Advances in vitamin D receptor function and evolution based on the 3D structure of the lamprey ligand binding domain

Rita Sigüeiro, Laurent Bianchetti, Carole Peluso-Iltis, Sandra Chalhoub, Annick Dejaegere, Judit Osz, Natacha Rochel

Advances in vitamin D receptor function and evolution based

on the 3D structure of the lamprey ligand binding domain

Rita Sigüeiro[1,2,3,4], Laurent Bianchetti[1,2,3,4], Carole Peluso-Iltis[1,2,3,4], Sandra Chalhoub[1,2,3,4], Annick Dejaegere[1,2,3,4], Judit Osz[1,2,3,4], Natacha Rochel[1,2,3,4]*

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), 67400 Illkirch, France; [2]Institut National de La Santé et de La Recherche Médicale (INSERM), U1258, 67400 Illkirch, France; [3]Centre National de Recherche Scientifique (CNRS), UMR7104, 67400 Illkirch, France; [4]Université de Strasbourg, 67400 Illkirch, France.

*Corresponding author, Email: rochel@igbmc.fr

Abstract

1α,25-dihydroxyvitamin D_3 (1,25D_3) regulates many physiological processes in vertebrates by binding to the Vitamin D Receptor (VDR). Phylogenetic analysis indicates that jawless fishes are the most basal vertebrates exhibiting a VDR gene. To elucidate the mechanism driving VDR activation during evolution, we determined the crystal structure of the VDR ligand binding domain complex from the basal vertebrate *Petromyzon marinus,* sea lamprey (lVDR). Comparison of 3D crystal structure of lVDR-1,25D_3 complex with higher vertebrates VDR-1,25D_3 structures suggest that 1,25D_3 binds to lVDR similarly to human VDR (hVDR), but with unique features for lVDR around linker regions between H11 and H12 and between H9 and H10. These structural differences may contribute to the marked species differences in transcriptional responses. Further, residue co-evolution analysis among vertebrates VDR identifies amino-acid positions in H9 and the large insertion domain (iD) VDR LBD specific.

Introduction

The Vitamin D Receptor (VDR) is a ligand-dependent transcription factor of the nuclear receptor superfamily. It belongs to the NR1I subfamily together with the Pregnane X Receptor (PXR) and the Constitutive Androstane Receptor (CAR), all involved in detoxification responses and acting as functional heterodimers with the retinoid X receptor (RXR)[1]. The taxonomic distribution of NR1I gene orthologs in vertebrates reveals that VDR is found in both cyclostomes and gnathostomes (Figure

1A), while PXR and CAR were not found in the two currently available genomes of cyclostomes from sea lamprey (*Petromyzon marinus*) and japanese lamprey (*Lampetra japonica*)[2-3]. All species exhibit one VDR gene except the teleost fishes that underwent the 3R genome duplication and exhibit 2 genes coding VDR α and β[4]. NR1I subfamily members originated from a single ancestral gene in the genome of chordate invertebrates such as found in *Ciona intestinalis* (sea squirt) that is related to both PXR and VDR, i.e. CiVDR/PXR[5]. The functions and the endogenous ligands of this receptor remain poorly characterized, although synthetic ligands and natural toxins have been reported to activate CiVDR/PXR[6-7].

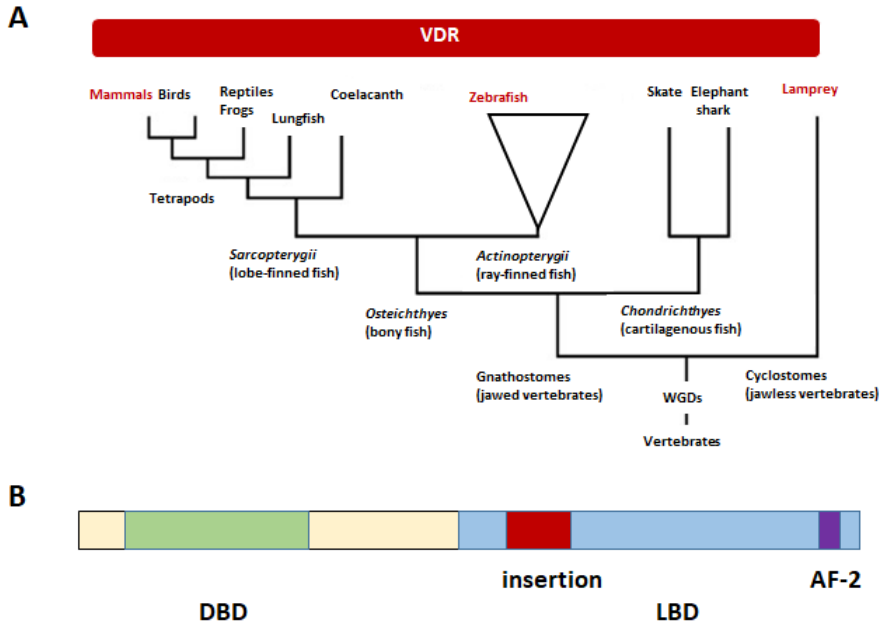

Figure 1: (A) Simplified phylogenetic tree of vitamin D receptor (VDR). The triangle corresponds to ray-finned fishes among which *Danio rerio* (zebrafish). WGDs: whole genome duplication. Species for which VDR LBD structural analysis are discussed are marked in red. **(B) Structural organization of VDR.** DBD: DNA binding domain; LBD: ligand binding domain; iD: insertion domain; AF-2: activation function 2.

Ligand selectivity of VDRs across vertebrates species is tightly conserved (8) and all vertebrates VDR bind to and are activated by 1α,25-dihydroxyvitamin D_3 ($1,25D_3$) generated from 7-dehydrocholesterol[9]. Lampreys contain significant levels of circulating $1,25D_3$ and binding of $1,25D_3$ to lVDR has been shown to be similar to that of hVDR[10]. Despite comparable affinity, $1,25D_3$ only induce partial efficacy of lVDR in transcriptional assays compared to higher vertebrate VDRs[5,10-12].

Before regulating the calcium endocrine system, ancient VDR functioned as a xenobiotic receptor mediating the degradation of marine biotoxins[2]. Evolutionary pressures led to the functional repurposing of the receptor with the acquisition of new roles including detoxification of endogeneous compounds, lipid metabolism, immunity and calcium homeostasis[13-14]. In humans, VDR retains the ability to detoxify toxic compounds such as the secondary bile acid, lithocholic acid (LCA) that binds and activate hVDR in the micromolar range[15]. As LCA is a product of the most recent bile acid pathways (2), LCA does not bind to the basal lVDR[8,16].

VDR shares the main structural characteristics of nuclear receptors: a highly conserved DNA-binding domain (DBD), a ligand-binding domain (LBD), and a hinge region connecting the DBD to the LBD (Figure 1B). In addition, VDRs exhibit a short N-terminal domain (20-30 AA) and an insertion domain (iD) (30-50 AA) in the LBD that is specific to VDR's members[17]. The iD located between helices H2 and H3 in the VDR LBD is poorly conserved in size and sequence, is disordered and does not play a major role in receptor selectivity for $1,25D_3$[17] nor bile acids[11]. lVDR sequence is the most divergent one of the VDRs, the DBD and LBD show 84% and 68% identity with hVDR (Supplementary Figure 1). Hundreds of X-ray crystal structures are available for VDR ligand complexes that correspond to VDR from three different species *Homo Sapiens* (hVDR), *Rattus norvegicus* (rVDR) and *Danio rerio* (zebrafish zVDRα) (review in [18]).

To elucidate the mechanism driving VDR activation during evolution, we determined the crystal structure of VDR LBD complex from the basal vertebrate *Petromyzon marinus* and analyzed the species-specific differences and the co-evolution of the amino acids among vertebrates VDR.

Results

Crystal structure of lVDR LBD-$1,25D_3$. We previously solved the crystal structures of the hVDR LBD [19] and of zVDR LBD in complex with $1,25D_3$[20-21], and the one with rVDR LBD-$1,25D_3$ has been reported[22]. Here we determined the crystal structure of the lVDR LBD in complex with $1,25D_3$ to investigate the molecular mechanism underlying ligand recognition and activation. The protein was crystallized in

the presence of excess of 1,25D_3 and of human NCoA1 peptide encompassing the second nuclear receptor LXXLL interacting motif. Note that NCoA coactivators are well conserved in vertebrates and the sequence of the peptide used in crystallization has 92% similarity with lamprey's sequence. The structure was solved by molecular replacement and refined to 2.5 Å. The data collection and refinement statistics of the structure are summarized in Supplementary Table 1. The overall structure (Figure 2A) is highly homologous to the other VDR-1,25D_3 structures with a three-layered α-helical sandwich composed of 12 helices (H1 to H12) and a three-stranded β-sheet. Although the iD of lVDR is the shortest among VDRs (Supplementary Figure 1), it is not visible in the electron density map, indicating local disorder. Similarly, the 3D structures of hVDR, zVDR and lVDR LBDs predicted by AlphaFold[23-24] indicate that the iD region, which is characterized by a low confidence in the prediction, is disordered (Supplementary Figure 2). The position and conformation of the activated helix H12 is in its agonistic position. Regarding the interface between lVDR LBD and the coactivator peptide, we observed that contacts between protein residues and the coactivator peptide are formed in agreement with the high conservation of the activation function-2 (AF-2) as shown by the ConSurf[25] representation on the VDR LBD structure (Supplementary Figure 3).

In addition to AF-2, the ligand binding pocket (LBP) is the most conserved region in VDRs. The ligand is buried in the predominantly hydrophobic LBP formed by helices H2, H3, H5, H6, H7, H10 and H12, as well as the β-sheet strands. The 1,25D_3 ligand adopts the same orientation as in the previously solved VDR-1,25D_3 structures. The C1-OH, C3-OH and C25-OH groups of 1,25D_3 form similar H-bonds (Figure 2B): the 1α-OH group interacts with lTyr146 (helix H1) and lSer257 (helix H5), the 3β-OH group contacts lSer216 (helix H3) and lArg253 (helix H5), and the 25-OH group interferes with lHis284 (loop between helices H6 and H7) and lHis376 (helix H11). The other interactions formed by 1,25D_3 with lVDR LBP (Figure 2C) are similar to the other VDR-1,25D_3 complexes, reflecting the high conservation of the LBP with 15 residues of the 16 at a cutoff of 4.0 Å being identical to hVDR LBP. Of note, only lLeu397 differs in sequence with a Valine residue in human sequence (hVal418). The

channel of water molecules near the position 2 of the A-ring of 1,25D_3 is observed as in the other VDR structures.

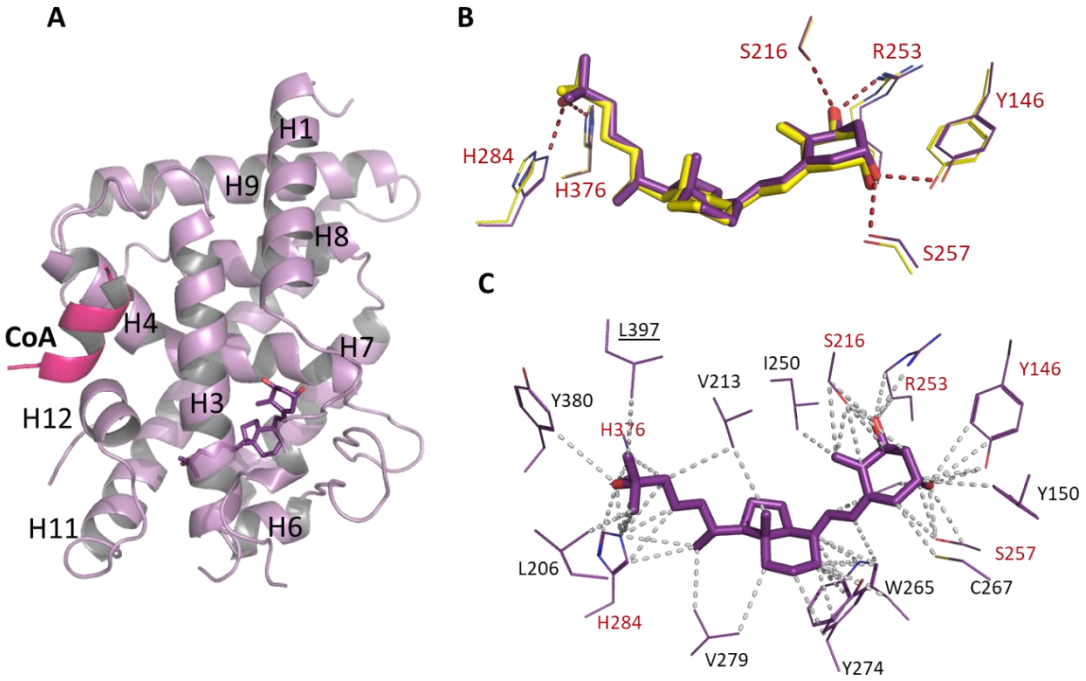

Figure 2: Crystal structure of lVDR LBD in complex with 1,25D_3 and NCoA1 coactivator peptide. A) Overall structure of the complex (PDB ID: 7QPI). **B)**. Hydrogen bonds formed by 1,25D_3 with residues of lVDR (purple) and hVDR (yellow; PDB ID: 1DB1, 7QPP) shown as red dashed lines. Residues numbers correspond to lVDR. **C)** Structural accommodation of 1,25D_3 in the lVDR LBP. Interactions between 1,25D_3 and residues lining the LBP of lVDR LBD at 4 Å cutoff, are shown as gray dashed lines. Residues involved in hydrogen bonds are shown in red and lVDR specific residue is underlined.

Structural comparison of lVDR with VDR orthologs. The LBP is conserved both in sequence and structure, and the ligand conformation and interactions observed in all VDR crystal structures are very similar. The lVDR LBD superimposes to hVDR LBD with rms deviation of 0.87 Å over 230 Cα atoms but significant local differences are observed for lVDR LBD compared to the structures of h, r, z VDR LBDs (Figure 3A). Among lVDR unique features, lVDR exhibits a one turn longer H11 and an extended loop connecting H11-H12 compared to the other structures where the L11-12 adopts a two-turn α-helix (Hx) leading to a shorter H11. In hVDR LBD, Hx packs against H3 and H11 (Figure 3A-B) and residues of Hx are forming intramolecular interactions with residues of H3, H11 and H12, helping to stabilize an active conformation (Figure 3B). In contrast in lVDR LBD, fewer interactions are observed and none with H12, suggesting a lower AF-2 stabilization.

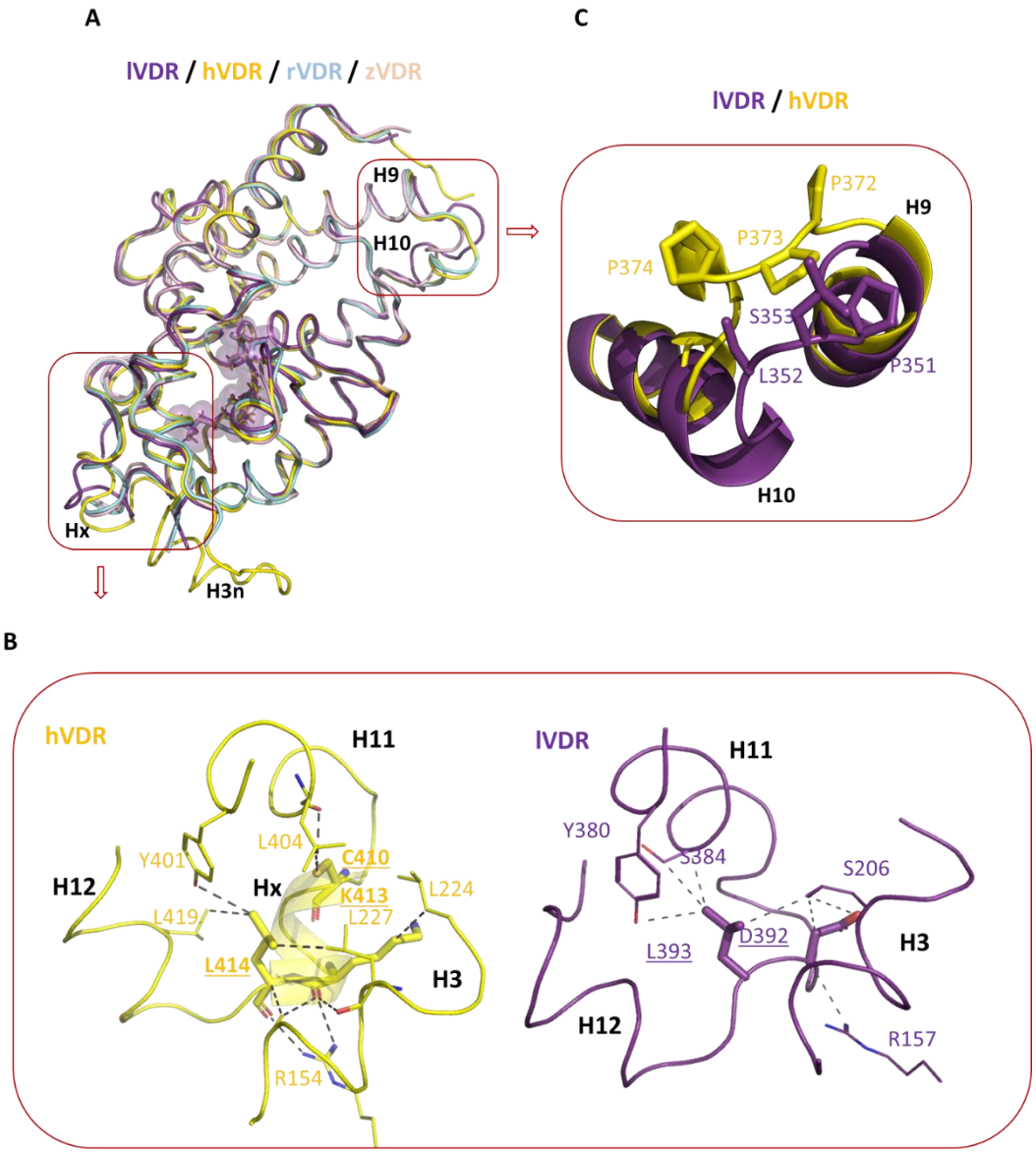

Figure 3: A) Overlay of VDR-1,25D_3 structures (lVDR-1,25D_3 PDB ID: 7QPI; hVDR-1,25D_3 PDB ID: 7QPP; rVDR-1,25D_3 PDB ID: 1RK3; zVDR-1,25D_3 PDB ID: 2HC4). **B**) Detailed view around C-terminus. **C)** Detailed view around loop 9-10 that differs the most between lVDR and hVDR.

A second difference in lVDR structure is the one turn longer H10 and a differentially positioned loop connecting H9 and H10 as compared to h, r, z VDR LBDs (Figure 3C). Interestingly this region is part of the dimeric interface with RXR (Figure 4A). As this difference may affect the interactions of lVDR with RXR, we next monitored the affinity between lVDR LBD and hRXRα LBD by surface plasmon resonance (SPR) and compared it to those of hVDR and zVDR LBDs. We choose hRXRα LBD as it was previously shown to interact with lVDR[12]. Apo hRXRα LBD was immobilized on CM5 chips using

standard amine coupling and increasing concentrations of apo or 1,25D_3-bound lVDR or zVDR LBD were run over as analytes (Figure 4B and Supplementary Figure 4). In the absence of ligand, lVDR LBD binds poorly to hRXRα LBD, whereas in the presence of saturating 1,25D_3 concentration, lVDR binds with a dissociation constant (K_D) of 2.5 μM. The affinity of lVDR is 4- to 6-fold lower than those of zVDR and hVDR LBDs (Figure 4C).

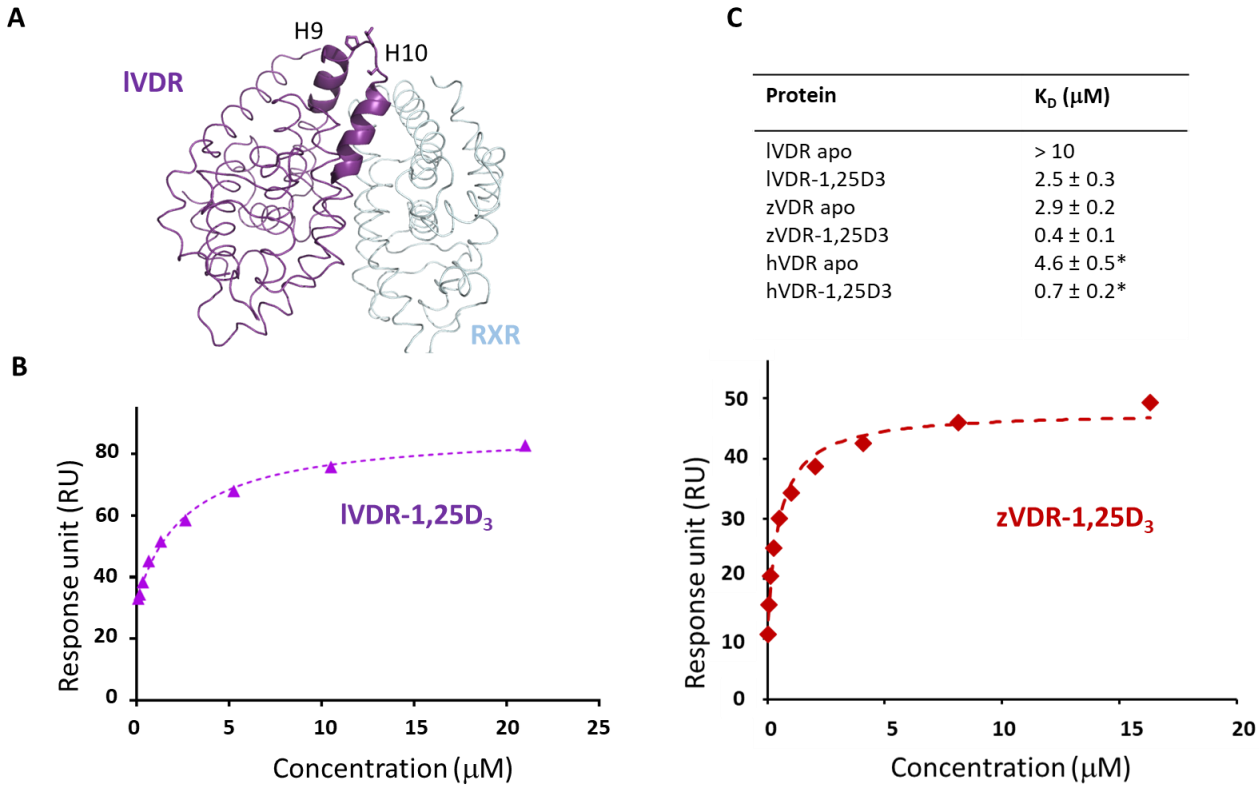

Figure 4: (**A**) 3D model of lVDR/RXR LBDs dimer based on PDB ID: 1XDK. **(B)** Analysis of the interactions between hRXRα LBD with 1,25D_3-bound lVDR LBD or zVDR LBD by SPR. Representative data set used for SPR analysis. Equilibrium responses are plotted as a function of total protein concentration and fit to simple 1:1 binding isotherms to determine the steady-state binding affinity. (**C**) Averaged equilibrium dissociation constants for interactions of VDR LBDs with hRXRα LBD from 2 to 3 independent experiments. * from [26].

VDR LBD shows correlated mutations in the insertion domain. The structures revealed that the active site residues are conserved to maintain their functional integrity. To identify covariant amino-acid positions or co-evolving amino-acid pairs occurring at specific positions that are maintained by evolutionary pressure to accommodate novel functionalities, we used the mutual information (MI) method[27] of amino-acids frequencies between positions in multiple sequence alignment (MSA). The MSA contained 331 vertebrate VDR LBD sequences originating from 312 distinct organisms from sea lamprey to human, with 39.6% mammals including 8.5% primate sequences, 17.5% birds, 7.3%

reptiles and 35% fishes. When available in protein sequence databases, the α and β paralogs were both added to the alignment. The quality of the MSA was assessed using the normalized mean distance (norMD) scores[28]. Full-length MSA and most regions except H2 and iD exhibited a maximum score of 1, supporting high quality alignment (Supplementary Table 2). The conservation profile exhibits three regions. First, at the N-terminus, the H1-H2 region (from hLeu124 to hGln152 in hVDR) is rather conserved. Second, the iD (hPhe153 to hSer216) is weakly conserved. Finally, the region from H3 to H12 (hVal217 to the C-terminal hSer427) is strongly conserved. Of particular note, although the iD was deemed weakly conserved (Supplementary Figure 5), its norMD score was above the quality cut-off (Supplementary Table 2), which indicated that the loop residues are not extensively variable but rather conserved inside groups of sequences, i.e. taxonomic groups.

MI scores were calculated between every pair of MSA positions and the distribution of scores was analyzed (Supplementary Figure 6). About half of the scores equaled zero (49.4%) and showed no correlation information between the compared positions in the MSA. The maximum score 0.474 was calculated between positions hSer177 and hSer200 that are both localized in the iD. In the MI method, scores between pairs of highly variable positions are known to be biased and slightly greater than zero[29]. To estimate the bias, 10 MSA of 331 sequences of two random residues were generated and a MI score was calculated for each MSA. Finally, the mean of MI scores was determined, i.e. 0.263 ± 0.02. Therefore, in the VDR MSA, any pair of positions which showed a MI score less than 0.263 was considered insignificant and set to zero. However, 1182 scores were still above the MI threshold. A score matrix was drawn with scores greater than the threshold (Figure 5A). The matrix clearly shows that iD positions, i.e. from hVal159 to hGln223 correlate with each other and with remote positions in the loop between H5 and H6 such as hGln291 and with hAla349 in H9. Finally, the number of correlated positions was reduced to the 99.5th percentile of top scores, i.e. 0.351 and 231 position pairs, to allow a readable structural representation (Figure 5B). In the 231 selected pairs, 31 positions were unique and most of them in the insertion domain.

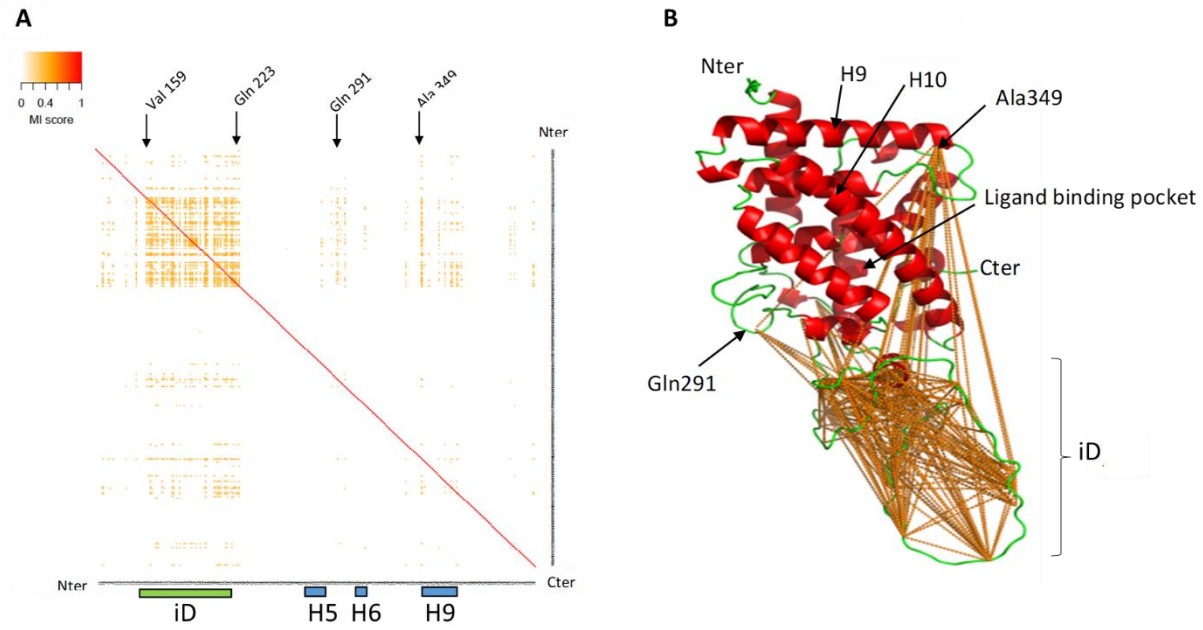

Figure 5: Mutual information (MI) analysis of covariant residue positions in the whole VDR LBD. A) MI score matrix. The insertion domain (iD) is indicated with a green rectangle. The matrix red diagonal represents default MI score values of one when a column of the MSA is compared to itself, and are not meaningful. **B)** 3D structure of the hVDR LBD (PDB ID: 1DB1) with iD modeled[30] with network of top scoring MI positions (99.5 percentile, i.e. MI score greater than 0.35). Every pair of position which shows a MI score greater than 0.35 are connected by an orange dashed line.

Interestingly, in a sequence cluster of birds from the *Passeriform* order (Supplementary Figure 7), specific correlated mutations are observed in H9, i.e. Gln374, Leu351, Asp354, Asp357, Thr362, Thr365, Cys369, Pro372 and Pro373, all located on the solvent exposed face of the H9 and facing the DBD-DNA in the VDR full-length structure (Supplementary Figure 8).

Discussion

Our results show structural differences in lVDR LBD in comparison with higher vertebrates VDR LBDs. These structural changes around linker regions between H11-H12 and H9-H10, differentially modulate AF-2 stabilization and RXR dimerization. Co-evolution sequence analysis of VDR orthologs reveals correlated mutations localized in iD and H9.

Vitamin D_3 was already present in ancient organisms such as phytoplankton, and the enzymes capable of producing the hydroxylated metabolites of vitamin D_3 appeared early in evolution[31]. VDR orthologs have been found in invertebrate Tunicates, the sister clade to vertebrates, *Ciona intestinalis* (Ci) and *Botryllus schlosseri* (Bs)[32]. Both encode at least two genes considered orthologous to the vertebrate PXR, CAR and VDR and abbreviated as VDR/PXR[6-7,32]. CiVDR/PXR has been shown not to be activated by vitamin D neither by bile acids but to be activated by marine biotoxins and to be closer functionally to PXR's current role in xenobiotics detoxification[2,32]. It is thought that the

VDR/PXR gene split occurred after the 1R WGD giving rise to separate VDR and PXR genes with different but also overlapping functions. The jawless fishes whose modern lampreys are descendants were first to diverge after the 1R WGD event and lampreys are considered to be the most basal vertebrates among extant organisms. As revealed by VDR gene repertoire analysis[33-34], its function has evolved along with more complex animal species from detoxification response in the basal vertebrates to lipid metabolism, immunity and calcium homeostasis in higher vertebrates[14]. Consistent with the high ligand-binding homology across various species and high affinity binding of $1,25D_3$ to all vertebrates VDRs[5,10], the crystal structure of lVDR-$125D_3$ complex revealed that $1,25D_3$ form similar interactions with lVDR LBP in comparison to the other crystal structures of VDR orthologs. Despite high affinity and transactivation potency in the low nanomolar range, $1,25D_3$ has been shown to exhibit a lower efficacy to activate lVDR compared to higher vertebrate VDRs[5,10,12].

Functional VDR needs to heterodimerize with RXR to exert biological effects and RXR has been shown to boost transactivation of lVDR[10,12]. Overexpresion of hRXR in transcriptional assays has been shown to increase VDR activation for all VDRs, lVDR being particularly sensitive[12]. RXR orthologs exist in all vertebrates and invertebrates, and RXR has coevolved with VDR[2-3]. Three RXRs (lRXR1, lRXR2, lRXR3) have been cloned from sea lamprey and their sequences exhibit a high conservation between them (92-96% identity)[35]. We show that $1,25D_3$ binding to lVDR stabilize the heterodimer, similarly to the other vertebrate VDRs. This is in agreement with published data showing that addition of ligand enhances heterodimer binding to canonical VDREs[12]. However, we show that lVDR LBD forms a weaker dimer with hRXRα LBD compared to hVDR or zVDR LBDs, accounting for the weak transcriptional maximal response in absence of excess of exogenous RXR. Note that the heterodimer interface is conserved between hRXRα and lRXRs LBDs (86% identity for the LBDs and 100% for the dimer interface), suggesting that the decreased affinity between lVDR and RXR is the consequence of the lVDR specific structural differences in H10 and loop 9-10 and of the lower affinity of $1,25D_3$ for lVDR [Kd of 0.7 nM for lVDR in comparison to 0.3 nM for hVDR [10]]. Our data show that VDR has evolved from a low affinity protein partner in lamprey into a higher affinity interactant in higher vertebrates. The lower stabilization of the lVDR agonist conformation together with lower affinity of lVDR for RXR explain the lower efficacy of $1,25D_3$ for lVDR, corroborating previous hypothesis on differential interactions with protein partners between basal and higher vertebrates VDRs[12,36].

In addition to $1,25D_3$, LCA has been shown to bind and to activate VDR with marked species differences. LCA binds and activates in the μM range VDRs from teleost to human but does not to activate lVDR[8,12,16]. The high conservation of VDR LBP suggest that LCA will bind to lVDR LBP similarly as to rVDR or zVDR[37-38], forming no direct interaction with AF-2. The absence of direct interaction with H12 together with the lower stabilization of lVDR C-ter region explain why LCA is unable to activate lVDR.

The protein structures are similar and the active site residues are under evolutionary pressure to maintain their functional integrity, exhibiting fewer mutations than less functionally important amino acids. Co-evolving residues that are maintained by evolution pressure allow the proteins to accommodate novel functionalities[39-40]. Co-evolution analysis by statistical coupling has been previously applied to the nuclear receptor LBDs[41] where nuclear receptors sequences containing paralogs and orthologs of the different subfamilies were used, revealing important allosteric communication across dimer interface[41]. By using MI methods applied on VDR subfamily we showed that residues of the insertion domain correlate with each other and with hGln291 in loop 5-6 and with hAla349 in H9. VDR's specific insertion domain exhibits the greatest divergence among VDRs and has been shown to be disordered and not to play a major role in receptor selectivity for $1,25D_3$[17] nor bile acids[11]. For crystallization of hVDR and rVDR LBDs, constructs where iD were deleted were used[19,22]. Similarly to crystalized zVDR, the iD of lVDR remains flexible and does not interact with the ligand. Our co-evolution analysis among vertebrates VDR reveals that this region exhibits most of covariant amino-acid positions. Interestingly, the identified co-evolved residues are part of the interacting interface between hVDR with the coactivator MED1[30]. Located in hVDR iD, hSer208 phosphorylation has been reported to enhance VDR interaction with MED1[42-43]. The corresponding residue in lVDR is lLeu187, suggesting a lower affinity of lVDR for MED1. In addition, several data indicate specific role of iD in higher vertebrates. A mutation in iD, Cys190Trp, was reported in human patients of herediditary vitamin D resistant ricket that results in loss of $1,25D_3$ binding[44]. A caspase-3 cleavage site (195-DMMD-198S199) that is involved in VDR inactivation and only present in primates is also found in iD[45] as well as a phospho-degron motif (175-TPSFS-179) only present in primates[46]. Thus, our results indicate that VDR iD residues have co-evolved to modulate VDR interactions and to gain new functionality in regulation of VDR signaling in higher vertebrates.

Conclusions

In summary, biophysical, structural and sequence co-evolution analysis allow us to infer ancestral and derived properties of more recent VDR function. Increasing sensitivity to RXR and coregulators during evolution leads to increase transactivation responses to $1,25D_3$ and VDR to be fully activated by 1,25D3 and to respond to LCA in higher vertebrates, facilitating novel functions. As lamprey may be used as an ideal model organism with less complex Vitamin D - regulated functions, future work should focus on structure-function correlation with *in vivo* data that could unravel novel aspect of Vitamin D signaling and function of VDR and provide new clues for future design of specific therapeutic drugs.

Experimental Section

Chemicals. 1,25D_3 (gift from Antonio Mourino) was dissolved in absolute ethanol at 10^{-2} M and stored at -20 °C. Tested compound was >95% pure as determined by HPLC (Supporting Information). The SRC1 NR2 peptide was synthesized at IGBMC peptide synthesis common facility.

Expression and purification. The cDNA encoding GST-tagged lVDR LBD (124-406) was cloned into pGEX-4T-3. The recombinant proteins were produced in Escherichia coli BL21 DE3 grown ON at 18 °C after induction with 1 mM IPTG at an OD600 of ~0.7. Soluble proteins were purified on Glutathione Sepharose 4B column, followed by GST tag removal by thrombin cleavage and by size exclusion chromatography on HiLoad Superdex 75 column equilibrated in Tris 20mM pH7, NaCl 200mM, TCEP 1mM. The proteins were concentrated to 3-5 mg/mL with an Amicon Ultra 10 kDa MWCO. Purity and homogeneity of the proteins were assessed by SDS and Native Page.

Crystallization and structure determination. The concentrated protein was incubated with a 2-fold excess of 1,25D_3 and a 3-fold excess of the coactivator NCoA1 NR2 (RHKILHRLLQEGS) peptide. Crystals were obtained in PEG 3350 25%, Tris 0.1M pH 8.5, magnesium chloride 0.2M at 293K. Protein crystals were mounted in a fiber loop and flash-cooled under a nitrogen flux after cryoprotection with PEG 3350 35%. Data collection from a single frozen crystal was performed at 100 K on the PX2 at SOLEIL (France). The raw data were processed with XDS[47] and scaled with AIMLESS[48] programs. The crystals belong to the space group $P2_12_12_1$, with one LBD complex per asymmetric unit. The structure was solved and refined using Phenix[49], BUSTER[50] and iterative model building using COOT[51]. Crystallographic refinement statistics are presented in Supplementary Table 1.

Surface Plasmon Resonance. Measurements were performed by a Biacore T100 sensitivity enhanced T200 equipment (GE Healthcare) using CM5 series S sensor chip (GE) (29-1496-03). The hRXRα LBD was immobilized on the chip surface using a standard amino-coupling protocol in 10 mM Na-acetate buffer pH 5.5. The resulting immobilized RXR was in the range of 150-250 response unit. The running buffer was 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP, 0.005% Tween 20. For regeneration 1 M sodium chloride solution was used. Interactions between hRXRα LBD and VDR LBD were analyzed by dose response using twofold dilution series of VDR LBDs ranging from 80 nM to 42 µM in the absence or presence of ligand. The association phase was 120 s and the dissociation phase was 120 s. After subtracting the reference and buffer signal, the steady state affinity were determined by the Biacore T200 Evaluation software (GE Healthcare).

Co-evolution analysis. The hVDR LBD sequence was used as a query to carry out a sequence similarity search with BLASTp in Swissprot, RefSeq (NCBI) and Genepep (NCBI) databases with default parameters. VDR orthologues were selected using significant blast scores, expect values and sequence annotations. LBD sequences were aligned with MAFFT version 7[52] and manually refined in Jalview 2.11.1.2[53]. The human sequence was used as a reference point and every column of the MSA

that introduced a gap in the human sequence was deleted before proceeding with MI calculations. Residue conservation scores were calculated with the Livingston & Barton (L&B) algorithm[54] implemented in Jalview. Sequences that introduced large gaps in the iD loop were discarded. Alignment quality was estimated with the norMD program[28]. NorMD scores range from zero (poor quality) to one (invariant or highly conserved position). We used the authors recommended 0.5 score cut-off as a marker of good alignment quality[28]. Correlated mutations were determined with the MI algorithm[27] implemented in an in-house developed PERL 5 script. The MI score data matrix was generated with a PYTHON 2.7.17 script while the graphical representation of the matrix was created in R using the gplots library and the heatmap.2 function.

All structural figures were prepared using PyMOL (www.pymol.org).

ASSOCIATED CONTENT

Accession Code. The atomic coordinates and structure factors have been deposited accession code 7QPI. Authors will release the atomic coordinates and experimental data upon article publication.

Supporting Information.

Table of the refinement of the crystal structure as well as $1,25D_3$ HPLC traces, additional supplementary figures and table can be found in SI appendix (PDF).

AUTHOR INFORMATION

Corresponding Author

N. Rochel (rochel@igbmc.fr)

ORCID

Natacha Rochel: 0000-0002-3573-5889

ACKNOWLEDGMENT

We acknowledge Pr A. Mourino, Universidad de Santiago de Compostela, Spain, for the $1,25D_3$. We thank P. Eberling (IGBMC) for peptide synthesis and M. Savko (PX2, Soleil) for X-ray data collection. The work was supported by institutional funds from Instruct-ERIC for support and the use of resources of the French Infrastructure for Integrated Structural Biology (ANR-10-LABX-0030-INRT and ANR-10-IDEX-0002-02). S.C. thanks ARTP for a doctoral fellowship.

ABBREVIATIONS USED

$1,25D_3$, 1α,25-dihydroxyvitaminD_3; AF-2, activation function; CAR, constitutive androstane receptor; DBD, DNA binding domain; iD, insertion domain; LBD, ligand binding domain: LBP, ligand binding pocket; LCA, lithocholic acid; MI, mutual information; MSA, multiple sequence alignment; norMD, normalized mean distance; PXR, pregnane X receptor; RXR, retinoid X receptor; VDR, vitamin D receptor

REFERENCES

1. Reschly, E.J.; Krasowski, M.D. Evolution and function of the NR1I nuclear hormone receptor subfamily (VDR, PXR, and CAR) with respect to metabolism of xenobiotics and endogenous compounds. *Curr. Drug Metab*. **2006,** *7*, 349-365.

2. Newmark, H.; Dantoft, W.; Ghazal, P. Evolutionary origin of the interferon-immune metabolic axis: the sterol-vitamin D link. *Front Immunol*. **2017**, *8*, 62.

3. Fonseca, E.S.S.; Ruivo, R.; Machado, A.M.; Conrado, F.; Tay, B.H.; Venkatesh, B.; Santos, M.M.; Castro, L.F.C. Evolutionary plasticity in detoxification gene modules: the preservation and loss of the pregnane X receptor in chondrichthyes lineages. *Int J Mol Sci*. **2019**, *20*, 2331.

4. Kollitz, E.M.; Hawkins, M.B.; Whitfield, G.K.; Kullman, S.W. Functional diversification of vitamin D receptor paralogs in teleost fish after a whole genome duplication event. *Endocrinology*. **2014**, *155*, 4641-4654.

5. Reschly,, E.J.; Bainy, A.C.; Mattos, J.J.; Hagey, L.R.; Bahary, N.; Mada, S.R.; Ou, J.; Venkataramanan, R.; Krasowski, M.D. Functional evolution of the vitamin D and pregnane X receptors. *BMC Evol Biol.* **2007**, *7*, 222.

6. Ekins, S.; Reschly, E.J.; Hagey, L.R.; Krasowski, M.D. Evolution of pharmacologic specificity in the pregnane X receptor. *BMC Evolutionary Biology* **2008**, *8*, 103.

7. Fidler, A.E.; Holland, P.T.; Reschly, E.J.; Ekins, S.; Krasowski, M.D. Activation of tunicate (Ciona intestinalis) xenobiotic receptor orthologue by both natural toxins and synthetic toxicants. *Toxicon*. **2012**, *59*, 365-372.

8. Krasowski, M.D.; Yasuda, K.; Hagey, L.R.; Schuetz, E.G. Evolutionary selection across the nuclear hormone receptor superfamily with a focus on the NR1I subfamily (vitamin D, pregnane X, and constitutive androstane receptors). *Nucl Recept.* **2005**, *3*, 2.

9. Tuckey, R.C.; Cheng, C.Y.S.; Slominski, A.T. The serum vitamin D metabolome: What we know and what is still to discover. *J Steroid Biochem Mol Biol*. **2019**, *186*, 4-21.

10. Whitfield, K.G.; Dang, H.T.L.; Schluter, S.F.; Bernstein, R.M.; Bunag, T.; Manzon, L.A. Hsieh, G.; Dominguez, C.E.; Youson, J.H.; Haussler, M.R. ; Marchalonis, J.J. Cloning of a functional vitamin D receptor from the lamprey (Petromyzon marinus), an ancient vertebrate lacking a calcified skeleton and teeth. *Endocrinology*. **2003**, *144*, 2704-2716.

11. Krasowski, M.D.; Ai, N.; Hagey, L.R.; Kollitz, E.M.; Kullman, S.W.; Reschly, E.J.; Ekins, S. The evolution of farnesoid X, vitamin D, and pregnane X receptors: insights from the green-spotted pufferfish (Tetraodon nigriviridis) and other non-mammalian species. *BMC Biochem*. **2011**, *12*, 5.

12. Kollitz, E.M.; Zhang, G.; Hawkins, M.B.; Whitfield, G.K.; Reif, D.M.; Kullman, S.W. Molecular cloning, functional characterization, and evolutionary analysis of vitamin D receptors isolated from basal vertebrates. *PLoS One*. **2015**, *10*, e0122853.

13. Bouillon, R.; Suda, T. Vitamin D: calcium and bone homeostasis during evolution. *BoneKEy Reports* **2014**, *3*, 480.

14. Hanel, A.; Carlberg, C. Vitamin D and evolution: Pharmacologic implications. *Biochem Pharmacol.* **2020**, *173*, 113595.

15. Makishima, M.; Lu, T.T.; Xie, W.; Whitfield, G.K.; Domoto, H.; Evans, R.M.; Haussler, M.R.; Mangelsdorf, D.J. Vitamin D receptor as an intestinal bile acid sensor. *Science*. **2002**, *296*, 1313-1316.

16. Krasowski, M.D.; Yasuda, K.; Hagey, L.R.; Schuetz, E.G. Evolution of the pregnane X receptor: adaptation to cross-species differences in biliary bile salts. *Mol Endocrinol*. **2005**, *19*, 1720-1739.

17. Rochel, N.; Tocchini-Valentini, G.; Egea, P.F.; Juntunen, K.; Garnier, J.M.; Vihko, P.; Moras, D. Functional and structural characterization of the insertion region in the ligand binding domain of the vitamin D nuclear receptor. *Eur J Biochem*. **2001**, *268*, 971-979.

18. Maestro, M.A.; Molnár, F.; Carlberg, C. Vitamin D and Its Synthetic Analogs. *J Med Chem.* **2019**, *62*, 6854-6875.

19. Rochel, N.; Wurtz, J.M.; Mitschler, A.; Klaholz, B.; Moras, D. The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. *Mol Cell*. **2000**, *5*, 173-179.

20. Ciesielski, F.; Rochel, N.; Mitschler, A.; Kouzmenko, A.; Moras, D. Structural investigation of the ligand binding domain of the zebrafish VDR in complexes with 1alpha,25(OH)2D3 and Gemini: purification, crystallization and preliminary X-ray diffraction analysis. *J Steroid Biochem Mol Biol.* **2004**, *89-90*, 55-59.

21. Ciesielski, F.; Rochel, N.; Moras, D. Adaptability of the Vitamin D nuclear receptor to the synthetic ligand Gemini: remodelling the LBP with one side chain rotation. *J Steroid Biochem Mol Biol.* **2007**, *103*, 235-242.

22. Vanhooke, J.L.; Benning, M.M.; Bauer, C.B.; Pike, J.W.; DeLuca, H.F. Molecular structure of the rat vitamin D receptor ligand binding domain complexed with 2-carbon-substituted vitamin D3 hormone analogues and a LXXLL-containing coactivator peptide. *Biochemistry*. **2004**, *43*, 4101-4110.

23. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S.A.A.; Ballard, A.J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A.W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

24. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Žídek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439-D444.

25. Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **2016**, *44*, 344-350.

26. Rovito, D.; Belorusova, A.Y.; Chalhoub, S.; Rerra, A.I., Guiot, E.; Molin, A.; Linglart, A.; Rochel, N.; Laverny, G.; Metzger, D. Cytosolic sequestration of the vitamin D receptor as a therapeutic option for

vitamin D-induced hypercalcemia. *Nat Commun*. **2020**, *11*, 6249.

27. Atchley, W.R.; Wollenberg, K.R.; Fitch, W.M.; Terhalle, W.; Dress, A.W. Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. *Mol Biol Evol*. **200**0, *17*, 164-178.

28. Thompson, J.D.; Pewniak, F.; Ripp, R.; Thierry, J.C.; Poch, O. Towards a reliable objective function for multiple sequence alignments. *J. Mol. Biol*. **2001**, *314*, 937-951.

29. Fodor, A.A.; Aldrich, R.W. Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins*. 2004, 56, 211-221.

30. Belorusova, A.Y.; Bourguet, M.; Hessmann, S.; Chalhoub, S.; Kieffer, B.; Cianférani, S.; Rochel, N. Molecular determinants of MED1 interaction with the DNA bound VDR-RXR heterodimer. *Nucleic Acids Res.* **2020**, *48*, 11199-11213.

31. Holick, M.F. Vitamin D: evolutionary, physiological and health perspectives. *Curr Drug Targets*. **2011**, *12*, 4-18.

32. Richter, I.; Fidler, A.E. Tunicate pregnane X receptor (PXR) orthologs: transcript characterization and natural variation. *Mar Genomics*. **2015**, *23,* 99-108.

33. Laizé, V. ; Martel, P. ; Viegas, C.S. ; Price, P.A. ; Cancela, M.L. Evolution of matrix and bone gamma-carboxyglutamic acid proteins in vertebrates. *J Biol Chem*. **2005**, *280*, 26659-26668.

34. Mazzaferro, S.; Pasquali, M. Vitamin D: a dynamic molecule. How relevant might the dynamism for a vitamin be? *Nephrol Dial Transplant*. **2016**, *31*, 23-30.

35. Manzon, L.A.; Youson, J.H.; Holzer, G.; Staiano, L.; Laudet, V.; Manzon, R.G. Thyroid hormone and retinoid X receptor function and expression during sea lamprey (Petromyzon marinus) metamorphosis. *Gen Comp Endocrinol.* **2014**, *204*, 211-222.

36. Kollitz, E.M.; Zhang, G.; Hawkins, M.B.; Whitfield, G.K.; Reif, D.M.; Kullman, S.W. Evolutionary and functional diversification of the vitamin D receptor-lithocholic acid partnership. *PLoS One*. **2016**, *11*, e0168278.

37. Masuno, H.; Ikura, T.; Morizono, D.; Orita, I.; Yamada, S.; Shimizu, M.; Ito, N. Crystal structures of complexes of vitamin D receptor ligand-binding domain with lithocholic acid derivatives. *J Lipid Res.* **2013**, *54*, 2206-2213.

38. Belorusova, A.Y.; Eberhardt, J.; Potier, N.; Stote, R.H.; Dejaegere, A.; Rochel, N. Structural insights into the molecular mechanism of vitamin D receptor activation by lithocholic acid involving a new mode of ligand recognition. *J Med Chem*. **2014**, *57*, 4710-4719.

39. Sandler, I.; Zigdon, N.; Levy, E.; Aharoni, A. The functional importance of co-evolving residues in proteins. *Cell Mol Life Sci*. **2014,** *71*, 673-682.

39. Iserte, J.A.; Lazar, T.; Tosatto, S.C.E.; Tompa, P.; Marino-Buslje, C. Chasing coevolutionary signals in intrinsically disordered proteins complexes. *Sci Rep.* **2020**, *10*, 17962.

41. Shulman, A.I.; Larson, C.; Mangelsdorf, D.J.; Ranganathan, R. Structural determinants of allosteric ligand activation in RXR heterodimers. *Cell.* **2004**, *116*, 417-429.

42. Barletta, F.; Freedman, L.P.; Christakos, S. Enhancement of VDR-mediated transcription by phosphorylation: correlation with increased interaction between the VDR and DRIP205, a subunit of the VDR-interacting protein coactivator complex. *Mol Endocrinol*. **2002**, *16*, 301-314.

43. Arriagada, G.; Paredes, R.; Olate, J.; van Wijnen, A.; Lian, J.B.; Stein, G.S.; Stein, J.L.; Onate, S.; Montecino, M. Phosphorylation at serine 208 of the 1alpha,25-dihydroxy vitamin D3 receptor modulates the interaction with transcriptional coactivators. *J Steroid Biochem Mol Biol*. **2007**, *103*, 425-429.

44. Malloy, P.J.; Feldman D. Vitamin D resistance. *Am J Med*. **1999**, *106*, 355-370.

45. Malloy, P.J.; Feldman, D. Inactivation of the human vitamin D receptor by caspase-3, *Endocrinology* **2009**, *150*, 679–686.

46. Salehi-Tabar, R.; Memari, B.; Wong, H.; Dimitrov, V.; Rochel, N.; White, J.H. The Tumor suppressor FBW7 and the vitamin D receptor are mutual cofactors in protein turnover and transcriptional regulation. *Mol Cancer Res*. **2019**, *17*, 709-719.

47. Kabsch, W. Software XDS for image rotation, recognition and crystal symmetry assignment. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125−132.

48. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr., Sect. D: Biol. Crystallogr*. **2006**, *62*, 72−82.

49. Afonine, P.V.; Grosse-Kunstleve, R.W.; Adams, P.D. *CCP4 Newsl* **2005**, 42, contribution 8

50. Bricogne, G.; Blanc, E.;Brandl, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P.; Sharff, A.; Smart, O.S.; Vonrhein, C.; Womack, T.O. BUSTER, version 2.11.2, **2011,** Global Phasing Ltd: Cambridge, U.K.

51. Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.

52. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol*. **2013**, *30*, 772-780.

53. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. **2009**, *25*, 1189-1191.

54. Livingstone, C.D.; Barton, G.J. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci*. **1993**, *9*, 745-756.

Table of Contents graphic