UNIVERSITÀ
DI SIENA
1240

Department of Biotechnology, Chemistry and Pharmacy

Doctoral research school in Chemical and Pharmaceutical Sciences

Cycle XXXIV

Design of small molecules interacting with wild-type and mutant rhodopsins

COORDINATOR OF THE DOCTORAL SCHOOL
Prof. Maurizio Taddei

TUTOR
Prof. Mattia Mori

CANDIDATE
Francesca Picarazzi

Academic year 2021/2022

Table of contents

Aim and summary of this thesis

Chemical transformations in a biological system are often modulated by the interaction between macromolecules (proteins, nuclei acids, etc.) and molecular partners, including small molecules of different origin. Pharmacological agents are thus envisioned as a tool to drive chemical transformations aimed at modifying some specific cell functions in a pathological context. However, the design of new drugs is frequently hampered by experimental limitations such as the very short time scale of the biological process and interactions (in the order of a few nanoseconds), and the difficult structural characterization of the regions of the protein where these interactions occur. In this Ph.D thesis work, molecular modeling tools were used to extensively study some pathological systems of therapeutic relevance, with the ultimate goal to accelerate the validation of a target, and the identification and optimization of small molecules as drug candidates. Overall, computational modeling techniques are a fast and reliable tool for overcoming experimental limitations and to extensively studying the factors that determine or influence biochemical processes at the atomic-molecular level, although the validation of theoretical results generally requires experimental efforts.

In the **introductory section** of this Ph.D thesis work, the overall molecular basis of *retinitis pigmentosa* (RP) are described, as well as physiological and pathological processes in which the rhodopsin protein is involved, starting from its biosynthesis to the execution of the photoreceptor activity, along with the pathological mutations implicated in RP. In addition, physiological features of aldehyde dehydrogenase (ALDH) 1A3, an enzyme that is responsible for the conversion of the photoisomerized all-*trans* retinal to retinoic acid in retinal tissue, are described. RP is an inherited degenerative disease of the retina that is mostly caused by pathological mutations in numerous genes, including that of rhodopsin. However, altered physiological levels of the endogenous retinal cofactor are also related to ocular diseases, including RP. Although the N-terminal P23H rhodopsin mutation causes structural destabilization and impairs the correct protein folding and localization, accounting for the most of RP cases, *in vitro* studies showed that the correct folding of P23H mutant rhodopsin can be partially rescued by small chaperone ligands binding within the 11-*cis* retinal binding site [1], validating rhodopsin as a suitable target for pharmacological treatment of RP by small chaperone ligands. On the other hand, C-terminal P347L/S/R/T/Q/A mutations in the VAPA-COOH motif, which is essential for the recognition of rhodopsin by the specific transporter ARF4 [2–4] that traffics the folded rhodopsin from the Golgi to the rod outer segment (OS), are associated to the most severe forms of RP. Unfortunately, little is known about the structural features of clinically relevant P347 mutants, and their chemical and physicochemical properties do not directly correlate with their pathological role.

Considering the topics summarized above, the first aim of this thesis is to identify new small molecular chaperone compounds able to bind the P23H mutant and to restore the correct rhodopsin folding and localization. The second objective of this thesis is to provide structural details on the role of rhodopsin C-terminal mutations in RP. Finally, by combining the hypothesis that retinal concentration in the cells may be modulated by ALDH1A3 enzymatic activity with the evidence that some forms of RP are associated with decreased levels of endogenous 11-*cis* retinal, we aimed at the understanding of ALDH1A3 role in the context of RP. Accordingly, the third aim of this thesis is indeed to support the design and development of selective and potent small molecule inhibitors of ALDH1A3 as biological tools for contributing to validate this enzyme as an alternative target in RP treatment.

Overall, results of this thesis work are expected to open novel possibilities for developing therapeutic strategies against RP, as well as to shed light into the pathological events that might affect the vision process. To enhance this understanding, three different molecular modeling approaches have been used in this thesis, as discussed in the respective sections, and briefly outlined below.

Section 2 focuses on the identification of new small molecular chaperone compounds able to bind the P23H mutant in the retinal binding site and to restore the correct rhodopsin folding and localization. In this context, a structure-based virtual screening was carried out using the P23H rhodopsin structure refined by molecular dynamics (MD) simulations as a receptor. Different libraries of compounds were screened in the search of potentially bioactive hits, including a high diversity *in house* library of natural products [5], the DrugBank database [6] and the commercially available MolPort database [7]. This latter database was further filtered to select compounds bearing up to three aldehyde groups, which are thought to mimic the main pharmacophore of the natural substrate 11-*cis* retinaldehyde in covalent bonding to rhodopsin K296. By virtual screening, compounds were prioritized based on a combination of docking score, docking pose, and polar/hydrophobic interactions with the receptor. Aldehyde compounds were selected also by comparing the non-covalent and covalent docking poses and choosing those compounds that show highly comparable poses in both settings. The most promising compounds emerged from this virtual screening were further investigated *in vitro* for their ability to restore the correct localization of mutant rhodopsin on the cell membrane by experimental validation in cell-based assay. In this section, the structure-based approach has emerged as a fast and valuable tool for the virtual screening of large libraries of compounds to identify molecular chaperone of P23H mutant rhodopsin with experimental validation.

Section 3 describes a computational workflow for studying clinically relevant C-terminal rhodopsin mutants in RP, and to provide structural insights into the molecular determinants that might be relevant for the trafficking of WT and mutant rhodopsin to the rod membrane. The conformational space WT and mutant rhodopsin in the phospholipidic membrane was explored by MD simulations. In addition to the full-length protein forms, the corresponding isolated N-capped C-terminal deca-peptides were also studied as a simplified and representative model system. In agreement with experimental evidence [2], MD simulations suggested that rhodopsin C-terminal tail is mostly involved in trafficking processes. The WT VAPA-COOH motif adopts a unique conformation that is not found in pathological mutants, suggesting that structural features could better explain the pathogenicity of P347 rhodopsin mutants than physicochemical or steric determinants. Interestingly, analysis of MD trajectories of isolated deca-peptides showed results that are highly comparable to the full-length system, suggesting that they are reliable model systems for the study of C-terminal rhodopsin mutations.

Section 4 focuses on the role of ALDH1A3 in the metabolism of retinal cofactor. Here, we hypothesized that inhibition of ALDH1A3 may increase the level of endogenous 11-*cis* retinal in RP whose molecular basis are low levels of retinal cofactor. Starting from the reference ALDH inhibitor GA11 previously developed by our research group, a novel series of imidazo[1,2-*a*]pyridines was developed and optimized by means of a structure-based approach. These novel compounds were evaluated *in vitro* for their inhibitory activity and selectivity against the ALDH1A family, and they were also investigated through X-ray crystallography and molecular modeling studies for their ability to interact with the catalytic site of the 1A3 isoform. Among the most promising ALDH1A3 inhibitors, NH51 did not show a sufficient electron density in crystallographic studies. Molecular docking, MD simulations and Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) [8] calculations helped to fill this structural gap. Furthermore, theoretical affinity of compound NR6 was evaluated by MD simulations, while alanine scanning calculations highlighted the thermodynamic role of Y472, specific for the ALDH1A3 isoform, which emerged as a fundamental residue in ALDH1A3/inhibitor interaction. Overall, the computational approach provided structural and thermodynamics information that are essential for the design and optimization of potent and selective small molecule inhibitors of ALDH1A3. This section is therefore emphasizing the advantage of *in silico* methods in drug discovery as useful alternative or supplementary aid to experimental approaches.

Section 5 summarizes two side projects and a review article carried out in collaboration with the research group of Prof. Maurizio Zazzi at the laboratory of Microbiology and Virology of the

Department of Medical Biotechnologies at the University of Siena. Driven by an ongoing collaboration with the research group in which I carried out my doctoral studies, and from the scientific interest to address topics related to the COVID-19 pandemic, these projects focused on antiviral drug discovery targeting the RNA-dependent RNA polymerase (RdRp) of emerging RNA viruses. Structural basis of RdRp drug resistance were also addressed through a combination of biology and computational modeling studies.

Finally, as an additional side project of this thesis related to the COVID-19 pandemic, an international collaboration with the Karolinska Institutet of Stockholm, Sweden was reported. Specifically, this project focused on the study of a peptide from SARS-CoV-2 that promotes the activation of NKG2A-expressing natural killer (NK) cells through impaired inhibition.

1. Introduction

1.2. Structural features of retinal photoreceptor cells

The retina is the light-sensitive tissue of the eyes and consists of several layers of nerve cells. It is the most important structure of the eye and is essential in the process of vision. In fact, light photons are captured by the retina and are transformed into electrochemical signals that, through the optic nerve, reach the brain that elaborates them as images. In RP, the degradation of photoreceptor cells starts from the outer side of the retinal tissue to progress towards the center, which results in an initial narrowing of the visual field and, in severe cases, in the loss of central vision [9].

In the retina there are about 131 million light-sensitive photoreceptor cells classified into two different types, i.e., cones and rods [10]. Cones are so called because of their shape and are mainly responsible for receiving the details of images and various colors. There are three types of cones, each of which contains a pigment that makes them sensitive to different wavelengths in the visible spectrum. Specifically, they have absorption peaks at 420, 534 and 564 nm, which correspond to blue, green, and red light, respectively [11]. In the retina there are about 6 million of cone cells and they are mainly concentrated in the macula, the central region of the retina [10]. Therefore, cones are essential for central vision (reading, face recognition, driving, etc.) as their ability to distinguish details is about one hundred times higher than the rods. Rods have an elongated and tapered shape, have a peak absorption of 498 nm [11] and react mainly to the contrast between light and dark as well as the movement of objects. There are about 125 million rod cells and they are concentrated in the peripheral area of the retina, where they are present in an amount twenty times higher than the cones [10]. However, both photoreceptors have approximately the same structure with an OS, an inner segment (IS), and a synaptic termination (Figure 1).

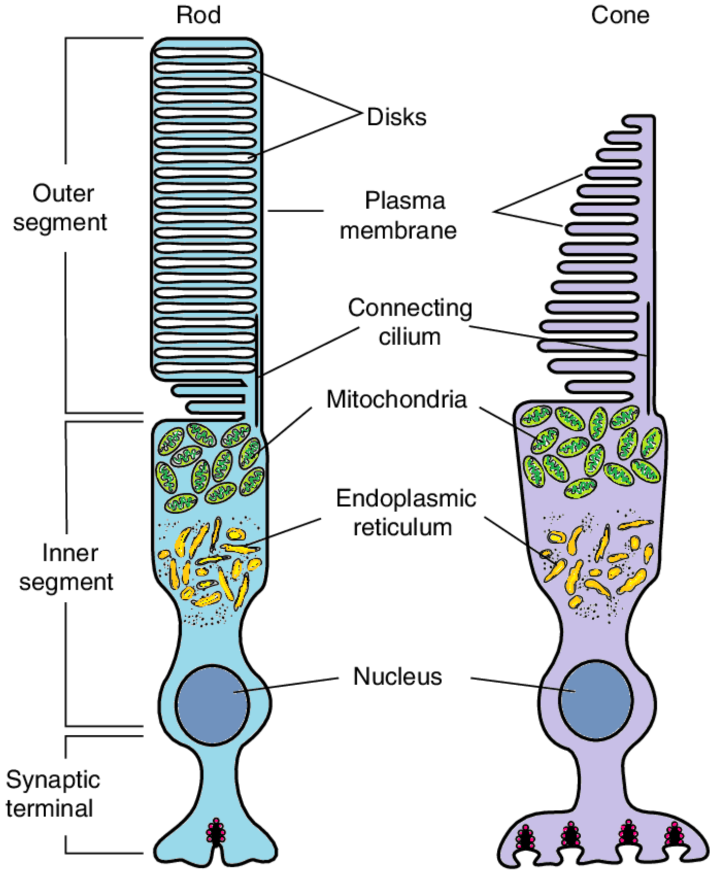


Figure 1. Schematic representation of rod and cone photoreceptor cells. The picture is adapted from Cote 2007 [12].

The OS of the cones is shaped like a truncated pyramid, while that of the rods is cylindrical and elongated; in both cases, this part is characterized by a layered series of lamellae, which delimits membranous flattened disks immersed in the cytoplasm of the cell [13,14]. These disks contain pigments that react to light and cause changes in the membrane potential of the photoreceptor. The OS of cones and rods is in contact with the retinal pigment epithelium, the outermost layer of the retina [15]. Here, ground state rhodopsin is restored by a series of reactions that involve the detachment of the all-*trans* retinal and its oxidation by mean of the ALDH1A3 enzyme, which metabolizes the photoisomerized all-*trans* retinal to retinoic acid with an enzymatic activity 10-fold higher with respect to other ALDH isoforms [16]. All-*trans* retinal can be eventually regenerated into the 11-*cis* form, which is thus suitable to participate in a new visual cycle.

The retinal cofactor is the most important molecule for the phototransduction of signal and its regeneration is fundamental for the vision process. The IS is characterized by the presence of intracellular organelles, such as mitochondria and granular endoplasmic reticulum (ER) membranes, indispensable for cellular metabolism. Their role is to produce new pigment molecules as they are cleaved. This portion continues by narrowing into an outer fiber (connecting cilium), which is followed by the part of the cell body containing the nucleus. The latter is connected to the synaptic termination, which has a bulb shape (spherule) in rods, flooded and branched (pedicle) in cones [13,14]. The synaptic termination allows the transmission of signals from the photoreceptor to the bipolar cells by synapse, i.e., by biochemical transmission between nerve cells. This region is analogous to the synaptic button of the axon terminals of neurons, where the vesicles containing the neurotransmitter are present [17].

1.3. *Retinitis Pigmentosa*

RP refers to a group of inherited retinal dystrophies characterized by the progressive degeneration of retinal photoreceptor cells. Clinical symptoms of RP include night blindness, loss of peripheral vision, decreased visual acuity, photopsia, and loss of central vision in more advanced stages [18]. RP is a rare disease, with a frequency of 1 case on 4000 [9], which originates from mutations in more than 70 genes including the rhodopsin gene (*RHO*) (genes and loci causing inherited retinal diseases can be found at RetNet, https://sph.uth.edu/RetNet/) [19]. In total, hereditary retinal degeneration affect about 2 million people worldwide [20], of which about 1.5 million are affected by RP. In most RP patients, symptoms first appear around young adulthood and then develop gradually over the time, although earlier and later forms of RP are also reported [21–24]. There are three genetic forms of RP: i) autosomal dominant, ii) autosomal recessive, iii) X-linked [25–27]. The X-linked RP can be either recessive, thus affecting mainly male individuals, or dominant, thus affecting individuals of both sexes, even if those of the female sex are, however, always affected in a lighter way [28]. Generally, RP is not syndromic although several forms associated with syndromes are known, the most frequent of which is Usher syndrome, i.e., RP combined with progressive deafness [29]. Sporadic forms (about 30% of all cases) involve the presence of a single case in a family. Sporadicity is only a familial finding, but it is very difficult to exclude recessive inheritance or X-linked RP, if the affected person is a male. It is also necessary to differentiate primary RP, in which there is only ocular involvement, from that associated with extraocular diseases [30].

1.3.1. Clinical symptoms, course, and outcome of the disease

The main symptoms that may lead to suspect that a patience is dealing with a case of RP are essentially two:

i) Twilight and night blindness. There may be difficulty seeing in low light conditions or a delayed adaptation as one moves from lighted to dark environments. In fact, in most cases, the rods are the cells most affected in the early stages of development of the disease.
ii) Narrowing of the visual field (tunnel vision), which is characterized by difficulty in perceiving objects located lateral to the visual field (Figure 2).

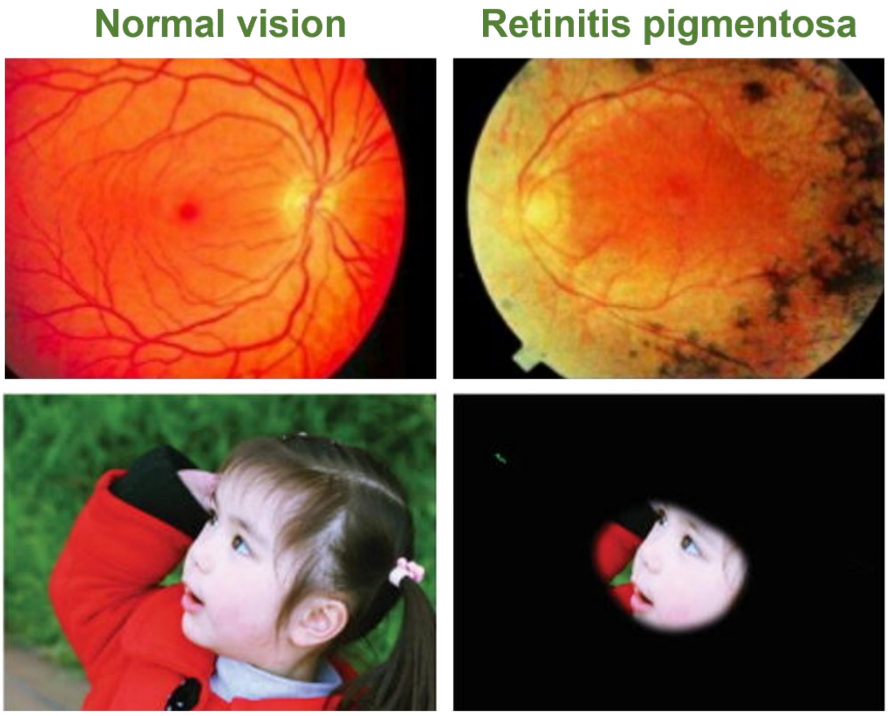
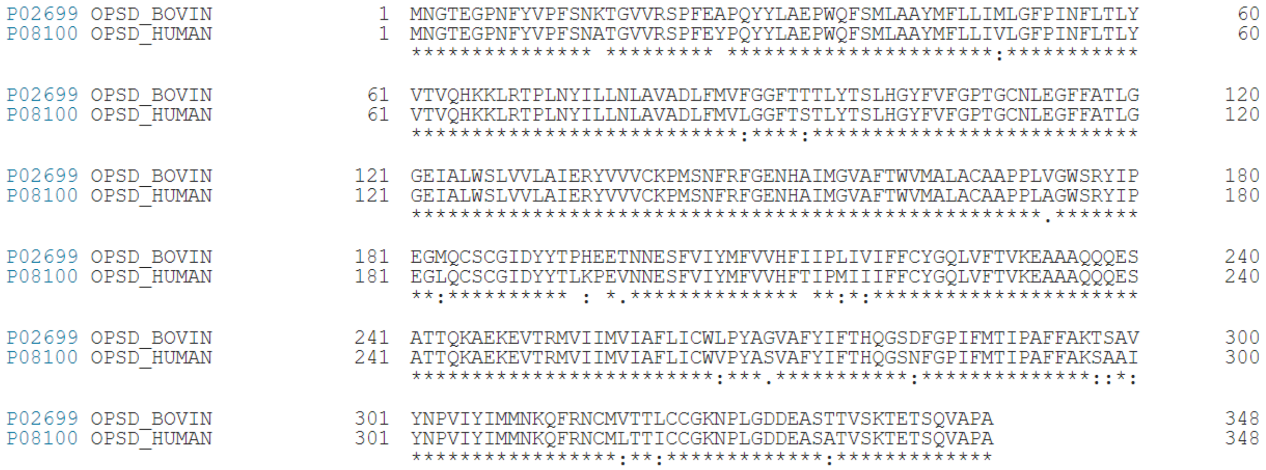


Figure 2. RP phenotypic and visual comparison. The upper left and right boxes show the retinal fundus. RP-affected retinae have fovea preservation amidst peripheral cell death, leading to tunnel vision as depicted in the lower right box. The picture is adapted from Justus et al., 2016 [31].

The alteration of the visual field is progressive and can also involve the central part of the retina, with loss of central visual abilities [32]. The rate of disease progression and age of onset of symptoms vary with many factors including the genetic transmission pattern. There is also an increased sensitivity to glare, contrasts fade, and it becomes difficult to perceive the surrounding environment. The disease can be diagnosed in childhood, adolescence and, not infrequently, even in adulthood [33]. In doubtful cases, the diagnosis is based on all available clinical data (age of onset, evolutionary mode, possible association with other ocular symptoms or with other organs and apparatuses) and on a thorough electroretinogram (ERG) and adaptive testing. The study of the chromatic sense and retinal fluorescein angiography may be useful and complementary. It is also necessary to examine the whole family, to define the type of hereditary transmission [34]. The course of the disease has an extremely variable duration, but it is always progressive and disabling. In most cases the symptoms worsen and unfortunately the field of vision narrows more and more until it closes completely to blindness [35]. Other disorders also appear, such as glare, the inability to distinguish colors, and a particular form of cataract [9]. Unfortunately, there is no cure available that can halt the progression of the disease or restore vision. Therefore, RP is associated with poor prognosis. Although new therapeutic strategies are emerging including gene therapy [36–38], retinal transplant [39] and retinal prostheses [40,41], no pharmacological treatment is available and the therapeutic approach allows only to slow down the degenerative process and consists in protection from sunlight, vitamin therapy [33,42,43], in the treatment of complications (cataracts and macular edema) and in helping patients to cope with social and psychological problems related to blindness.

1.4. Structure and biochemical features of rhodopsin

Rhodopsin is a class A member of the G protein-coupled receptor (GPCR) family and is the visual pigment of the rod photoreceptor cells. The human GPCR superfamily is a large group of transmembrane receptors that transduce an extracellular signal into intracellular pathways through the activation of heterotrimeric G proteins [44]. Their involvement in multiple physiological and pathological processes make them one of the most important targets in drug discovery [45]. The earliest high resolution crystallographic structure of rhodopsin dates back to 2000, when Palczewski and colleagues succeeded in crystallizing bovine rhodopsin in complex with 11-*cis* retinal cofactor with a resolution of 2.80 Å (PDB-ID: 1F88) [46], giving further insights into the mechanisms of receptor activation, and the specific ligand and G-protein interactions. Since bovine and human rhodopsin forms (Uniprot-ID: P02699 and P08100, respectively [47]) share a very high percentage of identity of the primary sequence (93.4%), bovine crystallographic structures are considered extremely reliable models to study the human orthologue (Figure 3).



```
P02699 OPSD_BOVIN          1  MNGTEGPNFYVPFSNKTGVVRSPFEAPQYYLAEPWQFSMLAAYMFLLIMLGFPINFLTLY      60
P08100 OPSD_HUMAN          1  MNGTEGPNFYVPFSNATGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTLY      60
                              *************** ********* **********************:***********

P02699 OPSD_BOVIN         61  VTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLG     120
P08100 OPSD_HUMAN         61  VTVQHKKLRTPLNYILLNLAVADLFMVLGGFTSTLYTSLHGYFVFGPTGCNLEGFFATLG     120
                              ***************************:****:***************************

P02699 OPSD_BOVIN        121  GEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLVGWSRYIP     180
P08100 OPSD_HUMAN        121  GEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLAGWSRYIP     180
                              ****************************************************.*******

P02699 OPSD_BOVIN        181  EGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFCYGQLVFTVKEAAAQQQES     240
P08100 OPSD_HUMAN        181  EGLQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMIIIFFCYGQLVFTVKEAAAQQQES     240
                              **:********** : *.************** **:*:**********************

P02699 OPSD_BOVIN        241  ATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSDFGPIFMTIPAFFAKTSAV     300
P08100 OPSD_HUMAN        241  ATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIFTHQGSNFGPIFMTIPAFFAKSAAI     300
                              *************************:***.***********:**************::*:

P02699 OPSD_BOVIN        301  YNPVIYIMMNKQFRNCMVTTLCCGKNPLGDDEASTTVSKTETSQVAPA                 348
P08100 OPSD_HUMAN        301  YNPVIYIMMNKQFRNCMLTTICCGKNPLGDDEASATVSKTETSQVAPA                 348
                              *****************:**:*************:*************
```

Figure 3. Sequence alignment between bovine and human rhodopsin FASTA sequences taken from the Uniprot database [47].

Over the years, X-ray structures of bovine opsin and rhodopsin in active conformations and human rhodopsin-arrestin complex were solved [48–54], increasing the understanding of photoactivation and signal transduction processes by GPCRs [55]. Bovine rhodopsin structure comprises 348 amino acids organized in a single polypeptide chain forming a seven transmembrane α-helices core architecture, common to all GPCRs, with three extracellular and intracellular loops (Figure 4) [56–59].

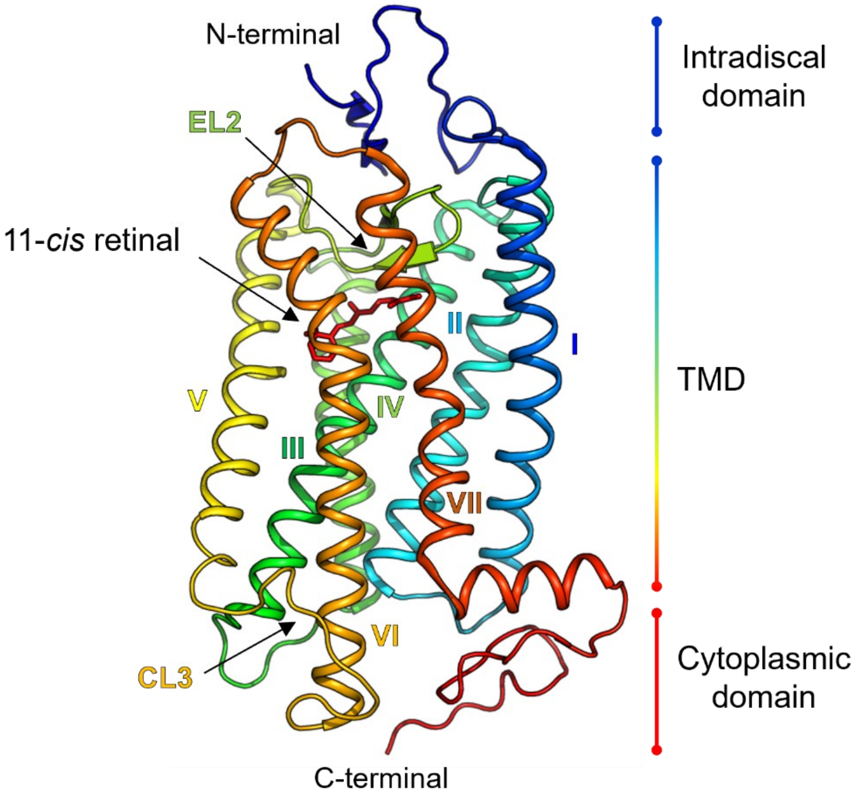


Figure 4. Schematic representation of bovine rhodopsin photoreceptor at 2.20 Å resolution (PDB-ID: 1U19) [60]. The N- and C- terminal tails are colored blue and red, respectively and the helices are labelled. The 11-*cis* retinal is highlighted by red sticks.

The N-terminal tail faces the intradiscal side of the membrane, and it is folded in a two-stranded β-sheet followed by a loop region of about 24 amino acid residues. Moreover, it contains the glycosylation sites on residue N2 and N15 and is involved in ligand recognition. It has been shown that the N-terminal tail is crucial for the correct folding of rhodopsin, while mutations in this region (i.e., T4, N5, T17, P23, and N28) can lead to the development of autosomal dominant RP. [61,62]. The second extracellular loop (EL2) ranges from G175 to N200 and has a crucial role in the interaction with ligands and folding [63]. EL2 is folded in a twisted two-stranded β-sheet, is located at the entrance of the transmembrane domain (TMD) and has been compared to a "lid" that opens to let in the ligands and closes to prevent them from exiting [64]. Its structure is stabilized both by hydrophobic interactions with residues Y178, P180, M183 (L183 in humans), C185, and C187 of EL2, by E181 and Y192 of EL2, Y268 and by a disulfide bond between C187 of the β-sheet and C110 of helix 3. Functional conformation of the lid is assured by a salt bridge between N190 and R177 [65] (Figure 5A). The C-terminal tail faces the cytoplasmic side of the membrane and ranges from M309 to the last amino acid of the protein, i.e., A348. Although this region is not folded into a well-defined secondary structure, it is involved in important functions such as recognition by specific transport carriers through the C-terminal VAPA-COOH motif, and in the interaction with G proteins, as it contains palmitoylation sites at cysteine residues C322 and C323 and phosphorylation sites at serine residues (S334, S338, and S343) useful for the deactivation of rhodopsin by binding arrestin protein [66]. Mutations of residues in this region are associated to the most aggressive form of RP due to mistrafficking and mislocalization of rhodopsin [4]. The cytosolic loop CL3 between helices 5 and 6 is the place for interaction with G proteins. In ground-state rhodopsin, it is structurally disordered, while in active-state rhodopsin model it is mostly folded [49–51,67]. The TMD represents the central structure of the receptor system, as it forms the binding site of the 11-*cis* retinal chromophore that is crucial to the vision process. Light-sensitive rhodopsin consists of the apoprotein opsin covalently bound to the 11-*cis* retinal cofactor through a protonated Schiff base with the side chain of K296 [46,60,68]. The retinal cofactor establishes an electrostatic interaction with the Schiff base to the E113 residue and is accommodated in a hydrophobic pocket of about 352 $Å^3$ of volume formed by M207, F208, and F212 residues from helix 5, and W265 and Y268 residues from helix 6 that stabilize the polyene backbone of the retinal with a parallel orientation with respect to the plane of the lipid membrane (Figure 5B).

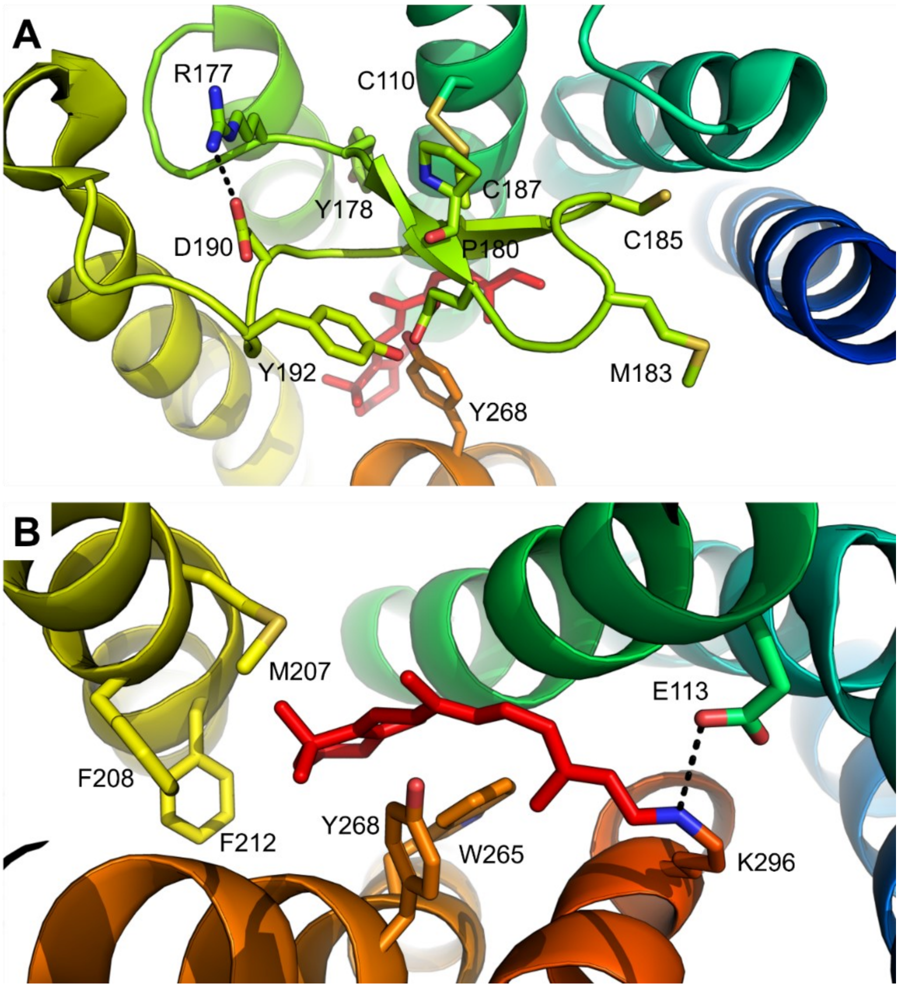


Figure 5. Graphical representation of PDB-ID 1U19 crystallographic structure with a particular detail on (A) the EL2 stabilizing interactions and (B) the hydrophobic retinal binding pocket within rhodopsin's TMD. Residues of interest are represented as sticks, the 11-*cis* retinal cofactor is represented as red sticks, and polar interactions are highlighted by black dashed lines.

Rhodopsin is activated through the photon-mediated isomerization of 11-*cis* retinal, to all-*trans* retinal [69]. Photon absorption by the π-electron system of the 11-*cis* retinal allows the transition from a curved conformation, given by the *cis* bond to an extended conformation, given by the *trans* bond, while the β-ionone moiety is shifted toward the space between helices 5 and 6. This photoswitch facilitates the binding of transducin to rhodopsin, and subsequently generates an hyperpolarization of the plasma membrane via channels gated by cyclic guanosine monophosphate [69]. Light-induced isomerization of 11-*cis* retinal to its all-*trans* form (Figure 6) induces several distinct conformational intermediates such as photorhodopsin, bathorhodopsin, lumirhodopsin, and metarhodopsin I, giving rise to active metarhodopsin II [69–73].

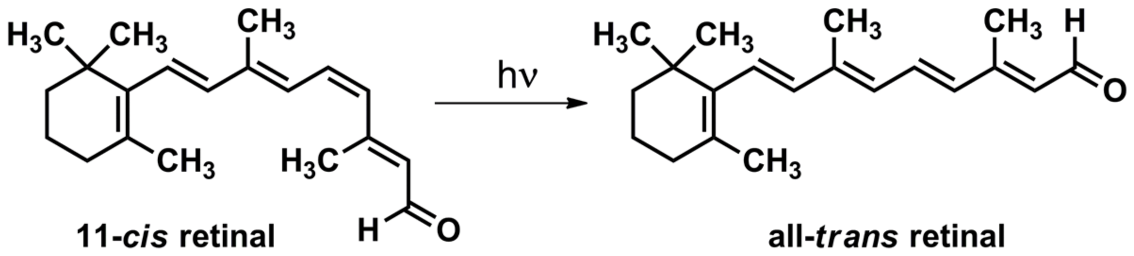


Figure 6. Schematic representation of light induced isomerization of 11-*cis* retinal to all-*trans* retinal.

Eventually, the all-*trans* retinal is metabolized by several enzymes, including ALDH1A3, which catalyzes its conversion to retinoic acid in retinal tissues, while it might be regenerated to the 11-*cis* retinal form, which is thus suitable to participate in a new visual cycle. Rhodopsin is synthesized and folded in the ER in the IS of the rod cells and passes to the Golgi membranes where it becomes glycosylated and palmitoylated. Rhodopsin-containing vesicles are then transported by specific carriers on the disc membrane into the OS of rod cells, where it is densely packed [74,75]. Indeed, the OS of rod cells has a high concentration of rhodopsin (~4 mM) for efficient photon capture [76]. Although mechanisms to achieve such a high concentration are mostly unclear [77], atomic force microscopy observations supported by biophysical studies have shown that WT rhodopsin can organize itself into multimeric superstructures, suggesting a role of these multimeric structures in the high packing density of the membrane discs [78]. The C-terminal tail of rhodopsin has been shown to play a key role in trafficking from the Golgi to the OS by directly binding the small GTPase ARF4 via the VAPA-COOH motif. ARF4 localizes in the trans-Golgi network where post-Golgi vesicles carrying rhodopsin are assembled. The binding between the VAPA-COOH motif and ARF4 is a crucial step for rhodopsin trafficking because it allows the recruitment of a complex of transporter components, which directs the targeting of rhodopsin toward the cilium [2–4].

1.5. Classification of rhodopsin RP mutant

More than 100 rhodopsin mutants have been identified in RP, including those affecting the binding of retinal, as well as structural and functional mutations that occur in sites that do not form the retinal binding pocket [75,79]. There is a high level of regulation in rhodopsin synthesis and degradation, and several mutations cause misfolding or aggregation [1]. Furthermore, some mutants that impair rhodopsin transport to the OS of rod cells or that result in constitutive activation of transducin have been described [64]. In general, rhodopsin point mutations are localized in different regions of the apoprotein opsin. As reviewed by Athanasiou et al., 2018 [75], rhodopsin mutants have been classified in seven different classes based on their effect on protein structure or functionality:

i) mutation affecting post-Golgi trafficking and outer segment targeting;
ii) mutation that causes misfolding, ER retention and instability;
iii) mutation causing disrupted vesicular traffic and endocytosis;
iv) mutation involved in altered post-translational modifications and reduced stability;
v) mutation that causes altered transducin activation;
vi) mutations involved in constitutive activation;
vii) mutation causing dimerization deficiency.

Finally, many other mutations have been identified that do not result in pathological phenotypes or that have not been thoroughly studied (Figure 7).

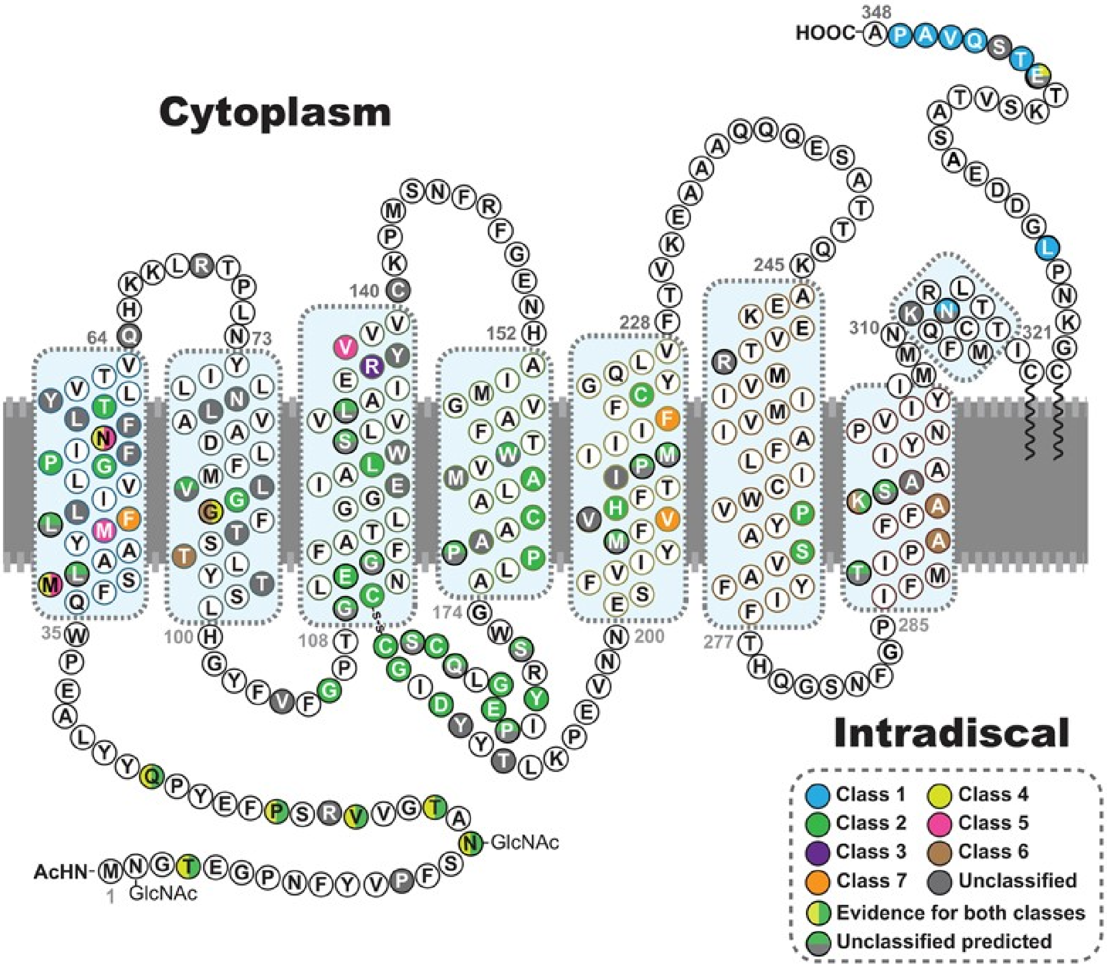


Figure 7. Rhodopsin structure and mutations causing RP. The picture is from Athanasiou et al, 2018 [75].

1.5.1. Class-1 mutant rhodopsin: mistrafficking and mislocalization

Rhodopsin mutations belonging to this class are focused on the C-terminal tail and are mechanistically linked to mistrafficking of folded rhodopsin from the Golgi to the rod OS, which is the final location where rhodopsin carries out its activity in the vision process [3]. Specifically, class-1 mutations are associated to the most severe forms of RP [80]. Many studies carried out on different animal models have shown that the VAPA-COOH motif is crucial for ciliary transport of rhodopsin as it directly binds to the specific transported ARF4, essential for the assembly of the transporting vesicles, and that mutations or truncations of this sequence leads to the accumulation of rhodopsin in the IS of rod cells, leading to photoreceptor degeneration [81–85].

1.5.2. Class-2 mutant rhodopsin: misfolding and structure instability

Class-2 rhodopsin mutant are mostly localized in the N-terminal tail of rhodopsin and in the TMD. Crystallographic studies have highlighted the importance of the N-terminal tail in maintaining the structural stability of rhodopsin by binding the EL2. Changes in the correct network of interactions between these loops can result in misfolded forms of rhodopsin that lead to relatively mild form of RP endowed with a slow progression over time in most of the cases. Overexpression of these mutants leads to the activation of unfolded protein response (UPR), with aggregation and accumulation in the ER with consequent cellular degradation [86–88]. Mutations that affect the TMD can also result in a class-2 pathological phenotype. A particular case of pathological TMD mutation is the modification of K296, which is the crucial anchor point for the covalent binding of the 11-*cis* retinal cofactor. Experimental evidence has shown that this mutation has a divergent pathological profile that overlaps with different classes of mutants. In fact, it causes misfolding with retention in the ER (class-2) [89], has a role in altering the mechanisms that manage the binding to arrestin (class-3) [90] and finally can promote constitutive activation of transducin (class-6) [91].

1.5.3. Class-3 mutant rhodopsin: disrupted vesicular trafficking and endocytosis

Similar to C-terminal mutations, class-3 rhodopsin mutants can affect vesicular transport, and are associated with fast progressing forms of RP [92]. Although pathological mechanisms remain unknown due to the difficulty of expression of class-3 mutants in animal models, it has been observed that some of them, specifically R135L, result in a constitutive phosphorylation that allows a high affinity form arrestin. This interaction causes a deregulation of vesicular and endocytic transport pathways which results in a lack of rhodopsin transport from the OS [93,94].

1.5.4. Class-4 mutant rhodopsin: altered post-translational modifications and reduced stability

Pathological mutations belonging to class-4 affect the glycosylation and phosphorylation sites on which the post-translational modifications are known to occur. In particular, the glycosylation of residues N2, T4, N15 and T17, which are found in the N-terminal domain, is essential for the correct protein folding, and T4K, N15S and T17M mutants have been associated with RP [61]. Phosphorylation of photoactivated rhodopsin through rhodopsin kinase enzyme occurs on specific site in the C-terminal domain and promotes rhodopsin-arrestin interaction. Mutations of arginine at position 135 and threonine at position 342 are also associated with RP [62].

1.5.5. Class-5 mutant rhodopsin: altered transducin activation

This class of mutants have altered stoichiometric balance of the different proteins involved in the phototransduction biochemical reactions. The M44T and V137M mutations are associated with RP as they have a high rate of transducin activation that alters the signaling transduction pathway equilibria [95]. However, some studies have shown that these mutants are not active in the dark or in the absence of the 11-*cis* retinal chromophore, suggesting that their pathological mechanism does not involve constitutive activation [90].

1.5.6. Class-6 mutant rhodopsin: constitutive activation

Unlike class-5 mutants, class-6 rhodopsin mutants are constitutively activated both in the dark and in the absence of the retinal cofactor, suggesting a potential role in the cell death in RP. Among them, mutations of glycine at position 90 in aspartate and valine (G90D and G90V) have shown to have this pathological mechanism. *In vitro* studies showed that the G90D mutant activates the visual cascade in the dark or without the chromophore [96]. Constitutive activity of rhodopsin was also observed in the G90D transgenic mice and frogs, acting as in the presence of continuous background light [97]. G90V mutation was also reported to be associated with RP, decreasing the thermal stability of the protein and partially overlapping with class-4 mutants [98]. Regarding the K296E mutant, its pathological profile is currently unclear. This mutant was proposed to cause RP due to its ability to sequester cellular components needed for the shut-off process, such as arrestin, similar to class-3 mutants, rather than its constitutive activity [91]. However, other reports suggest that K296E is also prone to misfolding and be retained in the ER, similar to class-2 mutants [89,99].

1.5.7. Class-7 mutant rhodopsin: dimerization deficiency

Although the role of multimeric organization of rhodopsin is poorly understood, class-7 mutations may affect multimerization [100,101]. Specifically, fully functional F45L, V209M and F220C mutant, identified in patients with clinical manifestations of RP, were tested by scramblase activity studies and it was demonstrated that while WT rhodopsin functionally reconstitutes into liposomes as dimers or multimers, F45L, V209M and F220C mutant rhodopsins behave as monomers. Although their pathological role it is not clear yet, it was suggested that these mutants could be potentially associated to RP due to their inability to dimerize [102].

1.6. Aldehyde dehydrogenase 1A3

1.6.1. Structural and biochemical features of ALDH1A3

In humans, during numerous physiological processes, aldehydes are generated from a wide variety of precursors, including the biotransformation of most compounds such as amino acids, neurotransmitters, carbohydrates, and lipids. In addition, aldehydes can be taken with the diet or in the form of environmental pollutants or external contaminants. These exogenous and endogenous compounds are tightly related with an increase of the oxidative stress and with DNA damage in cells, causing most of the time cytotoxicity and inflammation [103]. The ALDHs superfamily consists of 19 different isoenzymes involved in the irreversible detoxification of aldehydes to the respective carboxylic acid, using $NAD(P)^+$ as a cofactor [104]. The different ALDHs enzymes act in various metabolic pathways, including detoxification of aldehydes derived from lipid peroxidation, ethanol consumption and the metabolism of several neurotransmitters [105,106]. Moreover, ALDHs have an essential role in the biosynthesis of key metabolic regulators of cellular homeostasis, such as retinoic acid, γ-aminobutyric acid, dopamine and betaine [107]. In particular, the ALDH1A subfamily is composed by three different isoforms (ALDH1A1, ALDH1A2, and ALDH1A3) which are involved in the conversion of retinal, an important mediator of vision, to retinoic acid, a potent tissue differentiation factor for cellular development [108]. Indeed, the retinoic acid signaling pathway utilizes two classes of retinoid receptors, RARs and RXRs that belong to nuclear hormone receptor family. These proteins are ligand-regulated transcription factors that bind 9-*cis* (RXR, RAR) or all-*trans* (RAR) retinoic acid via a ligand-binding domain, and direct the transcription of target genes via a DNA-binding domain [109]. Retinoic acid is derived from vitamin A (retinol) and its pleiotropic effects include spinal cord and retina development during embryogenesis, neuronal cell differentiation and maintenance of epithelial cell type in adult tissues [110]. Synthesis of retinoic acid proceeds through two steps: i) the oxidation of retinol by alcohol dehydrogenase producing the relative aldehyde, followed by ii) its irreversible conversion to acid using retinal dehydrogenases [111].

The ALDH1A3 isozyme shares more than 70% sequence identity with ALDH1A1 and ALDH1A2, where a very high degree of conservation is found particularly in residues involved in catalysis. Although ALDH1A1, 1A2 and 1A3 recognize a common substrate, their expression pattern does not overlap entirely, and reflects their preference for specific substrates. Indeed, ALDH1A1 preferentially processes aldehydes produced by lipid peroxidation while retinal is the preferred substrate for both ALDH1A2 and ALDH1A3 [112]. In addition, murine ALDH1A3 processes *all*-trans retinal with a catalytic efficiency that is 10-fold higher than that of ALDH1A1 and ALDH1A2 [16]. ALDH1A3 is differentially activated during early embryonic head and forebrain development, and it is expressed at high levels in the differentiating keratinocytes of human and murine hair shafts [113,114]. Moreover, knockout of the murine *Aldh1a3* gene was associated with perinatal lethality that could be rescued by maternal treatment with retinoic acid [115]. In the rod photoreceptors cells, retinoic acid levels are finely controlled by a series of enzymes, including ALDH1A3, which catalyze the oxidation of the all-*trans* retinal to retinoic acid and for which the most efficient retinal dehydrogenase activity was reported

[16]. The role of retinoic acid signaling in eye development is well established, and there is a genetic evidence of a direct link between retinoic-acid-synthesis dysfunction and early-eye-development anomalies in humans [116,117]. Furthermore, some studies have shown a correlation between the enzymatic activity of ALDHs and retinal degenerative disease, including RP [118,119]. Although its role in retinal diseases is still largely unclear, it was suggested that the enzymatic activity of ALDH1A3 may be correlated to the concentration of the physiological retinal cofactor, and may have role in those retinal dystrophies, including RP, whose molecular basis are related to retinal concentration.

From a structural standpoint, ALDH1A3 shares a similar arrangement with other members of the ALDHs family. Its structure is assembled as a tetramer (dimer of dimer) in which each monomer functions autonomously by oxidizing a substrate molecule into its corresponding acid by reducing the NAD$^+$. Each monomer is arranged in three functional domains: the N-terminal domain containing the NAD$^+$ binding site, which is arranged in five-stranded parallel β-sheet; the catalytic site, which is arranged in six-stranded parallel β-sheet, and the oligomerization domain that consists of three stranded antiparallel β-sheet (Figure 8) [120].

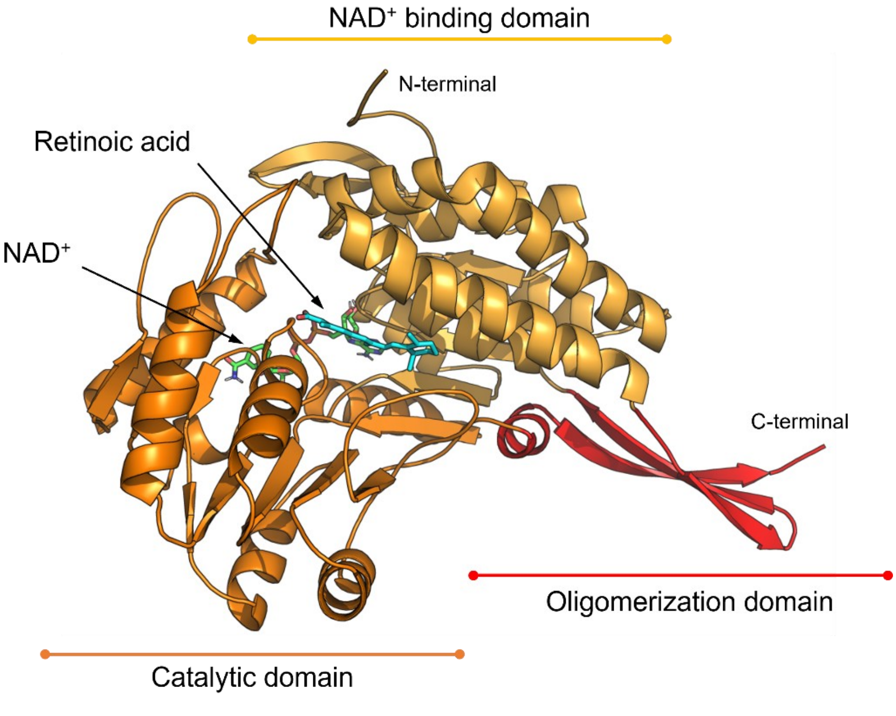


Figure 8. Graphical representation of the ALDH1A3 monomeric structure in complex with retinoic acid (cyan sticks) and NAD$^+$ (green sticks) (PDB-ID: 5FHZ) [120]. The N- terminal NAD$^+$ binding domain is represented by light orange cartoon; the catalytic domain is represented by orange cartoon and the C-terminal oligomerization domain is represented by red cartoon.

2. Identification of small molecular chaperones binding P23H mutant rhodopsin through *in silico* virtual screening

2.1. Background and rationale of the work

Rhodopsin is a GPCR membrane protein that is activated by light of different wavelengths in organisms ranging from single-celled eubacteria and archaea to humans. In vertebrates, rhodopsin is responsible for vision, a process activated by light-induced photoisomerization of retinal chromophore from the 11-*cis* form to the all-*trans* isomer. In this process, the retinal cofactor is linked to the K296 residue of opsin through a protonated Schiff base. Photoisomerization induces a number of conformational modifications to rhodopsin that finally activate the visual transduction cascade. The correct folding and localization of rhodopsin in the OS of photoreceptor cells are essential to this physiological process. In this context, rhodopsin mutations are the most common cause of inherited ocular diseases, including RP. The P23H mutation in the N-terminal tail of rhodopsin causes structural destabilization and impairs the correct protein folding, accounting for the most of RP cases with prevalence in North America (∼10% of RP cases) [19]. In addition, incorrect protein folding can be also caused by low cellular levels of 11-*cis* retinal, which fails in its function as a molecular chaperone. Indeed, when rhodopsin is not properly folded, it is poorly translocated to the OS, and it is accumulated on the ER where it causes aggregation and cellular degradation [86–88]. Although the precise mechanisms leading to photoreceptor degradation and disease development are yet to be elucidated, recent *in vitro* studies showed that the correct folding and trafficking of P23H mutant rhodopsin can be partially rescued by small chaperone ligands, able to bind rhodopsin in the 11-*cis* retinal binding site and to restore the putative core stability and the correct receptor localization. Thanks to these studies, P23H rhodopsin is currently considered as a profitable and validated target for pharmacological treatment by small chaperone ligand approach [1]. Unfortunately, the main limitation in the study of the P23H mutant is represented by the lack of high-resolution structural details due to the high instability of the mutant protein and by the experimental difficulties of studying the events associated with the transduction of the light signal [121]. In this work, these limitations have been addressed using computational modeling tools. Specifically, the structure of P23H mutant in complex with 11-*cis* retinal was generated by homology modeling and its conformational behavior was studied throughout MD simulations. Then, the representative structure of P23H mutant rhodopsin identified by cluster analysis was used as a rigid receptor in the structure-based virtual screening of large chemical libraries of compounds to identify new small molecules that could bind within the retinal binding site of P23H mutant rhodopsin. These molecules are expected to act as a pharmacological chaperone in restoring rhodopsin folding due to mutation causing disease (RP) or due to a reduced level of endogenous retinal cofactor, paving the way to a further development of novel therapeutical strategies against RP.

2.2. Results and discussion

2.2.1. Homology modeling of P23H mutant rhodopsin receptor

Given the lack of structural details, the full-length 3D structure of the P23H rhodopsin mutant was reconstructed by carrying out a homology model. The high-resolution crystallographic structure of the bovine full-length rhodopsin (Protein Data Bank [122] accession code PDB-ID: 1U19 [60]) was selected as a template for homology modelling and MD refinement. Analysis of the all-atom root mean square deviation (RMSD) of the P23H mutant rhodopsin model along 500 ns MD trajectory confirmed that the system is stable during MD simulation (Figure 9A). Cluster analysis of MD trajectories was then used to identify the centroid frame of the most populated cluster, which was finally used as a rigid receptor in the subsequent virtual screening. Although the P23H mutation is relatively distant from the binding site of 11-*cis* retinal (approximately 16 Å) (Figure 9B), by comparing the crystallographic structure of rhodopsin WT (PDB-ID 1U19) and the representative MD frame of the P23H mutant rhodopsin, it is evident that mutation P23H can affect the structural conformation of the TMD, and the conformational arrangement of residue involved in the binding of 11-*cis* retinal (Figure 9C).

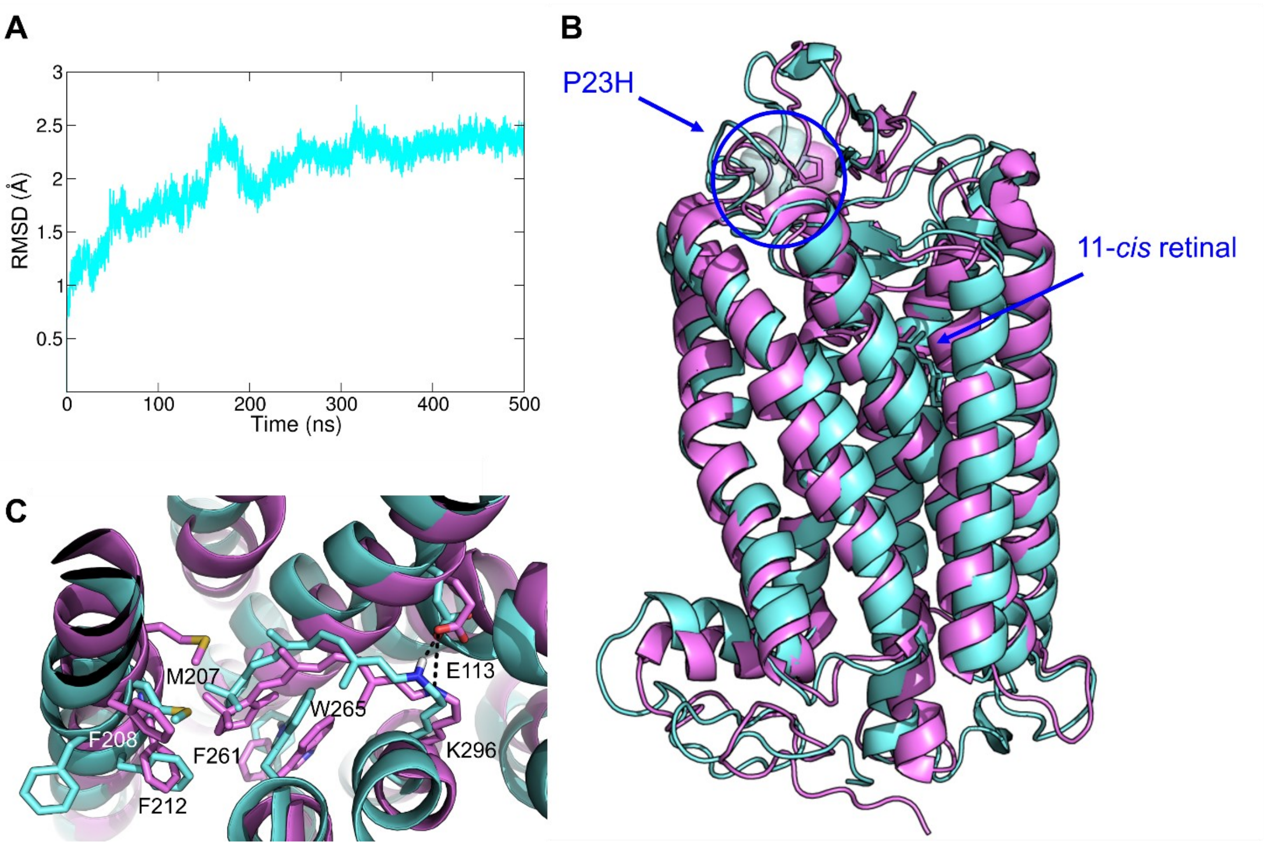


Figure 9. All-atom RMSD of the P23H mutant rhodopsin in complex with the 11-*cis* retinal (A). Structural superposition of crystallographic structures of WT rhodopsin coded by PDB-ID 1U19 (pink) and P23H mutant rhodopsin (cyan) depicted in front view (B) and close-up view of the 11-*cis* retinal binding site (C). Protein structures are aligned and showed in cartoon. The 11-*cis* retinal cofactor and the residues of interest are represented as sticks, while polar interactions are represented by the black dashed lines.

2.2.2. Docking-based virtual screening

In this approach, small molecular chaperones were designed to bind the 11-*cis* retinal binding site of the P23H mutant rhodopsin to restore the correct protein folding. Specifically, these compounds should compete with the natural substrate that binds with poor affinity in pathological mutants due to conformational differences that modify the binding site structure. As shown by *in vitro* studies, the conformational changes induced by molecular chaperones could restore the functionality of mutant rhodopsin and its trafficking to the OS [1]. To identify new small molecular chaperones, a structure-based virtual screening was carried out. Different libraries of compound were selected to this aim, including: i) a high diversity in house library of 867 natural products, ii) the DrugBank database (2509 approved compounds) [6] and iii) the commercially available MolPort database (7934460 compounds) [7]. Natural products are a unique and privileged source of hit/lead compounds with a wide chemical diversity as well as are tools for understanding chemical biology of target proteins. In this context, natural products are largely used in virtual screening and drug design approaches carried out by our research group, showing pharmaceutical importance by providing lead compounds that, in some cases, have been patented as antiviral or anti-cancer drug candidates [5,123–128]. The DrugBank database was selected with the purpose of repositioning approved drug, which is a widely used strategy that allows the therapeutic repositioning of a drug for a therapeutic indication that is different from that of its approval. The advantage of this strategy, when successful, is that it reduces the time and costs associated with the development of a new molecule, since at least the toxicological and safety studies have already been completed and approved. Finally, the MolPort database is a rich source of commercially available diverse structures. While screening libraries i) and ii) required a relatively low computation cost, screening of the MolPort database containing around 8 million compounds required much more time and resources to be completed. Based on the nature of the physiological substrate 11-*cis* retinal that reacts with the K296 residue of opsin through the aldehyde group, giving rise to the protonated Schiff base, the MolPort database was first preliminarily filtered to identify compounds bearing up to three aldehyde groups. The remaining non-aldehyde compounds were further filtered based on drug-like molecular descriptors common to GPCRs binder compounds (Table 1). In total, this operation led to the generation of two sub-libraries, containing 3195 aldehyde compounds and 5807381 GPCRs binder-like compounds.

Table 1. GPCRs binder-like molecular properties [129].

	Minimum value	Maximum value
Molecular mass	250	450
Rotatable bonds	0	10
Hydrogen Bond donors	0	5
Hydrogen Bond acceptors	0	10
ClogP	-1.5	5.0
Polar Surface Area (Å^2)	0	150
Chiral centers	0	2

Compounds were docked within the in the 11-*cis* retinal binding site of P23H rhodopsin. The validity of the docking/scoring function was preliminarily assessed by covalently and non-covalently re-docking the 11-*cis* retinal in its binding site. Binding poses were highly similar to each other: according to the covalent docking pose, when non-covalently docked, the oxygen atom from the aldehyde group is oriented toward the K296. In both cases, hydrophobic interaction of the ionone moiety with residues M207, F212, F261, and W265 were maintained (Figure 10A and 10B).

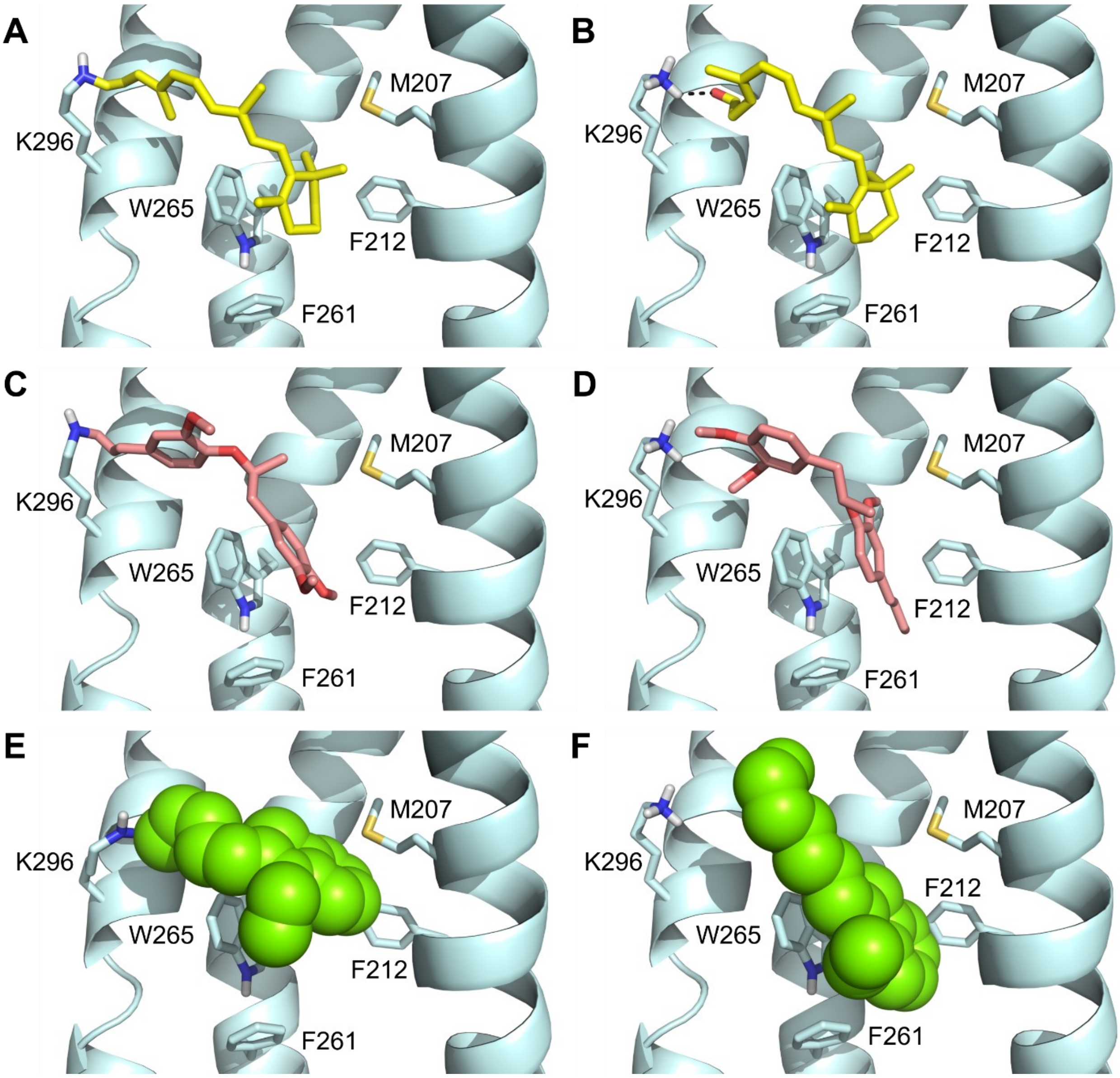


Figure 10. Binding poses of the 11-*cis* retinal cofactor (yellow), the MolPort 019-937-085 compound (pink) and compound A3 (green) within the binding site of the P23H rhodopsin mutant, predicted by covalent docking (A, C and E) and non-covalent docking (B, D and F). The 11-*cis* retinal and the MolPort 019-937-085 were used as positive and negative reference, respectively.

Docked molecules were sorted according to their scoring value, and the 100 top-ranked molecules from each library were visually inspected. It is worth noting that among docked compounds the best and worst molecule have a score value of -9.60 and -5.54 respectively, while the retinal has a -8.10 of score value. However, due to the lack of a training set to refer to for data comparison, the score was not considered as a parameter during the compound selection process. Therefore, compounds selection was based on a combination of factors such as the docking pose of the molecules within the binding site and the hydrophobic and polar interactions with the receptor. Specifically, compounds from the GPCR-binders library, the DrugBank, and the natural products library were docked through a non-covalent approach, while aldehyde compounds were selected by comparing the non-covalent docking pose, which simulates the protein-ligand recognition, and the covalent docking pose, which simulates the complex after the formation of the covalent bond with K296 following the recognition step. In this context, similar to 11-*cis* retinal, compounds for which non-covalent and covalent poses had a similar orientation were selected, while compounds whose non-covalent and covalent poses were found to be oriented differently were discarded. For example, the compound with MolPort accession code 019-937-085 showed a totally different orientation between the two docking poses (Figure 10C and 10D). Indeed, in covalent docking, the aldehyde group binds to the K296, with the dimethoxybenzene moiety being oriented toward helices 6 and 5. In contrast, in non-covalent docking pose, the dimethoxybenzene moiety is oriented toward the K296 residue and the aldehyde group toward the two helices. The covalent linkage involves a molecular geometry constraint that forces the molecule to adopt a certain orientation and to form a Schiff base by reacting with the side chain of K296, which may not be reproduced in the absence of the covalent geometric constraint. Given the different orientation of its docking poses, the MolPort 019-937-085 compound was used as a negative reference, while the 11-*cis* retinal was used as a positive reference to select the molecules to be experimentally investigated. Although the chemical structure of selected compounds will not be shown for confidentiality reasons, as they have not yet been published, the binding pose of compound A3 is shown in Figure 10 E and 10F. This compound binds within the retinal binding site establishing the same interactions as the natural substrate and maintaining the same orientation with both the covalent and non-covalent docking approaches, which makes it suitable for experimental validation. Finally, following the process described above, 7 aldehydes and 11 natural products were selected for experimental validation. Unfortunately, no molecules from the DrugBank database exhibited an interesting profile in binding to rhodopsin *in silico*, thus none of the drugs included in this database was selected for further investigations.

The experimental validation of selected compounds was entrusted to our collaborators at Cardiff University. Unfortunately, the lack of resources and organizational difficulties caused by the COVID-

19 pandemic delayed the work, and forced us to focus our efforts on natural products and aldehydes compounds that could covalently bind rhodopsin. In contrast, experimental validation of compounds selected from the MolPort database was deemed with a lower priority, also considering that a few non-covalent binders of opsin are already reported in the literature [1,130]. Nevertheless, these molecules are currently in the testing pipeline and will be investigated as soon as possible.

2.2.3. Single cell fluorescent microscopy – Preliminary results

To evaluate the localization of P23H human rhodopsin on cell membrane, the single cell fluorescent microscopy test was carried out by screening selected compounds NP2-NP13, A1-A7 and LI65-04 on U2OS cells transiently transfected with hRHO bearing the pathological P23H mutation. Rhodopsin trafficked to the cell membrane was detected with the anti-rhodopsin antibody RetP1, which recognize an extracellular epitope. Compounds were tested at a final concentration of 10 µM in 0.1% DMSO. The 9-*cis* retinal, b-ionone and 0.1% DMSO, were selected as positive and negative controls, respectively. The 9-*cis* retinal was used instead of 11-*cis* retinal because of its increased stability. Indeed, the 9-*cis* retinal stimulates the expression of rhodopsin uniformly throughout the cell membrane, while the β-ionone does not show a significant increase in fluorescence on the membrane in the first repeat (Figure 11).

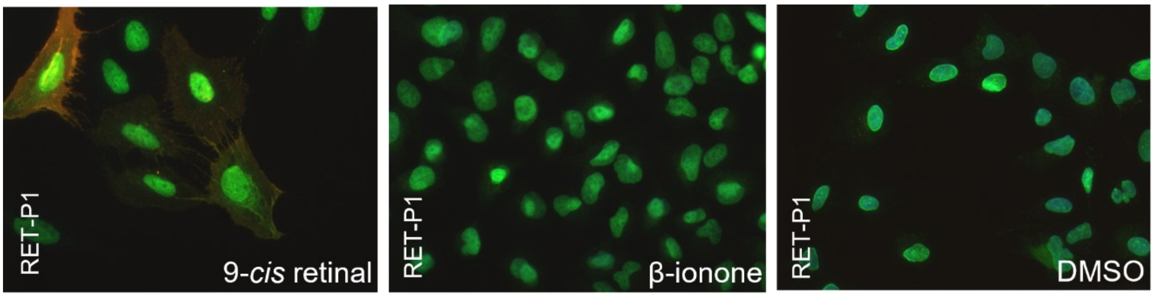


Figure 11. Fluorescent microscopy imaging of subcellular localization of hRHO P23H in presence of controls compounds 9-*cis* retinal, b-ionone and 0.1% DMSO.

Preliminary data show that, similar to the 9-*cis* retinal, compound A3 promote opsin trafficking to the cell membrane with a homogeneous distribution of the antibody across the surface of the plasma membrane. Interestingly, compound A3, showing the higher chaperone activity, is an aliphatic aldehyde with an aromatic moiety, allowing good interaction with the receptor. The same was noted in two of the images obtained from the NP9 compound, but in a milder way than in A3. It is worth noting that natural products may not have a direct structural correlation, but they belong to different chemical classes. On the other hand, compounds A1, A2, NP2, NP4, NP7, and NP13 showed a particular cell morphology, although they do not exhibit a distinct signal like A3, seem to increase the fluorescence linked to RET-P1 (Figure 12). However, the signal was inhomogeneous and localized in the cytoplasmic compartment. Some compounds such as A3 and NP6, showed an unexpected staining for the red channel in the first repeat, which highlights the need of at least an additional repeat as well as cytotoxicity assays. It is worth noting that these tests are running by our collaborators at Cardiff University at the time of submission of this thesis work and are aimed at ascertaining whether these compounds exhibit effective activity, if they are cytotoxic, or if the fluorescence signal might be due to incorrect staining. Finally, mentioned compounds will be tested at a higher concentration (50 µM) to confirm their activity. In addition, their cytotoxicity will be evaluated since the morphology of some cells, especially for compounds NP2, NP7, NP11, NP12, and NP13, appeared altered after the treatment.

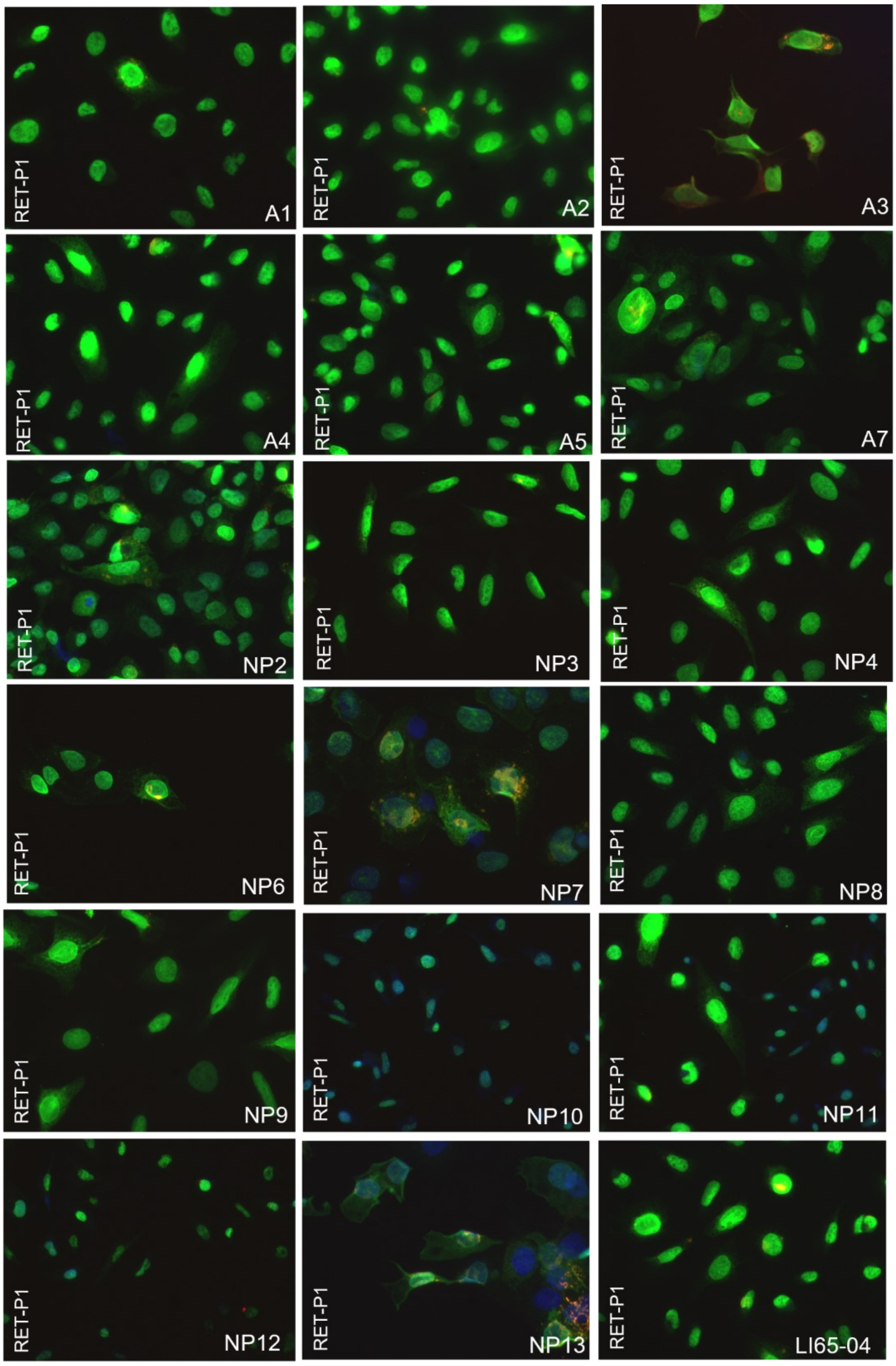


Figure 12. Fluorescent microscopy imaging of subcellular localization of hRHO P23H in presence of tested compounds.

2.3. Conclusions

In this work, a reliable computational model was developed with the aim of identifying small molecules that can possibly bind the P23H rhodopsin mutant and restore the correct protein folding and localization. In this context, molecular chaperones are an interesting perspective to deal with photoreceptors degeneration that characterizes ocular dystrophies, including RP. Computational techniques such as the homology model, MD simulations and docking-based virtual screening provide a fast and valuable tool to address the lack of structural details that prevent the in-depth study of rhodopsin mutants as well as the experimental screening of large chemical libraries. The docking-based virtual screening of aldehyde and natural product compounds led to the identification of 18 molecules belonging to different chemical classes that were selected for experimental investigations against the P23H mutant rhodopsin. Noteworthy, due to the peculiar structural features of the rhodopsin receptor, the screening of the DrugBank database did not provide any molecules with an interesting profile to be investigated *in vitro*. Furthermore, among GPCR binder-like compounds, selected molecules have not yet been tested for priority reasons. The fluorescent immunohistochemistry cell-based assay was carried out to evaluate the ability of selected compound to induce the correct folding and localization of rhodopsin on the membrane. Among them, compound A3 emerged as promising molecular chaperone and structural stabilizer of rhodopsin, representing a promising starting point for further investigation and optimization. In general, collected data strongly support the usefulness of molecular chaperones as a pharmacological approach for the treatment of RP whose molecular basis is opsin misfolding and mistrafficking.

2.4. Materials and methods

2.4.1. Homology modelling and MD simulation

Since the Protein Data Bank lack of 3D structure of the human P23H mutant rhodopsin, homology modelling was carried out. The X-ray structure of WT bovine rhodopsin (PDB-ID: 1U19) [60] was selected as a structural template based on its high resolution (2.20 Å), and the model was built using the Prime software in the Schrödinger Release 2019-1 [131]. MD simulation was carried out on the structural model of human P23H mutant rhodopsin in complex with the 11-*cis* retinal covalently bound to K296 by mean of a protonated Schiff base. The apoprotein opsin was parameterize using the ff14SB force field [132] while for the retinal cofactor, covalent bond parameters were partially retrieved from Ferrè et al. [133]. The system was then embedded in lipid bilayer of 138 molecules of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) [72] added by mean of the CHARMM-GUI membrane builder [134,135], solvated in a rectangular box of 14629 TIP3P water molecules, neutralized with the addition of counter-ions and finally relaxed through two steps of energy minimization. The first step involved only water molecules which were energy minimized for the first 1500 steps using the Steepest Descent algorithm (SD), and for the last 3500 steps using the Conjugate Gradient algorithm (CG). In the second step, energy minimization of the entire system was carried out for 10000 steps (1500 SD and 8500 CG). Langevin thermostat was used to heat the system from 0 to 300 K over 1 ns at constant volume, while the Berendsen barostat was used to control the density equilibration over 1 ns. Finally, the system was preliminary equilibrated for 50 ns and MD simulation was carried out to generate a trajectory of 500 ns at constant pressure. The MD simulation was carried out by the Amber18 program [136], and the MD trajectory analyzed by the CPPTRAJ software [137]. The hierarchical agglomerative (bottom-up) approach was used to run cluster analysis on MD frames. The representative MD frame of the most populated cluster was extrapolated from MD trajectory and eventually used as a rigid receptor for further virtual screening. Although an extended conformational space has been sampled, the cluster analysis shows that the selected structure is widely representative since the cluster leader has a population of 86% of the MD frames.

2.4.2. Docking-based virtual Screening

The docking-based virtual screening was carried out using three different libraries of compounds: an in house collection of 867 natural products, the Drug Bank database [6] and the commercially available MolPort database containing 7934460 compounds in April 2021[7]. Molecules were retrieved in SMILES format and filtered through the FILTER (Openeye) version 3.1.0.3 [138]. Compounds were converted in 3D coordinates with OMEGA (OpenEye) version 3.1.0.3 [139], their protonation state was assigned by QUACKPAC (OpenEye) version 2.0.0.3 [139], and then compounds were energy minimized by Szybki (OpenEye) version 1.10.0.3 [139], using the MMFF94S force field. Molecular docking simulations of the Drug Bank database and natural products were carried out with FRED (OpenEye) version 3.3.0.3 [140] using the Chemgauss4 fitness function with the high docking accuracy and keeping only the best pose for each compound. Aldehyde compounds were docked both covalently and non-covalently within the binding pocket of rhodopsin using Glide docking program from the Schrödinger Release 2019-1 [141]. The structure of the P23H rhodopsin mutant corresponding to the centroid of the most populated cluster extracted from MD simulations was used as a rigid receptor in molecular docking simulations.

2.4.3. Single cell fluorescent microscopy

Cells were plated on 13 mm polylysine-treated round glass coverslips and let to adhere for at least 4 h. Media was replaced with media containing 1 µg/mL of tetracycline and 5 µM of 9-*cis*-retinal or compound at a final concentration of 10 µM, not exceeding 0.1 % of DMSO. The 9-cis retinal, b-ionone and 0.1% DMSO, were selected as positive and negative controls, respectively. Cells were incubated overnight in the dark to allow the tetracycline-induced expression of rhodopsin. On the following day, under dim red light conditions, cells were washed twice in PBS and fixed for 25 min using a solution of methanol-free 4% paraformaldehyde, then washed in PBS before blocking 1 h with TBS blocking buffer (LI-COR, Lincoln, NE, USA) solution followed by incubation with RET-P1 antibody (Invitrogen, 1:500, Carlsbad, CA, USA) in TBS buffer and, subsequently, with anti-mouse IgG (H + L), F(ab') 2 Fragment Alexa Fluor ® 555 Conjugate (Cell Signalling Technologies, 1:1000, TBS buffer, Danvers, MA, USA) for 1 h. Coverslips were then carefully washed before permeabilizing cell membranes with a solution of Triton 0.1% for 20 min. Nuclei were stained with DAPI before mounting the coverslips on glass microscope slides. Immunofluorescence images were captured with an Olympus BX50 epifluorescence microscope, Southend-on-Sea, UK, (from 3 independent experiments). Gain and exposure times for 488 nm and 555 nm channels were maintained constant across all coverslips from the same experiment, which always included 9-*cis* retinal, b-ionone and DMSO as controls. Images were processed with ImageJ software: the contrast was enhanced in the 405 nm channel for clarity, whilst no manipulations were made to the channels relevant in the experiments (488 and 555 nm). RGB colors were assigned to each channel: channel 488 (green) corresponds to the total opsin staining (intracellular and on the surface), while channel 555 (red) detects the presence of opsin trafficked to the cell membrane. Finally, the blue (405) channel highlights the cell nucleus. Then, images were merged with the software in-built function.

3. Conformational insights into the C-terminal mutations of human Rhodopsin in *retinitis pigmentosa*

The data and information reported in this chapter were published in:

Picarazzi F, Manetti F, Marigo V, Mori M. Conformational insights into the C-terminal mutations of human rhodopsin in retinitis pigmentosa. *J Mol Graph Model*. 2022; 110:108076. https://doi.org/10.1016/j.jmgm.2021.108076

3.1. Background and rationale of the work

Rhodopsin is a light-sensitive transmembrane receptor involved in the visual transduction cascade. Among the several rhodopsin mutations related to RP, those affecting the C-terminal VAPA-COOH motif that is implicated in rhodopsin trafficking from the Golgi to the rod outer segment are notably associated with more aggressive RP forms. Several pathological mutations of P347 in the C-terminal tail of rhodopsin have been described, including P347L, P347S, P347R, P347T, P347Q, and P347A that originate from single-point mutations in the triplet of bases codifying for P347 in the *RHO* gene [142–147]. Among them, the P347L is the most prevalent in Europe [148]. Most notably, patients bearing the P347L rhodopsin mutant show a more severe phenotype of RP, not only with respect to other mutations at position P347, but also compared to mutations on different portions of rhodopsin including the retinal binding site, which points to the crucial role of the rhodopsin C-terminal tail integrity in the visual cascade [149]. The N- and C-terminal tails of GPCRs and the loops connecting transmembrane helices are crucially involved in the control of numerous cellular functions by mediating the recognition of extracellular ligands and initiating the intracellular signaling cascade [150–152]. Unfortunately, structural studies often fail to resolve the structural details of these GPCR regions, which are also known as intrinsically disordered protein regions (IDPs) due to their particularly dynamic and disordered nature, encoded by their sequence [56,153]. In the specific case of rhodopsin, recent studies underscored the crucial role of the last four amino acids of the C-terminal tail characterized by the VAPA-COOH motif, whose deficiency or mutations are linked to rod degeneration and blindness via rhodopsin mistrafficking and mislocalization [81,154,155]. The VAPA-COOH motif is known to interact with the GTPase ARF4, a small guanine nucleotide-binding protein that plays a pivotal role in rhodopsin trafficking. In particular, the α-helix 3 of ARF4 has been reported to be potentially involved in the direct interaction with rhodopsin C-terminal tail [2]. Although molecular details of ARF4/rhodopsin interaction and the impact of C-terminal mutations on rhodopsin structure are still unknown, modulation of this mistrafficking mechanism would be an attractive way to impair rhodopsin toxicity caused by mislocalization [4]. In this context, chemical and physicochemical features of clinically relevant P347 mutants do not directly correlate with their pathological role (Table 2).

Table 2. Main physicochemical descriptors of proline (WT) and clinically relevant mutant amino acids at position 347.

Amino acid	Hydropathy index[a]	Net charge at pH 7.4[b]	Molecular mass	Isotropic surface area[c]	van der Waals volume (nm^3)[d]	Hydrophobicity[d]	Dipole moment (μ, D)[e]	Hydrogen donor atoms[f]	Hydrogen acceptor atoms[f]
P (WT)	-1.6	0	115.132	122.35	0.2279	0.98	2.17	–	–
L	3.8	0	131.175	154.35	0.37876	2.32	1.66	–	–
S	-0.8	0	105.093	19.75	0.09204	-0.05	3.7	OG	OG (2)
R	-4.5	+1	174.203	52.98	0.58946	-1.37	2.37	NE, NH1 (2), NH2 (2)	–
T	-0.7	0	119.119	59.44	0.19341	0.35	2.46	OG1	OG1 (2)
Q	-3.5	0	146.146	19.53	0.34861	-0.30	2.84	NE2 (2)	OE1 (2)
A	1.8	0	89.094	62.90	0.05702	0.42	2.38	–	–

[a] According to the Kyte and Doolittle scale [156].

[b] (Vella, 1998) [157]

[c] (Collantes and Dunn, 1995) [158]

[d] (Lin et al., 2008) [159]

[e] (Kereselidze et al., 2018) [160]

[f] Hydrogen donor and acceptor atoms in amino acid side chains [161].

Whether the increased steric hindrance of P347R and P347Q mutants might impair the interaction of rhodopsin C-terminal tails with ARF4 or other carrier proteins, this hypothesis is disproved by the P347S and P347T mutants, which however bear polar side chains compared to the WT P347, as well as by P347A and P347L which share a similar shape and lipophilicity with the WT proline. Anyway, the investigation of the steric and polarity effects of P347 mutants would require robust molecular details of mutant rhodopsin C-terminal tail as well as rhodopsin/ARF4 interaction, which are currently unavailable. In contrast, proline is the only natural amino acid that forms a five-member ring connecting the side chain with the backbone, which is known to possibly induce conformational restraints [162,163]. Accordingly, structural changes that occur on the VAPA-COOH motifs at the C-terminal tail of rhodopsin upon mutation of P347 could impair rhodopsin binding to ARF4, which results in a lower or incorrect mistrafficking to the rod OS.

Driven by this evidence, in this work we applied computational structural methods such as homology modeling and extensive MD simulations to provide conformational insights into the role of clinically relevant P347 rhodopsin mutants in RP. Since no drugs for the therapy of RP caused by C-terminal rhodopsin mutations are available or in advanced development stages, structural details disclosed in this work might promote further investigations aimed at the understanding of the mechanisms governing rhodopsin mistrafficking as well as the identification of innovative therapeutic solutions for RP.

3.2. Results and discussion

3.2.1. MD simulations of rhodopsin full-length systems

Due to the lack of structural information on the human full-length rhodopsin (i.e., opsin covalently bound to 11-*cis* retinal), the high-resolution crystallographic structure of the bovine full-length rhodopsin (PDB-ID: 1U19) [60] was selected as a template for homology modelling and MD refinement. Notably, at the time of this work, the selected template was the only structure of rhodopsin in which structural details of the C-terminal tail were solved. Similarly, the structure of human opsin was not available from structural biology studies, whereas bovine native opsin was characterized by X-ray crystallography (PDB-ID: 3CAP) [67], although details of the C-terminal tail were unavailable. Therefore, the human full-length opsin was generated by removing the retinal chromophore from the rhodopsin model described above, and relaxing the resulting rough structure by MD simulations. A representative structure was extracted from MD trajectories by cluster analysis, which groups MD frames based on structural similarity and distance metrics using a specific clustering algorithm. The most representative cluster of opsin was populated by 50.8% of MD frames, while that of rhodopsin had a frame population of 75.4%. Each cluster is represented by a centroid frame, and the centroid of the most populated cluster was used in this work for structural analysis and graphic representations. Specifically, in the case of rhodopsin, structural alignment of the representative frame showed a good matching with the starting crystallographic structure, particularly within the TMD, giving a Root Mean Square Deviation (RMSD) of ~ 1.2 Å for the superposition of backbone Cα atoms. Similarly, the representative MD frame of opsin nicely overlap with the experimental crystallographic structure of native opsin, particularly within the TMD, with an RMSD of ~ 1.5 Å for the superposition of backbone Cα atoms. By comparing the structures of opsin and rhodopsin, it is evident that the conformation adopted by TMD is dependent on the presence or the absence of the retinal cofactor (Figure 13). In this respect, the computational protocol used herein can be considered reliable and robust in reproducing the main structural features of opsin and rhodopsin, including the conformational changes due to the binding of retinal within the TMD as known from experiments.

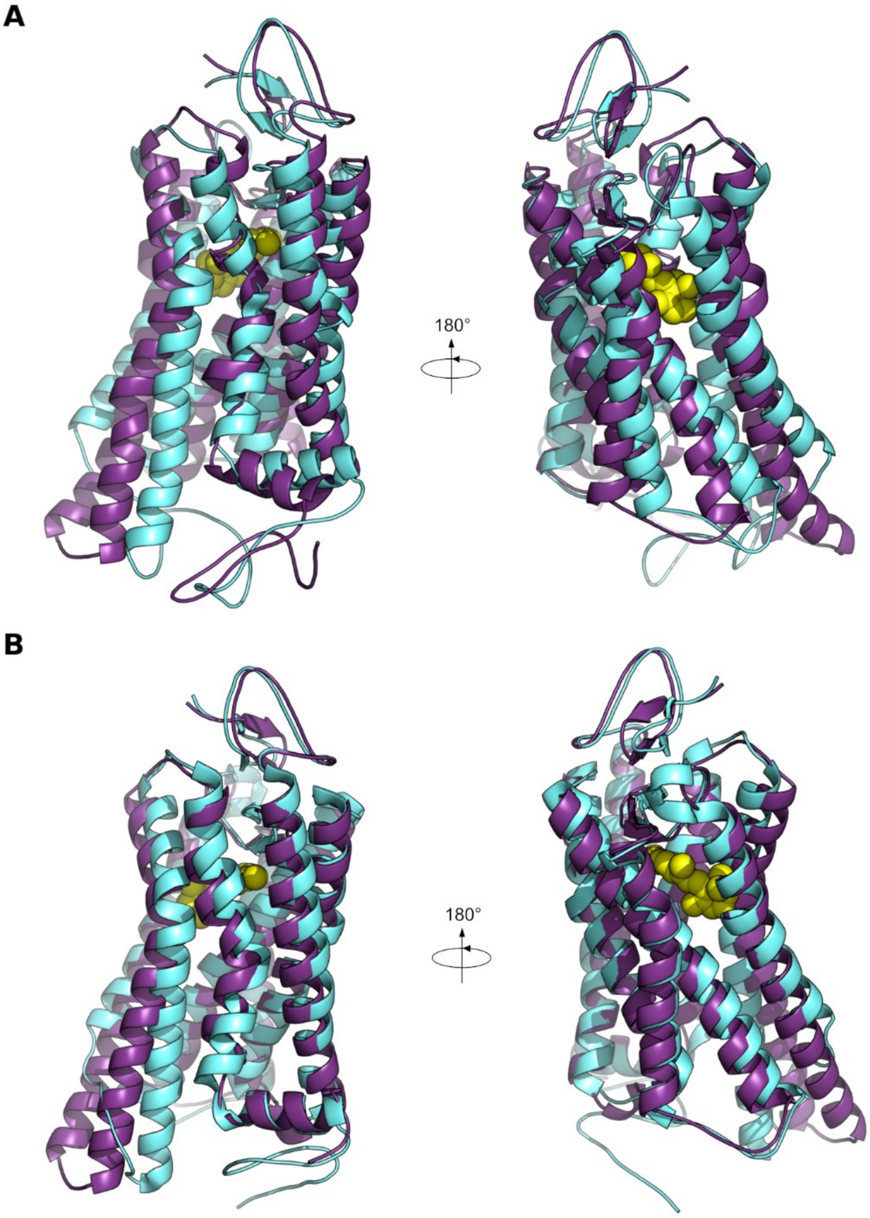


Figure 13. Front and rear view of MD (A) and crystallographic (B) structures of opsin (purple) and rhodopsin (cyan). The representative frames extracted from MD trajectories are showed in A), while

X-ray crystallographic structures coded by PDB-IDs 3CAP and 1U19 are showed in B). Protein structures are aligned, and showed in cartoon. The retinal cofactor is represented as yellow spheres.

3.2.2. Analysis of WT and mutant VAPA-COOH motifs in opsin/rhodopsin

Since in this work we focused on clinically relevant rhodopsin mutants in position P347 of the C-terminal tail, further information on the conformational features that characterize the VAPA-COOH motif of WT rhodopsin and opsin was obtained by measuring the distance between the Cα atoms of V345 and A348 in MD trajectories, the first and last amino acid of the VAPA-COOH motif, respectively. Both systems showed similar values of the maximum, minimum, and average distance ($\text{Max}_{\text{distance}}$, $\text{Min}_{\text{distance}}$ and $\text{Mean}_{\text{distance}}$, respectively), as well as an identical value of the Δ_{distance} (calculated as the difference between $\text{Max}_{\text{distance}}$ and $\text{Min}_{\text{distance}}$, Table 3). Moreover, the distance between the considered atoms was constant over the simulation time (Figure 14A) and the most representative structures extrapolated from the MD trajectories of opsin and rhodopsin nicely resulted in very similar conformations of the VAPA-COOH near the $\text{Mean}_{\text{distance}}$ value (Figure 14B). The distribution of dihedral angles Phi (Φ) and Psi (Ψ) of the three residues V345, A346, and P347 of the VAPA-COOH motif was computed along MD trajectories and plotted in a Ramachandran-like graph (to be noted that the C-terminal residue A348 was not accounted for in this analysis due to the lack of the Ψ dihedral angle). Results of dihedral angles distribution (Figure 15) showed that the backbone of monitored residues explores the same conformational space in rhodopsin and opsin. Coupling this result with the distance analysis of Figure 14, clearly suggests that the backbone conformation of the VAPA-COOH motif is stable along MD trajectories – although belonging to a disordered and unfolded region – as well as it is independent from the rest of the protein. Moreover, the structural behavior of the opsin/rhodopsin C-terminal tail is not affected by retinal binding, as demonstrated by the comparison of distance plots and dihedral angle distributions calculated along MD trajectories of opsin and rhodopsin.

Table 3. Distance between Cα of V345 and A348 calculated along MD trajectories of the full-length WT rhodopsin and opsin systems.

	$\mathrm{Max}_{\mathrm{distance}}$ (Å)	$\mathrm{Min}_{\mathrm{distance}}$ (Å)	$\Delta_{\mathrm{distance}}$ (Å)	$\mathrm{Mean}_{\mathrm{distance}}$ (Å) ± SD*
WT rhodopsin	10.7	7.3	3.4	9.2 ± 0.38
WT opsin	10.5	7.1	3.4	8.9 ± 0.32

*Standard deviation

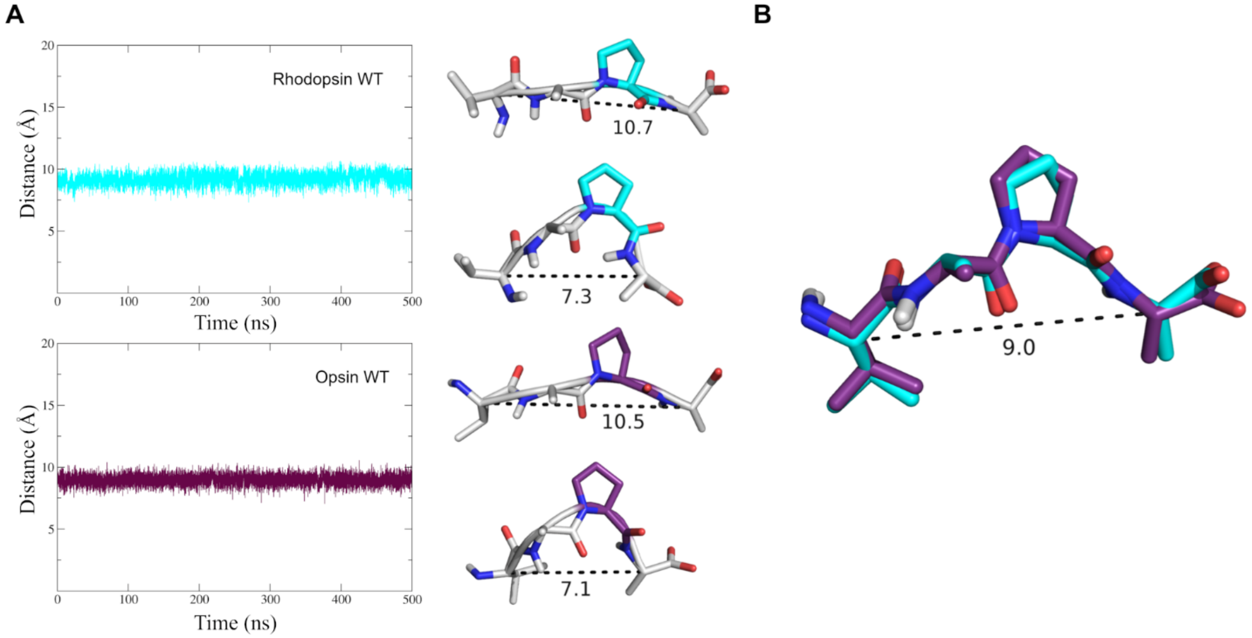


Figure 14. (A) Distance calculated between the Cα atoms of residue V345 and A348 of WT rhodopsin (cyan) and WT opsin (purple) full-length systems. The VAPA-COOH motif is represented as sticks and $\mathrm{Min}_{\mathrm{distance}}$ and $\mathrm{Max}_{\mathrm{distance}}$ between the Cα of V345 and A348 residues are highlighted by the black dashed lines. (B) Superimposition of the most representative structure of the VAPA-COOH from full-length WT rhodopsin (cyan) and WT opsin (purple). The distance between the Cα of V345 and A348 is highlighted by the black dashed line.

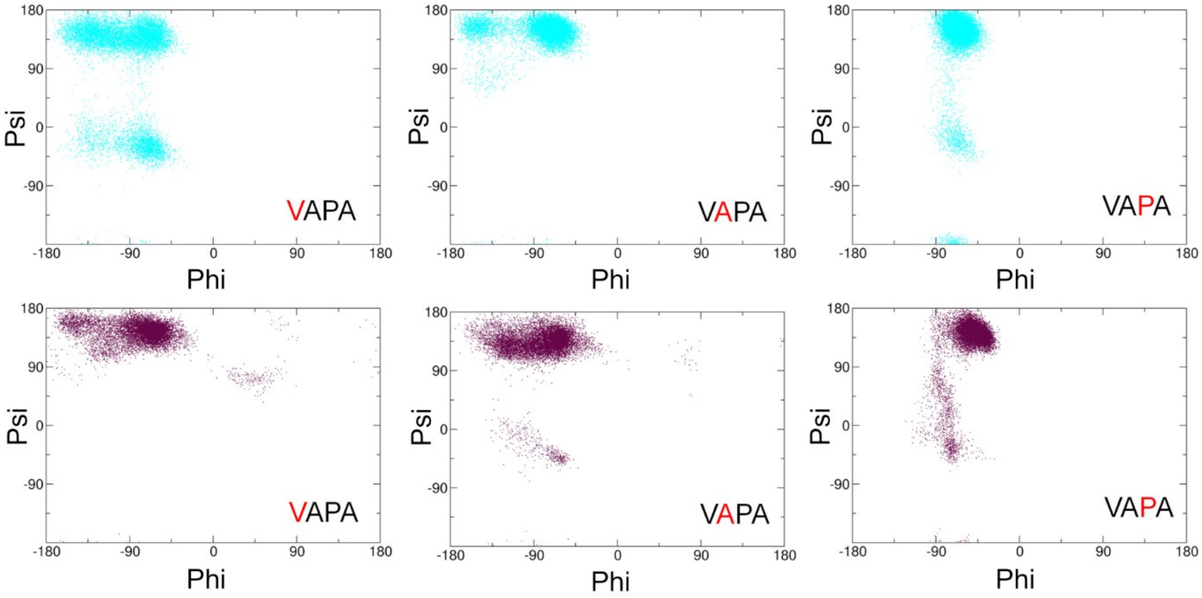


Figure 15. Φ and Ψ angles distribution of each residue of the VAPA-COOH motif of WT rhodopsin (cyan) and opsin (purple).

In agreement with experimental evidence [3], MD simulations on WT opsin and rhodopsin also suggested that the C-terminal tail does not affect the overall structure of the TMD and retinal binding properties. Since it is known that retinal photoisomerization and signal transduction are not impaired by P347 mutations, C-terminal rhodopsin mutants become of particular relevance in binding to partner proteins such as the ARF4 that promotes protein trafficking from the Golgi to the rod OS. In this context, the conformational features of clinically relevant rhodopsin mutants on the P347 site were investigated by MD simulations. Based on the results discussed above (Figures 14 and 15), which showed no difference in the behavior of VAPA-COOH in opsin and rhodopsin, these systems were simulated in the absence of the retinal chromophore (opsin form).

The distance between the Cα atoms of residues V345 and A348 was calculated among MD trajectories of pathological P347 mutants and compared to that of the WT opsin (Table 4 and Figure 16).

Table 4. Distance between Cα atoms of V345 and A348 along MD trajectories of full-length systems.

	$\mathrm{Max}_{\mathrm{distance}}$ (Å)	$\mathrm{Min}_{\mathrm{distance}}$ (Å)	$\Delta_{\mathrm{distance}}$ (Å)	$\mathrm{Mean}_{\mathrm{distance}}$ (Å) ± SD*
WT	10.5	7.1	3.4	8.9 ± 0.32
P347L	11.0	5.1	5.9	9.2 ± 0.55
P347S	11.2	5.2	6	9.2 ± 0.71
P347R	11.0	5.0	6	9.0 ± 0.96
P347T	11.0	5.0	6	9.0 ± 0.76
P347Q	11.1	4.6	6.5	9.1 ± 0.73
P347A	11.1	4.9	6.2	9.0 ± 0.73

*Standard deviation

Among the tested systems, WT opsin exhibited the lowest $\Delta_{\mathrm{distance}}$ suggesting a narrow range of fluctuations, as well as a more stable conformation of the V345-A348 sequence over MD simulations compared to P347 mutants. Indeed, in the same simulation conditions, the C-terminal tail of opsin bearing P347 mutations was remarkably more flexible than the WT. Although mutant opsins had $\mathrm{Mean}_{\mathrm{distance}}$ values similar to that of the WT, the $\Delta_{\mathrm{distance}}$ in mutants was remarkably higher than in the WT and the distance between Cα atoms of V345 and A348 largely fluctuated in MD trajectories. This suggests that, differently from the mutants, P347 plays a crucial role in maintaining the conformational stability of the VAPA-COOH motif.

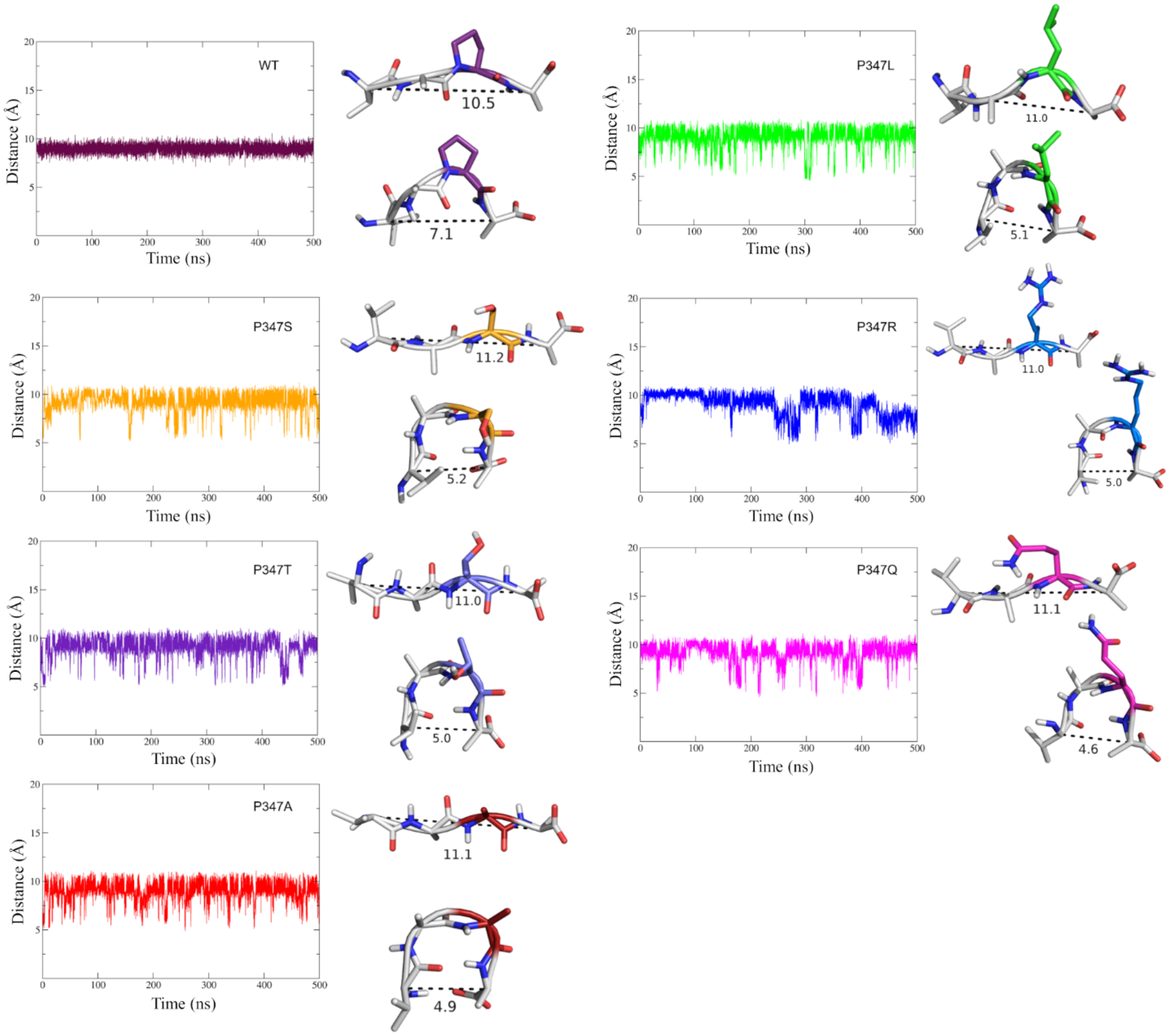


Figure 16. Distance calculated between the Cα atoms of V345 and A348 of WT (purple), P347L (green), P347S (orange), P347R (blue), P347T (violet), P347Q (magenta), and P347A (red) full-length opsin systems. The VAPA-COOH motif is represented as sticks and $\text{Min}_{\text{distance}}$ and $\text{Max}_{\text{distance}}$ between the Cα of V345 and A348 are highlighted by black dashed lines.

To further investigate the structural effects of P347 mutations on the backbone conformation of the VAPA-COOH motif, the distribution of Φ and Ψ of V345, A346, and P347 was computed along mutants' MD trajectories, and compared to the unique structural conformation of the WT systems described above (Figure 15). Following the trend of the distance plots (Figure 16), Ramachandran-like plots showed that the mutant VAPA-COOH motifs explored a larger conformational space with respect to that explored by the WT VAPA-COOH motif in MD simulations (Figure 17).

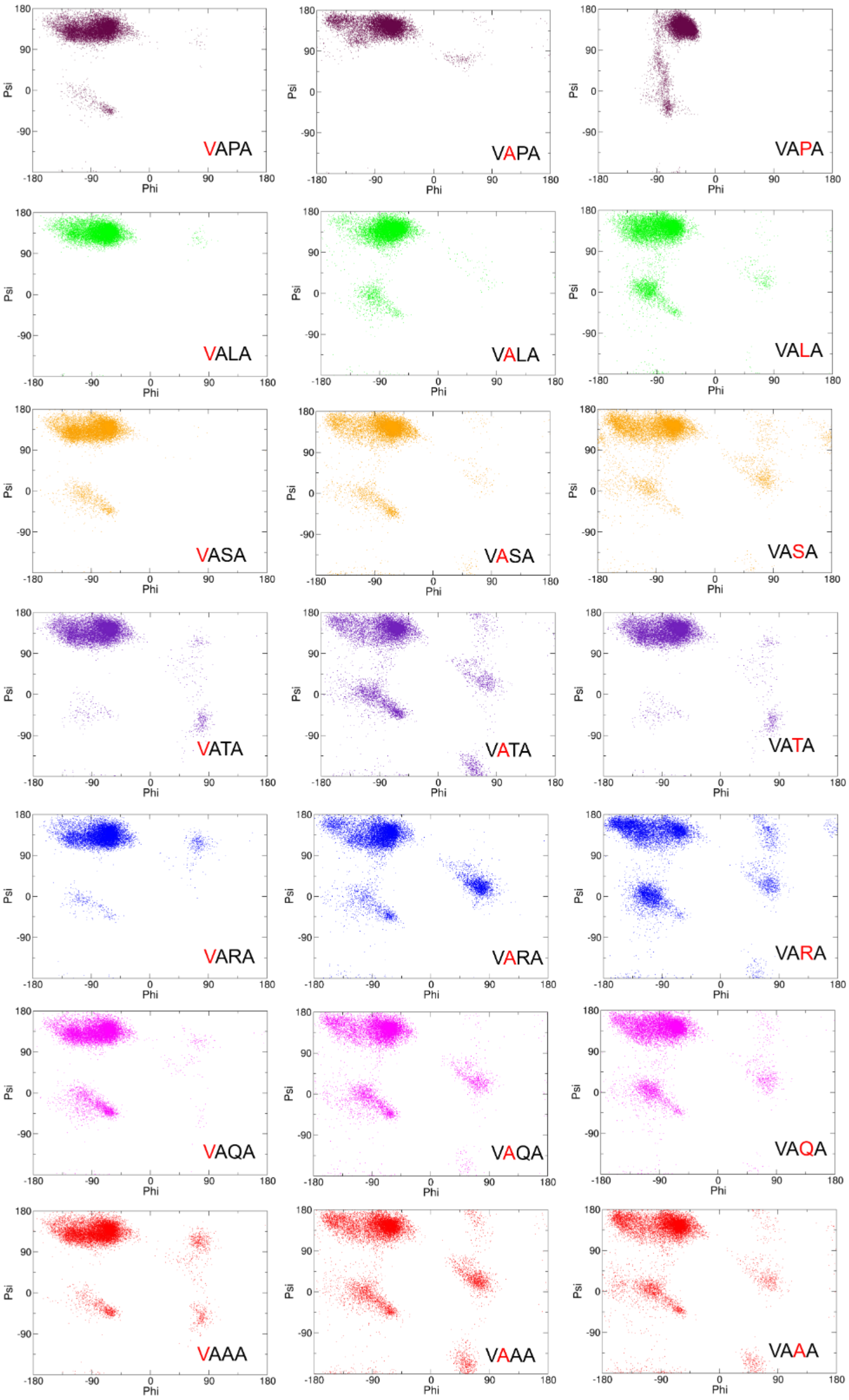


Figure 17. Phi (Φ) and Psi (Ψ) angles distribution along MD trajectories of each residue from the VAPA-COOH motif of full-length WT (purple), P347L (green), P347S (orange), P347T (violet), P347R (blue), P347Q (magenta), and P347A (red) mutant opsins.

Overall, these results clearly indicate that, in full-length opsin, mutation of P347 into clinically relevant variants has a strong impact on the overall flexibility of the VAPA-COOH motif, which might impair its interaction with specific transporter proteins such as the ARF4. In contrast, physicochemical or steric features of P347 mutants resulted less relevant, because different residues in position 347 provoke similar conformational variations in comparison to the WT P347. These data shed further light on the mistrafficking of P347 opsin/rhodopsin mutants in RP and may trigger further investigations on the recognition and binding with the specific transporter ARF4.

3.2.3. MD simulations on C-terminal rhodopsin deca-peptides

The simulation system that includes opsin/rhodopsin, the membrane, water molecules, and ions consists of a very large system. Since our efforts were focusing mainly on the last four amino acids of the protein (i.e., the VAPA-COOH motif) which showed a conformational behavior that is apparently independent from the rest of the protein in MD simulations described above, we also simulated shorter N-capped C-terminal peptides to justify use of simpler MD simulations. To this aim, simulations on the C-terminal deca-peptides of WT and mutant systems were carried out and analyzed as described for the full-length proteins. Calculation of the distance between Cα atoms of V345 and A348 (Table 5 and Figure 18) and the distribution of V345, A346, and P347 dihedral angles (Figure 19) showed a very similar trend to that obtained in full-length systems. Although the $\mathrm{Mean}_{\mathrm{distance}}$ values are very similar in all systems, the WT peptide has a $\Delta_{\mathrm{distance}}$ value significantly lower than the other investigated deca-peptides, indicating again a greater conformational rigidity of the WT VAPA-COOH motif compared to mutant systems. In these latter, conformational variations are more accessible in the MD timescale and conditions. It is worth noting that Ramachandran-like graphs (Figure 19) showed deca-peptides exploring a larger conformational space than full-length systems (Figure 17). One may note that the conformational freedom of isolated deca-peptides might be enhanced by the lack of interactions with the molecular environment composed by the membrane and the cytosolic edge of rhodopsin transmembrane domain. Notably, in our full-length systems we did not observe significant or persistent interactions between the VAPA-COOH and the remaining part of the macromolecular ensemble, indicating that the smaller size of the deca-peptides allowed a more efficient sampling of the intrinsic conformational space in MD time compared to the larger full-length systems. Results provided by MD simulations of WT full-length and deca-peptide opsin were highly comparable to each other, suggesting that the use of simplified peptides are particularly reliable to study the conformational behavior of the C-terminal tail of rhodopsin.

Overall, the peculiar conformation of the proline residue allows the unique arrangement of the C-terminal tail of rhodopsin to be maintained throughout the MD simulation timescale, which any other mutation failed to maintain. The characterization of conformational and steric changes caused by the P347 mutation could help understanding the reasons why P347 mutant rhodopsin cannot bind specific transporters and is linked to RP, and possibly to design pharmacological strategies aimed at restoring the functional rhodopsin signaling in the present of P347 mutations.

Table 5. Distance between Cα atoms of V345 and Cα of A348 in deca-peptide systems calculated along MD trajectories.

	$\text{Max}_{\text{distance}}$ (Å)	$\text{Min}_{\text{distance}}$ (Å)	Δ_{distance} (Å)	$\text{Mean}_{\text{distance}}$ (Å) ± SD*
WT	10.7	7.0	3.7	9.2 ± 0.37
P347L	11.1	4.0	7.1	8.2 ± 1.19
P347S	11.2	4.2	7	8.6 ± 0.85
P347R	11.3	4.7	6.6	9.2 ± 0.72
P347T	11.3	3.9	7.4	8.7 ± 1.14
P347Q	11.2	4.1	7.1	8.9 ± 0.81
P347A	11.2	4.1	7.1	8.3 ± 1.31

*Standard deviation

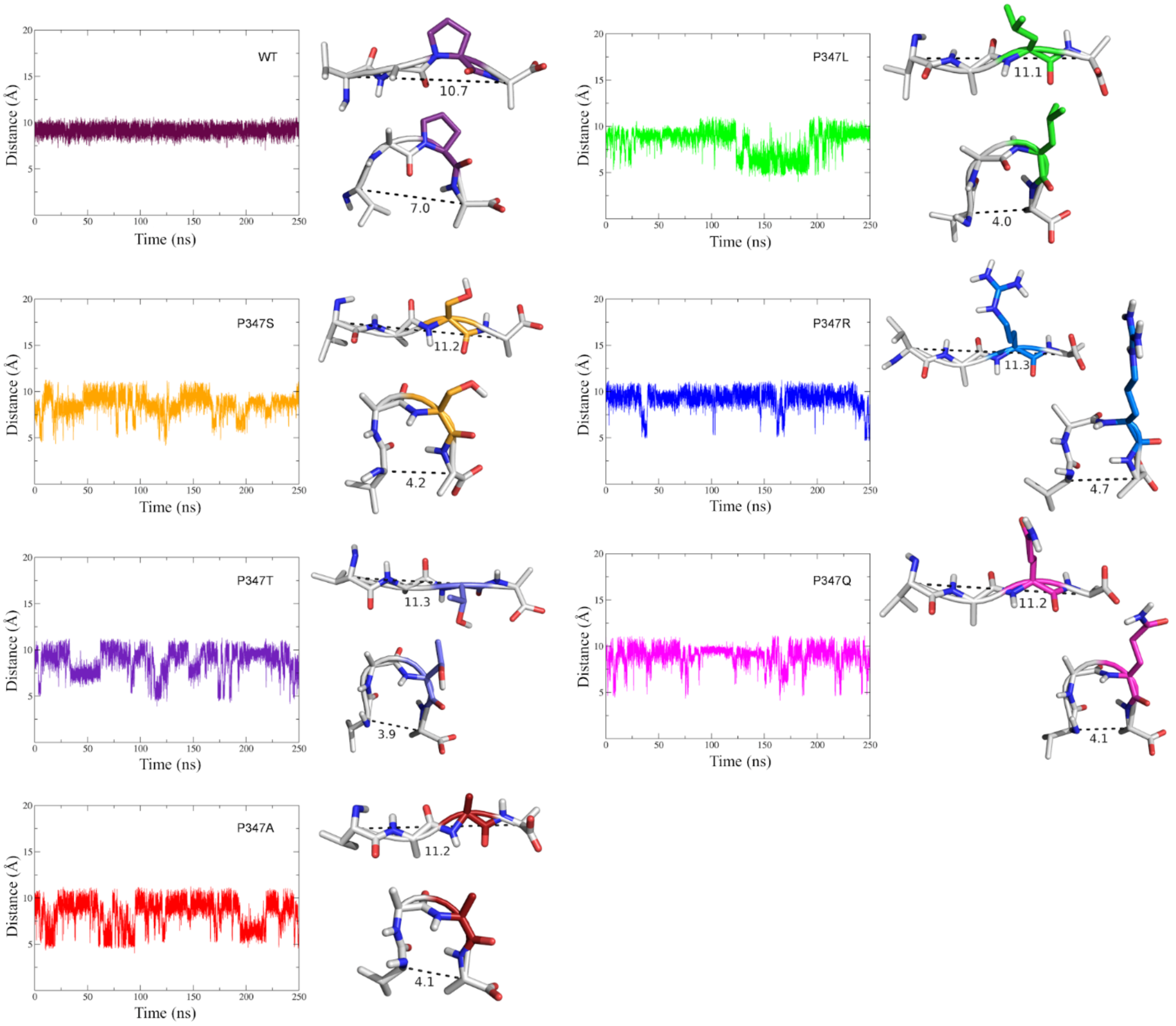


Figure 18. Distance calculated between Cα atoms of V345 and A348 of WT opsin (purple), P347L (green), P347S (orange), P347R (blue), P347T (violet), P347Q (magenta), and P347A (red) decapeptide systems. The VAPA-COOH motif is represented as sticks and $\mathrm{Min}_{\mathrm{distance}}$ and $\mathrm{Max}_{\mathrm{distance}}$ between the V345 and A348 residues are represented by the black dashed lines.

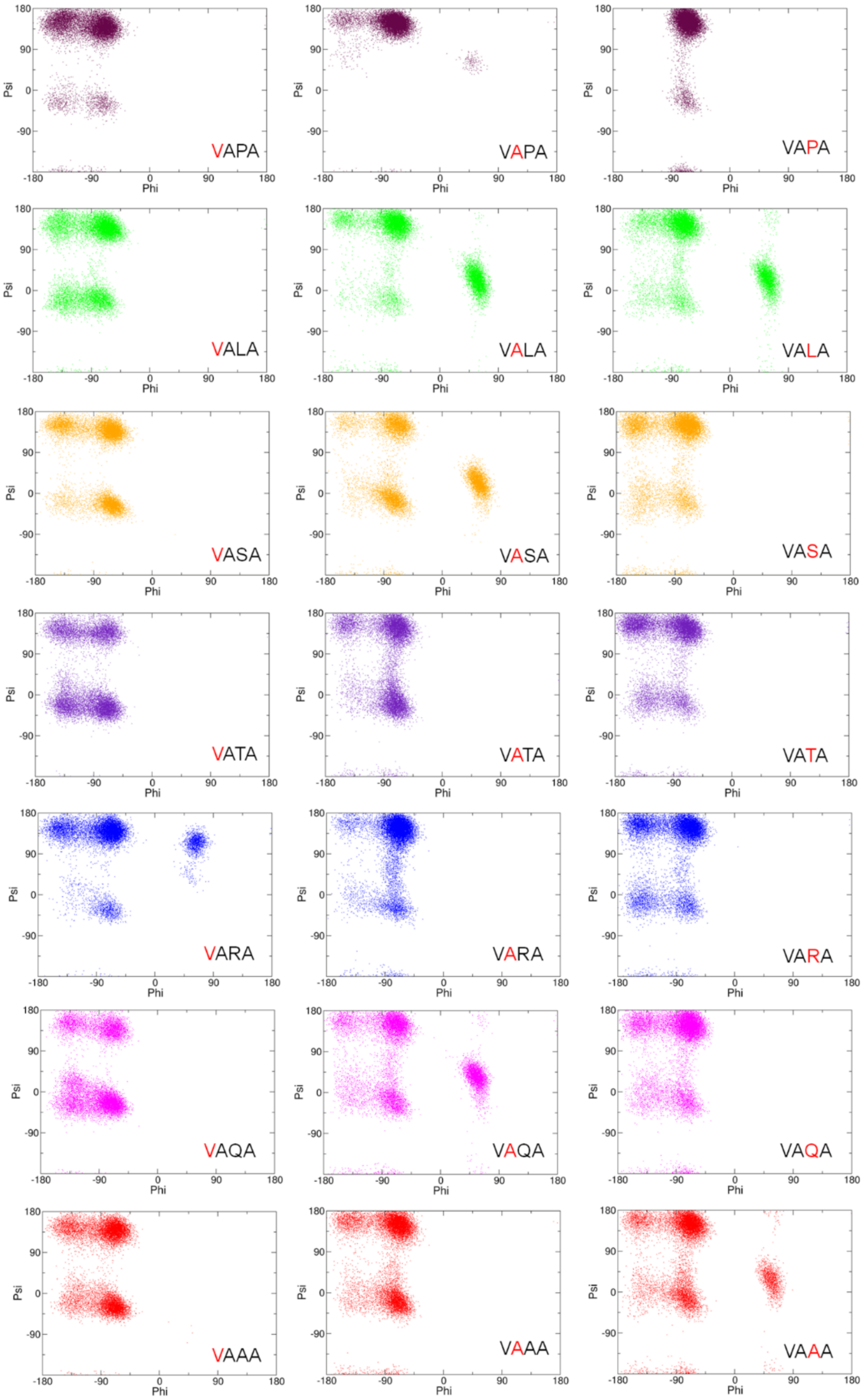


Figure 19. Phi (Φ) and Psi (Ψ) angles distribution of each residue from the VAPA-COOH motif of WT deca-peptide opsin (purple), P347L (green), P347S (orange), P347T (violet), P347R (blue), P347Q (magenta), and P347A (red) mutants.

3.3. Conclusions

Several lines of experimental evidence indicate that the trafficking of human rhodopsin to the rod OS, which is a crucial process for the vision cascade, is mediated by the direct interaction between the rhodopsin VAPA-COOH motif and the specific transporter ARF4 [3]. Nevertheless, structural details of this interaction as well as on the C-terminal VAPA-COOH motif of WT and mutant opsin/rhodopsin that are relevant for RP have not been elucidated, mostly because this sequence belongs to an IDP. In this work, the conformational features of the VAPA-COOH from WT rhodopsin and the structural effects caused by pathological mutations at position P347 have been characterized using MD simulations.

First, MD simulations on human full-length rhodopsin and opsin homology models were carried out, clearly indicating that the conformation of the VAPA-COOH motif is not affected by the presence or the absence of 11-*cis* retinal chromophore in its binding site. Secondly, comparison of the WT system with six mutants found in RP patients showed that Pro347 plays an essential role in the conformational arrangement of the VAPA-COOH motif, as this residue confers a remarkable structural rigidity to the backbone compared to pathological mutants. Since MD simulations indicated that the conformational freedom of the rhodopsin/opsin C-terminal tail is independent from the rest of the protein, to decrease the overall size of the studied systems and to increase the computational efficiency, we present data from the simulation of the corresponding C-terminal deca-peptides. Remarkably, isolated deca-peptides emerged as extremely reliable models in the conformational study of the VAPA-COOH motif, providing results that are very similar to those obtained with full-length systems. Since the pathological mutants herein investigated have heterogeneous physicochemical features, different properties could affect the binding to ARF4, including steric hindrance and polarity. In contrast, our results showed that the backbone plasticity of the VAPA-COOH WT motif is more relevant than physicochemical features of P347 mutants, and that changes in backbone flexibility may cause the defective interaction of P347 mutant rhodopsin to ARF4. Results of this study shed further lights into the mistrafficking of mutant rhodopsin, and could represent a first step for further investigations towards the development of therapeutic strategies for RP.

3.4. Materials and methods

3.4.1. Homology modelling

Since no crystallographic structure of human rhodopsin was available in the Protein Data Bank, the crystallographic structure of full-length bovine rhodopsin (PDB-ID: 1U19) [60] was selected as a structural template for homology modeling based on its high resolution (2.20 Å) as well as the structural details of the C-terminal tail. Alignment of the FASTA sequences of human and bovine rhodopsin (Uniprot-ID: P08100 and P02699, respectively [47]) resulted in a very high percentage of identity of the primary sequence (93.4%), which guarantees the generation of an extremely reliable homology model [164]. For this reason, structures of human WT and P347L, P347S, P347R, P347T, P347Q, and P347A mutant rhodopsin, were generated by homology model using the Prime software in the Schrödinger Release 2019-1 [131]. The protonation state of the histidine residues was predicted by H++ 3.0 [165].

3.4.2. MD simulations

Both full-length WT rhodopsin (opsin together with 11-*cis* retinal) and opsin were simulated, while pathological mutants were simulated only as opsin systems. The retinal chromophore in rhodopsin is covalently bound to the K296 through a protonated Schiff base. The ff14SB force field [132] was used to parameterize the protein, while the 11-*cis* retinal and the covalent bond parameters were partially retrieved from Ferrè et al., [133]. Each system was included in a DOPC-based membrane (138 DOPC molecules) [72] using the CHARMM-GUI membrane builder [134,135], neutralized in charge with the appropriate number of Na^+ counter-ions, and then solvated in a rectangular box of 14629 TIP3P water molecules. Each full-length system was relaxed as follows: first, solvent was energy minimized for 1500 steps using the SD followed by 3500 steps using the CG. Then, the whole system was energy minimized for 10000 steps (1500 SD and 8500 CG), heated from 0 to 300 K over 1 ns (Langevin thermostat) at constant volume with a collision frequency of 2 ps^{-1}. Density equilibration was achieved over 1 ns using the Berendsen barostat to control the anisotropic pressure scaling with a pressure relaxation time of 1 ps. The SHAKE algorithm was used to constraint the covalent bonds between heavy and hydrogen atoms. Long-range electrostatic interactions were computed with a cut-off of 10 Å using the Particle Mesh Ewald (PME) method. After a preliminary equilibration of 50 ns, unrestrained MD simulations were run for 500 ns with a time step of 2 fs at constant pressure by the Amber18 program [136], and analyzed by the CPPTRAJ software [137]. MD frames were clustered with the hierarchical agglomerative (bottom-up) approach. All MD simulations were performed on Nvidia Tesla V100 GPU. Simplified deca-peptide systems were generated by taking only the last 10 residues from each of the previous full-length systems and adding an ACE N-terminal cap. Deca-peptides were solvated in a rectilinear box of 4171 TIP3P water molecules and prepared following the same procedure used for the full-length systems. Then, 250 ns of unrestrained MD simulations were carried out. For all simulated systems, two replicas were run starting from slightly different initial coordinates.

4. Structural elucidation of novel Imidazo[1,2-*a*]pyridine inhibitors of Aldehyde Dehydrogenase 1A Family

The data and information reported in this chapter were published in:

Quattrini, L., Gelardi, E., Petrarolo, G., Colombo, G., Ferraris, D. M., **Picarazzi, F.**, Rizzi, M., Garavaglia, S., & La Motta, C. (2020). Progress in the Field of Aldehyde Dehydrogenase Inhibitors: Novel Imidazo[1,2-*a*]pyridines against the 1A Family. *ACS medicinal chemistry letters*, *11*(5), 963–970. https://doi.org/10.1021/acsmedchemlett.9b00686

Gelardi, E., Colombo, G., **Picarazzi, F.**, Ferraris, D. M., Mangione, A., Petrarolo, G., Aronica, E., Rizzi, M., Mori, M., La Motta, C., & Garavaglia, S. (2021). A Selective Competitive Inhibitor of Aldehyde Dehydrogenase 1A3 Hinders Cancer Cell Growth, Invasiveness and Stemness In Vitro. *Cancers*, *13*(2), 356. https://doi.org/10.3390/cancers13020356

4.1. Background and rationale of the work

The ALDHs superfamily, also known as retinaldehyde dehydrogenase, consists of 19 different isoenzymes involved in the irreversible detoxification of aldehydes to the respective carboxylic acids, using $NAD(P)^+$ as a cofactor [104]. The different ALDHs enzymes act in various metabolic pathways, including detoxification of aldehydes derived from lipid peroxidation, ethanol consumption and the metabolism of several neurotransmitters [105,106]. The 19 human ALDHs protein members are divided in families in which every isoform shares at least 40% of sequence identity, and in subfamilies that show 70% amino acid identity [166]. Moreover, ALDHs have an essential role in the biosynthesis of key metabolic regulators of cellular homeostasis, such as retinoic acid, γ-aminobutyric acid, dopamine and betaine [107]. In particular, the ALDH1A subfamily is composed by three different isoforms (ALDH1A1, ALDH1A2 and ALDH1A3) which are involved in the conversion of retinal, an important mediator of vision, to the retinoic acid, a potent tissue differentiation factor for cellular development. Members of the ALDH1A family are commonly acknowledged as hallmarks of cancer stem cells (CSCs) and the overexpression of the ALDH1A3 isoform is significantly associated with tumorigenesis and chemotherapeutic resistance [167]. However, within the rod photoreceptors cells, retinoic acid levels are finely controlled by a series of enzymes, including ALDH1A3, which catalyse the oxidation of the all-*trans* retinal to retinoic acid and for which the most efficient retinal dehydrogenase activity was reported [16]. Although their role still unclear, and would deserve in-depth investigations, some studies have shown the correlation between the enzymatic activity of ALDHs and RP [118,119]. Based on this evidence, we hypothesized that retinal concentration may be correlated to ALDH1A3 enzymatic activity, and that increasing retinal concentration in cell may partially restore the physiological activity of mutant rhodopsin, therefore we speculated here that modulation of the enzymatic activity of ALDH1A3 may have therapeutic implications in RP, especially in those mutations that are dependent from retinal concentration. In this regard, starting from the reference ALDH inhibitor GA11 previously developed by our group [168], and making use of distinctive substitution patterns on the two pendant phenyl rings, a novel series of imidazo[1,2-*a*]pyridines was designed and optimized by means of a structure-based approach. A close inspection of the complex obtained co-crystallizing GA11 with the human recombinant ALDH1A3 revealed that the reference compound gets settled at the entrance of the catalytic tunnel of the enzyme [168]. Arranged in this pose, GA11 forecloses access to the natural substrate retinal, thus allowing inhibition of the catalytic activity of the enzyme. However, this binding mode formally occupies only a part of the enzyme catalytic site, thus offering interesting clues for the rational optimization of GA11. Based on the crystallographic binding mode, the insertion of suitable substituents on the two pendant rings of GA11may help to fill in further the catalytic site of ALDH1A3, thus getting to novel effective and more selective analogues. Moreover, the obtained binding mode led to speculation that shifting the 6-phenyl ring to the 8-position could come in very handy to cast these derivatives into the deepest part of the enzyme binding site, in close proximity to the key catalytic residue C314. Prompted by these speculation, we embarked in the synthesis of a novel series of 8-substituted derivatives and 6-substituted-imidazo[1,2-*a*]pyridines, which were investigated for their functional efficacy and mode of interaction against the human recombinant isoforms of the ALDH1A family. Among them, compounds NH51 and NR6 emerged as the most promising selective ALDH1A3 inhibitors, and they were further evaluated *in silico* for their ability to interact with the catalytic site of the 1A3 isoform [166,169]. Overall, computational approaches provided structural and thermodynamics insights to the design and optimization of small molecule inhibitors of ALDH1A3 that could act as therapeutic and/or biological tools, not only as an anticancer agents for which ALDH1A3 is a validated target, but also to deepen the knowledge of its implication in the pathological contest of RP with the future ambition of a co-administration of molecular chaperones and selective inhibitors of ALDH1A3 with synergistic effects.

4.2. Results and discussion

4.2.1. Functional evaluation of the novel imidazo[1,2-a]pyridine derivatives

The imidazo[1,2-*a*]pyridine derivative NR6 and NH51 were tested for both their inhibitory efficacy against the human recombinant ALDH1A3 and selectivity towards the ALDH1A1 and ALDH1A2 isoforms. The kinetic parameters were determined for all compounds, which proved to inhibit the target ALDH1A3 when tested at 25 μM concentration, displaying different degrees of efficacy towards the 1A1 and 1A2 isoenzymes (Table 6). NH51 compound was highly selective against the human isoenzyme ALDH1A3 (K_i 0.12 ± 0.01 μM) with respect to the parent 1A1 (K_i 79.5 ± 1.8 μM), but not completely toward the 1A2 isoform. However, it showed a good inhibitory profile (IC_{50} 0.66 ± 1.3 μM, Table 6), with an activity in the submicromolar range. NR6, was investigated further to determine its IC_{50} (5.3 ± 1.5 μM) and K_i (3.7 ± 0.4 μM) values, which revealed a potent and competitive model of inhibition. Furthermore, the high IC_{50} and Ki values of NR6 towards ALDH1A1 and ALDH1A2 confirm its high selectivity against ALDH1A3 as shown in Table 6.

Table 6. ALDH1As inhibitory activity of compounds NH51 and NR6.

	ALDH1A1		ALDH1A2		ALDH1A3	
Compound	K_i μM	IC_{50} μM	K_i μM	IC_{50} μM	K_i μM	IC_{50} μM
GA11	n.a[a]	19.6 ± 1.6	n.a	54.6 ± 1.9	0.54 ± 0.11	4.7 ± 1.7
NH51	79.5 ± 1.8	n.a	0.92 ± 1.5	n.a	0.12 ± 0.01	0.66 ± 1.3
NR6	262.2 ± 76.4	57.6 ± 2.2	257.6 ± 26.4	123.6 ± 2.2	3.7 ± 0.4	5.3 ± 1.5

[a] Not active. No significant enzyme inhibition was observed at 25 μM.

4.2.2. Molecular modeling studies on ALDH1A3-NH51 complex

Compounds NH51 and NR6 were then investigated further, to clarify their mode of interaction with the ALDH1A3 isoform. Unfortunately, different attempts to co-crystallize NH51 in complex with human recombinant ALDH1A3 were fruitless. Therefore, in the absence of a reliable crystal structure, the binding mode of NH51 within the catalytic site of ALDH1A3 has been assumed by molecular docking simulations. To assess the reliability of the docking protocol, the co-crystallized ligand MF13, successfully crystallized in this work [169], was self-docked into the catalytic site of ALDH1A3, providing a very good agreement with the crystallographic pose (RMSD < 0.5 Å). Molecular docking simulations identified two possible, although alternative, binding modes for NH51. The replacement of the chlorine atom with the ester moiety significantly modified the binding mode of the imidazopyridine scaffold. In the first pose (Figure 20A), the phenyl ring in position 2 is oriented toward the catalytic C314 and establishes a sandwich-like and a parallel-displaced π–π stacking interaction with F182 and with W189, respectively. The imidazopyridine core is accommodated into the hydrophobic pocket flanked by I132, F308, and L478, at the entrance of the catalytic tunnel, while the ester group establishes a H-bond with N469. In the second pose (Figure 20B), interactions of the phenyl group at position 2 and the imidazopyridine group remain unchanged, while the ester moiety is H-bonded to R139. It is worth mentioning that, regardless of the binding poses proposed by the GOLD program, NH51 is able to establish an additional H-bond interaction with key ALDH1A3 residues that are not present in the crystallographic binding mode of MF13, and this gives reason for the stronger ALDH1A3 inhibitory activity of NH51.

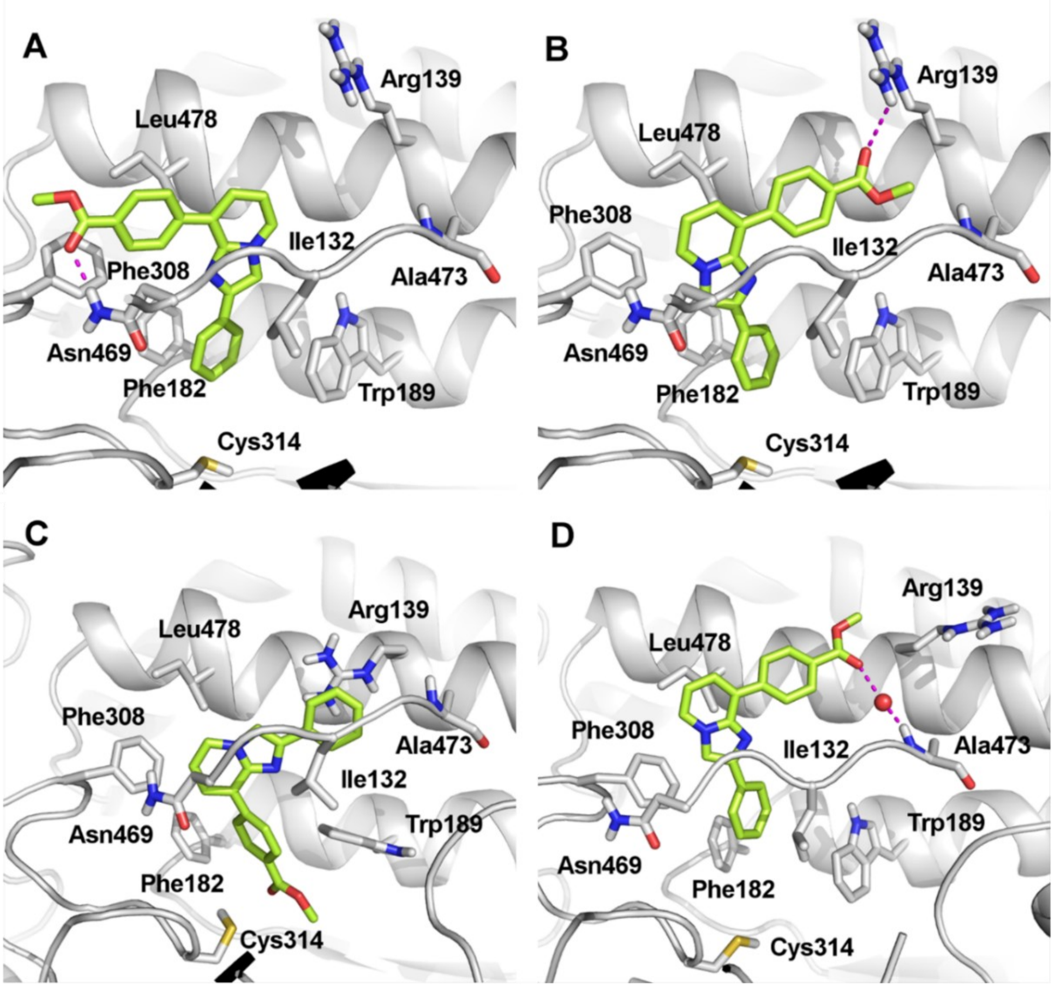


Figure 20. Stick representations of the proposed docking poses (A and B) and representative frames from MD simulations (C and D) of compound NH51 into the binding site of ALDH1A3 (PDB code:

6TRY), represented as gray ribbons. Residues relevant for the inhibitor binding have been labeled, and their side chains are shown as gray sticks.

The obtained docking poses were further investigated by MD simulations using Amber18 [136]. Each ALDH1A3-NH51 complex was simulated for 300 ns of unrestrained MD, while the most representative conformations were extracted from MD trajectories by cluster analysis. The docking pose shown in Figure 20A is highly destabilized in MD and converges to a new binding mode in which the ester group is oriented toward the catalytic C314, while the unsubstituted phenyl ring establishes a cation–π interaction with R139 (Figure 20C). Notably, this pose remains stable along the MD trajectory and is highly comparable to the crystallographic pose of MF13 (Figure 21).

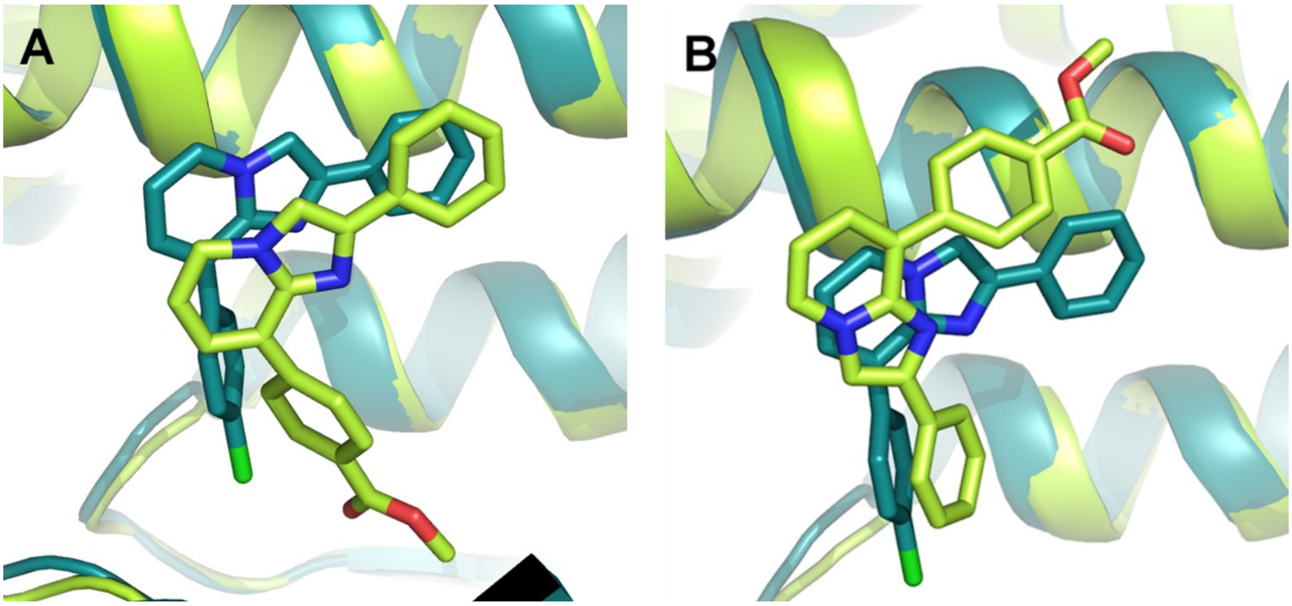


Figure 21. Overlap of the binding mode of crystallographic compound MF13 (aquamarine sticks) and the representative poses of compound NH51 (green sticks) extrapolated from MD trajectories.

The pose shown in Figure 20B, in which the ester group is H-bonded to R139, keeps the same orientation for the whole MD simulation. In the MD trajectory, NH51 loses the interaction with R139 and establishes a water-bridged H-bond with A473 (Figure 20D). The theoretical affinity of NH51 to ALDH1A3 was computed by the MM-GBSA approach [8]. Notably, the two poses are almost isoenergetic (delta energies of binding are -33.07 ± 0.26 kcal/mol and -35.88 ± 0.47 kcal/mol for the pose shown in Figure 20C and 20D, respectively). Overall, molecular modeling studies, carried out on the representative compounds NH51, evidence that the novel 2,8-diphenylimidazopyridines are able to bind ALDH1A3 protruding toward the catalytic C314 residue. Results obtained confirm the soundness of our initial working hypothesis, based on the suitable relocation of the 6-phenyl ring of GA11 to increase the structural complementarity of this class of compounds with the enzyme binding site.

4.2.3. ALDH1A3-NR6 Complex Crystallization Study

Differently from compound NH51, compound NR6 was successfully co-crystallized with ALDH1A3 (PDB ID: 7A6Q). Crystallographic data revealed that, in all monomers, NR6 binds to the enzyme at the entrance of the catalytic tunnel (Figure 22A), and no direct interactions between the catalytic cysteine and the ligand are present. The NR6 heterocyclic cores lie in a conserved hydrophobic pocket located at the entrance of the catalytic tunnel. In this conformation, the phenyl ring in position 6 of the nucleus establishes contacts with the protein backbone, with Van der Waals contacts and hydrophobic interactions with the residues G136, R139, W189, N469, A470, L471 and Y472 (Figure 22B, 22C, 22D).

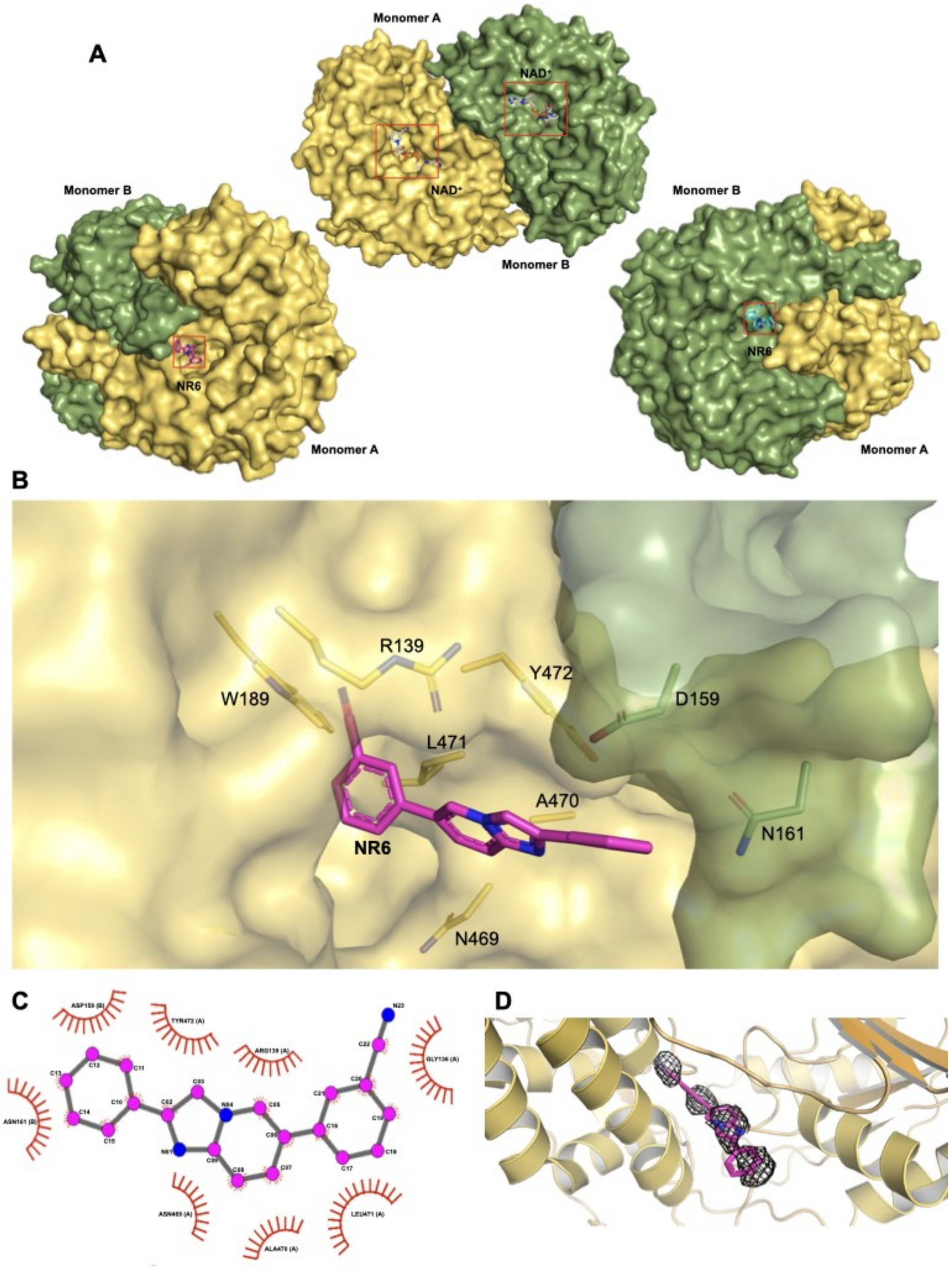


Figure 22. Structural analysis of ALDH1A3 in complex with NR6 (A) Representation of both monomer (A and B) as a surface, both are bonded with one molecule of NAD+ and one molecule of NR6, that are lodged with the same orientation. The inhibitor occupies only partially the entrance of the catalytic tunnel in both monomer; (B) Focus on the NR6 binding site: all side chains of the residues implicated in the NR6 binding are highlighted as sticks. The 2-phenyl ring and the backbone of the residues D159 and N161 from monomer B interact each other. (C) LigPlot+ analysis of the complex between NR6 and ALDH1A3, all the major interactions are hydrophobic, due to the nature of the molecule and the aminoacidic supply of the protein in that specific region. (D) NR6 2mFo-Fc map (isomesh = 0.8 carve = 1.6).

It is worth noting that this is the same binding-site occupied by the cyclohexene ring of the natural ligand retinoic acid, by phenyl group of GA11 and its derivatives, including NH51, how recently highlighted in previous published data [120,168,169], (Figure 23A). All mentioned amino acids, excluded the Y472, are conserved in each of the 1A1, 1A2, and 1A3 isozymes, highlighting this enzymatic area as "stabilizing ligand binding site" in the whole members of the 1A subfamily. A structure comparison between ALDH1A3 in complex with the progenitor of the imidazo[1,2-*a*]pyridine series, namely GA11, and NR6, provides interesting suggestions. GA11, being devoid of any substituents on the phenyl rings in position 6 of the heterocycle core, is a symmetric compound that can bind the macromolecule with two different poses, while its best derivative NR6 is able to bind ALDH1A3 with only one conserved pose. Indeed, the cyanide group blocks the inhibitor in a rigid and conserved pose in both the monomers (Figure 23B). As clearly shown in Figure 23C, the Y472 strongly binds and coordinates the pyridinic ring of NR6 with an edge-to-face π-π stacking interaction. All these interactions are further stabilized by a hydrogen bond that is established between the -OH group of the Y472 and the N1 of NR6 (3.7 Å). It is important to highlight that Y472 residue is not conserved in the ALDH1A1 and in the ALDH1A2, where it is substituted with a S461 and N478, respectively. Although ALDH1A3 shares 71% and 72% amino acid identity with human ALDH1A1 and ALDH1A2 isoenzymes, respectively, the Y472, essential in the coordination with the inhibitor, differentiates the binding pocket of the target ALDH1A3 from the one of the parent isoenzymes (Figure 23D) and from the other nineteen ALDHs members (Figure 24). These two amino acids would retain the hydrophobic interactions unchanged, but the strong coordination with the aromatic group become completely lost. The presence of the cyanide group promotes the distinct orientation of NR6 respect to Y472, thus stabilizing the binding with the protein and determining high selectivity towards ALDH1A3 (Figure 23D and 23E).

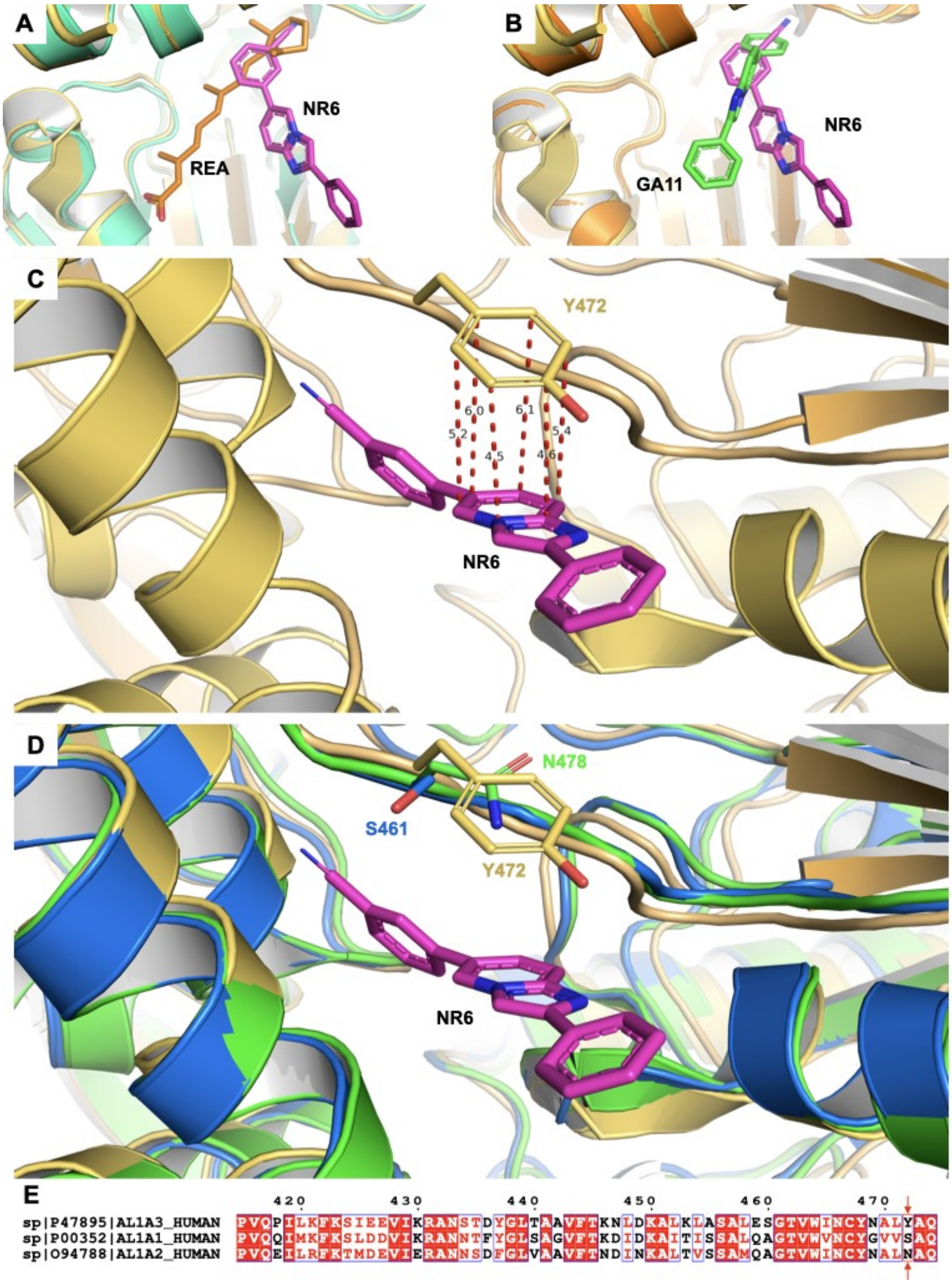
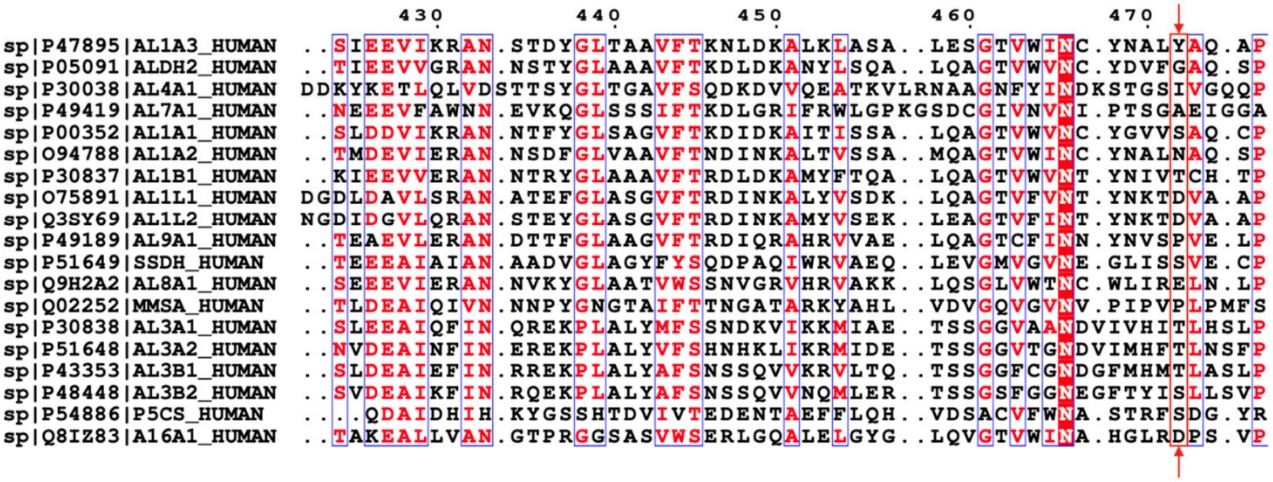


Figure 23. Structural superimposition between the complex ALDH1A3-NR6, other ALDH1A3 complexes (PDB 5FHZ and PDB 6S6W and the ALDH1A1 (PDB 4WB9) and ALDH1A2 (6B5G) to describe the key residue for the selectivity (A) Superimposition between monomers A from PDB 5FHZ (turquoise) and monomer A from PDB 7A6Q (yellow), with a focus on the binding region of the retinoic acid (REA) in the open conformation, shared with NR6. In particular, the 6-phenyl ring of NR6 ring conserved the main interactions as the cyclohexene ring of the natural product. (B) Superimposition between monomers A from PDB code 6S6W (orange) and monomer A from PDB code 7A6Q (yellow), highlighting the common pose at level of the 6-phenyl ring while the heterocycle is shifted to another region always at the entrance of the catalytic site (the main interaction

of GA11 is with F131). (C) Focus on the hydrophobic interaction mediated by the tyrosine and the pyridine ring of the heterocycle. The distances between all the single atoms participating to the edge-to-face π-π stacking between the Y472 from ALDH1A3 and NR6 are represented in red. (D) Superimposition between ALDH1A3 (yellow, PDB code 7A6Q), ALDH1A1 (blue, PDB code 4WB9) and ALDH1A2 (green, PDB code 6B5G), the tyrosine is not conserved in the other two isoforms. I Sequence alignment between the three isoforms, a single residue seems to give high selectivity to NR6 against ALDH1A3 toward the other members of the subfamily.



```
                                  430          440         450            460            470
sp|P47895|AL1A3_HUMAN    ..SIEEVIKRAN.STDYGLTAAVFTKNLDKALKLASA..LESGTVWINC.YNALYAQ.AP
sp|P05091|ALDH2_HUMAN    ..TIEEVVGRAN.NSTYGLAAAVFTKDLDKANYLSQA..LQAGTVWVNC.YDVFGAQ.SP
sp|P30038|AL4A1_HUMAN    DDKYKETLQLVDSTTSYGLTGAVFSQDKDVVQEATKVLRNAAGNFYINDKSTGSIVGQQP
sp|P49419|AL7A1_HUMAN    ..NEEEVFAWNN.EVKQGLSSSIFTKDLGRIFRWLGPKGSDCGIVNVNI.PTSGAEIGGA
sp|P00352|AL1A1_HUMAN    ..SLDDVIKRAN.NTFYGLSAGVFTKDIDKAITISSA..LQAGTVWVNC.YGVVSAQ.CP
sp|O94788|AL1A2_HUMAN    ..TMDEVIERAN.NSDFGLVAAVFTNDINKALTVSSA..MQAGTVWINC.YNALNAQ.SP
sp|P30837|AL1B1_HUMAN    ..KIEEVVERAN.NTRYGLAAAVFTRDLDKAMYFTQA..LQAGTVWVNT.YNIVTCH.TP
sp|O75891|AL1L1_HUMAN    DGDLDAVLSRAN.ATEFGLASGVFTRDINKALYVSDK..LQAGTVFVNT.YNKTDVA.AP
sp|Q3SY69|AL1L2_HUMAN    NGDIDGVLQRAN.STEYGLASGVFTRDINKAMYVSEK..LEAGTVFINT.YNKTDVA.AP
sp|P49189|AL9A1_HUMAN    ..TEAEVLERAN.DTTFGLAAGVFTRDIQRAHRVVAE..LQAGTCFINN.YNVSPVE.LP
sp|P51649|SSDH_HUMAN     ..TEEEAIAIAN.AADVGLAGYFYSQDPAQIWRVAEQ..LEVGMVGVNE.GLISSVE.CP
sp|Q9H2A2|AL8A1_HUMAN    ..SEEEVIERAN.NVKYGLAATVWSSNVGRVHRVAKK..LQSGLVWTNC.WLIRELN.LP
sp|Q02252|MMSA_HUMAN     ..TLDEAIQIVN.NNPYGNGTAIFTTNGATARKYAHL..VDVGQVGVNV.PIPVPLPMFS
sp|P30838|AL3A1_HUMAN    ..SLEEAIQFIN.QREKPLALYMFSSNDKVIKKMIAE..TSSGGVAANDVIVHITLHSLP
sp|P51648|AL3A2_HUMAN    ..NVDEAINFIN.EREKPLALYVFSHNHKLIKRMIDE..TSSGGVTGNDVIMHFTLNSFP
sp|P43353|AL3B1_HUMAN    ..SLDEAIEFIN.RREKPLALYAFSNSSQVVKRVLTQ..TSSGGFCGNDGFMHMTLASLP
sp|P48448|AL3B2_HUMAN    ..SVDEAIKFIN.RQEKPLALYAFSNSSQVVNQMLER..TSSGSFGGNEGFTYISLLSVP
sp|P54886|P5CS_HUMAN     ....QDAIDHIH.KYGSSHTDVIVTEDENTAEFFLQH..VDSACVFWNA.STRFSDG.YR
sp|Q8IZ83|A16A1_HUMAN    ..TAKEALLVAN.GTPRGGSASVWSERLGQALELGYG..LQVGTVWINA.HGLRDPS.VP
```

Figure 24. Sequence alignment between all the nineteen isoforms. The tyrosine is not conserved in any other ALDH isoenzymes. None of the residues close to the same region present a hydrophobic side chain that could supply the absence of the tyrosine to coordinate the heterocycle with the same kind of interaction (edge-to-face π-π stacking).

4.2.4. Molecular dynamics study of the mutant ALDH1A3 Y472A in complex with NR6

To gain further insights into the role of Y472 in the binding of NR6 near the entrance of ALDH1A3 catalytic tunnel, MD simulation was carried out starting from the crystallographic complex. The specific contribution of Y472 to NR6 theoretical affinity for ALDH1A3 was estimated by computational alanine scanning [170], while the delta energy of ligand binding was computed using the MM-GBSA approach [8]. Alanine scanning calculations allowed to quantify the role of a specific amino acid in a protein/ligand complex as the difference (delta-delta energy of binding $\Delta\Delta E_b$) between the ligand's theoretical affinity for the WT sequence and that for the alanine mutant (Y472A in our case).

Results showed that Y472 strongly contributes to the theoretical affinity of NR6 for ALDH1A3 with a $\Delta\Delta E_b$ of -2.19 kcal/mol (Figure 25A). To deeply investigate the nature of the energy contribution provided by Y472, the theoretical affinity was then decomposed in van der Waals and electrostatic terms. From these results, it clearly emerges that the greatest energetic contribution of Y472 compared to the A472 is attributable to van der Waals interactions (Figure 25B), in agreement with the π-π stacking interaction observed between Y472 and NR6 in the crystallographic structure.

A

Complex	Theoretical ΔE_b[kcal/mol]		$\Delta\Delta E_b$ [kcal/mol]
	WT	Y472A	
ALDH1A3/NR6	-31.29 ± 0.51	-29.10 ± 0.52	-2.19 ± 0.97

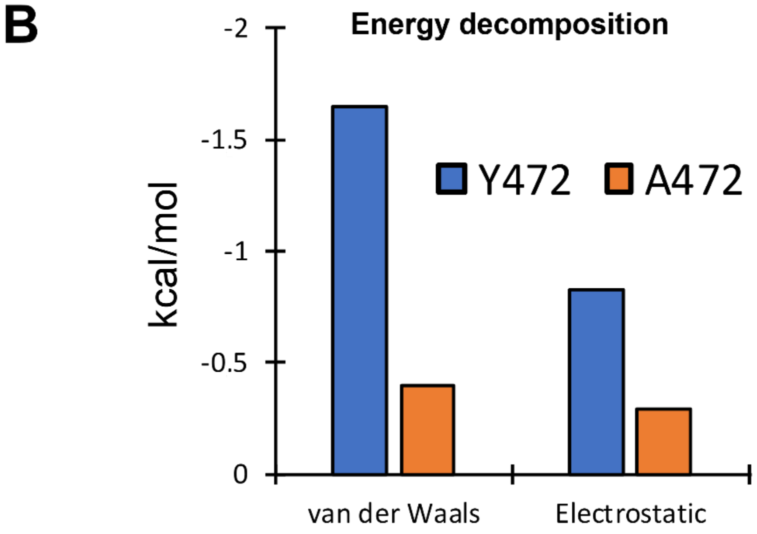


Figure 25. (A) Table of MM-GBSA-Alanine scanning theoretical affinity of ALDH1A3-NR6 complex. (B) Energy contribution of residue 472 to the predicted theoretical affinity of NR6 in the WT (blue) and Y472A mutant (orange) complexes.

4.3. Conclusions

Although the ALDHs family are a validated target for the development of anticancer therapies, recently these enzymes have been shown to be implicated in retinal dystrophies, arousing interest as a possible therapeutic target. Accordingly, compounds able to block their catalytic activity, that could be used in combination therapy with other drugs, may represent an effective strategy to fight RP. Pursuing our interest in the field of ALDH inhibitors, and taking advantage of a structure-based optimization, we succeeded in improving the functional activity of a previously developed hit compound (GA11, IC_{50} 4.7 ± 1.7 μM) [168] obtaining the novel lead NH51, methyl 4-(2-phenylimidazo[1,2-*a*]pyridin-8-yl)benzoate (IC_{50} 0.66 ± 1.3 μM) [169]. Characterized by a distinctive molecular geometry, NH51 proved to bind the catalytic site of ALDH1A3 protruding toward the key C314. Moreover, thanks to the methoxycarbonyl fragment on the 8-phenyl ring, it is able to H-bond the protein backbone, thus settling firmly into the binding site of the target protein. Such a structural complementarity gave reason of its significant functional efficacy. Showing both IC_{50} and K_i values in the sub-micromolar range, NH51 stands out as the most effective ALDH1A3 inhibitor ever described to date. Clearly, thorough investigations in complex structures like cell lines and animal models are now necessary, to prove its robustness as a novel prototype of therapeutic agent against RP. In addition, the potent and selective inhibitor NR6, able to inhibit selectively the ALDH1A3 isoform and to induce cell death, was designed by rational ligand optimization. Compared to the parent compound GA11, NR6 conserves the same heterocycle scaffold with the addition of a cyanide group in the meta position of the 6-phenyl ring. Thanks to the presence of this additional substituent, NR6 binds the protein with a unique arrangement, stabilized by the interaction with several hydrophobic residues. In particular, the interaction with the Y472 residue produces a strong edge-to-face π-π stacking that stabilizes the molecule at the entrance of the catalytic pocket. Even if the site appears partially occupied by the inhibitor, it prevents the binding of the cyclohexene ring belonging to the enzyme natural substrate. In addition, the hydroxyl group of Y472 stabilizes this specific binding mode thanks to a hydrogen bond with the N1 of NR6, that also interacts with a water molecule coordinated with the G136. Interestingly, our sequence analysis demonstrates that Y472 is present in ALDH1A3 and is non-conserved in all the other ALDHs isoforms (Figure 24). This is surprising, in particular for what concerns ALDH1A1 and ALDH1A2 that physiologically metabolize the same substrate, with similar K_M [108,171]. We believe that this residue is responsible for the high specificity of NR6 versus ALDH1A3, and the computational analysis strongly confirmed our hypothesis, providing structural and thermodynamics information that are essential for the design and optimization of small molecule ALDH1A3 inhibitors. Overall, we described the structure of the ALDH1A3/NR6 complex, highlighting a peculiar residue able to guarantee a marked selectivity toward the ALDH1A subfamily. Hence, NR6 represents a starting point for the design and synthesis of more potent ALDH1A3 inhibitor that could be used in combination therapy with small molecular chaperone targeting retinal dystrophies, including RP.

4.4. Materials and methods

4.4.1. Molecular modelling studies on ALDH1A3-NH51 complex

Docking calculations were performed with Gold the program [172], using the CHEMPLP fitness function. The crystallographic structure of ALDH1A3 in complex with MF13 molecule was used as rigid receptor in molecular docking simulations. The specific procedure for ALDH1A3-MF13 complex crystallization is available in Quattrini et al., 2020 [169]. The structure of NH51 was sketched in 2D format using Picto (OpenEye) version 4.4.0.4 [139], converted in 3D coordinates by OMEGA (OpenEye) version 3.1.0.3 [139] and then energy minimized using the MMFF94S force field [173]. GOLD default parameters were used, while docking efficiency was increase to the maximum value. MD simulations were carried out on docking complexes using Amber18 [136]. The ff14SB force field [132] was used to parameterize the protein and the general amber force field 2 (GAFF2) force field [174] was used for compound NH51. Docking complexes were solvated in a rectilinear box of explicit TIP3P water molecules. First, energy minimization was carried out only on water molecules for 5000 steps in which the first 1500 used the SD algorithm, and the last 3500 the CG algorithm. Then, whole systems were energy minimized for 10000 steps in which the first 1500 with the SD algorithm, and the last 8500 with the CG algorithm. Temperature was gradually raised at constant volume from 0 K to 300 K over 1 ns by the Langevin thermostat. Density equilibration was carried out at constant pressure by the Berendsen barostat, for 1 ns with periodic boundaries conditions. Finally, after an equilibration of 50 ns, MD trajectories were produced for of 300 ns. The CPPTRAJ software was used to analyse the MD trajectories [137].

4.4.2. MD simulation on ALDH1A3-NR6 complex

The crystallographic structure of the ALDH1A3-NR6 complex solved in the work Gelardi et al., 2021 was used as starting point in MD simulations. The protein was parametrized by the ff14SB force field [132] while the NR6 molecule was parametrize by the GAFF2 [174]. The complex was then solvated in a rectilinear box of TIP3P water molecules, and the solvent was energy minimized for 5000 steps of which the firs 1500 using the SD algorithm and the last 3500 using the CG algorithm, while keeping the complex frozen. Then, 10000 steps of energy minimization were carried out on the whole system applying the SD for the first 1500 steps, and the GC for the remaining 8500 steps. The system was then heated from 0 K to 300 K over 1 ns at constant volume by the Langevin thermostat, and the system density was equilibrated over 1 ns at constant pressure by the Berendsen barostat. After a 50 ns of preliminary equilibration, 500 ns of unrestrained MD trajectory was produced by the AMBER 18 program [136], and processed by the CPPTRAJ software [137]. To estimate the theoretical affinity of binding of the complex, and the energy contribution of Y472, 100 representative frames were extracted from MD trajectory by cluster analysis and were submitted to MM-GBSA alanine scanning.

Detailed information on the experimental procedures discussed in this chapter, including synthesis, functional evaluation and crystallographic studies on mentioned compounds are available in Quattrini et al., 2020 and Gelardi et al., 2021 [166,169].

5. Side projects

In the framework of an ongoing collaboration with the research group of Prof. Maurizio Zazzi from the Department of Medical Biotechnologies of the University of Siena (Italy), during my doctoral studies I had the opportunity to participate in some projects related to the study of the RdRp of emerging RNA viruses. The RdRp is an essential enzyme for the viral replication process, catalyzing the viral RNA synthesis using a metal ion-dependent mechanism. RdRp is considered as an attractive and validated target for the discovery of novel antiviral drugs for several reasons, resumed as follows: i) RdRp plays a crucial role in the replication cycle of most RNA viruses; ii) it is highly conserved among evolutionary distant RNA viruses [175]; iii) it is absent in mammalian cells as well as no homologous protein are expressed; iv) there is an extensive knowledge on RdRp structure and functions; v) biochemical assays for the rapid screening of large libraries of compounds against the RdRp are easy to be developed, and are consequently available. Nevertheless, the emergence of RdRp drug-resistant variants might limit the broad application of specific inhibitors, or might require their use in combination with drugs directed to other viral targets, if available.

In Picarazzi et al., 2020 [176], we reviewed the main sequence and structural features of the RdRp of emerging RNA viruses such as such the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV), Severe Acute Respiratory Syndrome coronavirus-2 (SARS-CoV-2), Middle East Respiratory Syndrome coronavirus (MERS-CoV), Zika virus (ZIKV), West Nile virus (WNV), Dengue virus (DENV) and Hepatitis C virus (HCV), as well as inhibition strategies implemented so far. Although the number of RNA viruses that have caused outbreaks in recent years is relatively high, we focused on those viruses that we believe might have the strongest clinical relevance, social impact, and scientific interest. Most notably, the current pandemic outbreak of COVID-19 that is due to infection by the SARS-CoV-2 is threatening the health systems on a global scale, and the scientific community is spending unprecedented multidisciplinary efforts to face the urgent need for both treatment and prevention of SARS-CoV-2 infections. On the other hand, HCV is a notable example of how coordinated and extensive research efforts can bring successful results in the development of potent and well tolerated specific antiviral drugs. The approval of sofosbuvir and remdesivir as specific RdRp inhibitors for the treatment of HCV and SARS-CoV-2 infections, respectively, marked a fundamental step in antiviral drug discovery [177,178]. In contrast, no drugs have been licensed against other RNA viruses coding for an RdRp, for example Flaviviruses. The success obtained so far with inhibitors of the RdRp from only two of the many viruses of this kind, probably reflects the disproportionate efforts devoted to these agents compared to others. Based on this evidence, it is highly plausible that drugs acting on the RdRp of Flaviviruses might be developed with success, if multidisciplinary and concerted efforts are dedicated to this task. It is also expected that these drugs will enhance the preparedness of health systems and international organizations to possible future pandemics. In this respect, the lack of structural details on catalytically competent or ligand-bound RdRps from Flaviviruses marks the difference with HCV and SARS-CoV-2, and might limit the successful application of structure-based drug design approaches either established on the repositioning or the conventional strategy.

However, to address the urgent need for anti-Flavivirus therapy, repurposing of licensed antivirals is under evaluation [179–182]. Given the high degree of structural homology observed among RdRp enzymes within the Flaviviridae family [183], sofosbuvir, a nucleotide analog licensed for HCV infection [184], has been recently evaluated as an anti-Flavivirus lead candidate. Indeed, the inhibitory activity of sofosbuvir has been documented *in vitro* against ZIKV and DENV and in mouse models against ZIKV [185–187]. Since the RdRp amino acid residues predicted to interact with sofosbuvir show approximately 80% conservation among WNV, DENV and ZIKV [188], sofosbuvir could also be active against WNV, providing a treatment option by itself or a lead structure for further development.

In Dragoni et al., 2020 [189], we described for the first time sofosbuvir antiviral activity against WNV in the low micromolar range, as well as its genetic barrier through *in vitro* resistance selection experiments. Moreover, we defined the *in vitro* enzymatic activity of sofosbuvir using the purified WNV RdRp. To assess the role of the mutations observed during *in vitro* selection experiments (S604T and M479K/A483G/L721M), an atomistic picture of the interaction between sofosbuvir and WNV RdRp was described by molecular modeling. Given the lack of experimental structural information, the 3D structures of catalytically competent wild type and mutant WNV RdRp forms in complex with RNA, metal ions and ADP were generated by homology modeling, relaxed through MD simulations, and used as receptors in molecular docking simulations. Compared to the wild type WNV RdRp, both mutant forms affected the sofosbuvir binding mode within the catalytic site of RdRp. These binding modes might account for the different efficacy of sofosbuvir against mutant WNV RdRp and highlight the key pharmacophores in WNV RdRp inhibition by small molecules.

In Boccuto et al., 2021 [190], we investigated the ZIKV resistance profile against sofosbuvir through cell-based *in vitro* selection experiments. A double mutant (V360L/V607I) was detected conferring a reduced susceptibility to sofosbuvir. MD simulations confirmed that the double mutant V360L/V607I impacts the binding mode of sofosbuvir, supporting its role in sofosbuvir resistance. By a combination of sequence analysis, phenotypic susceptibility testing, and molecular modeling, we characterized a double ZIKV RdRp mutant with decreased sofosbuvir susceptibility.

Overall, the recent findings in targeting the RdRp of RNA viruses and the data collected in these projects add important information to the profile of sofosbuvir as a possible lead for anti-Flaviviruses drug development, and might drive the design of additional drug candidates, also setting the bases for expanding structure-based approaches.

Finally, in the context of an international collaboration with the Karolinska Institutet of Stockholm, Sweden, I also had the opportunity to participate in a project related to the study of a peptide encoded by SARS-CoV-2 that promotes the activation of NKG2A-expressing NK cells through impaired inhibition. NK cells are innate immune effectors that contribute to host defense against virus infections [191]. NK cells are activated in patients with acute COVID-19, suggesting that they participate in the immune response against SARS-CoV-2 [192–195]. However, by which mechanisms NK cells detect SARS-CoV-2-infected cells and which signals drive their activation remains largely unknown. In this study, we have thoroughly investigated the effect of SARS-CoV-2 peptide on the NK cell response, and we showed that this peptide is presented by the HLA-E and forms stable complexes with it. Molecular modeling studies provided an in-depth understanding of the interactions of the HLA-E/peptide complex, being a valuable tool to improve our understanding of host–pathogen interactions between innate immune cells and SARS-CoV-2. Further information about this research work is available in Hammer et al., 2022 [196].

6. Epilogue

Although new therapeutic strategies are emerging, no pharmacological treatment is available for RP.

In **Section 2**, the extensive use of computational modeling tools allowed the screening of large libraries of compounds to identify small molecules binding the 11-*cis* retinal binding site of P23H rhodopsin, and possibly acting as a molecular chaperone. Preliminary *in vitro* validation of computational results supports the usefulness of molecular chaperones as a pharmacological approach for the treatment of RP whose molecular basis is opsin misfolding and mistrafficking.

Section 3 provided the extensive conformational study of the intrinsically disordered C-terminal tail of rhodopsin with a further insight into the conformational behavior of the C-terminal tail of both WT and mutant rhodopsin forms. Structural details elucidated in this work might facilitate the understanding of the pathological mechanisms of rhodopsin mutants at P347, which will be instrumental to the development of new therapeutic strategies.

In **Section 4**, computational techniques strongly supported the rational optimization of ALDHs, which are involved in the conversion of all-*trans* retinal to retinoic acid in retinal tissue. Starting from the hit inhibitor GA11, potency and selectively toward the 1A3 isoform was notably improved thanks to a combination of biochemical, computational, and structural studies. Experimental gaps in the structural characterization of the binding mode of NH51, and the elucidation of the identified interaction of NR6 within the catalytic site of ALDH1A3 were filled, highlighting the usefulness of the molecular modeling tools in the drug design and drug discovery process.

Section 5 collects the side projects of this thesis. We reviewed the main sequence and structural features of the RdRp of emerging RNA viruses and discussed the main limitations that prevent the successful application of the structure-based drug design approach in antiviral drug discovery process. In this context, the repositioning of sofosbuvir, approved for HCV infections, as an anti-Flaviviruses drug, was proposed through experimental and computational evaluations.

Moreover, we provided computational support to the characterization of the interaction between HLA-E protein and a peptide encoded by SARS-CoV-2 which is involved in immune responses.

In general, the data collected in this work enhance the use of computational techniques in a context of multidisciplinary effort to improve drug design and drug discovery processes in different research fields.

List of publications

1. Hammer et al., 2022. SARS-CoV-2 Nsp13 encodes for an HLA-E-stabilizing peptide that abrogates inhibition of NKG2A-expressing NK cells. **Cell Rep**. 38, 110503
2. **Picarazzi** et al., 2022. Conformational insights into the C-terminal mutations of human rhodopsin in *retinitis pigmentosa*. **J. Mol. Graph.** 108076
3. Boccuto et al., 2021. Sofosbuvir selects for drug-resistant amino acid variants in the Zika virus RNA-dependent RNA-polymerase complex in vitro. **Int. J. Mol. Sci**. 22(5), 2670.
4. Gelardi et al., 2021. A Selective Competitive Inhibitor of Aldehyde Dehydrogenase 1A3 Hinders Cancer Cell Growth, Invasiveness and Stemness In Vitro. **Cancers**. 13(2):356.
5. **Picarazzi** et al., 2020. Targeting the RdRp of Emerging RNA Viruses: The Structure-Based Drug Design Challenge. **Molecules**. 3;25(23):5695.
6. Ballone et al., 2020. Experimental and Computational Druggability Exploration of the 14-3-3ζ/SOS1pS^{1161} PPI Interface. **J. Chem. Inf. Model**. 28;60(12):6555-6565.
7. Quattrini et al., 2020. Progress in the Field of Aldehyde Dehydrogenase Inhibitors: Novel Imidazo[1,2-*a*]pyridines against the 1A Family. **ACS Med. Chem. Lett**. 25;11(5):963-970.
8. Dragoni et al., 2020. Evaluation of sofosbuvir activity and resistance profile against West Nile virus in vitro. **Antiviral Res**. 175:104708.

References

1. Behnen, P. *et al.* A Small Chaperone Improves Folding and Routing of Rhodopsin Mutants Linked to Inherited Blindness. *iScience* **4**, 1–19 (2018).
2. Deretic, D. *et al.* Rhodopsin C terminus, the site of mutations causing retinal disease, regulates trafficking by binding to ADP-ribosylation factor 4 (ARF4). *Proc Natl Acad Sci U S A* **102**, 3301–3306 (2005).
3. Deretic, D. A role for rhodopsin in a signal transduction cascade that regulates membrane trafficking and photoreceptor polarity. *Vision Research* **46**, 4427–4433 (2006).
4. Deretic, D., Lorentzen, E. & Fresquez, T. The ins and outs of the Arf4-based ciliary membrane-targeting complex. *Small GTPases* **0**, 1–12 (2019).
5. Ghirga, F. *et al.* A unique high-diversity natural product collection as a reservoir of new therapeutic leads. *Org. Chem. Front.* **8**, 996–1025 (2021).
6. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* **34**, D668–D672 (2006).
7. Easy compound ordering service - MolPort. https://www.molport.com/shop/index.
8. Miller, B. R. *et al.* MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **8**, 3314–3321 (2012).
9. Hamel, C. Retinitis pigmentosa. *Orphanet Journal of Rare Diseases* **1**, 40 (2006).
10. Jonas, J. B., Schneider, U. & Naumann, G. O. H. Count and density of human retinal photoreceptors. *Graefe's Arch Clin Exp Ophthalmol* **230**, 505–510 (1992).
11. Bowmaker, J. K. & Dartnall, H. J. Visual pigments of rods and cones in a human retina. *J Physiol* **298**, 501–511 (1980).
12. Cote, R. H. Photoreceptor Phosphodiesterase (PDE6): A G-Protein-Activated PDE Regulating Visual Excitation in Rod and Cone Photoreceptor Cells. in *Cyclic Nucleotide Phosphodiesterases in Health and Disease* (CRC Press, 2007).
13. Wensel, T. G. *et al.* Structural and Molecular Bases of Rod Photoreceptor Morphogenesis and Disease. *Prog Retin Eye Res* **55**, 32–51 (2016).

14. Mustafi, D., Engel, A. H. & Palczewski, K. Structure of Cone Photoreceptors. *Prog Retin Eye Res* **28**, 289–302 (2009).
15. Dratz, E. A. & Hargrave, P. A. The structure of rhodopsin and the rod outer segment disk membrane. *Trends in Biochemical Sciences* **8**, 128–131 (1983).
16. Sima, A., Parisotto, M., Mader, S. & Bhat, P. V. Kinetic characterization of recombinant mouse retinal dehydrogenase types 3 and 4 for retinal substrates. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1790**, 1660–1664 (2009).
17. Hsu, A., Tsukamoto, Y., Smith, R. G. & Sterling, P. Functional architecture of primate cone and rod axons. *Vision Research* **38**, 2539–2549 (1998).
18. Hartong, D. T., Berson, E. L. & Dryja, T. P. Retinitis pigmentosa. *The Lancet* **368**, 1795–1809 (2006).
19. Daiger, S. P. Identifying retinal disease genes: how far have we come, how far do we have to go? *Novartis Found Symp* **255**, 17–178 (2004).
20. Lobanova, E. S. *et al.* Increased proteasomal activity supports photoreceptor survival in inherited retinal degeneration. *Nat Commun* **9**, 1738 (2018).
21. Dr, J. I. A.-M., Dr, M. G. P., Dr, M. A.-G., Dr, A. A.-W. & Dr, S. A.-M. Registered Blindness and Low Vision in Kuwait. *Ophthalmic Epidemiology* **12**, 251–257 (2005).
22. Alabdulwahhab, K. M. & Ahmad, M. S. Visual Impairment and Blindness in Saudi Arabia’s School for the Blind: A Cross-Sectional Study. *Clin Optom (Auckl)* **12**, 169–173 (2020).
23. Buch, H. *et al.* Prevalence and causes of visual impairment and blindness among 9980 Scandinavian adults: The Copenhagen City Eye Study. *Ophthalmology* **111**, 53–61 (2004).
24. Tsujikawa, M. Age at Onset Curves of Retinitis Pigmentosa. *Arch Ophthalmol* **126**, 337 (2008).
25. Bunker, C. H., Berson, E. L., Bromley, W. C., Hayes, R. P. & Roderick, T. H. Prevalence of Retinitis Pigmentosa in Maine. *American Journal of Ophthalmology* **97**, 357–365 (1984).
26. GRøNDAHL, J. Estimation of prognosis and prevalence of retinitis pigmentosa and Usher syndrome in Norway. *Clinical Genetics* **31**, 255–264 (1987).
27. Novak-Lauš, K., Kukulj, S., Zorić-Geber, M. & Bastaić, O. Primary Tapetoretinal Dystrophies as the Cause of Blindness and Impaired Vision in the Republic Of Croatia. *Acta clinica Croatica* **41**, 23–25 (2002).

28. Prokisch, H., Hartig, M., Hellinger, R., Meitinger, T. & Rosenberg, T. A Population-Based Epidemiological and Genetic Study of X-Linked Retinitis Pigmentosa. *Investigative Ophthalmology & Visual Science* **48**, 4012–4018 (2007).
29. Mets, M. B., Young, N. M., Pass, A. & Lasky, J. B. Early diagnosis of Usher syndrome in children. *Trans Am Ophthalmol Soc* **98**, 237–245 (2000).
30. Martin-Merida, I. *et al.* Genomic Landscape of Sporadic Retinitis Pigmentosa: Findings from 877 Spanish Cases. *Ophthalmology* **126**, 1181–1188 (2019).
31. Justus, S. *et al. Genome Editing*. 149–162 (Springer, 2016).
32. Geller, A. M. & Sieving, P. A. Assessment of foveal cone photoreceptors in Stargardt's macular dystrophy using a small dot detection task. *Vision Research* **33**, 1509–1524 (1993).
33. Berson, E. L. *et al.* A Randomized Trial of Vitamin A and Vitamin E Supplementation for Retinitis Pigmentosa. *Archives of Ophthalmology* **111**, 761–772 (1993).
34. Chang, S., Vaccarella, L., Olatunji, S., Cebulla, C. & Christoforidis, J. Diagnostic Challenges in Retinitis Pigmentosa: Genotypic Multiplicity and Phenotypic Variability. *Curr Genomics* **12**, 267–275 (2011).
35. Clarke, G. *et al.* A one-hit model of cell death in inherited neuronal degenerations. *Nature* **406**, 195–199 (2000).
36. Sahel, J.-A. *et al.* Partial recovery of visual function in a blind patient after optogenetic therapy. *Nat Med* **27**, 1223–1229 (2021).
37. Bainbridge, J. W. B. *et al.* Effect of Gene Therapy on Visual Function in Leber's Congenital Amaurosis. *New England Journal of Medicine* **358**, 2231–2239 (2008).
38. Maguire, A. M. *et al.* Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial. *Lancet* **374**, 1597–1605 (2009).
39. Graham-Rowe, D. Retinal transplants see fleeting success. *Nature* (2008) doi:10.1038/news.2008.1088.
40. Humayun, M. S. *et al.* Interim Results from the International Trial of Second Sight's Visual Prosthesis. *Ophthalmology* **119**, 779–788 (2012).
41. Stingl, K. *et al.* Artificial vision with wirelessly powered subretinal electronic implant alpha-IMS. *Proc Biol Sci* **280**, 20130077 (2013).

42. Schwartz, S. G., Wang, X., Chavis, P., Kuriyan, A. E. & Abariga, S. A. Vitamin A and fish oils for preventing the progression of retinitis pigmentosa. *Cochrane Database of Systematic Reviews* (2020) doi:10.1002/14651858.CD008428.pub3.

43. Berson, E. L. Long-Term Visual Prognoses in Patients with Retinitis Pigmentosa The Ludwig von Sallmann Lecture. *Exp Eye Res* **85**, 7–14 (2007).

44. Hubbard, R. & Kropf, A. THE ACTION OF LIGHT ON RHODOPSIN*. *Proc Natl Acad Sci U S A* **44**, 130–139 (1958).

45. Fredriksson, R., Lagerström, M. C., Lundin, L.-G. & Schiöth, H. B. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol Pharmacol* **63**, 1256–1272 (2003).

46. Palczewski, K. *et al.* Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science* **289**, 739–745 (2000).

47. Consortium, T. U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506–D515 (2019).

48. Salom, D. *et al.* Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *PNAS* **103**, 16123–16128 (2006).

49. Standfuss, J. *et al.* The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471**, 656–660 (2011).

50. Scheerer, P. *et al.* Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502 (2008).

51. Choe, H.-W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471**, 651–655 (2011).

52. Kang, Y. *et al.* Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**, 561–567 (2015).

53. Zhou, X. E. *et al.* X-ray laser diffraction for structure determination of the rhodopsin-arrestin complex. *Sci Data* **3**, 160021 (2016).

54. Zhou, X. E. *et al.* Identification of Phosphorylation Codes for Arrestin Recruitment by G Protein-Coupled Receptors. *Cell* **170**, 457-469.e13 (2017).

55. JASTRZEBSKA, B., TSYBOVSKY, Y. & PALCZEWSKI, K. Complexes between photoactivated rhodopsin and transducin: progress and questions. *Biochem J* **428**, 1–10 (2010).

56. Jaakola, V.-P., Prilusky, J., Sussman, J. L. & Goldman, A. G protein-coupled receptors show unusual patterns of intrinsic unfolding. *Protein Engineering, Design and Selection* **18**, 103–110 (2005).

57. Rasmussen, S. G. F. *et al.* Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387 (2007).

58. Cherezov, V. *et al.* High-Resolution Crystal Structure of an Engineered Human β2-Adrenergic G Protein–Coupled Receptor. *Science* **318**, 1258–1265 (2007).

59. Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist-β2 adrenoceptor complex. *Nature* **469**, 236–240 (2011).

60. Okada, T. *et al.* The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure. *J Mol Biol* **342**, 571–583 (2004).

61. Sung, C. H. *et al.* Rhodopsin mutations in autosomal dominant retinitis pigmentosa. *PNAS* **88**, 6481–6485 (1991).

62. Al-Maghtheh, M., Gregory, C., Inglehearn, C., Hardcastle, A. & Bhattacharya, S. Rhodopsin mutations in autosomal dominant retinitis pigmentosa. *Human Mutation* **2**, 249–255 (1993).

63. Karnik, S. S. & Khorana, H. G. Assembly of functional rhodopsin requires a disulfide bond between cysteine residues 110 and 187. *Journal of Biological Chemistry* **265**, 17520–17524 (1990).

64. Zhou, X. E., Melcher, K. & Xu, H. E. Structure and activation of rhodopsin. *Acta Pharmacol Sin* **33**, 291–299 (2012).

65. Janz, J. M., Fay, J. F. & Farrens, D. L. Stability of Dark State Rhodopsin Is Mediated by a Conserved Ion Pair in Intradiscal Loop E-2*. *Journal of Biological Chemistry* **278**, 16982–16991 (2003).

66. Ohguro, H. *et al.* Structural and Enzymatic Aspects of Rhodopsin Phosphorylation (∗). *Journal of Biological Chemistry* **271**, 5215–5224 (1996).

67. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H.-W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183–187 (2008).

68. Teller, D. C., Okada, T., Behnke, C. A., Palczewski, K. & Stenkamp, R. E. Advances in Determination of a High-Resolution Three-Dimensional Structure of Rhodopsin, a Model of G-Protein-Coupled Receptors (GPCRs). *Biochemistry* **40**, 7761–7772 (2001).

69. Filipek, S., Stenkamp, R. E., Teller, D. C. & Palczewski, K. G Protein-Coupled Receptor Rhodopsin: A Prospectus. *Annu Rev Physiol* **65**, 851–879 (2003).

70. Okada, T., Ernst, O. P., Palczewski, K. & Hofmann, K. P. Activation of rhodopsin: new insights from structural and biochemical studies. *Trends in Biochemical Sciences* **26**, 318–324 (2001).

71. Menon, S. T., Han, M. & Sakmar, T. P. Rhodopsin: Structural Basis of Molecular Physiology. *Physiological Reviews* **81**, 1659–1688 (2001).

72. Kong, Y. & Karplus, M. The Signaling Pathway of Rhodopsin. *Structure* **15**, 611–623 (2007).

73. Hubbell, W. L., Altenbach, C., Hubbell, C. M. & Khorana, H. G. Rhodopsin structure, dynamics, and activation: A perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking. in *Advances in Protein Chemistry* vol. 63 243–290 (Academic Press, 2003).

74. Hargrave, P. A. Rhodopsin Structure, Function, and Topography The Friedenwald Lecture. *Investigative Ophthalmology & Visual Science* **42**, 3–9 (2001).

75. Athanasiou, D. *et al.* The molecular and cellular basis of rhodopsin retinitis pigmentosa reveals potential strategies for therapy. *Progress in Retinal and Eye Research* **62**, 1–23 (2018).

76. Nickell, S., Park, P. S.-H., Baumeister, W. & Palczewski, K. Three-dimensional architecture of murine rod outer segments determined by cryoelectron tomography. *Journal of Cell Biology* **177**, 917–925 (2007).

77. Lodowski, K. H. *et al.* Signals Governing the Trafficking and Mistrafficking of a Ciliary GPCR, Rhodopsin. *J Neurosci* **33**, 13621–13638 (2013).

78. Fotiadis, D. *et al.* Rhodopsin dimers in native disc membranes. *Nature* **421**, 127–128 (2003).

79. Mendes, H. F., van der Spuy, J., Chapple, J. P. & Cheetham, M. E. Mechanisms of cell death in rhodopsin retinitis pigmentosa: implications for therapy. *Trends in Molecular Medicine* **11**, 177–185 (2005).

80. Berson, E. L., Rosner, B., Weigel-DiFranco, C., Dryja, T. P. & Sandberg, M. A. Disease Progression in Patients with Dominant Retinitis Pigmentosa and Rhodopsin Mutations. *Invest. Ophthalmol. Vis. Sci.* **43**, 3027–3036 (2002).

81. Concepcion, F. & Chen, J. Q344ter Mutation Causes Mislocalization of Rhodopsin Molecules That Are Catalytically Active: A Mouse Model of Q344ter-Induced Retinal Degeneration. *PLOS ONE* **5**, e10904 (2010).

82. Green, E. S., Menz, M. D., LaVail, M. M. & Flannery, J. G. Characterization of rhodopsin mis-sorting and constitutive activation in a transgenic rat model of retinitis pigmentosa. *Investigative Ophthalmology and Visual Science* **41**, 1546–1553 (2000).

83. Tam, B. M., Moritz, O. L., Hurd, L. B. & Papermaster, D. S. Identification of an Outer Segment Targeting Signal in the Cooh Terminus of Rhodopsin Using Transgenic Xenopus laevis. *Journal of Cell Biology* **151**, 1369–1380 (2000).

84. Sung, C. H., Makino, C., Baylor, D. & Nathans, J. A rhodopsin gene mutation responsible for autosomal dominant retinitis pigmentosa results in a protein that is defective in localization to the photoreceptor outer segment. *J. Neurosci.* **14**, 5818–5833 (1994).

85. Deretic, D. & Wang, J. Molecular assemblies that control rhodopsin transport to the cilia. *Vision Res* **75**, 5–10 (2012).

86. Noorwez, S. M. *et al.* Retinoids Assist the Cellular Folding of the Autosomal Dominant Retinitis Pigmentosa Opsin Mutant P23H *. *Journal of Biological Chemistry* **279**, 16278–16284 (2004).

87. Lin, J. H. *et al.* IRE1 Signaling Affects Cell Fate During the Unfolded Protein Response. *Science* **318**, 944–949 (2007).

88. Kaushal, S. & Khorana, H. G. Structure and Function in Rhodopsin. 7. Point Mutations Associated with Autosomal Dominant Retinitis Pigmentosa. *Biochemistry* **33**, 6121–6128 (1994).

89. Tam, B. M. & Moritz, O. L. Characterization of Rhodopsin P23H-Induced Retinal Degeneration in a Xenopus laevis Model of Retinitis Pigmentosa. *Investigative Ophthalmology & Visual Science* **47**, 3234–3241 (2006).

90. Park, P. S.-H. Chapter One - Constitutively Active Rhodopsin and Retinal Disease. in *Advances in Pharmacology* (ed. Tao, Y.-X.) vol. 70 1–36 (Academic Press, 2014).

91. Chen, J. *et al.* Stable Rhodopsin/Arrestin Complex Leads to Retinal Degeneration in a Transgenic Mouse Model of Autosomal Dominant Retinitis Pigmentosa. *J Neurosci* **26**, 11929–11937 (2006).

92. Iannaccone, A. *et al.* Retinitis pigmentosa associated with rhodopsin mutations: Correlation between phenotypic variability and molecular effects. *Vision Research* **46**, 4556–4567 (2006).

93. Aguilà, M. *et al.* Hsp90 inhibition protects against inherited retinal degeneration. *Human Molecular Genetics* **23**, 2164–2175 (2014).

94. Chuang, J.-Z., Vega, C., Jun, W. & Sung, C.-H. Structural and functional impairment of endocytic pathways by retinitis pigmentosa mutant rhodopsin-arrestin complexes. *J Clin Invest* **114**, 131–140 (2004).

95. Andrés, A., Garriga, P. & Manyosa, J. Altered functionality in rhodopsin point mutants associated with retinitis pigmentosa. *Biochemical and Biophysical Research Communications* **303**, 294–301 (2003).

96. Rao, V. R., Cohen, G. B. & Oprian, D. D. Rhodopsin mutation G90D and a molecular mechanism for congenital night blindness. *Nature* **367**, 639–642 (1994).

97. Dizhoor, A. M. *et al.* Night Blindness and the Mechanism of Constitutive Signaling of Mutant G90D Rhodopsin. *J. Neurosci.* **28**, 11662–11672 (2008).

98. Toledo, D. *et al.* Molecular Mechanisms of Disease for Mutations at Gly-90 in Rhodopsin*. *Journal of Biological Chemistry* **286**, 39993–40001 (2011).

99. Saliba, R., Munro, P., Luthert, P. & Cheetham, M. The cellular fate of mutant rhodopsin: Quality control, degradation and aggresome formation. *Journal of cell science* **115**, 2907–18 (2002).

100. Mendes, H. F. & Cheetham, M. E. Pharmacological manipulation of gain-of-function and dominant-negative mechanisms in rhodopsin retinitis pigmentosa. *Human Molecular Genetics* **17**, 3043–3054 (2008).

101. Rajan, R. S. & Kopito, R. R. Suppression of Wild-type Rhodopsin Maturation by Mutants Linked to Autosomal Dominant Retinitis Pigmentosa*. *Journal of Biological Chemistry* **280**, 1284–1291 (2005).

102. Ploier, B. *et al.* Dimerization deficiency of enigmatic retinitis pigmentosa-linked rhodopsin mutants. *Nat Commun* **7**, 12832 (2016).

103. Mizumoto, A. *et al.* Molecular Mechanisms of Acetaldehyde-Mediated Carcinogenesis in Squamous Epithelium. *Int J Mol Sci* **18**, 1943 (2017).

104. Marchitti, S. A., Brocker, C., Stagos, D. & Vasiliou, V. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opinion on Drug Metabolism & Toxicology* **4**, 697–720 (2008).

105. LoPachin, R. M. & Gavin, T. Molecular Mechanisms of Aldehyde Toxicity: A Chemical Perspective. *Chem Res Toxicol* **27**, 1081–1091 (2014).

106. Vasiliou, V., Pappa, A. & Petersen, D. R. Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chemico-Biological Interactions* **129**, 1–19 (2000).

107. Koppaka, V. *et al.* Aldehyde Dehydrogenase Inhibitors: a Comprehensive Review of the Pharmacology, Mechanism of Action, Substrate Specificity, and Clinical Application. *Pharmacol Rev* **64**, 520–539 (2012).

108. Bchini, R., Vasiliou, V., Branlant, G., Talfournier, F. & Rahuel-Clermont, S. Retinoic acid biosynthesis catalyzed by retinal dehydrogenases relies on a rate-limiting conformational transition associated with substrate recognition. *Chemico-Biological Interactions* **202**, 78–84 (2013).

109. Germain, P. *et al.* International Union of Pharmacology. LX. Retinoic Acid Receptors. *Pharmacol Rev* **58**, 712–725 (2006).

110. Johnson, K. J. The retinoids: Biology, chemistry and medicine (second edition) Edited by M B Sporn, A B Roberts and the late D S Goodman. pp 311. Raven Press, NY. 1993. $250.00 ISBN 0-7817-0082-5. *Biochemical Education* **22**, 56–56 (1994).

111. Duester, G. Genetic dissection of retinoid dehydrogenases. *Chemico-Biological Interactions* **130–132**, 469–480 (2001).

112. Maly, I. P., Crotet, V. & Toranelli, M. The so-called ‘testis-specific aldehyde dehydrogenase’ corresponds to type 2 retinaldehyde dehydrogenase in the mouse. *Histochem Cell Biol* **119**, 169–174 (2003).

113. Everts, H. B., King, L. E., Sundberg, J. P. & Ong, D. E. Hair Cycle-Specific Immunolocalization of Retinoic Acid Synthesizing Enzymes Aldh1a2 and Aldh1a3 Indicate Complex Regulation. *J Invest Dermatol* **123**, 258–263 (2004).

114. Everts, H. B., Sundberg, J. P., King, L. E. & Ong, D. E. Immunolocalization of Enzymes, Binding Proteins, and Receptors Sufficient for Retinoic Acid Synthesis and Signaling During the Hair Cycle. *J Invest Dermatol* **127**, 1593–1604 (2007).

115. Dupé, V. *et al.* A newborn lethal defect due to inactivation of retinaldehyde dehydrogenase type 3 is prevented by maternal retinoic acid treatment. *Proc Natl Acad Sci U S A* **100**, 14036–14041 (2003).

116. Yahyavi, M. *et al.* ALDH1A3 loss of function causes bilateral anophthalmia/microphthalmia and hypoplasia of the optic nerve and optic chiasm. *Hum Mol Genet* **22**, 3250–3258 (2013).

117. Lin, S. *et al.* Novel mutations in ALDH1A3 associated with autosomal recessive anophthalmia/microphthalmia, and review of the literature. *BMC Medical Genetics* **19**, 160 (2018).

118. Yan, W. *et al.* Protection of retinal function and morphology in MNU-induced retinitis pigmentosa rats by ALDH2: an in-vivo study. *BMC Ophthalmology* **20**, 55 (2020).

119. Chen, Y., Thompson, D. C., Koppaka, V., Jester, J. V. & Vasiliou, V. Ocular aldehyde dehydrogenases: Protection against ultraviolet damage and maintenance of transparency for vision. *Progress in Retinal and Eye Research* **33**, 28–39 (2013).

120. Moretti, A. *et al.* Crystal structure of human aldehyde dehydrogenase 1A3 complexed with NAD+ and retinoic acid. *Sci Rep* **6**, 35710 (2016).

121. Noorwez, S. M. *et al.* Pharmacological Chaperone-mediated in Vivo Folding and Stabilization of the P23H-opsin Mutant Associated with Autosomal Dominant Retinitis Pigmentosa*. *Journal of Biological Chemistry* **278**, 14442–14450 (2003).

122. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).

123. Platella, C. *et al.* Identification of Effective Anticancer G-Quadruplex-Targeting Chemotypes through the Exploration of a High Diversity Library of Natural Compounds. *Pharmaceutics* **13**, 1611 (2021).

124. Zheng, Y. *et al.* Active Components from Cassia abbreviata Prevent HIV-1 Entry by Distinct Mechanisms of Action. *International Journal of Molecular Sciences* **22**, 5052 (2021).

125. Quaglio, D. *et al.* ent-Beyerane Diterpenes as a Key Platform for the Development of ArnT-Mediated Colistin Resistance Inhibitors. *J. Org. Chem.* **85**, 10891–10901 (2020).

126. Mascarello, A. *et al.* Naturally occurring Diels-Alder-type adducts from Morus nigra as potent inhibitors of Mycobacterium tuberculosis protein tyrosine phosphatase B. *European Journal of Medicinal Chemistry* **144**, 277–288 (2018).

127. Mori, M. *et al.* Structure-Based Identification of HIV-1 Nucleocapsid Protein Inhibitors Active against Wild-Type and Drug-Resistant HIV-1 Strains. *ACS Chem. Biol.* **13**, 253–266 (2018).

128. Mori, M. *et al.* Identification of a novel chalcone derivative that inhibits Notch signaling in T-cell acute lymphoblastic leukemia. *Scientific Reports* **7**, 2213 (2017).

129. GPCR Library - Enamine. https://enamine.net/compound-libraries/targeted-libraries/gpcr-library.

130. Pasqualetto, G. *et al.* Computational Studies towards the Identification of Novel Rhodopsin-Binding Compounds as Chemical Chaperones for Misfolded Opsins. *Molecules* **25**, 4904 (2020).

131. Jacobson, M. P. *et al.* A hierarchical approach to all-atom protein loop prediction. *Proteins: Structure, Function, and Bioinformatics* **55**, 351–367 (2004).

132. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

133. Ferré, N., Cembran, A., Garavelli, M. & Olivucci, M. Complete-active-space self-consistent-field/Amber parameterization of the Lys296–retinal–Glu113 rhodopsin chromophore-counterion system. *Theor Chem Acc* **112**, 335–341 (2004).

134. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859–1865 (2008).

135. Lee, J. *et al.* CHARMM-GUI supports the Amber force fields. *J. Chem. Phys.* **153**, 035103 (2020).

136. CASE, D. A. *et al.* The Amber Biomolecular Simulation Programs. *J Comput Chem* **26**, 1668–1688 (2005).

137. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).

138. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J Chem Inf Model* **50**, 572–584 (2010).

139. Software, O. S. Cheminformatics Software | Molecular Modeling Software | OpenEye Scientific. https://www.eyesopen.com.

140. McGann, M. FRED Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **51**, 578–596 (2011).

141. Friesner, R. A. *et al.* Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **49**, 6177–6196 (2006).

142. Gal, A. *et al.* Pro-347-Arg mutation of the rhodopsin gene in autosomal dominant retinitis pigmentosa. *Genomics* **11**, 468–470 (1991).

143. Dryja, T. P. *et al.* A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature* **343**, 364–366 (1990).

144. Li, T., Snyder, W. K., Olsson, J. E. & Dryja, T. P. Transgenic mice carrying the dominant rhodopsin mutation P347S: Evidence for defective vectorial transport of rhodopsin to the outer segments. *PNAS* **93**, 14176–14181 (1996).

145. Macke, J. P., Hennessey, J. C. & Nathans, J. Rhodopsin mutation proline347-to-alanine in a family with autosomal dominant retinitis pigmentosa indicates an important role for proline at position 347. *Hum Mol Genet* **4**, 775–776 (1995).

146. Dikshit, M. & Agarwal, R. Mutation analysis of codons 345 and 347 of rhodopsin gene in Indian retinitis pigmentosa patients. *J Genet* **80**, 111–116 (2001).

147. Vaithinathan, R., Berson, E. L. & Dryja, T. P. Further Screening of the Rhodopsin Gene in Patients with Autosomal Dominant Retinitis Pigmentosa. *Genomics* **21**, 461–463 (1994).

148. Jose, P. F.-S. *et al.* Prevalence of Rhodopsin mutations in autosomal dominant Retinitis Pigmentosa in Spain: clinical and analytical review in 200 families. *Acta Ophthalmologica* **93**, e38–e44 (2015).

149. Bonilha, V. L. *et al.* Retinal histopathology in eyes from patients with autosomal dominant retinitis pigmentosa caused by rhodopsin mutations. *Graefes Arch Clin Exp Ophthalmol* **253**, 2161–2169 (2015).

150. Flock, T. *et al.* Selectivity determinants of GPCR–G-protein binding. *Nature* **545**, 317–322 (2017).

151. Neves, S. R., Ram, P. T. & Iyengar, R. G Protein Pathways. *Science* **296**, 1636–1639 (2002).

152. Marinissen, M. J. & Gutkind, J. S. G-protein-coupled receptors and signaling networks: emerging paradigms. *Trends in Pharmacological Sciences* **22**, 368–376 (2001).

153. Oldfield, C. J. & Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **83**, 553–584 (2014).

154. Tam, B. M., Xie, G., Oprian, D. D. & Moritz, O. L. Mislocalized Rhodopsin Does Not Require Activation to Cause Retinal Degeneration and Neurite Outgrowth in Xenopus laevis. *J. Neurosci.* **26**, 203–209 (2006).

155. Sung, C.-H. & Chuang, J.-Z. The cell biology of vision. *J Cell Biol* **190**, 953–963 (2010).

156. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105–132 (1982).

157. Vella, F. The cell. A molecular approach; Edited by G H Cooper. pp 673. ASM Press, Washington DC, Sinauer Associates, Sunderland, MA. 1997 ISBN 0-87893-119-8. *Biochemical Education* **26**, 98–99 (1998).

158. Collantes, E. R. & Dunn, W. J. Amino Acid Side Chain Descriptors for Quantitative Structure-Activity Relationship Studies of Peptide Analogs. *J. Med. Chem.* **38**, 2705–2713 (1995).

159. Lin, Z., Long, H., Bo, Z., Wang, Y. & Wu, Y. New descriptors of amino acids and their application to peptide QSAR study. *Peptides* **29**, 1798–1805 (2008).

160. Kereselidze, J., Mikuchadze, G. & Bobokhidze, L. Quantum-Chemical Description of Some Physical-Chemical Properties of Proteinogenic Amino Acids. *J Proteomics Bioinform* **11**, (2018).

161. Lefranc, M.-P. *et al.* IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Research* **43**, D413–D422 (2015).

162. Balasubramanian, R. *et al.* Studies on the Conformation of Amino Acids. *International Journal of Protein Research* **3**, 25–33 (1971).

163. Williamson, M. P. The structure and function of proline-rich regions in proteins. *Biochem J* **297**, 249–260 (1994).

164. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng Des Sel* **12**, 85–94 (1999).

165. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* **40**, W537–W541 (2012).

166. Gelardi, E. L. M. *et al.* A Selective Competitive Inhibitor of Aldehyde Dehydrogenase 1A3 Hinders Cancer Cell Growth, Invasiveness and Stemness In Vitro. *Cancers (Basel)* **13**, (2021).

167. Durinikova, E. *et al.* ALDH1A3 upregulation and spontaneous metastasis formation is associated with acquired chemoresistance in colorectal cancer cells. *BMC Cancer* **18**, (2018).

168. Quattrini, L. *et al.* Imidazo[1,2-a]pyridine Derivatives as Aldehyde Dehydrogenase Inhibitors: Novel Chemotypes to Target Glioblastoma Stem Cells. *J. Med. Chem.* **63**, 4603–4616 (2020).

169. Quattrini, L. *et al.* Progress in the Field of Aldehyde Dehydrogenase Inhibitors: Novel Imidazo[1,2-a]pyridines against the 1A Family. *ACS Med. Chem. Lett.* **11**, 963–970 (2020).

170. Massova, I. & Kollman, P. A. Computational Alanine Scanning To Probe Protein−Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.* **121**, 8133–8143 (1999).

171. Nokin, M.-J. *et al.* Hormetic potential of methylglyoxal, a side-product of glycolysis, in switching tumours from growth to death. *Sci Rep* **7**, 11722 (2017).

172. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking11Edited by F. E. Cohen. *Journal of Molecular Biology* **267**, 727–748 (1997).

173. Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* **17**, 490–519 (1996).

174. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **25**, 1157–1174 (2004).

175. de Farias, S. T., dos Santos Junior, A. P., Rêgo, T. G. & José, M. V. Origin and Evolution of RNA-Dependent RNA Polymerase. *Front. Genet.* **8**, (2017).

176. Picarazzi, F., Vicenti, I., Saladini, F., Zazzi, M. & Mori, M. Targeting the RdRp of Emerging RNA Viruses: The Structure-Based Drug Design Challenge. *Molecules* **25**, 5695 (2020).

177. Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 — Final Report. *New England Journal of Medicine* **0**, null (2020).

178. Reardon, S. United States to approve potent oral drugs for hepatitis C. *Nature* (2013) doi:10.1038/nature.2013.14059.

179. Barrows, N. J. *et al.* A screen of FDA-approved drugs for inhibitors of Zika virus infection. *Cell Host Microbe* **20**, 259–270 (2016).

180. Xu, M. *et al.* Identification of small molecule inhibitors of Zika virus infection and induced neural cell death via a drug repurposing screen. *Nat Med* **22**, 1101–1107 (2016).

181. Giovannoni, F. *et al.* AHR is a Zika virus host factor and a candidate target for antiviral therapy. *Nat Neurosci* **23**, 939–951 (2020).

182. Song, W. *et al.* Repurposing clinical drugs is a promising strategy to discover drugs against Zika virus infection. *Front Med* 1–12 (2020) doi:10.1007/s11684-021-0834-9.

183. Lim, S. P. *et al.* A Crystal Structure of the Dengue Virus Non-structural Protein 5 (NS5) Polymerase Delineates Interdomain Amino Acid Residues That Enhance Its Thermostability and de Novo Initiation Activities *. *Journal of Biological Chemistry* **288**, 31105–31114 (2013).

184. Götte, M. & Feld, J. J. Direct-acting antiviral agents for hepatitis C: structural and mechanistic insights. *Nature Reviews Gastroenterology & Hepatology* **13**, 338–351 (2016).

185. Mesci, P. *et al.* Blocking Zika virus vertical transmission. *Sci Rep* **8**, 1218 (2018).

186. Sacramento, C. Q. *et al.* The clinically approved antiviral drug sofosbuvir inhibits Zika virus replication. *Scientific Reports* **7**, 40920 (2017).

187. Xu, H.-T. *et al.* Evaluation of Sofosbuvir (β-D-2′-deoxy-2′-α-fluoro-2′-β-C-methyluridine) as an inhibitor of Dengue virus replication#. *Sci Rep* **7**, 6345 (2017).

188. Appleby, T. C. *et al.* Structural basis for RNA replication by the hepatitis C virus polymerase. *Science* **347**, 771–775 (2015).

189. Dragoni, F. *et al.* Evaluation of sofosbuvir activity and resistance profile against West Nile virus in vitro. *Antiviral Research* **175**, 104708 (2020).

190. Boccuto, A. *et al.* Sofosbuvir Selects for Drug-Resistant Amino Acid Variants in the Zika Virus RNA-Dependent RNA-Polymerase Complex In Vitro. *Int J Mol Sci* **22**, 2670 (2021).

191. Björkström, N. K., Strunz, B. & Ljunggren, H.-G. Natural killer cells in antiviral immunity. *Nat Rev Immunol* 1–12 (2021) doi:10.1038/s41577-021-00558-3.

192. Osman, M. *et al.* Impaired natural killer cell counts and cytolytic activity in patients with severe COVID-19. *Blood Adv* **4**, 5035–5039 (2020).

193. Mazzoni, A. *et al.* Impaired immune cell cytotoxicity in severe COVID-19 is IL-6 dependent. *J Clin Invest* **130**, 4694–4703 (2020).

194. Zheng, M. *et al.* Functional exhaustion of antiviral lymphocytes in COVID-19 patients. *Cell Mol Immunol* **17**, 533–535 (2020).

195. Bao, C. *et al.* Natural killer cells associated with SARS-CoV-2 viral RNA shedding, antibody response and mortality in COVID-19 patients. *Exp Hematol Oncol* **10**, 5 (2021).

196. Hammer, Q. *et al.* SARS-CoV-2 Nsp13 encodes for an HLA-E-stabilizing peptide that abrogates inhibition of NKG2A-expressing NK cells. *Cell Reports* **38**, 110503 (2022).