Clustering of binding solutions for computer aided drug design

Alejandro Cuello-Rodriguez

Clustering of binding solutions for computer aided drug design

Alejandro Cuello-Rodriguez

A thesis presented for the degree of
Master of Science
2022

Supervised by:
Prof. Andrew Whiting

Chemistry Department
Durham University
United Kingdom

Declaration

The work in this thesis is based on research carried out in the Pohl and Whiting groups at the Chemistry Department of Durham University between November 2019 and January 2021. All of the work is attributed to the author unless referenced to the contrary in the text. No part of this thesis has been submitted elsewhere for any other degree or qualification at this or any other university.

Acknowledgements

To Dave, who had to do this manually. To Ehmke who suggested I learnt to automate it. To Andy, who believed I could do it.

Without the guidance, trust and patience of you all, I would not be where I am today.

Abstract

Clustering of binding solutions for computer aided drug design

Alejandro Cuello-Rodriguez

Amyotrophic lateral sclerosis (ALS) is a disease presenting with severe, progressive neurodegeneration. Binding of the retinoic acid receptors (RAR) and retinoic X receptors (RXR) with retinoids may promote neuroprotection and neurite outgrowth. With potential applications in ALS in mind, a method for screening potential synthetic retinoid drug candidates was developed in which binding complexes for the candidates are predicted in a statistical manner.

In this thesis, the statistical method of binding prediction is called clustering. Drug candidates were docked *in silico* and a python script was developed to group the multiple geometric solutions from docking into clusters of similar looking solutions.

The python script could reproduce the binding poses of synthetic retinoids, for which binding structures have already been resolved, with good accuracy and consistency. This new tool can be used to predict probable geometries of ligands in binding domains that have not yet been solved crystolographically, thus establishing a rapid understanding of how interactions may occur in such complexes without even entering a laboratory.

Abbreviations Used:

- **ALS** Amyotrophic lateral sclerosis
- **AMPA** hydroxy-5-methyl-4-isoxazolepropionic acid
- **ATRA** All-trans retinoic acid
- **CRABP** Cellular retinoic acid binding protein
- **C9ORF72** chromosome 9 open reading frame 72
- **DNA** Deoxyribonucleic acid
- **FTD** Fronto-temporal dementia
- **LBD** Ligand binding domain
- **LBHB** low-barrier hydrogen bonding
- **LMN** Lower motor neuron
- **MM2** molecular mechanics 2
- **PDB** Protein data bank
- **POW** Prisoner of war
- **RAR** Retinoic acid receptor
- **RMSD** root-mean squared deviation
- **RNA** Ribonucleic Acid
- **ROS** Reactive oxygen species
- **RXR** Retinoic X receptor
- **SOD1** Superoxide dismutase 1
- **TARDBP** Transactive response DNA binding protein (gene)
- **TDP-43** Transactive response DNA binding protein 43
- **TNF** Tumor necrosis factor
- **UMN** Upper motor neuron

Contents

1. INTRODUCTION TO ALS

Amyotrophic Lateral Sclerosis (ALS) is a rare subset of neurodegenerative disease but the most common of motor neuron diseases[1]. It was first described in 1874 by Jean-Martin Charcot but received major attention in the U.S. in 1939 when the American baseball player Lou Gehrig, after whom it is sometimes referred, was diagnosed with the condition[2]. It has a prevalence of up to 5 cases per 100,000 people and over 30,000 people in the U.S. are estimated to currently be affected[3].

Diagnosis of this universally fatal disease is difficult in the early stages. A lack of biomarkers, in addition to the many other diseases that can often masquerade as ALS, lends ambiguity to the symptoms[4]. A confirmed diagnosis often takes between 10-18 months after symptomatic onset and proceeds by the elimination of other possibilities[5]. Death usually follows 3-5 years of cognitive and motor decline, with 50% of deaths occurring within 30 months of diagnosis, usually by asphyxiation[4, 5].

ALS presents mostly as limbic onset which affects both upper and lower motor neurons (UMN, LMN)[6]. Bulbar onset, the next most common presentation, affects breathing and causes dysphagia, with limbic decline in the later stages[4]. It can also present as primary lateral sclerosis and progressive muscular atrophy, with pure UMN and LMN involvement respectively.

There are two subtypes to the disease, familial (fALS) which makes up 10% of cases and sporadic (sALS) which makes up the other 90%[6]. The only difference between the subtypes being the inheritance of mutations and the propensity for each gene to mutate and cause ALS. For example, superoxide-dismutase-1 (*sod1*) was the first gene linked to ALS[7] and mutations here account for 20% of familial cases but only 1-2% of the sporadic kind[6]. Both fALS and sALS are clinically identical and share common points of mechanistic convergence. Motor neuron death is universal. It is for this reason that a rodent model of mutant *sod1 (msod1*), while only accounting for 4% of all incidences of ALS, is the most widely used disease model in research[8]. Mutated *sod1* is a reliable initiator of most classical features of ALS, common among its other progenitors and aggressors.

Extensive efforts in search of genetic instigators and propagators of ALS have been high yielding (Fig.1). Yet the drugs approved for treatment to date, i.e. riluzole and Edaravone, only slow the progression and extend life modestly[9]. The effects from edaravone, which are still entirely theoretical and not understood, didn't reach significance in its first trial and only did so in its second when the population sample size was halved to 60[10]. There is no cure.

The varied and interdependent mechanisms which span the disease horizon are vast and so the pathogenesis is poorly understood. One notable shift in clinical trials is the growing emphasis on treatments which target more than one pathway and in the decades since the discovery of SOD1 in 1993, research into the mechanisms underlying ALS has shed light and contributed to the global understanding of other neurological disorders encompassed by dementia, the definitions and borders between which are becoming more synthetic.

Neurodegeneration in its many forms presents a genre of pathologies which remain stubbornly resistant to a relentless procession of scientific effort. This work presents ALS as a disease for which a significant advancement in treatment is long overdue and introduces a novel approach to computational analysis by which the process of lead candidate drug discovery can be eased. But, the approach to drug candidate simulation presented here is not limited to ALS. Instead, this work aims to produce a case study for the potential use of computational analysis to the benefit of any and all protein-target based therapies.

The hypothesis investigated here is the potential targets for ALS are the retinoic acid receptors (RARs). A group of human proteins from the family of nuclear receptors, these targets present a potential therapeutic direction that fills a gap in the current clinical landscape. As directors of gene expression, these ligand-activated targets could potentially impact on multiple aspects of ALS. Yet, they have gone largely unnoticed in the realm of neurodegeneration for some time despite their implementation in a wide range of other diseases including cancers[11].

Recent scientific research reports the anti-apoptotic benefits produced by the activation of these nuclear receptors as well as the thrilling potential for neurite outgrowth[12, 13]. If so, the retinoic acid receptors could be important targets for neuro-regenerative drugs, which would mark an

important advancement in ALS and for other neurodegenerative diseases, enabling a move away from merely managing symptoms, to disease reversal.

i. Genetic mutations and mechanisms of neurodegeneration in ALS

The exact causes of ALS, though well studied, are still poorly understood. How can the disease progress from one of many possible mutations[14-16], to a cascade of mini cellular events and critical motoneuron loss? It is perhaps easier to imagine the disease onset as the result of reaching a tipping point of a delicate balance between pro- and anti-apoptotic processes which underpin cellular homeostasis, rather than a single point event. It is certainly a disease with high collateral damage, that affects all aspects of the cell. Transcription, gene expression, cellular signalling, axonal transport and energetic malfunctions are all present, but initiated in a multitude of ways[16-19].

ii. Superoxide dismutase 1:

Superoxide dismutase 1 (SOD1) is a free radical scavenger whose function is to convert superoxide radicals to the less reactive dioxygen and dihydrogen peroxide. Since mutations of the *sod1* gene were first discovered in 1993, over 180 different mutations in *sod1* have been recorded, the most common being D90A, A4V and G93A. A4V alone is responsible for 50% of the ALS causing SOD1 mutations[20].

Mutations in SOD1 result in the misfolding and subsequent aggregation of the protein[7, 14, 19, 20]. Early thinking on SOD1 toxicity was that their folding would result in loss of function, leading to unchecked production of harmful reactive oxygen species (ROS) that go on to damage motoneurons.[19] While the concentration of ROS increases concurrently with disease progression, retains its role in scavenging radicals to near-usual capacity, instead lending toxicity by instigating an increase in radical production and by physical obstruction of sub-cellular and extra-cellular processes[21, 22].

Glial cells and astrocytes with SOD1 pathology show a marked decrease in glutamate transporter activity[14, 21]. These transporters are responsible for the removal of glutamate, a neurotransmitter, from the postsynaptic neuron. Normal reception of glutamate activates the calcium permeable AMPA receptors, resulting in an influx of Ca^{2+} and the initiation of an action potential. Excess glutamate, however, overwhelms the glutamate receptors. One of the

mechanisms by which this process, known as excitotoxicity, causes neuronal death is the subsequent impact on mitochondria[23].

An increased influx of calcium quickly overwhelms the calcium buffering capacity of mitochondria and when the mitochondrial buffering capacity is exceeded, mitochondria react by releasing aberrant amounts of ROS. Neighbouring mitochondria, affected by this surge in ROS production, react to the oxidative insult with the collapse of their membrane potential and subsequent pore opening and further release of these toxic species in a process called ROS-induced-ROS[17, 24]. In addition to their aggregation in glial cells, *sod1* may actually aggregate on the mitochondria[21], directly inhibiting their calcium buffering capacity. In a positive feedback loop, the large increase in ROS further directs *sod1* aggregation which may occur as a result of oxidative stress.

Mitochondrial dysfunction is a significant point in disease progression that deserves special attention and is not limited to *sod1* pathology. Motor neurons, because of their comparatively great size, are energetically demanding. The dysfunction of mitochondria impacts heavily on other mechanisms such as the expensive process of axonal transport, which in itself therefore cannot participate in the renewal and removal of healthy and damaged mitochondria respectively as it does under usual physiological conditions. As well as the harmful production of ROS by these important organelles, they actively participate in the activation of caspase mediated apoptosis.

iii. TAR DNA-binding protein 43 (TDP-43):

Among the most common of characteristics in ALS is the nuclear depletion and subsequent aggregation of the transactive response DNA binding protein 43, or TAR-DNA binding protein-43[25](TDP-43) . TDP-43 is encoded by the gene *tardbp* and is expressed in nearly all tissues. It resides in the nucleus of the cell where it is involved with the splicing of over 1500 RNAs, synaptic function and RNA stability. Its presence is critical for nuclear function[25, 26].

The interruption of usual TDP-43 activity is not limited to ALS. The group of diseases related to abnormal TDP-43 being referred to as TDP-43 proteinopathies and include frontotemporal lobar degeneration[27]. Its influence in so many areas of neurodegeneration are a consequence of its involvement in many downstream cellular processes.

ALS TDP-43 is recognized by tau-negative TDP-43 inclusions[28]. That being aggregates of TDP-43, and its subsequent depletion, in the absence of aggregated tau proteins which are also associated with neurodegenerative diseases such as Alzheimer's and Parkinson's. Ubiquitinated TDP-43 is found in the cytoplasm where it is misfolded, forming aggregates affecting the mitochondria and endoplasmic reticulum, interfering in their interactions[25,26]. The mitochondria of cells with TDP-43 pathology are malformed and enlarged, massively inhibiting their proper function and is thought to be the result of diminished interaction between TDP-43 and histone deacetylase 6 (HDAC6), a protein involved in the clearance of misfolded proteins[29]. The energetic depletion from dysfunctional mitochondria, as in *sod1* pathology, affects the energetically demanding processes of motoneurons. In addition, the presence of aggregates in the cytoplasm initiate endoplasmic reticular stress[30], which when prolonged results in neuronal cell death and the loss of nuclear signalling by the reduction of nuclear TDP-43 leaves cells helpless to respond to insult.

Mutations in the *tardbp* gene can occur but this itself is rarely the cause for TDP-43 related ALS. Mutations in *tardbp* make up for 3-5% of TDP-43 related ALS pathology; 97% of people with aberrant TDP-43 related ALS do not have this mutation[31]. The movement of TDP-43 to the cytoplasm is reported to be as a result of reduced solubility arising from ROS induced disulphide bond formation. Studies report that when exposed to ROS, TDP-43 was quickly localized to stress granules and lost their vital nuclear function[32]. Loss of TDP-43 from the nucleus then prevents its usual function in neuronal maintenance and protection from the ROS that has caused it's expulsion, furthering neurodegeneration.

iv. C9orf72 Gene:

Mutations resulting in repeat expansion sections of the *c9orf72* gene are the widest reported in ALS (C9-ALS) and also occurs in a subset of fronto-temporal dementia (C9-FTD) cases[33]. The

length of the repeat sections corresponds to disease severity, although with varying and disputed thresholds for disease initiation, ranging from tens to thousands of repeats. Rare cases of C9-ALS and C9-FTD have reported little to no repeat sections in people with neurodegeneration and conversely people with large repeat sections but no neurodegeneration[15] and no TDP-43 aggregation, which is an almost universal feature of C9-ALS.

C9-related toxicity is still poorly understood but two mechanisms are thought to be at play. Dipeptide repeat proteins (DPRs) are a product of repeat-associated translation of repeat transcripts like the expansion sections in *c9orf72* and form toxic cytoplasmic inclusions in a manner similar to TDP-43[33]. RNA foci are also the product of repeat expansions but occur mostly in the nucleus. They result in the formation of abnormal protein structures which sequester RNA binding proteins and become toxic[33, 34]. Like TDP-43, the loss of RNA binding proteins leaves the cell helpless against further insult and render it unable to maintain homeostasis.

Flaws with both mechanisms are reported, though. Some studies observe little to no connection between RNA foci and the clinical features of C9-ALS and others indicate the absence of a relationship between DPRs and neurodegeneration severity[33]. The only strong predictor of neurodegeneration in C9-ALS is the nuclear depletion of TDP-43, which arises sequentially after the appearance of RNA foci and DPRs. The relationship between the three is inconsistent, but TDP-43 offers a target in C9-related therapies.

v. Glia, Inflammation and Neuroprotection

Glia encompass a family of cells including astrocytes, microglia and oligodendrocytes. One of the main roles of glial cells is the protection and maintenance of neuronal and synaptic homeostasis by the production and release of neurotrophic factors[21, 35]. In response to cellular insults such as reactive oxygen species, glia take-on activated states that may lead to the production of pro-inflammatory species. The action of activated microglia depends on their activation state. M1 activation is characterised by the production of inflammatory cytokines

such as IL-1β and TNF-α. The M2 state instead drives anti-inflammatory neurotrophic factor release and promotes phagocytosis to clear abnormal aggregates and misfolded proteins[36].

Neuronal loss from activated glial cells is likely a combination of loss of homeostatic support and toxic gain of function mechanisms. Astrocytes have been shown to be activated by both *sod1* and TDP-43 and degenerate in their presence[21, 37]. Astrocytes with *sod1* pathology demonstrate reduced glutamate transporter activity and fail to support the postsynaptic neuron from excitotoxic death. When activated, their release of pro-inflammatory molecules is initially neuroprotective, but prolonged upregulation of these compounds becomes toxic. INF-γ directly induces motoneuron death and TNF-α is involved in necroptosis, a form of cell death, by interaction with specific kinases. TNF-α also increases the number of AMPA receptors in the postsynaptic neuron and their sensitivity to glutamate[38], contributing to an environment favouring excitotoxic neuronal death.

Therapies related to the glial mediated mechanisms of ALS alternate between the replacement of toxic, faulty astrocytes with mesenchymal stem cell replacements and the production of neurotrophic factors such as Bcl-2, demonstrated to improve mitochondrial function[39]. ALS mice demonstrate much reduced levels of neurotrophic factors which maintain the structural integrity of motoneurons and their synapses. Improving the production of these factors delays neurodegeneration and is a highly pursued avenue of therapy. Although, relative to other avenues of investigation, biological therapies are much fewer in number and none have been approved.

vi. Axonal Transport:

Altered axonal transport is a feature common among all forms of ALS. Microtubules, an essential component of neurons, exist in a state of constant equilibrium between assembly and disassembly known as dynamic instability[18]. The transport of materials is necessary for the survival of neurons, but axons are highly sensitive to cellular insult such as increased ROS and toxic aggregates[40]. Transport occurs in the intermission between assembly and disassembly but, in response to insult, the rest period is reduced and the switch between both processes is accelerated and known as hypertubule dynamics.

The energy required to move material by active transport across the distances spanned by motoneurons is massive. For this reason, mitochondrial dysfunction was primarily suspected to dramatically affect axonal transport, and while loss of energy from this damaged organelle will of course impact on transport, recent evidence shows altered transport before the onset of mitochondrial dysfunction in ALS[18]. Nevertheless, loss of energy production and dysfunctional mitochondria do produce large amounts of ROS which is known to instigate hypertubule dynamics. The reduction of this hyperactivity is the objective of therapies related to axonal transport.

2. CLINICAL LANDSCAPE

Clinical data surveying the different neurodegenerative diseases and therapeutics used to treat them, past, present and planned is available from numerous databases[41]. Understanding the current clinical landscape is important for anyone wishing to contribute to it, but no extensive reviews of ALS therapies exist to date. Here we present a brief overview of ALS from the review of one such database, which provides information on the varying stages of clinical trials in therapies aimed at ALS. The range of progress in these endeavours ranged from early discovery all the way to Phase III clinical trials.

i. Disease Overview:

The mechanisms related to disease progression in ALS are abundant and highly complex Lending to the complexity is the way these so-called mechanisms impact on each other and compound progression. While there are many ways in which ALS may be initiated, the mechanisms derived from these routes frequently converge. Endoplasmic reticular stress, the mis-localisation of nuclear substrates and their subsequent aggregation, loss of neurotrophic support, mitochondrial dysfunction and diminished axonal transport are just some of these consistent points of convergence.

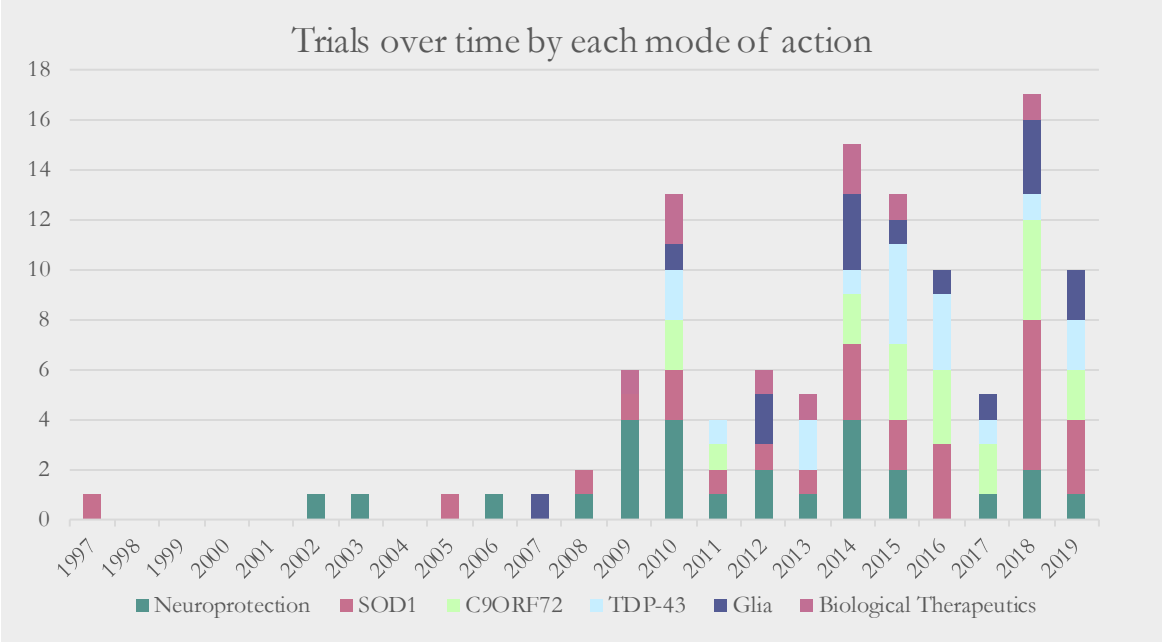


Figure 1 Neuroprotective trials are most consistent in terms of frequency over time. C9orf72 and TDP-43 related trials increased from 2010. sod1 therapies relate to those targeting the clearance of the misfolded protein and stimulation of the gene. Biological therapeutics consist mostly of methods of replenishing the function of glial cells by using mesenchymal stem cells.

Simplistically, the mechanistic targets of trials for the treatment of ALS fall under 6 broad categories, neuroprotection, *sod1*, *c9orf72*, TDP-43, glia and biological therapeutics. However, as must be stressed, the mechanisms in pathology and therefore in treatment constantly overlap and it is not possible in most of the trials in the database to designate the therapy or treatment to a single pathological mechanism. It is therefore entirely possible that the mechanisms the trials are associated with, as declared by the authors of the therapy, may alter as understanding of the disease mechanisms increases with time.

A general trend may be observed, nethertheless, that the number of trials relating to ALS is increasing. Between the years 1997 and 2009, there were 14 trials recorded for ALS directed therapies, whereas, between the years 2010 and 2011, 17 such trials were initiated. However, considering that in this time only two therapies (edaravone and riluzole, chapter 1) have been approved, the urgent need for new ideas and therapies becomes apparent.

ii. Types of Therapies

As all of the therapies relating to ALS at some point all desire to reduce the damage to motoneurons, they could all be considered "neuroprotective". However, the term neuroprotective therapies refers specifically to the stimulation or upregulation of neuroprotective growth factors such as the brain-derived neurotrophic factor or the insulin-like growth factor-1, which both exhibit neurotrophic properties[42].

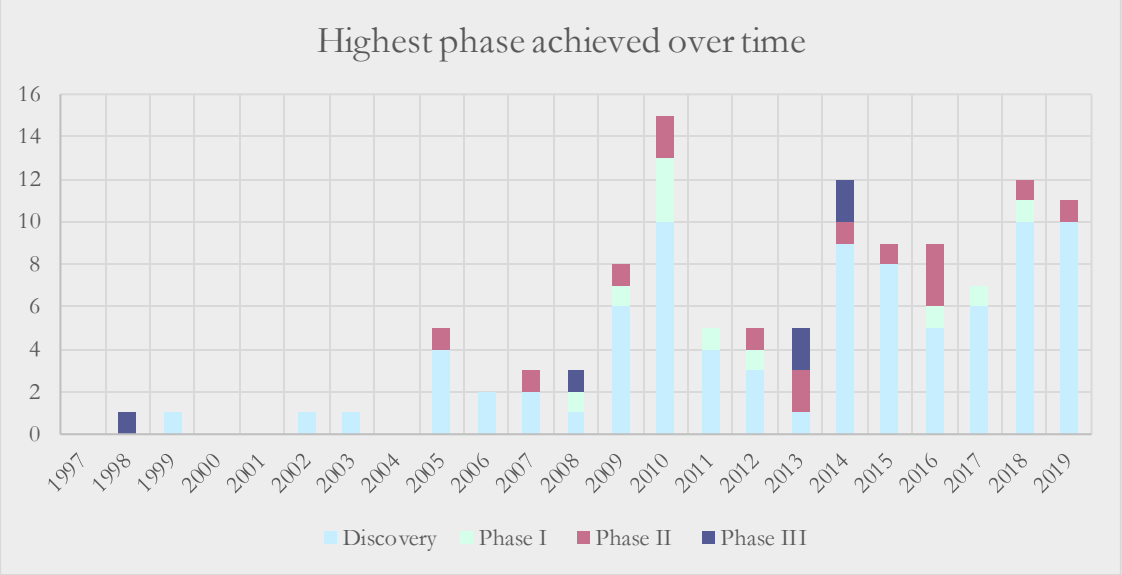


Figure 2 Despite the initial increase in trials, the pace of new trials has leveled off and no increase has occurred for the last 10 years. Very few efforts have reached phase III overall and none have reached phase III in the last 6 years .

Neuroprotective therapies are the most frequently undertaken and a good portion reach phase II and III clinical trials.

sod1 related therapies vary, with some targeting the aggregates by stimulating aggregate clearance mechanisms. Others seek to improve mitochondrial health or stimulate the *sod1* gene to produce more of the protein to counter abberant ROS production. These therapies have the greatest success of reaching Phase III clinical trials but also have the most terminations at discovery.

C9orf72 trials havent reached Phase III. These therapies mostly target the C9 gene. TDP-43 trials, however, have much greater success at reaching Phases II and III and aim to prevent mitochndrial damage, and reduce mislocalisation and aggregation.

Glial therapies and biological therapeutics mostly overlap, but biological therapeutics refer to the attempted replacement of glial cells by neutrotrophic-factor-rich stem cells. Glial therapies in comparison refer to the attempted reduction of pro-inflammatory compounds and the attempted stimulation of neurotrophic factor release by astrocytes.

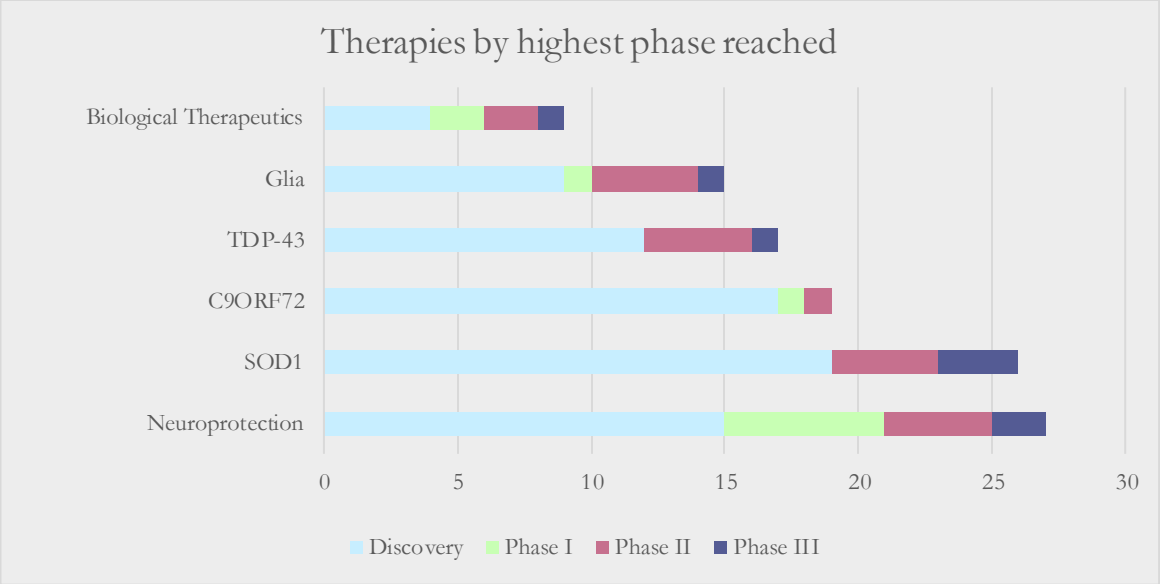


Figure 3 sod1 has the most at discovery and the most to reach phase III. No C9 related therapies have reached phase III yet.

While some of the therapies and trials for ALS succeed in the manipulation of one disease mechanism, the therapies fail at dramatically reducing the progression of disease or amelioration of symptoms to the benefit of remaining quality of life for patients. The ideal therapy must target the disease on multiple fronts, however, multiple drug treatments carry the risks associated with polypharmacy and lend further complexity to an already complex, multifaceted disease. Hence, there is a major unmet need for a single therapy which simultaneously targets many of the pathological junctions which when crossed, rapidly progress the disease

3. RETINOIC ACID RECEPTORS

i. Vitamin A and nuclear receptors

Historically, retinoids have been used for the treatment of a wide-range of diseases. Their use dates back thousands of years but, with the discovery of vitamin A, boomed in the early 20th century and later in World War II when in POW camps, hens were kept so that their eggs, a concentrated source of vitamin A, could be used to prevent and treat night blindness[43]. More recently, retinoids are used to treat a variety of cancers and skin conditions such as photoaging and acne[44].

Vitamin A (retinol) and its derivatives exert their biological effects through binding to their complimentary retinoic acid receptors (RARs) and retinoic X receptors (RXRs)[45]. For both the RARs and RXRs, there are three different isotypes, α, β and γ, so there are 6 receptors in total. The RARs consist of 6 sections, A-F and the RXRs 5, A-E. Section E contains the ligand binding domain (LBD) for both receptor types and this domain is where interactions occur between the retinoid and the protein[45]. It is also the site at which these proteins form dimers before binding to promotor regions on DNA, which occurs at section C. The structures of the two nuclear receptor types are quite well conserved but their LBDs vary significantly in shape. The RARs assume a more linear cavity whereas the RXRs are shorter and have a bend in the LBD mid-section[46]. This uniqueness in LBD brings about the opportunity to design compounds selective for a specific receptor that can direct different biological outcomes.

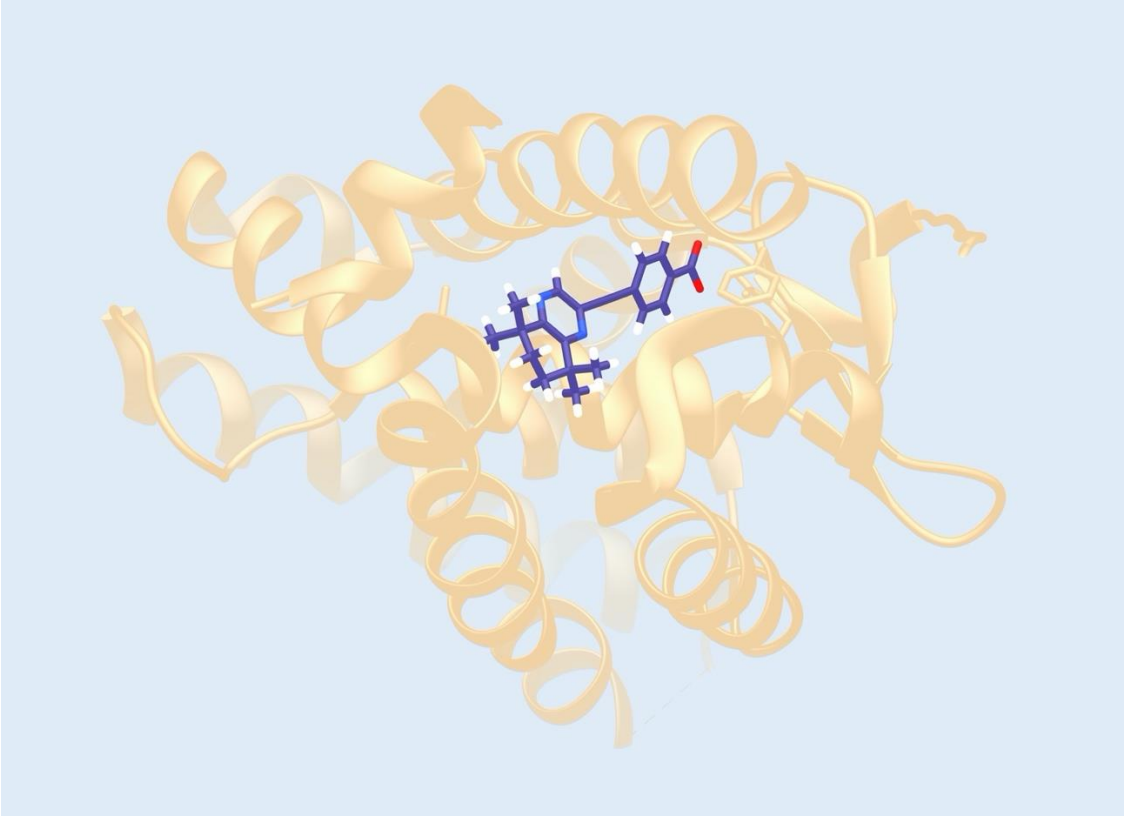


Figure 4 Image of a synthetic retinoid in retinoic acid receptor beta. Visualized with UCSF chimera.

As their name suggests, these nuclear receptors reside in the nucleus of the cell where they exert their influence on cellular processes such as differentiation, homeostasis and apoptosis, through transcriptional processes. Before binding to DNA-response elements, RARs must form heterodimers with an RXR and for this to happen, both the RAR and RXR must already be complexed with a retinoid ligand[47]. The necessity of ligand binding is due to the structural changes retinoids exert on their receptors.

Retinoids are drawn towards polar clusters at the base of the LBD in these proteins. Once the retinoid is secured by the interaction of its carboxylate group with this cluster, a hinge at section D swings across to encapsulate the ligand inside the binding pocket[48]. Images produced of bound RARs by X-ray diffraction show the ligands completely buried in the protein. It is this conformational change, i.e. the closing of the binding pocket, that produces a geometry favourable to dimerisation and subsequent association with DNA.

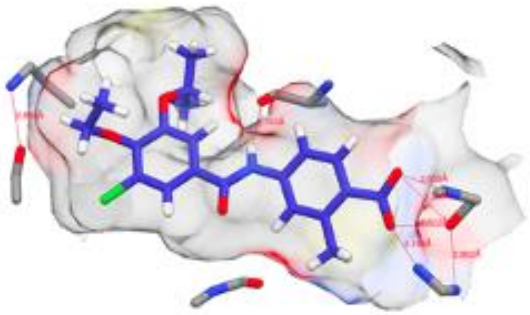


Figure 5 Interactions between the carboxylate group and the polar cluster at the bottom of the binding surface in RARα (PDB: 3KMR).

It has been repeatedly shown, by selectivity, that distinct retinoids can exert potently different downstream biological effects[49]. The unique shape of a ligand can in-turn bring about a unique folded shape of the receptor. If two receptors with distinct shape influenced by this ligand then dimerize, the resulting heterodimer can be uniquely different to that of the same two monomers complexed with a different ligand. The Whiting-Pohl group believes that these small and subtle changes in ligand geometry, size and shape may encourage larger, more amplified changes in the resulting dimer and ultimately increase or decrease the biological response exerted by its subsequent transcription.

This adds further complexity to understanding the interplay between selectivity and strength of activity. In this way, it is not only important to activate a specific retinoic receptor to achieve a desired response, but the receptor must also be activated in such a way that its subsequent heterodimer is shaped in a manner conducive to a specific activity. This can be taken one step further if more than one retinoid is introduced to the system in a competitive ratio. The existence of two compounds in equal proportions can yield 15 distinct heterodimer combinations and shapes across all six of the RARs and RXRs.

This theory can account for the phenomena of the different activities resulting from two ligands selective for the same receptor. Nuclear receptors in this sense can be thought of as generic enough in shape to influence a vast variety of processes, but with the careful curation of heterodimers, they can be made more specific and potent.

ii. Role in Neurodegeneration

Retinoic acid is associated with neurodegenerative disease via numerous cell signalling pathways. In addition to enhancing *sod1* levels and promoting the transcription levels of numerous growth factors, it was discovered that retinoic acid has the ability to generate neurite outgrowth[12, 50].

This outgrowth is seemingly dependent upon the upregulation and activation of RARβ[12], although RARα is implicated in the regulation of neuronal survival. In mouse models, increases in RARβ in the spinal cord induces increased neuronal survival, whereas mice fed a retinoid-free diet undergo similar pathology to ALS. The induction of neurite outgrowth, so desired in ALS since it could mimic neural regeneration, can only be achieved, however, by the activation of both genomic and non-genomic activities by retinoids[51].

Among the transcriptional effects of retinoid signalling, data from the Whiting-Pohl group has found the downregulation of pro-inflammatory elements TNF-α and IL-1β by the novel synthetic retinoid DC645[52]. In addition, the same retinoid strongly upregulated the gene ADAM10, a prerequisite for myelination and when upregulated, critical to axonal outgrowth[53].

Retinoids then present a treatment option which, by activation and regulation of their receptors, can direct large amounts of biological influence and transcriptional action. If the survival and regrowth of axons possible in rodent models is transferrable to humans, retinoids may represent a dramatic and highly impactful step in a new direction for not just ALS, but all neurodegenerative diseases.

Considering the significant implications of a retinoid mediated approaches to neurodegeneration and ALS, the chapters herein illustrate the potential influence and advancement of retinoid design by computational aid methods.

4. AN INTRODUCTION TO COMPUTATIONAL SIMULATION

Retinoic acid receptors (RARs) are targets for drug design, and therefore, it is necessary to design drugs aimed at eliciting a strong therapeutic response by way of strong interactions with the different RAR binding sites. For a drug to be effective, it must form a protein-ligand complex after entering the protein and finding a suitable binding position that is energetically favourable. This favourable state is denoted by the Gibbs free energy of the complex and is the product of the energy of the complex, enthalpy and distribution of this enthalpy and entropy, across the whole complex between the drug and the protein target. While enthalpy and entropy can be seen as directors of binding, commonly the strength of binding can also experimentally determined and its magnitude estimated, through the direct measurement of either the binding constant (Kb) (also known as the association constant (Ka)) or the inverse of this, the dissociation (Kd) constant, from which gibbs free energy can be derived[54].

$$\Delta G = \Delta H - T\Delta s$$

$$\Delta G^{\circ} = -\mathrm{R}T\ln \mathrm{K_b}$$

Figure 6 Change in Gibbs free energy is a determiner of binding and is calculated by considering the change in enthalpy and entropy. Bottom equation shows relationship between Gibbs free energy and binding constant.

i. Molecular Dynamics

The Gibbs free energy of a system can also be computationally calculated with high level Molecular Dynamics (MD) simulations, albeit exact values are still unattainable[55]. These calculations deliver energetic profiles of the protein ligand complex informing on aspects such as their binding energy and contributions to that energy like water, solvents, hydrogen bonding and conformational explorations of the protein and ligand during binding. Proteins and ligands possess multiple degrees of structural freedom and consequently move during binding[56].

Structural changes in proteins like the RARs can be initiated by the binding ligand, and the ligand itself can undergo conformational dynamics through rotation about its bonds within the protein to adopt different binding modes, known as conformers[55, 56]. Each conformer of the same compound will have differing interactions with the protein binding surface, the energy of each must be calculated to determine the optimum binding mode.

Advances in MD simulations correlate with the exponential progression in available computing power[57] and, in great part, with the conception of force-fields replacing more complex quantum mechanics in the calculation of, firstly small and then larger biological systems and chemical reactions. The latter winning a Nobel prize[58]. With appropriate power, simulations of systems containing 500,000 atoms are nowadays achievable[56].

While experimental methods can offer accurate binding information, the investigation of this is lengthy and not appropriate for screening masses of compounds[55]. Each ligand is the product of days, weeks and even months of design and synthesis; a process which is very expensive both in terms of time and resources. Computational analysis offers the potential of being able to simulate the interactions of thousands of compounds without ever having to synthesise anything. However, high-level MD analysis of the ligand-protein interactions is computationally intense and requires some of the largest computational facilities available to undertake effectively and on scale. Researchers rarely have access to the type computing power required to perform such analysis, and even when facilities are available, booking access at such facilities involves lengthy facilities applications rounds and time and resources allocating systems.

ii. Desktop Docking

Fortunately, MD simulations are not the only computational options available. Desktop binding software such as GOLD and AUTODOCK[59] are much more accessible, offering visualisation of ligand-protein interactions within minutes along with binding modes, ranked with fitness scores attributed to their relationship[57, 59]. However, computational analysis is a trade-off between speed and accuracy. MD simulations study the ligand protein complex as a whole,

dynamic system and calculate the corresponding energies for each compound and probe temporal, conformational and orientation effects, whereas desktop docking software will consider both the ligand and binding pocket as relatively rigid objects[57], i.e. there is no conformational or dynamic exploration of either the ligand, or the protein, or the protein by the ligand, and the protein remains in the same geometry as derived from its crystal structure, accessed from the Protein Data Bank (PDB). Hence, the process involves the ligands being 'placed' inside the binding domain of the protein structure from the PDB, and is then 'docked' with the protein in a two-step process of searching for orientations followed by comparing and scoring of the results[57, 59].

The searching process investigates the possible conformers of the ligand within the binding site typically using a genetic algorithm, which builds a user defined population. The members of this population consist of the same compound with varying differences in conformation and each member is assigned a fitness score based upon its interactions with the binding site of the protein. The members of this population are then progressively optimised[47] by the "mating" of members and the offspring can spontaneously "mutate" to develop geometrical changes and hence, produce new conformers with their own fitness scores. The result of this, is that for each compound inserted into a ligand binding domain or site, the combined structural input file is run by the software and a series of solutions are calculated as a result. Each solution consists of a possible way in which the compound might interact with the protein receptor or binding site, and they differ often by only small rotational changes to the conformation of the compound and orientation within the binding site. A fitness score is then also calculated for each solution, as a measure of how strong the interactions of the ligand, in that conformation and orientation, are with the protein binding domain.

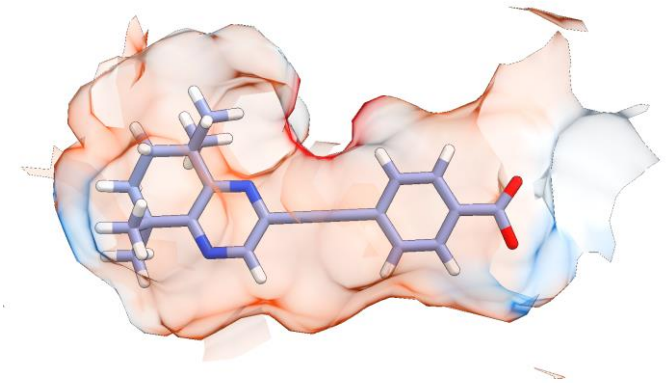


Figure 7 Image of a synthetic retinoid within binding site of RAR-alpha. Hydrophilic sections marked in blue and lipophilic in red. This is a single solution from a docking run, for any given run there can be as few as 3 solutions and as many as over 100.

One limitation to such software is the accuracy of the fitness score[55]. A compromise for accessibility is reached in which simplifications, like protein and ligand rigidity, reduce the reliability of the score. Crystal structures of proteins from X-ray diffraction come with bound ligands. A test of the accuracy of fitness scores is the redocking of the ligand which came with the protein to see if the same conformation is adopted by the solutions. Although a conformation consistent with the crystal structure is usually produced, this solution is often not the highest scoring one. If the objective of computational analysis is to determine the candidacy of drugs for investigation, the software needs to inform the user of the most likely way the drug might interact with the protein. Desktop analysis may offer an array of possible interactions with the protein, but manually inspecting each of the many conformations and orientations explored by the solutions is time consuming and discrimination between potentially good-looking candidates is difficult when many possible conformers are obscured by a number of both good and bad interactions with no indication as to which could actually be useful predictions or mimics of real ligand-receptor interactions and therefore of use for ongoing studies and potential use in drug design and development.

iii. Clustering

One useful approach to analysing these many docking solutions may lie in their grouping into structurally closely related conformations. Usually in a docking run, a user would upload one structure file for the compound in question and up to 10 solutions would usually be given by the software, in this case GOLD. A study by the Whiting-Pohl group has suggested that instead of docking one geometric instance of a compound, the user should dock multiple structure files of the same compound, i.e. in different starting conformations. We theorised that if a compound is docked with a protein using a large array of unique starting conformers and one conformation predicted by the docking programme is adopted by 90% of all solutions, whilst other conformers make up the remaining 10%, it is reasonable to suggest that this dominant, or "key" conformation warrants investigation and considering as the most preferred solution and potentially a real-world solution as well. Although not all compounds exhibit such extreme preferences for a single key conformation over others, the study by Whiting-Pohl group consistently established dominant key conformers for a series of synthetic retinoids with their complimentary retinoic acid receptors, many of which corresponded to the types of ligand-protein structures found from X-ray analysis.

Building on this work done by the Whiting-Pohl group, a clustering algorithm using Python, 18ct.py[60], was developed, automating the grouping process which was manually undertaken before. The automation is potentially important for numerous reasons, however, perhaps the primary being that automating the process offloads the laborious and tedious work of manually clustering the solutions, and is therefore much faster than the somewhat involved process of viewing each one by eye. Hence, the clustering approach can be queued overnight, or run quite reasonably in a matter of minutes and hours rather than days, as before. Secondly, with a specific computational threshold for similarity, discrimination between all conformers is accurate and repeatable, i.e. the consistency and reliability of the resulting analysis can be assured.

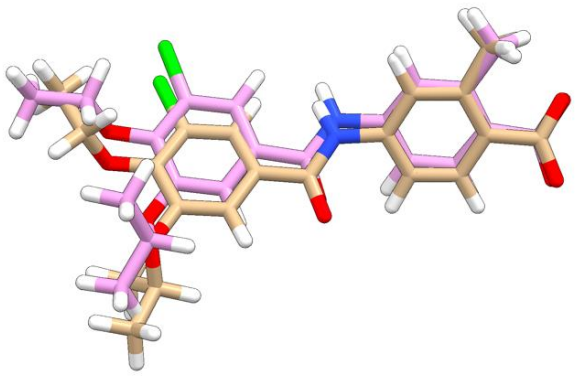


Figure 8 Two generic solutions are superimposed by the algorithm and their root-mean-squared-deviation is calculated as a measure of their similarity.

The score for similarity is calculated by a root-mean-squared-deviation equation (RMSD) between solutions. RMSD is a frequently used approach for structural comparison but is sometimes criticised because of algorithmic subtleties with structural superimposition, i.e. compounds with flipped symmetrical rings can be deemed to be structurally different despite being actually geometrically identical. The algorithm uses the RMSD python module from PyPi, created by Jimmy Kormann and Lars Bratholm, and their module utilises the Kabsch and Quaternion algorithm for superimposition and RMSD calculation. The Hungarian algorithm is also used for atom reordering and adaptations, which was used in order to identify flipped rings and minimise computational errors in the RMSD calculation. The algorithm was then optimised for use with GOLD but is nevertheless still quite compatible with other docking software.

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}$$

Figure 9 RMSD algorithm incorporated into the python script.

The code itself works by looping through all solutions from a docking experiment in an iterative pairwise fashion such that every possible combination of pairs is considered. In a list of

solutions *a*-*z*, solutions *b*-*z* are compared with *a* one-by-one. If the RMSD between the two solutions is within the user defined threshold, the two solutions are considered the same conformer and a cluster for *a* grows by 1. The process then repeats for *b* and *c* and so on until all solutions are assigned to a distinct conformer cluster. The largest conformer cluster at the end is the one with the most members of solutions. This is the Key conformer.

5. METHODS

i. Ligand preparation

All compounds (retinoid and synthetic retinoids) in this work were drawn in 2D using ChemDraw Prime and opened as 3D compounds with Spartan 18. Equilibrium geometries were calculated, for all compounds, using semi-empirical methods (AM1, gas-phase calculations). A conformer distribution was created from each equilibrium structure with MM2 methods. The number of conformers was limited to 50. The resulting lists of conformers were saved as .mol2 files and subsequently edited to exhibit carboxylate moieties rather than their carboxylic acid derivatives, which was necessary for compatibility with the next process, i.e. docking.

ii. Docking

Suitable protein structures were identified from the protein data bank and downloaded as PDB files (RARα, RARγ, RARβ files were 3KMR[61], 1XAP[62] and 2LBD[63] respectively. CRABPII files were 2FR3[64] and 2G78[65]). The protein structure files were then opened using the Hermes client from GOLD, the docking software programme with which all retinoid ligands were to be examined for their binding capabilities with the different protein receptor structures. The protein structures were protonated and water and existing bound ligands were removed to create the unbound receptor protein model. The original bound ligand resident in the PDB structure was noted as a positional reference to enable identification of the binding domain and hence, all atoms and residues within 8 Å (6 Å for CRABPII to limit the actionable surface in an otherwise large cavity) were selected to create a model to represent the binding cavity. Chemscore (within GOLD) was selected as the target analysis function, with a search efficiency of 200% and each retinoid was then docked into these binding domains. The number of solutions for each retinoid ligand was not limited and they were allowed full flexibility.

iii. Clustering and visualisation

Solution files from docking were collected and sorted using the clustering algorithm 18ct.py written by me. An RMSD value of 1.20 was used as a balance between speed and accuracy for the clustering process. A lower threshold RMSD value corresponds to a stricter regimen for grouping, resulting in a greater number of unique conformers, which is not necessarily useful as the difference energetically between the conformers decreases to the point that their interaction with the binding domain may be considered congruent and hence a higher threshold would be equally suitable and computationally faster as there are fewer total conformations to compute. The Key (most frequently occurring) conformer was then examined within the ligand binding domain (LBD) through visualisation using UCSF Chimera. The interaction pocket and relevant residues involved in the complex were identified using chimera's standard surface identification. Polar residues within 5A were kept in order to observe any hydrogen bond interactions.

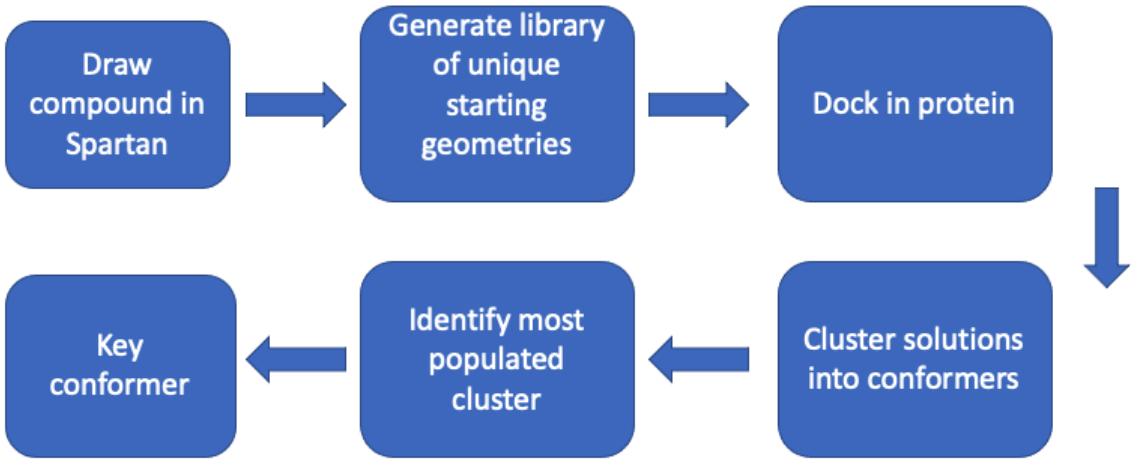


Figure 10 schema of process. The most frequently occurring conformer is known as the key conformer.

6. CLUSTERING VS CHEMSCORE ONLY

To validate this method, ligands were removed from the x-ray crystallographic .pdb files and redocked into their respective proteins using the methods in 6i-iii. This was done to compare the reliability of the Key conformer (most frequent solution derived from clustering) and highest scoring solution (from GOLD) in predicting the known bound conformation from the x-ray crystallography. The best method would give a prediction which matches the known ligand geometry with the highest accuracy.

The scoring function from GOLD, Chemscore, was chosen for it's complex considerations, such as hydrophobic-hydrophobic contact areas, hydrogen bonding and ligand flexibility, but also its speed[66].

If the key conformer can reliably predict the known bound conformation of ligands in their respective PDB structures, then the predicted key conformer for compounds with no known bound solution can be considered a reliable and reasonable starting place for further research. In this validation step, the accuracy of the predicted conformer was determined by the superimposition of the key conformer with the known binding conformer.

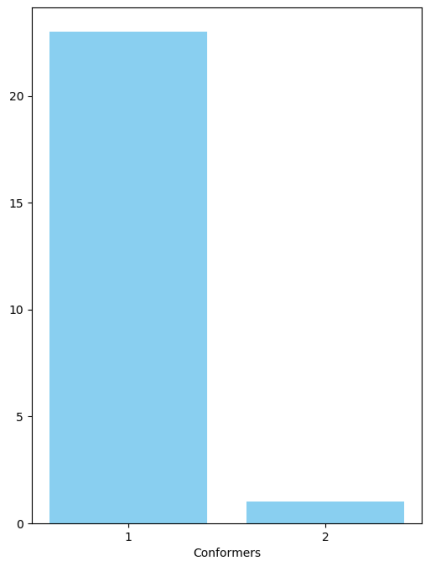
Here, the bound ligands from crystallographic structures on the pdb are put through the generic process detailed in part i. The names and chemistry of the ligands are deliberately omitted, they exert no influence on the function of the algorithm, instead the name of the whole complex from the PDB is given. What is noteworthy here is the ability of the algorithm to take solutions of any ligand, regardless of shape size or chemistry, and identify the most frequently occurring geometry from the solutions of a docking run. It is this most frequently occurring geometry, or Key conformer, that is then compared to the geometry of the same ligand from the original crystallographic image. If the two are in agreement then the theory that the most frequently occurring solution is a reliable prediction starts to accrue merit.

According to GOLD, the highest scoring solution from a docking run should be the best possible solution for that ligand in the target protein. As mentioned before, this is more often not the case. The purpose of this chapter is to compare the methods of highest score and

highest frequency in RARα and RARγ and determine which is the best predictor of ligand geometry.

RARα:

In RARα, both the key conformer and highest scoring solution were in agreement with the bound ligand **fig.11**. A good overlap was consistent across the entire length of the compound. But the compound itself doesn't have many points around which it can rotate and so has few degrees of freedom. The conformer distribution graph from the clustering of solutions shows only 2 conformers.



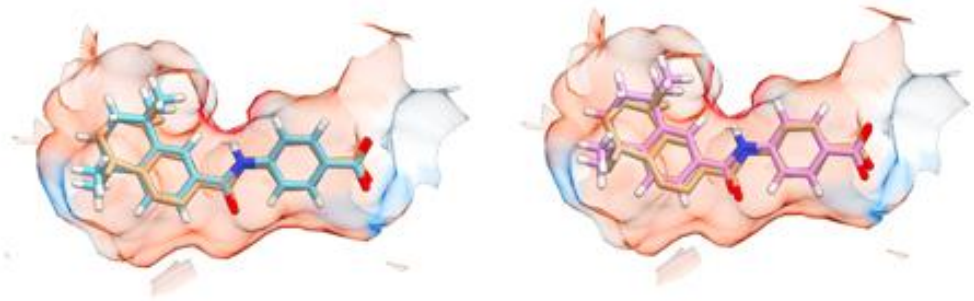


Figure 11 Only 2 conformers were produced by docking (top). Conformer 1 occured the most and is the "Key" conformer. The most frequently occuring conformer (blue, bottom left) and the highest scoring solution (pink, bottom right), both replicated the known binding mode (beige)

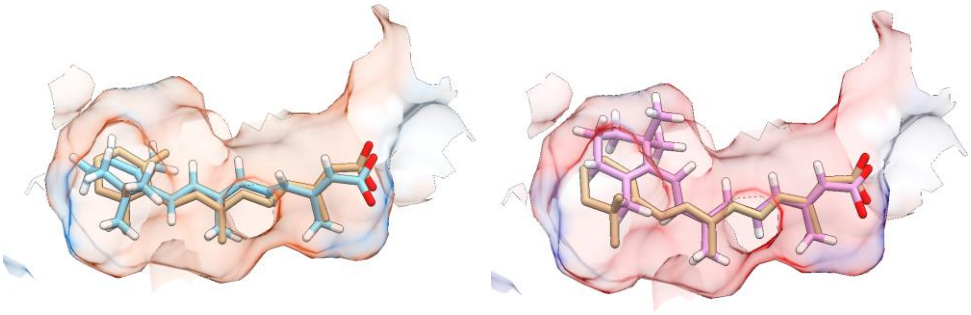


Figure 12 Most frequently occurring conformer (blue) is in excellent agreement with bound structure in RARγ (PDB: 2LBD). The double ring system from the highest scoring solution (pink) is not consistent with the bound geometry (beige).

The difference between the clustering algorithm and scoring function is better exemplified in the docking of the ligand from RARγ (2LBD) **fig.12**. Here, 5 conformations were identified in the solutions **fig.13**. The key conformer (conformer 1) was once again in excellent agreement with the bound position, and it was the most prominent conformer by a wide and comfortable margin. Excellent overlap was consistent across the length of the docked solution **fig.12 left**. The highest scoring solution, although in good agreement with the polar end of the bound structure, adopted a rotation about the middle of the compound and, more relevantly, was not in agreement with the known geometry **fig.12 right**.

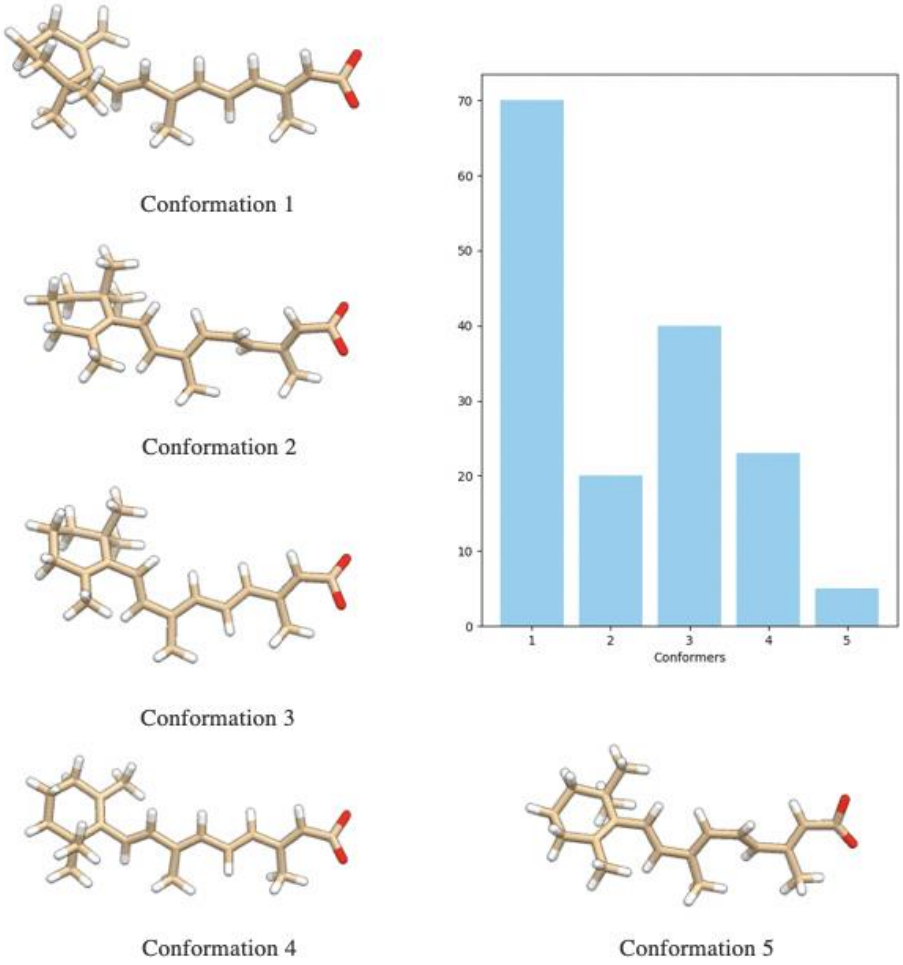


Figure 13 The solutions from the docking of the ligand in in RAR-gamma adopted one of 5 conformations. Conformer 1 was the most frequently adopted and is consistent with the bound position from the crystal structure.

The statistical distributions of conformers of each compound is useful not just in predicting a likely interaction complex, but in describing the behavior of the interaction itself. Such limited exploration of conformations in RARα, as seen in **fig.11**, is a visual representation of the limited freedom the compound has within the cavity. Although in this case the propensity of the compound to explore different binding modes is visibly sterically hindered by the ligand binding domain in RARα, in cases such as RARγ, it isn't so easy to tell. The total number of conformers

is therefore an important indicator of the behaviour of the ligand-protein complex. In RARγ, it is clear that a number of binding modes are likely, but that one is definitively preferred. In RARα, only one is probable and others extremely unlikely.

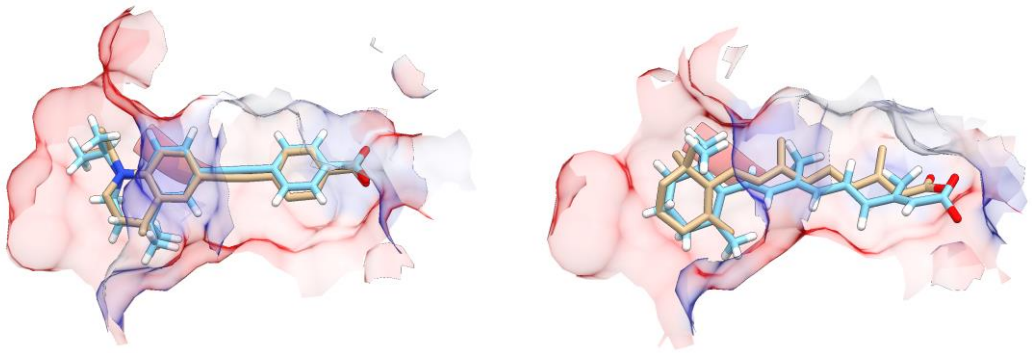


Figure 14 Again the Key conformers in both CRABP proteins (CRABPI left, CRABPII right) were in excellent agreement with the known binding position.

In CRABPII (6HKR), only one conformer was found among the solutions which was an exact replication of the bound compound **fig.14-left**. As another structure was available for CRABPII (2G78), this too was redocked. Here the key conformer among the solutions was once again in good agreement with the bound conformer **fig.14-right**, with consistent stereochemistry along the compound and correct orientation and overlap of the benzene ring. Analysis of the frequency distribution produced by the algorithm shows that the correct binding mode was more frequent and by an even greater margin than in RARγ despite there being far more conformations found in the solutions **fig.13** . The greater number of conformations is a result of the much larger and open cavity in CRABPII and suggests more frequent interconversion between binding modes but that one is preferred among the 11 discovered.

Such consistent replication by the Key conformer in the RAR's and CRABPs of the known binding modes is important. Whereas the score assigned to each solution from Chemscore uses an algorithmic simplification, the clustering algorithm uses only frequency to derive the most probable interaction. In all proteins examined so far the algorithm has correctly identified a key conformer, identical to the one procured from the protein crystal structure.

Chemscore too was able to replicate the known binding modes with good accuracy in most cases. But its precision was poor. To test it, a library of 50 starting files was generated of the bound ligand from 2G78 according to the procedure in chapter 5 part i. These solutions were docked according to the same procedure. The highest scoring solution for these 50 files only replicated the known binding mode in 44% of cases.

When accuracy is so vital for informing the prospective researcher, relying on a scoring system that is capable of 44% accuracy in some cases is risky. The clustering method proposed then offers a statement of the relative probabilities between solutions. Instead of a single score, the researcher is equipped with a statistical distribution that describes the overall scope of available ligand-protein complexes. Additionally, as this small accuracy experiment shows, GOLD can give different solutions for any given run of a compound. Its variation is both excellent and problematic. Excellent because it rules out the influence of the initial geometry on the final solution, but a hindrance in practice because the lack of repeatability. But this variation doesn't affect the clustering pipeline. Although there is small variation in the resulting conformers, the overall picture is consistent from run to run. It becomes a reliable and repeatable process.

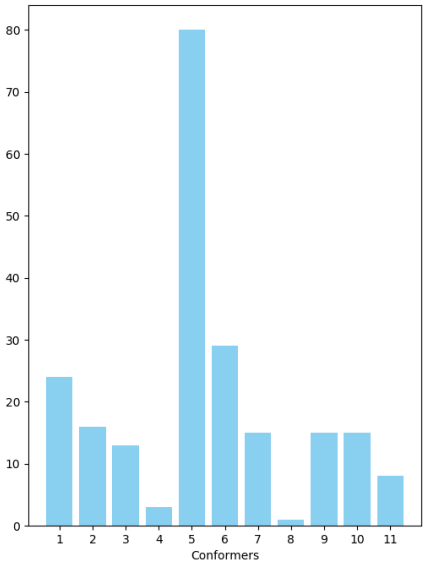


Figure 15 Solutions from docking with CRABPII had 11 possible conformations. The one dominant conformation, 5, was a good replication of the bound orientation and is shown in fig14-right..

The interactions computed using GOLD may not be as accurate when compared to the modelling using Molecular Dynamics (MD) simulations. But they are accurate enough to derive valid and probable interactions and weigh them against each other as competing forces to arrive at solutions using the genetic algorithm. Each solution is the result of competing forces driving the geometry of a ligand to a point of rest. By pushing multiple starting points into the docking software, you're effectively maximizing the number of times that ligand is put through the wash cycle of interactions to arrive at a local energy minimum. Dominant forces will win-out over a large sample range, regardless of which initial geometry is presented. The clustering algorithm, therefore, is a statistical distribution of the probability that a ligand, once in the binding domain, will find that particular point of rest.

The proposed method, then, is an extension of the existing scoring function. The scoring function drives the optimisation of the conformers before arriving at a solution and so each solution is the product of multiple progressive iterations of interactions between the drug and protein. Clustering these solutions is a numerical representation of the number of agreed endpoints from the scoring function and should be considered an additional level of theory to scoring.

If the score given to solutions is not consistently an accurate predictor of binding, then the use of that score alone to compare multiple compounds with one another may not be the most reliable way to discriminate between good and bad compounds. A good way to begin might be by the clustering of a large sample of solutions and visual inspection of the key conformer for each compound and cross comparison, where available, using compounds with known binding data.

Thanks to decades of research and discovery, thousands of protein structures with bound ligands are available on the protein data bank. The bound structures, along with ligands known to interact with that target offer a good starting point for drug design and now, any research group can produce a large database of drug candidates and run a high throughput docking and clustering analysis to investigate the best looking ligands. Docking of known agonists is invaluable to discovery as they give useful information on what to look for in potential candidates.

7. A NEW PIPELINE

Computational investigation is not routinely employed in synthetic laboratory environments by synthetic chemists and access to high level computation is extremely limited. The much simpler docking processes presented here, if accurate, would mean that actionable computational prediction would be accessible to anyone regardless of computational expertise and with any budget. Synthetic chemists would be able to dock, cluster and derive the best looking compounds from a database to corroborate binding data for compounds already made and inform on what to adjust for future ones. Such efficient feedback between compound behavior and understanding is out of reach for the time being.

But docking could potentially go one step further. Knowing how a compound may spatially occupy a binding domain is useful, but more useful still is knowing if that compound is better or worse than another in any one protein. Using GOLD as instructed enables docking of a compound in a protein, the result of which is a score which indicates the suitability of that given compound for that given protein. In the same way, scores from different compounds can be used to create a ranked order of the compounds by comparing the highest scores achieved by each and ordering them accordingly.

This is the current way in which GOLD might be used in a high-throughput screening experiment. Although, as demonstrated in the previous chapter, the highest scoring solution is not necessarily accurate in terms of how a compound interacts with the target and so ranking in this way may not be accurate.

In this chapter, the clustering process described in chapter 6 is compared with the existing GOLD process to predict the order ranking for a series of compounds with known binding data to identify which, if either, is accurate in this way.

i. Methods

The following methods are repeated for each of the compounds in **fig.16-17**

a. GOLD only method

Docking was as described in 6i-ii.

After docking, the conformer with the highest fitness score (given by GOLD) is treated as the best solution. Repeat for each compound and order scores to achieve a rank order.

b. Clustering GOLD solutions (experimental)

This method was as described in 6i-iii for each compound. The chemscore value for the key conformer is used to rank the compounds.

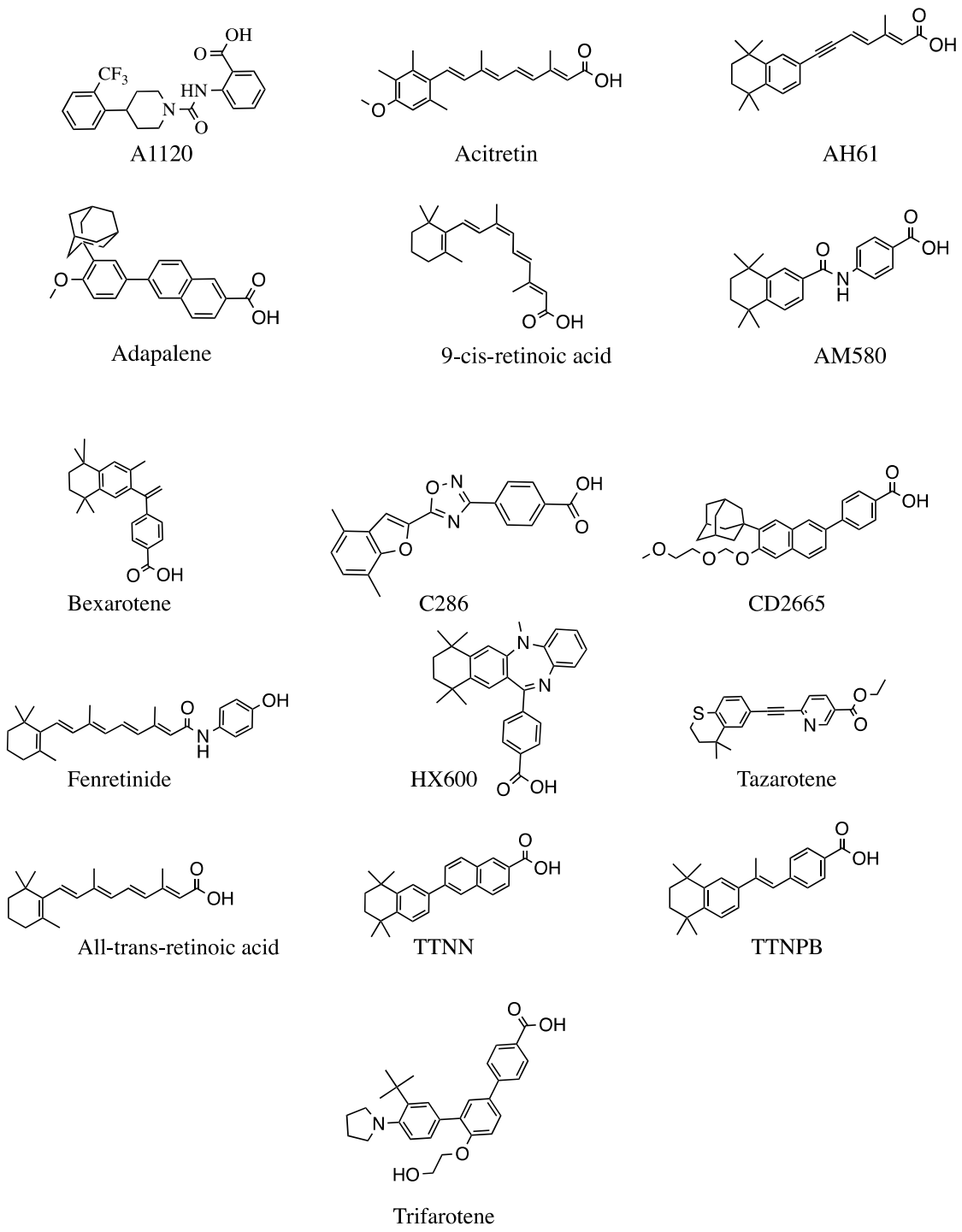
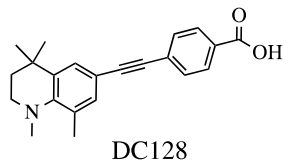
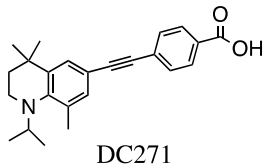
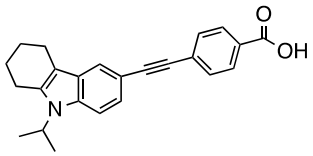
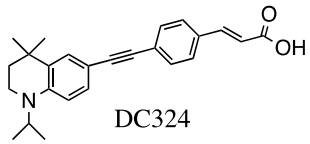


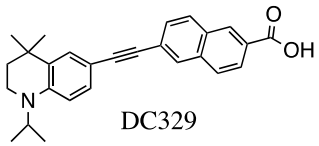
Figure 16 Structures of synthetic retinoids A1120[67], Acitretin[68], AH61[69], Adapalene[70], AM580[71], Bexarotene[72], C286[73], CD2665[74], Fenretinide[75], HX600[69], Tazarotene[76], TTNN[69], TTNPB[69] and Trifarotene[76].

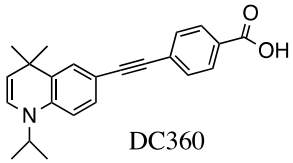

DC128

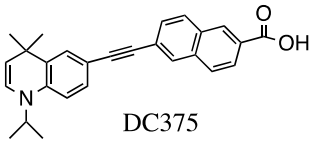

DC271

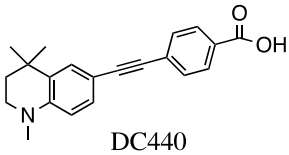

DC318

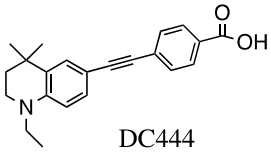

DC324

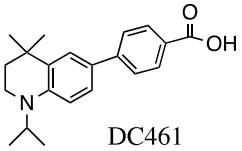

DC329

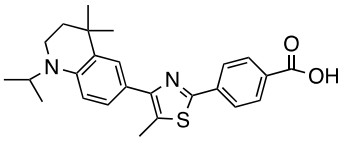

DC360

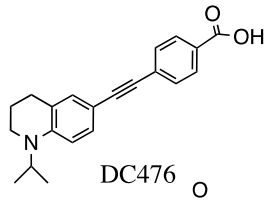

DC375

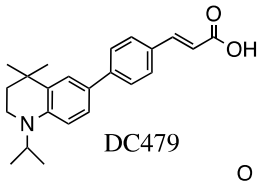

DC440

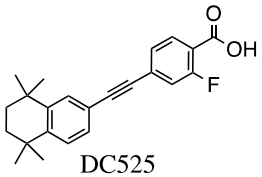

DC444

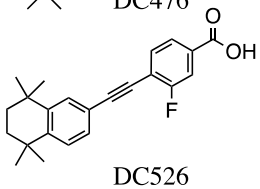

DC461

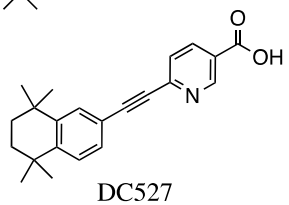

DC472

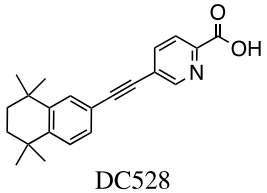

DC476

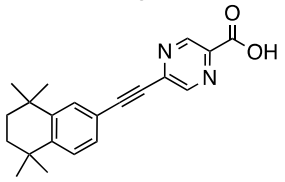

DC479

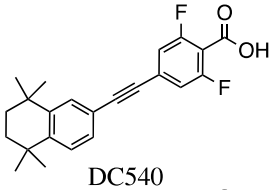

DC525

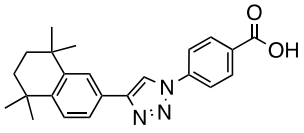

DC526

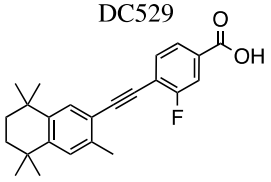

DC527

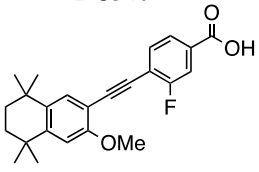

DC528

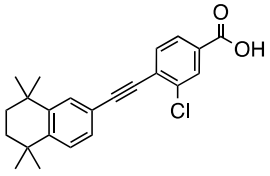

DC529

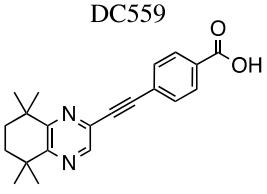

DC540

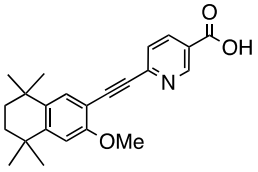

DC547

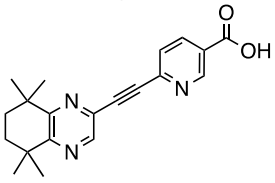

DC559

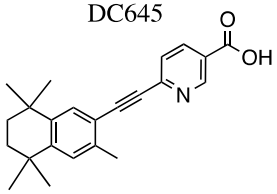

DC564


DC567


DC645


DC646


DC650


DC656

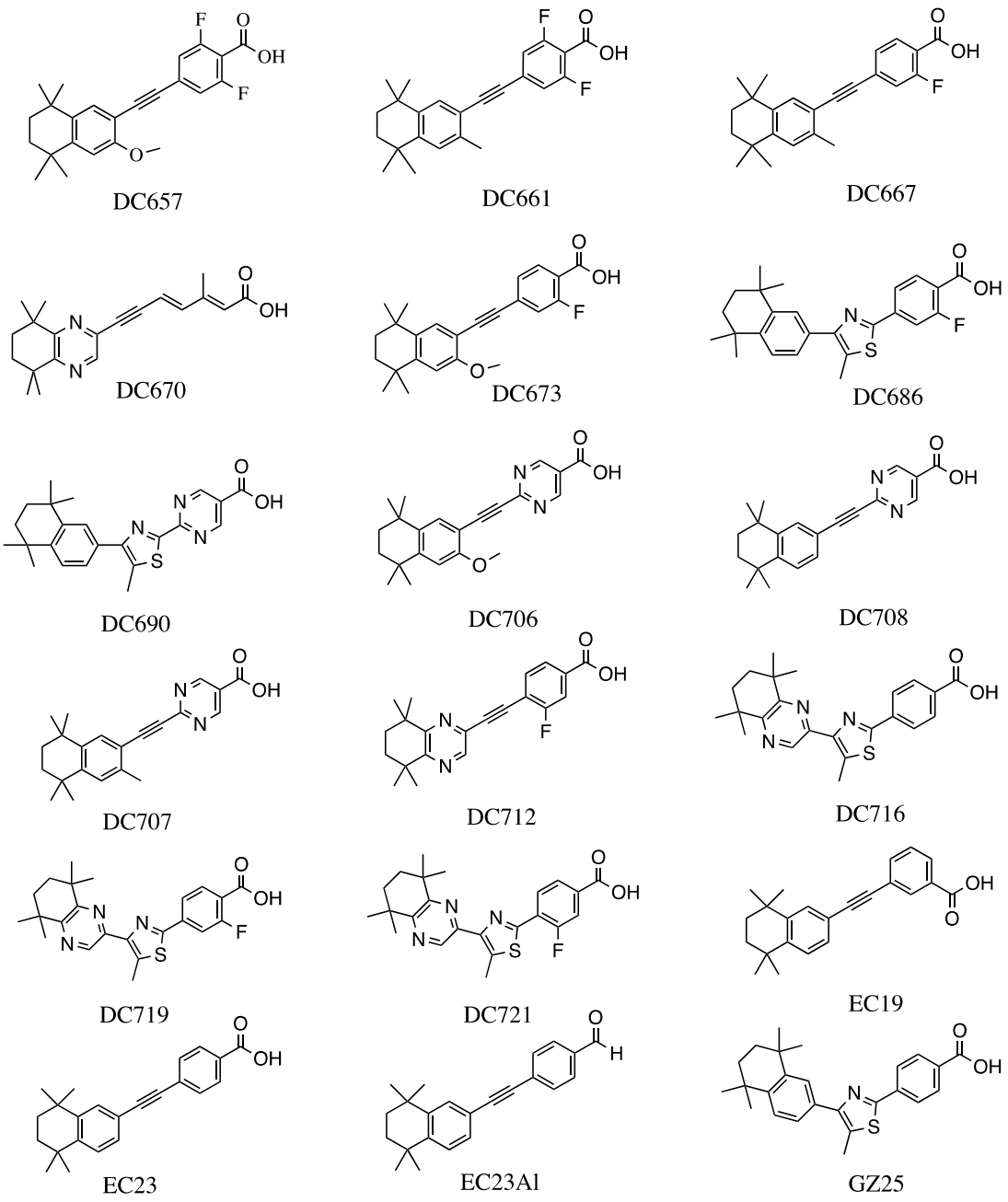


Figure 17 Synthetic retinoids created by the Whiting group (unpublished).

ii. New pipeline results

For each experimentally determined binding data, the predicted rank orders from the two pipeline processed were compared using the Kendal Tau correlation algorithm. This correlation calculation determines concordance of order, ranging in values from completely negatively correlated (-1), to completely positively correlated (1) and no correlation at all (0).

calculated order RARα	Kd Alpha	95% conf
DC540	1.37	0.80
DC667	2.82	7.09
DC567	13.03	8.07
DC527	29.00	13.14
DC528	30.83	12.67
DC657	78.67	19.30
DC645	93.00	12.91
DC719	133.00	68.89
DC712	148.67	10.04
DC716	369.33	68.54
DC707	416.67	62.52

calculated order RARγ	Kd Gamma	95% conf
DC567	10.90	1.55
DC540	14.00	2.48
DC667	52.67	21.70
DC645	92.00	8.61
DC527	97.67	34.27
DC712	101.50	120.71
DC528	106.67	36.20
DC719	240.00	65.72
DC657	313.33	100.40
DC716	336.67	87.24
DC707	893.33	37.95

Figure 18 Table showing binding affinity for a series of synthetic retinoids. Pohl, Tomlinson et al (unpublished).

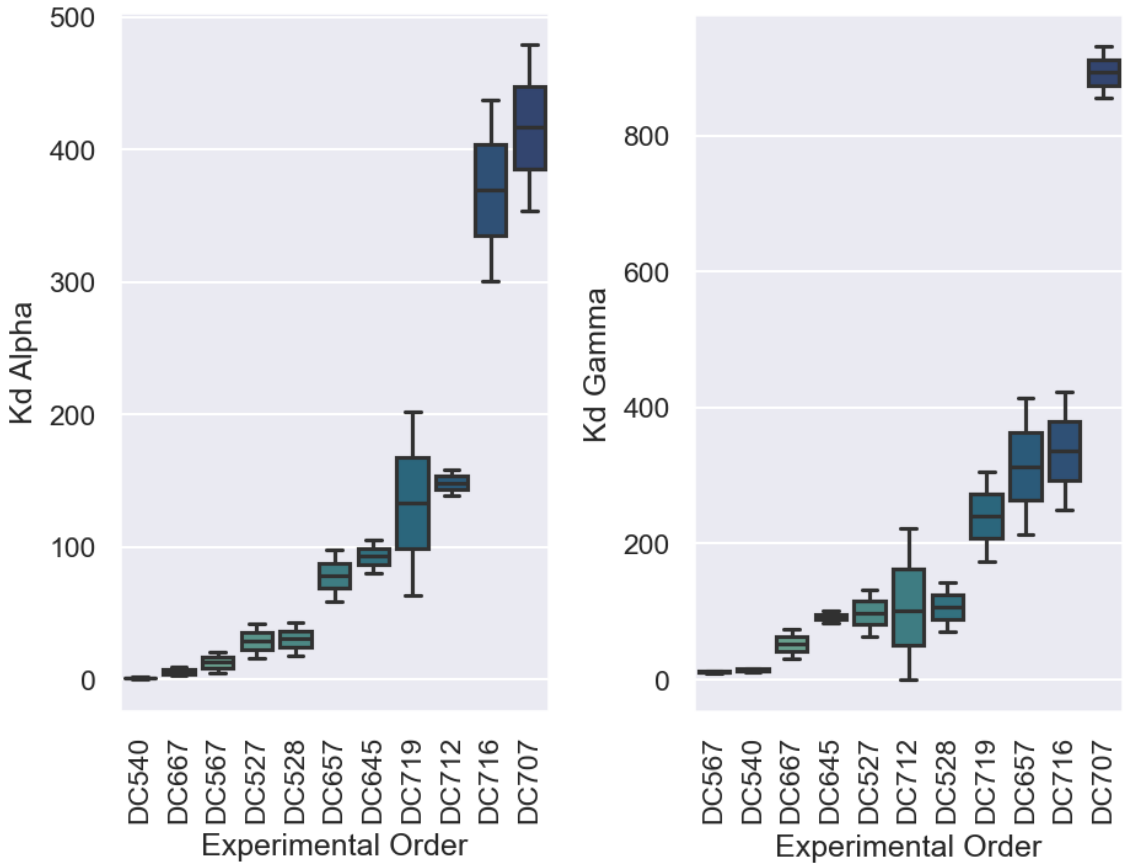


Figure 19 Experimentally determined rank order of synthetic retinoids in RAR-alpha and RAR-gamma Those with better binding are further left.

This series of ligands exhibited a wide range of experimentally determined binding affinities. This is useful. If affinites were too similar, it would have been difficult to determine the accuracy of any kind of predictive method. Their values would overlap and any number of predicted rank orders would have been equally valid.

Instead, here we have affinites that range from ligands that are completely cemented to the binding domain, to those which are relatively indifferent to any interaction.

a. GOLD only method

The rank order as determined using Chemscore demonstrated mixed results. In RARa, it succeeded in identifying DC667 as a high affinity compound, but DC540, with the highest experimentally determined affinity, was predicted to be one of the worst compounds.

Rank Order RARα	Compound
1	DC667
2	DC716
3	DC527
4	DC719
5	DC567
6	DC528
7	DC645
8	DC540
9	DC712
10	DC707
11	DC657

Rank Order RARγ	Compound
1	DC667
2	DC567
3	DC716
4	DC527
5	DC719
6	DC528
7	DC645
8	DC712
9	DC707
10	DC540
11	DC657

Figure 20 Predicted order from GOLD only method.

When compared with the experimentally determined rank order, it gave a low correlation score of 0.169 and did not reach statistical significance. Although the exact order as proposed by the data produces a low correlation, other rank orders are allowed within the error margins of the experimental data, suggesting that in reality, the exact value for Kd may differ in such a way that the rank order could move to either improve or decrease the correlation. Although this is true for all the pipelines and does not affect their intercomparison.

In RARγ, the order for this pipeline had much improved correlation. The Kendal Tau Correlation algorithm gave a correlation of 0.325, and reached statistical significance ($P<0.05$).

b. Clustering pipeline

In this pipeline, the rank orders were best correlated with the experimentally determined order in both RAR isotypes. In RARα the correlation score was 0.229, but didn't reach statistical significance. But in RARγ, it exhibited the best correlation with a score of 0.429 and achieving excellent significance ($P<0.005$).

Rank order RARα	Compound
1	DC667
2	DC567
3	DC528
4	DC716
5	DC527
6	DC540
7	DC719
8	DC645
9	DC712
10	DC657
11	DC707

Rank order RARγ	Compound
1	DC667
2	DC567
3	DC540
4	DC528
5	DC527
6	DC645
7	DC657
8	DC712
9	DC707
10	DC719
11	DC716

Figure 21 Predicted order from clustering pipeline

8. UNDERSTANDING BINDING AND SELECTIVITY

There are a large number of conformations available to all-trans-retinoic-acid (tretinoin). The poly-ene chain delivers multiple degrees of freedom within the compound, compromising its specificity to the RARs by allowing it to adopt geometries compatible with other receptors. Efforts to develop synthetic retinoids have therefore adopted the approach of restricting the flexibility of tretinoin. It is thought that the key to potency and selectivity between the RARs lies in the attainment of a specific rigid geometry for each isotype.

Achieving selectivity in the RARs would not only achieve specificity in terms of downstream activation of effects, such as the previously discussed neurite outgrowth, but also limit the well-known side effects of retinoids, namely their teratogenicity. To this end, a number of synthetic retinoids with selectivity between RAR isotypes and RXR proteins have been developed over the past few decades. Early efforts aimed at RARα selectivity built on the premise that the serine residue mid-cavity which is absent in RARβ and RARγ, would be available for hydrogen bonding. Alpha selectivity was achieved to great success before the turn of the century with the synthesis of AM580 and similar compounds. Amide linking groups made use of the serine residue and achieved stability in RARα while destabilizing the same compound in the highly lipophilic mid-sections of RARβ and RARγ.

The assumption from the clustering algorithm is that if a compound is reported to have particularly strong or weak interactions with a target, computational binding could reveal why. The key conformer, if the most accurate prediction of binding, would be best to demonstrate the particular attributes of that complex and therefore understand the behaviour of the complex.

It has already been established that the clustering of solutions yields a good prediction of the likely binding mode of a compound. Now we can use what are reliable predictive images of protein ligand binding and infer a positive strong, or weak and negative relationship. Additionally, it has now been shown that the chemscore score of the Key conformer derived

from clustering is a good useful predictor of rank order for a series of compounds. Together, this can be taken to mean that the clustering pipeline, previously tested, is a good qualitative method of ligand binding analysis.

In this chapter, influencers of binding are first summarized. Then, compounds with known selectivity, which have been docked in their preferred receptor, are clustered and the resulting Key conformer analysed to determine, if any, observable reasons for selectivity.

i. Influencers of binding

During visual inspection of the selective retinoids, a simplified overview of considerations are here summarized and were considered when the visual inspection was performed.

a. Enthalpy and entropy

The affinity for a target protein describes the strength of attraction to that target from the ligand. Enthalpy and entropy direct this affinity and make up the binding free energy of the ligand-protein complex given by the equation below.

$$\Delta G = \Delta H - T\Delta S$$

Better binding, simplistically, is denoted by a more negative value of ΔG and is the product of both enthalpy and entropy. These terms in turn are directed by ligand-protein interactions such as Van Der Waals forces, hydrogen bonding, the hydrophobic effect and the displacement of water molecules in the binding cavity by the introduction of the ligand itself. One major hurdle in drug design is optimization of these forces. Usually efforts to increase the favorability of enthalpy is met with a corresponding decrease in entropy, known as the entropic penalty[77].

Entropy in this context is the measure of disorder in a given compound. A greater degree of entropy is observed in larger ligands with more rotational degrees of freedom and more branched chains. A greater level of entropy would favor binding, but a significant decrease in entropy is observed by the loss of ligand freedom, and therefore disorder, when bound to a protein[77]. It is this loss in entropy that harms the affinity of the compound[78]. Minimizing this entropic loss is achieved by the reduction of the entropy of the ligand in its pre-bound state so the overall decrease in entropy when binding, ΔS, is reduced. Favorable enthalpic gains are made by the introduction of hydrogen bonds and good spatial occupation of the binding domain.

Computational algorithmic considerations have been introduced to reflect these influences. But the underlying physics in computation is still in development. In the desktop docking used here, the protein is kept rigid and so global entropic changes of the protein-ligand complex can't be considered and the pre-existing waters are removed before docking and so can't be considered either.

However a few entropic statements should be observed. Firstly multiple studies have noted no trend with respect to ligand size and conformational entropy in binding and, secondly, most of the entropy loss during binding is due to a loss in vibrational freedom and not loss of accessible conformers. In fact, a computational study calculating entropic loss of the same ligand in complex with different receptor and the same ligand in different conformations within the same target found no correlation between the number of rotatable bonds and entropy change[78].

It is impossible to identify and quantify the entropic loss of a ligand from visual inspection only, but these observations taken together suggest that actually, for a database of ligands of similar size, as is presented here, the entropic loss upon binding should be similar, regardless of size and conformational exploration. Instead, the only observable differences between the ligands should be of an enthalpic nature, i.e. the number and strength of hydrogen bonds and the special occupation of the binding cavity. In this way, entropy can be dismissed during the visual inspection of the selective retinoids.

b. Steric Effects

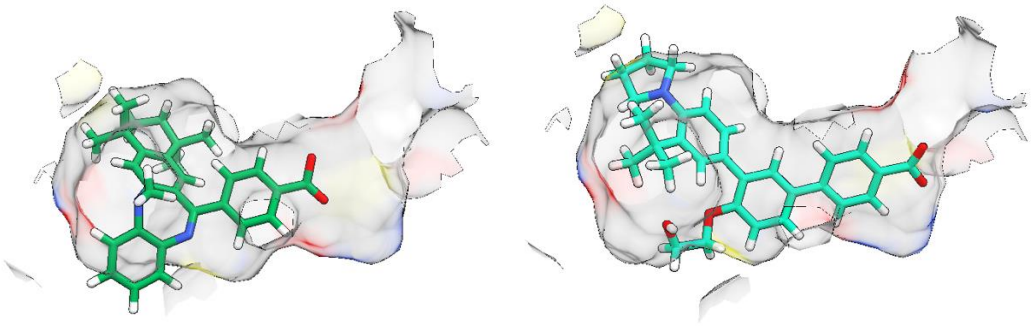


Figure 22a Category A. Minor perforation of the binding pocket by one or two hydrogen atoms that could reasonably occur with relatively limited harm to the overall binding complex. .

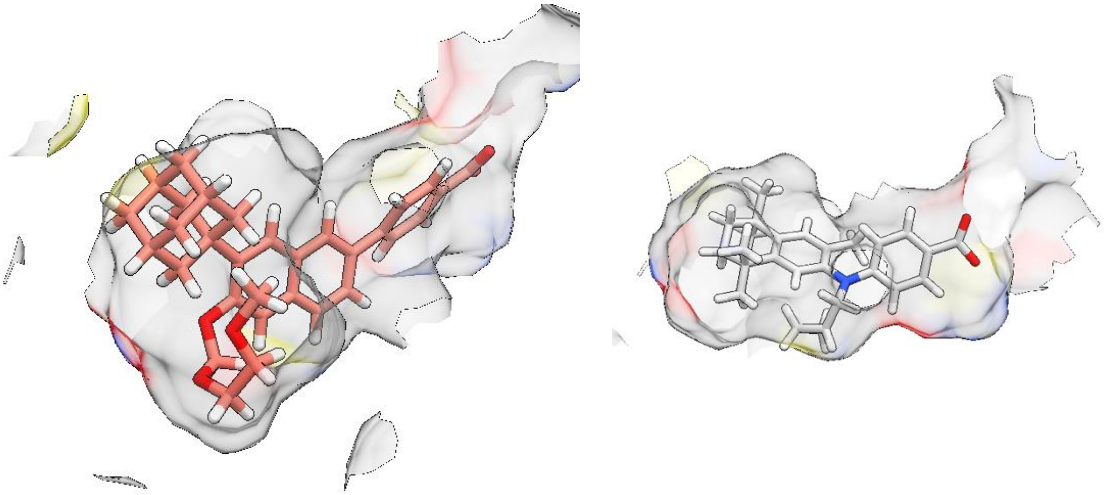


Figure 22b Category B. Minor perforation of the binding pocket by one or two hydrogen atoms that could reasonably occur with relatively limited harm to the overall binding complex. .

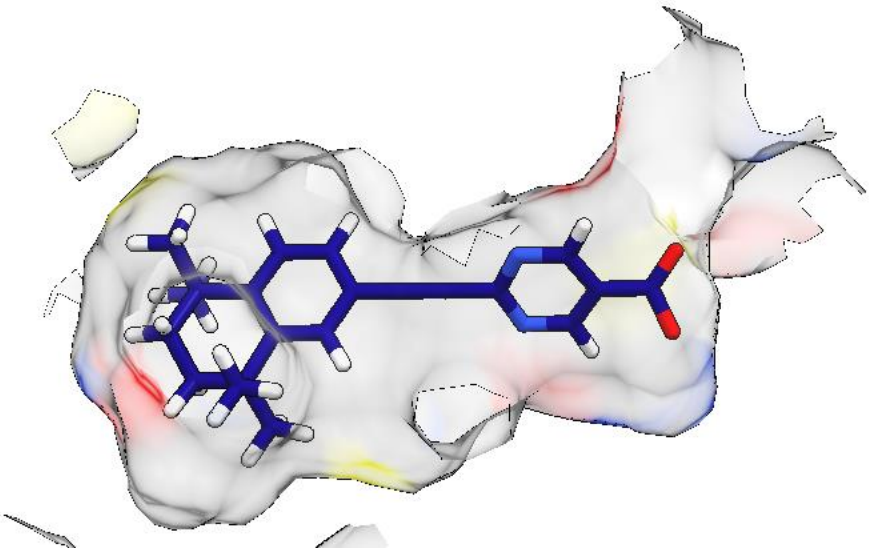


Figure 22c Category A. Good and reasonably balanced exploration of the binding domain with no perforation of the pocket.

Accordingly, binding data from a series of displacement assays show that compounds which fit poorly with the cavity unsurprisingly reveal much hindered affinities towards that target. Compounds exhibited dramatically poor binding when perforating the binding pocket. This information can applied to the observation of key conformers and the suitability of their interactions inferred and scored as one of three categories **fig.22**.

c. Hydrogen Bonding

Hydrogen bonds occur when a donor atom shares their bonded hydrogen bond with an acceptor atom. The distance between the donor atom and the hydrogen atom is usually shorter than the distance between the hydrogen atom and the acceptor. The bond distance is measured from donor atom to acceptor atom.

Although a lot of the bond distances quoted from Chimera in this investigation are below 2.0A, general wisdom indicates that intermolecular hydrogen bonds of this length are not realistic.

The crystal structures from the protien data bank come with varying degrees of coordinate error denoted by their resolution. In addition the protein is held rigid in the docking performed here whereas it would be able to move with the ligand *in vivo*. The unrealistic distances desrcibed in some instances therefore are most likely a defect resulting from these two limitations. Instead the shortest bond length will be considered at 2.2Å

Bond lengths are separated based on their strengths into three distance categories, 2.2-2.5Å, 2.5-3.2Å and 3.2-4.0Å with energies 40-15kcal/mol, 15-4kcal/mol and <4kcal/mol respectively.. A graph showing the relationship between bond energies and intermolecular distance is shown in **fig.23.**

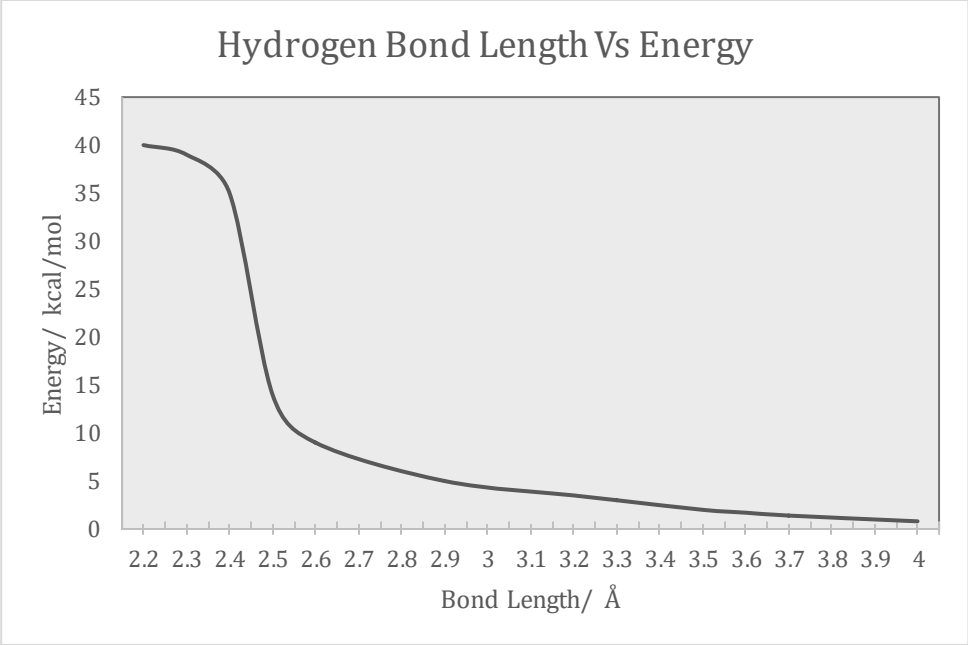


Figure 23 Bond energy variation with bond length.

The relationship between bond length and energy is not linear. The prompt increase in energy and subsequent plateu in **fig.23** for bonds shorter than 2.5Å denotes a state known as low-barrier hydrogen bonding (LBHB). Simplistically, this occurs when the hydrogen atom interacts with equal energy by both the donor and acceptor atoms and becomes equidistant from both.

Instead of referring to the hydrogen bonds of ligands by their distances with varying accuracy, the mean hydrogen bond length for a ligand was placed in categories of either STRONG(<2.50Å), MEDIUM(2.6-3.2Å) WEAK (3.3-4Å)or N/A for bonds of negligible energy (>4Å).

ii. Inspection of selective compounds

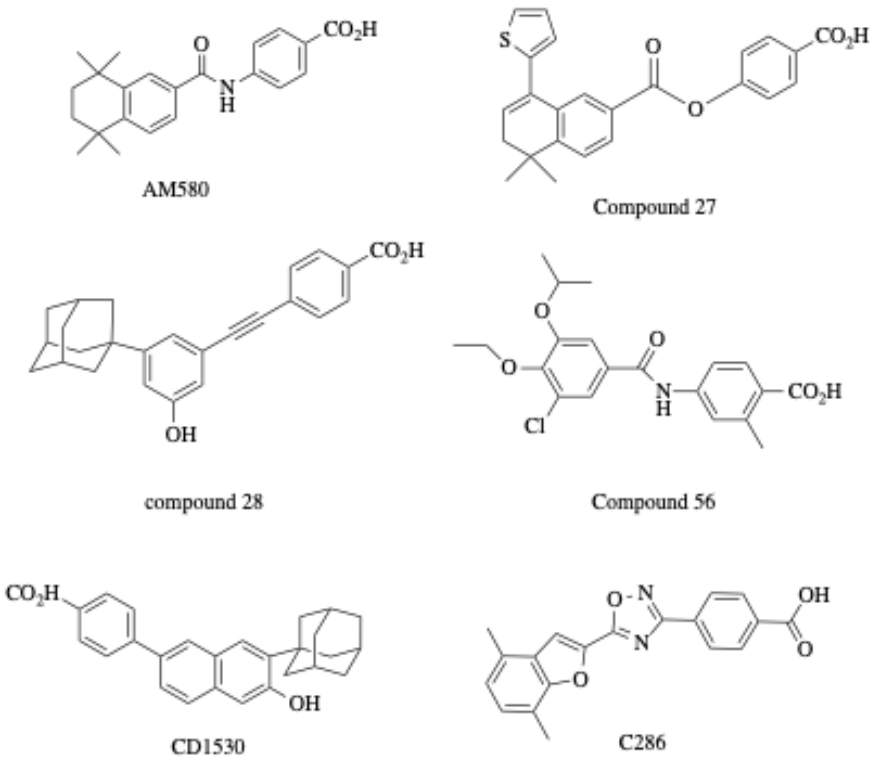


Figure 24 Selective Compounds[79 73, 80]

a. Alpha selective compounds

Compounds that are reported to have alpha selectivity are shown in **fig.24** . These compounds were drawn and docked according to the procedure in (whatever section it was), the solutions from docking were run through the algorithm and the key conformers selected for inspection with UCSF chimera. The binding pocket for RARa was identified and the rest of the protein hidden. Polar groups within 5A of the bound ligand were retained and all other atoms hidden for clarity and hydrogen bonds meeting Chimera's standard criteria were highlighted and interaction distance measured.

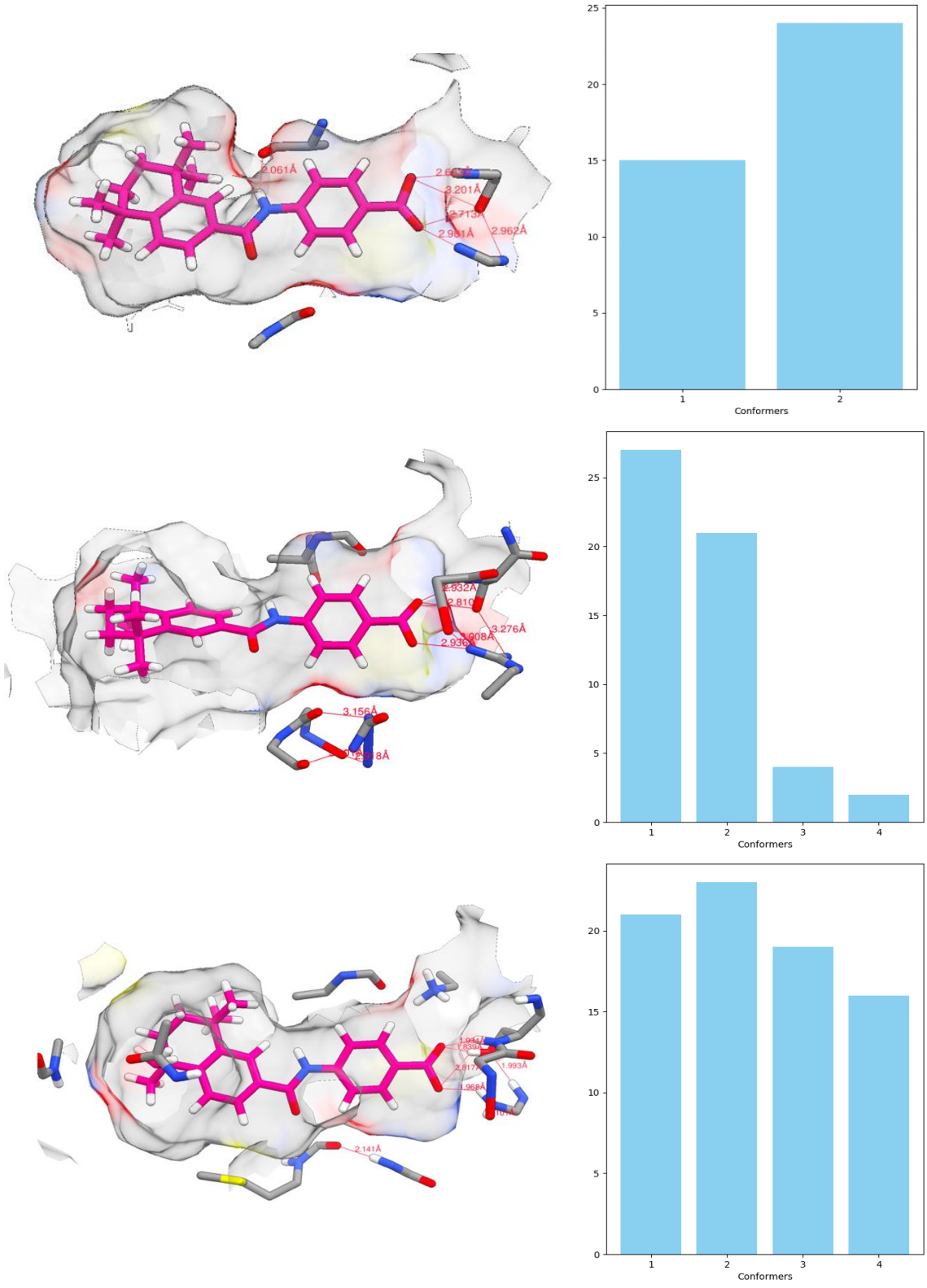


Figure 25 The conformer distribution and frequency of AM580 across the RARs. Fewer conformers explored in RARa indicating either some dominant and preferred solutions or some barrier to other conformers.

Both α selective compounds make use of the hydroxyl group from the serine residue in RARα to form hydrogen bonds, as reported. The number of hydrogen bonds vary between all compounds but all have more hydrogen bonds in RARα than in the other isotypes. The number of hydrogen bonds alone isn't conclusive but what makes a difference in RARα is the positioning of the Serine group mid-cavity, absent in the other isotypes. This polar group, lends a degree of stability to the compound by the formation of hydrogen bonds[79], fixing its geometry as seen by the much fewer conformations explored in α by both AM580[79] and compound 56[80].

Hydrogen bond distances vary between isotypes. For all compounds, the mean hydrogen bond distance is much less and therefore stronger in RARγ, than for RARα. This is likely due to the slightly longer cavity in RARα and the action of the double ring system acting as a steric anchor in the hydrophobic section, limiting the distance between the polar section and the carboxylate group. If the H-bonds are stronger in RARγ than RARα, how is it their affinity for α is greater than for γ. Sterically, the two receptors aren't offended by these compounds, the only difference between them is the quantity, strength and positioning of H-bonds. Subsequently, a consideration must be made here as to what is more important when considering selectivity towards RARα, the potential for an extra hydrogen bond, or the position of those bonds.

Compounds selective for the other two isotypes, β and γ appear to avoid this mid-cavity interaction by either making the polar linkage inaccessible or removing the polar linking groups altogether. Compound 27[79] avoids the interaction by the inclusion of a hugely bulky ring system. The only way the ring system fits within the pocket is by angling the polar section away from the serine group and although rotation and interconversion is possible for other compounds, rotating such a bulky group would be too hindered. It's too wide to rotate within the pocket. One of its solutions does in fact make use of the polar mid-section by adopting a geometry like that within RARγ, rudely perforating the binding pocket in a manner consistent with extremely poor or completely inactive binding.

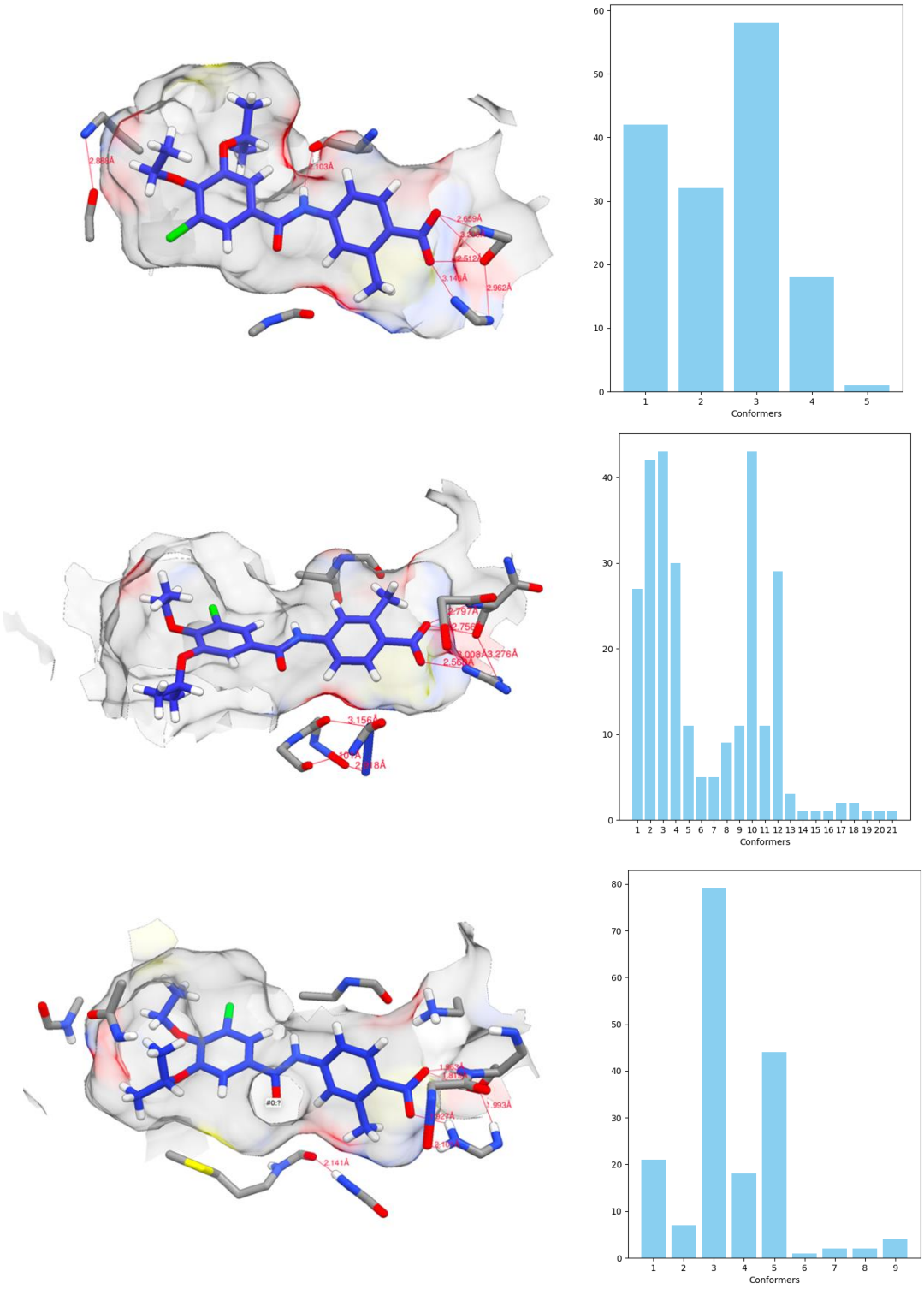


Figure 26 Compound 56 exhibits many more conformations in RAR-beta. This is a common observation and relates to the far larger biding pocket compared to the other isotypes

b. *Beta Selective Compounds:*

Hydrogen bonding doesn't seem as dominant for beta selective comounds as it was observed to be in alpha. C286[73] actually forms more hydrogen bonds in alpha, consistent with alpha selective compounds.

Compound 27 has more hydrogen bonds in RARβ than in RARα, but fewer than in RARγ. The mean distance of those bonds is less in gamma than in beta. So the shorter and closer distance of the hydrogen bonds alone can't be directing selectivity.

But selectivity becomes obvious when it is observed that these compounds perforate the LBD in RARα and RARγ. Compound 27 maintains a flexible ester linkage and so is able to just about fit in the RARα cavity, but with poor hydrogen bonding. Other conformations which produce better hydrogen bonding involve a completely unrealistic binding mode, violating the pocket surface. This is true also for RARγ.

It seems that unlike RARα selectivity, which is achieved through the introduction of an especially strong interaction specific to that isotype, RARβ selectivity is achieved by making binding with the other isotypes sterically impractical.

c. *Gamma Selective Compounds:*

Gamma selective compounds CD1530[81] and compound 28[79] avoid forming interactions with the serine group in RARα, much like the beta selective compounds. CD1530 consists of a bulky and lipophilic linking ring system. It fits all three isotypes reasonably well with very minor perforation of the binding pockets and it forms 4 hydrogen bonds in all RARs. The only distinguishing feature is the mean bond lengths of 2.87A, 3.15A and 2.02A for RARs α, β and γ reepctively.

Compound 28 does actually possess a polar group capable of forming interactions with the serine in the alpha isotype, but much like compound 27, the only way it fits in RARα is by

angling this polar group away from the serine. It could form the mid-cavity interaction by shifting much further to the left of the LBD, but presumably at the expense of the hydrogen bonds at the polar cluster at the end of the cavity, which is why this binding mode wasn't even explored by any of the solutions. What's interesting about this compound is its selectivity towards RARγ despite actually achieving fewer hydrogen bonds than in the other two isotypes. The strength of the bonds alone must be sufficient to establish a preference for this isotype. This is quite likely as the mean bond length for this compound in RARγ is 1.73A, making it the third lowest mean bond length observed across all isotypes of all 66 compounds docked.

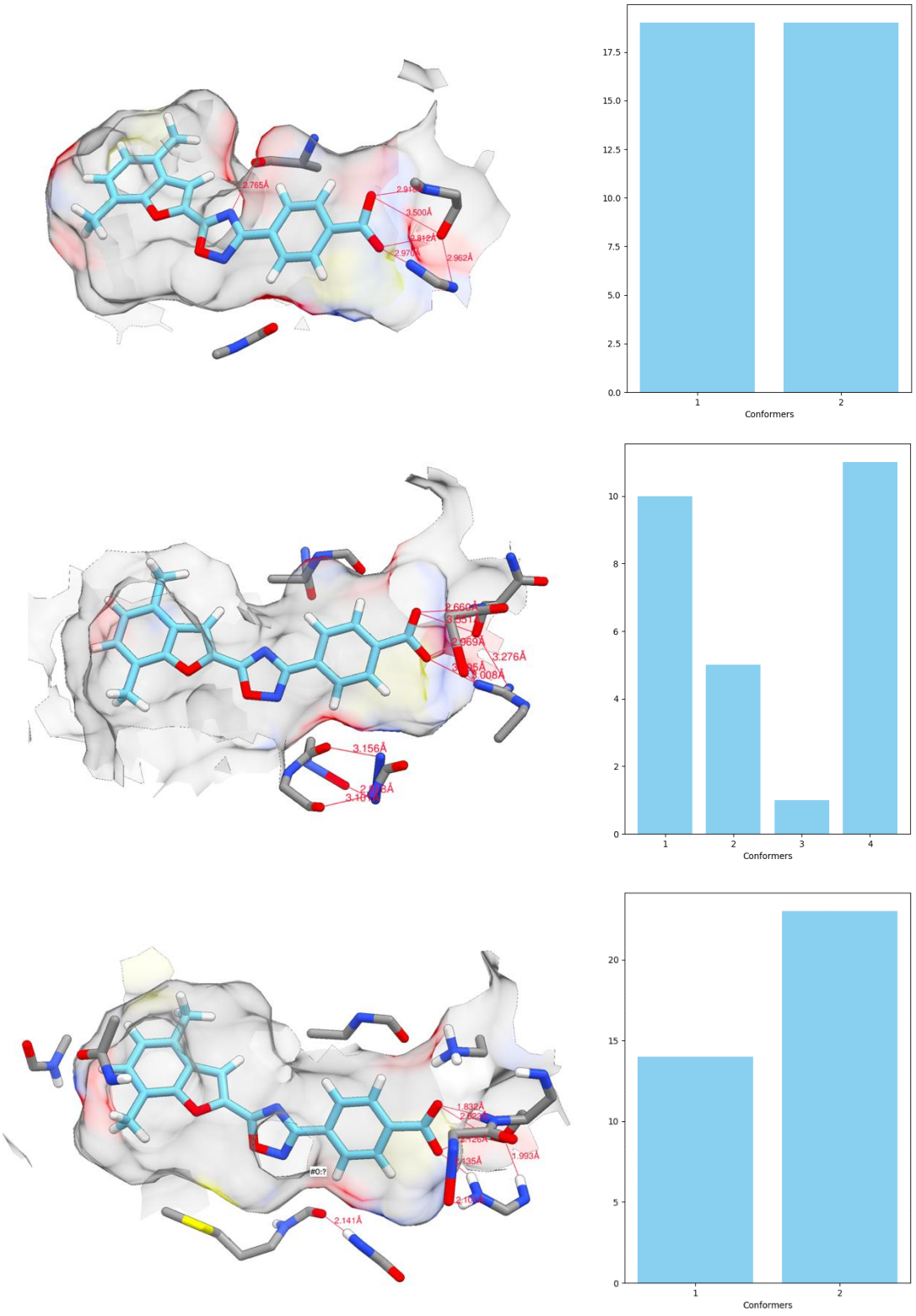


Figure 27 Compound C286 a compound with reported beta selectivity, interestingly forms the same mid-cavity hydrogen bond in RAR-alpha

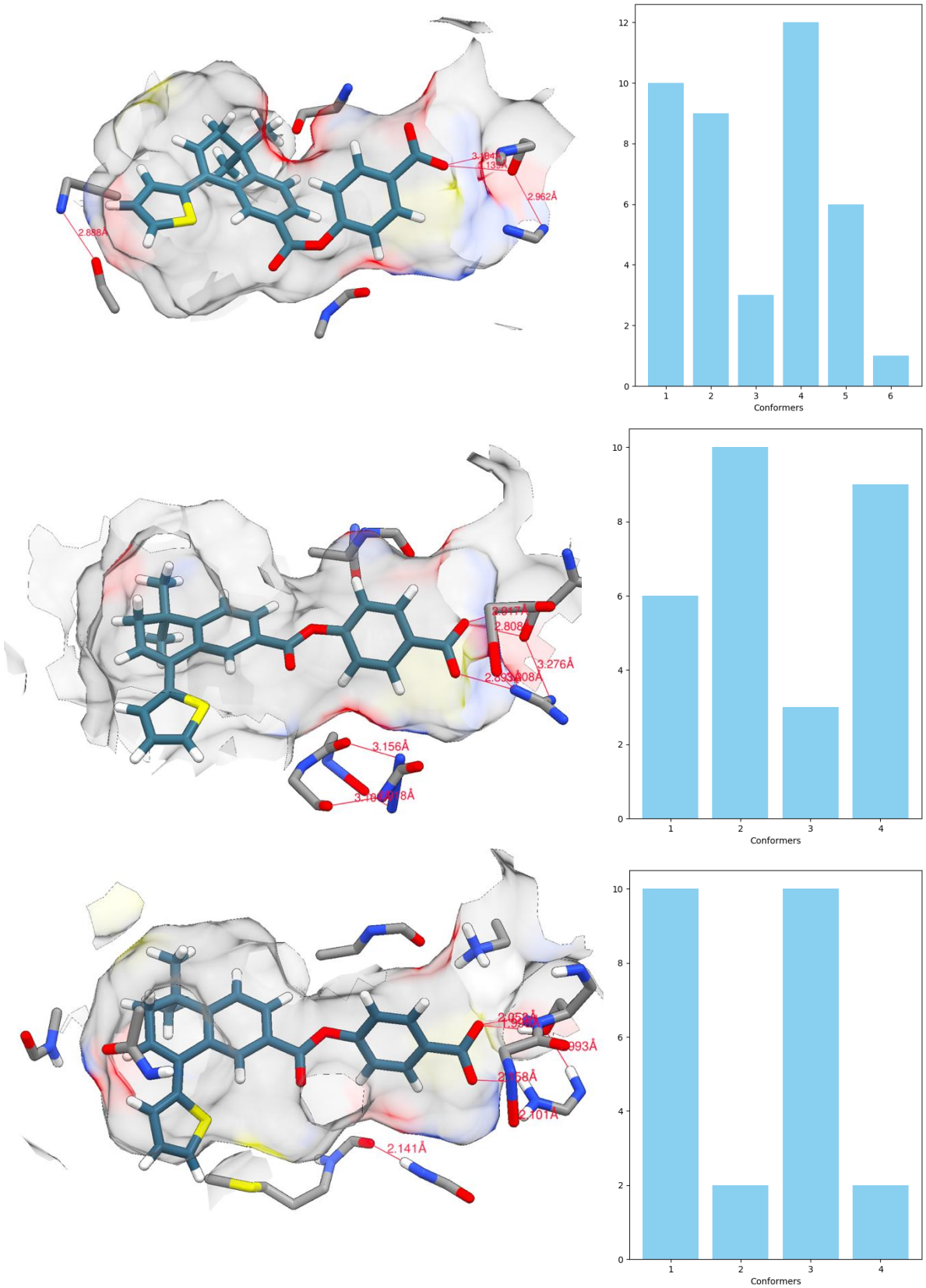


Figure 28 Compound 27 perforates the pocket surface in RAR α and, to a greater extent in RAR γ. Its selectivity for β is therefore the product of a process of elimination rather than raw affinity.

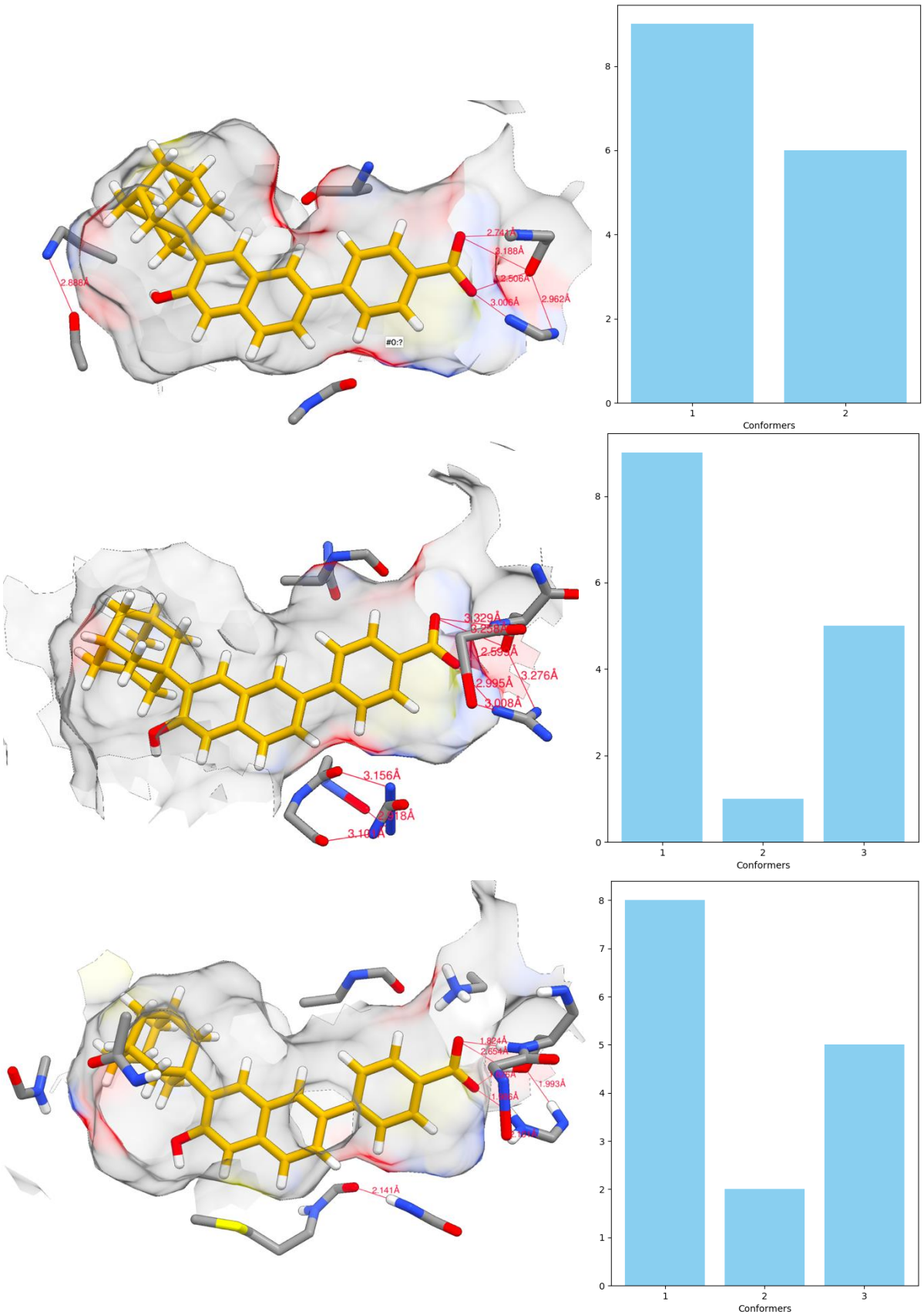


Figure 29 CD1530 perforates the binding surface in both RAR-α and γ, but is still γ-selective.

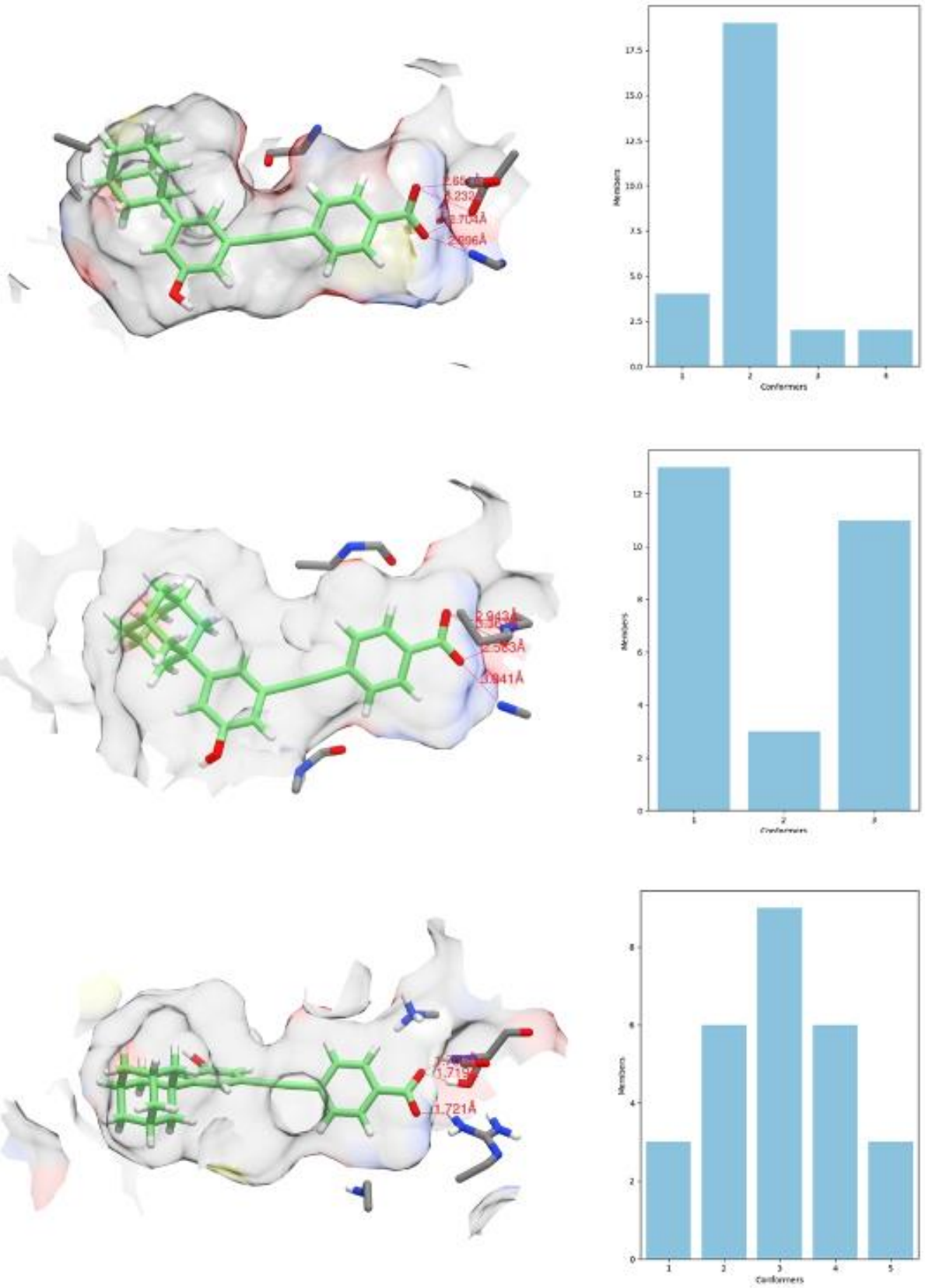


Figure 30 Compound 28 only fits in RARα by angling the only group that can form the mid-section h bond away from the serine group.

iii. Summary of Selective Compounds:

Despite being members of the same family of nuclear receptors, the retinoic acid receptors are quite unique in their characteristics and preferences for a ligand. RARβ has a much larger LBD that can accommodate correspondingly larger compounds, whereas RARγ, narrow and shorter than the others, limits the variety of compounds with which it is compatible, all the while optimizing hydrogen bond distance with the well conserved terminal polar cluster. RARα, being of intermediate width and length, is unique in its polar midsection.

From the observations made so far of interactions with their corresponding selective ligands, it is quite clear that each receptor has its penchants and foibles and that the likely reasons for selectivity can be observed reasonably well using images of the key conformers produced by the clustering algorithm.

There is competing equilibrium between the number of hydrogen bonds accessible to a ligand, the mean distance of such bonds and the positioning of a particular bond. Most compounds with RARα selectivity produce much shorter and stronger interactions with RARγ and yet have a higher affinity for alpha. The only observable reason for such selectivity is the use of the mid-cavity serine residue. The idea of this interaction acting as a potent override for seemingly weaker overall energetics than in RARγ is bolstered by the observation that compounds seeking selectivity with the other isotypes, in one way or another, avoid this interaction.

Beta selective compounds, as previously commented upon, aren't necessarily of high beta affinity, just of lesser affinity for the others. These compounds aren't tailored for beta binding, instead characteristics appropriate for the other two are removed. If these compounds only fit in beta then they are beta selective. While this may be an oversimplification, one should note that this conclusion is derived from repeated qualitative observation alone.

A similar approach for gamma selective compounds is observed in which the compounds are made devoid of mid-section polarity or augmented in such a way that this group is positioned to minimise this interaction with RARα.

In this manner, the receptors are observed to prioritise their ligands in terms of mid-section polarity, sterics and a lack of either combined with optimal bond distance for RARα, RARβ and RARγ respectively.

9. FINAL PREDICTION

With the ability of the algorithm to rank a series of compounds with good accuracy established, all compounds currently under investigation by the Whiting-Pohl group and others are here docked in each of the retinoic acid receptors and their resulting solutions clustered and scored for ranking. The top 30 compounds for each isotype are shown below.

Table 1 Predicted ranking of binding affinity derived from the chemscore score of the key conformer from each compound.

RARα		RARβ		RARγ	
Compound	**Score**	**Compound**	**Score**	**Compound**	**Score**
Adapalene	63.13	CD2665	64.5	CD2665	65.46
CD2665	61.28	Adapalene	63.26	Adapalene	64.45
TTNPB	61.08	Trifarotene	61.9	TTNPB	63.56
TTNN	61.06	TTNN	60.55	DC271	61.87
DC667	58.83	TTNPB	58.33	DC673	61.46
Trifarotene	58.49	DC271	58.18	DC667	61.16
DC656	58.24	DC559	57.78	TTNN	61.12
EC23	58.19	DC567	57.63	DC559	60.34
DC271	58.15	DC667	57.63	DC646	60.14
DC4672	58.02	DC673	57.28	EC23	60.04
DC559	57.9	DC661	57.26	DC564	59.99
Fenretinide	57.44	DC479	57.22	DC547	59.78
AM580	57.39	DC564	57.19	DC479	59.68
DC686	57.1	DC318	57.11	DC525	59.55
DC525	57.04	GZ25	57.02	Bexarotene	59.47
DC360	56.94	Bexarotene	56.9	DC567	59.14
DC128	56.77	DC686	56.73	DC661	58.98
DC527	56.71	DC360	56.62	DC360	58.95
DC479	56.27	EC23	56.2	DC444	58.83
DC567	56.05	DC656	56.14	DC128	58.46
DC528	56.02	EC19	56.01	AH61	58.42

DC564	55.96	DC525	55.94	DC656	58.39
DC716	55.92	EC23Al	55.92	GZ25	58.18
EC23Al	55.89	DC646	55.72	EC23Al	58.17
DC673	55.76	DC547	55.51	DC528	58.02
DC122	55.73	DC526	55.26	AM580	57.87
Tazarotene	55.71	DC540	55.25	DC540	57.8
DC646	55.69	DC657	55.25	EC19	57.75
DC661	55.55	DC716	55.16	DC526	57.48
DC526	55.48	DC128	54.85	DC527	57.42

Interestingly, all-trans-retinoic-acid and its other stereoisomers did not make it into the top 30 compounds in any of the retinoic acid receptor isotypes. The implication being that each of these compounds that are in the top are predicted to demonstrate better binding than ATRA.

The alpha selective AM580 is much higher up the rank order in RARα than in RARγ and did not make it into the top 30 in RARβ, corresponding to its observed higher selectivity over β than γ. Interestingly, the score given to AM580 in RARγ is higher than in RARα despite its lower rank. What is significant here is that, as concluded in the previous chapter, the score given to these compounds by GOLD across the proteins cannot be used to derive selectivity, but it seems that when put into perspective by a large group of retinoids, it may be possible for selectivity to be hinted at by the relative rank position of a compound in each of the proteins. Albeit with varying and probably random success.

Some compounds are indicated to be strong candidates consistently. CD2665, adapalene and TTNPB are all in the top 5 across all the receptor isotypes. With the algorithm pipeline indicated to be a potential predictor of rank order, it is expected that these compounds would have better binding interactions than, for example, the corresponding bottom 5 compounds for each of the respective isotypes.

The synthetic retinoids developed by this group are in good competition with other commercially available compounds according to this prediction. A good next step would be the

experimental determination of binding constants for all the compounds presented here to corroborate or disprove the rank order given by the algorithmic pipeline.

10. CONCLUSIONS

i. A new docking pipeline

Here was presented a new method of analysing the output from a computational simulation using by way of clustering the results. The clustering algorithm was used as a way of summarising the output from docking in a statistical manner and it was shown that this method is valid and useful in a variety of applications.

The algorithm works by the pairwise comparison of two solutions, finding their root-mean-squared-deviation of atomic positions. In doing so it can organise the solutions from a simulated docking into groups of like-compounds. But this data alone is limited and it is with the contributions from the Chemscore system employed by GOLD that this data becomes actionable.

In this study, the clustering alrogithm was introduced as a novel pipeline process. Instead of the docking of a single input file for each compound, as suggested by GOLD, multiple unique starting points of the same compound were docked. The resulting many solutions were therefore collectively exposed to as many interactions as possible, giving dominant forces the opportunity to win out in the resulting Key conformer.

ii. Replicating the geometry of bound compounds

The algorithm was shown to successfully predict the geometries of known bound compounds in their respective crystal structures. In all cases, the geometry was accurately predicted and was more accurate than simply taking the highest scoring solution from GOLD. It was also shown that the highest scoring solution from GOLD alone only replicated the highest occuring conformer in some cases with an accuracy of 40%.

The implication of this alone is meaningful. Researchers in a lab setting can quickly dock a series of compounds and know with confidence that the resulting Key conformer is the most accurate prediction available with desktop docking software. This information is invaluable to such

researchers. An extra layer of understanding is now cheaply available and exceedingly accessible. The entire process from start to finish is possible on a single desktop or laptop.

This new tool is also essential for understanding the biology of compounds already made. The ability to pair biological data confidently with an image of the very interaction is impactful.

iii. Predicting affinity ranking

The pipeline process suggested here suggests two major extra steps in computational docking, the creation of a library of starting compounds and the clustering of final solutions. The addition of steps is in itself a mark against this process. However, the results are incredibly useful and surprisingly accurate. This pipeline was consistently a good indicator of rank order of binding affinity for a series of compounds. In rarγ expecially, the pipeline yielded good correlation with experimental data and reached excellent statistical significance.

These results were not achieved alone, though. It must be emphasized that this pipeline would not exist without a scoring system such as Chemscore. It is not built as a replacement to the GOLD scoring system. Rather, it was built to maximise the use of Chemscore by running it as many times as possible and taking the score from the endpoint which Chemscore arrived at most often. These two pieces work together as pieces of a puzzle to derive the overall landscape of interactions available to a given compound-protein complex.

iv. Selectivity

Limitations of this pipeline were reached when it was applied to deriving selectivity. It is out of the remit of desktop software to predict and compare interactions from a single compound with several proteins and so any quantitative prediction of selectivity is not possible.

But while this data alone cannot be used to predict selectivity, it was shown that visual analysis of key conformers is still useful. By visual inspection, logical reasons for selectivity, consistent with the literature were possible to derive in RARα. Combined with additional biological data, the Key conformer can assist in the discovery of features necessary for selectivity.

Additionally, when large banks of compounds were ranked by score for each isotype, the changing rank for each compound in each of the receptors could be indicative of selectivity. However the repeatability of this observation and accuracy requires further investigation.

v. Future work

Having shown the ability of the python script to succesfully cluster solutions, further studies into the significance of cluster sizes should take place. Regression analyses to investigate association between cluster size and the correct geometry could be performed, as well as regression analyses of clustering and native chemscore to investigate the relative association between the two predictive methods and the correct solutions.

The python script, 18ct.py, itself will be adapted and converted into a jupyter notebook and a google collab notebook for ease of use and access. When adapted, it will be made available on github under the page https://github.com/alexrodri01 .

11. BIBLIOGRAPHY

(1) Yedavalli, V. S.; Patil, A.; Shah, P. Amyotrophic Lateral Sclerosis and its Mimics/Variants: A Comprehensive Review. *J Clin Imaging Sci* **2018**, *8*, 53. DOI: 10.4103/jcis.JCIS_40_18.

(2) Walling, A. D. Amyotrophic lateral sclerosis: Lou Gehrig's disease. *Am Fam Physician* **1999**, *59* (6), 1489-1496.

(3) Mehta, P.; Kaye, W.; Raymond, J.; Punjani, R.; Larson, T.; Cohen, J.; Muravov, O.; Horton, K. Prevalence of Amyotrophic Lateral Sclerosis - United States, 2015. *MMWR Morb Mortal Wkly Rep* **2018**, *67* (46), 1285-1289. DOI: 10.15585/mmwr.mm6746a1.

(4) Kiernan, M. C.; Vucic, S.; Cheah, B. C.; Turner, M. R.; Eisen, A.; Hardiman, O.; Burrell, J. R.; Zoing, M. C. Amyotrophic lateral sclerosis. *Lancet* **2011**, *377* (9769), 942-955. DOI: 10.1016/S0140-6736(10)61156-7.

(5) Andersen, P. M.; Abrahams, S.; Borasio, G. D.; de Carvalho, M.; Chio, A.; Van Damme, P.; Hardiman, O.; Kollewe, K.; Morrison, K. E.; Petri, S.; et al. EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS)--revised report of an EFNS task force. *Eur J Neurol* **2012**, *19* (3), 360-375. DOI: 10.1111/j.1468-1331.2011.03501.x.

(6) Zarei, S.; Carr, K.; Reiley, L.; Diaz, K.; Guerra, O.; Altamirano, P. F.; Pagani, W.; Lodin, D.; Orozco, G.; Chinea, A. A comprehensive review of amyotrophic lateral sclerosis. *Surg Neurol Int* **2015**, *6*, 171. DOI: 10.4103/2152-7806.169561.

(7) Rosen, D. R.; Siddique, T.; Patterson, D.; Figlewicz, D. A.; Sapp, P.; Hentati, A.; Donaldson, D.; Goto, J.; O'Regan, J. P.; Deng, H. X. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **1993**, *362* (6415), 59-62. DOI: 10.1038/362059a0.

(8) Morrice, J. R.; Gregory-Evans, C. Y.; Shaw, C. A. Animal models of amyotrophic lateral sclerosis: A comparison of model validity. *Neural Regen Res* **2018**, *13* (12), 2050-2054. DOI: 10.4103/1673-5374.241445.

(9) Paspe Cruz, M. Edaravone (Radicava)

A Novel Neuroprotective Agent for the Treatment of Amyotrophic Lateral Sclerosis. *Drug Forecast* **2018**, *43*. Jaiswal, M. K. Riluzole and edaravone: A tale of two amyotrophic lateral sclerosis drugs. *Med Res Rev* **2019**, *39* (2), 733-748. DOI: 10.1002/med.21528.

(10) Cruz, M. P. Edaravone (Radicava): A Novel Neuroprotective Agent for the Treatment of Amyotrophic Lateral Sclerosis. *P T* **2018**, *43* (1), 25-28.

(11) Doan, T. B.; Graham, J. D.; Clarke, C. L. Emerging functional roles of nuclear receptors in breast cancer. *J Mol Endocrinol* **2017**, *58* (3), R169-R190. DOI: 10.1530/JME-16-0082. Dhiman, V. K.; Bolt, M. J.; White, K. P. Nuclear receptors in cancer - uncovering new and evolving roles through genomic analysis. *Nat Rev Genet* **2018**, *19* (3), 160-174. DOI: 10.1038/nrg.2017.102.

(12) Trigo, D.; Goncalves, M. B.; Corcoran, J. P. T. The regulation of mitochondrial dynamics in neurite outgrowth by retinoic acid receptor β signaling. *FASEB J* **2019**, *33* (6), 7225-7235. DOI: 10.1096/fj.201802097R.

(13) Clark, J. N.; Whiting, A.; McCaffery, P. Retinoic acid receptor-targeted drugs in neurodegenerative disease. *Expert Opin Drug Metab Toxicol* **2020**, *16* (11), 1097-1108. DOI: 10.1080/17425255.2020.1811232.

(14) Rojas, F.; Cortes, N.; Abarzua, S.; Dyrda, A. v. Z., Brigitte. Astrocytes expressing mutant SOD1 and TDP43 trigger motoneuron death that is mediated via sodium channels and nitroxidative stress. *Frontiers in Cellular Neuroscience* **2014**, *8* (24).

(15) Balendra, R.; M. Isaacs, A. C9orf72 -mediated ALS and FTD: multiple pathways to disease. *Nat Rev Neurol* **2018**

14 (9

), 14.

(16) Leko, M. B.; Župunski, V.; Kirincich, J.; Smilović, D.; Hortobágyi, T.; Hof, P. R. Š., Goran. Molecular Mechanisms of Neurodegeneration Related to C9orf72 Hexanucleotide Repeat Expansion. **2019**.

(17) Gautam, M.; Jara, J. H.; Kocak, N.; Rylaarsdam, L. E.; Kim, K. D.; Bigio, E. H.; Hande Özdinler, P. Mitochondria, ER, and nuclear membrane defects reveal early mechanisms for upper motor neuron vulnerability with respect to TDP-43 pathology. *Acta Neuropathol* **2019**, *137* (1), 47-69. DOI: 10.1007/s00401-018-1934-8.

(18) Millecamps, S.; Jean-Pierre, J. Axonal transport deficits and neurodegenerative diseases. *Nature Reviews* **2013**, *14*.

(19) Minako, T.; Hisako, S.; Mika, T.; Shigeyoshi, I.; Ryong-Moon, S.; Masami, M.; Masao, M.; Toshihiko, A.; Makoto, U.; Hidemi, M.; et al. Calcium-permeable AMPA receptors promote misfolding of mutant SOD1 protein and development of amyotrophic lateral sclerosis in a transgenic mouse model. *Human Molecular Genetics* **2004**, *13* (19), 13.

(20) Pansarasa, O.; Bordoni, M.; Diamanti, L.; Sproviero, D.; Gagliardi, S. C., Cristina. SOD1 in Amyotrophic Lateral Sclerosis: "Ambivalent" Behavior Connected to the Disease. *International Journal of Molecular Sciences* **2018**, *19*.

(21) Michal, I.; Shalom Guy, S.; Itskovitz-Eldor, J.; Michel, R. Astrocytes in Pathogenesis of ALS Disease and Potential Translation into Clinic. In *Astrocyte*.

(22) Saccon, R. A.; Bunton-Stasyshyn, R. K.; Fisher, E. M.; Fratta, P. Is SOD1 loss of function involved in amyotrophic lateral sclerosis? *Brain* **2013**, *136* (Pt 8), 2342-2358. DOI: 10.1093/brain/awt097.

(23) Grlach a, A.; Bertram a, K.; Hudecova c, S.; Krizanova , O. Calcium and ROS: A mutual interplay. *Redox Biology* **2015**, *6*.

(24) Zhou, J.; Li, A.; Li, X.; Yi, J. Dysregulated mitochondrial Ca. *Arch Biochem Biophys* **2019**, *663*, 249-258. DOI: 10.1016/j.abb.2019.01.024.

(25) Gao, J.; Wang, L.; Huntley, M. L.; Perry, G.; Wang, X. Pathomechanisms of TDP-43 in neurodegeneration. *J Neurochem* **2018**. DOI: 10.1111/jnc.14327.

(26) Cohen, T. J.; Lee, V. M.; Trojanowski, J. Q. TDP-43 functions and pathogenic mechanisms implicated in TDP-43 proteinopathies. *Trends Mol Med* **2011**, *17* (11), 659-667. DOI: 10.1016/j.molmed.2011.06.004.

(27) Glass, J. D. Stathmin-2: adding another piece to the puzzle of TDP-43 proteinopathies and neurodegeneration. *J Clin Invest* **2020**, *130* (11), 5677-5680. DOI: 10.1172/JCI142854.

(28) Neumann, M.; Sampathu, D. M.; Kwong, L. K.; Truax, A. C.; Micsenyi, M. C.; Chou, T. T.; Bruce, J.; Schuck, T.; Grossman, M.; Clark, C. M.; et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **2006**, *314* (5796), 130-133. DOI: 10.1126/science.1134108.

(29) Guo, W.; Naujock, M.; Ordovs , L.; Vandoorne, T.; Vanden Berghe , P.; Boon, R.; Patel, A.; Welters, M.; Vanwelden, T.; Geens, N.; et al. HDAC6 inhibition reverses axonal transport defects in motor neurons derived from FUS-ALS patients. *Nature Communications* **2017**, *8* (861).

(30) Atkin, J. D.; Farg, M. A.; Walker, A. K.; McLean, C.; Tomas, D.; Horne, M. K. Endoplasmic reticulum stress and induction of the unfolded protein response in human sporadic amyotrophic lateral sclerosis. *Neurobiol Dis* **2008**, *30* (3), 400-407. DOI: 10.1016/j.nbd.2008.02.009. Jaronen, M.; Goldsteins , G.; Koistinaho, J. ER stress and unfolded protein response in amyotrophic lateral sclerosis—a controversial role of protein disulphide isomerase. *Frontiers in Cellular Neuroscience* **2014**.

B Medinas, D.; V Gonzlez, J.; Falcon, P.; Hetz, C. Fine-Tuning ER Stress Signal Transducers to Treat Amyotrophic Lateral Sclerosis. *Frontiers in Molecular Neuroscience* **2017**, *10*.

(31) Scotter, E. L.; Chen, H. J.; Shaw, C. E. TDP-43 Proteinopathy and ALS: Insights into Disease Mechanisms and Therapeutic Targets. *Neurotherapeutics* **2015**, *12* (2), 352-363. DOI: 10.1007/s13311-015-0338-x.

(32) Cohen, T. J.; Hwang, A. W.; Unger, T.; Trojanowski, J. Q. L., Virginia MY. Redox signalling directly regulates TDP-43 via cysteine oxidation and disulphide cross-linking. *The EMBO journal* **2012**, *31*.

(33) C Vatsavayai, S.; L Nana, A.; S Yokoyama, J.; W Seeley, W. C9orf72 -FTD/ALS pathogenesis: evidence from human neuropathological studies. *Acta Neuropathologica* **2019**

(34) Tran, H.; Almeida, S.; Moore, J.; Gendron, T. F.; Chalasani, U.; Lu, Y.; Du, X.; Nickerson, J. A.; Petrucelli, L.; Weng, Z.; et al. Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. *Neuron* **2015**, *87* (6), 1207-1214. DOI: 10.1016/j.neuron.2015.09.015.

(35) Pöyhönen, S.; Er, S.; Domanskyi, A.; Airavaara, M. Effects of Neurotrophic Factors in Glial Cells in the Central Nervous System: Expression and Properties in Neurodegeneration and Injury. *Front Physiol* **2019**, *10*, 486. DOI: 10.3389/fphys.2019.00486.

(36) Zhou, T.; Huang, Z.; Sun, X.; Zhu, X.; Zhou, L.; Li, M.; Cheng, B. L., Xialin. Microglia Polarization with M1/M2

Phenotype Changes in rd1 Mouse

Model of Retinal Degeneration. *Frontiers in Neuroanatomy* **2017**, *11* (77).

(37) Rojas, F.; Cortes, N.; Abarzua, S.; Dyrda, A.; van Zundert, B. Astrocytes expressing mutant SOD1 and TDP43 trigger motoneuron death that is mediated via sodium channels and nitroxidative stress. *frontiers in cellular neuroscience* **2014**, *8* (24).

(38) Reddy, S.; Seth, R. Role of Tumor Necrosis Factor in Neurodegeneration. *EC Endocrinology and Metabolic Research* **2019**.

(39) Vercelli, A.; Mereuta, O. M.; Garbossa, D.; Muraca, G.; Mareschi, K.; Rustichelli, D.; Ferrero, I.; Mazzini, L.; Madon, E.; Fagioli, F. Human mesenchymal stem cell transplantation extends survival, improves motor performance and decreases neuroinflammation in mouse model of amyotrophic lateral sclerosis. *Neurobiol Dis* **2008**, *31* (3), 395-405. DOI: 10.1016/j.nbd.2008.05.016.

(40) Clark, J. A.; Yeaman, E. J.; Blizzard, C. A.; Chuckowree, J. A.; Dickson, T. C. A Case for Microtubule Vulnerability in Amyotrophic Lateral Sclerosis: Altered Dynamics During Disease. *Front Cell Neurosci* **2016**, *10*, 204. DOI: 10.3389/fncel.2016.00204.

(41) https://clinicaltrials.gov (accessed.

(42) Wen, D.; Cui, C.; Duan, W.; Wang, W.; Wang, Y.; Liu, Y.; Li, Z.; Li, C. The role of insulin-like growth factor 1 in ALS cell and mouse models: A mitochondrial protector. *Brain Res Bull* **2019**, *144*, 1-13. DOI: 10.1016/j.brainresbull.2018.09.015.

(43) Robson, D.; Welch, E.; Beeching, N. J.; Gill, G. V. Consequences of captivity: health effects of far East imprisonment in World War II. *QJM* **2009**, *102* (2), 87-96. DOI: 10.1093/qjmed/hcn137.

(44) Yeh, L.; Bonati, L. M.; Silverberg, N. B. Topical retinoids for acne. *Semin Cutan Med Surg* **2016**, *35* (2), 50-56. DOI: 10.12788/j.sder.2016.024.

(45) Nagpal, S.; Friant, S.; Nakshatri, H.; Chambon, P. RARs and RXRs: evidence for two autonomous transactivation functions (AF-1 and AF-2) and heterodimerization in vivo. *EMBO J* **1993**, *12* (6), 2349-2360. Pohl, E.; Tomlinson, C. W. E. Classical pathways of gene regulation by retinoids. *Methods Enzymol* **2020**, *637*, 151-173. DOI: 10.1016/bs.mie.2020.03.008.

(46) Haffez, H.; Chisholm, D. R.; Valentine, R.; Pohl, E.; Redfern, C.; Whiting, A. The molecular basis of the interactions between synthetic retinoic acid analogues and the retinoic acid receptors. *Medchemcomm* **2017**, *8* (3), 578-592. DOI: 10.1039/c6md00680a.

(47) de Lera, A. R.; Bourguet, W.; Altucci, L.; Gronemeyer, H. Design of selective nuclear receptor modulators: RAR and RXR as a case study. *Nat Rev Drug Discov* **2007**, *6* (10), 811-820. DOI: 10.1038/nrd2398.

(48) Sommer, K. M.; Chen, L. I.; Treuting, P. M.; Smith, L. T.; Swisshelm, K. Elevated retinoic acid receptor beta(4) protein in human breast tumor cells with nuclear and cytoplasmic localization. *Proc Natl Acad Sci U S A* **1999**, *96* (15), 8651-8656. DOI: 10.1073/pnas.96.15.8651.

(49) Chisholm, D. R.; Whiting, A. Design of synthetic retinoids. *Methods Enzymol* **2020**, *637*, 453-491. DOI: 10.1016/bs.mie.2020.02.011.

(50) Clagett-Dame, M.; McNeill, E. M.; Muley, P. D. Role of all-trans retinoic acid in neurite outgrowth and axonal elongation. *J Neurobiol* **2006**, *66* (7), 739-756. DOI: 10.1002/neu.20241.

(51) Khatib, T.; Marini, P.; Nunna, S.; Chisholm, D. R.; Whiting, A.; Redfern, C.; Greig, I. R.; McCaffery, P. Genomic and non-genomic pathways are both crucial for peak induction of neurite outgrowth by retinoids. *Cell Commun Signal* **2019**, *17* (1), 40. DOI: 10.1186/s12964-019-0352-4.

(52) Tomlinson, C. W. E.; Cornish, K. A. S.; Whiting, A.; Pohl, E. Structure-functional relationship of cellular retinoic acid-binding proteins I and II interacting with natural and synthetic ligands. *Acta Crystallogr D Struct Biol* **2021**, *77* (Pt 2), 164-175. DOI: 10.1107/S2059798320015247.
(53) Brummer, T.; Müller, S. A.; Pan-Montojo, F.; Yoshida, F.; Fellgiebel, A.; Tomita, T.; Endres, K.; Lichtenthaler, S. F. NrCAM is a marker for substrate-selective activation of ADAM10 in Alzheimer's disease. *EMBO Mol Med* **2019**, *11* (4). DOI: 10.15252/emmm.201809695.
(54) Du, X.; Li, Y.; Xia, Y. L.; Ai, S. M.; Liang, J.; Sang, P.; Ji, X. L.; Liu, S. Q. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int J Mol Sci* **2016**, *17* (2). DOI: 10.3390/ijms17020144.
(55) Tatum, N. J.; Duarte, F.; Kamerlin, S. C. L.; Pohl, E. Relative Binding Energies Predict Crystallographic Binding Modes of Ethionamide Booster Lead Compounds. *J Phys Chem Lett* **2019**, *10* (9), 2244-2249. DOI: 10.1021/acs.jpclett.9b00741.
(56) Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. L. Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem* **2015**, *8*, 37-47. DOI: 10.2147/AABC.S70333.
(57) Stanzione, F.; Giangreco, I.; Cole, J. C. Use of molecular docking computational tools in drug discovery. *Prog Med Chem* **2021**, *60*, 273-343. DOI: 10.1016/bs.pmch.2021.01.004.
(58) Karplus, M. Development of Multiscale Models for Complex Chemical Systems From H+H 2 to Biomolecules. 2013

(59) Śledź, P.; Caflisch, A. Protein structure-based drug design: from docking to molecular dynamics. *Curr Opin Struct Biol* **2018**, *48*, 93-102. DOI: 10.1016/j.sbi.2017.10.010.
(60) Cuello-Rodriguez. https://github.com/alexrodri01 (accessed.
(61) le Maire, A.; Teyssier, C.; Erb, C.; Grimaldi, M.; Alvarez, S.; de Lera, A. R.; Balaguer, P.; Gronemeyer, H.; Royer, C. A.; Germain, P.; et al. A unique secondary-structure switch controls constitutive gene repression by retinoic acid receptor. *Nat Struct Mol Biol* **2010**, *17* (7), 801-807. DOI: 10.1038/nsmb.1855.
(62) Germain, P.; Kammerer, S.; Pérez, E.; Peluso-Iltis, C.; Tortolani, D.; Zusi, F. C.; Starrett, J.; Lapointe, P.; Daris, J. P.; Marinier, A.; et al. Rational design of RAR-selective ligands revealed by RARbeta crystal stucture. *EMBO Rep* **2004**, *5* (9), 877-882. DOI: 10.1038/sj.embor.7400235.
(63) Renaud, J. P.; Rochel, N.; Ruff, M.; Vivat, V.; Chambon, P.; Gronemeyer, H.; Moras, D. Crystal structure of the RAR-gamma ligand-binding domain bound to all-trans retinoic acid. *Nature* **1995**, *378* (6558), 681-689. DOI: 10.1038/378681a0.

(64) Vaezeslami, S.; Mathes, E.; Vasileiou, C.; Borhan, B.; Geiger, J. H. The structure of Apo-wild-type cellular retinoic acid binding protein II at 1.4 A and its relationship to ligand binding and nuclear translocation. *J Mol Biol* **2006**, *363* (3), 687-701. DOI: 10.1016/j.jmb.2006.08.059.

(65) Vasileiou, C.; Vaezeslami, S.; Crist, R. M.; Rabago-Smith, M.; Geiger, J. H.; Borhan, B. Protein design: reengineering cellular retinoic acid binding protein II into a rhodopsin protein mimic. *J Am Chem Soc* **2007**, *129* (19), 6140-6148. DOI: 10.1021/ja067546r.

(66) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **2003**, *52* (4), 609-623. DOI: 10.1002/prot.10465.

(67) Motani, A.; Wang, Z.; Conn, M.; Siegler, K.; Zhang, Y.; Liu, Q.; Johnstone, S.; Xu, H.; Thibault, S.; Wang, Y.; et al. Identification and characterization of a non-retinoid ligand for retinol-binding protein 4 which lowers serum retinol-binding protein 4 levels in vivo. *J Biol Chem* **2009**, *284* (12), 7673-7680. DOI: 10.1074/jbc.M809654200.

(68) Liu, P.; Peng, C.; Chen, X.; Wu, L.; Yin, M.; Li, J.; Qin, Q.; Kuang, Y.; Zhu, W. Acitretin Promotes the Differentiation of Myeloid-Derived Suppressor Cells in the Treatment of Psoriasis. *Front Med (Lausanne)* **2021**, *8*, 625130. DOI: 10.3389/fmed.2021.625130.

(69) Khatib, T.; Whiting, A.; Chisholm, D. R.; Redfern, C.; Müller, B.; McCaffery, P. A Bioluminescence Reporter Assay for Retinoic Acid Control of Translation of the GluR1 Subunit of the AMPA Glutamate Receptor. *Mol Neurobiol* **2019**, *56* (10), 7074-7084. DOI: 10.1007/s12035-019-1571-9.

(70) Bernerd, F.; Ortonne, J. P.; Bouclier, M.; Chatelus, A.; Hensby, C. The rhino mouse model: the effects of topically applied all-trans retinoic acid and CD271 on the fine structure of the epidermis and utricle wall of pseudocomedones. *Arch Dermatol Res* **1991**, *283* (2), 100-107. DOI: 10.1007/BF00371617.

(71) Heusinkveld, H. J.; Schoonen, W. G.; Hodemaekers, H. M.; Nugraha, A.; Sirks, J. J.; Veenma, V.; Sujan, C.; Pennings, J. L. A.; Wackers, P. F.; Palazzolo, L.; et al. Distinguishing mode of action of compounds inducing craniofacial malformations in zebrafish embryos to support dose-response modeling in combined exposures. *Reprod Toxicol* **2020**, *96*, 114-127. DOI: 10.1016/j.reprotox.2020.06.002.

(72) Hacioglu, C.; Kar, F.; Kacar, S.; Sahinturk, V.; Kanbak, G. Bexarotene inhibits cell proliferation by inducing oxidative stress, DNA damage and apoptosis via PPARγ/ NF-κB signaling pathway in C6 glioma cells. *Med Oncol* **2021**, *38* (3), 31. DOI: 10.1007/s12032-021-01476-z.

(73) Goncalves, M. B.; Clarke, E.; Jarvis, C. I.; Barret Kalindjian, S.; Pitcher, T.; Grist, J.; Hobbs, C.; Carlstedt, T.; Jack, J.; Brown, J. T.; et al. Discovery and lead optimisation of a potent, selective and orally bioavailable RARβ agonist for the potential treatment of nerve injury. *Bioorg Med Chem Lett* **2019**, *29* (8), 995-1000. DOI: 10.1016/j.bmcl.2019.02.011.

(74) le Maire, A.; Teyssier, C.; Balaguer, P.; Bourguet, W.; Germain, P. Regulation of RXR-RAR Heterodimers by RXR- and RAR-Specific Ligands and Their Combinations. *Cells* **2019**, *8* (11). DOI: 10.3390/cells8111392.

(75) Torrisi, R.; Pensa, F.; Orengo, M. A.; Catsafados, E.; Ponzani, P.; Boccardo, F.; Costa, A.; Decensi, A. The synthetic retinoid fenretinide lowers plasma insulin-like growth factor I levels in breast cancer patients. *Cancer Res* **1993**, *53* (20), 4769-4771.

(76) Han, J. J.; Faletsky, A.; Barbieri, J. S.; Mostaghimi, A. New Acne Therapies and Updates on Use of Spironolactone and Isotretinoin: A Narrative Review. *Dermatol Ther (Heidelb)* **2021**, *11* (1), 79-91. DOI: 10.1007/s13555-020-00481-w.

(77) Crespo, A.; Fernández, A. Induced disorder in protein-ligand complexes as a drug-design strategy. *Mol Pharm* **2008**, *5* (3), 430-437. DOI: 10.1021/mp700148h.

(78) Chang, C. E.; Chen, W.; Gilson, M. K. Ligand configurational entropy and protein binding. *Proc Natl Acad Sci U S A* **2007**, *104* (5), 1534-1539. DOI: 10.1073/pnas.0610494104.

(79) Beard, R. L.; Chandraratna, R. A. S. RAR-Selective Ligands: Receptor Subtype and Function Selectivity. In *Retinoids: The Biochemical and Molecular Basis of Vitamin A and Retinoid Action*, Nau, H., Blaner, W. S. Eds.; Springer Berlin Heidelberg, 1999; pp 185-213.

(80) Clarke, E.; Jarvis, C. I.; Goncalves, M. B.; Kalindjian, S. B.; Adams, D. R.; Brown, J. T.; Shiers, J. J.; Taddei, D. M. A.; Ravier, E.; Barlow, S.; et al. Design and synthesis of a potent, highly selective, orally bioavailable, retinoic acid receptor alpha agonist. *Bioorg Med Chem* **2018**, *26* (4), 798-814. DOI: 10.1016/j.bmc.2017.12.015.

(81) Di Rocco, A.; Uchibe, K.; Larmour, C.; Berger, R.; Liu, M.; Barton, E. R.; Iwamoto, M. Selective Retinoic Acid Receptor γ Agonists Promote Repair of Injured Skeletal Muscle in Mouse. *Am J Pathol* **2015**, *185* (9), 2495-2504. DOI: 10.1016/j.ajpath.2015.05.007.