Research Article

Structural insights into the interaction of the intrinsically disordered co-activator TIF2 with retinoic acid receptor heterodimer (RXR/RAR)

Lucile Senicourt[a,†], Albane le Maire[a,†], Frédéric Allemand[a], João E. Carvalho[b], Laura Guee[a], Pierre Germain[a], Michael Schubert[c], Pau Bernadó[a], William Bourguet[a,*], Nathalie Sibille[a,*,1]

[a] Centre de Biochimie Structurale (CBS), CNRS, INSERM, Univ Montpellier, Montpellier, France
[b] Institute for Research on Cancer and Aging (IRCAN), INSERM U1081 - CNRS UMR 7284, Université Cotê d'Azur, 28, Avenue de Valombrose, 06107 Nice, France
[c] Evolution of Intercellular Signaling in Development (EvoInSiDe), UMR 7009 - CNRS/Sorbonne Université), Institut de la Mer de Villefranche, 181 Chemin du Lazaret, 06230 Villefranche-sur-Mer, France
[†] These authors contributed equally to this work.
[*] Corresponding authors. Email addresses: william.bourguet@cbs.cnrs.fr, nathalie.sibille@cbs.cnrs.fr
[1] Lead contact. Tel: +33 4 67 41 77 05; Fax: +33 4 67 41 79 13; Email address: nathalie.sibille@cbs.cnrs.fr

ABSTRACT

Retinoic acid receptors (RARs) and retinoid X receptors (RXRs) form heterodimers that activate target gene transcription by recruiting co-activator complexes in response to ligand binding. The nuclear receptor (NR) co-activator TIF2 mediates this recruitment by interacting with the ligand-binding domain (LBD) of NRs trough the nuclear receptor interaction domain (TIF2NRID) containing three highly conserved α-helical LxxLL motifs (NR-boxes). The precise binding mode of this domain to RXR/RAR is not clear due to the disordered nature of TIF2. Here we present the structural characterization of TIF2NRID by integrating several experimental (NMR, SAXS, Far-UV CD, SEC-MALS) and computational data. Collectively, the data are in agreement with a largely disordered protein with partially structured regions, including the NR-boxes and their flanking regions, which are evolutionary conserved. NMR and X-ray crystallographic data on TIF2NRID in complex with RXR/RAR reveal a multisite binding of the three NR-boxes as well as an active role of their flanking regions in the interaction.

INTRODUCTION

The 48 human nuclear receptors (NRs) belong to a major family of ligand regulated-transcription factors that represent important drug targets for several diseases (cancer, diabetes, obesity, infertility) [1,2]. NRs are activated by small lipophilic ligands such as hormones, vitamins and dietary lipids to regulate diverse functions connected to homeostasis, reproduction, development and metabolism [1,3]. The retinoic acid receptor (RAR) and the retinoid X Receptor (RXR) form a heterodimer that binds to specific DNA sequences (called RAREs - Retinoic Acid Response Elements) of the targeted genes and regulate gene transcription in response to retinoids. RXR/RAR heterodimer regulated genes are involved in crucial physiological processes, such as embryo development and organ homeostasis [4,5]. The ligand-dependent transcriptional activity of the RXR/RAR heterodimer relies on the dynamical balance between binding and release of co-regulator proteins. Unliganded and inverse agonist bound heterodimers exert a repressive effect by interacting with co-repressors, such as N-CoR or SMRT, which act as molecular platforms to recruit proteins with a histone deacetylase activity that induces chromatin compaction and precludes gene transcription [6,7]. Conversely, binding of RAR agonists induces conformational changes in RAR that dissociate the co-repressor complex and subsequently, lead to the recruitment of members of the nuclear receptor coactivator 1-3 (NCOA1-3) families of proteins such as SRC1, TIF2 or RAC3, respectively. Those co-activators associate to histone acetylases to uncompact chromatin and trigger gene transcription [8]. Co-activators interact with the NR ligand binding domain (LBD) through a region named the nuclear receptor interaction domain (NRID) characterized by the presence of multiple conserved LxxLL motifs (the so-called NRboxes) essential for the interaction [9–14]. Similar motifs (LxxI/HIxxI/L, called CoRNR-boxes) are identified in co-repressors [15–17]. Importantly, NRIDs are mainly intrinsically disordered regions with pre-formed secondary structure elements corresponding to NR- and CoRNR-boxes [18–21].

The structural bases of the interactions between some co-regulator proteins and RAR have been revealed by crystallographic structures of RAR LBD in complex with peptides harbouring the interacting motifs [22–25]. These structural studies demonstrate that the core of NR- and CoRNRboxes adopt α-helical conformations and bind to the surface formed by residues from helices H3, H4 and H12 on the LBD of RAR. Importantly, the conformation of the LBD C-terminal helix (named helix 12 or H12) dictates the class of co-regulator recruited by NRs. Indeed, on the one hand, co-repressor binding to unliganded RAR is mediated by the formation of an extended β-strand (β1) in CoRNR1 that forms an antiparallel β-sheet with RAR residues (S3), while H12 remains flexible in solution [18,22,26–29]. On the other hand, binding of agonist induces the transition of S3 to H11 and the folding of H12 in its active position that triggers co-repressor release and co-activator recruitment. It is worth noting that the LBD-interacting surface with NR-box overlaps the co-repressor one [22]. Despite this atomistic information available for interacting peptides bound to RAR, the mechanism by which RAR in the context of a heterodimer with RXR recognizes co-regulator fragments harbouring multiple boxes is still poorly understood. Moreover, the contribution of RXR in the recruitment of co-activators is still under debate. Some structural and functional studies argue in favour of a binding of the NR-boxes to both subunits (deck model) [20,25,30–32], whereas other studies claim that the binding of the co-activator is exclusively to RAR (asymmetric model) [24,33]. On the side of co-repressors, a complete structural study on the interaction of N-CoR NRID with RXR/RAR allowed us to show an equilibrium between a major population of asymmetric binding of the co-repressor to RXR/RAR

and a minor population of doubly bound N-CoR in which both CoRNR-boxes simultaneously interact with the heterodimer, accounting for the cooperativity of the interaction [18]. The binding mode of co-activators remains unclear due to the inherent disorder of NRIDs and the presence of several similar (or equivalent) motifs within this domain even if affinities toward RAR and RXR vary between the different NR-boxes [25]. The presence of a variable number of boxes in NRID of co-regulator proteins found in different eukaryotes might suggests that the disordered part acts as an entropic chain in order to have an efficient cooperative binding to heterodimeric NRs [34].

Recently, TIF2 NRID (residues 624 to 823) was assigned by Nuclear Magnetic Resonance (NMR) in order to study its interaction with RXR/PPAR heterodimer [20]. By combining NMR and hydrogen/deuterium exchange mass spectrometry, the authors showed that all three NR-boxes interact with the heterodimer. However, the atomistic details of this interaction could not be unveiled. Thus, in order to gain insights into the molecular mechanism of NR regulation, we have structurally characterised the human co-activator TIF2 NRID fragment (from residues 624 to 773, hereafter named TIF2NRID) and its interaction with RXR/RAR heterodimer. This minimal fragment encompasses the three NR-boxes of the protein (Figure 1), and is still functional in terms of interaction with NRs [12,35]. To highlight its conformational preferences, we have used multiple NMR observables including backbone chemical shifts, 15N NMR relaxation parameters and 1DNH residual dipolar couplings (RDCs), in combination with other biophysical techniques and computational tools [36]. Altogether, the results are consistent with a largely disordered protein containing multiple transiently structured elements. In addition to the three NR-boxes, supplementary upstream and/or downstream regions with conformational preferences are identified within TIF2NRID. We show that these flanking regions play a relevant role in the interaction with the heterodimer. The crystallographic structures of RAR and RXR with an extended peptide including NR-box2 and the downstream flanking region unveil a mechanism of specificity in the interaction of co-activators with NRs. Finally, NMR interaction experiments show that the three TIF2NRID NR-boxes are involved in the interaction through multisite binding mechanism enabled by the inherent flexibility within the co-regulator. Our study provides new insights into the structural bases of the multifaceted transcriptional regulation mechanism of NRs by co-regulators.

RESULTS
TIF2NRID is an intrinsically disordered protein with partially structured elements

We produced a fragment of the co-activator TIF2 spanning from residues Glu624 to residue Thr773. This construct encompasses the three nuclear receptor binding motifs, NR-box1 (from 641 to 645), NR-box2 (from 690 to 694) and NR-box3 (from 745 to 749) (Figure 1, top panel). *Biophysical characterisation of TIF2NRID.* The recombinant protein TIF2NRID elutes from a SEC column as a single peak (Figure 1A) at a volume that corresponds to an apparent molar mass of 45 kDa according to the molecular mass standards. However, the mass derived from MALS analysis is 16.4 kDa (± 0.5%), in agreement with the expected theoretical mass of 16.6 kDa, demonstrating that TIF2NRID is monomeric in solution. The elution volume is smaller than that of a globular protein of the same weight, a feature that is typical of either a folded protein with an elongated shape or a disordered one [40]. The low backbone amide 1H chemical shifts dispersion in the 15N-HSQC spectrum of TIF2NID is also a signature of the disordered nature of

TIF2NRID (Figure 1B). Finally, Far-UV CD spectrum of TIF2NRID recorded at 283 K shows that the TIF2NRID adopts a random coil conformation with some secondary structure contributions with a minimum shifted above 200 nm rather than at 198 nm. Furthermore, a shoulder around 220 nm, which is a typical signature of helical regions, is also observed (Figure 1C) [41,42]. SAXS data were collected in order to probe the overall properties of TIF2NRID in solution. The SAXS profile of TIF2NRID (Figure S1A) and its Kratky representation (Figure S1B) with no clear maximum and a monotonic increase along the momentum transfer range, are typical of a disordered protein. Guinier's analysis of the initial part of the curve ($sRg < 1.3$; where s is the momentum transfer and R g is the radius of gyration) indicates that TIF2NRID has an Rg of 37.4 ± 0.09 Å (Figure S1C). This value is slightly larger than that expected for an IDP of 154 residues (Rg_{IDP} = 35.2 Å) confirming that TIF2NRID transiently adopts secondary structures that extend its overall shape [43,44]. Indeed, the presence of transiently populated secondary structures expands the overall size of the protein by restraining the conformational space sampled and hampering the chain to fold back. Altogether, these data confirm that TIF2NRID is essentially disordered although it contains low populations of structured elements.

Consistently with experimental data presented above, bioinformatics analyses of TIF2NRID sequence using various computational tools (Table S1) indicated that TIF2NRID is characterized by a high intrinsic disorder propensity and also revealed some structural features. First, the amino acid composition has a typical feature of IDPs (Figure S2A) [45–47]. Also, according to the charge-hydropathy plot, TIF2NRID lies on the boundary between globular and disordered proteins [48], suggesting the presence of some structuration in the protein (Figure S2B).

The structural and dynamic characterization of TIF2NRID, at the residue level by NMR highlights the importance NR-box flanking regions

In order to identify structural features at the residue level, bioinformatics (Figure S2) and NMR studies (Figure S3) of TIF2NID in solution were performed thanks to the almost complete peak assignment of the protein (93% of residues of its primary sequence: i.e. 133 of 143 expected NMR peaks of the 154 residues as the 10 prolines and the first residue G1 are not visible in a 15N-HSQC).

Bioinformatics characterisation of TIF2NRID. The degree of disorder for TIF2NRID was predicted using several servers as shown in Figure 2A; all predictors indicated that TIF2NRID is globally disordered except for regions encompassing the three NR-boxes (i.e. residues L641-L645, L690-L694 and L745-L749). In addition, secondary structure predictions highlighted the existence of helical secondary structure elements in the three NR-boxes, as expected from the literature [10,22,25,31,49], although with different boundaries (Figure 2B). Additional secondary structure elements were also predicted for NR-box2 and NR-box3 flanking regions (P700-A710 and E727-K739, respectively). Interestingly, regions predicted to have a higher tendency of being ordered/structured were found to be evolutionary conserved, according to the GREMLIM sequence conservation prediction (Figure 2C). Overall, bioinformatics analyses converge in showing that TIF2NRID is an IDP with partially structured regions, the three NR-boxes and their flanking regions, that in addition show evolutionary conservation. These characteristics suggest an important functional role for these regions.

Secondary structure propensity and dynamic features of TIF2NRID. Quantitative analysis of the secondary structure populations using the secondary structure propensity (SSP) algorithm

[50] and backbone secondary chemical shift exploration of TIF2NRID shown in Figure 2D and Figure S4, respectively, are consistent with a mostly random coil polypeptide chain [51], and report multiple weakly populated segments containing secondary structures (see below). Further evidence for transient secondary structure elements came from 3JHNHA scalar coupling that strongly correlate with backbone dihedral Φ angles [52]. These couplings (Figure S5) are consistent with an overall unstructured protein (65% of the residues) and smaller proportions of helical type (34%) and extended beta (1%) conformations. Regarding TIF2NRID backbone dynamics [53], the heteronuclear NOEs (Figure S6A) display the typical bell shape profile of an IDP with small but positive values (average of 0.21 +/- 0.02), with the exception of the highly flexible N- and C-termini, which present negative values. Moreover, 15N longitudinal (T1) profile is relatively smooth along the sequence with an average value of 572 +/- 20 ms (Figure S6B), whereas more variability is observed for the transversal (T2) relaxation times (Figure S6C) and, consequently, for T1/T2 ratios (Figure 2E), with observed average values of 219.2 ± 20 ms and 2.88 +/- 0.15, respectively. Interestingly, these observations indicate restricted mobility in regions identified as partially structured by chemical shifts and bioinformatics analyses. In practice, the region G680-S697, which contains NR-box2, and the region E741-D750, which contains NR-box3, display helical propensities (< 15% and < 25%, respectively). For NR-box1, the helical tendency is shifted upstream of the LxxLL motif. Remarkably, the highest helix propensity appears for the flanking region (P700-A710) downstream of the NR-box2 (named α-helical flanking region, see bellow). Interestingly, residues in the three NR-boxes and the upstream and downstream NR-box2 α-helical flanking regions present large T1/T2 values suggesting chemical exchange phenomena due to conformational fluctuations on the μs-ms time-scale. In addition, SSP analysis highlights extended beta structures at TIF2NRID C-terminus (G758-T774) and upstream NR-box3 (E727-K739).

Ensemble description. To further study the presence of transient structured regions in TIF2NRID, we analysed 1DHN RDCs measured in two different media (alcohol mixture and filamentous Pf1). While 1DHN RDCs measured in alcohol mixture (Figure 2F, black line) are mainly negative, as normally observed in disordered proteins, segment-encompassing NR-box3 displays large positive values. As we found the same features for RDCs measured in Pf1, which does not interact with the protein, we concluded that the helical propensity in the NR-box3 fragment was present in solution but was slightly amplified by the alcohol alignment medium, which displays very small perturbations in the chemical shifts (0.04-0.05 ppm) only in this specific region. To precisely localize transient conformations present in TIF2NRID, we evaluated deviations of the experimental RDC profile from that of a fully disordered model computed from 100,000 conformations built with Flexible-Meccano (FM) [54,55], (Figure 2F, blue line). Several regions presenting disagreement between experimental and backcalculated RDCs were identified: the NR-box3 (L745-L749) and its flanking regions (E727-K739), the C-terminal region (G758-T774) and, to a lower extent, the central NR-box2 plus the subsequent α- helical flanking region (L683-L706). Most of these regions were predicted to adopt secondary structures according to bioinformatics and SSP analysis of TIF2NRID (see above), and thus were used to build structurally biased ensembles of 100,000 conformers by imposing previously predicted/determined secondary structure populations. The quality of these ensembles was evaluated by comparing their back-calculated RDCs with the experimental ones (Figure 2F, red and black lines, respectively). The best agreement was found by imposing: 15%, 7% and 70% of α-helix in L642-L645 (NR-box1), L683-

L694 (upstream extended NR-box2) and A743-L749 (upstream slightly extended NR-box3) segments, respectively; and 15% and 30% of β-strand in V728-K731 (up) and G758-D770 (downstream NR-box3) segments, respectively. Additionally, 25%, 40% and 20% α-turn torsion angles for residues K736-G737, A704-K705 (in the α-helical flanking region) and S716-Q717, respectively, were also imposed. Introduction of these secondary structure preferences resulted in a much better description of the entire chain. Furthermore, SAXS was used to characterize the conformational space sampled by TIF2NRID in solution and to validate the structure ensemble derived from the RDC data. After adding side-chains to 2,000 conformers from the RDC-refined ensemble, the average theoretical SAXS profile was calculated and compared with the experimental one (represented by black lines in Figure S1A). An excellent agreement between the two curves was observed ($\chi 2 = 1.38$), which notably improved that obtained from a pure random coil model ($\chi 2 = 1.57$) (data not shown).

Intramolecular long-range contacts. Finally, the presence of long-range contacts in TIF2NRID was explored using Paramagnetic Relaxation Enhancement (PRE) NMR experiments. As TIF2NRID does not contain any native cysteine, Ser716, which lies just after the partially structured flanking region at the C-terminus of NR-box2 (α-helical flanking region), was mutated to a cysteine (S716C) and subsequently labelled with a paramagnetic probe (nitroxide radical MTSL) to measure residue-specific PRE ratios (Ipara/Idia) (Figure 2G). Excluding residues directly adjacent to C716, the mapping shows a significant intensity reduction (Ipara/Idia < 0.50) for the region encompassing the NR-box 2 until NR-box 3, indicating the presence of extensive long-range contacts in the C-terminus of TIF2NRID. Interestingly, PREs measured in the segment containing the NR-box2 and its downstream flanking region (shaded in light blue) display a bell-shape with reduced PRE ratios with respect to the random coil profile (Figure 2G, blue curve). Indeed, the reduction of PRE ratios is accentuated towards the Nterminal part of NR-box2 and the C-terminal part of the flanking helix where the probe is localized. Thus, this bell-shape suggests a transient interaction between those two helical regions (L690-L694, NR-box2 and P700-A710, α-helical flanking region) in the free state.

Interaction of TIF2NRID with RXR/RAR heterodimer

Ligand modulation of the TIF2 affinity for RXR/RAR followed by steady-state fluorescence anisotropy. Firstly, we measured the affinity of TIF2NRID for the heterodimer RXR/RAR in different liganded states (Figure 3 and Figure S7). In the absence of ligand, the affinity is low (KD = 1.80 µM) but higher than the one measured with the equivalent fragment of the co-activator SRC1 [25] (KD = 4.80 µM), suggesting a preference of the heterodimer for the co-activator TIF2. The addition of RAR or RXR agonists (AM580 and CD3254, respectively) significantly increases the affinity of the liganded heterodimer for TIF2NRID, although more efficiently for the RAR agonist. Similar measurements done in the presence of RAR and RXR antagonists (BMS614 and UVI3003, respectively) confirmed the predominant role of RAR in the interaction with the co-activator. In fact, the simultaneous addition of a RAR agonist (AM580) and RXR antagonist (UVI3003) still allows a strong interaction of the liganded heterodimer with TIF2NRID (KD = 0.2 µM). On the contrary, the affinity of the heterodimer for TIF2NRID in the presence of a RAR antagonist (BMS614) and RXR agonist is nearly equal to that of unliganded state. Finally, a two-site interaction is observed in the presence of both agonists, suggesting that at least two interaction motifs of TIF2NRID can be simultaneously involved in the interaction with the fully activated heterodimer.

NMR exploration of the TIF2:RAR/RXR complex. To map TIF2NRID regions affected by the interaction with the heterodimer RXR/RAR (Figure 4), 15N-TIF2NRID was incubated with unlabelled RXR/RAR (RXR/RAR:TIF2NRID molar ratio of 1.1:1) using concentrations around the KD of the interaction measured by fluorescence anisotropy (Figure 3). Comparison of 15N-HSQC spectra of 15N-TIF2NRID with and without the apo RXR/RAR LBD heterodimer (Figure 4A) shows that addition of RXR/RAR LBD resulted in the attenuation or disappearance of several peaks, confirming the interaction of TIF2NRID with the heterodimer. Although to a different extent, a clear decrease of peak intensity appears in the 15N-HSQC of TIF2NRID upon binding to RXR/RAR for residues within the three NRboxes and also for their flanking residues. Attenuation or disappearance of peaks in the interacting region is generally attributed to an increase in local rotational correlation time (τc) for the attached region to a larger partner; or indicates a certain degree of flexibility of the chain upon binding, forming a "fuzzy" interaction [56]. More precisely, the decrease in intensity and thus interaction is localized around the NR-box1 (T639-K648) and NR-box3 that extends towards its C-terminal flanking region (K739-T763). Notably the NR-box2 binding region extends towards both flanking regions (spanning from G680 to E718), especially the downstream one (P700-A710, α-helical flanking region) (Figure 4B). Indeed, the strongest attenuation with peaks vanishing from the spectrum happens for residues within NR-box2 and its flanking regions, (Figure 4A), in accordance with higher affinity of RAR for NRbox2 of co-activator [25]. Note that these regions involved in the interaction were identified by NMR relaxation and secondary structure analysis as transiently structured with chemical exchange phenomena in unbound TIF2NRID. These results clearly demonstrate that all three LxxLL motifs are involved in the interaction with the heterodimer, highlight the leading role of the NR-box2 and the participation of several flanking regions in the heterodimer recognition.

To monitor the effect of ligands on the interaction, RAR and RXR agonists (AM580 and CD3254, respectively) were sequentially added to RXR/RAR:TIF2NRID sample and changes in TIF2NRID peak intensities in 15N-HSQC spectra were monitored (Figure 4C and 4D). As shown in Figure 4C, addition of RAR agonist induces a slight decrease of peak intensity relatively to the apo complex for residues belonging to NR-boxes 1 and 3 but more significantly for their flanking residues. The strongest attenuation appears for NR-box3 flanking regions whereas NR-box2 interacting region seems not to be further affected by addition of the ligand. We observed more accentuated changes in intensity after addition of the second ligand (RXR agonist) (Figure 4D). These data indicate that the presence of agonists into the two subunits of the heterodimer strengthens its interaction with TIF2NRID involving the different NR-boxes and their flanking regions. Therefore, NMR data are in agreement with affinity measurements (Figure 3A), showing that the presence of RAR agonist leads to a 9-fold decrease of the KD value with respect to the apo form and it is further increased, although moderately, in the additional presence of RXR agonist.

Binding mode of TIF2 to RXR/RAR heterodimer. To further understand the role of the three individual NR-boxes in the interaction of TIF2NRID with RXR/RAR, NMR interaction experiments were performed for three TIF2NRID LL/AA double variants in which the LxxLL motifs of NR-boxes-1, -2 and -3 were individually mutated into LLQA644A645, LHRA693A694 and LLRYA748A749, respectively. Note that these mutations strongly diminish the interaction of the individual boxes with the heterodimer [57]. The affinities of each TIF2NRID variant for the RXR/RAR heterodimer in the absence of ligand were measured by fluorescence anisotropy (Figure 3B and Figure S7).

Interestingly, no significant changes with respect to the wild-type were observed, suggesting that the presence of only two NR-boxes in TIF2NRID is sufficient for the interaction with the heterodimer. Similarly, NMR intensity profiles demonstrated that even if one interaction box is suppressed by the double mutation, the two other ones take over the interaction (Figure 5). For instance, NR-box1 variant showed that while the mutated NR-box1 was not interacting with the heterodimer, the entire region encompassing NR-box2 and NRbox3 presents a strong decrease of peak intensities (Figure 5A). An identical behaviour was observed when mutating NR-box3 (Figure 5C). For NR-box2 variant, NR-box1 takes over less clearly even if we still see a decrease in intensity that confirms its involvement (Figure 5B). These results substantiate the contribution of the three NR-boxes in the interaction with RXR/RAR heterodimer with a dominant role of NR-box2.

Finally, in order to further disentangle the interacting mode of the three NR-boxes and their flanking regions with the heterodimer, we performed NMR mapping interaction of TIF2NRID and its three variants with RAR and RXR separately (Figure S8). In the absence of agonist, the interacting profiles of TIF2NRID WT and variants are the same with RAR LBD alone, whereas they show no interaction with the RXR LBD. Addition of specific agonist towards RAR LBD domain increases the NR-box3 binding to RAR for WT or variants containing the LxxLL motif in NR-box3, whereas RXR agonist towards RXR LBD domain increases the binding of TIF2NRID WT in a non-specific manner, but no interaction is detected if one of the NR-box is suppressed by mutation. These results reinforce the idea that the specific and dominant interaction of NR-box2 towards RAR is necessary to properly drive the interaction.

A new interaction motif of TIF2NRID with RAR that could account for NR specificities in the different co-activator paralogues.

NMR experiments have highlighted the involvement of NR-boxes flanking regions of TIF2NRID in its interaction with RXR/RAR heterodimer (Figure 4). Of particular interest, the region following NR-box2, spanning from residues 700 to 710, shows a high propensity to be ordered (Figure 2D) and a high sequence conservation (Figure 2C), suggesting a particular functional role. To evaluate the involvement of this sequence in the interaction with the receptors, we prepared crystals of both agonist-bound RAR and RXR LBDs in complex with a TIF2 peptide comprising the NR-box2 (NR2) and its C-terminal α-helical flanking region spanning from residue K686 to K713 (here after named Ext for extended part). Data collection and refinement statistics are summarized in Table S3. The 2.4 Å resolution crystal structure of RAR LBD in complex with AM580 and TIF2 NR2-Ext displays two complexes per asymmetric unit, both showing that the interaction of the LHRLL helix of NR-box2 (named α1) with the co-activator binding groove of the receptor is preserved and that the extension folds as a long α-helix (residues P700 to T711, named α2) making hydrophobic contacts with residues from both helix α1 and RAR (Figure 6A and Figure S9). In particular, L703, L706, T707, E709 and A710 lie on the same side of α2 and make hydrophobic interactions with I689, R692 and L693 of NR2 α1, and with T233, I237 and L409 from RAR. In addition, a weak electrostatic interaction (4.0 Å in length) between R692 (α1) and E709 (α2) may stabilise further the positioning of helix α2 (Figure 6B). Interestingly, using fluorescence anisotropy we were able to show that the TIF2 NR2-Ext peptide has a three-fold higher affinity for the agonist-bound RAR LBD compared to the shorter TIF2 NR2 peptide (Figure 6C), confirming the involvement of α2 in the interaction between TIF2 and RAR. Conversely, similar experiments with RXR suggested that α2 plays no or very little role in the interaction with this receptor as the short and longer versions of TIF2 peptides display very similar affinities (Figure 6C). Competition

experiments of each unlabelled peptide against a fluorescently labelled SRC1 NR2 peptide allowed us to confirm that RAR interacts more avidly with TIF2 NR2-Ext whereas RXR shows no preference (Figure S10). Next, we solved the crystal structure of RXR LBD in complex with the TIF2 NR2-Ext peptide and the agonist LG100268 at a resolution of 2.8 Å (Figure 6D). The structure shows two molecules of liganded-RXR LBD per asymmetric unit, each bound to one TIF2 NR2-Ext peptide. Whereas in both complexes TIF2 NR2 (α1) interacts classically with RXR, the extension is either devoid of regular secondary structure elements (Figure 6D) or too disordered to be visible (not shown), confirming that it is not involved in the interaction with RXR. The superposition of RAR and RXR LBD structures reveals that steric hindrance generated by the two large aromatic RXR residues F277 and F450 (corresponding to I237 and L409 in RAR) could account for the differential involvement of the extension in the interaction with the two receptors (Figure S11). These structural data further substantiate the transient nature of the helical folding of TIF2 NR2 α2, which adopts a helical conformation in complex with RAR but is fully disordered in complex with RXR.

Having shown the structural properties and functional role of the N-terminal extension of NR2 in TIF2, we next examined whether this feature is unique to TIF2 (NCOA2) or conserved in other coactivator paralogues, namely SRC1 (NCOA1) and RAC3 (NCOA3). We first collected sequences for TIF2, SRC1 and RAC3 of different vertebrate species, from lampreys to humans (Table S4), aligned them and analysed their evolutionary relationships by phylogenetic tree reconstruction (Figure S12). The phylogenetic tree revealed that the vertebrate NCOA paralogues fall into two distinct groups: (i) SRC1 and (ii) TIF2 plus RAC3. Furthermore, within the TIF2 plus RAC3 group, the three representatives of the sea lamprey Petromyzon marinus, a jawless vertebrate, formed a separate branch at the base of the clade, which is indicative of an origin by lineage-specific gene duplication. To evaluate, whether the vertebrate NCOA genes have been under positive, neutral or negative selection, we used the full-length NCOA sequences to calculate the substitution rates at non-synonymous and synonymous sites. The resulting dN/dS ratio (or ω value) was well below 1 for all three NCOA paralogues (Figure S12), suggesting that SRC1, TIF2 and RAC3 are under negative (i.e. purifying) selection. We next focused our attention on the TIF2 NR2-Ext sequences and corresponding regions in SRC1 and RAC3. Sequence logos calculated for the three peptides revealed specific evolutionary signatures for each of the NCOA paralogues, with the SRC1 NR2-Ext peptide seemingly less conserved than the corresponding domains of TIF2 and RAC3, a result that was coherent with the phylogenetic tree analysis (Figure 7A and Figure S12). Similar to the results obtained for the full-length NCOA sequences, the dN/dS ratio analysis for the three NR2-Ext domains revealed, once again, a very strong negative (i.e. purifying) selection. However, when comparing the ω values for each of the three NR2-Ext paralogues, we noticed that the ω value for TIF2 was more than three times higher than those for SRC1 and RAC3 (Figure 7A). This suggests that, in evolutionary terms, TIF2 NR2-Ext is subjected to weaker purifying selection than the corresponding domains of its two paralogues, SRC1 and RAC3. Taken together, our sequence analyses are consistent with a conserved role for the NR2-Ext peptides in TIF2 and RAC3, but not in SRC1. To test this hypothesis, we solved the crystal structure of the RXR/RAR LBDs heterodimer bound to RAR and RXR specific agonists and to the SRC1 NR2-Ext peptide (Figure 7B). As expected, the structure shows that the peptide interacts classically with both receptors through the helical LxxLL motif (α1), while the rest of the sequence remains highly mobile with no interpretable electron density on both sides of the dimer. In line with these data,

10

fluorescence anisotropy measurements revealed no noticeable difference in the affinities of SRC1 NR2 and SRC1 NR2-Ext peptides for RAR (Figure 7C). As a whole these data support the notion that NR box flanking regions may play different roles and may drive NR specificities in the different co-activator paralogues.

DISCUSSION

Recruitment of co-activators is a critical step in ligand-dependent activation of gene transcription by NRs but little is known structurally about their interactions with NRs and their dynamic properties. In this study, we have first provided a complete structural characterization of the NRID of the co-activator TIF2 (TIF2NRID). Combination of solution-state structural methods and biophysical experiments show that, as suggested by its amino acid sequence and in accordance with a previous study [20], TIF2NRID behaves as an intrinsically disordered protein (IDP) with multiple transiently structured elements. The large conformational heterogeneity of IDPs usually makes their detailed structural characterization challenging, since it requires an ensemble representation to appropriately describe averaged experimental observables [58]. The combination of different high-resolution NMR parameters extracted from TIF2NRID, such as chemical shifts, RDCs and 15N-relaxation, as well as low resolution information provided by SAXS, together with computationally generated conformational ensembles allowed to precisely highlight functional regions with conformational preferences and rigidity that significantly deviate from pure random-coil. In addition to the three evolutionary conserved NR-boxes that adopt various degrees of pre-structured helices when free in solution (Figure 2), we have identified previously undescribed regions flanking both sides of NR-box2 and NR-box3 with preferential conformations and motional restriction. Notably, the upstream (T681-I689) and downstream (V701-A710, α2) NR2 flanking regions present a significant degree of evolutionary conservation (Figure 2C) that could anticipate a functional role of these regions.

The interaction mode of the LxxLL motif with NRs is known for some time and has been described in many structural studies [20,22,25,30,31,41,49,59–64]. Moreover, several functional and biochemical analyses of the interaction of NRs with different mutated or chimeric LxxLL-containing peptides from TIF2 or SRC1 have suggested that residues immediately flanking the core motif could play important roles in the affinity and, in NR specificity [10,65–68]. This phenomenon has been observed in other IDP interacting systems [69,70,71]. In this line, our structural and biophysical interaction studies highlighted the participation of the pre-structured flanking regions in the interaction of TIF2 with the RXR/RAR LBD heterodimer and gave new insights into the role of these less well-characterized stretches in the specificity of the interaction between co-activators and NRs. More specifically, NMR data showed that the recruitment of TIF2NRID by the RXR/RAR LBD heterodimer is preferentially driven by the interaction between RAR and NR-box2 and its downstream flanking region, α2 (Figure 4B and Figure S8). These data were fully confirmed by a crystallographic analysis revealing that the TIF2 NR2-Ext peptide adopts a helix-turn-helix conformation and that both helices α1 (LxxLL core motif) and α2 (C-terminal flanking region) are involved in the interaction with RAR (Figure 6). Interestingly, the PRE data also suggested that the two helical elements found in the crystallographic structure present transient contacts in the unbound form (Figure 2G), possibly reducing the entropic cost of the recognition event [72,73]. Further sequence, structural and interaction analyses revealed that, by bringing a few additional contacts with RAR, α2 confers a three times higher affinity to the NR-

box2-Ext for this receptor, whereas it has no effect on the interaction with RXR. Furthermore, they also showed that the mechanism described here for TIF2, a member of the NCOA2 family of co-activators, is very likely to apply to members of the NCOA3 family, but not to the NCOA1 family members as shown in the present study with SRC1. As a whole, these results provide a structural basis supporting the notion that NR-box flanking regions contribute to the specificity of coactivators for retinoid receptors (RARs and RXRs), even though the modulatory role of the α2 flanking region does not reach the level observed in other systems [71]. Further analyses are required to address the question whether or not this mechanism may apply to a broader range of coactivators and nuclear receptors.

The interplay of the different NR-boxes in the interaction with NR dimers is also a matter of debate. Two models of the interaction between co-activators and NR heterodimers have been previously proposed. In the first one, the so called "asymmetric model", TIF2 would be recruited preferentially by RAR, most likely using NR-box2, which presents the highest affinity [24,33]. In the second one, the "deck model", two NR-boxes can simultaneously recognise the two members of the heterodimer, providing a cooperative binding mechanism [20,25,30–32]. Our results unambiguously demonstrate that the three NR-boxes can interact with the heterodimer, in line with the interaction of TIF2 with RXR/PPAR [20]. The absence of kinetic measurements precludes the assessment of the molecular steps leading TIF2 to the complex formation. However, the stronger binding observed for NR-box2 suggests that TIF2NRID is initially recruited by the liganded heterodimer through the NR-box2 and its downstream flanking region (α2) into the hydrophobic groove generated on RAR. This anchoring point restricts the exploration of the conformational space facilitating interaction of the NR-box3 or NR-box1 with RXR. Although this seems the most reasonable mechanism, we cannot discard that the other NRboxes could initially recognize RAR to subsequently evolve to the final thermodynamically most stable conformational state. Indeed, this first recognition step, given that the three NR-boxes can interact with RAR, might present allovalency [74] as previously described for other disordered systems, such as the multiple phosphorylable sites of Sic1 in interaction with Cdc4 [75], and the multiple FG motifs of Nups in interaction with transport factors (TFs) [76]. In the thermodynamic equilibrium, the ensemble of our data suggests the co-existence of multiple forms of the complex in which singly- and doubly-bound TIF2 conformations involving the three NR-boxes are in continuous exchange. The specific population of these configurations will depend of the individual affinities. This interaction model would resemble the "shuffle complex" proposed for the NHE1/ERK2 [77] and GCN4/Med15 [78] complexes. Although our observations could also be compatible with the previously mentioned allovalency model in which RAR acts as the only binding partner of the individual NR-boxes, we privilege the shuffle complex model due to the observed changes in the complex in the presence of ligands. Concretely, the addition of either the RXR or the RAR agonist results in a decrease in the KD, which is further decreased when both agonists are added simultaneously, indicating that both RAR and RXR binding sites are active. Note that the addition of ligands, which enhance the individual interaction affinities, most probably reduces the number of possible configurations and doubly-bound forms become preeminent. Then, we used TIF2NRID LxxAA variants to more precisely access to the role of each NR-box and its flanking regions. Our results suggest as well that two NR-boxes are sufficient to trigger a double-site binding interaction with RXR/RAR. Indeed, differences in the intensities of the individual NR-boxes upon mutations as well as in the length of the entropic chain between interacting NR-box pairs most likely account for this

observation. Indeed, the sequence distance between two interacting sites in a disordered chain decreases the effective concentration and inevitably affects cooperativity [79]. This observation advocates for a binding mode encompassing two neighbouring NR-boxes and where the NR-box2 is still the primary driver of the interaction.

In summary, our study delineates the structural bases of the regulation of gene transcription by RAR and highlights the importance of disorder in this process. The presence of multiple NR interaction boxes in co-regulator proteins provides a two-site binding mode to the complex enabling the capacity to transit from an active to an inactive state or vice versa. Although the mechanisms observed for TIF2 and RAR seem to be general in NRs, specific structural features in the LBDs, co-regulator boxes and associated flanking regions most probably define the efficiency and selectivity of gene transcription.

MATERIALS AND METHODS
TIF2NRID expression, purification and preparation
The TIF2NRID polypeptide studied here corresponds to the NRID fragment of the human TIF2 coactivator (Uniprot entry Q15596-1) ranging from residue 624 to residue 773. Four extra residues (GPHM) at the N-terminus are remnants from the His-tag 3C protease cleavage site (G and P residues) and the NdeI restriction site (H and M residues) chosen for cloning the construct. TIF2NRID recombinant protein is composed of 154 residues with a theoretical molecular mass of 16.6 kDa.

The DNA encoding human TIF2NRID, optimized for bacterial expression (purchased to Integrated DNA Technologies) was cloned (InFusion cloning kit) into a pDB vector between the NdeI and XhoI restriction sites. pDB-TIF2NRID construct codes for a protein containing a hexa histidine-tag and a HRV 3C cleavage site (Leu-Glu-Val-Leu-Phe-Gln/Gly-Pro) at the N terminus, where a specific cleavage occurs between Gln and Gly. Uniformly 15N- or 13C/15N-labelled recombinant TIF2NRID was expressed in *E. coli* BL21 (DE3), grown in 2 L of LB containing 50 µg/mL of kanamycin at 37°C, harvested at an OD600nm of 3 and resuspended in 2 L of M9 minimal medium with isotopic enrichment (1 g/L of 15NH4Cl and/or 3 g/L of 13C6-glucose, Cambridge Isotope Laboratories). After 30-45 min, overexpression was induced with 0.5 mM IPTG for 4 h at 37°C. Cells were harvested and resuspended in 300 mM NaCl, 20 mM Tris pH 7.5, 2 mM DTT (buffer A), plus one tablet of Complete EDTA free protease inhibitor cocktail and sonicated on ice. The lysate was centrifuged at 18,000 x g at 4°C for 20 min to remove inclusion bodies. The supernatant was clarified by a 5 µm filter followed by a 0.45 µm filter and loaded on a cOmplete His-Tag purification Column (Roche) with a column volume (CV) of 1 mL. The column was washed and equilibrated with buffer A and the protein was eluted with a linear gradient using buffer A supplemented with 300 mM of imidazole in 25 CV. Fractions containing the protein (monitored by SDS-PAGE) were pooled, incubated with GST-tagged 3C protease (at a protease-to-target protein ratio of 1:100 (w/w), to cleave the His-tag), and at the same time dialyzed overnight in buffer A supplemented with 10 mM imidazole at 4°C. To remove proteases and not digested proteins, the pool was loaded in 0.5 mL GST resin mixed with 1 mL of cOmplete His-tag purification resin (Roche) stacked in a 12 mL polyprep column. Resin was washed and equilibrated with 3 CV of buffer A supplemented with 10 mM imidazole. The flow through was collected, concentrated and loaded in a HiLoad 16/60 Superdex 75 PG (GE Healthcare) equilibrated with 150 mM NaCl, 50 mM BisTris pH 6.8, and 0.5 mM EDTA (NMR buffer) at 1 mL/min. The protein eluted

after 53 mL of run approximately. The purity of the sample was evaluated by SDS-PAGE analysis. The pure fractions containing proteins were pooled and concentrated by centrifugation using a Vivaspin™ protein concentrator with a 5 kDa cut-off. Protein concentration was determined by UV absorbance at 280 nm using the extinction coefficient calculated from the amino acid composition (1,490 M-1.cm-1) and by refractometry (0.1814 mL/g as refractive index according to the software SEDFIT (Schuck, 2000). The yield of pure doubly labelled protein was around 2.5 mg per litter of bacterial culture. TIF2NRID was stable in solution for months and resisted to un/freezing without detecting signs of degradation. NMR spectra at low (5 µM) and high (350 µM) concentrations are superimposable, supporting a homogeneous state of TIF2NRID that is independent of the concentration.

For paramagnetic relaxation enhancement experiments (PRE), single-cysteine variant (S716C) of TIF2NRID (corresponding gene purchased to Integrated DNA Technologies) was expressed and purified as the wild type. Paramagnetic spin labelling was carried out by incubating the variant with 20-fold molar excess of nitroxide spin label MTSL (Toronto Research Medical) solubilized in acetonitrile. To remove any trace of DTT, the protein solution was passed through a ZebaTM Spin desalting column (7K) equilibrated in NMR buffer pH 7.5 before labelling. After incubation at 4°C overnight, excess of free MTSL was removed by passage through a ZebaTM Spin desalting column (7K) equilibrated with NMR buffer pH 6.8. The yield of labelling close to 100% was estimated with DTNB assays (Ellman's reagent from ThermoFisher Scientific).

Preparation of LXXAA variants of TIF2NRID

The motifs in the three TIF2NRID NR-boxes (LxxLL) were double mutated into LxxAA variants. Each variant was carried out by PCR QuikChangeTM Site-Directed Mutagenesis reactions from AgilentTM. A pair of complementary oligonucleotides carrying the desired double mutation has been ordered from Integrated DNA TechnologiesTM. The mutated motifs LLQA644A645, LHRA693A694 and LLRYA748A749, were respectively obtained for NR-boxes 1, 2 and 3. All the variants were isolated and characterized by sequencing. These variants were purified following the same protocol as wild-type TIF2NRID.

RXR/RAR expression and purification

The RAR LBD studied here corresponds to a 246-residue long fragment (~ 27.8 kDa) of the RARα protein (50.8 kDa full-length protein amino acid sequence as described in Uniprot entry P10276-1) coded by the human RARA gene. The RXR LBD studied here corresponds to a 241-residue long fragment (~ 26.8 kDa) of the RXRα protein (51.2 kDa full-length protein amino acid sequence as described in Uniprot entry P28700-1) coded by the mouse RXRA gene. Consequently, the RXR/RAR LBD recombinant heterodimer has a theoretical molecular mass around 54.6 kDa.

The hexa histidine-tagged LBD of human RARα LBD (residues 176-421 in a pDB-His 3C vector) and mouse RXRα LBD (residues 227-467 in a pET-3a vector) were expressed in *E. coli* BL21(DE3) and copurified. The heterodimer RXR/RAR was prepared as previously described in [25] with some modifications. Briefly, after a nickel affinity column (5 mL His Trap FF, GE Lifescience), fractions containing RXR/RAR heterodimer were pooled with 3C to remove the His-tag and dialyzed overnight at 4°C against NMR buffer. The sample was further purified using a Superdex 75 26/60 gel filtration column (Amersham Biosciences) equilibrated with NMR buffer. The peak fractions were analysed by SDS-PAGE and the fractions containing pure RXR/RAR heterodimer were pooled. Protein concentration was determined by UV absorbance at 280 nm using the extinction coefficient calculated from the amino acid composition (28,045 M-1.cm-1)

and by refractometry (0.188 mL/g as refractive index according to the software SEDFIT [80]. The affinity of the RXR/RAR heterodimer is expected to be in the nM range, like the closely related RXR/TR heterodimer (KD value of 65 nM [82])

Ligands and peptides

All the ligands (AM580, CD3254, LG100268, BMS614 and UVI3003) were purchased from Tocris Bioscience. All compound stock solutions were prepared at 10 mM in DMSO. The peptides TIF2 NR2 (KHKILHRLLQDSS and CTSLKEKHKILHRLLQDSS), TIF2 NR2-Ext (KHKILHRLLQDSSSPVDLAKLTAEATGK and CTSLKEKHKILHRLLQDSSSPVDLAKLTAEATGK), FITC-SRC1 NR2 (FITC-LTERHKILHRLLQEGSP), SRC1 NR2 (RHKILHRLLQEGS) and SRC1 NR2-Ext (RHKILHRLLQEGSPSDITTLSVEPDKK) were purchased from EZbiolab.

Bioinformatics sequence analyses

Amino acid composition of TIF2NRID was analysed using Composition Profiler Tool [83] that detects protein amino acid composition bias by comparing to the average amino acid frequencies reference values of a representative set of folded proteins from the PDB [84]. Uversky's plot (Charge-Hydropathy - CH) was generated using PONDR [85]. The CH border is described by $H=(C+1.151)/2.785$ [48].

Predictions of disorder in the TIF2NRID protein were computed using several tools: the metaPrDOS web server that integrates the results of eight different methods (PrDOS, DISOPRED2, DisEMBL, DISPROT, DISpro, IUpred, POODLE-S and DISOclus) [86]; the meta-predictor PONDR-FIT that combines the results of six different methods (PONDR-VLXT, PONDR-VSL2, PONDR-VL3, FoldIndex, IUPred, and TopIDP) [87]; PrDOS a structure-based method [88]; SPOT-disorder based on a windowbased neural network (SPINE-D) [89]; and finally DISOPRED3 that combines three machine learning models: support vector machine, neural network and nearest neighbour [90].

Secondary structure predictions were obtained using different servers: PSIPRED v3.3 [91]; Jpred4 [92] SOPMA [93]; PSSpred [94] and a α-turns predictor ALPHAPRED [95].

Sequence conservation was analysed using GREMLIN (Generative Regularized Models of proteins) software [96]. Multiple sequence alignment were generated using HHblits algorithm [96,97] (with an Evalue cut-off of 10-10 and sequences having > 75% gaps are filtered out), which performed eight sequence search iterations, obtaining 86 sequences of TIF2NRID homologs (Table S2).

SEC-MALS

Size-Exclusion Chromatography-Multi Angle Light Scattering (SEC-MALS) experiment was performed at 25°C using a Superdex 200 10/300 GL column (GE HealthCare) connected to a miniDAWN-TREOS light scattering detector and an Optilab T-rEX differential refractive index detector (Wyatt Technology, Santa Barbara, CA). The column was equilibrated with 0.1 μM filtered NMR buffer and the SEC-MALS system was calibrated with a sample of Bovine Serum Albumin (BSA) at 1 mg/mL. A sample of 40 μL of TIF2NRID protein at 3.3 mg/mL (200 μM) was injected at 0.5 mL/min. Data acquisition and analyses were performed using the ASTRA software (Wyatt). Based on measurement on BSA sample under the same or similar conditions, we have estimated an experimental error in molar mass around 5%.

Far-UV Circular Dichroism spectroscopy (CD)

Far-UV CD spectrum of a 1 mg/mL (60 μM) sample of TIF2NRID in NMR buffer was recorded on a Chirascan Plus spectropolarimeter using quartz cuvette (0.1 mm path length) at 10°C. Spectra were scanned from 185 to 250 nm with an increment of 0.2 nm, an integration time

of 2 s. The signal from the scan of the buffer was subtracted from the corresponding sample scan and signal was converted to mean residue ellipticity [deg.cm².dmol-1.res-1] using a mean residue weight of 104 Da.

Steady-state fluorescence anisotropy

For all the experiments, the buffer solution was 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 5 mM DTT and 10% (v/v) glycerol. The reported data are the average of three independent experiments.

Effect of ligands on the recruitment of TIF2NRID by the heterodimer RXR/RAR. TIF2NRID was aminelabelled with a fluorescent probe (fluorescein isothiocyanate, FITC, Sigma) according to the manufacturer's instructions. Assays were performed using a CLARIOstar microplate reader (BMG Labtech) with the excitation wavelength set at 485 nm and emission measured at 530 nm. Measurements were initiated at the highest concentration of protein (7 μM) and the protein sample was diluted successively with the buffer solution. For each point of the titration curve, the protein sample was mixed with 10 nM of fluorescent TIF2NRID and 21 μM of ligand (three molar equivalents).

Effect of LXXAA mutations on the interaction of TIF2NRID with the heterodimer RXR/RAR. Wild-type and the three LXXAA variants of TIF2NRID were amine-labelled with a fluorescent probe (Alexa 488 NHS Ester, ThermoFisher) according to the manufacturer's instructions. Measurement of binding affinities was performed using a Safire2 microplate reader (TECAN) with the excitation wavelength set at 470 nm and emission measured at 530 nm. Here the highest concentration of protein was 20 μM. For each point of the titration curve, the protein sample was mixed with 10 nM of fluorescent TIF2NRID (wt or variants).

Interaction of RAR LBD with TIF2 NR2 and TIF2 NR2-Ext peptides. The peptides were labelled with Alexa Fluor 488 C5 Maleimide (ThermoFisher) on the additional cysteine at the N-terminus and the labelled peptides were subsequently HPLC purified to separate them from free fluorophore. Measurement of binding affinities was performed using a Safire2 microplate reader (TECAN) with the excitation wavelength set at 470 nm and emission measured at 530 nm. The highest concentration of protein (RAR or RXR) was again 20 μM and the protein sample was diluted successively in the presence of two-fold molar excess of RAR or RXR ligands (AM580 and CD3254, respectively). For each point of the titration curve, the protein sample was mixed with 4 nM of fluorescent TIF2 peptide. For these three types of experiments, binding data were fitted using a sigmoidal dose-response model (GraphPad Prism, GraphPad Software).

Competition experiments. The protein (RAR at 0.5 μM or RXR at 5 μM) was mixed with a constant amount of FITC-SCR1 NR2 (4 nM) and two-fold molar excess of agonist ligands (AM580 and CD3254, respectively). Titrations were initiated at the highest concentration of unlabelled peptide (20 μM) and the peptide sample was diluted successively two-fold with the same buffer solution as mentioned above. Binding data were fitted using a One site competition model (GraphPad Prism, GraphPad software).

NMR spectroscopy data collection and analysis

All NMR experiments were recorded at 283 K on a Bruker Avance III spectrometer equipped with a 5 mm triple (TCI) resonance z-gradient cryoprobe operating at 18.8 T (800 MHz 1H Larmor frequency) and at 16.4 T (700 MHz 1H Larmor frequency). Spectra were processed using TopSpin NMR (Bruker) and analysed using Sparky [98]. 1H chemicals shifts were referenced

directly and 13C and 15N indirectly [99] with 2,2-dimethyl-2-silapentane-5-sulfonate (DSS, methyl 1H signal at 0.00 ppm).

Spectral assignment was done on a 350 µM of 15N/13C-TIF2NRID sample in NMR buffer. All NMR samples were supplemented with 5-10% (v/v) D2O. 1H, 15N, 13C O, 13Cα, 13Cβ were assigned using a two dimensional 15N-HSQC, and a set of three dimensional HNCO, HN(CA)CO, HN(CO)CACB, HNCACB, HN(CO)CA, HNCA experiments. A partial automatic assignment of the backbone resonances was obtained using the program MARS [100] and then completed manually. Resulting backbone chemical shift data are available on the Biological Magnetic Resonance Bank (BMRB accession code 50477).

NMR carbon chemical shifts are highly sensitive to backbone conformations [101,102], 13Cα and 13Cβ, and to a lower extent 1Hα, 13CO, 15N and 1HN secondary chemical shifts report on the secondary structural preferences at residue level. Secondary chemical shift (Δδ) values were calculated as the difference between the experimental chemical shifts and their amino-acid specific random-coil values. We used random coil chemical shifts from two different databases: one based on a set of pentadisordered peptides (Ac-QQXQQ-NH2) (referred as "Poulsen"; Poulsen IDP/IUP random coil chemical shifts) [103] and the other one based on a set of IDPs (referred as "Potenci") [104]. Both use nearest neighbour amino acid sequence corrections [51,105] and also include corrections for temperature and pH effects [103,104]. Secondary structure propensity (SSP) score was calculated using the webserver ncSPC [106] using the Cα and Cβ secondary chemical shifts. The program uses random coil reference data from ncIDP [51] and a algorithm for converting the measured data into secondary structure propensities [107]. For a given residue, a positive or negative SSP score indicate a propensity for helical or β-strand structures, respectively (with a score of 1 or -1 for a fully formed helical or β-strand structure).

Backbone amide 15N longitudinal (T1) and transverse (T2) relaxation times and heteronuclear 15N{1H} NOEs are good indicators of local backbone mobility on the ps-ns timescale [108,109], and T2 is in addition sensitive to exchange processes taking place on the µs to ms timescale. They were measured using the methods described in [108] using a 200 µM 15N-labelled TIF2NRID sample in NMR buffer. For 15N T1 and T2, a series of eight 15N-HSQC spectra with relaxation delays of 10, 50, 100, 200, 400, 600, 800 and 1000 ms for T1, and 16, 32, 64, 96, 160, 240, 480 and 640 ms for T2 were recorded. T1 and T2 spectra were recorded with 8 scans and 2.5 s of recycle delay, 2048 direct complex points and 256 indirect complex points with an echo/antiecho acquisition scheme. The intensity of the backbone amide signals was fitted to a single exponential decay using Sparky [110]. Signal overlap prevented the reliable measurement of the signal intensity for several residues, which were excluded from the analysis. 15N{1H} NOE values were obtained by recording two sets of spectra in the presence and in the absence of a 5 s proton saturation period. Heteronuclear NOE spectra were acquired with 16 scans, 2048 direct complex points and 256 indirect complex points. 15N{1H} NOEs were calculated from the ratio of intensities measured in the saturated (I) and unsaturated spectra (I0).

Three-bond HN-Hα J-coupling constants (3JHNHα) were extracted from the 3D HNHA experiment [52]. 3JHNHα values were obtained from intensity ratios of HN- Hα cross-peaks (Scross) and the corresponding HN-HN diagonal peaks (Sdiag) using the equation:

$$^3J_{HNH\alpha} = \frac{\arctan\sqrt{S_{cross}/S_{diag}}}{2\pi\zeta f}$$

where 2ζ, the total evolution time for the homonuclear 3JHNHα coupling, was set to 26.1 ms. The empirical scaling factor f, required to correct for differential T1 relaxation effects [52], was set to 1 since the apparent effective T1 of the Hα spin is long enough for the highly flexible TIF2NRID, which has an apparent correlation time of approximately 3.4 ns, as estimated from the T1/T2 ratio [111]. In all cases, errors in the measurement of intensities were estimated from the standard deviation of the experimental noise level.

Residual Dipolar Couplings (RDCs) are linked to the relative orientation of internuclear bond vectors with respect to the applied magnetic field so they are sensitive to both local and global conformations and are good indicator of presence of secondary structures in IDPs [44,112]. Two different alignment media have been used to achieve a partial alignment in the magnetic field: alcohol mixture, containing C8E5 glycol (polyoxyethylene 5 Octyl Ether, C18H36O6) and 1-octanol (r = 0.87) [113], and filamentous phages Pf1 [114–116] provided by ASLA biotech at an estimated concentration of 18 mg/mL previously dialyzed in NMR buffer. The homogeneity of the partially aligned samples was confirmed by the well shaped 2H doublet from the solvent [113]. A 2H quadrupolar splitting of 36.6 Hz was observed for the alcohol mixture and 10 Hz for the Pf1 phages. Both anisotropic and isotropic samples contained 250 μM of TIF2NRID in NMR buffer. RDC values (1DHN) were extracted from the differences between the 1JHN+1DHN couplings and the 1JHN scalar couplings measured using15N-HSQC-DSSE (In Phase Anti Phase IPAP) [117] on an anisotropic and isotropic sample, respectively (1DHN RDCs were corrected for the negative gyromagnetic ratio of 15N). Errors were estimated from the line-width and signal to noise ratio.

Ensembles of explicit models were generated for TIF2NRID using Flexible-Meccano, which sequentially builds peptide planes based on amino acid specific conformational propensity and a simple volume exclusion term [54,55]. The algorithm randomly selected torsion angle (,) pairs from a database of amino-acid-specific conformations present in loop regions of high-resolution x-ray structures. To account for deviations from a random-coil description of TIF2NRID, different structure ensembles were computed including user-defined local conformational propensities in different regions of the protein. Each ensemble comprised 100,000 conformers. 1DHN RDCs were calculated for each conformer of individual ensembles, and the per-residue averages over all conformers were computed and defined as back-calculated RDC values. Comparing the back-calculated and experimental values according to the equation assessed the quality of each ensemble:

$$\chi^2 = \frac{1}{N} \sum (\text{RDC}_{\text{exp}} - \kappa.\text{RDC}_{\text{back\_cal}})^2$$

where RDCexp and RDCback_cal are the experimental and back-calculated RDC values, respectively, and N is the number of analysed RDCs. All back-calculated RDCback_cal values were scaled uniformly by a factor κ to account for the uncertainty in the degree of sample alignment.

Paramagnetic Relaxation Enhancement (PRE) experiments of the MTSL labelled TIF2NRID S716C variant were performed to obtain long-range contact information, since the unpaired electron of a nitroxyl radical MTSL causes line broadening of the NMR signals in a distance-dependent manner up to 20 Å [118]. PRE measurements were performed using samples at a concentration of TIF2NRID S716C variant adjusted to 50 μM to minimize the effects of nonspecific intermolecular interactions. Reference diamagnetic samples were obtained after adding 5-fold molar excess of fresh ascorbic acid, with pH adjusted to 6.8, to the paramagnetic sample. PRE

effects were determined from the peak intensity ratios (Ipara/Idia) between two 15N-HSQC spectra of paramagnetic and diamagnetic samples. For both states, spectra were recorded using a recycling delay of 2 s and the same NMR parameters. No significant chemical shift perturbations were observed between the 15N-HSQC spectra of the unlabelled and the MTSL labelled protein in its diamagnetic form, meaning that the MTSL does not induce conformational changes in TIF2.

Interaction mapping was performed using unlabelled RXR/RAR LBDs, and NMR 15N-HSQCs were recorded for complex with 15N-TIF2NRID: RXR/RAR concentrations of 5 μM:6 μM (1:1.1). After this first experiment, AM580 RAR agonist was added to have a 1.2-fold molar excess of ligand. Subsequently, after the second experiment, CD3254 RXR agonist (1.2-fold molar excess) was added to the sample. All ligands were purchased in Tocris Bioscience and stock solutions were prepared at a concentration of 10-3 M in DMSO.

Small-angle X-ray Scattering (SAXS) measurement and analysis

Synchrotron radiation small-angle X-ray scattering (SAXS) data were acquired for TIF2NRID at the P12 beamline of the European Molecular Biology Laboratory (EMBL) at the storage ring PETRA-III at the Deutsches Elektronen-Synchrotron (DESY), Hamburg [119]. Using a Pilatus 2M detector at a sampleto-detector distance of 3.1 m and a wavelength of 0.124 nm, a momentum transfer range of 0.026 < s < 5.04 nm-1 was covered (s = 4π.sinθ/λ, where 2θ is the scattering angle). SAXS data were acquired for a dilution series (1.7, 5.2 and 7.1 mg/mL) measured with the robotic sample changer at 10°C. Radial averaging, frame averaging and buffer subtraction were done using standard protocols [120]. No concentration effects were observed between the three measured concentrations and curves from all measurements were merged to reduce interparticle interactions appearing at the beginning of the curve while preserving a good signal to noise ratio at the end and averaged curve was used for the subsequent structural analysis. The forward scattering intensity, I(0), and the radius of gyration, Rg, were evaluated using Guinier's approximation [121], assuming that at very small angles (s < 1.3/Rg), the intensity can be described as . The molecular mass of TIF2 I(s) = I(0)e — (sRg)² 3 NRID was evaluated by comparing its forward scattering intensity with that of a reference solution of Bovine Serum Albumin (MW = 66 kDa).

Simulated SAXS curves from RDC-derived ensemble models were calculated using CRYSOL [122] after addition of side chains using SCRWL [123] for each conformer. Then, the individual SAXS curves were averaged to obtain the ensemble scattering curve. Both curves were directly compared by scaling the theoretical curve using a momentum transfer range of 0.015 < s < 0.50 Å-1. Constant subtraction was used. A total of 2,000 conformers were used to calculate the average curve, which was sufficient to reach convergence.

Crystallization of the RARα and RXRα LBDs in complex with TIF2 NR2-Ext peptide

Expression and purification of the human RARα and mouse RXRα LBDs as well as oh the heterodimer RXR/RAR have already been described [22,25,29] The final buffers were 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% Glycerol and 2 mM DTT for RAR and RXR/RAR LBDs and 20mM Tris-HCl pH 7.5, 150 mM NaCl, 10% Glycerol and 5 mM DTT for RXR LBD. Fractions containing the purified receptor were pooled, mixed with a threefold molar excess of ligand, AM580 or LG100268, for RAR or RXR, respectively, and a two-fold molar excess of the TIF2 NR2-Ext peptide and concentrated to 8 mg/ml.

Similarly, the RXR/RAR heterodimer was mixed with threefold molar excess of both ligands (AM580 and LG100268) and a two-fold molar excess of the SCR1 NR2-Ext peptide and

concentrated to 8 mg/ml. Crystals were obtained by vapor diffusion at 293 K. The well buffer contained 0.1 M Na Hepes, 0.2 M sodium acetate pH 7.5, 27% PEG 3350 for RXR complex, 0.2 M lithium chloride, 0.1 M MES pH 6.0, 20% PEG 6000 for RAR complex and 0.2 M Na acetate, 0.1 M Bis-Tris propane pH 7.0, 24% PEG 3350 for RXR/RAR complex. Crystals grew in a few days and were of space group P1, P21 and P43212 for RXR, RAR and RXR/RAR complexes, respectively. For each complex, a single crystal was mounted from the mother liquor onto a cryoloop, soaked in the reservoir solution containing an additional 25% glycerol and frozen in liquid nitrogen.

Crystallographic data collection, processing and structure refinement

Diffraction data were collected at the BL13-XALOC beamline of the synchrotron ALBA (Barcelona, Spain) for RXR and RXR/RAR complexes at 2.8 and 2.55 Å resolution, respectively, and at the PXI beamline of the synchrotron SLS (Villigen, Switzerland) at 2.4 Å resolution, for RAR complex. Diffraction data were processed using XDS [124] and scaled with SCALA or Aimless from the CCP4 program suite [125]. The structures were solved by using the previously reported structures 3E94 [126], 3KMR [22] and 1XDK [25], for RXR, RAR and RXR/RAR complexes, respectively, from which the ligand was omitted. Initial Fo–Fc difference maps showed a strong signal for the ligand, which could be fitted accurately into the electron density. The structure was modelled with COOT [127] and refined with REFMAC [128] or with Phenix (phenix.refine) [129] using rigid body refinement, restrained refinement, and individual B-factor refinements. Data collection and refinement statistics are summarized in Table S3. Figures were prepared with PyMOL (http://pymol.org/).

Phylogenetic analysis, signatures of selection and sequence logos

Full-length nucleotide sequences of NCOA family members were recovered from publicly available sequence databases. Accession numbers are provided in Table S4. The transcripts were translated into amino acid sequences and an initial amino acid alignment was performed using MAFFT [130], which was followed by automated refinement using Noisy [131] and final manual curation. The phylogenetic tree was calculated based on the alignment of 34 sequences using 1,562 amino acid positions and allowing gaps (Figure S12). Phylogenetic relationships were assessed using the Maximum Likelihood (ML) method as implemented in RAxML [132]. The ML tree was calculated applying a JTT matrix [133] for all 5 classes of amino acid profiles and a discrete Gamma distribution (+G) to model evolutionary rate differences among sites with empirical base frequencies (+F). The robustness of each node of the resulting tree was assessed by rapid bootstrap analyses (with 1,000 pseudoreplicates). The tree was visualized and midpoint rooted using FigTree (Rambaut, A. FigTree v1.3.1. 2010; Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. http://tree.bio.ed.ac.uk/software/figtree/), with branch lengths indicating the number of substitutions per site.

The amino acid alignment was further used as guide to generate a codon-based alignment for additional analyses. To detect signatures of selection, the ratios between non-synonymous and synonymous substitutions (ω) were estimated for each paralogous lineage (NCOA1, NCOA2 and NCOA3, excluding the Petromyzon marinus sequences) using the branch-site model analysis of the CodeML program in PAML [134]. A similar analysis was carried out for the NR2-Ext peptide sequences of NCOA1, NCOA2 and NCOA3. To highlight evolutionary conserved amino acid sites, a sequence logo was calculated for the NR2-Ext region of each of the three NCOA paralogs using WebLogo [135].

ACCESSION NUMBERS

NMR backbone chemical shift data are available from the Biological Magnetic Resonance Bank (BMRB accession code 50477).

Atomic coordinates and structure factors for the reported crystal structures are available from the Protein Data bank (PDB codes 7APO, 7BK4 and 7AOS).

SAXS data are available from the SAS biological databank (SAS accession code SASDK68).

Ensemble resulting from the use of RDCs data is available from the Protein Ensemble Database (PED accession code PED00184).

SUPPLEMENTARY DATA

CREDIT AUTHOR STATEMENT

Lucile Senicourt: Investigation, Methodology, Writing - Original Draft. Albane le Maire: Investigation, Writing - Review & Editing. Frédéric Allemand: Investigation, Resources. João E. Carvalho: Software, Formal Analysis. Laura Guee: Investigation, Resources. Pierre Germain: Conceptualization, Resources. Michael Schubert: Software, Formal Analysis. Pau Bernadó: Investigation, Writing - Review & Editing. William Bourguet: Conceptualization, Writing - Review & Editing. Nathalie Sibille: Conceptualization, Investigation, Supervision, Writing - Original Draft, Review & Editing.

FUNDING

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] P. Germain, B. Staels, C. Dacquet, M. Spedding, V. Laudet, Overview of nomenclature of nuclear receptors, Pharmacol. Rev. 58 (2006) 685–704.

[2] S. Khan, J.B. Lingrel, Thematic Minireview Series on Nuclear Receptors in Biology and Diseases, J. Biol. Chem. 285 (2010) 38741–38742. doi:10.1074/jbc.r110.196014.

[3] M. Mark, N.B. Ghyselinck, P. Chambon, Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis, Annu. Rev. Pharmacol. Toxicol. 46 (2006) 451–480.

[4] P. Germain, P. Chambon, G. Eichele, R.M. Evans, M.A. Lazar, M. Leid, et al., International Union of Pharmacology. LXIII. Retinoid X receptors, Pharmacol. Rev. 58 (2006) 760–772.

[5] P. Germain, P. Chambon, G. Eichele, R.M.M.M. Evans, M.A.A.A. Lazar, M. Leid, et al., International Union of Pharmacology. LX. Retinoic acid receptors, Pharmacol. Rev. 58 (2006) 712–725.

[6] V. Perissi, K. Jepsen, C.K. Glass, M.G. Rosenfeld, Deconstructing repression: evolving models of co-repressor action, Nat. Rev. Genet. 11 (2010) 109–123.

[7] H. Gronemeyer, J.A. Gustafsson, V. Laudet, Principles for modulation of the nuclear receptor superfamily, Nat Rev Drug Discov. 3 (2004) 950–964.

[8] S. Dasgupta, D.M. Lonard, B.W. O'Malley, Nuclear receptor coactivators: master regulators of human health and disease, Annu. Rev. Med. 65 (2014) 279–292.

[9] B. Le Douarin, A.L. Nielsen, J.M. Garnier, H. Ichinose, F. Jeanmougin, R. Losson, et al., A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors, EMBO J. 15 (1996) 6701–6715.

[10] D.M. Heery, E. Kalkhoven, S. Hoare, M.G. Parker, A signature motif in transcriptional coactivators mediates binding to nuclear receptors, Nature. 387 (1997) 733–736.

[11] J. Torchia, D.W. Rose, J. Inostroza, Y. Kamei, S. Westin, C.K. Glass, et al., The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function, Nature. 387 (1997) 677–684. doi:10.1038/42652.

[12] J.J. Voegel, The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP binding-dependent and -independent pathways, EMBO J. 17 (1998) 507–519. doi:10.1093/emboj/17.2.507.

[13] A. le Maire, W. Bourguet, Retinoic acid receptors: structural basis for coregulator interaction and exchange, Subcell. Biochem. 70 (2014) 37–54.

[14] M.J. Plevin, M.M. Mills, M. Ikura, The LxxLL motif: a multifunctional binding sequence in transcriptional regulation, Trends Biochem. Sci. 30 (2005) 66–69. doi:10.1016/j.tibs.2004.12.001.

[15] X. Hu, M.A. Lazar, The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors, Nature. 402 (1999) 93–96. doi:10.1038/47069.

[16] V. Perissi, L.M. Staszewski, E.M. McInerney, R. Kurokawa, A. Krones, D.W. Rose, et al., Molecular determinants of nuclear receptor–corepressor interaction, Genes Dev. 13 (1999) 3198–3208.

[17] L. Nagy, H.-Y. Kao, J.D. Love, C. Li, E. Banayo, J.T. Gooch, et al., Mechanism of corepressor binding and release from nuclear hormone receptors, Genes Dev. 13 (1999) 3209–3216.

[18] T.N. Cordeiro, N. Sibille, P. Germain, P. Barthe, A. Boulahtouf, F. Allemand, et al., Interplay of Protein Disorder in Retinoic Acid Receptor Heterodimer and Its Corepressor Regulates Gene Expression, Structure. 27 (2019) 1270–1285. doi:10.1016/j.str.2019.05.001.

[19] S. Devarakonda, K. Gupta, M.J. Chalmers, J.F. Hunt, P.R. Griffin, G.D. Van Duyne, et al., Disorder-to-order transition underlies the structural basis for the assembly of a transcriptionally active PGC-1α/ERRγ complex, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 18678–18683. doi:10.1073/pnas.1113813108.

[20] I.M.S. de Vera, J. Zheng, S. Novick, J. Shang, T.S. Hughes, R. Brust, et al., Synergistic Regulation of Coregulator/Nuclear Receptor Interaction by Ligand and DNA, Structure. 25 (2017) 1506–1518.e4. doi:10.1016/j.str.2017.07.019.

[21] M. Guillien, A. le Maire, A. Mouhand, P. Bernadó, W. Bourguet, J.-L. Banères, et al., IDPs and their complexes in GPCR and nuclear receptor signaling., Prog. Mol. Biol. Transl. Sci. 174 (2020) 105–155. doi:10.1016/bs.pmbts.2020.05.001.

[22] A. le Maire, C. Teyssier, C. Erb, M. Grimaldi, S. Alvarez, A.R. de Lera, et al., A unique secondary-structure switch controls constitutive gene repression by retinoic acid receptor, Nat. Struct. Mol. Biol. 17 (2010) 801–807.

[23] Y. Sato, N. Ramalanjaona, T. Huet, N. Potier, J. Osz, P. Antony, et al., The Phantom Effect of the Rexinoid LG100754: Structural and Functional Insights, PLoS One. 5 (2010) e15119. doi:10.1371/journal.pone.0015119.

[24] J. Osz, Y. Brelivet, C. Peluso-Iltis, V. Cura, S. Eiler, M. Ruff, et al., Structural basis for a molecular allosteric control mechanism of cofactor binding to nuclear receptors, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) E588--594.

[25] V. Pogenberg, J.F. Guichou, V. Vivat-Hannah, S. Kammerer, E. Perez, P. Germain, et al., Characterization of the interaction between retinoic acid receptor/retinoid X receptor (RAR/RXR) heterodimers and transcriptional coactivators through structural and fluorescence anisotropy studies, J. Biol. Chem. 280 (2005) 1625–1633.

[26] I.M. Chrisman, M.D. Nemetchek, I.M.S. de Vera, J. Shang, Z. Heidari, Y. Long, et al., Defining a conformational ensemble that directs activation of PPARgamma, Nat. Commun. 9 (2018) 1794. doi:10.1038/s41467-018-04176-x.

[27] D.J. Kojetin, T.P. Burris, Small Molecule Modulation of Nuclear Receptor Conformational Dynamics: Implications for Function and Drug Discovery, Mol. Pharmacol. 83 (2012) 1–8. doi:10.1124/mol.112.079285.

[28] L. Nagy, Mechanism of the nuclear receptor molecular switch, Trends Biochem. Sci. 29 (2004) 317–324. doi:10.1016/j.tibs.2004.04.006.

[29] V. Nahoum, E. Perez, P. Germain, F. Rodriguez-Barrios, F. Manzo, S. Kammerer, et al., Modulators of the structural dynamics of the retinoid X receptor to reveal receptor function, Proc. Natl. Acad. Sci. 104 (2007) 17323–17328. doi:10.1073/pnas.0705356104.

[30] V. Chandra, P. Huang, Y. Hamuro, S. Raghuram, Y. Wang, T.P. Burris, et al., Structure of the intact PPAR-gamma-RXR- nuclear receptor complex on DNA, Nature. 456 (2008) 350–356.

[31] V. Chandra, D. Wu, S. Li, N. Potluri, Y. Kim, F. Rastinejad, The quaternary architecture of RARβ-RXRγ heterodimer facilitates domain-domain signal transmission, Nat Commun. 8 (2017) 868.

[32] F. Meng, V.N. Uversky, L. Kurgan, Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions, Cell. Mol. Life Sci. 74 (2017) 3069–3090. doi:10.1007/s00018-017-2555-4.

[33] N. Rochel, F. Ciesielski, J. Godet, E. Moman, M. Roessle, C. Peluso-Iltis, et al., Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings, Nat. Struct. Mol. Biol. 18 (2011) 564–570. doi:10.1038/nsmb.2054.

[34] R. Van Der Lee, M. Buljan, B. Lang, R.J. Weatheritt, G.W. Daughdrill, A.K. Dunker, et al., Classification of intrinsically disordered regions and proteins, Chem. Rev. 114 (2014) 6589–6631. doi:10.1021/cr400525m.

[35] J. Leers, E. Treuter, J.-Å. Gustafsson, Mechanistic Principles in NR Box-Dependent Interaction between Nuclear Hormone Receptors and the Coactivator TIF2, Mol. Cell. Biol. 18 (1998) 6001–6013. doi:10.1128/mcb.18.10.6001.

[36] N. Sibille, P. Bernado, Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS, Biochem. Soc. Trans. 40 (2012) 955–962.

[37] D. Chen, S.M. Huang, M.R. Stallcup, Synergistic, p160 coactivator-dependent enhancement of estrogen receptor function by CARM1 and p300, J. Biol. Chem. 275 (2000) 40810–40816.

[38] C. Teyssier, D. Chen, M.R. Stallcup, Requirement for multiple domains of the protein arginine methyltransferase CARM1 in its transcriptional coactivator function, J. Biol. Chem. 277 (2002) 46066–46072.

[39] J.J. Voegel, M.J. Heine, C. Zechel, P. Chambon, H. Gronemeyer, TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors, EMBO J. 15 (1996) 3667–3675.

[40] V.N. Uversky, Size-Exclusion Chromatography in Structural Analysis of Intrinsically Disordered Proteins, in: Intrinsically Disord. Protein Anal., Springer New York, 2012: pp. 179–194. doi:10.1007/978-1-4614-3704-8_11.

[41] R.W. Woody, Theory of Circular Dichroism of Proteins, in: Circ. Dichroism Conform. Anal. Biomol., Springer US, 1996: pp. 25–67. doi:10.1007/978-1-4757-2508-7_2.

[42] R.W. Woody, Circular dichroism and conformation of unordered polypeptides, Adv. Biophys. Chem. 2 (1992) 37–79.

[43] P. Bernadó, M. Blackledge, A Self-Consistent Description of the Conformational Behavior of Chemically Denatured Proteins from NMR and Small Angle Scattering, Biophys. J. 97 (2009) 2839–2845. doi:10.1016/j.bpj.2009.08.044.

[44] M.R. Jensen, P.R.L. Markwick, S. Meier, C. Griesinger, M. Zweckstetter, S. Grzesiek, et al., Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings, Structure. 17 (2009) 1169–1185. doi:10.1016/j.str.2009.08.001.

[45] V.N. Uversky, Unusual biophysics of intrinsically disordered proteins, Biochim. Biophys. Acta. 1834 (2013) 932–951.

[46] V.N. Uversky, Intrinsically disordered proteins from A to Z, Int. J. Biochem. Cell Biol. 43 (2011) 1090–1103.

[47] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, et al., Intrinsically disordered protein, J. Mol. Graph. Model. 19 (2001) 26–59.

[48] V.N. Uversky, J.R. Gillespie, A.L. Fink, Why are "natively unfolded" proteins unstructured under physiologic conditions?, Proteins. 41 (2000) 415–427.

[49] R.T. Nolte, G.B. Wisely, S. Westin, J.E. Cobb, M.H. Lambert, R. Kurokawa, et al., Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma, Nature. 395 (1998) 137–143. doi:10.1038/25931.

[50] J.A. Marsh, V.K. Singh, Z. Jia, J.D. Forman-Kay, Sensitivity of secondary structure propensities to sequence differences between - and -synuclein: Implications for fibrillation, Protein Sci. 15 (2006) 2795–2804. doi:10.1110/ps.062465306.

[51] K. Tamiola, B. Acar, F.A.A. Mulder, Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins, J. Am. Chem. Soc. 132 (2010) 18000–18003. doi:10.1021/ja105656t.

[52] G.W. Vuister, A. Bax, Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins, J. Am. Chem. Soc. 115 (1993) 7772–7777. doi:10.1021/ja00070a024.

[53] S. Kosol, S. Contreras-Martos, C. Cedeno, P. Tompa, Structural characterization of intrinsically disordered proteins by NMR spectroscopy, Molecules. 18 (2013) 10802–10828.

[54] V. Ozenne, F. Bauer, L. Salmon, J.R. Huang, M.R. Jensen, S. Segard, et al., Flexiblemeccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables, Bioinformatics. 28 (2012) 1463–1470.

[55] P. Bernado, L. Blanchard, P. Timmins, D. Marion, R.W.H. Ruigrok, M. Blackledge, A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering, Proc. Natl. Acad. Sci. 102 (2005) 17002–17007. doi:10.1073/pnas.0506202102.

[56] P. Tompa, M. Fuxreiter, Fuzzy complexes : polymorphism and structural disorder in protein – protein interactions, Trends Biochem. Sci. 33 (2007) 2–8. doi:10.1016/j.tibs.2007.10.003.

[57] P. Germain, J. Iyer, C. Zechel, H. Gronemeyer, Co-regulator recruitment and the mechanism of retinoic acid receptor synergy, Nature. 415 (2002) 187–192. doi:10.1038/415187a.

[58] T.H. Click, D. Ganguly, J. Chen, Intrinsically Disordered Proteins in a Physics-Based World, Int. J. Mol. Sci. 11 (2010) 5292–5309. doi:10.3390/ijms11125292.

[59] R.T. Gampe, V.G. Montana, M.H. Lambert, A.B. Miller, R.K. Bledsoe, M. V Milburn, et al., Asymmetry in the PPARgamma/RXRalpha Crystal Structure Reveals the Molecular Basis of Heterodimerization among Nuclear Receptors, Mol. Cell. 5 (2000) 545–555. doi:10.1016/s1097-2765(00)80448-7.

[60] E. Hur, S.J. Pfaff, E.S. Payne, H. Grøn, B.M. Buehrer, R.J. Fletterick, Recognition and Accommodation at the Androgen Receptor Coactivator Binding Interface, PLoS Biol. 2 (2004) e274. doi:10.1371/journal.pbio.0020274.

[61] K. Suino, L. Peng, R. Reynolds, Y. Li, J.-Y. Cha, J.J. Repa, et al., The Nuclear Xenobiotic Receptor CAR, Mol. Cell. 16 (2004) 893–905. doi:10.1016/j.molcel.2004.11.036.

[62] P.E. Wright, H.J. Dyson, Linking folding and binding, Curr. Opin. Struct. Biol. 19 (2009) 31–38. doi:10.1016/j.sbi.2008.12.003.

[63] K.F. Ahmad, A. Melnick, S. Lax, D. Bouchard, J. Liu, C.L. Kiang, et al., Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain, Mol. Cell. 12 (2003) 1551–1564.

[64] J. Liu, N.B. Perumal, C.J. Oldfield, E.W. Su, V.N. Uversky, A.K. Dunker, Intrinsic disorder in transcription factors, Biochemistry. 45 (2006) 6873–6888.

[65] B.D. Darimont, R.L. Wagner, J.W. Apriletti, M.R. Stallcup, P.J. Kushner, J.D. Baxter, et al., Structure and specificity of nuclear receptor-coactivator interactions, Genes Dev. 12 (1998) 3343–3356. doi:10.1101/gad.12.21.3343.

[66] C. Chang, J.D. Norris, H. Grøn, L.A. Paige, P.T. Hamilton, D.J. Kenan, et al., Dissection of the LXXLL Nuclear Receptor-Coactivator Interaction Motif Using Combinatorial Peptide Libraries: Discovery of Peptide Antagonists of Estrogen Receptors alpha and beta, Mol. Cell. Biol. 19 (1999) 8226–8239. doi:10.1128/mcb.19.12.8226.

[67] E.M. McInerney, D.W. Rose, S.E. Flynn, S. Westin, T.-M. Mullen, A. Krones, et al., Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation, Genes Dev. 12 (1998) 3357–3368. doi:10.1101/gad.12.21.3357.

[68] B. He, E.M. Wilson, Electrostatic Modulation in Steroid Receptor Recruitment of LXXLL and FXXLF Motifs, Mol. Cell. Biol. 23 (2003) 2135–2150. doi:10.1128/mcb.23.6.2135-2150.2003.

[69] K. Bugge, I. Brakti, C.B. Fernandes, J.E. Dreier, J.E. Lundsgaard, J.G. Olsen, et al., Interactions by Disorder – A Matter of Context, Front. Mol. Biosci. 7 (2020) 1–16. doi:10.3389/fmolb.2020.00110.

[70] N. Palopoli, N.S.G. Foutel, T.J. Gibson, L.B. Chemes, Short linear motif core and flanking regions modulate retinoblastoma protein binding affinity and specificity, Protein Eng. Des. Sel. 31 (2018) 69–77. doi:10.1093/protein/gzx068.

[71] A. Prestel, N. Wichmann, J.M. Martins, R. Marabini, N. Kassem, S.S. Broendum, et al., The PCNA interaction motifs revisited: thinking outside the PIP-box, Cell. Mol. Life Sci. 76 (2019) 4923–4943. doi:10.1007/s00018-019-03150-0.

[72] A. Mohan, C.J. Oldfield, P. Radivojac, V. Vacic, M.S. Cortese, A.K. Dunker, et al., Analysis of Molecular Recognition Features (MoRFs), J. Mol. Biol. 362 (2006) 1043–1059. doi:10.1016/j.jmb.2006.07.087.

[73] R. Pancsa, M. Fuxreiter, Interactions via intrinsically disordered regions: What kind of motifs?, IUBMB Life. 64 (2012) 513–520. doi:10.1002/iub.1034.

[74] J.G. Olsen, K. Teilum, B.B. Kragelund, Behaviour of intrinsically disordered proteins in protein–protein complexes with an emphasis on fuzziness, Cell. Mol. Life Sci. 74 (2017) 3175–3183. doi:10.1007/s00018-017-2560-7.

[75] M. Borg, T. Mittag, T. Pawson, M. Tyers, J.D. Forman-kay, H.S. Chan, Suggest a Physical Basis for Ultrasensitivity, (2007).

[76] R. Hayama, S. Sparks, L.M. Hecht, K. Dutta, J.M. Karp, C.M. Cabana, et al., Thermodynamic characterization of the multivalent interactions underlying rapid and selective translocation through the nuclear pore complex, J. Biol. Chem. 293 (2018) 4555–4563. doi:10.1074/jbc.AC117.001649.

[77] R. Hendus-Altenburger, E. Pedraz-Cuesta, C.W. Olesen, E. Papaleo, J.A. Schnell, J.T.S. Hopper, et al., The human Na+/H+ exchanger 1 is a membrane scaffold protein for extracellular signal-regulated kinase 2, BMC Biol. 14 (2016) 1–17. doi:10.1186/s12915-016-0252-7.

[78] L.M. Tuttle, D. Pacheco, L. Warfield, J. Luo, J. Ranish, S. Hahn, et al., Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex, Cell Rep. 22 (2018) 3251–3264. doi:10.1016/j.celrep.2018.02.097.

[79] H.X. Zhou, Quantitative account of the enhanced affinity of two linked scFvs specific for different epitopes on the same antigen, J. Mol. Biol. 329 (2003) 1–8. doi:10.1016/S0022-2836(03)00372-3.

[80] P. Schuck, Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling, Biophys. J. 78 (2000) 1606–1619. doi:10.1016/s0006-3495(00)76713-0.

[81] No Title No Title, (n.d.).

[82] A. le Maire, P. Germain, W. Bourguet, Protein-protein interactions in the regulation of RAR–RXR heterodimers transcriptional activity, Methods Enzymol. 637 (2020) 175–207. doi:10.1016/bs.mie.2020.02.007.

[83] V. Vacic, V.N. Uversky, A.K. Dunker, S. Lonardi, Composition Profiler: a tool for discovery and visualization of amino acid composition differences, BMC Bioinformatics. 8 (2007) 211.

[84] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, et al., The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242. doi:10.1093/nar/28.1.235.

[85] P. Romero, Z. Obradovic, X. Li, E.C. Garner, C.J. Brown, A.K. Dunker, Sequence complexity of disordered protein, Proteins. 42 (2001) 38–48.

[86] T. Ishida, K. Kinoshita, PrDOS: prediction of disordered protein regions from amino acid sequence, Nucleic Acids Res. 35 (2007) W460–464.

[87] B. Xue, R.L. Dunbrack, R.W. Williams, A.K. Dunker, V.N. Uversky, PONDR-FIT: A metapredictor of intrinsically disordered amino acids, Biochim. Biophys. Acta - Proteins Proteomics. 1804 (2010) 996–1010. doi:10.1016/j.bbapap.2010.01.011.

[88] T. Ishida, K. Kinoshita, Prediction of disordered regions in proteins based on the meta approach, Bioinformatics. 24 (2008) 1344–1348. doi:10.1093/bioinformatics/btn195.

[89] J. Hanson, Y. Yang, K. Paliwal, Y. Zhou, Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks, Bioinformatics. 33 (2017) 685–692.

[90] D.T. Jones, D. Cozzetto, DISOPRED3: precise disordered region predictions with annotated protein-binding activity, Bioinformatics. 31 (2014) 857–863. doi:10.1093/bioinformatics/btu744.

[91] D.W. Buchan, F. Minneci, T.C. Nugent, K. Bryson, D.T. Jones, Scalable web services for the PSIPRED Protein Analysis Workbench, Nucleic Acids Res. 41 (2013) W349–357.

[92] A. Drozdetskiy, C. Cole, J. Procter, G.J. Barton, JPred4: a protein secondary structure prediction server, Nucleic Acids Res. 43 (2015) W389--W394. doi:10.1093/nar/gkv332.

[93] C. Geourjon, G. Deleage, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, Comput. Appl. Biosci. 11 (1995) 681–684.

[94] R. Yan, D. Xu, J. Yang, S. Walker, Y. Zhang, A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction, Sci. Rep. 3 (2013) 2619. doi:10.1038/srep02619.

[95] H. Kaur, G.P. Raghava, Prediction of alpha-turns in proteins using PSI-BLAST profiles and secondary structure information, Proteins. 55 (2004) 83–90.

[96] H. Kamisetty, S. Ovchinnikov, D. Baker, Assessing the utility of coevolution-based residueresidue contact predictions in a sequence- and structure-rich era, Proc. Natl. Acad. Sci. U.S.A. 110 (2013) 15674–15679.

[97] M. Remmert, A. Biegert, A. Hauser, J. Soding, HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment, Nat. Methods. 9 (2011) 173–175.

[98] T.D. Goddard, D.G. Kneller, Sparky - NMR Assignment and Integration Software, SPARKY 3, (2008) Sparky version (3.115). https://www.cgl.ucsf.edu/home/sparky/.

[99] D.S. Wishart, C.G. Bigam, A. Holm, R.S. Hodges, B.D. Sykes, 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, J. Biomol. NMR. 5 (1995) 67–81.

[100] Y.-S. Jung, M. Zweckstetter, Mars robust automatic backbone assignment of proteins, J. Biomol. NMR. 30 (2004) 11–23. doi:10.1023/b:jnmr.0000042954.99056.ad.

[101] D. Wishart, B. Sykes, The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data, J. Biomol. NMR. 4 (1994) 171–180. doi:10.1007/bf00175245.

[102] J. Kragelj, V. Ozenne, M. Blackledge, M.R. Jensen, Conformational propensities of intrinsically disordered proteins from NMR chemical shifts, Chemphyschem. 14 (2013) 3034–3045.

[103] M. Kjaergaard, S. Brander, F.M. Poulsen, Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH, J. Biomol. NMR. 49 (2011) 139–149. doi:10.1007/s10858-011-9472-x.

[104] J.T. Nielsen, F.A.A. Mulder, POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins, J. Biomol. NMR. 70 (2018) 141–165.

[105] S. Schwarzinger, G.J. Kroon, T.R. Foss, J. Chung, P.E. Wright, H.J. Dyson, Sequencedependent correction of random coil NMR chemical shifts, J. Am. Chem. Soc. 123 (2001) 2970–2978.

[106] K. Tamiola, F.A. Mulder, Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins, Biochem. Soc. Trans. 40 (2012) 1014–1020. doi:10.1042/bst20120171.

[107] J.A. Marsh, J.D. Forman-Kay, Ensemble modeling of protein disordered states: experimental restraint contributions and validation, Proteins. 80 (2012) 556–572.

[108] N.A. Farrow, R. Muhandiram, A.U. Singer, S.M. Pascal, C.M. Kay, G. Gish, et al., Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation, Biochemistry. 33 (1994) 5984–6003. doi:10.1021/bi00185a040.

[109] R. Konrat, NMR contributions to structural dynamics studies of intrinsically disordered proteins, J. Magn. Reson. 241 (2014) 74–85.

[110] W. Lee, M. Tonelli, J.L. Markley, NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy, Bioinformatics. 31 (2014) 1325–1327. doi:10.1093/bioinformatics/btu830.

[111] L.E. Kay, D.A. Torchia, A. Bax, Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, Biochemistry. 28 (1989) 8972–8979. doi:10.1021/bi00449a003.

[112] L. Salmon, M.R. Jensen, P. Bernadó, M. Blackledge, Measurement and Analysis of NMR Residual Dipolar Couplings for the Study of Intrinsically Disordered Proteins, in: Methods Mol. Biol., Humana Press, 2012: pp. 115–125. doi:10.1007/978-1-61779-927-3_9.

[113] M. Rückert, G. Otting, Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments, J. Am. Chem. Soc. 122 (2000) 7793–7797. doi:10.1021/ja001068h.

[114] M.R. Hansen, L. Mueller, A. Pardi, Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions, Nat. Struct. Biol. 5 (1998) 1065–1074.

[115] G.M. Clore, M.R. Starich, A.M. Gronenborn, Measurement of Residual Dipolar Couplings of Macromolecules Aligned in the Nematic Phase of a Colloidal Suspension of Rod-Shaped Viruses, J. Am. Chem. Soc. 120 (1998) 10571–10572. doi:10.1021/ja982592f.

[116] M. Zweckstetter, A. Bax, Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage, J. Biomol. NMR. 20 (2001) 365–377.

[117] F. Cordier, A.J. Dingley, S. Grzesiek, A doublet-separated sensitivity-enhanced HSQC for the determination of scalar and dipolar one-bond J-couplings, J. Biomol. NMR. 13 (1999) 175–180.

[118] J.L. Battiste, G. Wagner, Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data, Biochemistry. 39 (2000) 5355–5365. doi:10.1021/bi000060h.

[119] M.W. Roessle, R. Klaering, U. Ristau, B. Robrahn, D. Jahn, T. Gehrmann, et al., Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg, J. Appl. Crystallogr. 40 (2007) s190–s194. doi:10.1107/s0021889806055506.

[120] D. Franke, M. V Petoukhov, P. V Konarev, A. Panjkovich, A. Tuukkanen, H.D.T. Mertens, et al., ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, J. Appl. Crystallogr. 50 (2017) 1212–1225. doi:10.1107/s1600576717007786.

[121] A. Guinier, X-ray diffraction at small angles: Application to the study of ultramicroscopic phenomena, Ann. Phys. (Paris). 11 (1939) 161–237. doi:10.1051/anphys/193911120161.

[122] D. Svergun, C. Barberato, M.H.J. Koch, CRYSOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, J. Appl. Crystallogr. 28 (1995) 768–773. doi:10.1107/S0021889895007047.

[123] A.A. Canutescu, A.A. Shelenkov, R.L. Dunbrack, A graph-theory algorithm for rapid protein side-chain prediction, Protein Sci. 12 (2003) 2001–2014. doi:10.1110/ps.03154503.

[124] W. Kabsch, XDS., Acta Crystallogr. D. Biol. Crystallogr. 66 (2010) 125–132. doi:10.1107/S0907444909047337.

[125] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, et al., Overview of the CCP4 suite and current developments., Acta Crystallogr. D. Biol. Crystallogr. 67 (2011) 235–242. doi:10.1107/S0907444910045749.

[126] A. le Maire, M. Grimaldi, D. Roecklin, S. Dagnino, V. Vivat-Hannah, P. Balaguer, et al., Activation of RXR-PPAR heterodimers by organotin environmental endocrine disruptors., EMBO Rep. 10 (2009) 367–373. doi:10.1038/embor.2009.8.

[127] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot., Acta Crystallogr. D. Biol. Crystallogr. 66 (2010) 486–501. doi:10.1107/S0907444910007493.

[128] A.A. Vagin, R.A. Steiner, A.A. Lebedev, L. Potterton, S. McNicholas, F. Long, et al., REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use., Acta Crystallogr. D. Biol. Crystallogr. 60 (2004) 2184–2195. doi:10.1107/S0907444904023510.

[129] D. Liebschner, P. V Afonine, M.L. Baker, G. Bunkóczi, V.B. Chen, T.I. Croll, et al., Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix., Acta Crystallogr. Sect. D, Struct. Biol. 75 (2019) 861–877. doi:10.1107/S2059798319011471.

[130] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability., Mol. Biol. Evol. 30 (2013) 772–780. doi:10.1093/molbev/mst010.

[131] A.W.M. Dress, C. Flamm, G. Fritzsch, S. Grünewald, M. Kruspe, S.J. Prohaska, et al., Noisy: identification of problematic columns in multiple sequence alignments., Algorithms Mol. Biol. 3 (2008) 7. doi:10.1186/1748-7188-3-7.

[132] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies., Bioinformatics. 30 (2014) 1312–1313. doi:10.1093/bioinformatics/btu033.

[133] D.T. Jones, W.R. Taylor, J.M. Thornton, The rapid generation of mutation data matrices from protein sequences., Comput. Appl. Biosci. 8 (1992) 275–282. doi:10.1093/bioinformatics/8.3.275.

[134] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood., Mol. Biol. Evol. 24 (2007) 1586–1591. doi:10.1093/molbev/msm088.

[135] G.E. Crooks, G. Hon, J.-M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator., Genome Res. 14 (2004) 1188–1190. doi:10.1101/gr.849004.

Highlights

- TIF2 is an IDP with evolutionary conserved partially structured NR-boxes and flanking regions
- NR-box2 and its flanking region fold as a helix-turn-helix motif specifically interacting with RAR
- Multisite binding of TIF2 to liganded RXR/RAR is driven by two consecutive NR-boxes

Figure Captions

Figure 1. TIF2NRID is monomeric, disordered and contains secondary structure elements in solution. A) Schematic representation of the functional domains identified in TIF2. The TIF2 sequence contains several functional domains including a basic helix–loop–helix (bHLH) similarity domain; a Per-Arndt-Sim (PAS) homology domain; a nuclear receptor interacting domain (NRID); a glutamine-rich (Q rich) domain and two autonomous activation domains (AD): a CBP interaction domain (CID/AD1) recruiting the transcriptional co-activators CBP or p300, and the AD2 to recruit an arginine methyltransferase (CARM-1) [12,37–39]. TIF2NRID construct (residues 624 to 773) contains the three LxxLL (i.e. LLQLL, LHRLL and LRYLL) NR binding motifs (NR-boxes 1-3) to NRs. B) Size exclusion chromatogram of TIF2NRID (dashed line and left axis), and molar mass derived from MALS (thick line and right axis) at room temperature in NMR buffer. C) TIF2NRID backbone 15NHSQC of 15N/13C TIF2NRID. D) Far-UV CD spectrum of TIF2NRID at 10°C in NMR buffer.

Figure 2. Bioinformatics predictions, secondary structure propensity, backbone dynamics, long-range contacts of TIF2NRID. A) Disorder predictions from PrDos (yellow line), meta-PrDOs (red line), SPOT-DISORDER (magenta line), DISOPRED3 (blue line) and PONDR-Fit (green line) servers. Lower values indicate greater likelihood of being structured (scores < 0.5; dashed line). B) Secondary structure predictions from Psipred, Jpred4, SOPMA, AlphaPred and PSSpred servers. Predicted α-helices (red boxes), β-strands (blue arrows) and turns (green bars). C) Sequence logo conservation from GREMLIN analysis of a multiple sequence alignment of 86 sequences (Table S2). Letter height indicates the relative frequency and conservation in the alignment. D) Secondary structure propensity (SSP) obtained from experimental Cα and Cβ chemical shifts (positive and negative scores for propensity to form, respectively α-helices and β-strands). E) 15N relaxation times along the sequence: T1 (longitudinal) over T2 (transverse) (blue dashed line represents the average value). F) Backcalculated and experimental 1DHN RDCs of TIF2NRID. Comparison of the experimental 1DHN RDCs measured in alcohol mixture (black line) with back-calculated values from a completely random-coil ensemble generated with FM (blue line) and from an ensemble including conformational preferences (red line) populated as indicated in the figure (regions shaded in red (helical structure), blue (extended beta structure) or green (α-turn)). G) PRE data of TIF2NRID S716C variant. Values were calculated from the ratio of intensity (Ipara/Idia) of 15N-HSQC spectra measured for the paramagnetic (Ipara) sample in the position indicated with a red arrow, and for the diamagnetic sample (Idia). Dark grey and light grey-hashed bars correspond to overlapped peaks and prolines residues, respectively. Star labelled residues correspond to disappearing peaks. NR2-box2 (L690-L694) is highlighted in grey and its downstream region including the α-helical flanking region (P700-A710) is highlighted in light blue. NRboxes (LxxLL motifs; i.e. residues L641-L645, L690-L694 and L745-L749) are highlighted in grey or hashed (F).

Figure 3. Recruitment of TIF2NRID by the heterodimer RXR/RAR LBDs. A) Dissociation constants derived from titration of FITC-labelled TIF2NRID by RXR/RAR in the absence of ligand (Apo) or in the presence of RAR agonist (AM580), RXR agonist (CD3254), or combination of ligands including RAR antagonist (BMS614) and RXR antagonist (UVI3003). An one-way ANOVA of LogEC50 analysis has been performed on the interaction in the presence of AM580 and of both AM580 and CD3254
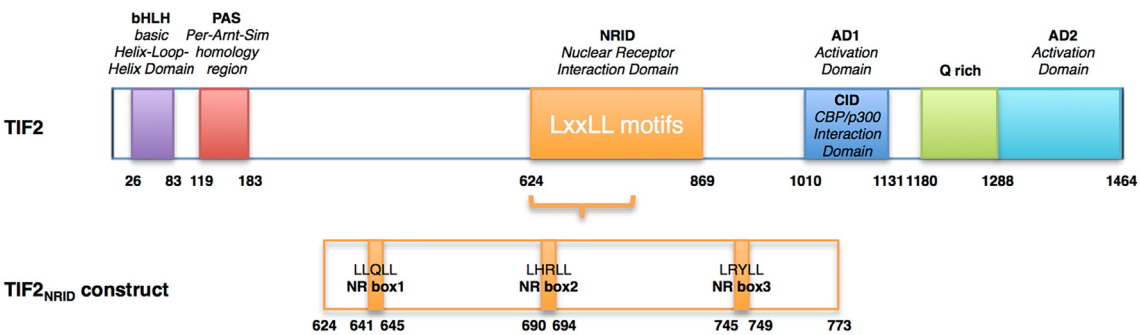
and shows no statistical difference, as highlighted by the bracket. B) Dissociation constants derived from titration of Alexa-labelled TIF2NRID LXXAA variants by RXR/RAR in the absence of ligand.

Figure 4. Interaction of TIF2NRID with RXR/RAR LBDs and agonists effects. A) Superposition of 15N-HSQC spectra of 5 µM of 15N labelled TIF2NRID in absence (blue) and in the presence (orange) of 6 µM of unlabelled RXR/RAR LBDs. Residues belonging to NR-boxes are indicated by grey bars. B) Relative peak intensity ratios IComplex/ITIF2NRID between the two spectra presented in A. C) and D) Peak intensity differences between 15N-HSQC spectra of RXR/RAR:TIF2NRID complex (1.1:1) in the presence of 1.2 molar excess of RAR agonist AM580 (cyan, C); then followed by the addition of 1.2 molar excess of RXR agonist CD3254 (green, D). Dashed bars correspond to prolines (grey) or ambiguous and overlapped peaks (coloured).
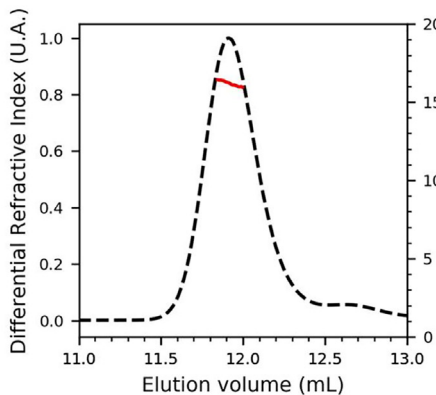
Figure 5. Interaction of individual TIF2NRID NR-box variants with unliganded RXR/RAR LBDs heterodimer. Relative peak intensity ratios of 15N-HSQC spectra of 5 µM of 15N labelled TIF2NRID-NR-box variant in absence (Ifree) and in the presence (IApo) of 6 µM of unlabelled RXR/RAR LBDs. A) NR-box1 variant. Red hashed bar correspond to non-assigned peaks, B) NR-box 2 variant, C) NR-box3 variant. Grey highlight boxes indicate NR-boxes. Hashed bars correspond to prolines (grey) or ambiguous and overlapped peaks (orange).

Figure 6. X-Ray crystallographic structures of the TIF2 NR2-Ext peptide in complex with RAR and RXR LBDs. A) Overall structure of RAR LBD in complex with the agonist AM580 and TIF2 NR2-Ext. The TIF2 NR2-Ext peptide is shown in yellow (α1) and orange (α2) cartoon and RAR is coloured in light blue with H12 coloured in red. B) Close-up view of the interaction surface of the TIF2 NR2-Ext peptide with RAR highlighting the hydrophobic contacts of L703, L706, T707 and A710 from helix α2 with L694, I689 from helix α1 or I237 in H3 and L409 in H12 from RAR. An electrostatic bond between E651 (α2) and R694 (α1) is also observed. C) Equilibrium dissociation constants (KD in µM) deduced from fluorescence anisotropy experiments between either RAR or RXR LBDs and TIF2 NR2 (yellow bars) or TIF2 NR2-Ext (orange bars). D) Crystal structure of RXR LBD in complex with TIF2 NR2-Ext and the RXR agonist LG100268.
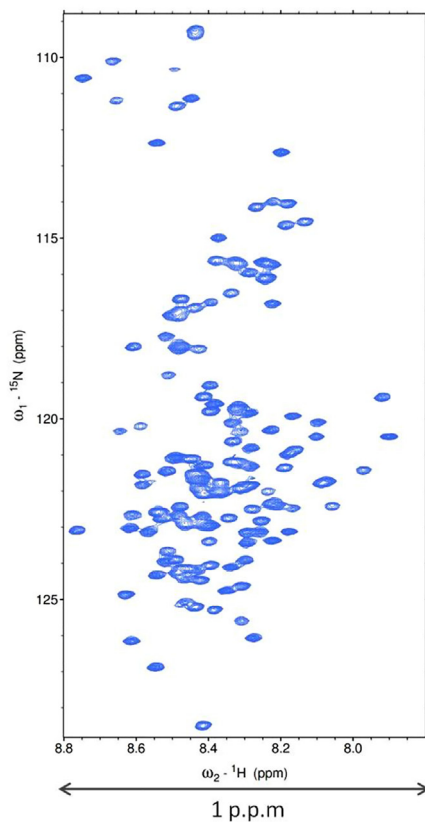
Figure 7. Interaction of SRC1 NR2-Ext peptide with RXR and RAR LBDs. A) Sequence logos for the NR2-Ext peptide sequence of nuclear receptor co-activator (NCOA) paralogues: SRC1 (NCOA1), TIF2 (NCOA2) and RAC3 (NCOA3). The ratio between non-synonymous and synonymous substitutions (ω) is shown for the NR2-Ext peptides of each of the three paralogues. B) Crystal structure of RXR/RAR LBD heterodimer in complex with the agonists AM580 and LG100268 and SRC1 NR2-Ext. The only part of SRC1 NR2-Ext peptide visible in the structure is shown in yellow (α1) cartoon, and RXR and RAR are coloured in light green and light blue, respectively, with H12 coloured in red. C) Competition curves at 0.5 µM of RAR LBDs of SRC1 NR2 by unlabelled SRC1 NR2 and SRC1 NR2-Ext peptides in the presence of AM580. The derived inhibition constant Ki is nearly the same for the two peptides.
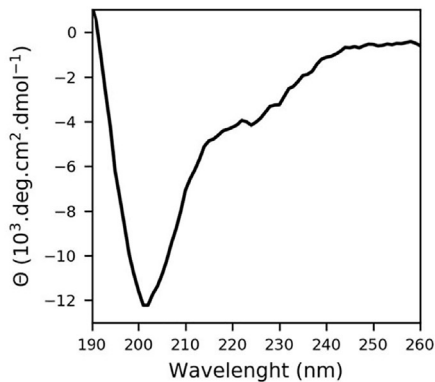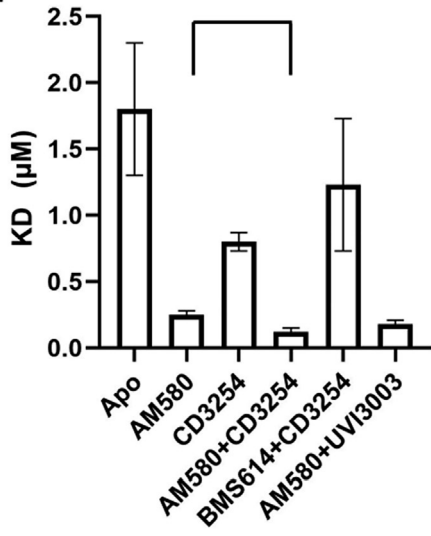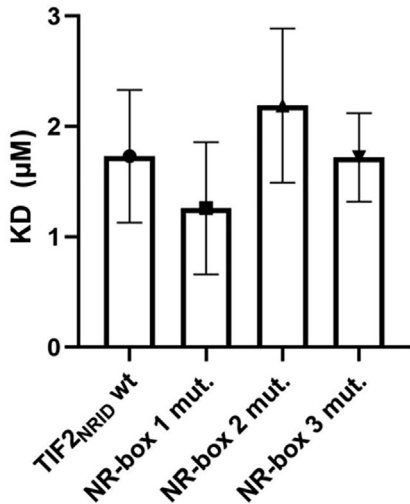
TIF2 domain organization showing: bHLH (basic Helix-Loop-Helix Domain) 26–83, PAS (Per-Arnt-Sim homology region) 119–183, NRID (Nuclear Receptor Interaction Domain) with LxxLL motifs 624–869, AD1 (Activation Domain) with CID (CBP/p300 Interaction Domain) 1010–1131, Q rich 1180–1288, AD2 (Activation Domain) 1288–1464.

TIF2$_{NRID}$ construct: LLQLL NR box1 (641, 645), LHRLL NR box2 (690, 694), LRYLL NR box3 (745, 749), spanning 624–773.

A

B

C

A

B

C

D

E

F

G

ERADGQSRLHDSKGQTKLLQLLTTKSDQMEPSPLASSLSDTNKDSTGSLPGSGSTHGTSLKEKHKILHRLLQDSSSPVDLAKLTAEATGKDLSQESSSTAPGSEVTIKQEPVSPKKKENALLRVLLLDKDDTKDIGLPEITPKLERLDSKT

A

KD (μM)

Apo, AM580, CD3254, AM580+CD3254, BMS614+CD3254, AM580+UVI3003

B

KD (μM)

TIF2$_{NRID}$ wt, NR-box 1 mut., NR-box 2 mut., NR-box 3 mut.

A NR-box 1 variant

B NR-box 2 variant

C NR-box 3 variant

I_{Apo}/I_{free}

Residue number

C

| RAR | | RXR | |

KD (µM)
2.5
2.0
1.5
0.50
0.25
0.00

A

SCR1 (NCOA1)

ω = 0.00669

TIF2 (NCOA2)

ω = 0.02149

RAC3 (NCOA3)

ω = 0.00604

B

RXR   RAR

SRC1   SRC

C

RAR

- SRC1 NR2
- SRC1 NR2-Ext

Anisotropy

[peptide] (μM)