Nucleic Acids Research, 2025, **53**, gkaf688
https://doi.org/10.1093/nar/gkaf688
Molecular and Structural Biology

The evolution of eukaryotic linear motifs governing the function of androgen receptor from fish to *Homo sapiens*

Antonella Falconieri[1,2,†], Giulia Boarolo[1,2,†], Chiara Boschelle[1,2,†], Giovanni Spagnolli[3], Caterina Marchioretti[1,2], Emanuela Zuccaro[1,2], Isabella Palazzolo[4], Laura Tosatto[3,5], Adham Kamaleldeen Omara Hegazy[1,2], Jessica Rosati[6,7], Kenneth Fischbeck[4], Manuela Basso[3], Hans Christoph Liedtke[8], Ivan Gomez-Mestre[8], Luisa Dalla Valle[9], Emiliano Biasini[3], Pietro Faccioli[10,11,*], Alessandro Grapputo[9,12,*], Maria Pennuto[1,2,*]

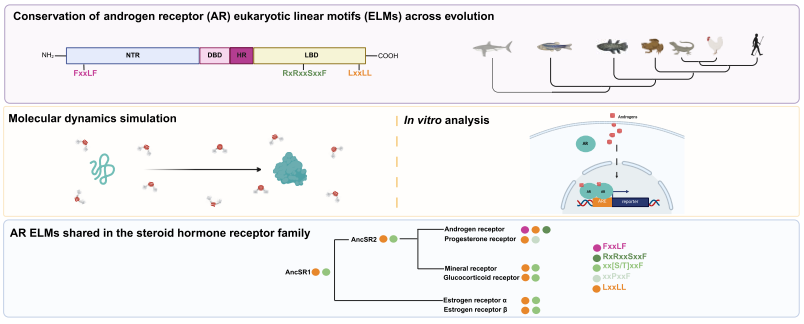
[1]Department of Biomedical Sciences, University of Padova, 35131 Padova, Italy
[2]Veneto Institute of Molecular Medicine (VIMM), 35129 Padova, Italy
[3]Department of Cellular, Computational, and Integrative Biology (CIBIO), University of Trento, 38123 Trento, Italy
[4]NINDS, National Institutes of Health (NIH), 20815 Bethesda, MD, United States
[5]Institute of Biophysics, Consiglio Nazionale delle Ricerche (CNR), Trento 38123, Italy
[6]Cellular Reprogramming Unit, Fondazione IRCCS Casa Sollievo della Sofferenza, Viale dei Cappuccini, 71013 San Giovanni Rotondo, (FG), Italy
[7]UniCamillus—Saint Camillus International University of Health Sciences, Via di Sant'Alessandro, 8, 00131 Rome, Italy
[8]Department of Ecology and Evolution, Doñana Biological Station, CSIC, Seville 41092, Spain
[9]Department of Biology, University of Padova, 35131 Padova, Italy
[10]Department of Physics, University of Milan-Bicocca, 20126 Milan, Italy
[11]Italian Institute for Nuclear Physics (INFN), Milan-Bicocca, 20126 Milan, Italy
[12]National Biodiversity Future Centre (NBFC), University of Palermo, 90133 Palermo, Italy

[*]To whom correspondence should be addressed. Email: maria.pennuto@unipd.it
Correspondence may also be addressed to Alessandro Grapputo. Email: alessandro.grapputo@unipd.it
Correspondence may also be addressed to Pietro Faccioli. Email: pietro.faccioli@unimib.it
[†]Equal contribution

Abstract

How essential *trans*-active factors that control gene expression evolved while maintaining their ability to interact with their natural effectors and *cis*-active elements on DNA is a fundamental question that requires multidisciplinary approaches. Focusing on a ligand-activated transcription factor (TF), namely androgen receptor (AR), we addressed how exonic eukaryotic linear motifs (ELMs), homorepeats (HReps), and amino acids subject to post-translational modifications (PTMs) evolved in 536 species from cartilaginous fish to *Homo sapiens*. By combining *in silico* (SWISS-Model and Alphafold3), computational (molecular dynamics), and biochemical and molecular approaches, we identify functionally active ELMs present in the *H. sapiens* AR and trace them back to Chondrichthyes, including one in the ligand-binding domain (LBD) required for androgen binding. Moreover, we provide evidence that an ELM dynamically regulated via PTM through a cryptic phosphorylation site is an old suppressive signal from fish to human. Furthermore, we identify a similar phospho-site in the LBD of the other steroid receptors and their ancestors. These findings shed light on the timing and enduring establishment of specific ELMs and their PTMs within the steroid hormone receptor (SHR) family, while highlighting early AR-specific ELMs conserved from fish to humans, as well as ELMs that emerged later in mammals.

Graphical abstract



Received: December 17, 2024. **Revised:** June 10, 2025. **Editorial Decision:** June 13, 2025. **Accepted:** July 3, 2025

Introduction

Genomes harbor essential and nonessential genes. Essential genes play a critical role in regulating gene expression, ensuring the survival of cells, tissues, and organisms. Although some genes are not indispensable for cell survival or division, they are crucial for reproduction and the overall maintenance of the organism. A prototypical example is the androgen receptor (*AR/NR3C4*), the transcription factor (TF) responsible for male sex differentiation, maturation, sexual dimorphism, and reproduction across the lifespan of most male vertebrates, ranging from fish to primates [1]. As a member of the steroid hormone receptor (SHR) family, which includes estrogen receptors α and β (ERα and ERβ), progesterone receptor (PR), glucocorticoid receptor (GR), and mineralocorticoid receptor (MR), AR serves as the primary mediator of the genomic signaling of androgens, mainly 11-keto-testosterone (11-KT) in fish and testosterone and dihydrotestosterone (DHT) in other vertebrates. The *ar* gene emerged early in vertebrate evolution, ∼400 million years ago, likely originating from a duplication of the *pr* gene, which itself evolved from an ancestral *er* [2, 3]. The presence of the *ar* gene in Chondrichthyes indicates that *ar* evolved within the Gnathostomata lineage before the divergence of cartilaginous fish from the other jawed vertebrates [4, 5]. The existence of AR in one of the oldest groups of jawed vertebrates, whose reproductive systems resemble more to those of amphibians and amniotes than to teleost fish, reflects the evolution of sexual reproduction through internal fertilization, sex determination and differentiation, sexual dimorphism, and reproductive endocrinology [6].

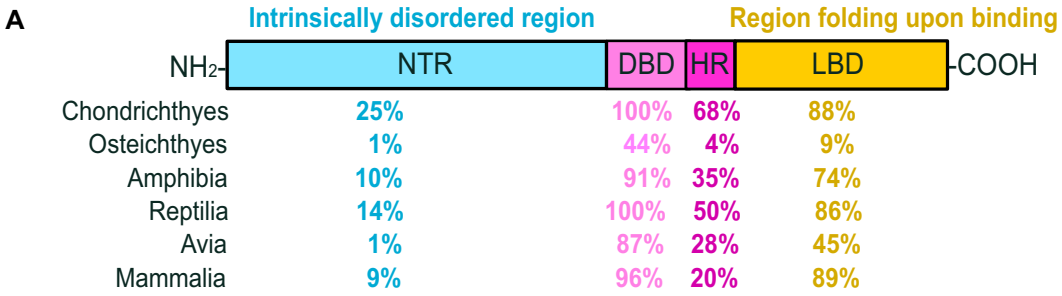
AR is a modular protein that comprises several key regions: an intrinsically disordered amino-terminal region (NTR), two zinc-finger DNA-binding domain (DBD), a flexible hinge region (HR), and a highly conserved carboxy-terminal ligand-binding domain (LBD) [7, 8]. The NTR serves as the primary *trans*-activation domain and includes the activating function 1 (AF-1) and AF-5 surfaces [9]. Upon binding to androgens, the LBD forms a globular structure consisting of 12 α-helices and 2 β-sheets [10, 11]. This configuration generates a weak AF-2 surface essential for co-regulator recruitment [12]. In its unliganded state, AR predominantly resides in the cytosol, where it is co-translationally associated with heat shock proteins (HSPs), such as HSP90 and HSP70. Ligand binding triggers a cascade of events that cause significant alterations in conformation, binding partners, and both intramolecular and intermolecular amino/carboxy (N/C)-terminal interactions [13, 14]. These changes coincide with protein stabilization [15] and nuclear translocation [16]. Once in the nucleus, AR regulates the expression of its target genes in a manner that depends on cell type and timing. Pure loss-of-function (LOF) mutations of AR are linked with androgen insensitivity syndrome [17], whose degree of penetrance depends on the residual activity left by the mutation, whereas gain-of-function (GOF) mutations may cause prostate cancer [18].

As a prototype ligand-activated TF, the function of AR at both the cellular and organismal levels is modulated by various factors, including gene transcription programs, protein translation rates, turnover, and notably, fluctuating ligand concentrations within target cells. AR activity is also influenced by *cis*-acting protein regions, such as exonic homorepeats (HReps), eukaryotic linear motifs (ELMs), and amino acids subjected to post-translational modifications (PTMs). Importantly, there is a positive correlation between the presence of HReps and PTMs in essential genes [19]. HReps, known for their hypervariability, serve as a significant source of genetic diversity, playing a crucial role in human development and brain function. On the contrary, once established, ELMs and their associated PTMs are subject to evolutionary constraints, which either limit variability or allow species-specific adaptations. In this study, we used AR as a model of a ligand-activated TF that undergoes folding upon binding to explore the evolutionary origins and relationships of HReps, ELMs, and PTM sites, which affect protein structure and function, thus influencing gene expression.

Materials and methods

Multiple sequence alignments

All available sequences that encode AR orthologs were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/, as of April 2025). To increase the number of sequences for amphibian species, all *de novo* Amphibian transcriptome assemblies available in the NCBI TSA database (https://www.ncbi.nlm.nih.gov/genbank/tsa/) were downloaded. The transcript sequences were then screened for AR orthologs using diamond BLASTx [20] and the *Xenopus tropicalis* Ensembl proteome version 109. Settings were set to ultra-sensitive, and the aligned parts of the query sequence were extracted and translated by including the qseq_translated output. Taxonomy follows the Amphibian Species of the World (https://amphibiansoftheworld.amnh.org/), with species names cross-checked using AmphiNom [21]. Each dataset was manually curated to exclude incomplete sequences from the analysis. Phylogenetic trees showing the relationships of the species for which an AR sequence was available were obtained separately for the classes of Vertebrata. For Chondrichthyes, Amphibia, Avia, and Mammalia a consensus tree was obtained with Treeannotator included in BEAST 2.7.4 [22] from 10 000 trees downloaded from the Vertlife phylogenetic subsets (https://vertlife.org/; last accessed on 3 August 2024). For the Avia tree, the relationships of Palaeognathae [23] were pruned with the package ape ver 5.8 [24] in R ver. 4.3.3 and then included in the Vertlife avian tree. The Reptilia tree was obtained from Colston [25] pruned with ape in R and the branch leading to Sphenodon punctatus was substituted with the Lepidosauria tree obtained from Vertlife. A tree for Actinopterygii was obtained from the Fish tree of life website (https://fishtreeoflife.org/) [26] and pruned for the species with available AR with the R package ape. The trees were drawn with Figtree ver. 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). For Fig. 1A, we selected a representative sequence for each group to assess the conservation of AR domains within each class. We used *Carcharodon carcharias* for Chondrichthyes, *Danio rerio* for Actinoperygii, *Xenopus laevis* for Amphibia, *Protobothrops mucrosquamatus* for Reptilia, *Gallus gallus* for Avia, and *Homo sapiens* for Mammalia. Multiple alignments were performed with MAFFT v7.490 using BLOSUM80 as scoring matrix, gap open penalty 1.53 and offset value 0.123 and visualized with Geneious Version 2024.0.7. By comparing the primary structure with that of *H. sapiens* AR, we manually annotated the NTD, DBD, HR, and LBD. The number of identical sites and the percentage of identity was calculated automatically by the software.



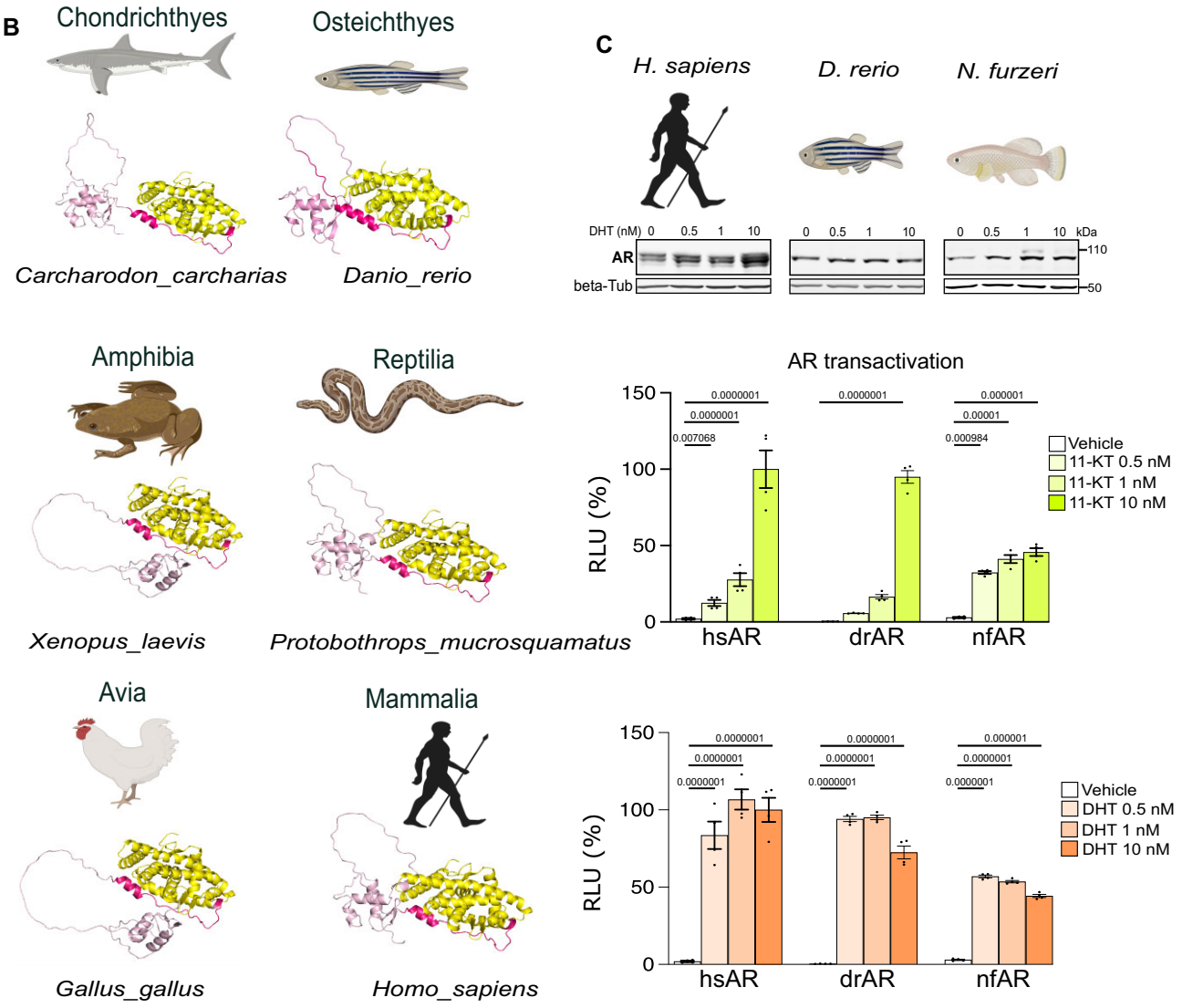


Figure 1. Phylogenetic conservation of AR primary-to-tertiary structure and function. (**A**) Schematic of AR domain organization: NTR, amino-terminal region; DBD, DNA-binding domain; HR, hinge region; LBD, ligand-binding domain. Alignment of AR primary structure (amino acidic sequence) in 518 vertebrates (10 Chondrichthyes sequences, 156 Osteichthyes sequences), Amphibia (35 sequences), non-avian Reptilia (37 sequences), Avia (113 sequences), and Mammalia (167 sequences) revealed high conservation of the DBD and LBD. The degree of conservation by AR protein regions is shown with the color corresponding to the protein region. (**B**) The SWISS-model prediction of the structure of the DBD-HR-LBD revealed that the secondary and tertiary structure of a representative AR ortholog for each vertebrate class is conserved. (**C**) Top panels: Western blot analysis of Flag-tagged *H. sapiens* (hsAR), *D. rerio* (drAR), and *N. furzeri* (nfAR) AR orthologs ectopically expressed in HEK293T cells treated with either vehicle (EtOH) or DHT (20 h). Shown is one experiment representative of three biological replicates. hsAR was detected with a specific antibody, and drAR and nfAR were detected with an anti-Flag antibody; β-tubulin (β-Tub) was used as a loading control. Middle and bottom panels: Transactivation assays measuring the activity of ectopically expressed hsAR, drAR, and nfAR in HEK293T cells treated with vehicle or 11-KT (middle panel) and DHT (bottom panel) for 20 h as indicated ($n = 4$ biological replicates). The graphs show the mean ± SEM; two-way ANOVA followed by Bonferroni HSD tests. "Created in BioRender. Falconieri, A. (2025) https://BioRender.com/m10f120".

ELM identification

To identify potential short linear motifs, we used the ELM database (http://elm.eu.org/) [27]. The H. sapiens AR amino acidic sequence was obtained from NCBI database and used as query for the prediction. The tool identified 356 ELM instances classified into 88 distinct categories (http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKOdVfejg&EXPECT_CUTOFF=100&r=1&bg=on).

AR structure prediction

To predict AR structure for *Carcharodon carcharias* for Chondrichthyes, *D. rerio* for Osteichthyes, *X. laevis* for Amphibia, *P. mucrosquamatus* for Reptilia, *G. gallus* for Avia, and *H. sapiens* for Mammalia, we generated a modeling using SWISS-Model software (https://swissmodel.expasy.org/) and Alphafold3 (https://alphafoldserver.com/). We used each AR amino acidic sequence as input and selected the structure with higher score. PyMol 2.5.7 was used for the visualization of predicted AR structures.

Plasmids

Plasmids encoding *H. sapiens* AR and AR mutants were previously generated in the lab. Vectors encoding *D. rerio* (drAR, ODa10403), *N. furzeri* (nfAR, ONo18063), and *H. sapiens* estrogen receptor α (ERα) (OHu26586D) and their mutants were obtained from Genscript Biotech (Genscript.com).

Transcriptional assays

Transcriptional assays were performed according to the manufacturer's instruction (Dual-Luciferase assay kit, Promega E1910). Briefly, HEK293T cells were transfected with plasmids expressing either AR or ERα together with TK-Renilla and pARE-Luciferase or pERE-luciferase plasmids, as previously described [28]. Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with charcoal-stripped fetal bovine serum (CDS) in presence of vehicle, DHT, 11-KT, and E2. After 20 h, the cells were washed with PBS at room temperature (RT), lysed in passive lysis buffer, and rocked for 30 min at RT. To measure luminescence, each sample was transferred to a white 96-well plate, mixed with luciferase substrate, and analyzed with Infinite 200Pro (Tecan Instruments). After the first measures, Renilla substrate was added to each sample to measure Renilla luciferase. All luciferase measures were normalized to Renilla.

Biochemical analysis

For western blotting, HEK293T cells were transfected using a protocol based on polyethyleneimine (PEI), as previously described [29], cultured in DMEM supplemented with charcoal-stripped CDS and treated with either vehicle or DHT and 11-KT. After 20 h, the cells were washed with ice-cold phosphate buffered saline (PBS) 1× and lysed in RIPA buffer [25 mM tris (pH 7.5), 150 mM NaCl, 0.1% sodium dodecyl sulfate (SDS), 0.5% sodium deoxycholate, and 1% NP-40] supplemented with protease inhibitor cocktail (Complete Tablets EDTA-free, Roche). Lysates were clarified by centrifugation at 15 000 rpm 4°C for 15 min. Protein concentration was measured using the bicinchoninic acid (BCA) assay (Pierce BCA Protein Assay, Thermo Fisher Scientific). Equal amounts of protein extracts were loaded on sodium dodecyl sulfate-polyAcrylamide gel electrophoresis (SDS–PAGE) upon addition of 5× Sample Buffer (62.5 mM Tris–HCl pH 6.8, 2% SDS, 25% glycerol, 0.05% bromophenol blue, and 5% β-mercaptoethanol) and denaturation for 5 min at 95°C. Proteins were transferred to 0.45 μm nitrocellulose membranes and blocked for 1 h in 5% nonfat dry milk or 5% bovine serum albumin (BSA) in tris-buffered saline (TBS). Membranes were incubated overnight with primary antibodies: AR (1:1000 Santacruz sc-7305) in 5% milk Tris/buffered saline (TBS) 1× with Tween 0.1%, or anti-FLAG in 5% BSA TBS Tween 0.1% (1:4000 Sigma 7425). Immunoreactivity was detected using IRDye secondary antibodies for the Odyssey Imaging system (LI-COR Biosciences), following the manufacturer's instructions.

Immunofluorescence assays

COS1 cells were transfected with plasmids expressing normal AR (AR24Q) and ARpolyQ (AR65Q) with and without the mutation of the LxxLL motif to LxxAA. Cells were treated with vehicle or DHT 10 nM for 20 h, fixed in 4% paraformaldehyde (PFA) for 20 min, washed with PBS 1×, and permeabilized in 0.1% Triton X-100 in PBS 1× for 10 min. Coverslips were covered with blocking solution for 1 h and then incubated with primary antibody: AR(N20) (1:250, Santa Cruz sc-816), and secondary antibody: Alexa Fluor 488-Goat AntiRabbit IgG antibody (A-11008, Molecular Probes). Images were acquired digitally with a DeltaVision microscope and deconvolved with the softWoRx algorithm (Applied Precision).

Folding pathways simulations

General features. The enhanced sampling method used in this work is based on combining two key ingredients: (i) an enhanced sampling technique to accelerate molecular dynamics (MD) simulations and efficiently generate protein folding pathways, and (ii) a ranking scheme to identify the most accurate trajectories among those generated by the enhanced sampling method. Specifically, we resorted to Ratchet-and-Pawl Molecular Dynamics (rMD) for the first task and to the Bias Functional (BF) method for the second [30]. The software to perform these simulations relies on the MD engine of Gromacs 4.6.5 patched with the plugin for collective variable analysis Plumed 2.0.2, which is freely available for download (https://www.plumed.org/)

Generation of denatured conditions. rMD simulations need to be initiated from unfolded configurations. To generate them, we performed several 3–5 ns trajectories of MD at high temperature (800 K) keeping fixed the number of particles the simulation box's volume (NVT ensemble) starting from the energy-relaxed native structure. For each trajectory, we extracted a single denatured conformation.

Generation of folding pathways by rMD. For each denatured conformation, we generated a set of folding trajectories with the rMD algorithm. In this scheme, the folding progress is described as a function of a reaction coordinate, defined as $z(X)$:

$$z(\boldsymbol{X}) \equiv \sum_{|i-j|>35}^{N} \left[C_{ij}(\boldsymbol{X}) - C_{ij}\left(X^{Native}\right) \right]^2 \qquad (1)$$

Where $C_{ij}(X)$ is the continuous contact map of the instantaneous system configuration $X(t)$, and $C_{ij}(X^{Native})$ is the

contact map in the target state. The latter is obtained by energy minimizing the experimental native structure retrieved from the protein data bank. The $C_{ij}(X)$ entries of $z(X)$ interpolate smoothly between 0 and 1 according to the following function:

$$C_{ij} = \begin{cases} \frac{1-\left(\frac{r_{ij}}{r_0}\right)^6}{1-\left(\frac{r_{ij}}{r_0}\right)^{10}} & \text{if} r_{ij} < r_c \\ \frac{6}{10} & \text{if} r_{ij} = r_0 \\ 0 & \text{if} r_{ij} \geq r_c \end{cases} \quad (2)$$

where r_{ij} is the Euclidean distance between the i_{th} and the j_{th} atom, r_0 is a typical distance defying residue contacts (set to 7.5 Å), and r_c is a cutoff distance (set to 12.3 Å) beyond which the contact is set to 0.

In rMD, the protein evolves according to plain MD as long as the reaction spontaneously proceeds towards the native state, i.e. if $z[X(t+\Delta t)] < z[X(t)]$. On the other hand, when the chain tries to backtrack along $z(X)$—i.e. if $z[X(t+\Delta t)] > z[X(t)]$—an external biasing force is introduced that redirects the dynamics towards the native state. The biasing force acting on a given atom, F_i^{rMD}, is defined as:

$$F_i^{rMD}(X,t) = \begin{cases} -k_r \nabla_i z(X) \cdot \{z(X) - z_m(t)\}, & if z(X) > z_m(t) \\ 0 & if z(X) \leq z_m(t) \end{cases} \quad (3)$$

where $z_m(t)$ indicates the smallest value of the reaction coordinate $z(X)$ up to time t, and k_r is a coupling constant.

Selection of the least biased trajectories. The history-dependent biasing force of the rMD scheme drives the system out of thermal equilibrium, potentially preventing us to access information about the protein folding dynamics, which is driven by spontaneous thermal fluctuation in equilibrium conditions. However, it was shown that, in the limit in which the biasing force is applied along the ideal reaction coordinate (namely, a parametrization of the so-called committor function), rMD simulations yield a correct sampling of the equilibrium distribution, in the region connecting the reactant and product states [30]. Hence, at least in principle, rMD can be used to infer the folding mechanism. In practice, however, simulations rMD rely on a phenomenologically inferred proxy of the reaction coordinate (the z(X) collective variable defined in Eq. 1), so the results may be affected by systematic errors. The BF approach is a variational scheme that enables keeping these systematic errors to a minimum. This is done by ranking the folding trajectories generated from a given unfolded conformation using the following functional:

$$T = \sum_{i=1}^{N} \frac{1}{m_i \gamma_i} \int_0^t d\tau \left| F_i^{rMD}(X, \tau) \right|^2 \quad (4)$$

Here, t is the trajectory folding time, m_i and γ_i are the mass and the friction coefficient of the i-th atom, respectively, and F_i^{rMD} is the force acting on it. It has been previously shown that the trajectories with smaller values of T have larger probability to occur in the absence of any biasing force [31]. In particular, the folding trajectory with the minimum value the bias functional for each set is referred to as Least Biased (LB) trajectory.

Computational analyses of the androgen receptor

Structure and topology of the AR LBD. Atomistic structure of the AR LBD was retrieved from PDB 3B68, the structure spans from residue 671 to 917 and contains the receptor bound to SARM-S4 (an AR agonist). Ligand structure was retained in the plain-MD simulations, while it was removed to perform folding pathway simulations. Protein topology was generated in Gromacs 4.6.5 using Amber99SB-ILDN force field in TIP3P water. Ligand topology was generated by employing the AM1-BCC method in Antechamber.

Plain MD simulations of the AR LBD. Setup and MD production was performed in Gromacs. The structure of the AR LBD in complex with SARM-S4 was positioned in a dodecahedral box with 12 Å minimum distance from the walls. The box was filled with TIP3P water molecules and neutralized with 5 Cl^- ions. The system was energy minimized using the steepest descent algorithm. Equilibration in the NVT ensemble was then performed for 500 ps at 310 K using a velocity-rescale thermostat ($\tau T = 0.1$ ps) with positional restraints on heavy atoms. The system was then equilibrated in the NPT ensemble for 500 ps at 310 K and 1 Bar by employing the velocity-rescale thermostat and the Parrinello Rahman barostat ($\tau P = 2$ ps) with positional restraints on heavy atoms. Finally, 500 ns of plain MD were performed keeping the number of particles, pressure, and temperature fixed (NPT ensemble) by releasing the positional restraints. All MD simulations were performed using a leap-frog integrator with time-step equal to 2 fs, bond constraints were treated using LINCS. Van der Waals and Coulomb cutoff was set to 12 Å, while long-range electrostatics were treated using Particle Mesh Ewald. Atomic configurations were saved every 200 ps.

Analysis of plain MD simulation. RMSD of the MD trajectory was computed using Gromacs, while the solvent accessible surface area (SASA) of serine 792 was computed by employing the Shrake-Rupley algorithm in the python library MDTraj. The relative surface area (RSA) was calculated by normalizing the SASA of S792 with the SASA value of the serine in a GSG linear tripeptide.

Folding simulations of the AR LBD. The native structure of the AR LBD (PDB 3B68) was deprived of the ligand and positioned in a dodecahedral box with 40 Å minimum distance from the walls. The box was filled with TIP3P water molecules and neutralized with 5 Cl^- ions. The system was energy minimized using the steepest descent algorithm. NVT equilibration was then performed for 500 ps at 800 K using the V-rescale thermostat with positional restraints on heavy atoms. Restraints were then removed, and three independent 5 ns of plain MD were performed in the NVT ensemble at 800 K, yielding three independent denatured conformations. Each initial condition was then repositioned in a dodecahedral box with 15 Å minimum distance from the walls, energy minimized using the steepest descent algorithm and then equilibrated first in the NVT ensemble (using the V-rescale thermostat at 350 K, $\tau T = 0.1$ ps) and then in the NPT ensemble (using the V-rescale thermostat at 350 K, $\tau T = 0.1$ ps, and the Parrinello-Rahman barostat at 1 bar, $\tau P = 2$ ps). For each initial condition, 20 trajectories were generated by employing the rMD algorithm in the NPT ensemble (350 K, 1 bar). Each trajectory consists of 2.5×10^6 rMD steps with integration time of 2 fs. Frames were saved every 5×10^3 steps. The ratchet constant k_r was set to 5×10^{-4} kJ/mol. Nonbonded interactions were treaded as follow: Van der Waals and Coulomb cutoff was set to 12 Å, while Particle Mesh Ewald was employed for long-range electrostatics. For each set, the bias functional scheme was applied yielding a total of 3 LB trajectories.

Analysis of trajectories. RMSD was computed using Gromacs while the fraction of native contacts (Q) was computed using VMD 1.9.2. A lower-bound approximation of the energy landscape *G(Q, RMSD)* was generated by plotting the negative logarithm of the 2D probability distribution of the collective variables Q and RMSD, obtained from the 60 rMD trajectories (115 × 115 bins). Protein conformations belonging to the LB trajectories and spanning over the energetic wells of interest ($G < 2.5\ k_BT$) were sampled. The RSA was calculated as in the AR plain MD analysis by employing the Shrake–Rupley algorithm in the python library MDTraj. Data were represented using the Matplotlib library in python. The 2D [Q, RMSD] energy plot was smoothed with a Gaussian kernel, while the 2D RSA plot of serine 792 was interpolated using an inverse distance weighting with a quartic Lorentzian metrics. Images of the protein conformations were produced using UCSF Chimera.

Statistical analysis

Statistical analysis was performed with Jamovi (Version 2.3.18.0) or Microsoft Office Excel (Microsoft, version 2017). To compare the mean difference of a dependent variable between independent groups, one-way analysis of variance (ANOVA) was used, and two-way ANOVAs were performed to evaluate the effects of two categorical predictors on a dependent variable. For all ANOVAs, follow-up Bonferroni post hoc tests were conducted for pairwise comparisons. All tests were two-tailed, and the significance threshold was set at $P < 0.05$.

Results

AR is a structurally and functionally conserved ligand-activated TF

The evolution of *trans*-active factors involved in gene expression regulation is constrained by the necessity for coevolution with binding effectors, such as androgens in the case of AR, and cognate chromatin *cis*-active elements, such as androgen-responsive elements (AREs) [32–34]. To better understand how an essential TF evolves, we examined the evolutionary conservation of AR across a range of species, from fish to mammals. To determine how the primary structure (amino acidic sequence) is conserved, we analyzed the available and newly annotated AR sequences from Fish (11 Chondrichthyes species, 124 Osteichthyes species), Amphibia (59 species), non-avian Reptilia (44 species), Avia (128 species), and Mammalia (170 species), for a total of 536 species (Fig. 1A and Supplementary Table S1). To analyze the AR domain conservation within each class, we manually curated each dataset to exclude incomplete sequences. We aligned 10 sequences for Chondrichthyes, 156 for Osteichthyes, 35 for Amphibia, 37 for Reptilia, 113 for Avia, and 167 for Mammalia. Sequences were first aligned separately by class and then all together. The greatest variability was observed in Osteichthyes [4]. The most conserved regions of AR were DBD and LBD, with amino acid composition showing remarkable conservation across the analyzed species. DBD was the most conserved domain, consistent with the coevolutionary constraints imposed on domains that directly interact with their cognate *cis*-active elements [32, 35]. LBD also exhibited a high degree of conservation, except in Osteichthyes, consistent with the coevolution of the receptor binding pocket and its cognate ligand [2, 3].

To further explore the evolutionary degrees of secondary (α-helix and β-sheet) and tertiary (globular) protein structure, we utilized the crystal structure information available for AR bound to ligand [10, 11], and we employed the SWISS-Model to predict whether the DBD-HR-LBD structure is conserved throughout evolution (Fig. 1B). We analyzed the DBD-HR-LBD structure in representative species of each class, namely *Carcharodon carcharias* for Chondrichthyes, *Danio rerio* for Osteichthyes, *Xenopus laevis* for Amphibia, *Protobothrops mucrosquamatus* for Reptilia, *Gallus gallus* for Avia, and *Homo sapiens* for Mammalia. SWISS-Model returned highly similar structures for the species analyzed. Similar results were obtained using Alphafold3 (data not shown). These analyses indicate that the secondary and tertiary structure of AR is conserved from fish to mammals.

To determine whether AR function is conserved throughout evolution, despite the divergence of NTR, we tested whether binding to DHT and 11-KT results in transactivation of AR from fish, specifically *D. rerio* (drAR) as representative of Otomorpha and *Notobranchius furzeri* as representative of Percomorphaceae, to *H. sapiens* (hsAR) in a heterologous system [36, 37]. We ectopically expressed hsAR, drAR, and nfAR in human embryo kidney 293 T positive cells (HEK293T), and we verified by Western blotting that these AR orthologs are expressed in eukaryotic cells (Fig. 1C, top panels). We then performed an AR transactivation assay using luciferase as a reporter under the control of a mammalian ARE (ARE-Luc), as previously described [29]. Treatment with physiological levels (0.5–10 nM) of 11-KT triggered hsAR transactivation in a dose-dependent manner (Fig. 1C, middle panel). As expected, DHT was much more potent than 11-KT in this assay (Fig. 1C, bottom panel). Treatment with DHT and 11-KT also resulted in the transactivation of drAR and nfAR, indicating that fish AR function is preserved in mammalian cells, as previously demonstrated for *D. rerio* and other fish species [36–38]. These findings collectively show that the structure and native function of AR are evolutionarily ancient and conserved.

Exonic microsatellite HReps are a convergent evolutionary feature of mammalian AR

Next, we investigated the evolution and conservation of AR HReps (Supplementary Table S1). The most common HReps, typically comprising stretches of at least three amino acids, include residues of glutamine (polyQ), alanine (polyA), serine (polyS), glycine (polyG), histidine (polyH), aspartic acid (polyD), glutamic acid (polyE), arginine (polyR), and proline (polyP) [39]. These microsatellite repeats are polymorphic in length, and, in some cases, their length correlates with species complexity. For example, the length of the huntingtin polyQ tract correlates with brain complexity [39–41]. hsAR contains three polyQ tracts [42], each separated by short sequences containing serine/proline (SP) phosphorylation sites, specifically ^{83}SP^{84} and ^{96}SP^{97} (based on NP_000035.2), which are highly conserved and crucial for AR biology [43, 44]. PolyQ tracts are often preceded by leucine residues (L), which are necessary for proper folding of these HReps [45]. In humans, the first polyQ tract ranges from 5 to 36 glutamines, with expansions beyond 38 glutamines leading to Kennedy's disease (OMIM 313200), a neurodegenerative disorder characterized by misfolding and aggregation of mutant AR [9]. The

comparison of AR across each group of vertebrates revealed that AR in fish lacks polyQ repeats (Fig. 2; Supplementary Figs S1--5 and Supplementary Table S1). However, a few species, such as *Pangasianodon hypophthalmus* and other Siluriformes, as well as the Mexican tetra fish *Astyanax mexicanus* (Otomorpha, Charcidae, also known as blind cave fish), possess three polyQ tracts similar to those of hsAR. Amphibia exhibit the LQ motif with a nearby SP phosphorylation site. Gymnophiona display short polyQ tracts flanked by SP sites. Lepidosauria have the LQ motif and have acquired a second polyQ tract flanked by two SP sites. Interestingly, Testudines have relatively longer polyQ tracts compared to other Reptilia. The evolutionary branch leading to Crocodilians and Avia acquired stretches rich in proline residues flanking the SP sites. Monotremata, namely *Ornithorhynchus anatinus* and *Tachyglossus aculeatus*, also feature polyP stretch. However, from Metatheria onwards, AR acquired three polyQ HReps of varying lengths.

hsAR has additional HReps (polyA, polyG, and polyP) in the NTR (Fig. 2; Supplementary Fig. S5 and Supplementary Table S1). The presence of HReps is common in mammalian AR orthologs, but less prevalent in other classes, with some exceptions in specific fish species. For example, *Pristis pectinata* has short polyG and polyE tracts, *D. rerio* and other related fish possess polyS and polyP HReps, while *A. mexicanus* and *P. hypophthalmus* (and other Siluriformes) exhibit polyH and polyE HReps, as well as polyG and polyS, respectively. *Gadus morhua* has a relatively long polyG tract in the LBD. In mammals, Monotremata have long HReps of polyA, polyP, polyE, and polyR but lack polyQ. From Marsupialia onwards, AR acquired polyP, polyA, and polyG HReps, which are present in all Eutheria. Interestingly, in *Meles meles* (Mustelidae) the first polyQ HRep is composed of 47Q, a length that is not tolerated in *H. sapiens* [9].

Conservation and evolution of the ELM mediating AR N/C interactions

To identify potential short linear motifs, we used the ELM database (http://elm.eu.org/) [27]. This analysis identified 356 ELM instances that could potentially affect AR structure and function (Supplementary Table S2). We initially focused on ELMs with known biological function, such as the hsAR ^{23}FxxLF^{27} motif (where F represents phenylalanine, L leucine, and x any amino acid, numbers refer to NP_000035.2) (Fig. 3A), which is required for the N/C interactions [46, 47]. In *H. sapiens*, these N/C interactions also require glycine 21 (G21) [48]. Therefore, we included this residue in our analysis and evaluated the conservation of the extended hsAR ^{21}GxFxxLF^{28} motif in the 536 available sequences of AR (Supplementary Table S1). We identified a potential N/C interaction motif, GxxFxxVF, within the first 30 amino acids in Chondrichthyes and Polypteriformes, Acipenseriformes, and Lepisosteriformes fish groups (Supplementary Fig. S1). This motif evolved to GxxYxxVF in four species of fish within Elopomorpha and Clupeiformes but was absent in Percomorphaceae fish. The GxFxxLF N/C interaction motif is present in Amphibia (Supplementary Fig. S2) and non-avian Reptilia (Supplementary Fig. S3), except in Lacertidae, where it evolved to SxFxxLF. In mammals, the GxFxxLF motif is conserved, whereas in Avia it evolved to GxFxxFF (Supplementary Figs S4--5). The divergence of this motif in Avia with respect to crocodiles indicates that this change occurred in the lineage leading to birds. Despite the variations in the N/C interaction motif across different classes of vertebrates, the motif remains conserved throughout evolution, with amino acid substitutions that have minimal or limited effects on the chemical and physical properties of the lateral chain.

The ELM spanning the LxxLL motif in the LBD is conserved and required for response to androgens

Next, we focused on an ELM located in the LBD, encompassing the ^{860}LxxLL^{864} motif (NP_000035.2). The LxxLL motif serves as a steroid hormone-binding motif in transcription cofactors that interact with SHRs [49] (Supplementary Tables S1 and 2). Mutations of the AR LxxLL motif are associated with complete androgen insensitivity syndrome, but the underlying mechanism is unknown [17, 50–52]. We thus investigated whether this motif is conserved and explored its role on AR function. The LxxLL motif is present in Chondrichthyes but diverged to the LxxLM motif in several Osteichthyes, including Olostei, Elopomorpha, and Otomorpha (Supplementary Fig. S1). In species where one AR isoform contains the LxxLL motif, the other isoform may contain the LxxLM motif. The LxxLL motif remains conserved in Amphibia (Supplementary Fig. S2), with the sole exception of the toad *Pelobates cultripes*, where it diverged to LxxVL. This motif is also found in Lepidosauria and Testudines (Supplementary Fig. S3), but it diverged to LxxVL in crocodiles and birds (Supplementary Figs S3 and 4), suggesting that this divergence occurred in their common ancestors. In Mammalia, the LxxLL motif is conserved (Supplementary Fig. S5), highlighting selective pressure to maintain the original motif across most groups. The remarkable conservation of the LxxLL motif in the LBD of AR prompted us to investigate the function of this ELM. Through site-directed mutagenesis, we substituted the two consecutive leucine residues with alanine (hsAR-LxxAA) in hsAR (Fig. 3A). Additionally, we generated a polyglutamine-expanded mutant AR with 65Q (hsARpolyQ-LxxAA). A unique feature of AR within the family of SHRs is that androgen binding results in protein stabilization [53]. To test whether the loss of the LxxLL motif affects AR stabilization, we expressed hsAR and hsARpolyQ with and without the LxxAA mutation in HEK293T cells and analyzed protein stabilization by western blotting (Fig. 3B). We found that the loss of the LxxLL motif prevents the androgen-dependent stabilization of hsAR and hsARpolyQ. We then analyzed nuclear translocation induced by androgen binding. We expressed the AR mutants in COS1 cells, and we analyzed AR subcellular localization by immunofluorescence analysis (Fig. 3C). In the absence of ligand, hsAR localized to cytosol, and it translocated to nucleus upon androgen treatment. Also hsAR-LxxAA localized to cytosol, but it failed to translocate to nucleus in the presence of androgens. Similar results were observed with hsARpolyQ (data not shown). Next, we asked whether loss of the LxxLL motif affects AR transactivation by androgens (Fig. 3D). In a transactivation assay in HEK293T cells, we found that the loss of the LxxLL motif suppressed hsAR and hsARpolyQ transactivation induced by ligand. Based on these observations, we examined whether the loss of the LxxLL motif affects androgen binding (Fig. 3E). We expressed hsAR and hsARpolyQ in HEK293T cells, we treated the cells with radioactive ligand, and we measured specific binding. We found that the LxxAA mutation nearly

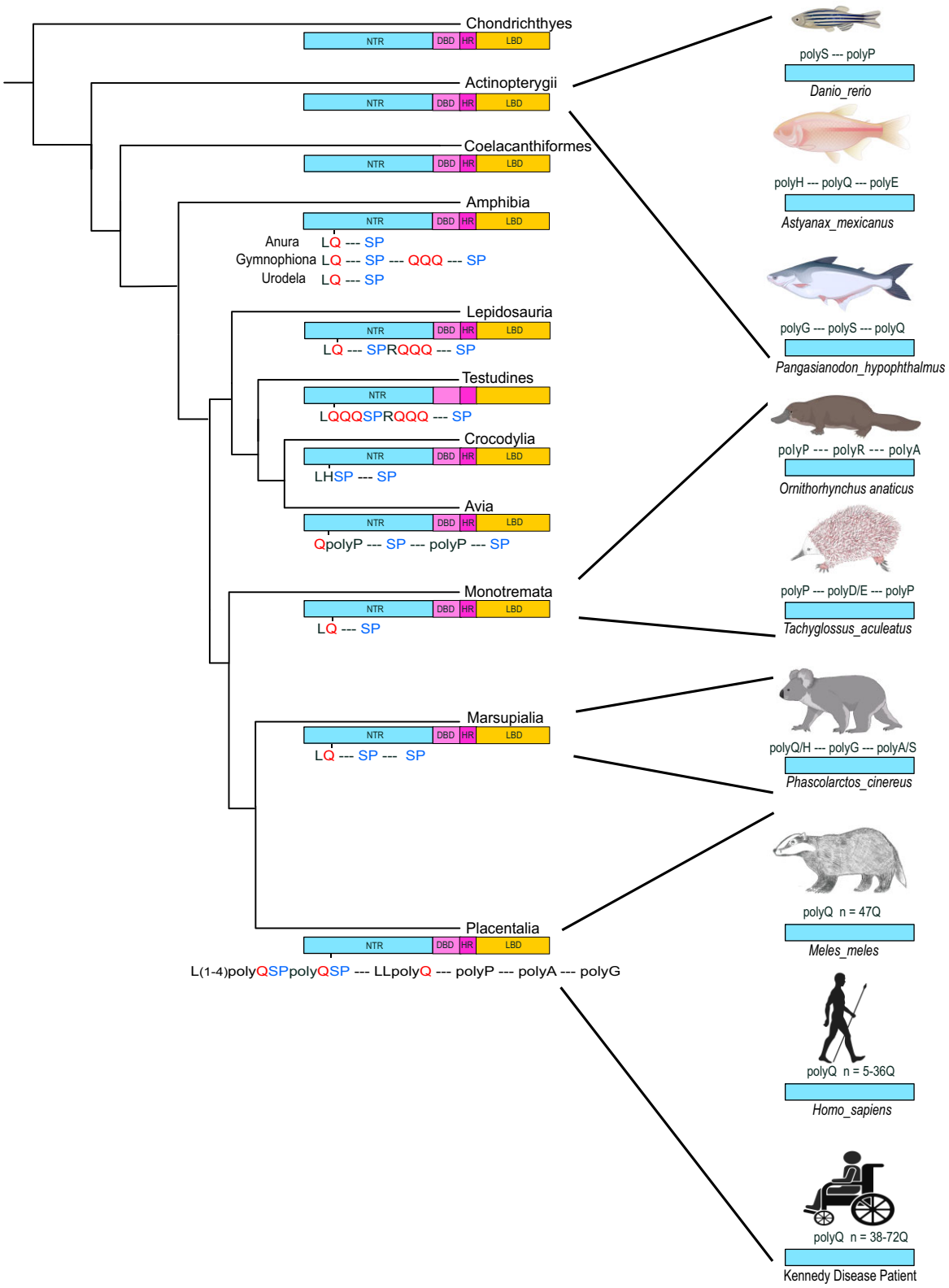


Figure 2. Evolution of AR heterogeneous HReps. Cladogram showing the phylogenetic relationships of the vertebrate groups according to [79] and the evolution of HReps in the *AR* orthologs. Putative SP phosphorylation sites corresponding to S83 and S96 of *H. sapiens* AR are indicated. *M. meles* has a polyQ tract composed of 47Q (first polyQ), while *H. sapiens* tolerates a maximum of 36Q, and expansions over 38Q cause Kennedy's disease. "Created in BioRender. Falconieri, A. (2025) https://BioRender.com/m10f120'' or handmade.

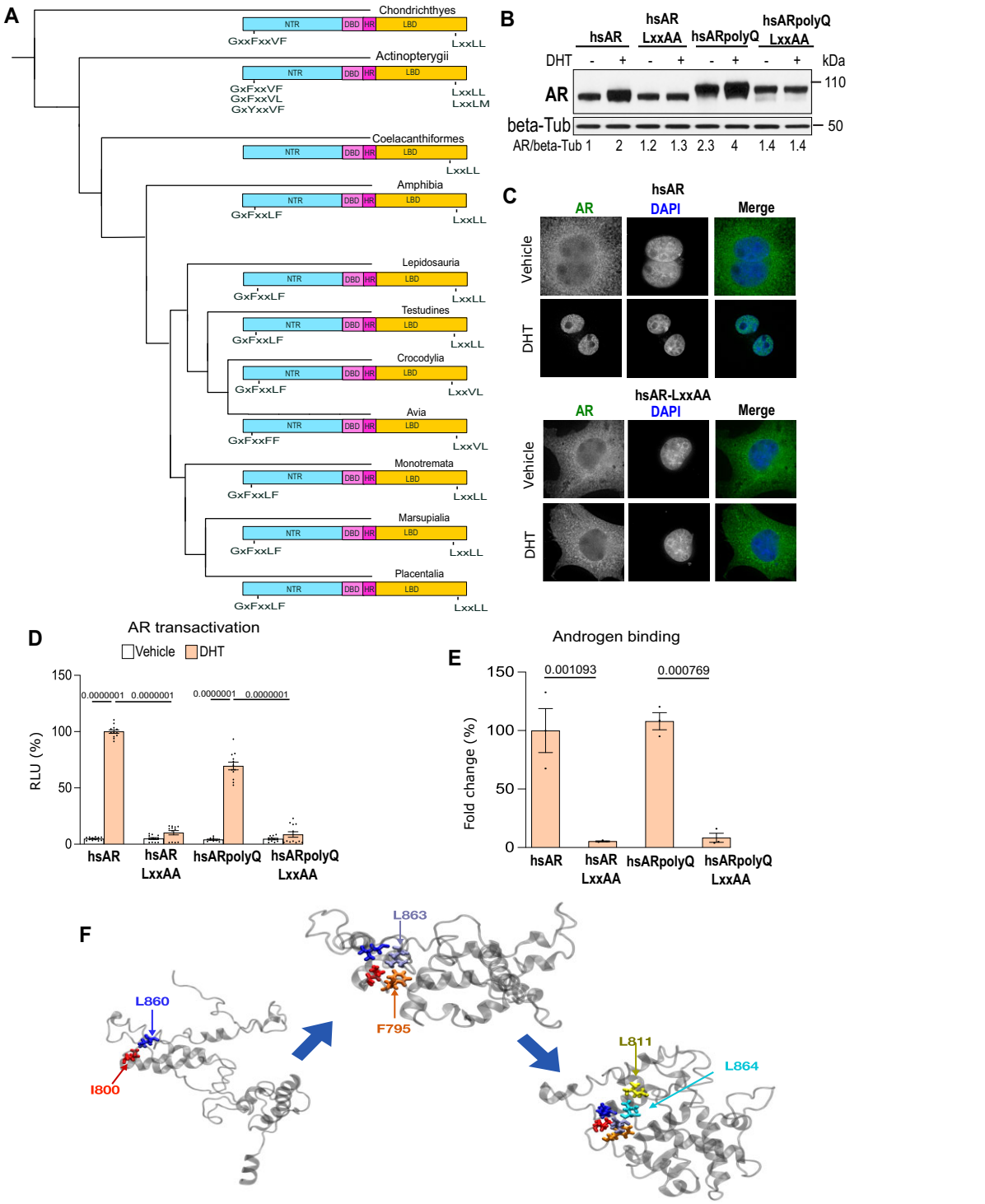


Figure 3. The ELM spanning the LxxLL motif is required for androgen binding. (**A**) Cladogram showing the phylogenetic evolution and conservation of the AR GxFxxLF and LxxLL motifs. (**B**) Western blot analysis of hsAR (with 24 Q residues) and ARpolyQ (with 65 Q residues) with and without LxxLL mutated to LxxAA ectopically expressed in HEK293T cells treated with vehicle (EtOH) and DHT (10 nM, 20 h). AR was detected with a specific antibody and β-tubulin (β-Tub) was used as the loading control. Quantification is shown at the bottom ($n = 2$ biological replicates). (**C**) Immunofluorescence analysis of the subcellular localization of AR and AR-LxxAA ectopically expressed in COS1 cells treated with vehicle and DHT (10 nM, 20 h). AR (green) was detected with a specific antibody, and nuclei (blue) were detected with DAPI (4′,6-diamidino-2-phenylindole). (**D**) Transactivation assay measuring the activity of AR and ARpolyQ with and without the LxxAA mutation ectopically expressed in HEK293T cells treated with vehicle and DHT (0.5 nM, 20 h) ($n = 12$ biological replicates). Statistical analysis: two-way ANOVA followed by Bonferroni HSD tests. (**E**) Ligand binding assay measuring ^{3}[H]-R1881 specific binding of AR and ARpolyQ with and without the LxxAA mutation ectopically expressed in HEK293T cells ($n = 3$ biological replicates). The specific binding of normal AR was set to 100% and the binding of the indicated AR mutants was normalized to that of AR. Statistical analysis: two-way ANOVA followed by Bonferroni HSD tests. (**F**) Our numerical simulations suggest that a contact between L860 and I800 occurs early in the folding process and represents a key step in the formation of the hydrophobic core (leftmost panel). This event is followed by the formation of a contact between L863 and F795 (central panel). Conversely, L864 acquires its native conformation and interactions at a later stage of the folding pathway when the core is already formed (rightmost panel).

abolishes androgen binding, accounting for the loss of androgen responsiveness.

To establish whether the loss of the LxxLL motif impinges on folding to prevent androgen binding, we applied MD (Fig. 3F). In our folding simulations, the side chains of L860 and L863 consistently assumes an inward conformation at a very early stage in the folding process, soon after the formation of secondary structures. This enables the formation of a first contact between L860 with isoleucine 800 (I800), followed by that of L863 with phenylalanine 795 (F795). These two contacts, which are consistently found to occur in order, arguably play a key role in nucleating the formation of the hydrophobic core. Conversely, L864 remains unstructured even when the hydrophobic core has already formed. The formation of a contact with L811, increasing the thermodynamic stability of the core, is typically one of the interactions along the folding process. These observations shed light onto the biological constraints of a key conserved ELM in the AR LBD necessary for ligand binding.

An ELM regulating androgen binding via a cryptic phosphorylation site

Next, we examined the conservation of ELMs that are dynamically regulated by PTMs [19] (Supplementary Tables S1 and 2). We focused on the ELM spanning serine 792 (S792) with the consensus sequence ^{787}RxRxxSxxF^{795}, which forms contacts with the LxxLL motif (Fig. 3F) and functions as a phospho-degron signal [29, 54]. This ELM represents an important site that undergoes two mutually exclusive PTMs: serine phosphorylation [29, 54] and arginine methylation [55]. Although phosphorylation suppresses transactivation, arginine methylation has the opposite effect, highlighting the dynamic role of this ELM in the regulation of AR activity. The crystal structure shows that S792 is buried in the LBD core in the ligand-bound state (Fig. 4A and B; Supplementary Table S3A and Supplementary Fig. S6). We have recently found that many proteins have cryptic phospho-sites buried inside the core of folded domains and are phosphorylated along the folding pathway [56]. Thus, we hypothesized that S792 phosphorylation occurs in a metastable state along the AR folding pathway. To test this hypothesis, we modeled the hsAR folding pathway using a computational scheme based on rMD [57, 58]. This method achieves an exponential reduction in the computational cost of generating protein folding pathways by introducing a biasing force. This force only switches on when the chain attempts to backtrack to an unfolded configuration, whereas no artificial bias is applied when it spontaneously progresses towards the native state, which needs to be provided in input. In our simulations, we used the LBD of hsAR (PDB 3B68, residues 671–917) as the target structure, excluding the ligand atoms. The artificial biasing force in the rMD can push the system out of equilibrium and introduce artifacts. To keep these systematic errors to a minimum, we resorted to the Bias Functional (BF) scheme [31] detailed in "Materials and methods" section, which enables us to identify among the set of generated folding pathways those that are minimally affected by the bias, i.e. those that have the largest probability to occur in the absence of the unphysical forces. To reconstruct the folding mechanism, we first generated three unfolded conformations by running short plain MD simulations at high temperature (800 K) and then created 20 folding trajectories per conformation using rMD. From the resulting ensemble (60 trajectories), we generated a probability distribution plot (Fig. 4C). Applying the BF scoring function, we ranked all reactive trajectories and selected the most realistic prediction—that is, the least biased prediction (LB)—for each set. We then sampled conformations from each trajectory that lay within the observed energy wells. This analysis identified three sequential non-native metastable states, each represented by three conformations: (i) a partially unfolded state (U) characterized by α-helical structures with minimal native contacts; (ii) an intermediate folding state (I1) partially similar to the native conformation; and (iii) a second intermediate folding state (I2) with a packing degree similar to the native state (Fig. 4D). S792 is exposed only in the U state in all conformations, while it remains buried in the I1 and I2 states (Fig. 4E). Interestingly, in the I1 state, the ligand-binding pocket is already structured and displays favorable druggability descriptors, as this site is not yet obstructed by helix-12 (Fig. 4F and Supplementary Table S3B). These data align with the idea that S792 phosphorylation during folding interferes with maturation towards a conformation capable of ligand binding, shifting the equilibrium towards inactivation and degradation. However, S792 becomes a cryptic phosphorylation site in the I1 and I2 states.

The ELM spanning S792 is an old and conserved suppressive signal from fish to *H. sapiens*

Next, we explored the evolutionary conservation of the ELM spanning hsAR S792 (Fig. 5A and Supplementary Table S1). Sequence alignment of the LBD of the AR orthologs revealed that Chondrichthyes possess the ExQxxSxxF motif, while early Actinopterygii species such as *Erpetoichthys calabricus*, *Polypterus senegalus*, *Acipenser ruthenus*, *Lepisosteus oculatus*, and *Amia calva* have the [R/A]x[K/Q]xxAxxF motif, indicating that in these species the serine corresponding to hsAR S792 is a nonphosphorylatable amino acid (Supplementary Fig. S1). Teleostei display the RxRxxSxxF motif, with absolute conservation of the serine. The motif diverged to QxKxxSxxF in Otomorpha, with an example being *D. rerio*. In certain species, such as *Scophthalmus maximus* (order Pleuronectiformes, family Scophthalmidae) and *N. furzeri*, the sequences of both AR onhologs are available, and notably, isoform 1 contains the RxRxxSxxF motif, while isoform 2 has the RxRxxAxxF motif, indicating that phosphorylatable and nonphosphorylatable AR variants can co-exist within the same species. *Latimeria chalumnae* has the RxRxxSxxL motif, while all tetrapods (Amphibia and Amniota) show conservation of the RxRxxSxxF motif in the LBD of AR (Supplementary Figs S2–5), except for toads in the genera *Bombina* and *Spea*, which possess the nonphosphorylatable RxRxxAxxF motif (Supplementary Fig. S2). SWISS-Model and Alphafold3 predictions for the structures of nfAR isoforms 1 and 2 yielded highly similar structures, suggesting that substitution of serine with alanine does not affect folding upon binding (Fig. 5B). These phylogenetic and structural analyses are consistent with our earlier findings that substituting S792 with alanine in hsAR does not alter androgen binding [29]. On the contrary, substituting S792 with aspartate blocks androgen binding, thus preventing any androgen response. To determine whether the cryptic phosphorylation site functions as a suppressive signal already established in fish, we generated phospho-defective

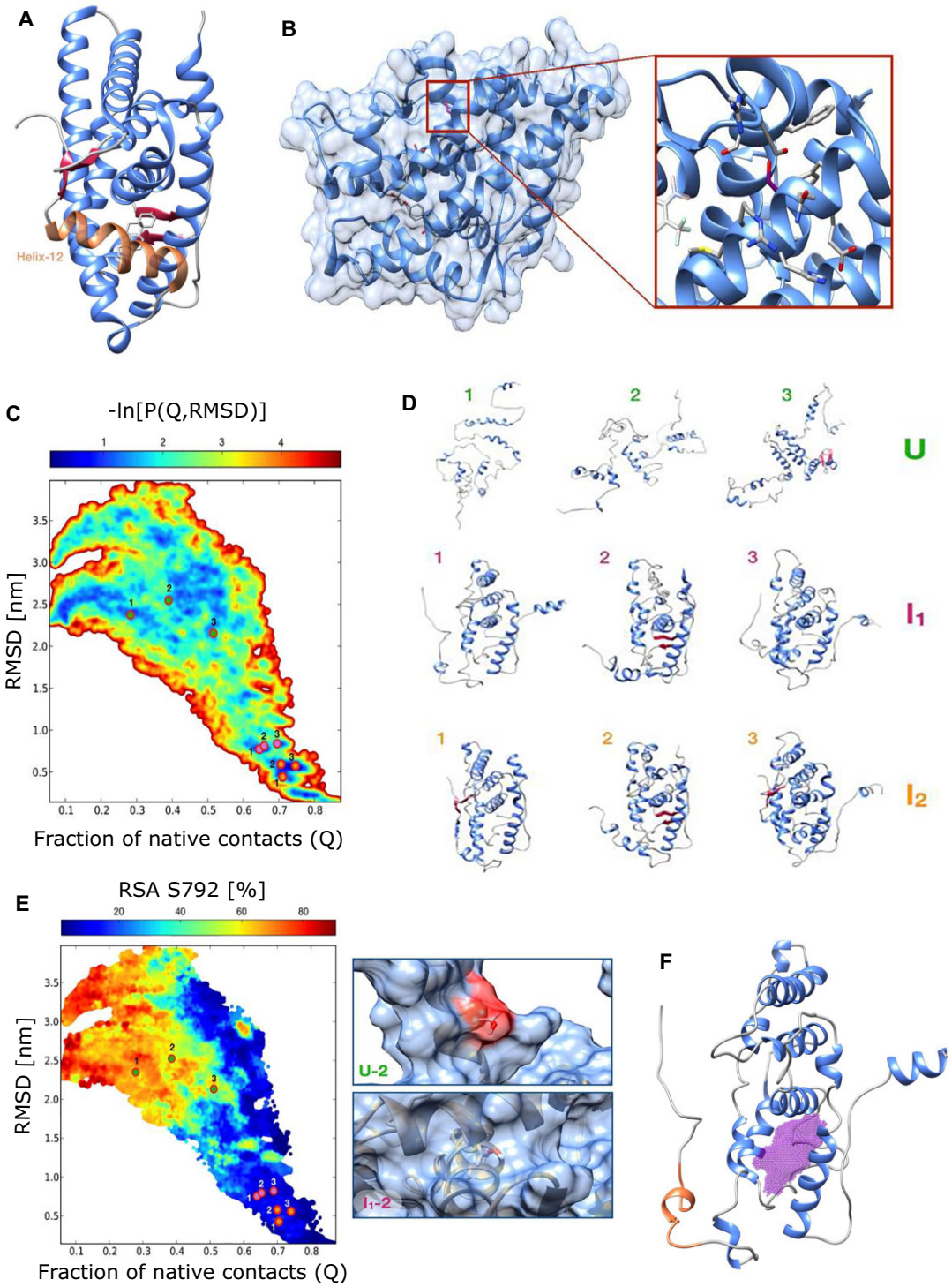


Figure 4. The ELM spanning S792 is a cryptic phosphorylation site. (**A**) Ribbon diagram of the native structure of the hsAR LBD in complex with DHT (PDB 2AMA). The picture shows the topology of helix 12 (depicted in orange) covering the ligand-binding pocket. (**B**) Ribbon diagram of the hsAR LBD in complex with the SARM-S4 agonist (PDB 3B68). The box shows the position of S792 inside the AR structure. S792 is barely visible because deeply buried in the hydrophobic core. (**C**) Reconstruction of the all-atom hsAR LBD folding pathway. Lower-bound approximation of the free energy landscape of hsAR folding obtained from 60 rMD trajectories, plotted as a negative logarithm of the probability distribution expressed as a function of the collective variables Q (fraction of native contacts) and RMSD. Namely, $G(Q, RMSD) = -kBT \ln[P(Q, RMSD)]$, where P(Q, RMSD) is the probability estimated from the frequency histogram in the Q versus RMSD plane. (**D**) Representative hsAR structures of non-native (RMSD > 4Å) conformations U, I1, and I2, each belonging to different free-energy wells, visited by three least biased trajectories. The borders of the relevant regions U, I1, and I2 are arbitrarily defined as the set of points in the Q versus RMSD plane with a free energy distance $\Delta G = 2.5\ kBT$ from the local nearby free energy minimum. (**E**) RSA of residue S792 along the hsAR folding pathway expressed as a function of Q and RMSD. Images on the right show the solvent-exposed surface of S792 (depicted in red) in the U-2 (upper panel) or I1-2 (lower panel) conformations, taken as representative examples of solvent accessibility or nonaccessibility of the residue in the U and I1 states, respectively. (**F**) Druggable pocket in the hsAR intermediate state (I1): Ribbon diagram of hsAR intermediate I1 with highlighted ligand-binding pocket (purple dots). Such a site is consistently detected as druggable by two different software, SiteMap and DogSiteScorer. The pocket, surrounded by helices 3 and 5, is accessible before the closure by helix 12 (depicted in orange).

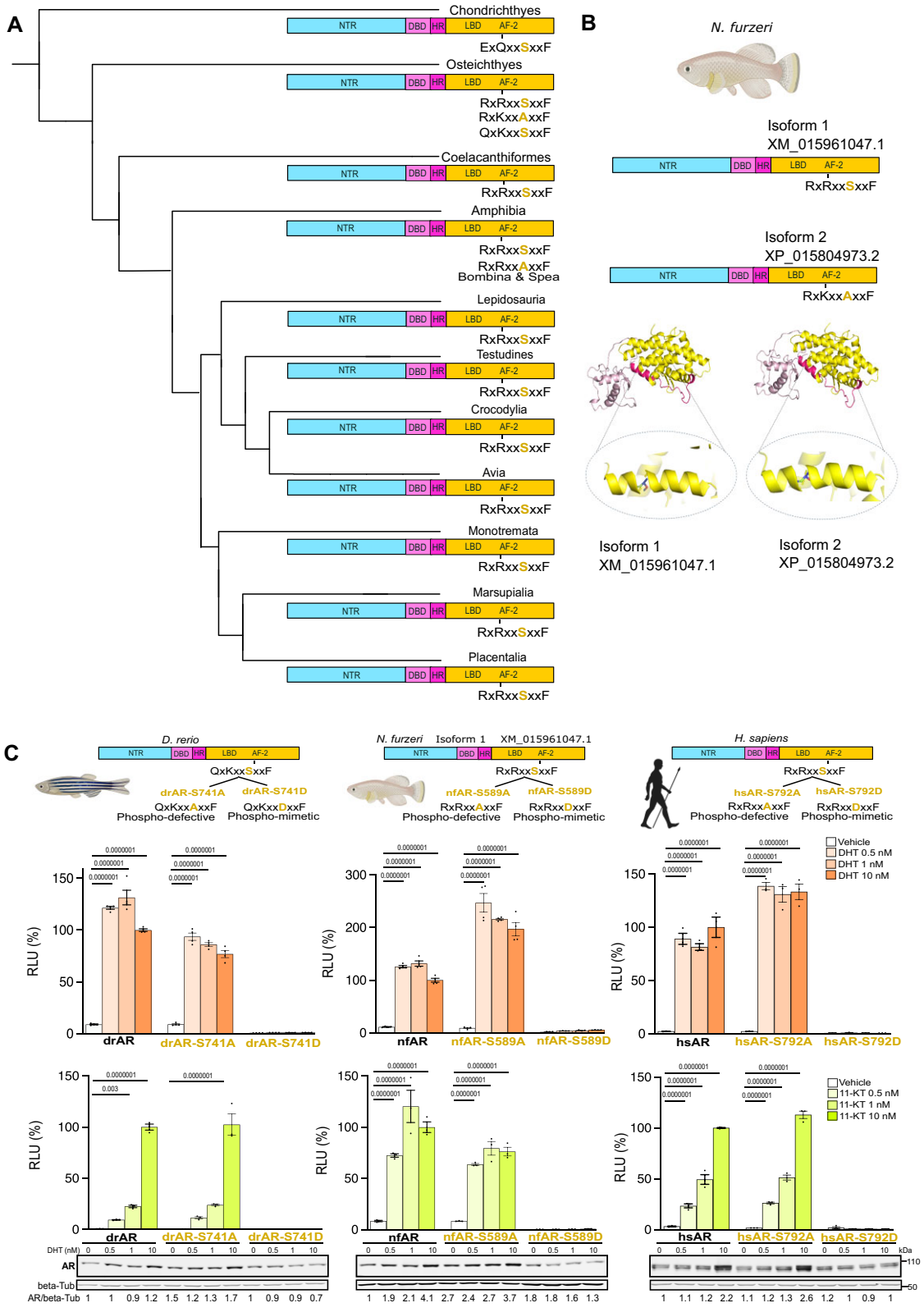


Figure 5. A cryptic phosphorylation site in the LBD is an old signal that suppresses AR activity from fish to mammals. (**A**) Cladogram showing the phylogenetic evolution and conservation of the ELM in the LBD spanning the RxRxxSxxF motif. (**B**) SWISS-model structure prediction of nfAR isoforms 1 with phosphorylatable serine (bearing the RxRxxSxxF motif) and isoform 2 with phospho-defective alanine (bearing the RxKxxAxxF motif). (**C**) Top panels: Experimental design illustrating the position of the serine residue with the ELM spanning the RxRxxSxxF motif. Experimental approach showing the generation of phospho-defective and phospho-mimetic drAR, nfAR, and hsAR variants. Middle panels: Transactivation assays measuring the activity of hsAR, drAR, and nfAR ectopically expressed in HEK293T cells treated with vehicle, 11-KT, and DHT (20 h) as indicated ($n = 3$ biological replicates). Graphs show mean ± SEM; two-way ANOVA followed by Bonferroni HSD tests. Bottom panels: Western blot analysis of Flag-tagged hsAR, drAR, and nfAR ectopically expressed in HEK293T cells treated with vehicle (EtOH) and DHT and 11-KT (20 h). Quantification is shown at the bottom ($n = 3$ biological replicates). hsAR was detected with a specific antibody, and drAR and nfAR were detected with an anti-Flag antibody; β-tubulin (β-Tub) was used as a loading control. "Created in BioRender. Falconieri, A. (2025) https://BioRender.com/m10f120''.

(alanine) and phospho-mimetic (aspartate) variants of fish AR (Fig. 5C), similar to the approach previously used for hsAR [29]. Using *D. rerio* as a representative species with the QxKxxSxxF motif (drAR-S741A and drAR-S741D) and *N. furzeri* as a representative with the RxRxxSxxF motif (nfAR-S589A and nfAR-S589D), we first verified the expression of these AR variants and then assessed their transactivation in androgen-treated HEK293T cells. Substituting S792 in hsAR and S589 in nfAR with alanine enhanced AR transactivation upon DHT, but not 11-KT treatment, an effect not observed with drAR. On the other hand, replacing serine with aspartate in hsAR, drAR, and nfAR suppressed AR transactivation in response to DHT and 11-KT treatment. These results suggest that the cryptic phosphorylation site in the LBD is an ancient suppressive signal conserved from fish to humans.

Specific ELMs present in AR are shared among SHRs and their ancestors

Finally, we explored the presence of specific ELMs in the other members of the SHR family and their ancestors. The SHRs evolved from an ancestral receptor (AncSR1), which diverged into ERα, ERβ, and a second ancestral receptor (AncSR2) through gene duplication (Fig. 6A). AncSR2 later gave rise to the PR and the ancestral receptor from which both the GR and MR evolved. AR likely originated from PR. The LBD sequences of AncSR1 and AncSR2 were reconstructed based on the phylogenetic relationships of modern SHRs [35, 59]. The FxxLF motif is specific to AR, as previously reported [60], whereas the LxxLL motif is present in all receptors, including the predicted ancestor receptors. These observations indicate that the degree of conservation of specific ELMs matches the essential functions that are an absolute prerequisite for folding upon binding, i.e. the ELM spanning the LxxLL motif, and SHR-specific ELMs, i.e. the ELM spanning the FxxLF motif.

Next, we asked whether the RxRxxSxxF motif or a similar motif is present in the other SHRs (Fig. 6A). We found that the arginine residues are exclusive to hsAR, while the serine residue is present in AncSR2, GR, MR, and ERα, and the phenylalanine residue is conserved in all receptors. AncSR1 and ERβ contain a threonine residue, which can be phosphorylated by serine/threonine kinases. PR is the only SHR with a nonphosphorylatable residue at the position corresponding to hsAR S792. To address whether the ELM spanning the putative phospho-site is a suppressive signal in other SHRs, we generated phospho-defective and phospho-mimetic ERα by substituting S432 with alanine (ERα-S432A) and aspartate (ERα-S432D), respectively (Fig. 6B). By western blotting, we found that phospho-defective substitution increased the steady-state levels of ERα, whereas phospho-mimetic substitution had the opposite effect. In a transcriptional assay using the luciferase under the control of an estrogen-responsive element (ERE-luc), we found that phospho-defective substitution of S432 does not modify ERα transactivation, whereas phospho-mimetic substitution significantly reduces ERα transactivation. This finding suggests that the ELM in the LBD of SHRs is an old and shared suppressive signal instated at the origin of evolutions of this family of essential TFs.

Discussion

Our phylogenetic analysis of 536 species, combined with computational, molecular, and biochemical studies, demonstrates that AR structure and function are highly conserved across vertebrates. We show that HReps are mostly, but not exclusively, present in mammals. We identified ELMs that are exclusive to AR and others shared by the other members of the SHR family, suggesting that they originated in the ancestral receptors that gave rise to modern SHRs. Furthermore, we found that a cryptic phosphorylation site in the LBD serves as a suppressive signal from fish to human AR and other SHRs. These findings highlight the value of interdisciplinary approaches to identify functionally active ELMs in essential genes.

The intrinsically disordered NTR of AR is essential for full receptor transactivation. Intrinsically disordered regions are composed of low-complexity sequences enriched in charged and polar amino acids, and HReps likely play important roles in cellular function [39, 40]. HReps represent a particularly interesting category of extremely low-complexity sequences ranging from a few to several dozen uninterrupted stretches of the same amino acid. For example, hsAR contains three polyQ tracts, along with polyA, polyG, and polyP HReps. Although it is well established that the number, position and length of HReps in other proteins are not random [61] and correlates with age-linked diseases [42], the relationship between the multiple polymorphic HReps present in the NTR of AR is not known. An emerging function of HReps is their involvement in cellular adaptation to environmental stressors [62], allowing organisms to rapidly adjust while evolving inheritable traits [63]. Proteins that influence environmental adaptability are conserved throughout evolution, often have extensive intrinsically disordered regions, and likely bind to nucleic acids [63]. In particular, the transgenerational inheritance of advantageous self-templating conformations in budding yeast is altered when HSP90 and HSP70 are inhibited. These chaperones are part of the foldosome complex that facilitates the stepwise maturation of AR from a disordered or unfolded state into a partially folded and transcriptionally active state [64]. Therefore, AR is a candidate to influence sex hormone-driven traits, which is particularly relevant for animals exposed to dramatic environmental changes, such as aquatic species. Teleost fish have explored various evolutionary strategies in reproductive systems to adapt to diverse environments, yet the gonadal system remains well conserved in these fish. Testes and ovaries originate from a bipotential gonadal primordium that undergoes genetic and environmental sex determination [65]. Genetic sex determination can be single gene, as in *Oryzias latipes* (Japanese medaka), or polygenic, as in *D. rerio*. *O. latipes* is a gonochoristic fish, with fish having XX chromosomes developing ovaries, and fish having XY chromosomes developing testes. Although environmental factors affecting sex determination, differentiation, or maintenance are more common in reptiles (e.g. crocodiles and turtles), such factors can influence the development of *D. rerio*, with elevated temperature, acidic pH, and hypoxia leading to male development. *Tachysurus fulvidraco* shows changes in gene expression in response to hypoxia [66], and masculinization of the XX genotype in response to a +3°C increase in water temperature [67]. Interestingly, the AR orthologs of these catfish contain polyQ stretches along with other HReps, which might contribute to their response and adaptation to stress

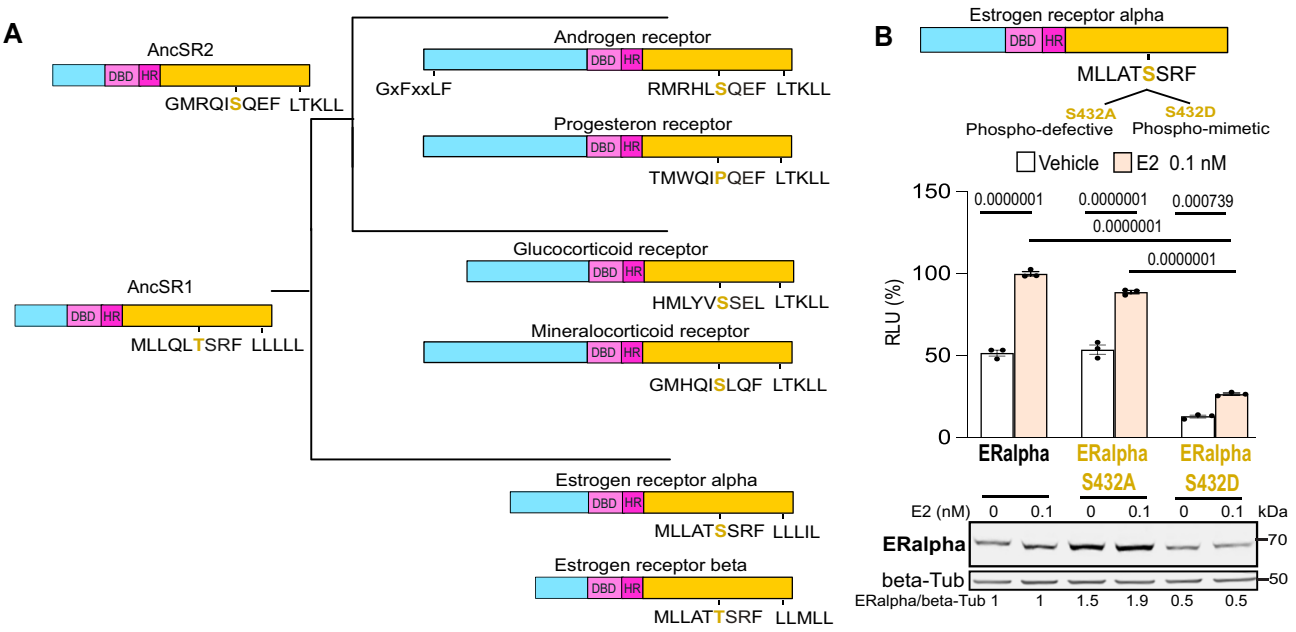


Figure 6. Specific ELMs are shared in the family of SHRs. (**A**) Cladogram showing the phylogenetic evolution and conservation of the six members of H. sapiens SHRs. The prediction of the sequence of ancestor 1 (AncSR1) and the derived ancestor (AncSR2) [59] revealed the presence of a phosphorylatable amino acid (S or T). The FxxLF N/C-interaction motif is present only in AR, whereas the LxxLL motif is present in all receptors, including the ancestors. Sequence alignment identifies a putative cryptic phosphorylation site in the LBD of all SHRs, except PR. (**B**) Top panel: Transactivation assay measuring the activity of ERalpha (ERα) phospho-variants expressed in HEK293T cells treated with vehicle and β-estradiol (E2, 20h) as indicated ($n = 3$ biological replicates). Bottom panel: Western blot analysis of Flag-tagged ERα phospho-variants expressed in HEK293T cells treated with vehicle (EtOH) and 17-β-estradiol (E2, 20 h). Quantification is shown at the bottom ($n = 3$ biological replicates). Graph shows mean ± SEM; two-way ANOVA followed by Bonferroni HSD tests.

conditions. Similarly, the *A. mexicanus* AR ortholog possesses three polyQ HReps, a polyH, and a polyE/D stretch. *A. mexicanus* is particularly intriguing in terms of adaptation to stressful conditions, as it has adapted to living in river surface waters or caves. To cope with long periods of starvation of nutrients, loss of sunlight, and temperature fluctuations different from those experienced by surface-dwelling fish, cave-dwelling populations underwent organ morphogenetic changes while maintaining interfertility with river-dwelling populations. Future research will determine the functional role of HReps in *A. mexicanus*, whether they contribute to environmental adaptation, and why they became a feature of the NTR of AR in mammals.

AR and other SHRs form homodimers upon activation by ligands. AR is unique among nuclear receptors because it forms N/C interactions via the FxxLF motif [12, 48], which trigger a conformational change in the AF-2 surface of the LBD [13, 68]. These N/C interactions initially intramolecularly and then intermolecularly stabilize AR in the presence of the ligand. The N/C interaction motif is conserved from cartilaginous fish to humans, indicating its importance for AR function, and the loss of this motif prevents the full activation by androgens. While the GxFxxLF motif is present only in AR, analysis of all SHRs and their ancestral forms indicates that all receptors possess the LxxLL motif. L860 is located in helix 10 of the LBD, and mutation to phenylalanine (L860F) was found in a case of androgen insensitivity syndrome and was shown to almost abolish androgen binding [50]. Mutation of leucine 864 to arginine (L864R) was also associated with complete androgen insensitivity syndrome [51], but the effect of loss of L864 was not tested at that time. This clinical and experimental evidence, together with our computational and experimental data, supports the concept that the LxxLL motif is absolutely required for androgen binding and suggests that the motif is necessary for folding upon binding in the entire family of SHRs. Of notice, mutations of threonine 861 to asparagine (T861N) and isoleucine (T861I) cause complete androgen insensitivity syndrome [17, 52]. A threonine residue is present also in the LxxLL motif of PR, GR, and MR, and thus likely in AncSR2, indicating that also the amino acids in between the leucine residues have a role in AR folding.

Phosphorylation is the most abundant PTM of AR. hsAR is phosphorylated at multiple serine residues through various signaling pathways. S792 is the only phosphorylation site known to inhibit androgen binding, indicating that its phosphorylation must be precisely controlled [29]. To further investigate this, we turned to physics-based MD simulations. Most biologically relevant conformational transitions are too complex or rare to be captured by standard MD, even resorting to the most powerful supercomputers. In some cases, however, the computational load of simulating biomolecular transitions can be significantly reduced using enhanced sampling techniques. These schemes introduce additional approximations and leverage the available experimental data. In particular, the BF approach is an enhanced sampling scheme suitable for simulating protein folding and requires input of the three-dimensional native structure (target structure). This method has been thoroughly validated against both standard MD simulations [31, 69, 70] and biophysical experiments [71–73]. Recently, this approach was applied to identify small molecules that inhibit target protein folding, thus inducing their degradation by the cell quality control machinery [74]. Phospho-mimetic substitution of S792 with

aspartate blocks androgen-induced AR stabilization, nuclear translocation and transactivation [54, 75], and it drives AR to degradation by the ubiquitin-proteasome system [76, 77]. Our computational analyses of the AR folding pathway revealed complex folding kinetics, including an intermediate state that exposes S792 to the solvent. Our results suggest that S792 is a cryptic regulatory site at the crossroads between a non-phosphorylated ligand-binding folding pathway of AR, and a phosphorylated, nonviable pathway leading to AR degradation [75, 76]. In its natively folded ligand-bound state, this residue is buried within the protein core, implying that phosphorylation likely occurs before AR binds to its ligand. Future research will explore whether this phosphorylation occurs early in the translation process when the nascent polypeptide exits the E site of the ribosome during or immediately after synthesis.

The importance of the S792 phosphorylation site stems from its role in regulating the fate of the AR. This is why we were surprised to discover that certain fish species express both versions of the motif (RxRxxSxxF and RxRxxAxxF). A Teleost-specific genomic duplication event led to the creation of onholog genes experiencing less selective pressure to maintain function [78]. However, in *Oryzias latipes* (Japanese Medaka), both AR ohnologs, ARalpha and ARbeta, are active and essential for reproduction [38], and notably they carry the RxRxxSxxF and RxRxxAxxF motifs, respectively. Another intriguing finding was the discovery that primitive bony fish, including the living fossil *Amia calva*, and the anuran genera *Bombina* and *Spea* carry the RxRxxAxxF motif. *Bombina* and *Spea* are phylogenetically distant and exhibit considerable ecological differences. *Bombina* species, found in forested regions of central Europe or Asia, depending on the specific species, are characterized by extended breeding periods during which they remain in or near water, hibernate during the winter, and reach sexual maturity within two to four years. *Spea* is a more terrestrial genus that inhabits the arid regions of North America and spend much of their time buried underground, emerging to breed explosively in temporary ponds after heavy summer rains. Neither genus exhibits pronounced sexual size dimorphism. Overall, the significant differences between *Bombina* and *Spea* suggest that they do not share distinctive traits that would group them separately from other anurans, and thus the occurrence of alanine replacing serine is an example of evolutionary convergence. Given the profound impact of phosphorylation at this site on AR biology, it is unlikely that the shift from RxRxxSxxF to RxRxxAxxF is merely a random mutation, for several reasons: (i) Alanine is the only amino acid found to replace this serine in vertebrates; (ii) Alanine shares chemical and physical properties similar to those of serine but cannot be phosphorylated; and (iii) Functionally, AR in these anurans cannot be controlled by this suppressive PTM in its ligand-free state. Alanine substitution probably arose through natural selection. The evolutionary forces driving this amino acid change are still unknown, and further research may reveal whether the RxRxxAxxF motif in the AR of these anurans played a role in their adaptation to extreme environments. Analysis of all SHRs and their ancestral forms revealed that the RxRxxSxxF motif is present only in AR. Nonetheless, the other SHRs, except PR, have a serine or threonine residue aligning with S792. Importantly, at least for ERα, such serine residue works as a suppressive signal, suggesting a common and old mechanism for inactivation of SHRs.

One limitation of this study is the limited number of AR sequences available in vertebrate groups. Expanding the dataset to include a broader range of sequences from various species could provide a more comprehensive understanding of the evolutionary dynamics of AR. An expanded dataset could be analyzed in conjunction with environmental factors, sexual dimorphism, and different modes of adaptation, providing deeper insights into how these elements influence the evolution of AR. Such an approach could reveal patterns of evolutionary pressure and adaptation, particularly in relation to AR functions in different ecological contexts and life histories. Ultimately, this could enhance our understanding of the evolutionary importance of specific amino acid substitutions and their impact on the biological roles of AR in diverse vertebrate lineages.

Acknowledgements

We are grateful to Andrea Astolfi (University of Perugia, Italy) for analysis of pocket druggability. We thank Morena Simonato for key technical support. Where indicated, pictures were created in BioRender. Falconieri, A. (2025) https://BioRender.com/42153z8.

This paper is dedicated to the memory of our beloved Laura Feltri.

Author contribution: A.F., G.B., C.B., P.F. (Conceptualization),A.G. (Investigation and Data curation) E.B., G.S., and P.F. (Data curation, formal analysis and methodology). I.G.M. & H.C.L (Data curation and Resources) A.F., G.B., I.P. and C.B (Investigation and Formal analysis). A.K.O.H. (Formal analysis). M.B., L.D.V., C.M, E.Z., L.T., K.F., and J.R. (Formal analysis). M.P. (Conceptualization, Funding acquisition, Supervision, writing-original draft, review & editing). All authors approved the final version of the manuscript.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

M.P. received support from Arvinas and Autifony. G.S., P.F., and E.B. are co-founders and shareholders of Sibylla Biotech S.p.A. G.S. is currently an employee of Sibylla. The other authors have no conflict of interest to declare.

Funding

This work was supported by Association Française contre les Myopathies (AFM-24337 to M.P. and M.B.), Kennedy's Disease Association (to E.Z. and M.B.), Fondazione Telethon (GMR24T1059 to M.P.), Fondazione AIRC-Italy (24423 to M.P.), Muscular Dystrophy Association (MDA 1288536 https://doi.org/10.55762/MDA.1288536.pc.gr.197898 to M.P.), PRIN-MIUR Italy (2022EFLFL8 to M.P.), Program "Rare diseases" CNCCS Scarl Pomezia (to M.P.). Figures were created with InkScape (https://inkscape.org/) and Biorender (Licence number UJ279AID40, https://www.biorender.com/). Funding to pay the Open Access publication charges for this article was provided by Fondazione Telethon.

Data availability

The data related to this article are available in the article and the Supplementary material.

References

1. Gronemeyer H, Gustafsson J-Å, Laudet V. Principles for modulation of the nuclear receptor superfamily. *Nat Rev Drug Discov* 2004;**3**:950–64. https://doi.org/10.1038/nrd1551
2. Baker ME. Steroid receptors and vertebrate evolution. *Mol Cell Endocrinol* 2019;**496**:110526. https://doi.org/10.1016/j.mce.2019.110526
3. Ogino Y, Katoh H, Kuraku S *et al.* Evolutionary history and functional characterization of androgen receptor genes in jawed vertebrates. *Endocrinology* 2009;**150**:5415–27. https://doi.org/10.1210/en.2009-0523
4. Ryu T, Okamoto K, Ansai S *et al.* Gene duplication of androgen receptor as an evolutionary driving force underlying the diversity of sexual characteristics in teleost fishes. *Zoolog Sci* 2024;**41**:68–76. https://doi.org/10.2108/zs230098
5. Stein RW, Mull CG, Kuhn TS *et al.* Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. *Nat Ecol Evol* 2018;**2**:288–98. https://doi.org/10.1038/s41559-017-0448-4
6. Maruska KP, Gelsleichter J. Hormones and reproduction in chondrichthyan fishes. In *Hormones and Reproduction of Vertebrates*. Elsevier, 2011;**1**:209–37. https://doi.org/10.1016/B978-0-12-375009-9.10011-6
7. Hunter I, Jamieson C, McEwan IJ. The androgen receptor amino-terminal domain: structure, function and therapeutic potential. *Endocr Oncol* 2025;**5**:e240061.
8. Sheikhhassani V, Scalvini B, Ng J *et al.* Topological dynamics of an intrinsically disordered N-terminal domain of the human androgen receptor. *Protein Sci* 2022;**31**:e4334. https://doi.org/10.1002/pro.4334
9. Pennuto M, Rinaldi C. From gene to therapy in spinal and bulbar muscular atrophy: are we there yet? *Mol Cell Endocrinol* 2018;**465**:113–21. https://doi.org/10.1016/j.mce.2017.07.005
10. Nadal M, Prekovic S, Gallastegui N *et al.* Structure of the homodimeric androgen receptor ligand-binding domain. *Nat Commun* 2017;**8**:14388. https://doi.org/10.1038/ncomms14388
11. Sack JS, Kish KF, Wang C *et al.* Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc Natl Acad Sci USA* 2001;**98**:4904–9. https://doi.org/10.1073/pnas.081565498
12. He B, Gampe RT, Kole AJ *et al.* Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Mol Cell* 2004;**16**:425–38. https://doi.org/10.1016/j.molcel.2004.09.036
13. Schaufele F, Carbonell X, Guerbadot M *et al.* The structural basis of androgen receptor activation: intramolecular and intermolecular amino-carboxy interactions. *Proc Natl Acad Sci USA* 2005;**102**:9802–7. https://doi.org/10.1073/pnas.0408819102
14. van Royen ME, van Cappellen WA, de Vos C *et al.* Stepwise androgen receptor dimerization. *J Cell Sci* 2012;**125**:1970–9.
15. Astapova O, Seger C, Hammes SR. Ligand binding prolongs androgen receptor protein half-life by reducing its degradation. *J Endocr Soc* 2021;**5**:bvab035. https://doi.org/10.1210/jendso/bvab035
16. Tyagi RK, Lavrovsky Y, Ahn SC *et al.* Dynamics of intracellular movement and nucleocytoplasmic recycling of the ligand-activated androgen receptor in living cells. *Mol Endocrinol* 2000;**14**:1162–74. https://doi.org/10.1210/mend.14.8.0497
17. Hughes I, Werner R, Bunch T *et al.* Androgen insensitivity syndrome. *Semin Reprod Med* 2012;**30**:432–42.
18. Rebello RJ, Oing C, Knudsen KE *et al.* Prostate cancer. *Nat Rev Dis Primers* 2021;**7**:1–27. https://doi.org/10.1038/s41572-020-00243-0
19. Singh AK, Amar I, Ramadasan H *et al.* Proteins with amino acid repeats constitute a rapidly evolvable and human-specific essentialome. *Cell Rep* 2023;**42**:112811. https://doi.org/10.1016/j.celrep.2023.112811
20. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 2021;**18**:366–8. https://doi.org/10.1038/s41592-021-01101-x
21. Liedtke HC. AmphiNom: an amphibian systematics tool. *Syst Biodivers* 2019;**17**:1–6. https://doi.org/10.1080/14772000.2018.1518935
22. Bouckaert R, Vaughan TG, Barido-Sottani J *et al.* BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2019;**15**:e1006650. https://doi.org/10.1371/journal.pcbi.1006650
23. Stiller J, Feng S, Chowdhury A-A *et al.* Complexity of avian evolution revealed by family-level genomes. *Nature* 2024;**629**:851–60. https://doi.org/10.1038/s41586-024-07323-1
24. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinforma Oxf Engl* 2019;**35**:526–8. https://doi.org/10.1093/bioinformatics/bty633
25. Colston TJ, Kulkarni P, Jetz W *et al.* Phylogenetic and spatial distribution of evolutionary diversification, isolation, and threat in turtles and crocodilians (non-avian archosauromorphs). *BMC Evol Biol* 2020;**20**:81. https://doi.org/10.1186/s12862-020-01642-3
26. Rabosky DL, Chang J, Title PO *et al.* An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* 2018;**559**:392–5. https://doi.org/10.1038/s41586-018-0273-1
27. Kumar M, Gouw M, Michael S *et al.* ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res* 2020;**48**:D296–306.
28. Palazzolo I, Nedelsky NB, Askew CE *et al.* B2 attenuates polyglutamine-expanded androgen receptor toxicity in cell and fly models of spinal and bulbar muscular atrophy. *J Neurosci Res* 2010;**88**:2207–16. https://doi.org/10.1002/jnr.22389
29. Palazzolo I, Burnett BG, Young JE *et al.* Akt blocks ligand binding and protects against expanded polyglutamine androgen receptor toxicity. *Hum Mol Genet* 2007;**16**:1593–603. https://doi.org/10.1093/hmg/ddm109
30. Bartolucci G, Orioli S, Faccioli P. Transition path theory from biased simulations. *J Chem Phys* 2018;**149**:072336. https://doi.org/10.1063/1.5027253
31. Beccara S A, Fant L, Faccioli P. Variational scheme to compute protein reaction pathways using atomistic force fields with explicit solvent. *Phys Rev Lett* 2015;**114**:098103. https://doi.org/10.1103/PhysRevLett.114.098103
32. Andrews G, Fan K, Pratt HE *et al.* Mammalian evolution of human *cis*-regulatory elements and transcription factor binding sites. *Science* 2023;**380**:eabn7930. https://doi.org/10.1126/science.abn7930
33. Berthelot C, Villar D, Horvath JE *et al.* Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat Ecol Evol* 2017;**2**:152–63. https://doi.org/10.1038/s41559-017-0377-2
34. Lindblad-Toh K, Garber M, Zuk O *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 2011;**478**:476–82. https://doi.org/10.1038/nature10530
35. Starr TN, Picton LK, Thornton JW. Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* 2017;**549**:409–13. https://doi.org/10.1038/nature23902
36. Hossain MS, Larsson A, Scherbak N *et al.* Zebrafish androgen receptor: isolation, molecular, and biochemical characterization. *Biol Reprod* 2008;**78**:361–9. https://doi.org/10.1095/biolreprod.107.062018
37. Olsson P-E, Berg AH, von Hofsten J *et al.* Molecular cloning and characterization of a nuclear androgen receptor activated by 11-ketotestosterone. *Reprod Biol Endocrinol* 2005;**3**:37. https://doi.org/10.1186/1477-7827-3-37
38. Ogino Y, Ansai S, Watanabe E *et al.* Evolutionary differentiation of androgen receptor is responsible for sexual characteristic

development in a teleost fish. *Nat Commun* 2023;**14**:1428. https://doi.org/10.1038/s41467-023-37026-6
39. Chavali S, Singh AK, Santhanam B *et al.* Amino acid homorepeats in proteins. *Nat Rev Chem* 2020;**4**:420–34. https://doi.org/10.1038/s41570-020-0204-1
40. Elena-Real CA, Mier P, Sibille N *et al.* Structure–function relationships in protein homorepeats. *Curr Opin Struct Biol* 2023;**83**:102726. https://doi.org/10.1016/j.sbi.2023.102726
41. Iennaco R, Formenti G, Trovesi C *et al.* The evolutionary history of the polyQ tract in huntingtin sheds light on its functional pro-neural activities. *Cell Death Differ* 2022;**29**:293–305. https://doi.org/10.1038/s41418-021-00914-9
42. Meszaros A, Ahmed J, Russo G *et al.* The evolution and polymorphism of mono-amino acid repeats in androgen receptor and their regulatory role in health and disease. *Front Med* 2022;**9**:1019803. https://doi.org/10.3389/fmed.2022.1019803
43. Piol D, Tosatto L, Zuccaro E *et al.* Antagonistic effect of cyclin-dependent kinases and a calcium-dependent phosphatase on polyglutamine-expanded androgen receptor toxic gain of function. *Sci Adv* 2023;**9**:eade1694. https://doi.org/10.1126/sciadv.ade1694
44. Polanco MJ, Parodi S, Piol D *et al.* Adenylyl cyclase activating polypeptide reduces phosphorylation and toxicity of the polyglutamine-expanded androgen receptor in spinobulbar muscular atrophy. *Sci Transl Med* 2016;**8**:370ra181. https://doi.org/10.1126/scitranslmed.aaf9526
45. Escobedo A, Topal B, Kunze MBA *et al.* Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor. *Nat Commun* 2019;**10**:2034. https://doi.org/10.1038/s41467-019-09923-2
46. He B, Bowen NT, Minges JT *et al.* Androgen-induced NH2- and COOH-terminal interaction inhibits p160 coactivator recruitment by activation function 2. *J Biol Chem* 2001;**276**:42293–301. https://doi.org/10.1074/jbc.M107492200
47. Langley E, Zhou ZX, Wilson EM. Evidence for an anti-parallel orientation of the ligand-activated human androgen receptor dimer. *J Biol Chem* 1995;**270**:29983–90. https://doi.org/10.1074/jbc.270.50.29983
48. Callewaert L, Verrijdt G, Christiaens V *et al.* Dual function of an amino-terminal amphipatic helix in androgen receptor-mediated transactivation through specific and nonspecific response elements. *J Biol Chem* 2003;**278**:8212–8. https://doi.org/10.1074/jbc.M210744200
49. Heery DM, Kalkhoven E, Hoare S *et al.* A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 1997;**387**:733–6. https://doi.org/10.1038/42750
50. Rajender S, Singh L, Thangaraj K. L859F mutation in androgen receptor gene results in complete loss of androgen binding to the receptor. *J Androl* 2007;**28**:772–6. https://doi.org/10.2164/jandrol.107.002691
51. Brown TR, Scherer PA, Chang YT *et al.* Molecular genetics of human androgen insensitivity. *Eur J Pediatr* 1993;**152**:S62–9. https://doi.org/10.1007/BF02125442
52. Cheikhelard A, Morel Y, Thibaud E *et al.* Long-term followup and comparison between genotype and phenotype in 29 cases of complete androgen insensitivity syndrome. *J Urol* 2008;**180**:1496–501. https://doi.org/10.1016/j.juro.2008.06.045
53. Zhou ZX, Lane MV, Kemppainen JA *et al.* Specificity of ligand-dependent androgen receptor stabilization: receptor domain interactions influence ligand dissociation and receptor stability. *Mol Endocrinol Baltim Md* 1995;**9**:208–18.
54. Lin HK, Yeh S, Kang HY *et al.* Akt suppresses androgen-induced apoptosis by phosphorylating and inhibiting androgen receptor. *Proc Natl Acad Sci USA* 2001;**98**:7200–5. https://doi.org/10.1073/pnas.121173298
55. Scaramuzzino C, Casci I, Parodi S *et al.* Protein arginine methyltransferase 6 enhances polyglutamine-expanded androgen receptor function and toxicity in spinal and bulbar muscular atrophy. *Neuron* 2015;**85**:88–100. https://doi.org/10.1016/j.neuron.2014.12.031
56. Gasparotto D, Zanon A, Bonaldo V *et al.* 2024; Mapping cryptic phosphorylation sites in the human proteome. bioRxiv, https://doi.org/10.1101/2024.12.03.626562, 27 January 2025, preprint: not peer reviewed.
57. Camilloni C, Broglia RA, Tiana G. Hierarchy of folding and unfolding events of protein G, CI2, and ACBP from explicit-solvent simulations. *J Chem Phys* 2011;**134**:045105. https://doi.org/10.1063/1.3523345
58. Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J Mol Biol* 1999;**288**:441–59. https://doi.org/10.1006/jmbi.1999.2670
59. Thornton JW. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc Natl Acad Sci USA* 2001;**98**:5671–6. https://doi.org/10.1073/pnas.091553298
60. He B, Minges JT, Lee LW *et al.* The FXXLF motif mediates androgen receptor-specific interactions with coregulators. *J Biol Chem* 2002;**277**:10226–35. https://doi.org/10.1074/jbc.M111975200
61. Mier P, Alanis-Lobato G, Andrade-Navarro MA. Context characterization of amino acid homorepeats using evolution, position, and order. *Proteins* 2017;**85**:709–19. https://doi.org/10.1002/prot.25250
62. Chavali S, Chavali PL, Chalancon G *et al.* Constraints and consequences of the emergence of amino acid repeats in eukaryotic proteins. *Nat Struct Mol Biol* 2017;**24**:765–77. https://doi.org/10.1038/nsmb.3441
63. Chakrabortee S, Byers JS, Jones S *et al.* Intrinsically disordered proteins drive emergence and inheritance of biological traits. *Cell* 2016;**167**:369–381.e12. https://doi.org/10.1016/j.cell.2016.09.017
64. Cano LQ, Lavery DN, Bevan CL. Mini-review: foldosome regulation of androgen receptor action in prostate cancer. *Mol Cell Endocrinol* 2013;**369**:52–62. https://doi.org/10.1016/j.mce.2013.01.023
65. Rajendiran P, Jaafar F, Kar S *et al.* Sex determination and differentiation in teleost: roles of genetics, environment, and brain. *Biology* 2021;**10**:973. https://doi.org/10.3390/biology10100973
66. Tao Y-F, Qiang J, Dagoudo M *et al.* Transcriptome profiling reveals differential expression of immune-related genes in gills of hybrid yellow catfish (Tachysurus fulvidraco ♀ × Pseudobagrus vachellii ♂) under hypoxic stress: potential NLR-mediated immune response. *Fish Shellfish Immunol* 2021;**119**:409–19. https://doi.org/10.1016/j.fsi.2021.10.023
67. Yu Y, Chen M, Lu Z-Y *et al.* High-temperature stress will put the thermo-sensitive teleost yellow catfish (Tachysurus fulvidraco) in danger through reducing reproductivity. *Ecotoxicol Environ Saf* 2022;**239**:113638. https://doi.org/10.1016/j.ecoenv.2022.113638
68. Dubbink HJ, Hersmus R, Verma CS *et al.* Distinct recognition modes of FXXLF and LXXLL motifs by the androgen receptor. *Mol Endocrinol Baltim Md* 2004;**18**:2132–50. https://doi.org/10.1210/me.2003-0375
69. Beccara S A, Škrbić T, Covino R *et al.* Dominant folding pathways of a WW domain. *Proc Natl Acad Sci USA* 2012;**109**:2330–5. https://doi.org/10.1073/pnas.1111796109
70. Terruzzi L, Spagnolli G, Boldrini A *et al.* All-atom simulation of the HET-s prion replication. *PLoS Comput Biol* 2020;**16**:e1007922. https://doi.org/10.1371/journal.pcbi.1007922
71. Dingfelder F, Macocco I, Benke S *et al.* Slow escape from a helical misfolded state of the pore-forming toxin cytolysin A. *JACS Au* 2021;**1**:1217–30. https://doi.org/10.1021/jacsau.1c00175
72. Ianeselli A, Orioli S, Spagnolli G *et al.* Atomic detail of protein folding revealed by an *ab initio* reappraisal of circular dichroism. *J Am Chem Soc* 2018;**140**:3674–82. https://doi.org/10.1021/jacs.7b12399
73. Wang F, Orioli S, Ianeselli A *et al.* All-atom simulations reveal how single-point mutations promote serpin misfolding. *Biophys J* 2018;**114**:2083–94. https://doi.org/10.1016/j.bpj.2018.03.027

74. Spagnolli G, Massignan T, Astolfi A *et al.* Pharmacological inactivation of the prion protein by targeting a folding intermediate. *Commun Biol* 2021;**4**:62. https://doi.org/10.1038/s42003-020-01585-x
75. Palazzolo I, Burnett BG, Young JE *et al.* Akt blocks ligand binding and protects against expanded polyglutamine androgen receptor toxicity. *Hum Mol Genet* 2007;**16**:1593–603. https://doi.org/10.1093/hmg/ddm109
76. Palazzolo I, Stack C, Kong L *et al.* Overexpression of IGF-1 in muscle attenuates disease in a mouse model of spinal and bulbar muscular atrophy. *Neuron* 2009;**63**:316–28. https://doi.org/10.1016/j.neuron.2009.07.019
77. Lin H-K, Wang L, Hu Y-C *et al.* Phosphorylation-dependent ubiquitylation and degradation of androgen receptor by Akt require Mdm2 E3 ligase. *EMBO J* 2002;**21**:4037–48. https://doi.org/10.1093/emboj/cdf406
78. Glasauer SMK, Neuhauss SCF. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics* 2014;**289**:1045–60. https://doi.org/10.1007/s00438-014-0889-2
79. Irisarri I, Baurain D, Brinkmann H *et al.* Phylotranscriptomic consolidation of the jawed vertebrate timetree. *Nat Ecol Evol* 2017;**1**:1370–8. https://doi.org/10.1038/s41559-017-0240-5

Received: December 17, 2024. **Revised:** June 10, 2025. **Editorial Decision:** June 13, 2025. **Accepted:** July 3, 2025