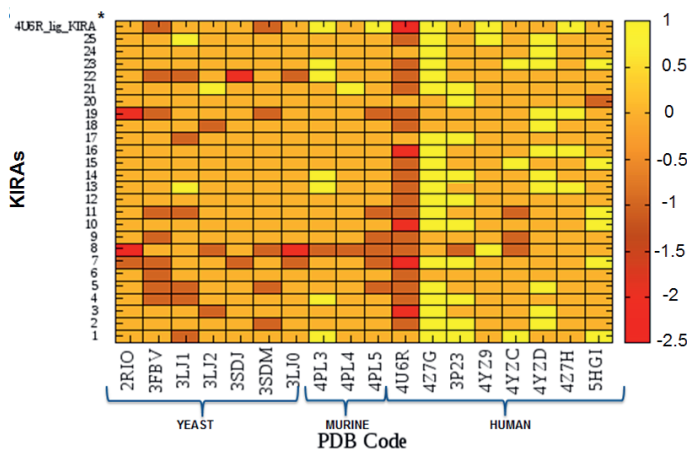


Figure 4. MASC score values, represented by a colorimetric scale, going from red (-2.5) to yellow (1), of KIRA analogues (1-25) for each IRE1 PDB structure. *KIRA co-crystallized in PDB structure 4U6R. Reproduced with permission of copyright owner¹⁵. Further permissions related to the material excerpted should be directed to the ACS (<https://pubs.acs.org/doi/10.1021/acsomega.8b01404>).

PAPER II. Merits and pitfalls of conventional and covalent docking in identifying new hydroxyl aryl aldehyde like compounds as human IRE1 inhibitors

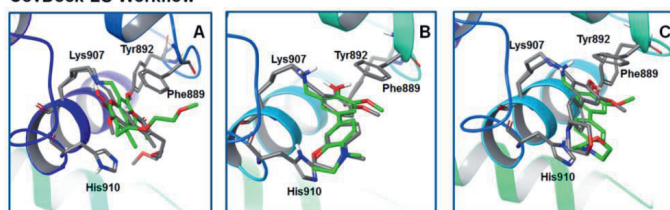
Crystal structures of murine IRE1 covalently bound with hydroxyl aryl aldehyde (HAA) inhibitors in the RNase domain prompted our interest in the development of improved HAA inhibitors. In order to understand the suitability of the RNase pocket as a candidate target site for virtual screening, a conventional molecular docking study was first performed.

The poor docking scores obtained for the HAA pre-reactive species in the RNase active site during the conventional non-covalent docking procedures highlights relevant issues for the docking and virtual screening of novel HAA inhibitors. These results should dissuade any users from choosing this pocket site as a suitable target for conventional virtual screening studies.

To select the right docking algorithm, we investigated the performance of Covalent docking analysis as a possible tool for the development of improved HAA inhibitors. Reproducing the covalently bound conformations of the co-crystallized ligands (Figure 5) and correctly ranking the experimental binding data of the three co-crystallized structures by using the docking score generated by the CovDock-LO and CovDock-VS modules^{52,53} in Schrödinger, unexpectedly turned out to be a challenging task.

This study oriented our focus on understanding important information regarding HAA reaction mechanisms, activation energies, and covalent binding energies for the rational design of novel HAA covalent inhibitors. This further research is currently ongoing.

CovDock-LO Workflow



CovDock-VS Workflow

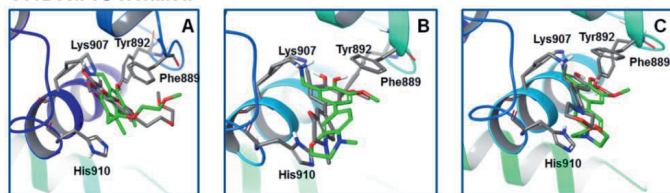


Figure 5. Superposition of the best-scoring docked poses from covalent docking (grey) onto the native crystal one (green). Lys907 and key residues in the RNase binding site are highlighted : (A) MKC9989 (PDB code: 4PL3) (B) OICR464 (PDB code: 4PL4) (C) OICR573 (PDB code: 4PL5). Reproduced with permission (<http://creativecommons.org/licenses/by/4.0/>).

PAPER III. Selective Inhibition of IRE1 Signalling mediated by MKC9989: New Insights from Molecular Docking and Molecular Dynamics Simulations

The characterization of the binding properties of MKC9989 to IRE1 is an important component of the IRE1 drug discovery program. To understand its selectivity towards two critical lysines, Lys907 and Lys599, located in the RNase and kinase domains of IRE1, respectively (Figure 6), we performed Molecular Docking and MD simulations, providing a computational approach to characterize MKC9989 selectivity towards Lys907 that triggers IRE1 inhibition. Although murine IRE1 crystal structures provided snapshots of the covalent MKC9989 binding mode with lysine, the structure of the non-covalent MKC9989 binding mode remains currently undetermined. By capturing reliable¹⁵ non-covalent MKC9989 binding poses within two representatives buried (*i.e.* the reactive Lys907⁵⁴ and the *in vitro* reactive Lys599⁵⁴ located in the RNase and kinase domain, respectively) and two solvent-exposed lysines in the IRE1 cytosolic domain (Figure 6), we performed 500 ns MD simulations in triplicate to observe the different recognition processes towards each representative lysine.

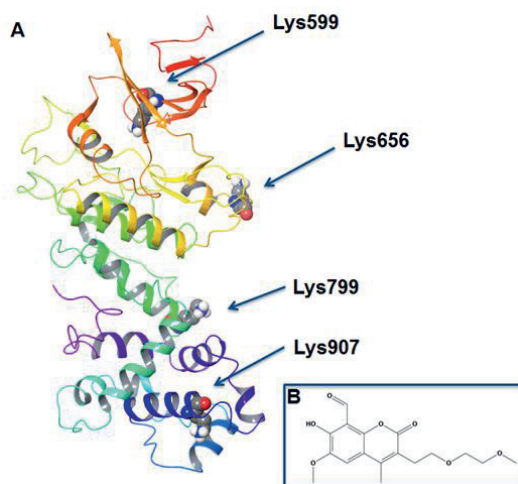


Figure 6. (A) Structure of IRE1 (PDB ID 4PL3) showing the four targeted lysine residues in space-filling model. (B) 2D chemical representation of MKC9989. Reproduced with permission of copyright owner⁵⁵.

In accordance with experimental findings, analysis of the recognition mode of MKC9989 illustrates that only the two *in vitro* reactive lysines⁵⁴ defining the kinase and RNase pockets are accompanied by stabilization of the ligand within the binding sites (Figure 7). Conversely, MKC9989 placed near Lys656 and Lys799, as two representative solvent exposed lysines, is unable to reach a sufficient energetic stabilization (Figure 7), resulting in rapid unbinding and delocalization from the putative binding sites.

This work sheds light on the understanding of the molecular mechanism of IRE1-MKC9989 recognition process, with important insights for the development of novel and possibly improved MKC9989-like compounds.

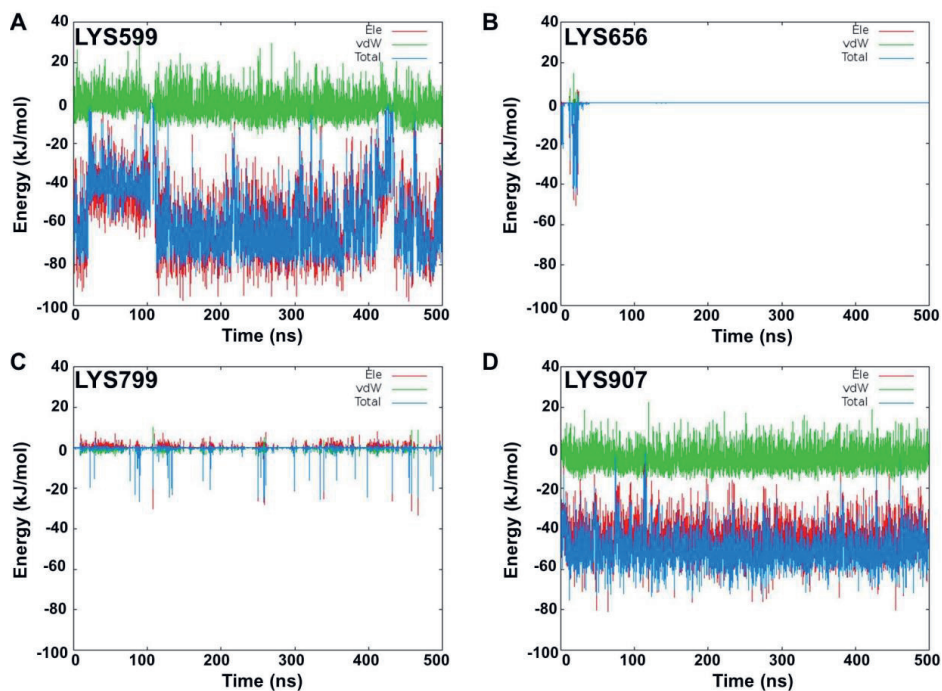


Figure 7. MKC9989 energetic stabilization towards (A) Lys599, (B) Lys656, (C) Lys799 and (D) Lys907, during the 500 ns classical MD simulation (Replica 1). Reproduced with permission of copyright owner⁵⁵.

PAPER IV. Effect of Kinase Inhibiting RNase Attenuator (KIRA) Compounds on the Formation of Face-to-Face Dimers of Inositol-Requiring Enzyme 1: Insights from Computational Modeling

IRE1 is a critical transmembrane protein for the UPR activation, and the development of IRE1 inhibitors is a strategy for the treatment of various human malignancies. A detailed and complete mechanistic understanding of IRE1 inhibition through a subset of IRE1 kinase inhibitors (known as “KIRA” compounds) that bind to the ATP-binding site and allosterically impede the RNase activity, has, however, been lacking.

Here, inspired by a previous model suggesting that KIRAs should stabilize both the DFG-out kinase domain conformation and the helix- α C displacement that make IRE1 incompatible with back-to-back dimer (Figure 8) formation, thereby leading to kinase and RNase inhibition²⁰, we performed computational analysis, protein-protein and protein-ligand docking studies, and molecular dynamics simulations, and observed that KIRA inhibitors act at an early stage of IRE1 activation by interfering with IRE1 face-to-face dimer formation.

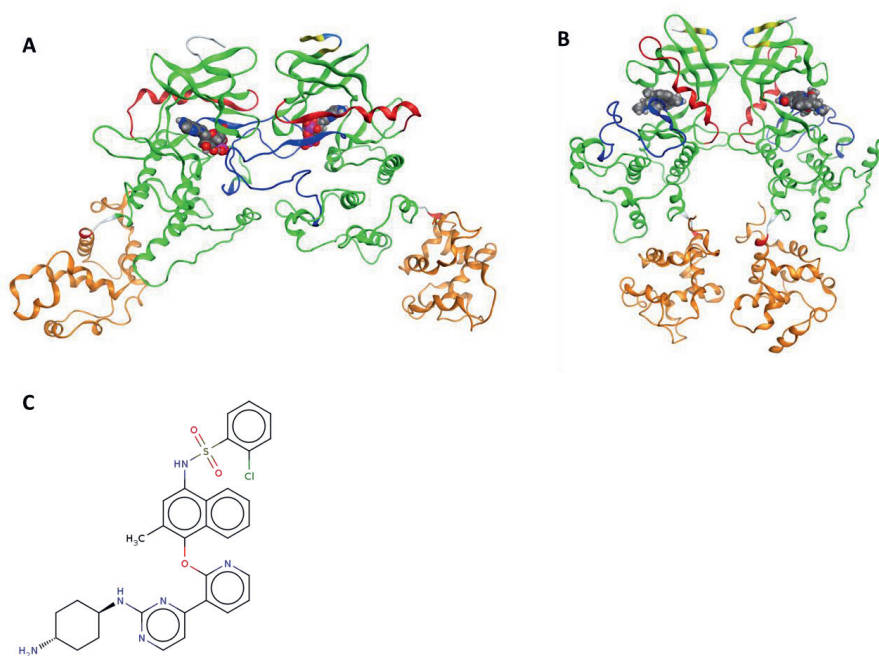


Figure 8. Dimeric structures of the IRE1 cytosolic regions. (A) Face-to-face dimer (PDB 3P23). (B) Back-to-back arrangement (PDB 4YZC). The kinase domain is shown in green (residues 571-832), the helix- α C in red (residues 603-623), the activation segment in blue (residues: 711-741) and the RNase domain in orange (residues 837-963). ADP (A) and staurosporine (B) are highlighted in space-filling models. (C) 2D molecular representation of KIRA. Reproduced with permission of copyright owner⁵⁶.

Using molecular dynamics simulations of the native crystal dimer structures (*i.e.* face-to-face and back-to-back), protein-protein docked pose of PDB 4U6R (*i.e.* KIRA-bound monomer) and the KIRA-docked dimers, we found that KIRA reduce energetic stabilization of face-to-face dimer (Figure 9), suggesting that the binding of KIRA is affecting the system already at an early stage of UPR activation.

This interesting result related with KIRA mode of action and its role in early stages of IRE1 activation is currently being exploited by our research group as novel information to design IRE1 allosteric modulators that could destabilize the formation of the face-to-face dimer.

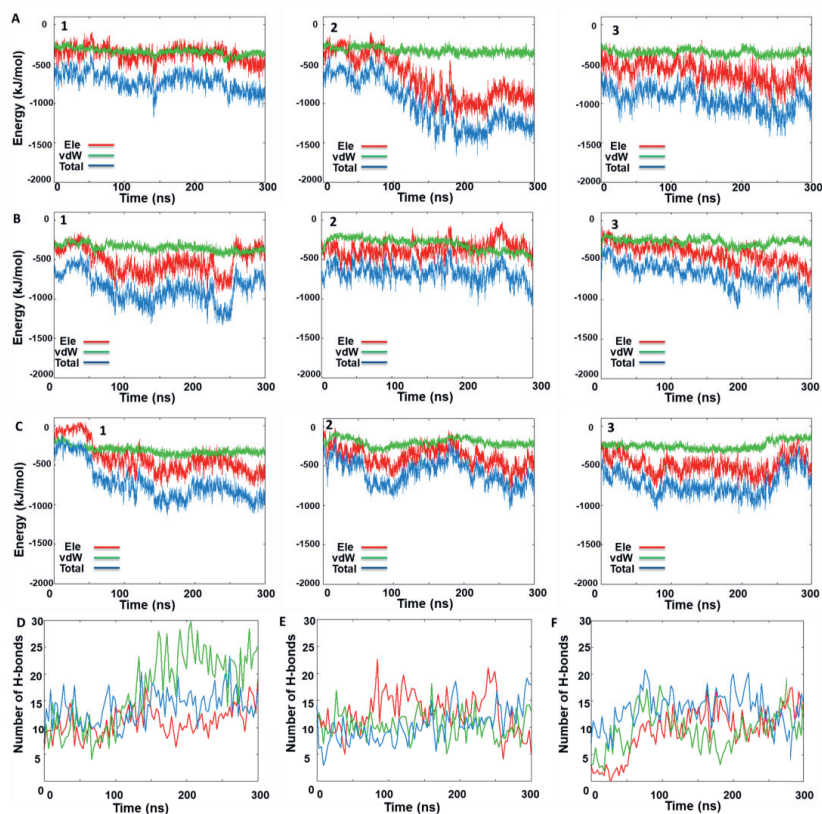


Figure 9. IRE1 face-to-face dimer MD simulations. Time-dependent interaction energy profiles for monomer A with monomer B during the three MD simulation replicas of (A) Native face-to-face crystal dimer (PDB code: 3P23), (B) KIRA-docked face-to-face dimer (PDB code: 3P23), (C) protein-protein docked pose of PDB 4U6R in face-to-face dimer form. Hydrogen bond analysis between monomers A and B during three MD replicas for (D) Native face-to-face crystal dimer (PDB code: 3P23), (E) KIRA-docked face-to-face dimer (PDB code: 3P23), (F) protein-protein docked pose of PDB 4U6R in face-to-face dimer form. Reproduced with permission of copyright owner⁵⁶.

PAPER V. Molecular modeling provides structural basis for PERK inhibitor selectivity towards RIPK1

GSK2606414, GSK2656157 and AMG44 are inhibitors of the PERK tyrosine kinase domain. While GSK2606414 and GSK2656157 target RIPK1⁵⁷ as well, a protein with high structure and sequence similarities with the PERK kinase binding site (Figure 10), AMG44 binds specifically to PERK.

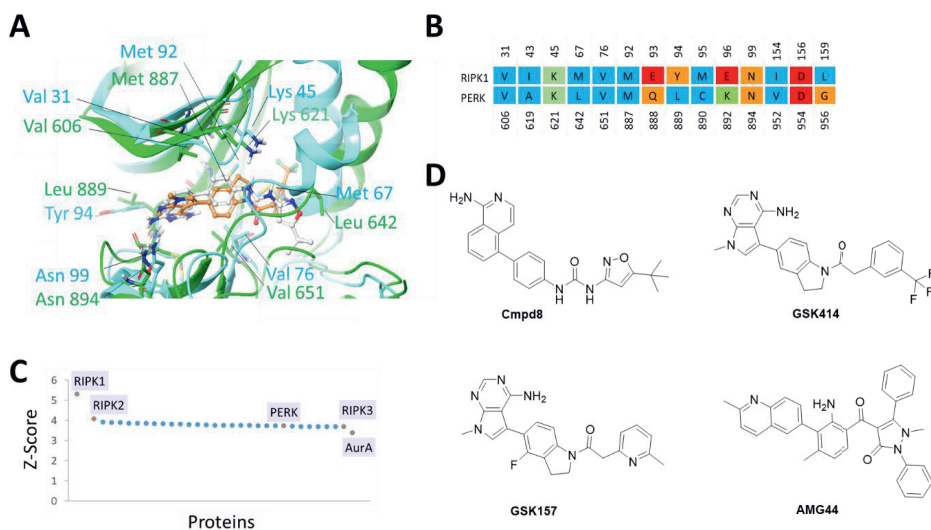


Figure 10. (A) Superposition of binding sites of RIPK1 (PDB 4NEU) and PERK (PDB 4G31) shown in cyan and green, respectively. (B). Structure-based sequence alignment of the active sites, the sequence numbering (above/below) is based on the PDB codes 4NEU and 4G31 for RIPK1 and PERK, respectively. (C) Binding sites similar to RIPK1 was calculated using ProBiS server⁵⁸. PDB code: 4NEU (Chain: A) was used as a query. Selected kinase hits are labelled. (D) 2D chemical structures of the small molecule inhibitors of RIPK1 (Cmpd8) and PERK (GSK414, GSK157 and AMG44).

Here we investigate the selectivity profiles of the compounds, by combining molecular docking followed by molecular dynamics simulations.

The origin of the high selectivity of AMG44 is suggested by the fact that AMG44 had a strikingly different predicted binding profile compared to its cognate ligand (*i.e.* Cmpd8) in the RIPK1 binding site with both rigid docking and induced fit docking settings, while GSK2606414 and GSK2656157 revealed a common binding mode in the RIPK1 binding site (Figure 11).

This novel piece of information had facilitated the process of selection of novel potential PERK ligands from the application of Virtual Screening.

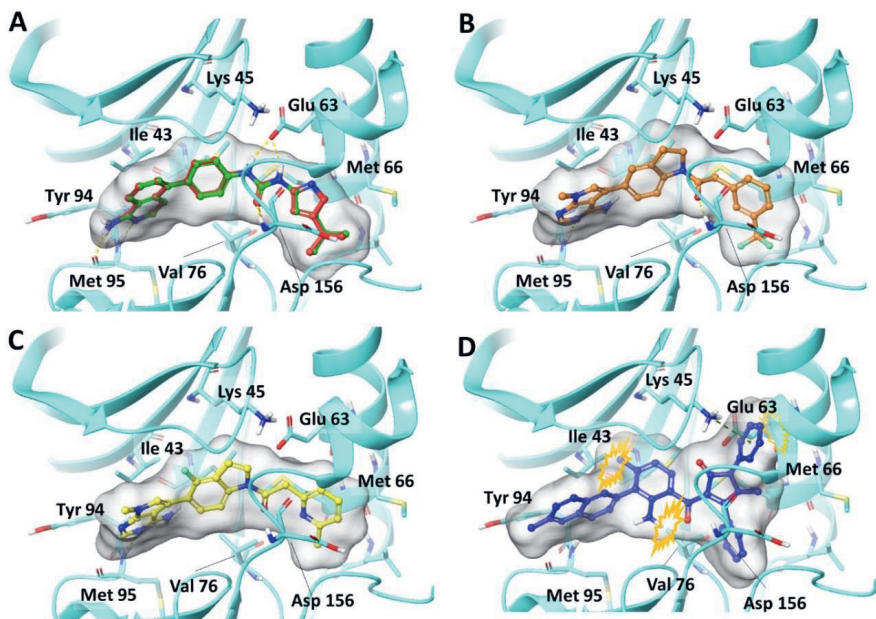


Figure 11. Glide Docking. Predicted binding mode of PERK inhibitors GSK414, GSK517 and AMG44 in RIPK1 active site. Close-up view of the RIPK1 active site (PDB code: 4NEU). (A) Superposition of docked pose (green) on the co-crystal structure of Cmpd8 (white). The docked pose of (B) GSK414 (orange), (C) GSK157 (yellow). (D) AMG44 did not dock in the RIPK1. Superposition of the AMG44 PERK bound pose into the RIPK1 active site (blue) shows steric clashes (orange).

Paper VI. Peptidomimetic-based identification of FDA approved compounds inhibiting IRE1 activity

In this study we applied both structure and peptide-based design approaches⁵⁹ to discover FDA approved compounds methotrexate, cefoperazone, folinic acid and fludarabine phosphate as novel IRE1 inhibitors (Figure 12). Our *in-silico* hypothesis has been validated by using a series of *in vitro* and in cell-based models (Figure 12). Lastly, MD simulations have been performed to investigate the possibility of the current FDA approved drugs to interact or not with the same sites of MKC9989.



Figure 12. Schematic representation of the approaches used for the identification of novel and FDA approved drugs as IRE1 inhibitors.

6. Concluding remarks

We believe that this thesis offers important insights relevant to hit-discovery and lead optimization of novel IRE1 modulators. As the amount of information regarding IRE1 signalling and modulation increases, the preliminary results presented here could be updated and used to develop novel approaches for IRE1 structure-based drug design.

Our analysis emphasizes the importance of structural information and the role played by crystallographers in this IRE1 drug discovery project. Using a combination of protein-ligand docking, protein-protein docking and MD simulations, we provide additional evidence that small organic molecules can allosterically and orthosterically modulate IRE1 activity.

While the basic IRE1 pathway has been thoroughly investigated^{2,60}, there is a major gap in our understanding of the molecular recognition process between IRE1 and currently available small organic molecules modulators of this pathway^{8,61}. This lack of knowledge hinders access to important information for speeding up the IRE1 drug discovery process. Further investigations are needed for a comprehensive understanding of important information regarding the IRE1-small organic molecules binding process, providing useful insights for IRE1 molecular pharmacology and drug discovery. Specifically, further studies will be required for the characterization of possible multiple binding events⁶² and the discovery of new potential meta-binding and allosteric sites to modulate IRE1 activity and the prediction of the associated rate constants⁶³ (*i.e.* unbinding rate k_{off} and binding rate k_{on}) of IRE1–ligands, a crucial piece of information in the IRE1 drug discovery process.

7. Author`s Acknowledgements

I would like to thank my supervisor, Professor Leif Eriksson, and my colleague team, Anna, Martin, Min, Patricia, Samuel, Chunxia, Johanna, Kevin, Sonali, Jalil, Melissa, Simone, Olof and Gabriella for all the interesting discussions during time as a PhD student.

Heartfelt thanks to all TRAINERS guys, especially Chetan, colleague and friend, for all the enlightening work-related talks and all the inebriating beverage-related activities.

A big thank you goes also to all my flat mates with whom I shared houses during these years: Laura, Marielle, Steffen, Fabiola, Maximilian, Tareq, Helge, Max, Anna, Johannes and Lisa for my first shared house; Amir, Sebastian and Jack for my shared accommodation in Galway; Milo for the great help during September/October and Mirey, Nan, Vivek and of course Yuri.

A special token of gratitude to all the friends in Göteborg and Galway: Pavel, Swaraj, Jae, Sergi, Matteo, Franc, Luca, Eric, Aaron, Eoghan, Tam, Gigi, Anton, Pepi, Veronica, Stephen, Cioin, Andrea, El Belin, Massimo, Co-coach and all the others.

A special thanks to Sibilla, my best place for eating hamburgers. Whatever the clock may be.

A special thanks to AZ, for hosting me, and to all the people inside the company, especially Giuseppina, Andrey and Anders: you made my AZ dream true.

A grateful acknowledgement to all the people who decided to read and comment this thesis.

Special thanks to "Salmon group" and our favourite emoticon in the group.

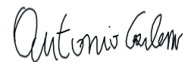
And to Cocktailskolan for the alternative "in vitro" assay optimization .

A huge thanks to Milo and Adriana for mathematical support and Giovanni for linguistic and caloric support.

A heartfelt thanks to all the people who told me I was from Scotland or Ireland and thus indirectly confirmed that my English had improved.

And to the Kantaller family. no further words needed.

And of course a huge thank you to my family and all my friends in Italy, who are always there, no matter how far.



Göteborg, November 2019

Bibliography

- (1) Walter, P. & Ron, D. The unfolded protein response: from stress pathway to homeostatic regulation. *Science* **334**, 1081-1086, doi:10.1126/science.1209038 (2011).
- (2) Almanza, A. *et al.* Endoplasmic reticulum stress signalling - from basic mechanisms to clinical applications. *FEBS J* **286**, 241-278, doi:10.1111/febs.14608 (2019).
- (3) Walter, P. & Ron, D. The unfolded protein response: from stress pathway to homeostatic regulation. *Science* **334**, 1081-1086, doi:10.1126/science.1209038 (2011).
- (4) Tabas, I. & Ron, D. Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress. *Nat Cell Biol* **13**, 184-190, doi:10.1038/ncb0311-184 (2011).
- (5) Rutkowski, D. T. & Hegde, R. S. Regulation of basal cellular physiology by the homeostatic unfolded protein response. *J Cell Biol* **189**, 783-794, doi:10.1083/jcb.201003138 (2010).
- (6) Wang, S. & Kaufman, R. J. The impact of the unfolded protein response on human disease. *J Cell Biol* **197**, 857-867, doi:10.1083/jcb.201110131 (2012).
- (7) Doultinos, D. *et al.* Control of the Unfolded Protein Response in Health and Disease. *SLAS Discov* **22**, 787-800, doi:10.1177/2472555217701685 (2017).
- (8) Hetz, C., Axten, J. M. & Patterson, J. B. Pharmacological targeting of the unfolded protein response for disease intervention. *Nat Chem Biol* **15**, 764-775, doi:10.1038/s41589-019-0326-2 (2019).
- (9) Pinkaew, D. *et al.* Fortilin binds IRE1 α and prevents ER stress from signaling apoptotic cell death. *Nature communications* **8**, 18 (2017).
- (10) Nadanaka, S., Okada, T., Yoshida, H. & Mori, K. Role of disulfide bridges formed in the luminal domain of ATF6 in sensing endoplasmic reticulum stress. *Mol Cell Biol* **27**, 1027-1043, doi:10.1128/MCB.00408-06 (2007).
- (11) Fedorov, O., Muller, S. & Knapp, S. The (un)targeted cancer kinome. *Nat Chem Biol* **6**, 166-169, doi:10.1038/nchembio.297 (2010).
- (12) Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-158, doi:10.1038/nature05610 (2007).
- (13) Liu, Y. & Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol* **2**, 358-364, doi:10.1038/nchembio799 (2006).
- (14) Muller, S., Chaikuad, A., Gray, N. S. & Knapp, S. The ins and outs of selective kinase inhibitor development. *Nat Chem Biol* **11**, 818-821, doi:10.1038/nchembio.1938 (2015).
- (15) Carlesso, A., Chintia, C., Gorman, A. M., Samali, A. & Eriksson, L. A. Binding Analysis of the Inositol-Requiring Enzyme 1 Kinase Domain. *Acs Omega* **3**, 13313-13322, doi:10.1021/acsomega.8b01404 (2018).
- (16) Sanches, M. *et al.* Structure and mechanism of action of the hydroxy-aryl-aldehyde class of IRE1 endoribonuclease inhibitors. *Nature Communications* **5**, 4202, doi:10.1038/ncomms5202

- (17) Wang, L. *et al.* Divergent allosteric control of the IRE1alpha endoribonuclease using kinase inhibitors. *Nat Chem Biol* **8**, 982-989, doi:10.1038/nchembio.1094 (2012).
- (18) Jha, B. K. *et al.* Inhibition of RNase L and RNA-dependent protein kinase (PKR) by sunitinib impairs antiviral innate immunity. *J Biol Chem* **286**, 26319-26326, doi:10.1074/jbc.M111.253443 (2011).
- (19) Harrington, P. E. *et al.* Unfolded Protein Response in Cancer: IRE1a Inhibition by Selective Kinase Ligands Does Not Impair Tumor Cell Viability. *ACS Med Chem Lett* **6**, 68-72, doi:10.1021/ml500315b (2015).
- (20) Feldman, H. C. *et al.* Structural and Functional Analysis of the Allosteric Inhibition of IRE1alpha with ATP-Competitive Ligands. *ACS Chem Biol* **11**, 2195-2205, doi:10.1021/acscchembio.5b00940 (2016).
- (21) Cramer, C. J. *Essentials of computational chemistry: theories and models.* (John Wiley & Sons, 2013).
- (22) Pielak, L. *Ideas of quantum chemistry.* (Elsevier, 2006).
- (23) Bottaro, S. & Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **361**, 355-360, doi:10.1126/science.aat4010 (2018).
- (24) Braun, E. *et al.* Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living J Comput Mol Sci* **1**, doi:10.33011/livecoms.1.1.5957 (2019).
- (25) Mobley, D. L. *et al.* Escaping atom types in force fields using direct chemical perception. *Journal of Chemical Theory and Computation* **14**, 6076-6092 (2018).
- (26) Wang, L. P. *et al.* Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J Phys Chem B* **121**, 4023-4039, doi:10.1021/acs.jpcc.7b02320 (2017).
- (27) Spomer, J. *et al.* RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem Rev* **118**, 4177-4338, doi:10.1021/acs.chemrev.7b00427 (2018).
- (28) Sagui, C. & Darden, T. A. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual review of biophysics and biomolecular structure* **28**, 155-179 (1999).
- (29) Cisneros, G. A., Karttunen, M., Ren, P. & Sagui, C. Classical electrostatics for biomolecular simulations. *Chem Rev* **114**, 779-814, doi:10.1021/cr300461d (2014).
- (30) Fasman, G. D. *Prediction of protein structure and the principles of protein conformation.* (Springer Science & Business Media, 2012).
- (31) Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242, doi:10.1093/nar/28.1.235 (2000).
- (32) Salmaso, V. & Moro, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front Pharmacol* **9**, 923, doi:10.3389/fphar.2018.00923 (2018).
- (33) Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry* **47**, 1739-

- 1749 (2004).
- (34) Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **267**, 727-748, doi:10.1006/jmbi.1996.0897 (1997).
- (35) Baxter, C. A., Murray, C. W., Clark, D. E., Westhead, D. R. & Eldridge, M. D. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins: Structure, Function, and Bioinformatics* **33**, 367-382 (1998).
- (36) Kjellander, R. *Thermodynamics Kept Simple—A Molecular Approach: What is the Driving Force in the World of Molecules?*, (CRC Press, 2015).
- (37) Dill, K. & Bromberg, S. *Molecular Driving Forces: Statistical Thermodynamics in Biology. Chemistry, Physics, and Nanoscience* (2010).
- (38) Guedes, I. A., Pereira, F. S. S. & Dardenne, L. E. Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges. *Front Pharmacol* **9**, 1089, doi:10.3389/fphar.2018.01089 (2018).
- (39) Liu, J. & Wang, R. Classification of current scoring functions. *J Chem Inf Model* **55**, 475-482, doi:10.1021/ci500731a (2015).
- (40) Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52**, 609-623, doi:10.1002/prot.10465 (2003).
- (41) Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V. & Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aid Mol Des* **11**, 425-445 (1997).
- (42) Davis, A. M., St-Gallay, S. A. & Kleywegt, G. J. Limitations and lessons in the use of X-ray structural information in drug design. *Drug Discov Today* **13**, 831-841 (2008).
- (43) Davis, A. M., Teague, S. J. & Kleywegt, G. J. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew Chem Int Ed Engl* **42**, 2718-2736, doi:10.1002/anie.200200539 (2003).
- (44) Kleywegt, G. J. On vital aid: the why, what and how of validation. *Acta Crystallogr D Biol Crystallogr* **65**, 134-139, doi:10.1107/S090744490900081X (2009).
- (45) Sondergaard, C. R., Garrett, A. E., Carstensen, T., Pollastri, G. & Nielsen, J. E. Structural artifacts in protein-ligand X-ray structures: implications for the development of docking scoring functions. *J Med Chem* **52**, 5673-5684, doi:10.1021/jm8016464 (2009).
- (46) Kleywegt, G. J. Validation of protein crystal structures. *Acta Crystallogr D Biol Crystallogr* **56**, 249-265, doi:10.1107/s0907444999016364 (2000).
- (47) Pascarella, S. & Paiardini, A. *Bioinformatica: dalla sequenza alla struttura delle proteine*. (Zanichelli, 2011).
- (48) Plach, M. G., Grasser, K. & Schubert, T. MicroScale Thermophoresis as a tool to study protein-peptide interactions in the context of large eukaryotic protein complexes. *Bio Protoc* **7**, 2632e (2017).

- (49) Hastie, C. J., McLauchlan, H. J. & Cohen, P. Assay of protein kinases using radiolabeled ATP: a protocol. *Nature protocols* **1**, 968 (2006).
- (50) Vigers, G. P. & Rizzi, J. P. Multiple active site corrections for docking and virtual screening. *J Med Chem* **47**, 80-89, doi:10.1021/jm030161o (2004).
- (51) Feldman, H. C. *et al.* Structural and Functional Analysis of the Allosteric Inhibition of IRE1 α with ATP-Competitive Ligands. *ACS Chem Biol* **11**, 2195-2205, doi:10.1021/acscchembio.5b00940 (2016).
- (52) Toledo Warshaviak, D., Golan, G., Borrelli, K. W., Zhu, K. & Kalid, O. Structure-based virtual screening approach for discovery of covalently bound ligands. *J Chem Inf Model* **54**, 1941-1950, doi:10.1021/ci500175r (2014).
- (53) Zhu, K. *et al.* Docking covalent inhibitors: a parameter free approach to pose prediction and scoring. *J Chem Inf Model* **54**, 1932-1940, doi:10.1021/ci500118s (2014).
- (54) Tomasio, S. M., Harding, H. P., Ron, D., Cross, B. C. & Bond, P. J. Selective inhibition of the unfolded protein response: targeting catalytic sites for Schiff base modification. *Mol Biosyst* **9**, 2408-2416, doi:10.1039/c3mb70234k (2013).
- (55) Carlesso, A., Eriksson, L. A. Selective Inhibition of IRE1 Signalling Mediated by MKC9989: New Insights from Molecular Docking and Molecular Dynamics Simulations. *ChemistrySelect* **4**, 3199-3203 (2019).
- (56) Carlesso, A., Chintha, C., Gorman, A. M., Samali, A. & Eriksson, L. A. Effect of Kinase Inhibiting RNase Attenuator (KIRA) Compounds on the Formation of Face-to-Face Dimers of Inositol-Requiring Enzyme 1: Insights from Computational Modeling. *Int J Mol Sci* **20**, doi:10.3390/ijms20225538 (2019).
- (57) Rojas-Rivera, D. *et al.* When PERK inhibitors turn out to be new potent RIPK1 inhibitors: critical issues on the specificity and use of GSK2606414 and GSK2656157. *Cell death and differentiation* **24**, 1100 (2017).
- (58) Konc, J., Cesnik, T., Konc, J. T., Penca, M. & Janezic, D. ProBiS-database: precalculated binding site similarities and local pairwise alignments of PDB structures. *J Chem Inf Model* **52**, 604-612, doi:10.1021/ci2005687 (2012).
- (59) Bouchecareilh, M., Higa, A., Fribourg, S., Moenner, M. & Chevet, E. Peptides derived from the bifunctional kinase/RNase enzyme IRE1 α modulate IRE1 α activity and protect cells from endoplasmic reticulum stress. *The FASEB Journal* **25**, 3115-3129 (2011).
- (60) Hetz, C. & Papa, F. R. The Unfolded Protein Response and Cell Fate Control. *Mol Cell* **69**, 169-181, doi:10.1016/j.molcel.2017.06.017 (2018).
- (61) Maly, D. J. & Papa, F. R. Druggable sensors of the unfolded protein response. *Nat Chem Biol* **10**, 892-901, doi:10.1038/nchembio.1664 (2014).
- (62) Limongelli, V., Bonomi, M. & Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proceedings of the National Academy of Sciences* **110**, 6358-6363 (2013).
- (63) Tiwary, P., Limongelli, V., Salvalaglio, M. & Parrinello, M. Kinetics of protein-ligand

unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc Natl Acad Sci U S A* **112**, E386-391, doi:10.1073/pnas.1424461112 (2015).

