

UNIVERSITY OF RIJEKA

FACULTY OF BIOTECHNOLOGY AND DRUG DEVELOPMENT

Master's university study

Biotechnology in medicine

Lana Žoldoš

**Comparative analysis of the aryl hydrocarbon receptor gene between  
*Homo sapiens sapiens* and *Homo sapiens neanderthalensis*.**

Master's thesis

Rijeka, 2025.

UNIVERSITY OF RIJEKA

FACULTY OF BIOTECHNOLOGY AND DRUG DEVELOPMENT

Master's university study

Biotechnology in medicine

Lana Žoldoš

**Comparative analysis of the aryl hydrocarbon receptor gene between  
*Homo sapiens sapiens* and *Homo sapiens neanderthalensis*.**

Master's thesis

Rijeka, 2025.

Mentor: prof. dr. sc. Ivor Janković

Co-mentor: prof. dr. sc. Željka Maglica

SVEUČILIŠTE U RIJECI  
FAKULTET BIOTEHNOLOGIJE I RAZVOJA LIJEKOVA  
Diplomski studij  
Biotehnologija u medicini

Lana Žoldoš

**Komparativna analiza gena koji kodira za receptor arilnih ugljikovodika  
između *Homo sapiens sapiens* and *Homo sapiens neanderthalensis***

Diplomski rad

Rijeka, 2025.

Mentor: prof. dr. sc. Ivor Janković

Komentorica: prof. dr. sc. Željka Maglica

Master's thesis was  
defended on in front of the  
committee:

- 1.
- 2.
- 3.

This thesis has 58 pages, 21 figures, 1 table and 49 references.

## **ACKNOWLEDGMENTS**

I sincerely thank everyone who contributed in any way to the creation of this thesis and supported the implementation of the research.

I owe special gratitude to my mentor, Prof. Dr. Sc. Ivor Janković, who, to me, is not only an exceptional scientist but also a person of great integrity and selfless support.

I would also like to thank Prof. Dr. Sc. Željka Maglica for her warmth, advice, and encouraging words, which were a valuable source of support throughout the entire research process.

My gratitude extends to Prof. Dr. Sc. Pere Gelabert, who made a significant contribution to the part of this work related to genetic research and database analysis, generously sharing his knowledge and expertise. Many thanks go to Assistant David Visentin, from whom I learned almost everything I know about molecular docking. David followed the entire course of this research, shared his knowledge, and introduced me to the world of molecular modeling and working on HPC systems. His dedication and approach left a strong impression; if I ever have the opportunity to teach someone, I hope to do so with the same commitment and clarity as David. I would also like to thank David for all the encouraging words and support.

Special thanks go to Dr. Sc. Mario Lovrić, whose support had a great impact on my personal and academic development, as well as on recognizing and shaping my own affinities and research interests. Through working with Dr. Sc. Lovrić, I had the opportunity to participate in exciting projects that significantly shaped and deepened my motivation and desire for further scientific work. It is difficult to express in words my gratitude for the trust, opportunities, and support.

The writing of this thesis was supported by the COST Action *Integrating Neanderthal Legacy – from past to present* (grant: CA19141) and the Horizon project *EDIAQI* (grant: 101057497), which focuses on indoor air quality research.

## SUMMARY

The aryl hydrocarbon receptor (AHR), a regulator of xenobiotic metabolism and detoxication, arose in early metazoans and has a deep evolutionary history. In living humans (*Homo sapiens sapiens*), the unique Val381 substitution characterizes AHR from extinct hominins, like Neanderthals (*Homo sapiens neanderthalensis*), which possess the ancestral Ala381 residue. This substitution is also anticipated to be an evolutionary adaptation to increased exposure to polycyclic aromatic hydrocarbons (PAHs) due to cultural developments like controlled fire burning, which escalated during human evolution. In this way, Val381 may regulate AHR activity in humans compared to other animals, possibly reducing AHR activation and resulting toxic responses. Current data are contradictory: two in vitro studies differ over whether the human AHR possesses lower ligand-binding affinity than Neanderthal AHR. This study addresses this disparity through *in silico* molecular docking of 42,995 ligands—that is, from AHR modulators, microbial metabolites, bioactive molecules, and natural compound datasets—to investigate binding affinities and interaction modes of the two species. The results aim to clarify the functional effect of Val381 and ascertain its role in human adaptation to environmental toxins.

**KEYWORDS:** aryl hydrocarbon receptor (AHR), evolution, genetics, molecular docking analysis, *Homo sapiens neanderthalensis*

## **SAŽETAK**

Receptor arilnih ugljikovodika (AHR), regulator ksenobiotičkog metabolizma i detoksikacije, pojavio se kod ranih metazoa i ima duboku evolucijsku povijest. Kod živućih ljudskih populacija (*Homo sapiens sapiens*) jedinstvena supstitucija Val381 karakterizira AHR, za razliku od izumrlih hominina poput neandertalaca (*Homo sapiens neanderthalensis*), koji posjeduju ancestralni ostatak Ala381. Ova se supstitucija također smatra evolucijskom adaptacijom na povećanu izloženost policikličkim aromatskim ugljikovodicima (PAH) zbog kulturnih dostignuća poput kontroliranog paljenja vatre, čija se uporaba intenzivirala tijekom ljudske evolucije. Na taj način Val381 možda regulira aktivnost AHR-a kod ljudi u usporedbi s drugim životinjama, vjerojatno smanjujući aktivaciju AHR-a i posljedične toksične odgovore. Postojeći podaci su kontradiktorni: dvije in vitro studije razlikuju se u pogledu toga posjeduje li ljudski AHR nižu afinitet vezanja liganada od neandertalskog AHR-a. Ova studija rješava tu nedosljednost kroz in silico metodu molekularnog dokiranja (eng. molecular docking) 42.995 liganda - od modulatora AHR-a, mikrobnih metabolita, bioaktivnih molekula do prirodnih spojeva - kako bi se istražili afinitet vezanja i načini interakcije kod dviju vrsta. Rezultati imaju za cilj razjasniti funkcionalni učinak Val381 te utvrditi njegovu ulogu u ljudskoj adaptaciji na okolišne toksine.

**KLJUČNE RIJEČI:** receptor arilnih ugljikovodika (AHR), evolucija, genetika, molekularni doking (eng. molecular docking), *Homo sapiens neanderthalensis*

## TABLE OF CONTENT

|   |    |
|---|----|
| <b>INTRODUCTION</b>   | 1  |
| 1.1. Homo sapiens neanderthalensis and Denisova hominins    | 1  |
| 1.1.1. Neanderthals (Homo sapiens neanderthalensis)         | 1  |
| 1.1.2. Denisovans (Homo denisova)                           | 8  |
| 1.1.3. Shared ancestry                                      | 9  |
| 1.2. Aryl hydrocarbon receptor                              | 12 |
| 1.2.1. The discovery of Aryl Hydrocarbon receptor           | 12 |
| 1.2.2. The AHR structure and the bHLH/PAS Family            | 12 |
| 1.2.3. Biochemical pathways involving AHR                   | 14 |
| 1.2.4. Evolution of Aryl hydrocarbon receptor               | 18 |
| 1.3. Molecular modeling and molecular docking               | 20 |
| 1.3.1 Molecular modeling                                    | 20 |
| 1.3.2. Molecular docking                                    | 21 |
| <b>2. AIM</b>   | 23 |
| <b>3. MATERIALS AND METHODS</b>                             | 24 |
| 3.1. Materials  | 24 |
| 3.1.1. Comparative analysis of the AHR-coding gene          | 24 |
| 3.1.2. Molecular docking structures                         | 25 |
| 3.2. Methods  | 26 |
| 3.2.1. Comparative analysis of the AHR-coding gene          | 26 |
| 3.2.2. Modeling Neanderthal AHR                             | 26 |
| 3.2.3. Protein and ligand preparation for molecular docking | 26 |
| 3.2.4. Machine learning model: random forest                | 29 |
| <b>4. RESULTS</b>   | 30 |

|           |  |    |
|-----------|--|----|
| 4.1.      | Comparative analysis of the AHR-coding gene .....            | 30 |
| 4.2.      | Molecular docking .....                                      | 32 |
| 4.2.1.    | Summary of docking data .....                                | 32 |
| 4.2.2.    | Correlations between molecular descriptors and docking score | 40 |
| 4.2.3.    | Random forest .....  | 41 |
| <b>5.</b> | <b>DISCUSSION</b> .....                                      | 44 |
| 5.1.      | Comparative analysis of the AHR-coding gene.....             | 44 |
| 5.2.      | Variations in the untranslated regions.....                  | 45 |
| 5.3.      | Evolution of AHR.....  | 45 |
| 5.4.      | Molecular docking .....                                      | 46 |
| 5.5.      | Choosing molecular docking software .....                    | 48 |
| 5.6.      | Limitations .....  | 49 |
| <b>6.</b> | <b>CONCLUSION</b> .....                                      | 50 |
| <b>7.</b> | <b>REFERENCES</b> .....                                      | 51 |
| <b>8.</b> | <b>CURRICULUM VITAE</b> .....                                | 57 |

## INTRODUCTION

### 1.1. *Homo sapiens neanderthalensis* and Denisova hominins

#### 1.1.1. Neanderthals (*Homo sapiens neanderthalensis*)

*Homo sapiens neanderthalensis* is an ancient population of Western Eurasia (Figure 1) that lived approximately 300,000 and 40,000 years ago. Within the genus *Homo*, they represent a sister lineage of *Homo sapiens sapiens*, which they share with a common ancestor, probably *Homo heidelbergensis*. (1) So far, no clear speciation event has been established; instead, it appears that Neanderthal characteristics built up gradually in the second half of the Middle Pleistocene, and the earliest phenotypic indicators date back as early as 600,000 years ago, while the characteristic Neanderthal morphology is confined to subsequent populations. (2)

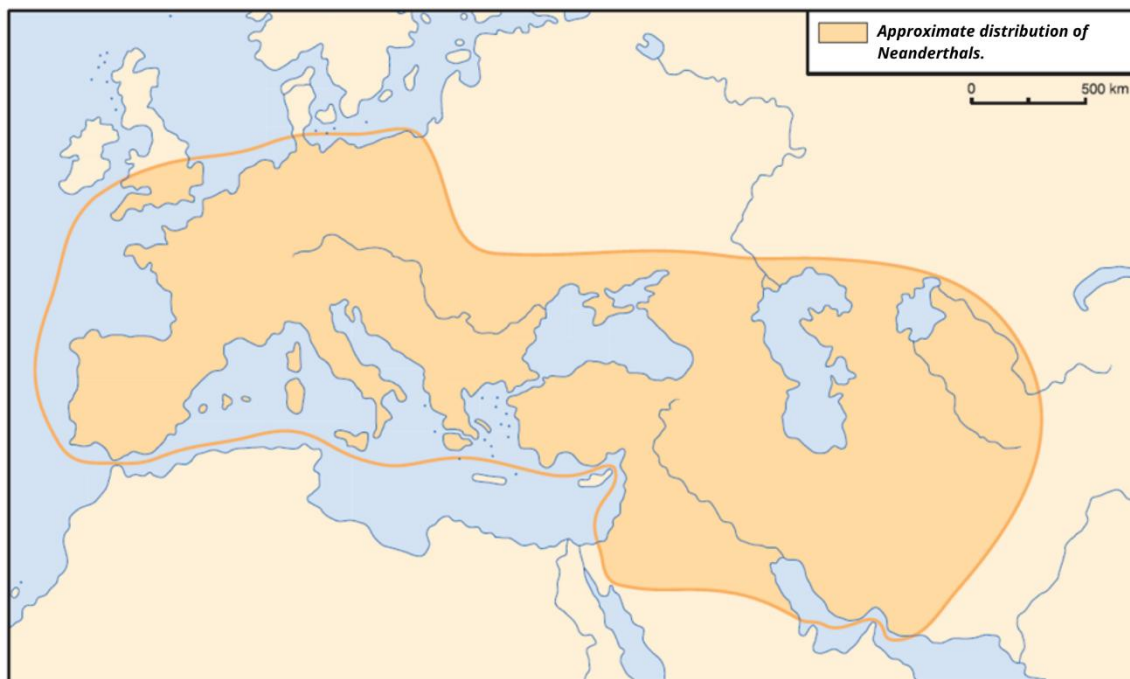


Figure 1 Geographic distribution of Neanderthals (Image taken from: Janković and Karavanić, 2009 (1))

Their taxonomic status and phylogenetic relationship to anatomically modern humans has been a subject of extensive debate over the years. A few authors, including Hrdlička, Gorjanović-Kramberger, and Schwalbe, viewed *Homo sapiens neanderthalensis* as an evolutionary stage towards modern humans.(1) They were thought by Virchow to belong to the species *Homo sapiens sapiens* and, from a morphological perspective, viewed their morphology as a consequence of illness.(1) Boule and Vallois held the view of total segregation of Neanderthals and contemporary Europeans.(1) Up-to-date morphometric and genetic analysis clearly indicates that Neanderthals and modern humans show some differences, reflected in details of craniofacial, dental, and postcranial anatomy, and ontogeny. Analysis of three-dimensional morphometric data has established that the morphological divergence among Neanderthals and modern humans is greater than among chimpanzees and bonobos—two taxonomically distinct species that never interbreed in the nature. Neanderthals also exhibit lower intra-population morphological variability, whereas modern humans already possessed high diversity in the Pleistocene. This suggests that there were different evolutionary processes that influenced nature of Neanderthals and modern humans like natural selection and genetic drift. (2)

Genetic data also attest to the deep divergence between these two lineages of genus *Homo*. More than 15 *Homo sapiens neanderthalensis* mitochondrial DNA exhibits divergence between *H. sapiens* and *H. neanderthalensis* lineages roughly 660,000 years ago, though autosomal genome evidence favors a population split roughly 370,000 years ago. (2)

### **Extinction Hypotheses**

The disappearance of Neanderthals from the European fossil record between 41,000 and 39,000 years ago was gradual and regionally differentiated, with a possible late survival of populations in certain areas, such as Iberian Peninsula. (3) What is certain is that their retreat substantially coincides with the expansion of anatomically modern humans into Europe, with the

implication of harsh biological and cultural interactions between the two groups during the final millennia of Neanderthal survival.

A number of theories have been suggested to explain the causes of this extinction. Among the more popular hypotheses is the Repeated Replacement Model (RRM), or "repeated replacement" hypothesis (4), which maintains that rapid and severe climatic oscillations in the Late Pleistocene led to repeated local extinctions and demographic replacements of human populations. The theory emphasizes that climate volatility was not an immediate cause of Neanderthal extinction, but rather created unfavorable demographic conditions in combination with the arrival of competing *Homo sapiens* populations.(4) In climatically favorable conditions, Neanderthal populations recovered, but, during periods of ecological crisis, they were increasingly supplanted. Other than climatic stress, a body of cumulative evidence identifies the determining factor of competition with anatomically modern humans. *Homo sapiens* was a very successful predator with the capacity to expand rapidly and more efficiently exploit resources—particularly impacting the viability of other large mammals, Neanderthals included. (5) In this instance, Neanderthals were less victims of outright violence than they were competed out for the same ecological niches over time. More recent findings suggest that the replacement of Neanderthals by modern humans during the GI 10 and GS 10 stages (Greenland Interstadial 10 – a warm phase; Greenland Stadial 10 – a cold phase) was not so much climate-induced as induced by the ability of *Homo sapiens sapiens* to more effectively exploit winter resources during glacial climates. (6)

Neanderthals did not "disappear" in the biological sense of the term, but were absorbed to a certain degree into modern human populations. The assimilation model (Figure 2) is based on morphological and genetic evidence bearing witness to intermixture between the two populations and the existence of individuals of hybrid features. (7) Although most of the

distinct Neanderthal anatomical features disappeared soon after the arrival of anatomically modern humans in Europe, genetic traces, but also some morphological traits of Neanderthals may have persisted in European populations for several millennia. In this case, the extinction of the Neanderthals is not viewed as complete biological extinction but instead as the result of the demographic and genetic absorption of a minority population by the much larger *Homo sapiens* population. A further factor is the demographic vulnerability of Neanderthals. Interglacial phases, especially the Eemian, through their environmental conditions (dense forests, extended watercourses, and reduced hunting grounds), generated population fragmentation, breakdown of social networks, and genetic bottlenecks, further reducing their adaptive potential. (6) In this context, Neanderthals began the contact period with modern humans already weakened in population structure. The demise of Neanderthals was the result of the interaction of interrelated factors—climatic change, competition with modern humans, and partial assimilation. The evolution of the genus *Homo* in the Late Pleistocene was not linear or unilateral but dynamic, with constant replacements, interactions, and adaptations.

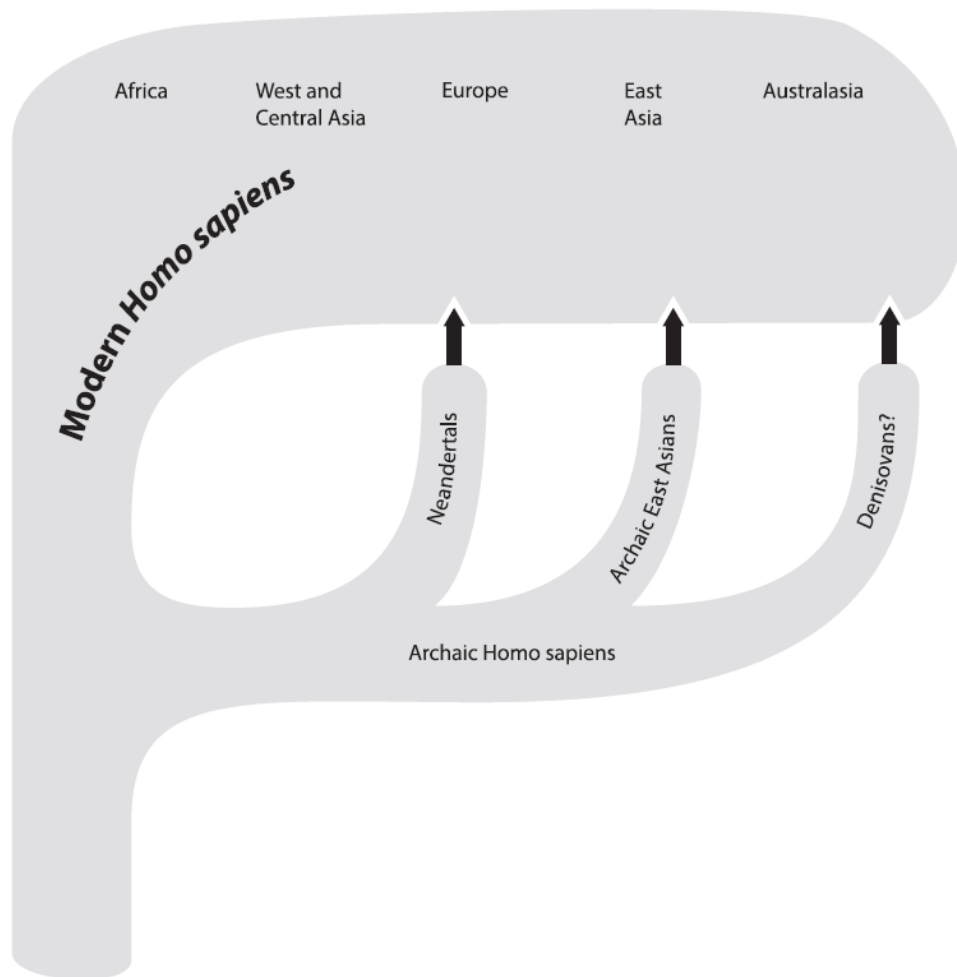


Figure 2 The assimilation model (Image taken from: Smith et. al, 2005. (7))

## Fire use

Archaeological data from Middle Paleolithic provides conclusive proof of Neanderthals' use of fire. Most sites, such as the Levant (Tabun, Kebara, Hayonim) and European sites (St. Césaire, La Quina, Pech de l'Azé IV, Roc de Marsal, Abric Romani), show the presence of hearths, charcoal, ash, and thermally altered artifacts, all of which indicate the frequent application of controlled fire. (6) (8) (Figure 3) In south-west Germany, such as Große Grotte and Bockstein, hearth patches and concentrations of burned bones occur, denoting controlled fire use in daily subsistence—such as heating and



practiced. This practice is known since at least 200,000 years ago. (6) (8) Apart from its utilitarian uses, fire may have had social and spatial functions in Neanderthals. All but a few sites have multi-hearth architecture. Few hearths suggest the concept of a "central fire" around which individuals gathered—patterns observed among current hunter-gatherer populations. One instance of possibly more sensational complex use of fire is that of Bruniquel Cave, where six circular enclosures were constructed deep in the cave interior by Neanderthals, and all of these contain evidence of fire. (6) It is very likely that Neanderthals had the ability to produce fire on demand. Some researchers postulate that, rather than initiating fires from scratch, they preferred to maintain ongoing fires and transported embers—possibly explaining why there is little affirmative archaeological evidence for fire-initiation technology. Others argue that fire use may have been more a consequence of natural availability of fire sources, such as lightning and forest fires, rather than controlled burn technologies. (6)

Evidence from Neanderthal diets suggests a high degree of dietary flexibility in employing varying diets as a response to climatic and environmental fluctuations during glacial and interglacial times. Analysis of dental microwear suggests continuity among patterns of tooth use that reflect climate-determined shifts in diet. (6) When Neanderthals survived during the colder phases of the Late Pleistocene, they had supplemented their meat diets with additional abrasive plant foods such as roots and tubers. These dietary changes were strongly determined by the type of habitat that Neanderthals occupied. In forest habitats, Neanderthals could access a variety of resources and thus had better-balanced diets with higher proportions of plant food. In mosaic vegetation, they consumed both animal and plant food. In cold steppe environments, their diet comprised almost all meat-based food, similar to the abundance of large herbivores. An interesting point is that Mediterranean landscape-dwelling Neanderthals also had marine shellfish as part of their diet. Apart from geography, age

structure further influenced diet: children and the elderly consumed softer, lower-fiber food than was consumed by prime-aged adults. (6)

### **1.1.2. Denisovans (*Homo denisova*)**

Denisovans were found in DNA samples from Denisova Cave, Altai Mountains, where fragmentary remains—primarily teeth and a phalanx—were found. Preservation of DNA in the Denisova 3 sample was unusually high, with over 70% of the DNA found to be endogenous, i.e., belonging to the person himself. This allowed for extensive genome sequencing. (9)

The nuclear genome of the Denisovans shows that they are a sister group of Neanderthals. Unlike mitochondrial DNA, which exhibits increased Neanderthal and Denisovan divergence, the nuclear genomes reveal an identical level of divergence, indicating that these two lineages diverged around 640,000 years ago. Estimates based on comparisons with differences between the Denisovan and modern human genomes versus chimpanzees suggest that Denisovans diverged from the common ancestors of modern humans around 800,000 years ago. Mitochondrial DNA is more Neanderthal-Denisovan divergent than nuclear DNA. Two main hypotheses have been put forward to explain this. The first is that there is incomplete lineage sorting, whereby genetic markers are randomly inherited and may display older divergences in some parts of the genome. The second hypothesis is that Denisovans inherited their genetic contribution from an archaic group of hominins—a supposed super-archaic group—whose branch arose from the lineage that leads to modern humans earlier than Neanderthals and Denisovans diverged. (6) (10)

Although genetic evidence for the presence of the Denisovans is definite, their fossil record is extremely fragmentary. They have only left behind minuscule skeletal remains and, aside from the data obtained from their DNA, nothing much can be deduced about their form. Outside of the Altai region, the only fossil more securely attributed to the Denisovans is the

Xiahe mandible, discovered in Tibet, that confirmed that Denisovan hominins were geographically dispersed in Asia. (11)

### **1.1.3. Shared ancestry**

Genetic and fossil evidence show that Neanderthals, Denisovans, and early anatomically modern humans genetically intermingled with one another, and that such interactions happened in different temporal and spatial contexts. 1% to 4% Neanderthal DNA is present in the genomes of living Eurasians. (12) Take into consideration that significant gene pool shifts between modern humans occurred during and after the Pleistocene, so the initial proportion of Neanderthal genetic heritage must have been higher than can be determined today. Morphological data also suggest possible admixture of populations: some later European Neanderthals possess features closer to early modern humans, and the earliest anatomically modern Europeans possess features that suggest some degree of Neanderthal heritage. (7) The Europe-based chronology for a possible overlap of Neanderthals and modern humans spans many thousand years—possibly as long as 7000 years, but regionally variable. (13) This overlapping period is in accordance with both genetic and archaeological accounts of intergroup contact, for example, interbreeding episodes that left a lasting influence on the genetic composition of modern humans. (14)

Genomic analyses also show significant Denisovan influx into modern human gene pool, and indicate that they passed on 4-6% of genetic material to modern human beings in Oceania and parts of Southeast Asia. A striking example of adaptive introgression is the presence of a specific Denisovan haplotype in the EPAS1 gene that facilitates high-altitude adaptation among Tibetans. This haplotype is completely lacking in other populations, providing evidence for adaptively beneficial and localized gene flow. (6)

The Denisova 11 individual is a unique find that reveals the intermixing between Neanderthals and Denisovans. Collagen peptide analysis identified the individual that had a Neanderthal mother and a Denisovan father, clearly showing that these populations lived together and intermingled in the Altai area. (15) The Oase 1 individual's genome from Romania (16), instead, illustrates the interbreeding between early modern humans and Neanderthals. Oase 1 shared a Neanderthal ancestor just four or six generations prior, indicating that the admixture took place less than 200 years prior to the time at which the individual lived. Yet since Oase 1 is not obviously associated with later European populations, it is thought to be an early modern human population that had interbred with Neanderthals but contributed no significant genetic inheritance to later populations. (16)

It is important to note that gene exchanges between groups cannot be in absolute terms equal to percentages of gene input. They need to be described within the overall framework of population processes, migration, reproductive isolation, and natural selection. For example, the Denisovans' and Neanderthals' mitochondrial DNA is highly divergent from the recent human line but with their nuclear DNA having intervals of admixture clearly indicated. These trends reveal the existence of gene flow between these three species. (16) (Figure 4)

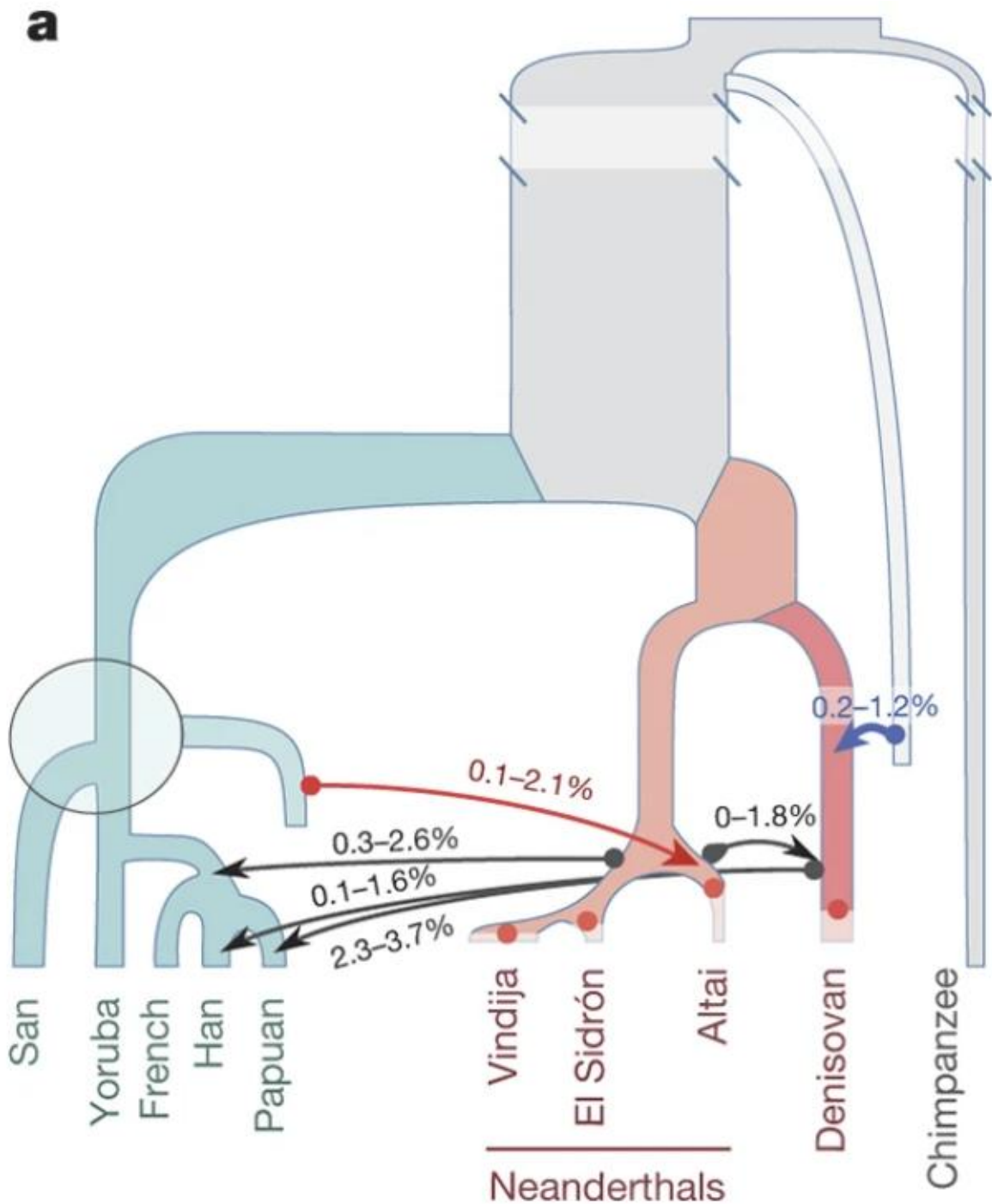


Figure 4 A model of genealogical relationships between modern humans and archaic populations (Image taken from: Kuhlwilm et al., 2016 (10))

## **1.2. Aryl hydrocarbon receptor**

### **1.2.1. The discovery of Aryl Hydrocarbon receptor**

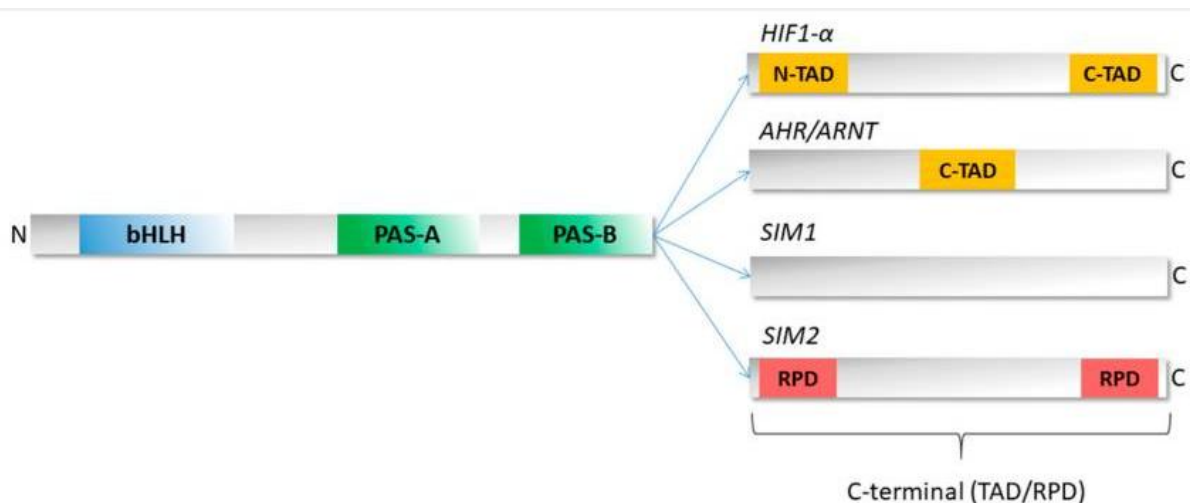
The aryl hydrocarbon receptor (AHR) was first hypothesized in the 1970s through studies investigating xenobiotic-mediated enzyme activity in rodent liver. Early observations illustrated strain-specific differences in the induction of cytochrome P450 (CYP) enzyme by polycyclic aromatic hydrocarbons (PAHs). With additional testing, it was discovered that 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) was approximately 36,000-fold more potent than PAHs, providing a compelling rationale for the existence of a high-affinity receptor. Comparative dose-response studies in C57BL/6 and DBA/2 mice provided early evidence of a receptor mediating these effects; - an idea that initially challenged the prevailing view that intracellular receptors could not bind exogenous chemicals. (17)

In 1976, high-affinity binding of TCDD to a cytosol factor in mouse liver provided direct chemical support for the existence of such a receptor. Cloning of the Ahr gene in 1992, - and the creation of Ahr deficient mouse models in the mid-1990s revealed essential roles for AHR not only xenobiotic metabolism, but also vascular development, immune regulation, and cellular homeostasis. These findings established AHR as a multifunctional transcription factor responsive to both environmental and endogenous signals. (17)

### **1.2.2. The AHR structure and the bHLH/PAS Family**

The aryl hydrocarbon receptor (AHR) belongs to the Basic-helix-loop-helix Per-Arnt-Sim (bHLH/PAS) family, which is a highly conserved class of transcription factors present in all kingdoms of life. Members of this family function as intracellular sensors of diverse physicochemical stimuli, including xenobiotics, gases, redox potential, temperature, light, and osmotic pressure. Signal perception is typically via the PAS domains, while DNA binding and dimerization are facilitated via the bHLH domain. (17)

bHLH–PAS proteins contain conserved bHLH, PAS-A and PAS-B domains at the N-terminal, which allow DNA recognition, dimerization and signal detection. Signal perception is typically via the PAS domains, while DNA binding and dimerization are facilitated via the bHLH domain. The C-terminal part is variable and contains a transcriptional activation (TAD) or repression (RPD) domain, depending on the protein. (18) (Figure 5)

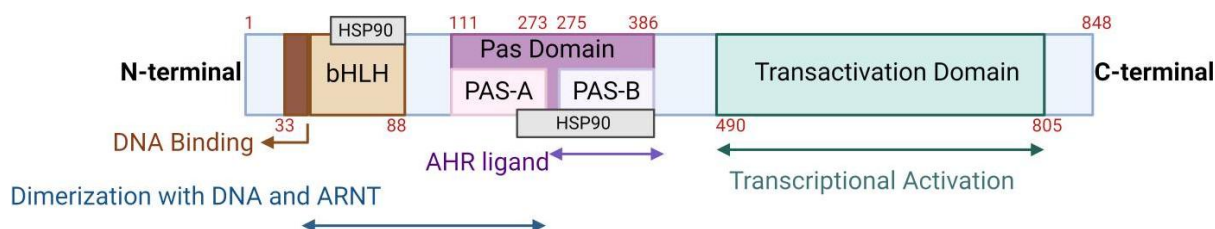


*Figure 4 Figure shows the organization of functional domains between the proteins in bHLH/PAS Family, the conserved motifs towards the amino-terminal domain, and the variable regulatory sequences towards the the carboxyl terminus highlighted. (Image taken from Kolonko and Greb-Markiewicz, 2019 (18))*

bHLH/PAS proteins are generally divided into two functional classes: class I proteins, whose expression is environmentally or developmentally regulated, and class II proteins, which are constitutively expressed and function as dimerization partners. Transcriptional activity requires heterodimerization, usually between one class I and one class II member. (18)

AHR is a representative class I bHLH/PAS protein. It is cytoplasmic in its inactive form. Ligand binding results in the relocation of AHR to the nucleus, dimerization with ARNT (a class II partner), and regulation of target gene expression by xenobiotic response elements (XRE). AHR specifically

contains a PAS-B domain for ligand binding and a TAD domain that enables the regulation of genes involved in detoxification and response to xenobiotics. AHR thus bridges environmental signal to transcriptional response, in this case, to detoxification and cellular stress. (18) (19) (Figure 6)

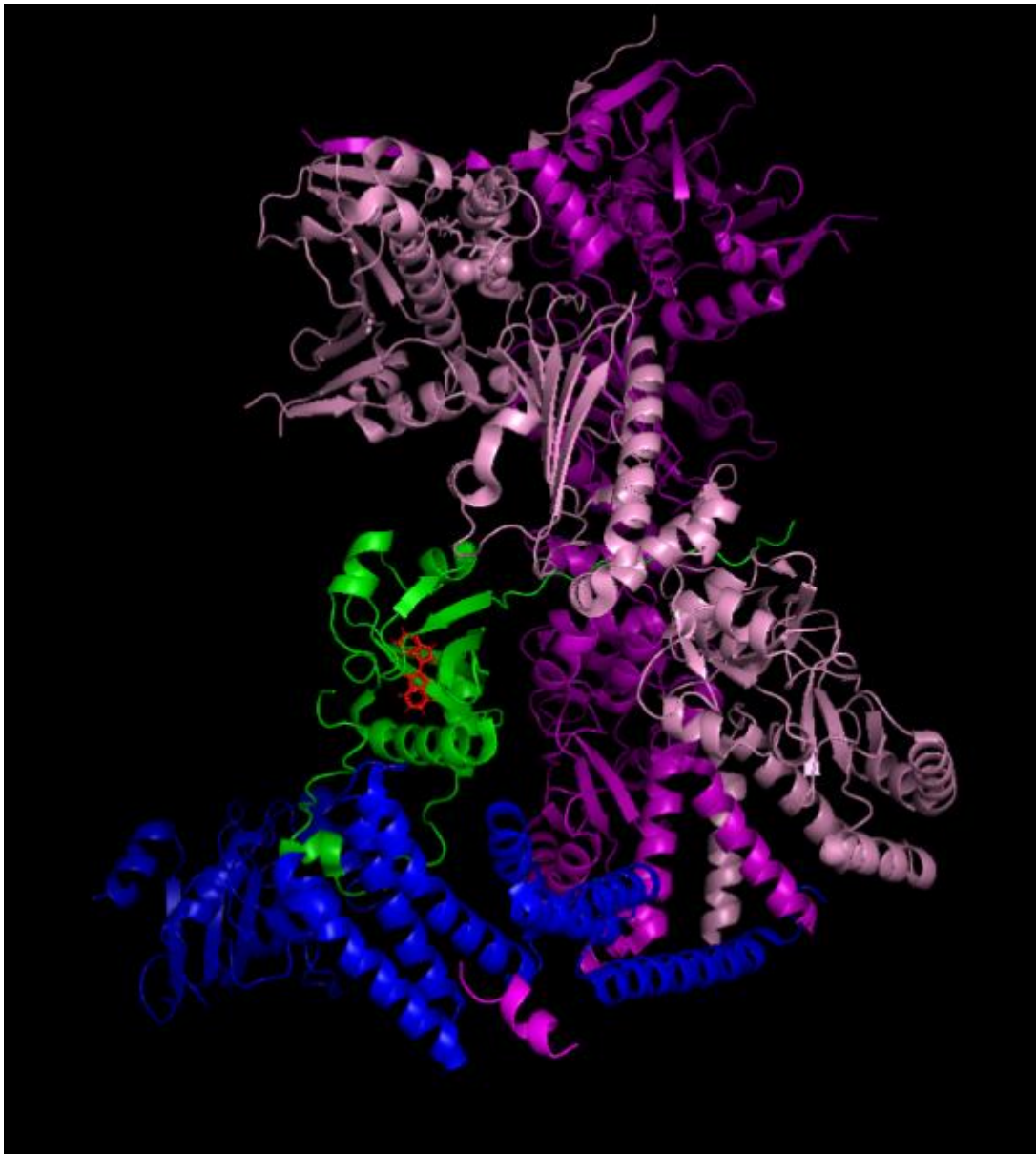


*Figure 5 The scheme represents the domain structures of the AHR protein. (Image taken from: Bahman et al., 2024 (19))*

Structural data on full-length bHLH/PAS proteins remain limited. Available crystal and NMR structures predominantly include isolated PAS domains, while C-terminal transactivation/repression domains contain intrinsically disordered regions (IDRs), representing a current limit to the understanding of full protein function. (18)

### 1.2.3. Biochemical pathways involving AHR

Among the diverse biological roles attributed to the aryl hydrocarbon receptor (AHR), its function in xenobiotics sensing and detoxification remains most often emphasized. (17) Xenobiotics are exogenous compounds that are not part of normal metabolism, and may include naturally occurring secondary metabolites, combustion products, synthetic compounds, etc. Besides its well-characterized response to polycyclic aromatic hydrocarbons (PAHs)(20), AHR is also known to react to halogenated aromatic compounds. (21)



*Figure 6 AHR in the cytoplasm. Different shades of pink show the subunits of the HSP90 dimer, blue is AIP, green is AHR, and red is the ligand indirubin in the AHR binding site. Image created using PyMol software.*

Upon ligand binding, AHR performs a structural change, dissociates from the cytoplasmic AIP-cochaperone complex (Figure 7), associates with importin proteins, and imports into the nucleus. (22) There, it heterodimerizes with ARNT (HIF-1 $\beta$ ), and the complex binds to xenobiotic response elements (XREs) in the promoters of the target gene to initiate transcription. (Figure 8)

It is inhibited negatively by the repressor AHR (AhRR), which acts by competing with ligand-bound AHR in binding ARNT. AhRR, which has a similar structure to AHR but lacks the PAS-B domain, cannot bind ligands.

(19)

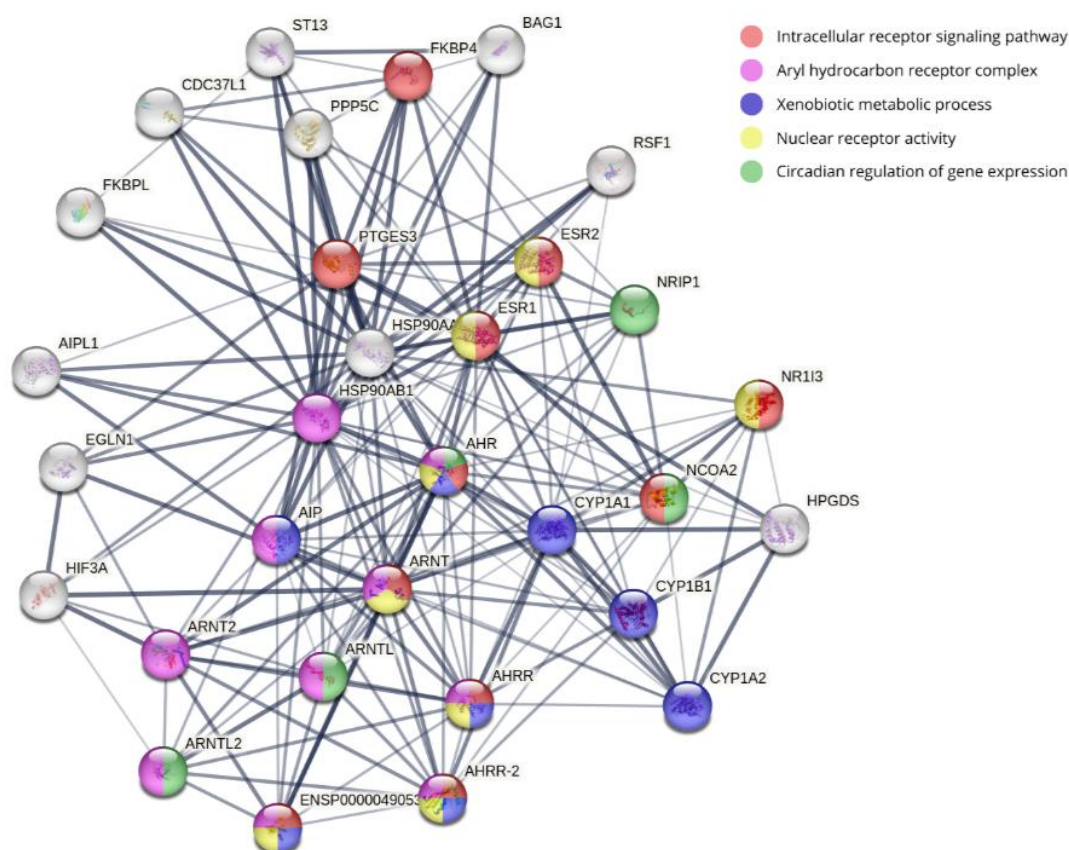


*Figure 7 AHR in the nucleus bound to DNA. AHR is presented in green, and ARNT in red. Image created using PyMol software.*

Primary AHR targets include genes that code for xenobiotic-metabolizing enzymes (XREs) such as CYP1A1, CYP1A2, and CYP1B1, which catalyze the lipophilic compound biotransformation into more soluble, excretable

metabolites. (22) CDKN1A (p21) and CDKN1B (p27), which are cyclin-dependent kinase inhibitors that cause G1/G0 cell cycle arrest, are also AHR-induced, and transcription is suppressed of E2F-dependent cell cycle progression genes. (23) AHR interacts with estrogen receptors ESR1 and ESR2, modulating their transcriptional activity, with development- and hormone-dependent tumorigenesis consequences. (24)

To better explain AHR signaling, a protein-protein interaction (PPI) network was constructed. (Figure 9) It is composed of direct interactors (AHR, ARNT, HSP90AA1), xenobiotic-metabolizing enzymes (CYP1A1, CYP1B1), co-regulators (NCOA2, NRIP1), and nuclear receptors (ESR1, ESR2). In addition, computational predictions based on literature, databases, and experiments also disclosed other proteins with potential implication in AHR signaling. These *in silico* predictions augment and refine the existing known AHR interactome.



*Figure 8 Network represents AHR and its molecular interaction landscape. Multicolored nodes represent proteins most closely functionally associated with AHR. White nodes indicate proteins involved in secondary or indirect interactions. Thicker edges (connections) imply a higher confidence level of interaction. This visualization was generated using STRING and Cytoscape software.*

AHR is a highly active and multifunctional protein involved in numerous biological pathways. However, many of these roles fall outside the scope of this chapter and are therefore not discussed in detail here.

#### **1.2.4. Evolution of Aryl hydrocarbon receptor**

The evolution of AHR proteins in metazoans is a complex process that took millions of years. This protein is highly conserved throughout evolution. (25) To understand AHR evolution, it is necessary to compare its presence and adaptation across environmental niches and taxonomic groups.

The evolutionary history of AHR begins with the earliest metazoans, where homologues are present in the genomes of deuterostome animals, such as

echinoderms (e.g., sea urchins), hemichordates (e.g., acorn worms), and cephalochordates (e.g., amphioxus). (26) In these species, AHR homologues do not bind dioxins, suggesting ancestral AHR had functions unrelated to xenobiotic sensing. The AHR of invertebrates like *Ciona* also lacks the ability to bind TCDD, indicating a functional divergence from modern AHR proteins. This suggests that ancestral AHR may have participated in processes such as hypoxia response or sensory neuron development. (26)

Diversification in vertebrates is marked by the first AHR homologues capable of binding TCDD, representing a key adaptation to environmental toxins. During vertebrate evolution, AHR adapted to regulate xenobiotic-metabolizing genes, including cytochrome P450 enzymes (e.g., CYP1 family). (26) Over 450 million years ago, genome duplications in early vertebrates allowed the expansion of AHR genes, enabling functional diversification. Jawless vertebrates (agnathans) like lampreys possess up to five AHR genes, reflecting early gene expansion. This redundancy likely facilitated specialization to diverse environmental conditions. (26)

In modern vertebrates, AHR has diversified across phyla. Teleosts, the most species-rich group of fish, carry tandemly duplicated AHR1 and AHR2 genes. For example, zebrafish possess an AHR1b–AHR2b pair, while basal species like gar, which diverged prior to the teleost-specific genome duplication, retain a single AHR gene. In vertebrates, the AHR repressor (AHRR) evolved as a regulator of AHR signaling. AHRR inhibits AHR/ARNT activity, maintaining homeostasis and preventing cytotoxic overactivation. (26)

In humans, a unique evolutionary feature is the fixation of Valine at position 381 in the AHR gene. (27) This substitution differs from all other known animals, including archaic hominins like Neanderthals and Denisovans, which retain Alanine at this site. (27) It remains unclear whether Val381 is a neutral mutation or an adaptation to environmental toxins, such as fire

smoke exposure. Some studies report functional consequences of this mutation (28), while others suggest minimal impact. (29)

The evolution of AHR reflects a complex interplay of genetic, physiological, and environmental forces over deep evolutionary time. Information about the evolution of AHR may advance our understanding of the mechanisms by which evolutionary pressure acts on biochemical pathways and how proteins like AHR can respond to environmental pressures over extended periods of time. (26)

### **1.3. Molecular modeling and molecular docking**

#### **1.3.1 Molecular modeling**

Molecular modeling applies calculations grounded in physical principles to analyze molecular structures, properties and interactions. It enables characterization of atomic composition, bonding, geometry, energy landscapes, vibrational frequencies, and electronic properties such as nucleophilicity, electrophilicity, and electrostatic potential. Parameters like this can then give further insight into biological processes elucidating things like structure–activity relationships. (30) Spectroscopy is key to supporting molecular modeling methods, and the outcome has led to deep new physical, chemical, and biological understanding.

Quantum mechanical coupling with spectroscopy has more accurately educated us on atomic and molecular structure, forming field of computational chemistry. Molecular modeling and computational chemistry are bridging the gap between theory and experiment in chemistry using quantum mechanics and mathematics to describe complex chemical and biochemical processes. Experimental spectroscopy provides direct information, whereas computational methods provide prediction and analysis of chemical properties and reactions like unstable intermediates and transition states, which cannot be analyzed experimentally. Molecular modeling is becoming more and more crucial for visualization, analysis, and

prediction of molecular structure and interactions, for use in docking studies, energy calculations, and receptor mapping. (31)

Due to the progress in spectroscopic techniques and computer technology, molecular modeling is becoming increasingly a routine technique in different fields of research. It is used in the design of sensors, new materials, astrophysics, and drug design. (32) (33)

### **1.3.2. Molecular docking**

Molecular docking is a type of molecular modeling, which is a computationally driven process that simulates the interaction of target proteins and small molecules. Given a solved receptor structure, it can be used to predict ligand poses and estimate binding affinity. Advances in X-ray crystallography and NMR spectroscopy have contributed to, thousands of experimentally determined 3D protein structures which are available at the Protein Data Bank (PDB). (34) (35)

The docking process uses a search algorithm to evaluate ligand conformations, identifying minimal energy states. The ligand poses are ranked according to a scoring function of affinity, based on electrostatic and van der Waals interactions within protein-ligand complex. (36)

Docking is used for three main purposes: virtual screening, pose prediction, and binding affinity evaluation. Virtual screening identifies potential ligands without precise estimation of pose or affinity and is useful in the search for large compound databases. Pose prediction mimics the manner a ligand binds to a protein, and its accuracy is validated by comparison with experimental co-crystal structures. Binding affinity calculation quantifies the binding of the protein and ligand based on thermodynamic criteria, but precise prediction of binding free energy remains a challenge. (35)

One of the major drawbacks of docking is receptor rigidity. Proteins are flexible molecules, and their flexibility can affect ligand binding. Most docking algorithms are not considering receptor as flexible, which can lead

to incorrect predictions of interactions. Some programs have limited flexibility, but these improvements are often insufficient to replicate natural behavior of protein-ligand complex. (35)

When computing ability improved, the flexible-ligand docking became standard, although rigid proteins remained common in practice due to the difficulty in simulating protein flexibility. Ligands should ideally be totally flexible while docking to better reflect actual binding. Ligand as well as receptor alter their conformations to attain a minimum energy and form a best-fit complex. Small conformational changes of ligand can produce significant differences in docking results. (37) (38)

## 2. AIM

The aim of this study is to investigate structural and functional differences in the Aryl Hydrocarbon Receptor (AHR) in *Homo sapiens neanderthalensis* and *Homo sapiens sapiens*, focusing on comparison of their abilities in ligand binding. In the active site of the AHR protein, alanine is found in *Homo sapiens neanderthalensis* and valine in *Homo sapiens sapiens*. The amino acid substitution may have the potential to affect the functionality of the AHR receptor. It is probable that it can alter its ligand binding as well as modulation of the downstream biological processes.

Because the AHR receptor is involved in detoxification, functional variation may be significant in the context of human use of fire. Exposure to combustion byproducts may have placed selective pressure on AHR function, suggesting an evolutionary adaptation.

Three main elements of this research:

**Comparative genomic analysis** – Identification of key genetic differences in the AHR-coding gene from publicly available Neanderthal genome data.

**Structural modeling** – Reconstructing the Neanderthal AHR protein in 3D to assess structural differences from the modern human variant.

**Computational ligand-binding analysis** – Using molecular docking and molecular dynamics to test tens of thousands of compounds binding to both receptor variants under the same conditions.

### **3. MATERIALS AND METHODS**

#### **3.1. Materials**

##### **3.1.1. Comparative analysis of the AHR-coding gene**

In order to perform a comparative analysis of the AHR-coding gene in *Homo sapiens sapiens* and *Homo sapiens neanderthaliensis*, we used public genomic databases. Three Neanderthal and one Denisova hominin genomes were compared, found at the sites of Vindija, Altai, and Chagyrskaya. (Figure 10)

##### **Altai Neanderthal**

Denisova Cave in Russia, in the Altai Mountains, is one of the significant sites of archaic hominins. A phalanx of the toe was found at this site in 2010, with an absolute date of  $50,300 \pm 2,200$  years. This sample was used for sequencing the genome of the Altai Neanderthal, a female *Homo sapiens neanderthaliensis*. The genome was sequenced to  $\sim 50$ -fold coverage. (39)

##### **Altai Denisovan**

In the same cave, in 2008, a phalanx of a finger was discovered. It is found in the same layer as toe phalanx, so it is also dated to  $50,300 \pm 2,200$  years. This individual provided the first genome sequence of the previously unknown Denisovan hominin. The genome was sequenced to  $\sim 30$ -fold coverage. (39)

##### **Chagyrskaya Neanderthal**

Chagyrskaya Cave, a further important location in the Altai Mountains, revealed remains of a female Neanderthal in 2011. In 2020, the genome was recovered to  $\sim 27$ -fold coverage from a finger phalanx. The fossil is  $\sim 80,000$  years old. (40)

## Vindija Neanderthal

At Vindija cave in Croatia, remains of several Neanderthals were discovered. (41) Among these are the remains of a female Neanderthal, dated to approximately  $\sim 50,000$  years. (nova ref) DNA sequencing was performed on 12 samples that belonged to the same individual. The genome was sequenced to  $\sim 30$ -fold coverage. (42)

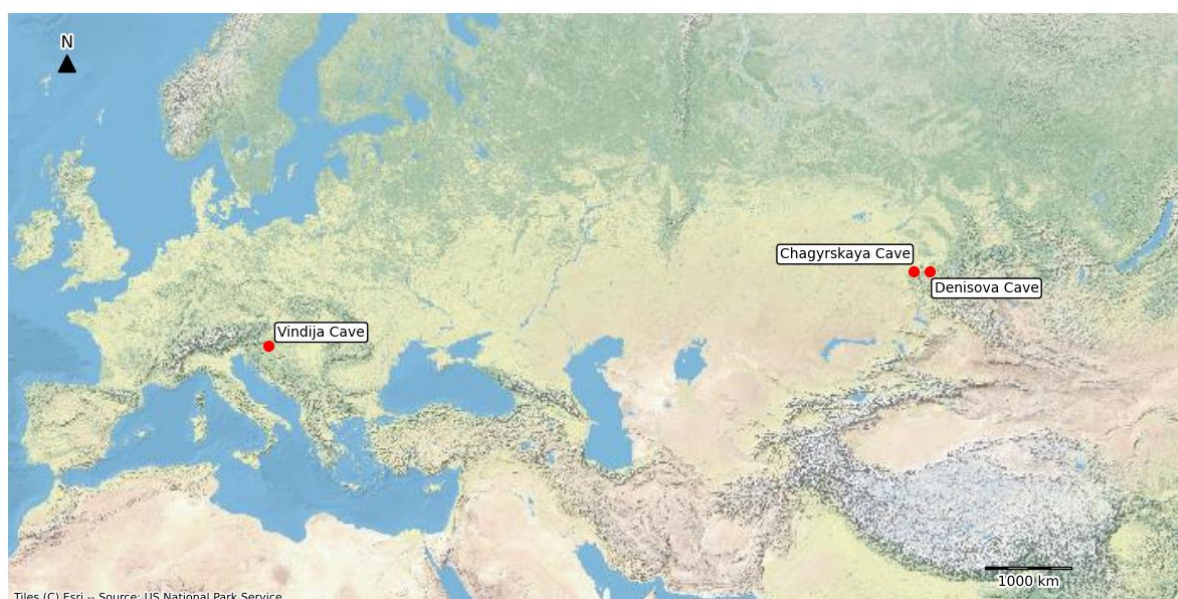


Figure 9 Archaeological sites with Neanderthal and Denisovan skeletal remains analyzed in this study. The map shows the locations of Vindija Cave (Croatia), Chagyrskaya Cave (Russia), and Denisova Cave (Russia). Image was created with Python.

### 3.1.2. Molecular docking structures

Docking analysis was performed on the AHR receptor structure obtained from the Protein Data Banks (PDB). (34) The structure is resolved by Cryo-EM NMR imaging and has the PDB ID 7ZUB. It is a complex of AHR receptor, chaperone Hsp90, and co-chaperone XAP2. The reference ligand used in the study was indirubin. A total of 44,995 ligands from diverse libraries, including polycyclic aromatic hydrocarbons (PAHs), microbial metabolites,

bioactive molecules, natural products, and known AHR modulators, were screened.

## **3.2. Methods**

### **3.2.1. Comparative analysis of the AHR-coding gene**

To compare the AHR gene between two hominin populations, Integrative Genomics Viewer (IGV) software was used. The Neanderthal and Denisovan genomic sequences were plotted against the reference human genome to identify variations in base pairs. Other Neanderthal and archaic *Homo sapiens* sequences were also plotted, but coverage for this locus was not found in those sequences.

### **3.2.2. Modeling Neanderthal AHR**

For modeling the Neanderthal version of the AHR protein, PyMOL software was used. The Neanderthal sequence was modeled by adjusting the amino acids based on the genomic sequence information identified above. Specifically, at position 381, human valine was replaced by Neanderthal alanine. Lysine replaced arginine at position 554. It is important to note that arginine at this position was found in the human reference genome, but PDB structure of AHR had different variation. Blast analysis is also performed, in order to find differences in AHR proteins among animal kingdom.

### **3.2.3. Protein and ligand preparation for molecular docking**

The human protein and the modeled Neanderthal protein were prepared and optimized for docking using Chimera software. A reference ligand was also generated using the Chimera program. Ligands were prepared in python with the RDkit package. The input data were Smiles records from the used compound libraries. With Smiles records, we generated a 3D structure, minimized the energy and filtered them based on the desired parameters.

### **3.2.4. Molecular docking programs**

For molecular docking analysis, we utilized RDock and RxDock software. RDock is an efficient, general-purpose, and open-source program that is designed for docking ligands to proteins and nucleic acids. RxDock, which is a derivative of RDock, was developed to revise the codebase such that it can easily work on modern computing equipment, including supercomputers.

#### **RxDock**

RxDock was designed particularly for high-throughput virtual screening. Ligands in RxDock are handled as flexible, while the receptor is handled as a rigid body.

One of the key features of RxDock is the cavity search engine, which helps identify binding pockets. In other words, it defines and maps areas on macromolecules to which ligands can bind. (43)

There are two methods for specifying the docking site in RxDock:

- a) Two-sphere method: detects cavities that are accessible to a small sphere but not to a large one, eliminating flat surface and shallow pockets as potential binding sites.

- b) Reference ligand method: defines a docking volume of a specific size around the binding site of a known ligand. (44)

#### **Reference ligand method**

In this study, the reference ligand method was used to determine the AHR binding site. The method is dependent on the use of a ligand whose position within the binding site is well-established and described and forms the basis upon which the LBD volume is computed before molecular docking. The technique is found most useful in high-throughput investigations since it enables standardization and reproducibility of binding cavity

measurements. We used the previously determined binding site for all ligands to standardize the results. Indirubin was the reference compound. Based on the position of indirubin atoms, a three-dimensional grid is established which covers the region of the binding site.

Spheres of a given radius are constructed around all the reference ligand atoms. Beyond these spheres, grid points are automatically excluded from consideration. Single-contact points with protein atoms were also excluded. The van der Waals radii of receptor atoms are, in the process, inflated by a given parameter to represent the true three-dimensional space occupied by the protein. The remaining grid points are sampled by small-radii probes, which check if the point is accessible or whether there is enough space for a ligand to be fit. Any points that clash with protein atoms and probe accessible points are retained to be subjected to further processing. The cavity is resolved based on the remaining accessible points. These sites fall under spatially adjacent regions, and within each region there is a potential binding site. (44)

#### **3.2.4. Molecular docking**

For each of the ligands, 50-iteration molecular docking runs were performed in an attempt to extensively sample potential binding conformations in the AHR ligand-binding domain. The ligand was made fully flexible in each iteration. The receptor flexibility was set at 3.0 Å to allow partial side-chain movement to help in the accommodation of ligand binding. The results of docking encompass a set of essential energy terms used within the framework of ligand binding affinity and quality of interaction criteria:

- Total docking score (SCORE): Overall estimated binding energy taking into account all interaction terms.
- Interaction energy terms (SCORE.INTER): Including van der Waals forces (SCORE.INTER.VDW), polar interactions (SCORE.INTER.POLAR), and repulsive forces (SCORE.INTER.REPUL).

- Intramolecular ligand strain (SCORE.INTRA): Penalties in energy because of ligand conformational strain upon binding.
- System energy terms (SCORE.SYSTEM): Accounting for dihedral strain and other intramolecular contributions.
- Normalized scores (SCORE.norm and SCORE.INTER.norm): Compare binding affinities among ligands on a normalized scale.

These scores provided a general indication of the stability of ligand binding and were used to rank the ligands based on predicted receptor binding pocket fit and affinity. (44) (45)

### **3.2.4. Machine learning model: random forest**

For developing Random forest model, we used Scikit-learn package in Python. The dataset, comprising molecules from various chemical libraries, was separated into two independent sets: one for model training and internal validation (90/10 split) and the other for external model performance testing only.

Molecular descriptors were computed from SMILES strings, including Morgan fingerprints (radius = 2, 2048 bits) and a wide range of molecular descriptors computed using RDKit. Descriptors showing a significant correlation with the target variable (SCORE) were retained from the total descriptor set. To remove redundancy, the highly correlated descriptors were filtered to avoid having multiple descriptors with similar information.

Hyperparameter optimization was performed using GridSearchCV and five-fold cross-validation. The hyperparameters optimized are the number of trees in the forest (n\_estimators), the maximal depth of the individual trees (max\_depth), the minimum number of samples to split an internal node (min\_samples\_split), the minimum number of samples at a leaf node (min\_samples\_leaf), the number of features to consider at each split

(max\_features), and whether bootstrap sampling was used (bootstrap).

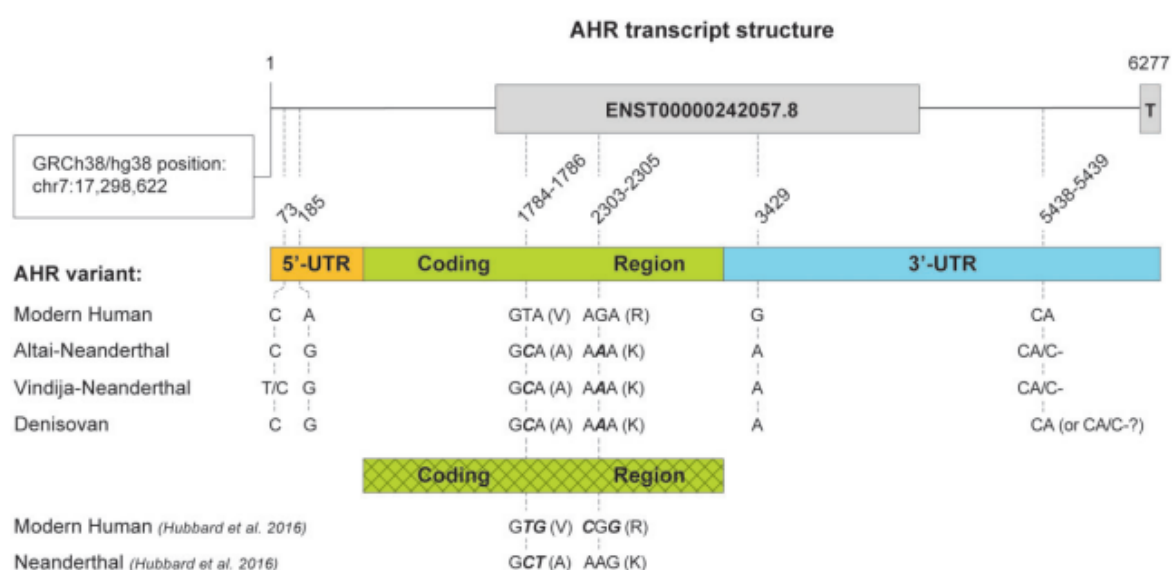
The predictive performance of the ultimate model was evaluated on the test set according to standard regression metrics, including the coefficient of determination ( $R^2$ ), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the effect size measure  $f^2$ . The relative significance of molecular descriptors was also extracted from the trained model to identify the most influential descriptors underlying SCORE prediction.

## **4. RESULTS**

### **4.1. Comparative analysis of the AHR-coding gene**

In a comparison of genomic areas between *Homo sapiens sapiens* and archaic hominins (*Homo sapiens neanderthalensis* and Denisovans), two

differences were identified. All genomes of archaic hominins that were examined were fixed for alanine at position 381, while modern humans are fixed for valine at this position. (Figure 11) We performed Basic Local Alignment Search Tool (BLAST) analysis which showed that all other species examined so far are also fixed for alanine at this position. (Figure 12) The BLAST analysis was performed using the UniProt database with an E-value threshold set at 0.01. At codon 554, Neanderthals and the Denisovan individual are fixed for lysine at position 554. Modern humans are polymorphic at this site, possessing either arginine or lysine. (28)



*Figure 10 Visualization of AHR mRNA sequence differences between modern humans and archaic hominins. In this thesis, only the coding region was analyzed. (Image taken from: Aarts et al., 2020(29))*

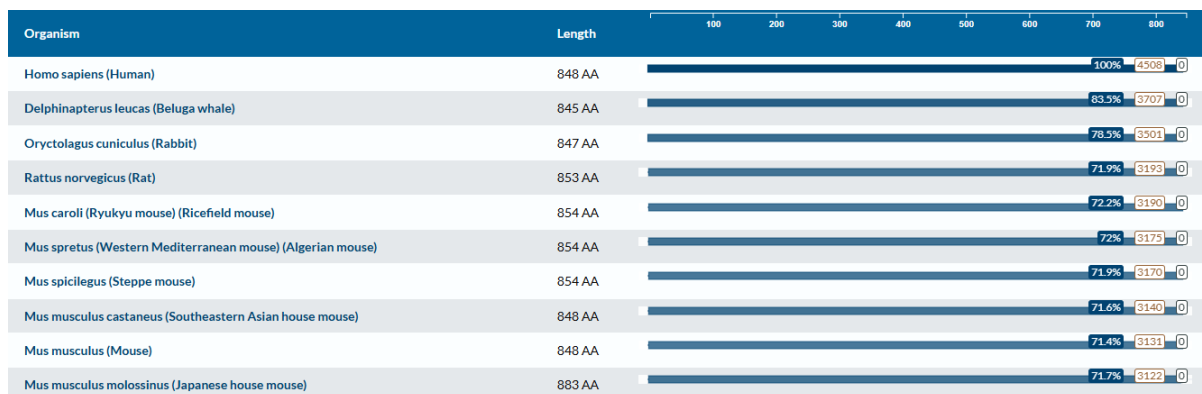


Figure 11 BLAST alignment results from UniProt showing organisms with homologous AHR protein sequences.

## 4.2. Molecular docking

### 4.2.1. Summary of docking data

The results of the molecular docking analysis are presented below. Molecular docking was performed on two different variants of the AHR gene: *Homo sapiens sapiens* (HS) and *Homo sapiens neanderthalensis* (NS). The analysis included different groups of ligands: polycyclic aromatic hydrocarbons (PAHs), microbial compounds, natural products, AHR modulators, and bioactive compounds.

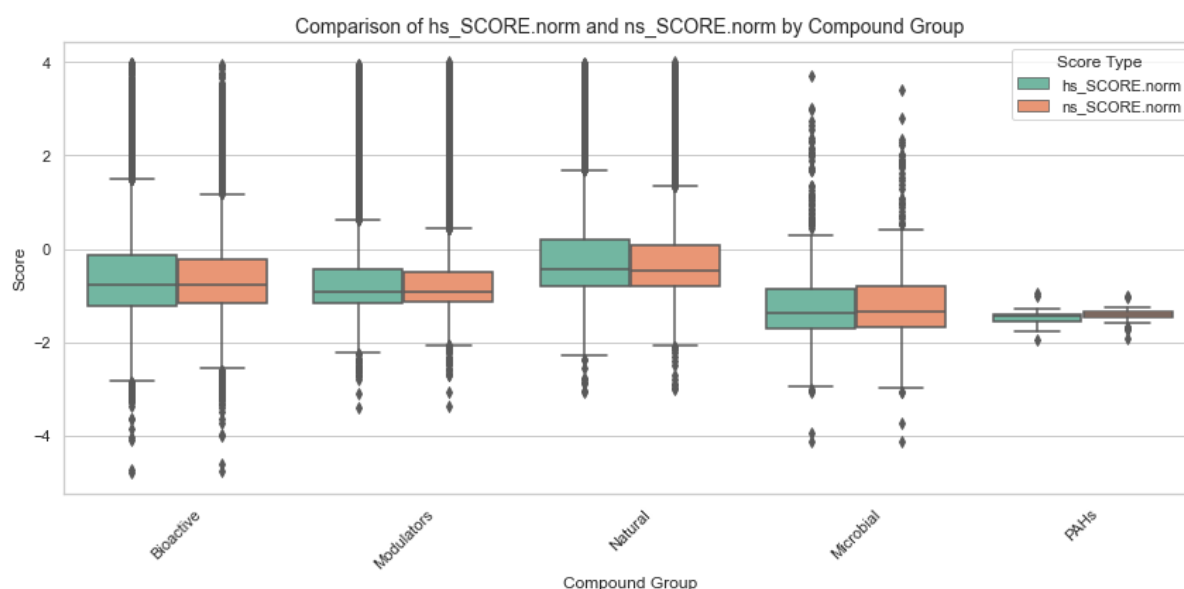
Table 1 Descriptive statistics of molecular docking results between *Homo sapiens* (HS) and *Homo sapiens neanderthalensis* (NS) AHR proteins for different groups of ligands.

| Ligand library                        | receptor                          | score_type   | count | mean     | median   | sd     | min      | max      | IQR    |
|---------------------------------------|-----------------------------------|--------------|-------|----------|----------|--------|----------|----------|--------|
| Polycyclic aromatic hydrocarbon (PAH) | Homo sapiens AHR                  | SCORE        | 29    | -27.7902 | -28.5299 | 4.2318 | -33.4432 | -19.6225 | 6.2191 |
|                                       |                                   | SCORE.INTER  | 29    | -21.3030 | -22.1104 | 4.3146 | -26.9085 | -12.6945 | 6.3752 |
|                                       |                                   | SCORE.INTRA  | 29    | 0        | 0        | 0      | 0        | 0        | 0      |
|                                       |                                   | SCORE.SYSTEM | 29    | -6.4872  | -6.5436  | 0.295  | -6.9280  | -5.4417  | 0.2365 |
|                                       |                                   | SCORE.norm   | 29    | -1.4648  | -1.4369  | 0.205  | -1.9626  | -0.9380  | 0.1537 |
|                                       | Homo sapiens neanderthalensis AHR | SCORE        | 29    | -27.0589 | -27.6041 | 3.883  | -32.1939 | -19.2418 | 6.9239 |
|                                       |                                   | SCORE.INTER  | 29    | -20.990  | -21.7730 | 4.000  | -26.0637 | -12.7706 | 6.9170 |
|                                       |                                   | SCORE.INTRA  | 29    | 0        | 0        | 0      | 0        | 0        | 0      |
|                                       |                                   | SCORE.SYSTEM | 29    | -6.0689  | -6.1347  | 0.3866 | -6.5214  | -4.6806  | 0.4191 |
|                                       |                                   | SCORE.norm   | 9     | -1.4267  | -1.3972  | 0.1941 | -1.9242  | -1.0105  | 0.1295 |

|  |  |              |       |          |          |         |          |          |         |
|--|--|--------------|-------|----------|----------|---------|----------|----------|---------|
| Microbial ligands                          | Homo sapiens<br>AHR                      | SCORE        | 534   | -11.3439 | -17.2903 | 21.8177 | -32.1084 | 150.0503 | 7.1817  |
|  |  | SCORE.INTER  | 534   | -7.0527  | -11.3611 | 19.1165 | -26.0192 | 145.1723 | 6.5651  |
|  |  | SCORE.INTRA  | 534   | 0.7519   | 0.3615   | 2.9126  | -9.8497  | 26.1419  | 1.7083  |
|  |  | SCORE.SYSTEM | 534   | -5.0681  | -5.7107  | 1.5980  | -6.8107  | 1.8297   | 1.4283  |
|  |  | SCORE.norm   | 534   | -1.1252  | -1.3780  | 1.0596  | -4.1286  | 4.1681   | 0.8433  |
|  | Homo sapiens<br>neanderthalens<br>is AHR | SCORE        | 534   | -11.7684 | -16.7643 | 19.7821 | -30.9279 | 152.6790 | 6.9561  |
|  |  | SCORE.INTER  | 534   | -7.7299  | -11.2769 | 17.2311 | -24.5212 | 146.3020 | 6.4231  |
|  |  | SCORE.INTRA  | 534   | 0.6623   | 0.3273   | 2.7976  | -10.2092 | 22.9149  | 1.7710  |
|  |  | SCORE.SYSTEM | 534   | -4.7211  | -5.3102  | 1.5569  | -6.4216  | 1.6448   | 1.4953  |
|  |  | SCORE.norm   | 534   | -1.1203  | -1.3359  | 0.9844  | -4.1413  | 4.2411   | 0.8884  |
| Natural ligands                            | Homo sapiens<br>AHR                      | SCORE        | 17170 | 1.3412   | -10.8316 | 33.8300 | -61.1827 | 276.4876 | 25.1499 |
|  |  | SCORE.INTER  | 17170 | 1.8891   | -9.5830  | 30.6160 | -25.9395 | 267.1691 | 19.8387 |
|  |  | SCORE.INTRA  | 17170 | 2.6894   | 2.4115   | 3.5919  | -36.7105 | 23.2808  | 3.5194  |
|  |  | SCORE.SYSTEM | 17170 | -3.3784  | -3.6962  | 1.7734  | -6.5439  | 7.1212   | 2.4383  |
|  |  | SCORE.norm   | 17170 | -0.0856  | -0.4242  | 1.1261  | -3.0591  | 7.8825   | 1.0393  |
|  | Homo sapiens<br>neanderthalens<br>is AHR | SCORE        | 17170 | -0.7473  | -12.0937 | 32.0581 | -61.3021 | 273.3729 | 20.9417 |
|  |  | SCORE.INTER  | 17170 | -0.2009  | -10.6277 | 28.9920 | -25.9029 | 261.1852 | 16.0920 |
|  |  | SCORE.INTRA  | 17170 | 2.4383   | 2.1928   | 3.5361  | -37.1017 | 20.4460  | 3.3081  |
|  |  | SCORE.SYSTEM | 17170 | -3.1083  | -3.4003  | 1.7058  | -6.2504  | 6.6810   | 2.3072  |
|  |  | SCORE.norm   | 17170 | -0.1516  | -0.4686  | 1.0659  | -3.0091  | 7.8107   | 0.8910  |
| Aryl hydrocarbon<br>receptor<br>modulators | Homo sapiens<br>AHR                      | SCORE        | 14645 | -10.1012 | -20.7121 | 30.9010 | -54.9035 | 294.142  | 13.518  |
|  |  | SCORE.INTER  | 14645 | -7.1715  | -16.4562 | 27.6013 | -28.5392 | 270.7444 | 9.2704  |
|  |  | SCORE.INTRA  | 14645 | 1.1323   | 0.8370   | 3.3503  | -31.1442 | 33.5028  | 3.0942  |
|  |  | SCORE.SYSTEM | 14645 | -4.1707  | -4.6251  | 1.7452  | -6.8609  | 4.4587   | 2.2750  |
|  |  | SCORE.norm   | 14645 | -0.5601  | -0.9205  | 1.1199  | -3.4004  | 8.6512   | 0.7461  |
|  | Homo sapiens<br>neanderthalens<br>is AHR | SCORE        | 14645 | -10.8620 | -20.4113 | 29.8224 | -54.1852 | 294.1529 | 11.4237 |
|  |  | SCORE.INTER  | 14645 | -8.0817  | -16.4479 | 26.6577 | -27.9113 | 261.7069 | 7.8277  |
|  |  | SCORE.INTRA  | 14645 | 0.9595   | 0.6994   | 3.2936  | -32.0016 | 41.6074  | 2.9632  |
|  |  | SCORE.SYSTEM | 14645 | -3.8376  | -4.2336  | 1.6849  | -6.4711  | 6.7148   | 2.2072  |
|  |  | SCORE.norm   | 14645 | -0.5788  | -0.8992  | 1.0800  | -3.3670  | 8.6516   | 0.6577  |
| Bioactive<br>compounds                     | Homo sapiens<br>AHR                      | SCORE        | 11944 | -6.1393  | -16.0768 | 27.8898 | -56.4993 | 348.8437 | 17.7242 |
|  |  | SCORE.INTER  | 11944 | -4.2706  | -12.3595 | 24.5632 | -30.9986 | 327.2584 | 12.1315 |
|  |  | SCORE.INTRA  | 11944 | 2.1086   | 1.3099   | 3.9090  | -34.4383 | 35.5633  | 3.7969  |
|  |  | SCORE.SYSTEM | 11944 | -4.0527  | -4.4737  | 1.8998  | -6.9919  | 4.5512   | 2.7826  |
|  |  | SCORE.norm   | 11944 | -0.5315  | -0.7688  | 1.0903  | -4.7936  | 9.9669   | 1.0934  |
|  | Homo sapiens<br>neanderthalens<br>is AHR | SCORE        | 11944 | -7.9403  | -16.0862 | 25.4756 | -55.6735 | 361.0781 | 14.5927 |
|  |  | SCORE.INTER  | 11944 | -6.0752  | -12.7296 | 22.4377 | -30.8535 | 334.1726 | 10.2184 |
|  |  | SCORE.INTRA  | 11944 | 1.8598   | 1.1594   | 3.7327  | -35.1372 | 33.2305  | 3.4142  |
|  |  | SCORE.SYSTEM | 11944 | -3.7855  | -4.1594  | 1.8090  | -6.6027  | 5.0952   | 2.6094  |
|  |  | SCORE.norm   | 11944 | -0.5827  | -0.7673  | 1.0002  | -4.7532  | 10.3165  | 0.9568  |

Descriptive statistics were used to quantify the distribution of results across different data groups. For each group, key statistical parameters were calculated, including sample size (count), arithmetic mean, median, standard deviation (SD), minimum and maximum values, and interquartile range (IQR). (Table 1)

To compare results between related groups, the Wilcoxon signed-rank test was applied. This non-parametric test is appropriate when the assumption of normality is not met and is used to evaluate whether there is a statistically significant difference between the medians of two paired samples. The test results indicated statistically significant differences between the compared groups, accompanied by a large effect size, suggesting a strong practical significance of the observed differences.



*Figure 12 shows boxplots comparing the distributions of normalized docking scores (hs\_SCORE.norm and ns\_SCORE.norm) across ligand groups: bioactive compounds, modulators, natural compounds, microbial compounds, and polycyclic aromatic hydrocarbons (PAHs).*

PAHs had the narrowest range of scores and lowest variability. This was the smallest library with just 29 compounds, but they all possessed high binding affinity towards both NS and HS AHR proteins. Bioactive, natural, modulator, and microbial compound libraries possessed wider distributions of scores with obvious outliers. Microbial ligands have a relatively stronger mean binding affinity for HS and NS AHR proteins compared to bioactive, natural and AHR modulators compound libraries, as confirmed by descriptive statistics in Table 1. Differences in score distributions for HS and NS were noted within ligand groups, as also illustrated in Figure 13, confirming the Wilcoxon test outcomes.

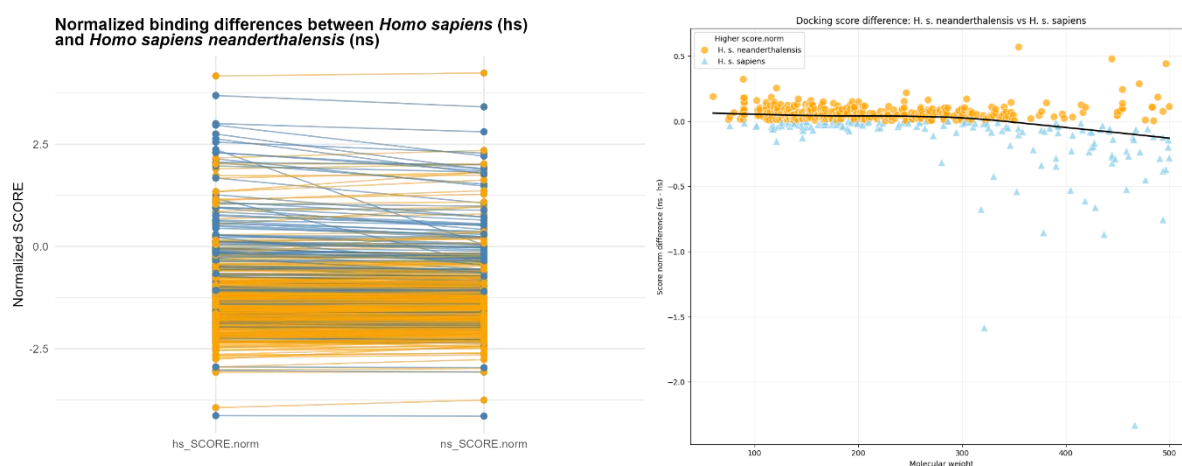
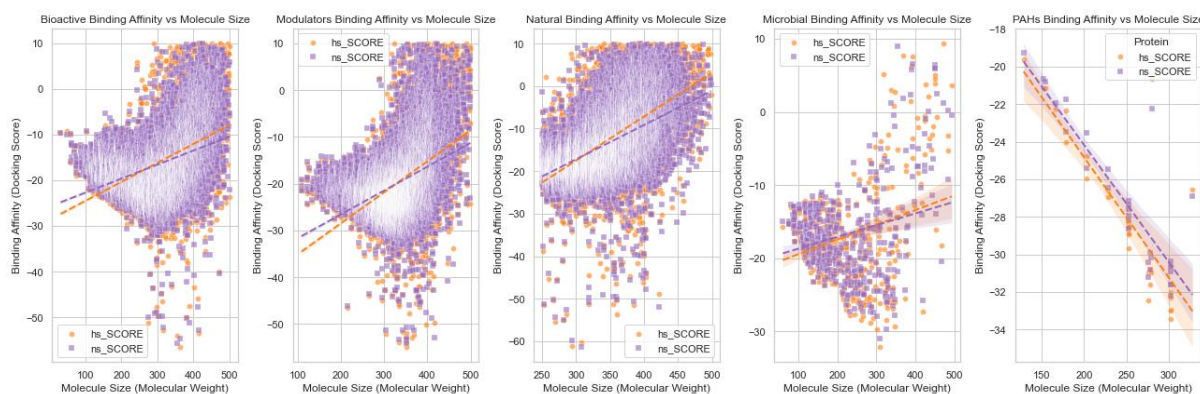


Figure 13 (left) displays paired lines connecting normalized docking scores for HS and NS. Ligands with higher scores in HS are shown in blue, while those with higher scores in NS are shown in orange. Figure 15 (right) plots the difference in docking scores ( $ns\_SCORE - hs\_SCORE$ ) against molecular weight, with a fitted trend line.

Microbial ligand library data are visualised in Figure 14 and Figure 15. The other libraries contained too many compounds to graph like that. Both receptors favored HS AHR for higher affinity ligands ( $SCORE.norm < -0.5$ ), and both receptors favored NS AHR variant for lower affinity ligands ( $SCORE.norm > -0.5$ ).

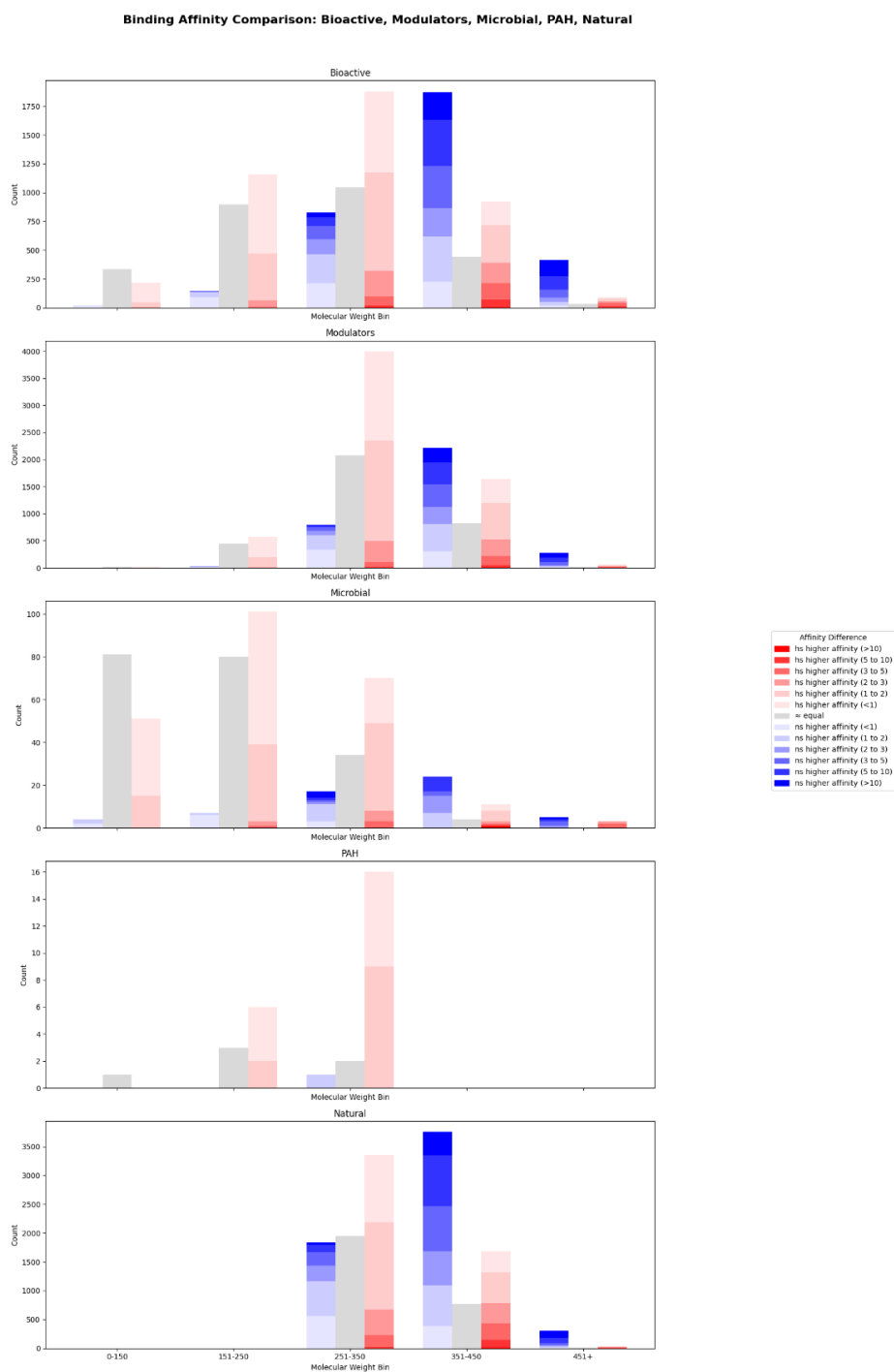
The trend was according to molecular weight: HS was preferred by lower-molecular-weight ligands and NS preferred higher-molecular-weight ligands. This trend was noticed in other ligand libraries as well (Figure 16).



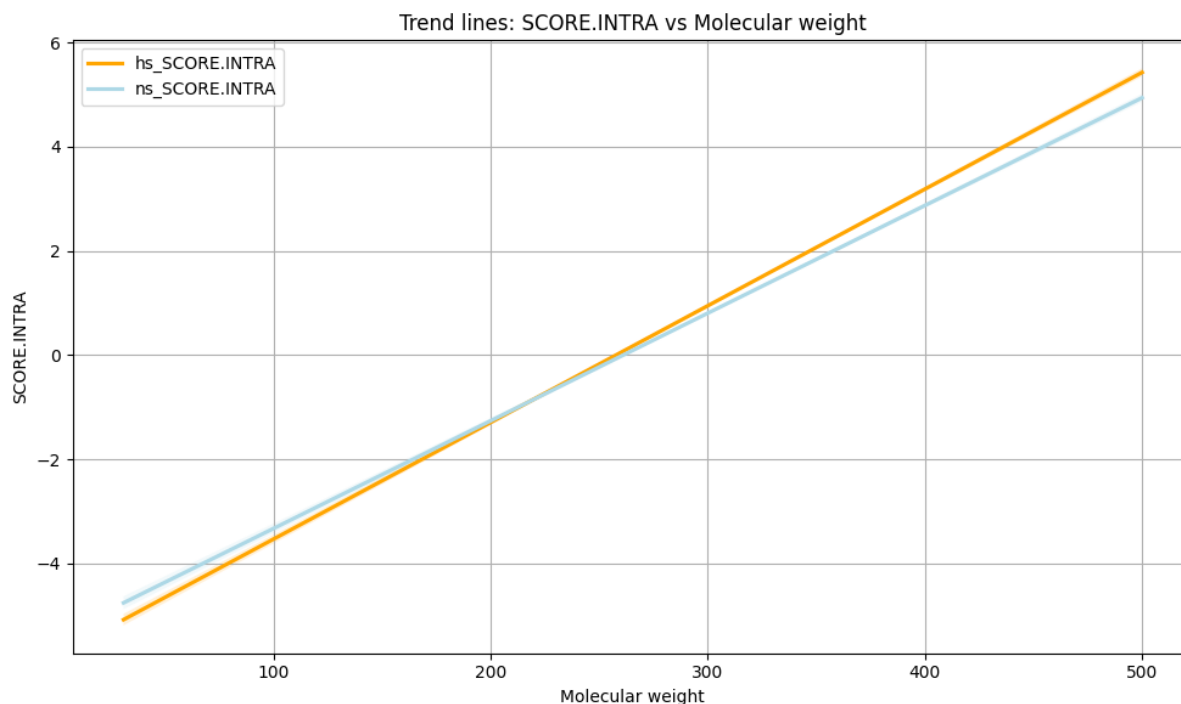
*Figure 16 presents scatter plots of the correlation between molecular weight and binding affinity for five ligand libraries: natural compounds, modulators, bioactive compounds, microbial compounds, and PAHs. Docking scores for HS (hs\_SCORE, orange) and NS (ns\_SCORE, purple) AHR variants for each ligand are shown, with regression lines characterizing trends in affinity as a function of molecular weight.*

In all libraries except PAHs, increasing trend is observed for ns\_SCORE with increased molecular weight. There is a distinct divergence of the two receptor variants in the higher mass zone (above ~300 Da), where the NS has increasing predicted affinity. There is increased stronger binding in lower-mass compounds to the HS variant. This is best observed in the PAH group, where there are no high-mass compounds. Here once again, the HS has uniformly higher binding affinity. Both receptor variants have decreasing affinity with increasing mass, but hs\_SCORE values are more positive overall. This pattern is found again in Figure 17, where compounds are grouped by molecular weight and their relative binding affinities are shown as stacked bars.

Every mass category has three bars: one for the compounds with higher affinity for the HS (red shades), one for higher affinity for the NS (blue shades), and one for equal affinity for both variants (gray shades). Intensity of shade depends on size of difference in affinity, darker for larger. The graph illustrates the dominance of HS preference for low-mass ligands and increasing prevalence of NS preference for high-mass ligands.



*Figure 17 Stacked bar plots display the distribution of docking affinity differences (NS vs. HS) across molecular weight bins. For each bin, three stacked bars are shown: one for compounds with higher affinity toward the HS (red shades), one for the NS (blue shades), and one for compounds with similar affinity for both variants (gray).*



*Figure 18. Relationship between molecular weight and internal energy upon binding (SCORE.INTRA) across all ligand compounds.*

SCORE.INTRA is the internal energy change in ligands during a binding event. A value of 0 indicates no change in energy, i.e., the molecule binds in its native form. Negative values represent energy release in association, while positive values represent that the molecule had to take in energy to reach a binding-capable conformation. Among low-mass molecules (<200 Da), HS has more favorable internal energy profiles. In high-mass molecules, ligands tend to require more energy to reach a binding-capable conformation in both receptors. (Figure 18) As Table 1 shows, all 29 PAH compounds share the same SCORE.INTRA of 0, indicating that they bind both AHR variants without conformation change.

## 4.2.2. Correlations between molecular descriptors and docking score

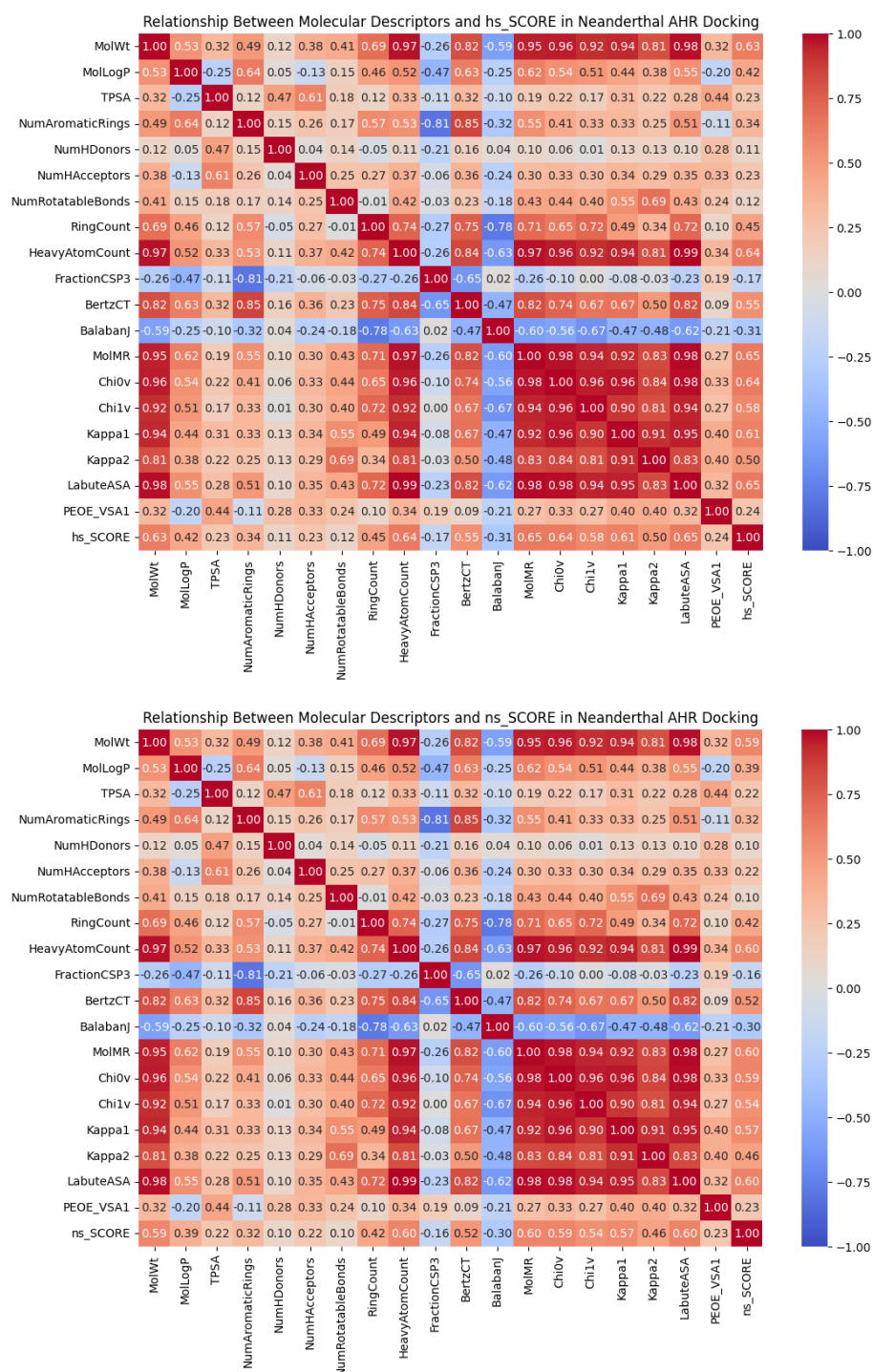


Figure 19. Heatmap showing the correlations between docking scores (*hs\_SCORE* and *ns\_SCORE*) and various molecular descriptors.

Docking scores were correlated with a set of molecular descriptors before constructing the machine learning model. The descriptors were computed with the RDKit Python package. Figure 19 represents a heatmap of Pearson correlation coefficients of docking scores (hs\_SCORE and ns\_SCORE) with selected molecular descriptors. Descriptors with strongest correlations to docking scores were selected as random forest model input variables. In case descriptors showed high intercorrelation (e.g., MolWt and HeavyAtomCount), the descriptor showing the strongest correlation to the docking score was retained to avoid redundancy in the feature space.

#### **4.2.3. Random forest**

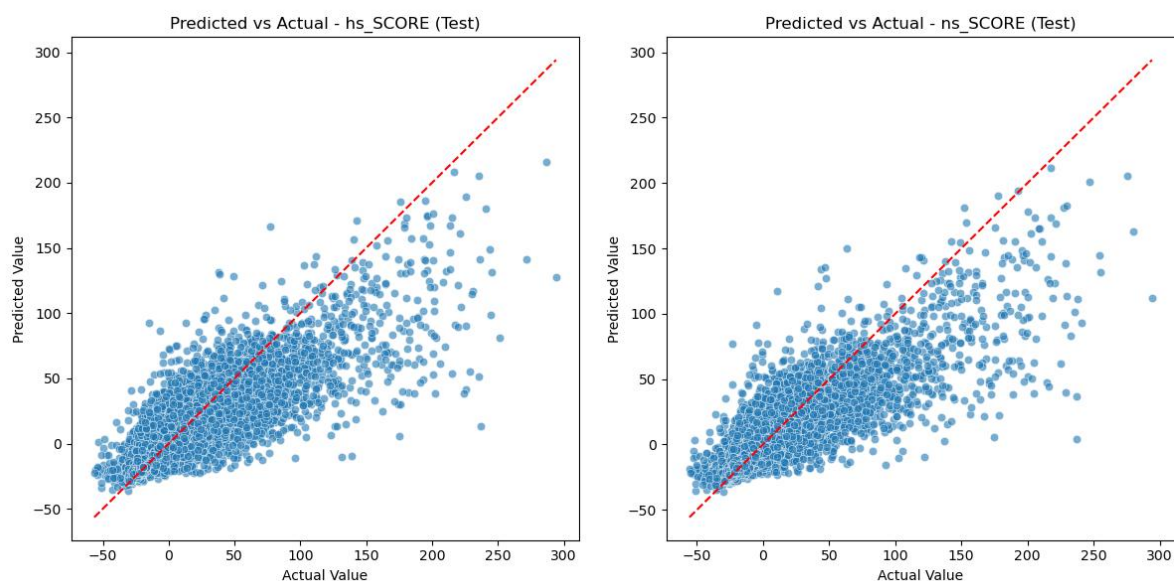
Random Forest regression model performance was compared between training, validation, and independent test datasets on a range of regression measures such as the coefficient of determination ( $R^2$ ), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and effect size measure  $f^2$ . The two target variables, hs\_SCORE and ns\_SCORE, were each modeled and compared separately.

On the training set, the model performed well for both targets with  $R^2 = 0.948$  (hs\_SCORE) and  $R^2 = 0.943$  (ns\_SCORE), along with low RMSEs of 7.30 and 7.24, respectively. The MAE was below 4 for both scores, and  $f^2$  values of 18.27 for hs\_SCORE and 16.49 for ns\_SCORE.

Validation set performance showed a trended drop in performance due to regularization and generalization effects. In the case of hs\_SCORE, the model attained  $R^2 = 0.749$  with RMSE = 15.39 and MAE = 7.99. ns\_SCORE trended in the same direction with  $R^2 = 0.721$ , RMSE = 15.30, and MAE = 7.57. The corresponding  $f^2$  values were 2.98 and 2.59.

Test set performance on the ultimate independent test set was consistent with validation outcomes. The hs\_SCORE model converged to  $R^2 = 0.750$ ,

and the ns\_SCORE model converged to  $R^2 = 0.736$ . (Figure 20) RMSE levels were comparable to those in the validation phase (15.66 for hs\_SCORE and 15.18 for ns\_SCORE) and the MAE levels were consistent ( $\sim 8.00$  and  $\sim 7.57$ , respectively). The test set  $f^2$  statistics (2.99 for hs\_SCORE and 2.79 for ns\_SCORE) corroborated the model's ability to explain variance without overfitting.

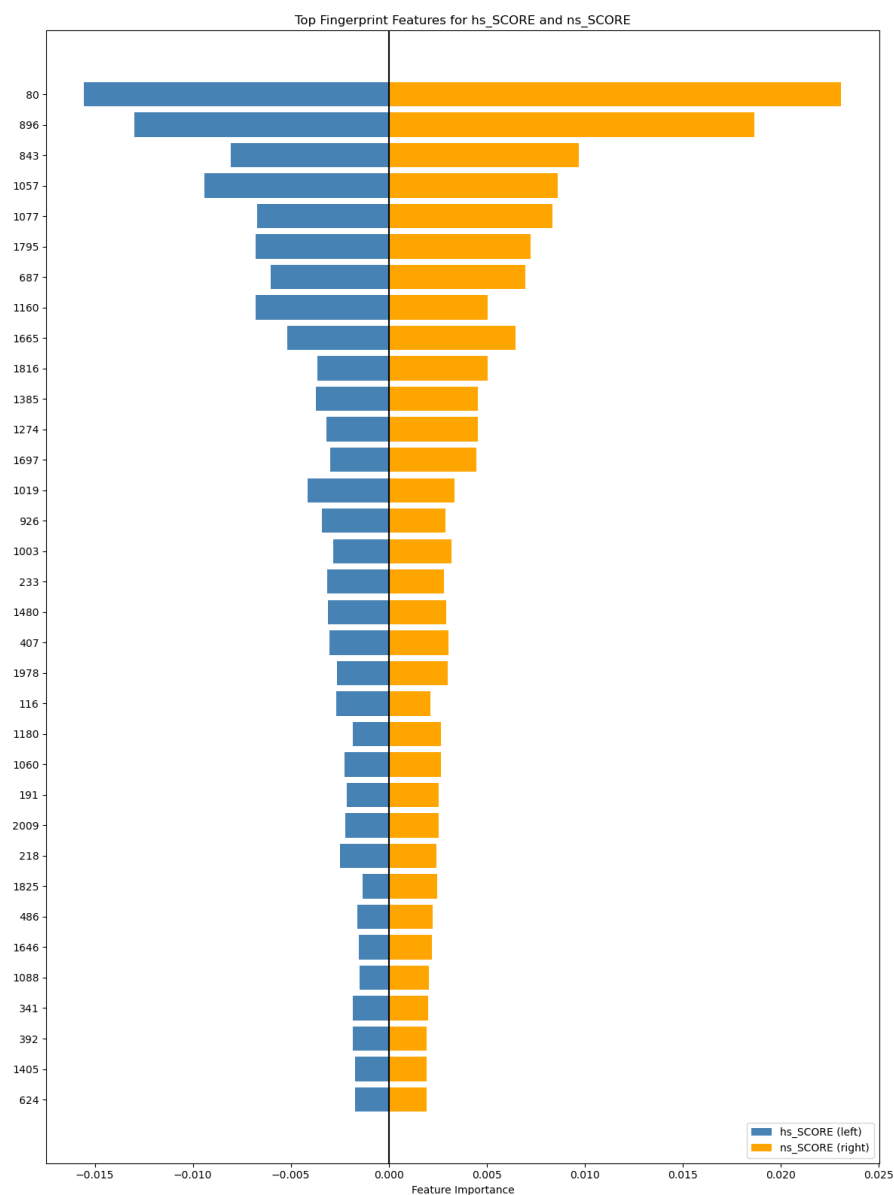


*Figure 20 Predicted vs. actual values for hs\_SCORE and ns\_SCORE on the independent test set. Scatter plots illustrate the relationship between predicted and observed values for both scoring systems. Each point represents an individual compound. The dashed diagonal line indicates the ideal scenario where predicted values perfectly match actual values (i.e.,  $y = x$ ).*

#### **4.2.4. Feature importance**

Figure 21 shows individual Morgan molecular fingerprint contribution to HS and NS protein docking scores. Morgan fingerprints were implemented using RDKit for the analysis. ID 80 (occurring in 27,337 molecules) and ID 896 (occurring in 4,174 molecules) were two most important fingerprints discovered. All fingerprint 80 molecules contain at least one polar functional

group, and fingerprint 896 is invariably associated with the presence of at least one aromatic ring.



*Figure 21 Molecular Morgan fingerprints with the highest contribution to docking score prediction for the human (left) and Neanderthal (right) AHR proteins.*

## **5. DISCUSSION**

### **5.1. Comparative analysis of the AHR-coding gene**

The AHR of archaic hominins differs from the modern human AHR by two bases at positions 381 and 554. These results are consistent with findings described in the literature. (29) It is presumed that the arginine-lysine polymorphism at position 554 in modern humans has no functional consequence. As all the archaic hominins we analyzed had arginine at this position, we proceeded and changed the sequence to reflect this for our study, bringing it in line with the genetic makeup of archaic hominins. (28)

High-coverage archaic genomes remain rare, limiting our ability to determine with certainty whether all Neanderthals were fixed to the Ala381 allele. It is important to remain open to the possibility that this mutation arose earlier in hominin evolution, but only became established later — in modern humans. However, the currently available data suggests the opposite.

There are two possible scenarios: first, that all Neanderthals were fixed for Ala381; and second, that genetic polymorphism existed at AHR position 381, allowing for the presence of the Val381 variant. While we may never be able to decipher the complete archaic genomic sequence, the currently available data suggest that the alanine variant may have been present long before the emergence of Neanderthals — possibly representing the ancestral form. However, one thing is clear: at some point after the divergence from other archaic lineages, something occurred in the evolution of modern humans that resulted in the emergence and complete fixation of the Val381 variant — a variant that, as far as is known, does not exist in any other animal species. (29)

## **5.2. Variations in the untranslated regions**

Variations exist between the untranslated regions (UTR) of the AHR gene in humans and Neanderthals. It is known that variations in UTR regions affect the functionality and expression of proteins. 5' and 3' UTR sequence variations may affect mRNA stability, translation efficiency, secondary structure, and post-translational modification, and this can alter the levels of expression and function of proteins. It is still not possible to predict how changes in the UTR regions affect protein function using only *in silico* methods. Our Neanderthal protein model of the human protein may not be entirely faithful to the real Neanderthal protein version since UTR regions can cause changes that affect protein function. Therefore, there exists a possibility that the real Neanderthal protein may possess different biological characteristics that cannot be mimicked by our model. (46)

## **5.3. Evolution of AHR**

It is risky to confidently claim that the unique characteristics of the human AHR arose as a consequence of fire use. Nonetheless, there is some basis for such a hypothesis, and it is understandable to attempt to draw a causal connection between fire use and the specific fixation of Val381 in the AHR gene in humans. To date, however, no clear and unequivocal scientific evidence exists to explain the functional impact of Val381 fixation on the AHR receptor, making it difficult to hypothesize why this change became established.

This may be the result of a neutral mutation that persisted in the population due to genetic drift—possibly reinforced by a founder effect or population bottleneck—both of which are common mechanisms in the population dynamics of early human groups. On the other hand, given that humans were undoubtedly increasingly exposed to toxic combustion by-products through fire use, it remains plausible that an environmental shift imposed a selective pressure on the receptor.

An interesting alternative hypothesis is proposed by Aarts et al., who suggest that the fixation might have been driven by selection acting not directly on the AHR gene itself, but on a nearby gene. Specifically, the HDAC9 gene, located approximately 1 Mb from AHR, is a potential candidate. HDAC9 is involved in epigenetic regulation and could have been under selective pressure, with Val381 becoming fixed indirectly due to reduced recombination in that genomic region.

The results of this research suggest that significant differences in ligand binding appear when considering molecules with higher molecular mass and volume. Valine, being bulkier than alanine and located right at the "entry" to the binding site, likely creates steric hindrance that interferes with the binding of larger ligands. On the other hand, literature generally describes AHR as primarily binding smaller ligands, making it difficult to assess the biological relevance of these results. (47)

It is important to remember that AHR biology is still relatively poorly understood. There remains a possibility that AHR also participates in biochemical pathways involving larger ligands, but further research is essential to examine this. (25)

#### **5.4. Molecular docking**

The literature describes two key in vitro studies attempting to explain functional variation between the Neanderthal and human AHR. Experiment by Hubbard et al. (2016) (28) demonstrated that the Neanderthal form of AHR (Ala381) is 150 to 1,000 times more sensitive to polycyclic aromatic hydrocarbons (PAHs), such as benzo(a)pyrene, compared with the human form (Val381). The increased sensitivity resulted in higher levels of production of the enzymes CYP1A1/CYP1B1. Conversely, Aarts et al. (2021) (27) had sampled TCDD (dioxin) and found there was no significant difference in expression of CYP1A1 between human and Neanderthal AHR variants. The dose-response curves were essentially identical.

It is important to note that both studies used an extremely limited set of ligands with very similar chemical structures, mainly PAH compounds, and these are only directly linked to the traditional detoxification role of AHR. However, AHR is involved in an enormous list of signal pathways, and there are many more AHR-binding ligands than the detoxification function. (48)

Perhaps most striking among the methodological differences in these studies is the cell source used in the experiments: while in experiments by Hubbard et al. (2016) (28) rat cells were used, experiments by Aarts et al. (2020) (29) used human HeLa cells. This could have a decisive effect on results and interpretation.

The results of molecular docking analysis showed that smaller molecular weight ligands had a better affinity for the human variant of AHR and larger molecular weight ligands favored the AHR of the archaic hominins. This is due to structural differences between the protein variants. After Aarts' hypothesis, valine at position 381 in human variant AHR forms a steric obstacle against large molecules entering the binding site. Since valine is more spatially extensive than alanine (found in the Neanderthal form), this leads to restricted accessibility and requires higher conformational adaptation on binding, increasing the internal energy of the molecule and leading to less stable binding.

Conversely, minor ligands typically require fewer conformational changes, or their binding is accompanied by a liberation of energy in the form of an energetically more favorable state. In human AHR variant, valine will also be capable of stabilizing smaller molecules within the binding cleft and lead to more tighter binding. Such interactions are stabilized by various forms of molecular forces: hydrogen bonds, Van der Waals forces, hydrophobic interactions,  $\pi$ - $\pi$  stacking, aromatic, and electrostatic interactions.

Relative docking score comparison of two homologous proteins reduces bias from scoring function inadequacy. Isoforms of AHR were docked under the same computing conditions so that systemic error equally impacts, hence

differences in the binding energy reflecting true sequence-based effects and not methodological problems. Since both isoforms share a conserved PAS-B domain topology, docking algorithms should be capable of detecting retained binding features, such as hydrophobic contacts and cavity shape, enhancing discriminatory ability even in the absence of a complete structural model. While absolute docking interpretations remain limited by structural ambiguity, relative binding comparisons between homologous proteins provide a firm foundation for evolutionary and functional inference, as shown in our AHR analysis.

### **5.5. Choosing molecular docking software**

While choosing an adequate molecular docking software, we considered several programs in use like AutoDock Vina and rDock. We opted for rDock, keeping in view the biological and structural aspects of the aryl hydrocarbon receptor (AHR) target domain, specifically its ligand-binding domain (LBD). AHR LBD possesses a stable secondary and tertiary structure, and therefore, targeted docking analysis can be carried out. In this context, the most critical requirement for selecting the docking program was being correct in identifying the binding site (cavity), especially since this involves a comparison of binding to the active site of two protein variants. One advantage of rDock is that it ensures all compounds are docked to automatically generated sites, which remain the same during the docking of each new ligand. Another advantage of rDock is the ability to finely adjust parameters. Next advantage is its penalty system, which allows us to immediately detect which ligands did not even enter the binding site.

Both rDock and AutoDock Vina are generally characterized in literature as having high ligand binding prediction accuracy tools. Their performance usually differs only in specific cases, depending on the ligand, the protein, or the complexity of the binding site. In our example, the strategy of defining the binding site turned out to be a determining point, and that is why we employed rDock, which provided greater control over the geometry of cavities and was more suitable for the LBD of the AHR protein. (36) (49)

## **5.6. Limitations**

One of the biggest limitations to structural analysis in bHLH/PAS transcription factors is the lack of full-length resolved structures available, especially with intrinsically disordered regions (IDRs) in the C-terminal part of the proteins. These regions, which usually contain transactivation or repression domains, are eluded by crystallographic or NMR structural characterization, so it is impossible to have an in-depth analysis of conformational dynamics for the full-length protein. For AHR, while high-resolution structures of isolated PAS-B domains can give an insight into the ligand-binding pocket, the lack of structural information for flexible C-terminal domains restricts mechanistic explanation of downstream transcriptional consequences. (18)

But even under these constraints, molecular docking can be a useful tool for research of such proteins when appropriate scope and interpretation is applied. In this research, we were not trying to predict absolute binding affinities or identify novel ligands, which would require a more precise and rigid structure model, but rather to contrast relative binding tendencies between Neanderthal and human AHR variants, which only vary at two amino acids. In such a case, comparative docking has some methodological advantages.

## 6. CONCLUSION

In this study, we identified key differences in the coding region of the AHR protein between *Homo sapiens sapiens*, *Homo sapiens neanderthalensis*, and Denisovan hominins. Humans are fixed for valine at position 381, while *Homo sapiens neanderthalensis*, Denisovans, and other organisms are fixed for alanine. Human valine at position 381 is unique in the animal kingdom.

Although both the human and Neanderthal forms of the protein react similarly when binding different ligands, statistically significant differences were observed in the docking of molecules with relatively larger molecular weights and volumes. These findings suggest that the change in the amino acid sequence may introduce steric hindrances upon binding large ligands in the active site. As a result, humans would be less responsive to large ligands due to potential interference on the binding site. This could represent an evolutionary adaptation to living and using fire in enclosed environments, such as caves.

## **7. REFERENCES**

1. Janković I, Karavanić I. Osvit čovječanstva. Zagreb: Školska knjiga; 2009.
2. Hublin JJ. The origin of Neandertals. *Proc Natl Acad Sci U S A* 2009;106(38):16022–7.
3. Higham T, Douka K, Wood R, Ramsey CB, Brock F, Basell L, et al. The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* 2014;512:306–9.
4. Müller UC, Pross J, Tzedakis PC, Gamble C, Kotthoff U, Schmiedl G, et al. The Repeated Replacement Model—Rapid climate change and population dynamics in Late Pleistocene Europe. *Quat Sci Rev* 2011;30(9–10):1159–77.
5. Hortolà P, Martínez-Navarro B. The Quaternary megafaunal extinction and the fate of Neanderthals: An integrative working hypothesis. *Quaternary International* 2013;295:69–72.
6. Romagnoli F, Rivals F, Benazzi S. Updating Neanderthals: Understanding Behavioural Complexity in the Late Middle Palaeolithic. London: Academic Press; 2022.
7. Smith FH, Janković I, Karavanić I. The assimilation model, modern human origins in Europe, and the extinction of Neandertals. *Quat Int* 2005;137(1):7–19.
8. Conard NJ. The demise of the Neanderthal cultural niche and the beginning of the Upper Paleolithic in southwestern Germany. *Quatern Int* 2012;259:93–104.
9. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 2012;338(6104):222–6.

10. Kuhlwilm M, Gronau I, Hubisz MJ, de Filippo C, Prado-Martinez J, Kircher M, et al. Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* 2016;530:429–33.
11. Chen F, Welker F, Shen CC, Bailey SE, Bergmann I, Davis SJM, et al. A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. *Nature* 2019;569:409–12.
12. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. *Science* 2010;328(5979):710–22.
13. Iasi LNM, Chintalapati M, Skov L, Mesa AB, Hajdinjak M, Peter BM, et al. Neanderthal ancestry through time: Insights from genomes of ancient and present-day humans. *Science* 2024;386(6727):eadq3010.
14. Djakovic I, Key A, Soressi M. Optimal linear estimation models predict 1400–2900 years of overlap between *Homo sapiens* and Neandertals prior to their disappearance from France and northern Spain. *Sci Rep* 2022;12(1):15000.
15. Slon V, Mafessoni F, Vernot B, de Filippo C, Grote S, Viola B, et al. The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature* 2018;561:113–6.
16. Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 2015;524(7564):216–9.
17. Nebert DW. Aryl hydrocarbon receptor (AHR): “pioneer member” of the basic-helix/loop/helix per-Arnt-sim (bHLH/PAS) family of “sensors” of foreign and endogenous signals. *Prog Lipid Res* 2017;67:38–57.
18. Kolonko M, Greb-Markiewicz B. bHLH–PAS proteins: Their structure and intrinsic disorder. *Int J Mol Sci* 2019;20(15):3653.

19. Bahman F, Choudhry K, Al-Rashed F, Al-Mulla F, Sindhu S, Ahmad R. Aryl hydrocarbon receptor: current perspectives on key signaling partners and immunoregulatory role in inflammatory diseases. *Front Immunol* 2024;15:1421346.
20. Rieger PG, Meier HM, Gerle M, Vogt U, Groth T, Knackmuss HJ. Xenobiotics in the environment: present and future strategies to obviate the problem of biological persistence. *J Biotechnol* 2002;94(1):101–23.
21. Sun JL, Zeng H, Ni HG. Halogenated polycyclic aromatic hydrocarbons in the environment. *Chemosphere* 2013;90(6):1751–9.
22. Haidar R, Shabo R, Möser M, Luch A. The nuclear entry of the aryl hydrocarbon receptor (AHR) relies on the first nuclear localization signal and can be negatively regulated through IMP $\alpha$ / $\beta$  specific inhibitors. *Sci Rep* 2023;13(1):47066.
23. Chang CC, Sue YM, Yang NJ, Lee YH, Juan SH. 3-Methylcholanthrene, an AhR agonist, caused cell-cycle arrest by histone deacetylation through a RhoA-dependent recruitment of HDAC1 and pRb2 to E2F1 complex. *PLoS One* 2014;9(3):e92793.
24. Szaefer H, Licznarska B, Baer-Dubowska W. The aryl hydrocarbon receptor and its crosstalk: A chemopreventive target of naturally occurring and modified phytochemicals. *Molecules* 2024;29(18):4283.
25. Stinn A, Furkert J, Kaufmann SHE, Moura-Alves P, Kolbe M. Novel method for quantifying AhR ligand binding affinities using microscale thermophoresis. *bioRxiv* 2021;2021.01.26.428246.
26. Hahn ME, Karchner SI, Merson RR. Diversity as opportunity: Insights from 600 million years of AHR evolution. *Curr Opin Toxicol* 2017;2:58–71.
27. Aarts JMMJG, Alink GM, Scherjon F, MacDonald K, Smith AC, Nijveen H, et al. Fire usage and ancient hominin detoxification genes: Protective

ancestral variants dominate while additional derived risk variants appear in modern humans. PLoS One 2016;11(9):e0161102.

28. Hubbard TD, Murray IA, Bisson WH, Sullivan AP, Sebastian A, Perry GH, et al. Divergent Ah receptor ligand selectivity during hominin evolution. Mol Biol Evol 2016;33(10):2648–58.

29. Aarts JMMJG, Alink GM, Franssen HJ, Roebroeks W. Evolution of hominin detoxification: Neanderthal and modern human Ah receptor respond similarly to TCDD. Mol Biol Evol 2020;38(4):1292–305.

30. Saleh NA, Elhaes H, Ibrahim M. Chapter 2 – Design and development of some viral protease inhibitors by QSAR and molecular modeling studies. In: Gupta SP, editor. Viral Proteases and Their Inhibitors. London: Academic Press; 2017. p. 25–58.

31. Sahraei AA, Azizi D, Mokarizadeh AH, Boffito DC, Larachi F. Emerging trends of computational chemistry and molecular modeling in froth flotation: a review. ACS Eng Au 2023;3(3):128–64.

32. Aydın SG. Review on molecular modeling and docking. SAR J Sci Res 2022;5(4):206–10.

33. Castillo Robles JM, de Freitas Martins E, Ordejón P, Cole I. Molecular modeling applied to corrosion inhibition: a critical review. npj Mater Degrad 2024;8(1):72.

34. RCSB Protein Data Bank. 2025. Available from: <https://www.rcsb.org/>

35. Adelusi TI, Oyedele AQK, Boyenle ID, Ogunlana AT, Adeyemi RO, Ukachi CD, et al. Molecular modeling in drug discovery. Inform Med Unlocked 2022;29:100880.

36. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. Biophys Rev 2017;9(2):91–102.

37. Huang SY. Comprehensive assessment of flexible-ligand docking algorithms: current effectiveness and challenges. *Brief Bioinform* 2018;19(5):982–94.
38. Khan T, Lawrence AJ, Azad I, Raza S, Khan AR. Molecular docking simulation with special reference to flexible docking approach. *SciFed Drug Deliv Res J* 2018;6(1). doi:10.47739/2334-1831/1053.
39. Prüfer K, Racimo F, Patterson N, Jay F, et al. The complete genome sequence of a Neandertal from the Altai Mountains. *Nature* 2014;505(7481):43–9.
40. Mafessoni F, Grote S, de Filippo C, Slon V, et al. A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc Natl Acad Sci U S A* 2020;117(26):15132–6.
41. Smith FH, Karavanić I, Janković I, Mauch Lenardić J. *Vindija Cave*. Zagreb: Faculty of Humanities and Social Sciences; 2024.
42. Prüfer K, de Filippo C, Grote S, Mafessoni F, Korlević P, Hajdinjak M, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* 2017;358(6363):655–8.
43. Miletic V, Katic MA, Svedruzic Z. High-throughput virtual screening web service development for SARS-CoV-2 drug design. In: 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO). Opatija, Croatia: IEEE; 2020. p. 371–6.
44. RxDock. Cavity mapping — RxDock 0.1.0 documentation. 2025. Available from:  
<https://rxdock.gitlab.io/documentation/devel/html/reference-guide/cavity-mapping.html>
45. rDock. Documentation. 2025. Available from:  
<https://rdock.sourceforge.net/documentation/./documentation/>

46. Wieder N, D'Souza EN, Martin-Geary AC, Lassen FH, Talbot-Martin J, Fernandes M, et al. Differences in 5' untranslated regions highlight the importance of translational regulation of dosage sensitive genes. *Genome Biol* 2024;25(1):111.
47. Diao X, Shang Q, Guo M, Huang Y, Zhang M, Chen X, et al. Structural basis for the ligand-dependent activation of heterodimeric AHR-ARNT complex. *Nat Commun* 2025;16(1):1282.
48. Sládeková L, Mani S, Dvořák Z. Ligands and agonists of the aryl hydrocarbon receptor AhR: facts and myths. *Biochem Pharmacol* 2023;213:115626.
49. Gaillard T. Evaluation of AutoDock and AutoDock Vina on the CASF-2013 benchmark. *J Chem Inf Model* 2018;58(8):1697–706.

## 8. CURRICULUM VITAE

PERSONAL INFORMATION Lana Žoldoš

✉ [lane@chemical.codes](mailto:lane@chemical.codes)

+385 99 402 2469

Date of birth 05/12/2000 | Nationality Croatian

### WORK EXPERIENCE

2025 - **Data analyst**  
Chemical codes, d.o.o., Zagreb  
Bioinformatics and statistical analyses under the supervision of dr. sc. Mario Lovrić

07/2024 - 10/2024 **Work&Travel program USA**  
The Little Nell, Aspen, Colorado

Cook III at banquets team

2024. **Group tutoring**  
Oratorij Marija Pomoćnica, Rijeka (Volunteer Activity)

Group tutoring for primary and secondary school students (mathematics, chemistry)

2022/2024 **Coordinator of Student Project "Biotech Magazine"**  
Biotechnology Student Association at University of Rijeka, (Volunteer Activity)

The "Biotech Magazine" is written and edited by students, with the focus of the project being on the popularization of science.

2023/2024 **Coordinator of Student Project "PosteRi"**  
Biotechnology Student Association at University of Rijeka, (Volunteer Activity)

Biotechnology Student Symposium "PosteRi" is a project that invites all students of biotechnology and related fields to share their experimental results in the form of posters.

2021-2022 **Member of the USBRI pr team**  
Biotechnology Student Association at University of Rijeka, (Volunteer Activity)

Content development and creation for the social media platforms of the faculty and the student association.

2021 **Co-coordinator of Student Project "Biotech Magazine"**  
Biotechnology Student Association at University of Rijeka, (Volunteer Activity)

The "Biotech Magazine" is written and edited by students, with the focus of the project being on the popularization of science.

09/2021 **Student Intern - Microbiology laboratory**  
Opća županijska bolnica Pakrac i bolnica hrvatskih veterana, Pakrac

Student internship in a hospital microbiology laboratory lasting 2 weeks.

2020 - 2024 **Member of the Appeals Disciplinary Committee**  
Faculty of Biotechnology and Drug Development, University of Rijeka

## EDUCATION AND TRAINING

2022 - **Biotechnology in Medicine student, master program**  
Faculty of Biotechnology and Drug Development, University of Rijeka

2024 - 2025 **Coursera courses**  
Mathematics for machine learning,  
Toxicology 21: Scientific Applications,  
Supervised Machine Learning: Regression and Classification (currently enrolled)

04/2024 **Short-term scientific mission**  
Title: Comparative analysis of the AHR gene in *Homo sapiens neanderthalensis* and *Homo sapiens sapiens*  
Biologiezentrum der Universität Wien, Austria

19/09/2023-21/09/2023 **Paleogenomics: From sampling and data collection to multidisciplinary analyses, Training school**  
Biologiezentrum der Universität Wien, Austria

2019 – 2022 **bacc. biotechn. et pharm. Inv.**  
Bachelor's thesis title: Evolution of human brain  
Faculty of Biotechnology and Drug Development, University of Rijeka

## PERSONAL SKILLS

Mother tongue(s) Croatian

Other language(s)

|         | UNDERSTANDING |         | SPEAKING           |                   | WRITING |
|---------|---------------|---------|--------------------|-------------------|---------|
|         | Listening     | Reading | Spoken interaction | Spoken production |         |
| English | B1            | B1      | B1                 | B1                | B1      |
| German  | A2            | A2      | A2                 | A2                | A2      |

Levels: A1/2: Basic user - B1/2: Independent user - C1/2 Proficient user  
Common European Framework of Reference for Languages