Poly-glutamine-dependent self-association as a potential mechanism for regulation of androgen receptor activity

Carlos M. Roggero[1,2,3*], Victoria Esser[1,2,3], Lingling Duan[4], Allyson M. Rice[1], Shihong Ma[6], Ganesh V. Raj[6], Michael K. Rosen[1,7], Zhi-Ping Liu[4,5] and Josep Rizo[1,2,3*]

[1]Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, Texas, United States; [2]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas, United States; [3]Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas, United States; [4]Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, United States; [5]Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, Texas, United States; [6]Department of Urology, University of Texas Southwestern Medical Center, Dallas, Texas, United States; [7]Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas, United States

[*]Email: Carlos.Roggero@UTSouthwestern.edu or Jose.Rizo-Rey@UTSouthwestern.edu

Short title: Regulation of androgen receptor activity by self-association

Abstract

The androgen receptor (AR) plays a central role in prostate cancer. Development of castration resistant prostate cancer (CRPC) requires androgen-independent activation of AR, which involves its large N-terminal domain (NTD) and entails dramatic epigenetic changes depending in part on histone lysine demethylases (KDMs) that interact with AR. The AR-NTD is rich in low-complexity sequences, including a polyQ repeat. Longer polyQ sequences were reported to decrease transcriptional activity and to protect against prostate cancer. However, the molecular mechanisms underlying these observations are unclear. Using NMR spectroscopy, here we identify weak interactions between the AR-NTD and the KDM4A catalytic domain, and between the AR ligand-binding domain and a central KDM4A region that also contains low-complexity sequences. We also show that the AR-NTD can undergo liquid-liquid phase separation in vitro, with longer polyQ sequences phase separating more readily. Moreover, longer polyQ sequences hinder nuclear localization in the absence of hormone and increase the propensity for formation of AR-containing puncta in the nucleus of cells treated with dihydrotestosterone. These results lead us to hypothesize that polyQ-dependent liquid-liquid phase separation may provide a mechanism to decrease the transcriptional activity of AR, potentially opening new opportunities to design effective therapies against CRPC.

Introduction

Prostate cancer is one of the most common forms of cancer in men(1). Survival and proliferation of prostate cancer cells depend critically on signaling through the androgen receptor (AR) and, consequently, standard therapies against prostate cancer involve androgen reduction through chemical or surgical castration(2, 3). Unfortunately, prostate cancer cells eventually develop the ability to activate AR independently of androgen and patients invariably develop the more aggressive castration-resistant prostate cancer (CRPC), which is associated with a much worse prognosis(4-6). Hence, understanding the mechanisms underlying AR function and how AR is activated in a ligand-independent manner is crucial to develop effective therapies against CRPC.

One aspect of AR biology that is most likely involved in prostate cancer is the complicated functional interplay between AR and epigenetic enzymes, including histone deacetylases (HDACs) and lysine demethylases (KDMs). Thus, AR activity can be downregulated or upregulated by HDACs and KDMs (e.g. (7-11)), and the development and progression of prostate cancer is accompanied by extensive abnormalities in the levels of these enzymes and in histone modification marks(12, 13), which has been referred to as an 'epigenetic catastrophe'(14). Moreover, AR was reported to bind to several KDMs, including KDM4A, although only limited information was reported about the regions of the KDMs and of AR implicated in the interactions(9-11). Hence, the nature of these interactions remains unclear and it is unknown how they affect AR activity.

AR contains a long N-terminal domain (NTD), a central DNA-binding domain (DBD), a short hinge region (H) and a ligand-binding domain (LBD) (Fig. 1A). The LBD is normally key for receptor activation, but the NTD is required for full transcriptional activity, and interactions between NTD and LBD regulate androgen-dependent gene expression(15-17). Activation involves the translocation of AR to the nucleus when androgen binding to the LBD induces a large conformational change that exposes a nuclear localization signal (NLS) present between the DBD and LBD(18). The NTD has attracted much attention

not only because of its known participation in transcriptional activity but also, and in particular, because diverse C-terminally truncated splice variants of AR lacking the LBD have been identified in prostate cancer cell lines and in clinical specimens(19). These variants can explain the development of androgen-independent AR activity in CRPC. Indeed, multiple studies have suggested that AR-V7, which is the most abundant of these variants and is constitutively localized to the nucleus, mediates androgen-independent cell growth and resistance to androgen deprivation (reviewed in ref. (20)). Interestingly, our recent finding that KDM4B regulates the generation of AR-V7 by alternative splicing upon androgen deprivation(21) provided another connection between AR and KDM function.

The AR NTD contains abundant low-complexity sequences that are common in transactivation domains of transcription factors and that include polyQ and poly-G regions with variable length in the human population(22). Exceedingly long polyQ sequences (> 38 Qs) can lead to neurodegenerative disease(23), which may be associated to the formation of oligomeric fibrils(24). Conversely, short polyQ length (< 20 Qs) was shown to increase the risk for prostate cancer(25, 26), indicating that longer polyQ sequences may have a protective effect. These findings may arise because the transcriptional potency of AR is inversely correlated with polyQ length(27-29). As expected, the NTD is largely unstructured in solution(30), but some sequences do have defined propensities to form secondary structure, particularly in regions that are important for transactivation(31, 32). The polyQ sequence has a clear tendency to form α-helical structure that increases with the length of the sequence and is stabilized by interactions of the glutamine side chains with the backbone(33-35). These observations led to the proposal that the effects of polyQ length on transcriptional activity may arise from changes in the strength of protein-protein interactions, for instance those with transcriptional co-regulators or general transcription factors(34). It is also important to note that the AR NTD has been shown to undergo liquid-liquid phase separation at high concentrations (100 μM)(36), and that a propensity for phase separation appears to be a fundamental property of many low-complexity sequences that are involved in transcriptional activation

and/or mediate formation of membraneless organelles(37-39). However, the importance of this property for AR function and the influence of polyQ length on the tendency of the AR NTD to undergo phase separation have not been explored.

The study presented here was designed to examine how AR binds to KDM4A and to investigate how self-association of the AR NTD and nuclear localization of AR depend on the length of the polyQ sequence. NMR experiments reveal weak interactions between the AR-NTD and the KDM4A catalytic domain, and between the AR LBD and a central KDM4A region that also contains low-complexity sequences. We also show that the drive for the AR-NTD to undergo phase separation in vitro increases progressively with polyQ length. We verify that polyQ length is inversely correlated with transcriptional activity in two prostate cancer cell lines. Moreover, we find that nuclear localization is decreased by longer polyQ sequences in the absence of hormone, and that such sequences increase the formation of AR-containing puncta in nuclei of cells treated with dihydrotestosterone (DHT). Based on these results, we propose that longer polyQ sequences in AR may have a protective role against prostate cancer by enhancing self-association processes that reduce the AR transcriptional activity.

Results

Direct AR-KDM4A binding through regions containing low-complexity sequences

The present study was initiated with the goal of gaining insights into how KDMs affect AR function by investigating the nature of the interactions between these proteins. KDM4A and its homologues contain an N-terminal catalytic domain, a central region that contains low-complexity sequences, a C-terminal region that includes two PHD domains and two Tudor domains (Fig. 1A). Co-immunoprecipitation experiments suggested that AR binds to KDM4A and KDM4D through its LBD, while the catalytic domain and the C-terminus of KDM4A, or the C-terminus of KDM4D, mediate AR binding(11). However, it is difficult to derive definitive conclusions with regard to direct physical interactions from co-

116 immunoprecipitation. To examine whether we could observe direct binding between purified
117 recombinant proteins, we performed extensive pulldown assays and hold up assays(40) using fragments
118 that spanned different regions of AR and KDM4A. However, we were unable to obtain definitive results
119 using these approaches, suggesting that potential interactions between AR and KDM4A are too weak to
120 be reliably detected by these methods.

121 We turned to an NMR method based on the analysis of perturbations caused by an unlabeled
122 protein on transverse relaxation optimized (TROSY) ^{1}H-^{15}N heteronuclear single quantum coherence
123 (HSQC) spectra of a uniformly ^{15}N-labeled protein, which provide a sensitive tool to detect interactions
124 between the two proteins(41). In a first set of experiments, we acquired ^{1}H-^{15}N TROSY-HSQC spectra of a
125 ^{15}N-labeled fragment spanning the catalytic domain of KDM4A (residues 1-350) in the absence and
126 presence of different unlabeled fragments of AR. DHT was included in experiments with fragments
127 containing the AR LBD, which is unstable without a ligand. A fragment spanning the NTD of AR (AR-NTD;
128 residues 1-559) caused noticeable but limited broadening on the ^{1}H-^{15}N TROSY-HSQC spectrum of ^{15}N-
129 KDM4A(1-350), while a longer fragment that included the AR NTD and DBD (AR-NTD-DBD; residues 1-632)
130 caused stronger broadening (Fig. 1B,C). In contrast, almost no perturbations on the ^{1}H-^{15}N TROSY-HSQC
131 spectrum of ^{15}N-KDM4A(1-350) were caused by fragments containing the AR LBD alone (AR-LBD; residues
132 663-919) or from the DBD to the LBD (AR-DBD-LBD; residues 554-919) (Fig. 1D,E). These results show that,
133 under these conditions, the catalytic domain of KDM4A binds weakly to AR-NTD and that such binding is
134 enhanced by the DBD, while there is no appreciable binding to the LBD or to the DBD in the absence of
135 the NTD. Hence, binding of the AR-NTD-DBD fragment to KDM4A(1-350) appears to be driven by weak
136 interactions with the NTD that can cooperate with additional interactions involving the DBD. The latter
137 might be too weak to be observable without such cooperation or might be hindered by intramolecular
138 binding of the DBD to the hinge region or the LBD in the context of the AR-DBD-LBD fragment.

Because the AR-LBD had been previously implicated in binding to KDM4s(11) but we did not observe binding to KDM4A(1-350), we tested whether the LBD might interact with other regions of KDM4A. For this purpose, we prepared ^{15}N-labeled AR-LBD bound to DHT and acquired ^{1}H-^{15}N TROSY-HSQC spectra in the absence and presence of fragments spanning the central region of KDM4A (residues 301-708) or its C-terminal region containing the PHD and Tudor domains (residues 703-1064). The central region of KDM4A containing the low complexity sequences caused substantial broadening on the ^{1}H-^{15}N TROSY-HSQC spectrum of ^{15}N-AR-LBD, indicative of binding, while the C-terminal KDM4A fragment induced very little perturbations (Fig. 1F,G).

We also analyzed whether the AR-NTD-DBD fragment binds to the catalytic domains of two other KDM4 isoforms, KDM4B and KDM4C. The ^{1}H-^{15}N TROSY-HSQC spectrum of ^{15}N-labeled catalytic domain of KDM4C exhibited dramatic broadening upon addition of AR-NTD-DBD, while the broadening was less overt in the ^{1}H-^{15}N TROSY-HSQC spectrum of ^{15}N-labeled catalytic domain of KDM4B (Supporting Fig. S1). Hence, there is some level of specificity in the interaction of AR-NTD-DBD with KDM4 isoforms.

To confirm the interaction of AR-NTD-DBD with the catalytic domain of KDM4A by another method, we performed chemical cross-linking experiments with bis(sulfosuccinimidyl)suberate (BS3), an agent that cross-links primary amines with primary amines. When we treated mixtures of AR-NTD-DBD and KDM4A(1-350) with BS3 and analyzed the mixture by SDS-PAGE followed by coomassie blue staining, we observed the almost complete disappearance of the bands of AR-NTD-DBD and KDM4A(1-350) and the appearance of bands corresponding to high molecular weight cross-linked complexes (Fig. 2A). An AR fragment spanning most of the NTD but lacking the DBD (residues 93-495) was also cross-linked efficiently with KDM4A(1-350) (Fig. 2B), albeit not as quantitatively as the AR-NTD-DBD fragment. Hence, these data correlate with the NMR results, which revealed binding of AR-NTD to KDM4A(1-350) that was strengthened by inclusion of the AR-DBD (Fig. 1B,C).

Overall, our results indicate that there are indeed direct interactions between AR and KDM4A. The interactions appear to be weak, but may cooperate with other interactions among components of the transcriptional machinery for recruitment of these proteins and/or control of their activities. Interestingly, the two types of interactions that we observed involve regions containing low complexity sequences of AR (binding of AR-NTD to the KDM4A catalytic domain) or KDM4A (binding of the KDM4A central region to AR-LBD).

Longer polyQ sequences enhance gel formation by the AR-NTD

Increasing experimental evidence shows that many proteins that are multivalent for protein-protein interactions have a tendency to undergo liquid-liquid phase separation and that this property underlies the formation of membraneless subcellular compartments that serve a wide variety of biological functions(38). In such systems, the drive to phase separate increases with protein length and/or interaction valence(38, 42). The tendency to phase separate is particularly common for low-complexity sequences present in transcription factors and other proteins involved in formation of complexes with nucleic acids (e.g. (37, 39)). Indeed, the AR-NTD (residues 1-559) was shown to phase separate into liquid droplets at 100 μM concentration(36). These findings, together with the known importance of the AR-NTD for development of CRPC, the notion that longer polyQ sequences in the AR-NTD may protect against prostate cancer, and our finding that low complexity sequences mediate AR-KDM4A interactions, led us to investigate how polyQ length affects the ability of AR-NTD to phase separate. For this purpose, we prepared and purified recombinant fragments corresponding to the AR-NTD that contained polyQ sequences with 12, 20, 31 and 49 glutamines (AR-NTD-12Q, -20Q, -31Q and -49Q, respectively).

The four versions of AR-NTD each exhibited some turbidity at 500-600 μM concentrations and 200 mM NaCl, consistent with the previous observation of phase separation by this AR fragment(36), but the turbidity was markedly stronger for the AR-NTD fragments with longer polyQ sequence (Fig. 3A). After

a sample of AR-NTD-31Q was allowed to reach high turbidity, the sample became clear upon addition of 10% 1,6-hexanediol (Fig. 3B). This observation shows that phase separation by NTD is reversible and does not result from irreversible formation of amyloid aggregates(43). To have a more quantitative measure of the influence of polyQ length on the drive to phase separate, we prepared different dilutions of the four AR-NTD fragments from freshly prepared concentrated solutions, incubated the resulting samples for 1 h at 4 °C, room temperature (RT; ca. 23 °C) or 37 °C, and measured the absorption at 400 nm. Analysis by SDS PAGE illustrated the purity of the samples used in these experiments (Supporting Fig. S2). Precipitation prevented reliable quantification of the turbidity at 37 °C, but consistent results were obtained in the experiments performed at 4 °C and RT, which generally revealed progressive increases of absorption at 400 nm with the concentration for the four proteins. Importantly, the turbidity correlated with the length of the polyQ sequence under the majority of conditions (Fig. 3C,D; note that the lower OD at 400 nm observed for 100 μM AR-NTD-49Q after 1 h at RT, compared to that observed for 50 μM AR-NTD-49Q or 100 μM AR-NTD-31Q, might arise because droplets may settle out of solution). These results indicate that longer polyQ sequences increase the drive for the AR-NTD to phase separate.

The length of the polyQ sequence in AR was found to be inversely correlated with transcriptional activity(27-29). To verify these findings with the same polyQ tracts that we used for the turbidity assays, we used a luciferase assay in two prostate cancer cell lines (LNCaP and PC3) transfected with vectors expressing full-length AR that contained 12, 20, 31 and 49 glutamines (AR-12Q, -20Q, -31Q and -49Q, respectively). In LNCaP cells, which express endogenous AR, we observed robust enhancement of transcriptional activity upon addition of DHT even when cells were transfected with control vector (Fig. 4A). The DHT-induced increase in activity was markedly stronger for cells transfected with AR-12Q, but the stimulation gradually decreased as the polyQ length increased, and there was no additional stimulation in cells expressing AR-49Q (Fig. 4A). To ensure that these results did not arise from lower expression of the AR variants with longer polyQ sequences, we analyzed the total AR levels in Western

blots with AR antibodies and the levels of transfected AR proteins, which included a FLAG tag, with FLAG antibodies (Fig. 4B). These experiments showed that the expression levels of the transfected AR proteins were comparable to each other, and similar to the levels of endogenous AR. In experiments performed with PC3 cells, which do not express endogenous AR, we did not observe a substantial DHT-induced enhancement of transcriptional activity when the cells were transfected with control vector, but transfection with AR-12Q caused robust DHT-dependent increase in activity and such increase again decreased gradually with the length of the polyQ sequence (Fig. 4C). Hence, these results confirm that the length of the polyQ sequence can indeed have strong effects on AR activity.

Longer polyQ sequences decrease nuclear localization of AR and increase DHT-dependent formation of AR-containing puncta in the nucleus

Transcriptional activation by AR requires the transport of AR to the nucleus, which is normally induced by ligand binding and exposure of a hidden NLS(18). However, C-terminally truncated AR variants such as AR-V7 can be constitutively localized to the nucleus(20). To analyze how the length of the polyQ sequence influences the nuclear localization of AR, we transfected LNCaP cells with Cherry-labeled AR variants containing 12, 20, 31 and 49 glutamines, and analyzed their distribution in the cytoplasm and the nucleus. The results of these experiments need to be examined with caution, as we observed nuclear localization of the transfected proteins even in the absence of hormone (Fig. 5A,B), which may arise because fluorescent proteins have a tendency to translocate to the nucleus(44). Nevertheless, these experiments can still provide information regarding how the length of the polyQ sequence affects the propensity of AR for nuclear localization. Importantly, we observed that, under castration-mimicking conditions lacking hormone, longer polyQ sequences clearly favored the cytoplasmic localization of AR in detriment to its nuclear localization (Fig. 5A,B). This finding did not arise from the differences in expression, as the levels of the expressed cherry-AR proteins assessed by Western blot exhibited some variability (Fig. 5C) but this

variability did not correlate with the percentage of protein localized to the nucleus. In the presence of DHT, most of the AR was localized to the nucleus regardless of polyQ length (Fig. 6A). However, we observed that the nuclei contained puncta with higher Cherry fluorescence than the diffuse background, and that longer polyQ sequences led to more, brighter and larger puncta (Fig. 6A). This conclusion was confirmed by measuring the ratio of fluorescence intensity in puncta to that of the diffuse background, which showed that the ratio correlated with polyQ length (Fig. 6B). These results suggest that there are at least two potential mechanisms by which the increased tendency of AR to self-associate as polyQ length increases may decrease its transcriptional activity, i.e. a decreased tendency to translocate to the nucleus and an increased tendency to form nuclear puncta that might sequester AR and thus limit its access to the transcriptional machinery.

Discussion

AR plays a key role in the development of CRPC. Understanding the mechanisms that control the transcriptional activity of AR and that underlie the emergence of androgen-independent AR activity is thus critical to design effective therapies to treat this devastating disease. The presence in AR of a large, unstructured NTD rich in low complexity sequences, together with growing evidence that this type of sequence often mediates incorporation of proteins into phase-separated membraneless organelles(38), suggest that one function of the AR NTD may be to mediate phase separation or at least to enhance the localization of AR into biomolecular condensates, thus modulating AR availability and/or activity. The study presented here now shows that low-complexity regions of both AR and KDM4A are involved in interactions between the two proteins in vitro and hence might contribute to epigenetic regulation of AR activity. Moreover, we find that the length of the polyQ sequence correlates with the tendencies of AR-NTD to phase separate in vitro, to remain in the cytoplasm and to form puncta inside the nucleus in the presence of DHT. Altogether, these findings lead us to propose the hypothesis that polyQ-dependent

phase separation of AR may provide a means of negatively regulating its transcriptional activity. Other explanations of our data are possible, however, and further studies will be needed to establish whether the nuclear foci observed here indeed form by liquid-liquid phase separation in cells, and to learn whether and how focus formation may be related mechanistically to effects on AR-mediated transcription.

Early work showed that the NTD is largely unstructured(30), but some of its sequences exhibit clear propensities to adopt secondary structure(45). In particular, the polyQ sequence of the AR NTD has an intrinsic tendency to form α-helical structure that increases in stability with the length of the sequence(33-35). A leucine-rich region preceding the polyQ sequence also increases its propensity to form α-helical structure and impairs aggregation(35). Conversely, exceedingly long polyQ sequences in AR favor aggregation that most likely involves beta-amyloid structure and can lead to formation of toxic fibrils, causing neurodegenerative disease(23, 24). Since longer polyQ sequences protect against prostate cancer(25, 26) and lead to lower transcriptional activity(27-29) (Fig. 4), it appears that the optimal length of the polyQ sequence may be the result of a trade-off between preventing over-activation of AR (if polyQ is too short) and minimizing the possibility of aggregation that leads to neurodegeneration (if polyQ is too long)(34, 46). The increased stability of the helical structure formed by longer polyQ sequences was proposed to influence aggregation through formation of helical oligomers and to affect the transcriptional activity of AR by altering the strength of protein-protein interactions(34). Our finding that the length of the polyQ sequence correlates with the tendency of the AR NTD to phase separate (Fig. 3) shows that longer polyQ sequences increase self-association. However, it is unclear whether such self-association involves direct formation of oligomers with beta-structure or oligomerization via helix-helix interactions that may or may not convert to beta-amyloid-like structures. It is even plausible that there are multiple pathways to aggregation and the different pathways have different biological consequences (see below). Regardless of these possibilities, the observation that addition of 1,6-hexanediol reverses gel formation

(Fig. 3B) indicates that the self-association of the AR-NTD is reversible and therefore does not involve the irreversible formation of amyloid-like aggregates.

The correlation between polyQ length and propensity of the AR-NTD to phase separate (Fig. 3), together with the inverse correlation between polyQ length and transcriptional activity(27-29) (Fig. 4), suggest that polyQ-dependent phase separation of AR could provide a mechanism to reduce its activity. However, it is still unclear how the activity is reduced, and whether reduction directly arises from increased self-association. Phase separation can increase the efficiency of a biological process by helping to concentrate multiple factors involved in the process in a defined compartment, but can also decrease biological activity by sequestering a protein and thus limiting access to its site of action, or by actively inhibiting it(47). Our analyses of nuclear versus cytoplasmic localization lead us to two hypotheses that are not necessarily exclusive of each other. One hypothesis is based on the observation that longer polyQ sequences hinder the nuclear localization of AR in the absence of hormone (Fig. 5), and postulates that longer polyQ sequences enhance the formation of AR oligomers in the cytoplasm, hindering access to the nuclear import machinery. Note that, although we did not observe formation of puncta containing AR in the cytoplasm, oligomers of limited sizes would not be distinguishable in our experiments. The second hypothesis arises from the observation that longer polyQ sequences led to increased formation of puncta in the nucleus when DHT was present (Fig. 6), and proposes that localization of AR to these puncta sequesters the protein and hence prevents its incorporation into transcriptional complexes. One prediction of this hypothesis is that the AR foci do not correspond to AR-responsive genes, and that localization of AR to its proper loci is decreased with increasing polyQ length.

It is important to realize that other mechanisms may also underlie the reduction of AR activity caused by long polyQ sequences, such as increased affinity for transcriptional repressors. Moreover, it is unclear whether AR can undergo phase separation by itself in a cellular environment, as high concentrations of the AR-NTD are required for phase separation(36). It is also plausible that the AR NTD

facilitates recruitment to membraneless organelles that are formed through phase separation of other proteins. Indeed, the AR NTD was readily incorporated in phase separated liquid droplets formed by speckle-type POZ protein (SPOP), an adaptor for a ubiquitin ligase that helps to recruit substrates for proteasomal degradation(36). Thus, increased recruitment of AR to subcellular compartments for protein degradation might also underlie a decrease in AR activity with increased polyQ length. An additional possibility is that the AR NTD mediates differential incorporation into more than one type of membraneless organelle based on its length. For instance, AR might be recruited to organelles where it is sequestered or degraded and to other organelles where transcription of AR-dependent genes is activated, and where interactions between AR and KDM4s help recruiting either one of the two proteins to the condensate. Longer polyQ sequences may tilt the balance in favor of recruitment to the former type of condensate. Intriguingly, overexpression of the AR NTD was reported to delay progression of prostate cancer tumors and CRPC(48, 49). The rationale behind these experiments was to use the AR NTD as a decoy molecule that competes with endogenous AR for binding to proteins required for its activation, and no interaction between the overexpressed AR NTD and endogenous AR was observed by co-immunoprecipitation. However, it is plausible that weak interactions underlying phase separation might not be detected by this method and that the decoy molecules helped to sequester endogenous AR through phase separation. If this notion is correct, overexpression of AR NTD containing a moderately long polyQ sequence (e.g. with 31 Qs) might provide a more effective means to delay progression of CRPC.

Clearly, there are many uncertainties about the ideas discussed above and much needs to be learned to understand how AR becomes activated in an androgen-independent manner in CRPC. Our results bring new hypotheses to this area that need to be tested with further research and that could open new therapeutic avenues.

Materials and methods

Protein expression and purification

Standard recombinant DNA techniques were used to prepare pGEX-KG or pET28 expression vectors encoding the AR and KDM4 fragments used in this study starting from vectors encoding the full-length proteins. His_6-AR-NTD fragments (aa 1-559) containing 12Q, 20Q, 31Q or 49Q were expressed with pET28 vectors at 25 °C in *E. coli* BL21 (DE3) for 20 h with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were resuspended in buffer A (50 mM Tris pH7.5, 250 mM NaCl, 10 mM Imidazole, 6 M Guanidinium HCl, protease inhibitors and 10 mM 2-Mercaptoethanol) and frozen in liquid N2. Cells were thawed and disrupted using an Avestin cell disruptor and centrifuged at 20K, the supernatant was treated with protamine sulfate for 1 h at 4 °C and centrifuged again. The supernatant was loaded onto NiNTA beads and incubated for 2 h at 4 °C. Beads were later washed with 10 column volumes (CVs) of buffer A, 10 CVs buffer A containing 1M NaCl, 5 CVs buffer A and 5 CVs buffer B (50 mM Tris pH 7.5, 250 mM NaCl, 10 mM Imidazole, 6 M Urea and 10 mM 2-Mercaptoethanol). Proteins were eluted with buffer B containing 150 mM Imidazole, and subjected to anion exchange on a HiTrap Q column in 20mM Tris pH7.4 and 6M Urea. Purified proteins were concentrated and diluted 10x in buffer C (20 mM Tris pH7.5, 200 mM NaCl, 1 mM TCEP and 0.1 mM PMSF), concentrated again and dialyzed overnight at 4 ^{o}C in buffer C.

His_6-AR-DBD-LBD (aa 554-919) was expressed with a pET28 vector at 18 °C in *E. coli* BL21 (DE3) for 20 h with 0.2 mM isopropyl IPTG, with 10 μM of DHT and 10 μM $ZnCl_2$ added to the culture media. Cells were re-suspended in buffer A (20 mM Tris-HCl pH 7.5, 0.2 M NaCl, 10% glycerol, 10 μM zinc acetate, and 10 μM DHT) and lysed using an Avestin cell disruptor, centrifuged at 20K and the supernatant incubated with Ni-NTA beads for 2h at 4 °C. The resin was washed with 500 ml of wash buffer (20 mM Tris-HCl pH 7.5, 0.2 M NaCl, 10% glycerol, 10 μM zinc acetate, 10 μM DHT, and 20 mM imidazole), treated with 20U/ml of Benzonase for 1 h at RT and then eluted with wash buffer containing 250 mM imidazole. TEV protease was added to the eluted protein and dialyzed overnight against 20 mM Tris-HCl pH 7.5, 150 mM NaCl,

353 10% glycerol, 10 μM $ZnCl_2$ and 10 μM DHT to remove the His6-tag. The protein was further purified using
354 a Hiload 16/60 Superdex S200 gel filtration column in 20 mM Tris-HCl, pH 7.5, 150mM NaCl, 5% glycerol,
355 1 mM TCEP and 10 μM DHT.

356 Glutathione-S-transferase (GST)-AR-LBD (aa 663-919) was expressed with a pGEX-KG vector at 15
357 °C in *E. coli* BL21 (DE3) for 18-24 h with 0.06 mM IPTG in LB media with 10 μM DHT. Cells were re-
358 suspended in 50 mM Tris pH 7.2, 150 mM NaCl, 2 mM DTT, 10% Glycerol, protease inhibitor cocktail, 5
359 mM EDTA and 10 μM DHT. Cells were lysed on an Avestin cell disruptor and spun at 20K and incubated
360 on a glutathione agarose-resin at 4 °C overnight. The resin was washed with 5 CVs washing buffer (50 mM
361 Tris, pH 8.0, 150 mM NaCl, 5 mM EDTA, 10% Glycerol, 10 μM DHT, 1 mM DTT), washed with 5 CVs
362 benzonase buffer and incubated with 20U/ml of Benzonase at RT for 1-2 h. The resin was washed with
363 TCB (50mM Tris, pH 8.0, 150 mM NaCl, 2.5 mM $CaCl_2$, 10 μM DHT, 10% Glycerol, 0.1% beta-Octyl
364 Glucoside). Eluted protein was further purified by gel filtration on a Hiload16/60 Superdex S200 column
365 in 25 mM Tris pH 7.4, 125 mM NaCl, 1 mM TCEP, 10 μM DHT, 5% glycerol, 0.1% beta-Octyl glucoside.

366 His_6-AR-NTD-DBD (aa 1-632) was expressed with a pET28 vector at 18 °C in *E. coli* BL21 (DE3) for
367 20 h with 0.2 mM IPTG. Cells were re-suspended in buffer A (50 mM Tris pH7.5, 250 mM NaCl, 10 mM
368 Imidazole, protease inhibitors and 10 mM 2-Mercaptoethanol) and frozen in liquid N2. Cells were thawed
369 and disrupted using an Avestin cell disruptor and centrifuged at 20K and supernatant incubated with Ni-
370 NTA beads for 2 h at 4 °C. Beads were washed with 20 CVs of re-suspension buffer and treated with
371 20U/ml of Benzonase for 1 h at RT with rotation. Protein was eluted with re-suspension buffer containing
372 250 mM imidazole and subjected to Gel-filtration chromatography on a Superdex S200 column in 50 mM
373 Tris, 150 mM NaCl and 1 mM TCEP.

374 GST-KDM4A-CD (aa 1-350), GST-KG-KDM4B-CD (aa 1-348) and GST-KG-KDM4C-CD (aa 1-350)
375 were expressed with pGEX-KG vectors at 18 °C in *E. coli* BL21 (DE3) for 20 h with 0.2 mM IPTG in LB media
376 with 0.1 mM $ZnCl_2$ and 0.1 mM $FeSO_4$. Cells were re-suspended in 50 mM Tris pH 7.2, 300 mM NaCl, 2 mM

DTT and protease inhibitors. After fast freezing in liquid N2, cells were broken on an Avestin cell disruptor, spun at 20K and loaded on glutathione agarose-resin at 4°C overnight. The resin was extensively washed with 10 CVs phosphate buffer saline (PBS), 10 CVs PBS containing 1 M NaCl, 5 CVs PBS and 5 CVs Benzonase buffer (25 mM Tris pH 8.0, 50 mM NaCl, 2 mM $MgCl_2$), and treated with 20 U/ml of Benzonase for 1 h at RT with slow rotation. The resin was then washed with 10 CVs PBS containing 1M NaCl, 10 CVs PBS and 5 CVs of thrombin cleavage buffer, then incubated with thrombin to remove the GST-tag for 3 h at RT. The protein was purified by Cation exchange chromatography on a HiTrap S column with 50 mM MES buffer pH6.5, followed by size exclusion chromatography on a Hiload 16/60 Superdex S75 column in 25 mM Tris pH 7.4, 125 mM NaCl and 1 mM TCEP.

His_6-KDM4A(301-708) was expressed with a pET28 vector at 18 °C in *E. coli* BL21 (DE3) for 20 h with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were re-suspended in buffer A (50 mM Tris pH 7.5, 250 mM NaCl, 10 mM Imidazole, 6 M Guanidinium HCl, protease inhibitors and 10 mM 2-Mercaptoethanol). Cells were disrupted and the supernatant was loaded onto NiNTA beads and washed with 10 CVs buffer A, 10 CVs buffer A containing 1M NaCl, 5 CVs buffer A and 5 CVs buffer B (50 mM Tris pH 7.5, 250 mM NaCl, 10 mM Imidazole, 6 M Urea and 10 mM 2-Mercaptoethanol). Proteins were eluted with buffer B containing 150 mM Imidazole and further purified by gel filtration on a S75 column in 20 mM Tris pH7.5, 200 mM NaCl, 1 mM TCEP.

His_6-KDM4A(703-1064) was expressed with a pET28 vector at 16 °C in *E. coli* BL21 (DE3) for 20 h with 0.5 mM IPTG in LB media and pellet re-suspended in buffer RB (50mM Tris pH=7.5, 250 mM NaCl 4 mM Imidazole, protease inhibitors). After protein binding, the Ni-NTA beads were washed with RB and treated with Benzonase in 10 ml Benzonase buffer for 2h RT, washed again with RB and eluted with elution buffer (50 mM Tris pH 7.5, 250 mM NaCl 250-500 mM Imidazole). Proteins were then subjected to exclusion chromatography on a S200 column in 25 mM Tris pH 7.0, 125 mM NaCl and 1 mM TCEP.

GST-AR (93-495) was expressed with a pGEX-KG vector at 16 °C in *E. coli* BL21 (DE3) for 20 h with 0.5 mM IPTG in LB media. Cells were re-suspended in PBS, protease inhibitors cocktail and 5 mM DTT. Cells were broken on an Avestin cell disruptor, spun at 20K and incubated with glutathione agarose-resin at 4°C overnight. The resin was extensively washed with 10 CVs PBS, 10 CVs PBS containing 1 M NaCl, 5 CVs PBS and 5 CVs Benzonase buffer (25mM Tris pH8.0, 50mM NaCl, 2 mM MgCl2), and treated with 20U/ml of Benzonase for 1-2 h at RT with rotation. The beads were washed with 5 CVs PBS and treated with 20U/ml of TEV protease to remove the GST tag overnight at 4°C. Collected fractions were subjected to anion exchange chromatography on a HiTrap-Q column with 20mM Bis-Tris buffer pH 6.4 and size exclusion chromatography on a S75 column in 25 mM Tris pH 7.4, 125 mM NaCl and 2 mM TCEP.

To generate uniformly ^{15}N-labeled proteins for NMR studies, bacteria were grown in M9 minimal medium supplemented with $^{15}NH_4Cl$ (CIL, Andover, MA) as the sole nitrogen source. Proteins were purified as described above.

NMR spectroscopy

All NMR spectra were acquired at 25 °C on a Varian INOVA 600 MHz spectrometer. Samples for ^{1}H-^{15}N TROSY-HSQC measurements contained 40-50 μM of the ^{15}N labelled protein and unlabeled protein. Experiments with ^{15}N-KDM4A(1-350) or ^{15}N-KDM4B(1-348) were performed in 25 mM Tris (pH 7.4), 125 mM NaCl and 1 mM TCEP with 5% D_2O. For samples containing ^{15}N-AR-LBD or ^{15}N-KDM4C(1-350), the buffer was 25 mM Tris (pH 7.4), 125 mM NaCl, 0.1% Octyl β-D-glucopyranoside and 1 mM TCEP with 5% D2O. For the NMR experiments including AR-LBD or AR-DBD+LBD, 1 μM DHT was included in the buffer. Total acquisition times were 2–10 h. NMR data were processed with NMRPipe(50) and analyzed with NMRView(51).

Chemical Cross-linking

Cross-linking reactions were performed with mixtures of 15 μM AR NTD-DBD (aa 1-628) with 15 or 30 μM KDM4A(1-350), or with mixtures of 15 μM KDM4A(1-350) with 15 μM, 30 μM or 60 μM AR-NTD (aa 93-495). The mixtures were incubated with 1 mM bis-(sulfosuccinimidyl)suberate (BS3) for 1 h at RT and the reactions quenched with 25 mM Tris (pH 8.0). Samples of 18 μg of protein were loaded onto SDS gels for PAGE and coomassie blue staining.

Gel formation and turbidity assays

AR-NTD proteins containing 12Q, 20Q, 31Q or 49Qs at ~50uM concentration were dialyzed against a gelation buffer containing 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 20 mM BME, 0.5 mM EDTA and 0.1 mM PMSF overnight. The proteins were then concentrated to 500 μM and incubated at RT for 24-48 h until gel formation was observed. For turbidity assays, protein samples in 25 mM Tris pH 7.5, 150 mM NaCl and 1 mM TCEP were concentrated to 100 μM and then diluted in the same buffer to the various concentrations shown in Figs. 3C,D. The proteins were incubated at 4 °C or RT for 1 h, and the OD at 400 nm was then measured.

Luciferase Assay

LNCaP and PC3 cells were co-transfected with an ARE-luc construct containing three androgen response elements (ARE) ligated in tandem to a luciferase reporter(52), along with a control vector pcDNA3.1-FLAG or an AR-expression vector (pcDNA3.1-FLAG-AR) containing different polyQ sequence in the presence or absence of DHT. LNCaP cells were maintained in RPMI phenol red free with 5% charcoal stripped FBS during the assay. Transfections were carried out using FuGENE HD (Promega, Catalog. E2311) for LNCaP cells and Lipofectamine 3000 Kit for PC3 cells (Thermo Fisher, Catalog. L30000) following the instructions from the manufacturer. Cells were treated with vehicle or 10 nM Dihydrotestosterone (DHT) 24 h after transfection. Cell extracts were prepared 48 h after transfection and assayed for luciferase activity using

the Promega luciferase detection kit. Luciferase activities were normalized to co-transfected β-galactosidase activity.

Western blot analysis

Protein samples from LNCaP cells used to measure luciferase activity (described in Fig. 4A) were loaded into 4% to 15% SDS-PAGE (Bio-Rad) and subjected to electrophoretic analysis and subsequent blotting. Nitrocellulose membranes were incubated with the primary antibody (overnight at 4°C) and the relevant secondary antibodies (1 h at room temperature). The antibodies were purchased from MilliporeSigma: anti-AR (catalog 06-680); anti-FLAG M2 (catalog F3165) and anti-β-actin (catalog no. A5441). For the experiments shown in Fig. 5C, LNCaP cells transiently transfected with Cherry-AR containing different polyQ lengths were blotted against the same anti-AR and anti-β-actin antibodies.

AR localization assessment

LNCaP cells transiently transfected with Cherry-AR (full-length) containing different lengths of glutamine repeat (12Q, 20Q, 31Q or 49Q) were seeded onto coverslips in RPMI-1640 with 10% Charcoal-stripped fetal bovine serum (FBS), and continued to grow for 48 h under the same starvation conditions. Coverslips were fixed with methanol for 10 min at -20°C, washed with PBS and mounted on slides with mounting media containing 4′,6-diamidino-2-phenylindole (DAPI). Cells were imaged as z-stacks using a widefield Delta Vision Fluorescence microscope. Maximum projections were done before nuclear/cytoplasmic localization of cherry-AR-FL was quantified using the ImageJ Macro “Intensity Ratio Nuclei Cytoplasm Tool”. The data were plotted as the percentage of nuclear and cytoplasmic distribution of Cherry AR. More than 50 cells were analyzed per experimental condition.

Puncta formation assay

LNCaP cells stably expressing Cherry-AR containing different lengths of glutamine repeat (12Q, 20Q, 31Q or 49Q) were seeded onto coverslips in RPMI-1640 with 10% Charcoal stripped FBS and continued to grow for 48 h under the same starvation conditions. Twenty four h before fixation, the cells were treated with 10 nM DHT to induce AR nuclear localization. Coverslips were fixed with methanol for 10 min at -20 °C and mounted on coverslips with mounting solution containing 4',6-diamidino-2-phenylindole (DAPI). A widefield Delta Vision Fluorescence microscope was used to take z-stacks images through whole cells, maximum projections were performed, and the puncta intensity of cherry-AR-FL was quantified using ImageJ. More than 50 cells were analyzed per experiment per condition, and there were ~5-45 puncta per cell. Intensity ratio (IR) was calculated as the ratio of fluorescent intensity in the puncta minus the background divided by the fluorescent intensity in the nucleoplasm minus the background. The background intensity was obtained from an image of an area where there were no cells.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

We thank Yilun Sun for technical assistance. This work was supported by grants from the Cancer Prevention and Research Institute of Texas (CPRIT) (RP120717-AC and RP120717-P1 to Z.P.L.; RP120717-P3 to J.R.), NIH (RO1 CA215063 to Z.P.L.), and from the Welch Foundation (I-1304 to J.R.).

References

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin. 2010;60(5):277-300.
2. Maximum androgen blockade in advanced prostate cancer: an overview of 22 randomised trials with 3283 deaths in 5710 patients. Prostate Cancer Trialists' Collaborative Group. Lancet. 1995;346(8970):265-9.
3. Attard G, Cooper CS, de Bono JS. Steroid hormone receptors in prostate cancer: a hard habit to break? Cancer Cell. 2009;16(6):458-62.
4. Eisenberger MA, Walsh PC. Early androgen deprivation for prostate cancer? N Engl J Med. 1999;341(24):1837-8.
5. Chen CD, Welsbie DS, Tran C, Baek SH, Chen R, Vessella R, et al. Molecular determinants of resistance to antiandrogen therapy. Nat Med. 2004;10(1):33-9.
6. Harris WP, Mostaghel EA, Nelson PS, Montgomery B. Androgen deprivation therapy: progress in understanding mechanisms of resistance and optimizing androgen depletion. Nat Clin Pract Urol. 2009;6(2):76-85.
7. Nagy L, Kao HY, Chakravarti D, Lin RJ, Hassig CA, Ayer DE, et al. Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase. Cell. 1997;89(3):373-80.
8. Shang Y, Myers M, Brown M. Formation of the androgen receptor transcription complex. Mol Cell. 2002;9(3):601-10.
9. Yamane K, Toumazou C, Tsukada Y, Erdjument-Bromage H, Tempst P, Wong J, et al. JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. Cell. 2006;125(3):483-95.
10. Wissmann M, Yin N, Muller JM, Greschik H, Fodor BD, Jenuwein T, et al. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. Nat Cell Biol. 2007;9(3):347-53.
11. Shin S, Janknecht R. Activation of androgen receptor by histone demethylases JMJD2A and JMJD2D. Biochem Biophys Res Commun. 2007;359(3):742-6.
12. Perry AS, Watson RW, Lawler M, Hollywood D. The epigenome as a therapeutic target in prostate cancer. Nat Rev Urol. 2010;7(12):668-80.
13. Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, et al. Global histone modification patterns predict risk of prostate cancer recurrence. Nature. 2005;435(7046):1262-6.
14. Yegnasubramanian S, Kowalski J, Gonzalgo ML, Zahurak M, Piantadosi S, Walsh PC, et al. Hypermethylation of CpG islands in primary and metastatic human prostate cancer. Cancer Res. 2004;64(6):1975-86.
15. Simental JA, Sar M, Lane MV, French FS, Wilson EM. Transcriptional activation and nuclear targeting signals of the human androgen receptor. J Biol Chem. 1991;266(1):510-8.
16. Doesburg P, Kuil CW, Berrevoets CA, Steketee K, Faber PW, Mulder E, et al. Functional in vivo interaction between the amino-terminal, transactivation domain and the ligand binding domain of the androgen receptor. Biochemistry. 1997;36(5):1052-64.
17. Wilson EM. Analysis of interdomain interactions of the androgen receptor. Methods Mol Biol. 2011;776:113-29.
18. Gelmann EP. Molecular biology of the androgen receptor. J Clin Oncol. 2002;20(13):3001-15.
19. Cao S, Zhan Y, Dong Y. Emerging data on androgen receptor splice variants in prostate cancer. Endocr Relat Cancer. 2016;23(12):T199-T210.
20. Shao C, Yu B, Liu Y. Androgen receptor splicing variant 7: Beyond being a constitutively active variant. Life Sci. 2019;234:116768.

538 21. Duan L, Chen Z, Lu J, Liang Y, Wang M, Roggero CM, et al. Histone lysine demethylase KDM4B
539 regulates the alternative splicing of the androgen receptor in response to androgen deprivation. Nucleic
540 Acids Res. 2019;47(22):11623-36.
541 22. Tan MH, Li J, Xu HE, Melcher K, Yong EL. Androgen receptor: structure, role in prostate cancer and
542 drug discovery. Acta Pharmacol Sin. 2015;36(1):3-23.
543 23. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene
544 mutations in X-linked spinal and bulbar muscular atrophy. Nature. 1991;352(6330):77-9.
545 24. Jochum T, Ritz ME, Schuster C, Funderburk SF, Jehle K, Schmitz K, et al. Toxic and non-toxic
546 aggregates from the SBMA and normal forms of androgen receptor have distinct oligomeric structures.
547 Biochim Biophys Acta. 2012;1822(6):1070-8.
548 25. Stanford JL, Just JJ, Gibbs M, Wicklund KG, Neal CL, Blumenstein BA, et al. Polymorphic repeats in
549 the androgen receptor gene: molecular markers of prostate cancer risk. Cancer Res. 1997;57(6):1194-8.
550 26. Kumar R, Atamna H, Zakharov MN, Bhasin S, Khan SH, Jasuja R. Role of the androgen receptor
551 CAG repeat polymorphism in prostate cancer, and spinal and bulbar muscular atrophy. Life Sci.
552 2011;88(13-14):565-71.
553 27. Chamberlain NL, Driver ED, Miesfeld RL. The length and location of CAG trinucleotide repeats in
554 the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Res.
555 1994;22(15):3181-6.
556 28. Tut TG, Ghadessy FJ, Trifiro MA, Pinsky L, Yong EL. Long polyglutamine tracts in the androgen
557 receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. J
558 Clin Endocrinol Metab. 1997;82(11):3777-82.
559 29. Irvine RA, Ma H, Yu MC, Ross RK, Stallcup MR, Coetzee GA. Inhibition of p160-mediated
560 coactivation with increasing androgen receptor polyglutamine length. Hum Mol Genet. 2000;9(2):267-74.
561 30. Lavery DN, McEwan IJ. Structural characterization of the native NH2-terminal transactivation
562 domain of the human androgen receptor: a collapsed disordered conformation underlies structural
563 plasticity and protein-induced folding. Biochemistry. 2008;47(11):3360-9.
564 31. De Mol E, Fenwick RB, Phang CT, Buzon V, Szulc E, de la Fuente A, et al. EPI-001, A Compound
565 Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen
566 Receptor. ACS Chem Biol. 2016;11(9):2499-505.
567 32. De Mol E, Szulc E, Di Sanza C, Martinez-Cristobal P, Bertoncini CW, Fenwick RB, et al. Regulation
568 of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General
569 Transcription Regulators. Structure. 2018;26(1):145-52 e3.
570 33. Davies P, Watt K, Kelly SM, Clark C, Price NC, McEwan IJ. Consequences of poly-glutamine repeat
571 length for the conformation and folding of the androgen receptor amino-terminal domain. J Mol
572 Endocrinol. 2008;41(5):301-14.
573 34. Escobedo A, Topal B, Kunze MBA, Aranda J, Chiesa G, Mungianu D, et al. Side chain to main chain
574 hydrogen bonds stabilize a polyglutamine helix in a transcription factor. Nat Commun. 2019;10(1):2034.
575 35. Eftekharzadeh B, Piai A, Chiesa G, Mungianu D, Garcia J, Pierattelli R, et al. Sequence Context
576 Influences the Structure and Aggregation Behavior of a PolyQ Tract. Biophys J. 2016;110(11):2361-6.
577 36. Bouchard JJ, Otero JH, Scott DC, Szulc E, Martin EW, Sabri N, et al. Cancer Mutations of the Tumor
578 Suppressor SPOP Disrupt the Formation of Active, Phase-Separated Compartments. Mol Cell.
579 2018;72(1):19-36 e8.
580 37. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: low
581 complexity sequence domains form dynamic fibers within hydrogels. Cell. 2012;149(4):753-67.
582 38. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular
583 biochemistry. Nat Rev Mol Cell Biol. 2017;18(5):285-98.

584 39. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, et al. Transcription Factors
585 Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell.
586 2018;175(7):1842-55 e16.
587 40. Charbonnier S, Zanier K, Masson M, Trave G. Capturing protein-protein complexes at equilibrium:
588 the holdup comparative chromatographic retention assay. Protein Expr Purif. 2006;50(1):89-101.
589 41. Rizo J, Rosen MK, Gardner KH. Enlightening molecular mechanisms through study of protein
590 interactions. J Mol Cell Biol. 2012;4(5):270-83.
591 42. Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, et al. Valence and patterning of
592 aromatic residues determine the phase behavior of prion-like domains. Science. 2020;367(6478):694-9.
593 43. Peskett TR, Rau F, O'Driscoll J, Patani R, Lowe AR, Saibil HR. A Liquid to Solid Phase Transition
594 Underlying Pathological Huntingtin Exon1 Aggregation. Mol Cell. 2018;70(4):588-601 e6.
595 44. Seibel NM, Eljouni J, Nalaskowski MM, Hampe W. Nuclear localization of enhanced green
596 fluorescent protein homomultimers. Anal Biochem. 2007;368(1):95-9.
597 45. McEwan IJ. Intrinsic disorder in the androgen receptor: identification, characterisation and
598 drugability. Mol Biosyst. 2011.
599 46. Oppong E, Stier G, Gaal M, Seeger R, Stoeck M, Delsuc MA, et al. An Amyloidogenic Sequence at
600 the N-Terminus of the Androgen Receptor Impacts Polyglutamine Aggregation. Biomolecules. 2017;7(2).
601 47. Lyon AS, Peeples WB, Rosen MK. A framework for understanding the functions of biomolecular
602 condensates across scales. Nat Rev Mol Cell Biol. 2020.
603 48. Quayle SN, Mawji NR, Wang J, Sadar MD. Androgen receptor decoy molecules block the growth
604 of prostate cancer. Proc Natl Acad Sci U S A. 2007;104(4):1331-6.
605 49. Myung JK, Wang G, Chiu HH, Wang J, Mawji NR, Sadar MD. Inhibition of androgen receptor by
606 decoy molecules delays progression to castration-recurrent prostate cancer. PLoS One.
607 2017;12(3):e0174134.
608 50. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. Nmrpipe - A Multidimensional Spectral
609 Processing System Based on Unix Pipes. Journal of Biomolecular Nmr. 1995;6(3):277-93.
610 51. Johnson BA, Blevins RA. Nmr View - A Computer-Program for the Visualization and Analysis of
611 Nmr Data. Journal of Biomolecular Nmr. 1994;4(5):603-14.
612 52. Roggero CM, Jin L, Cao S, Sonavane R, Kopplin NG, Ta HQ, et al. A detailed characterization of
613 stepwise activation of the androgen receptor variant 7 in prostate cancer cells. Oncogene.
614 2021;40(6):1106-17.
615

616

617

618

Figure legends

Figure 1. Low complexity sequences are involved in binding between AR and KDM4A. (A) Domain diagrams of AR and KDM4A. Selected residue numbers indicating the protein N- or C-termini, or approximate domain boundaries, are indicated above the diagrams. Domain names are shown below the diagrams (Tu = Tudor). (B-E) Superpositions of ^{1}H-^{15}N TROSY-HSQC spectra of ^{15}N-KDM4A(1-350) alone (black contours) or in the presence of equimolar amounts of unlabeled AR-NTD (B), AR-NTD-DBD (C), AR-LBD (D) or AR-DBD-LBD (E). (F-G) Superpositions of ^{1}H-^{15}N TROSY-HSQC spectra of ^{15}N-AR-LBD alone (black contours) or in the presence (red contours) of equimolar amounts of KDM4A(301-708) or KDM4A(703-1064).

Figure 2. The catalytic domain of KDM4A can be efficiently cross-linked with the AR N-terminal region. (A) Samples containing 15 μM AR-NTD-DBD were incubated with 1 or 2 equivalents of KDM4A(1-350) and 1 mM BS3 for 1 h at RT, and the reactions were analyzed by SDS-PAGE followed by coomassie Blue staining (two right-most lanes). The three lanes on the left show samples of AR-NTD-DBD and KDM4A(1-350) alone or together without BS3. (B) Samples containing 15 μM KDM4A(1-350) were incubated with 1, 2 or 4 equivalents of AR(93-495) and 1 mM BS3 for 1 h at RT, and the reactions were analyzed by SDS-PAGE followed by coomassie Blue staining (three right-most lanes). The three lanes on the left show samples of KDM4A(1-350) and AR(93-495) alone or together without BS3. Note that the staining efficiency of AR fragments containing the NTD with coomassie Blue is substantially lower than is usual in proteins because of the paucity of basic residues in its sequence.

Figure 3. Longer polyQ sequences increase the tendency of AR-NTD to form gels. (A) Samples containing AR-NTD with different polyQ length were concentrated to 600 μM (12Q), 520 μM (20Q), 520 μM (31Q) and 480 μM (49Q) and incubated at RT for 24-48 h until gel formation was observed. Images of the

643 resulting samples are shown. (B) Images of a 500 μM sample of AR-NTD-31Q that was allowed to form
644 gels for 48 h before (left) and after (right) adding 10% 1,6-hexanediol (HD). (C-D) Turbidity assays with
645 samples of AR-NTD containing different polyQ length. Samples were concentrated to 100 μM,
646 immediately diluted to the indicated concentrations and incubated for 1 h at 4 °C (B) or RT (C). The OD at
647 400 nm was then measured. Bars show averages of values measured in three independent experiments
648 performed under the same conditions and error bars show standard deviations.
649
650 **Figure 4.** Longer polyQ sequences decrease the transcriptional activity of AR. (A,C) Luciferase reporter
651 assays were performed with LNCaP cells and PC3 cells transfected with vectors expressing full-length AR
652 containing 12, 20, 31 or 49 glutamines in the polyQ repeat. Luciferase activity was measured in the
653 absence and presence of DHT, and the activities were normalized to co-transfected β-galactosidase
654 activity. Bars show averages of values measured in three independent experiments performed under the
655 same conditions and error bars show standard deviations. (B) Western blot showing the levels of AR in
656 the LNCaP cells used for the assays shown in panel (A). Top: total AR using an anti-AR antibody; middle:
657 overexpressed FLAG-AR using an anti-Flag antibody; and bottom: anti-β-actin used as a loading control.
658
659 **Figure 5.** Longer polyQ sequences decrease the nuclear localization of AR in the absence of hormone. (A)
660 Representative images of LNCaP cells transiently transfected with Cherry-AR containing different lengths
661 of glutamine repeat (12Q, 20Q, 31Q or 49Q), in the absence of DHT. Red corresponds to Cherry
662 fluorescence and blue to DAPI fluorescence. (B) AR distribution in the nucleus (grey bars) or the cytoplasm
663 (red bars) normalized to the total in cells transiently transfected with Cherry-AR-12Q, -20Q, -31Q or -49Q.
664 More than 50 cells were analyzed for each condition. Bars show averages and error bars show standard
665 deviations. (C) Western blot showing the expression levels of Cherry-AR and endogenous AR proteins
666 using an anti-AR antibody and anti-β-actin antibody as a loading control.

667

668 **Figure 6.** Longer polyQ sequences increase the tendency of AR to localize to nuclear puncta in the
669 presence of DHT. (A) Representative images of cells stably expressing Cherry-AR-12Q, -20Q, -31Q or -49Q,
670 treated with DHT 24 h before fixation. Red corresponds to Cherry fluorescence and blue to DAPI
671 fluorescence. (B) Puncta intensity ratios (IRs) in cells stably expressing Cherry-AR-12Q, -20Q, -31Q or -49Q
672 were calculated as the ratio of fluorescent intensity in the puncta minus the background divided by
673 fluorescent intensity in the nucleoplasm minus the background. The background intensity corresponds to
674 an image taken from an area where there were no cells. Bars represent average IRs calculate from
675 measurements performed in at least 50 cells under each condition (5-45 puncta per cell) and errors bars
676 show standard deviations. Statistical significance and P values were determined by one-way analysis of
677 variance (ANOVA) with Holm-Sidak test (** $P < 0.01$; *** $P < 0.001$).

678

679

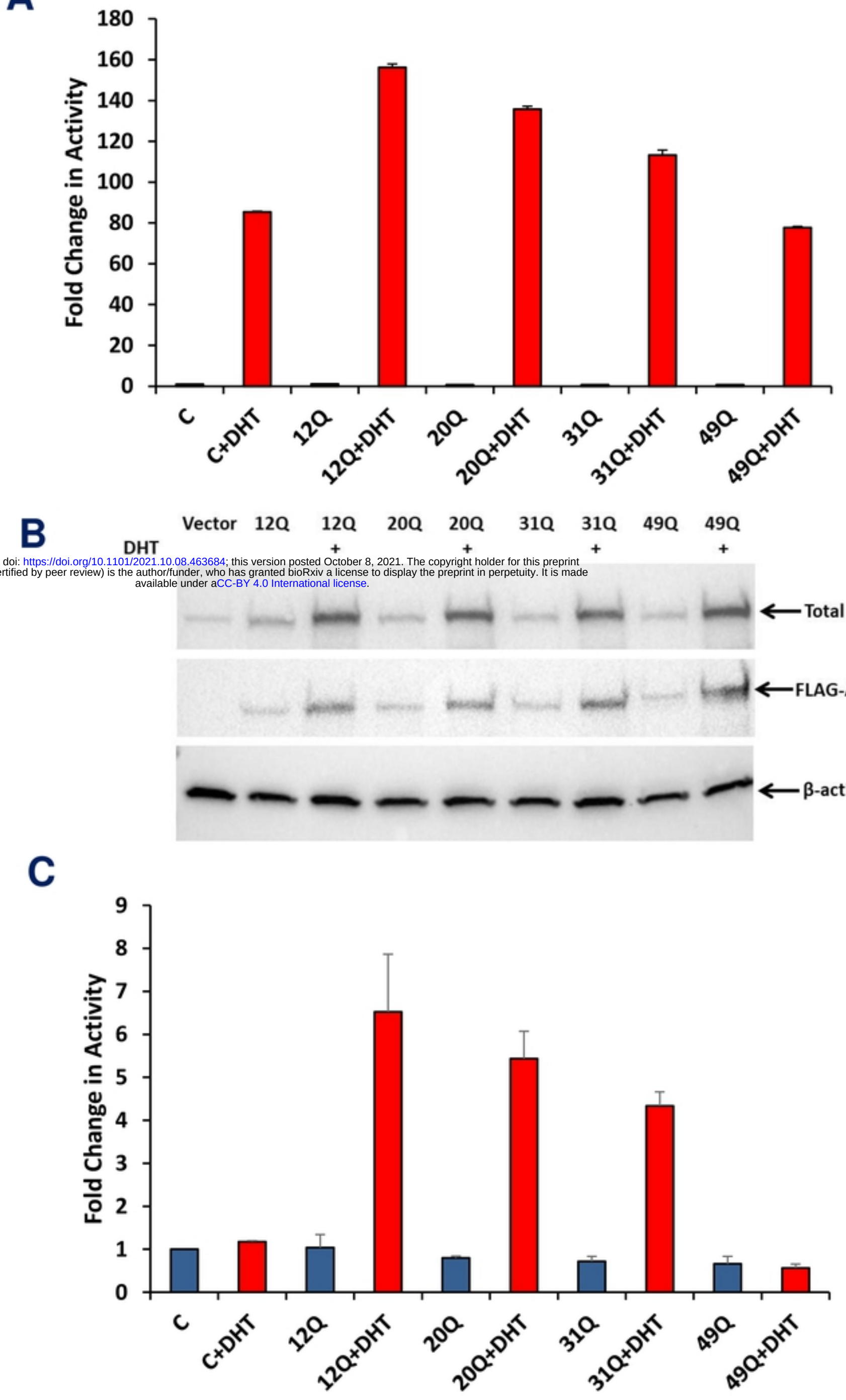
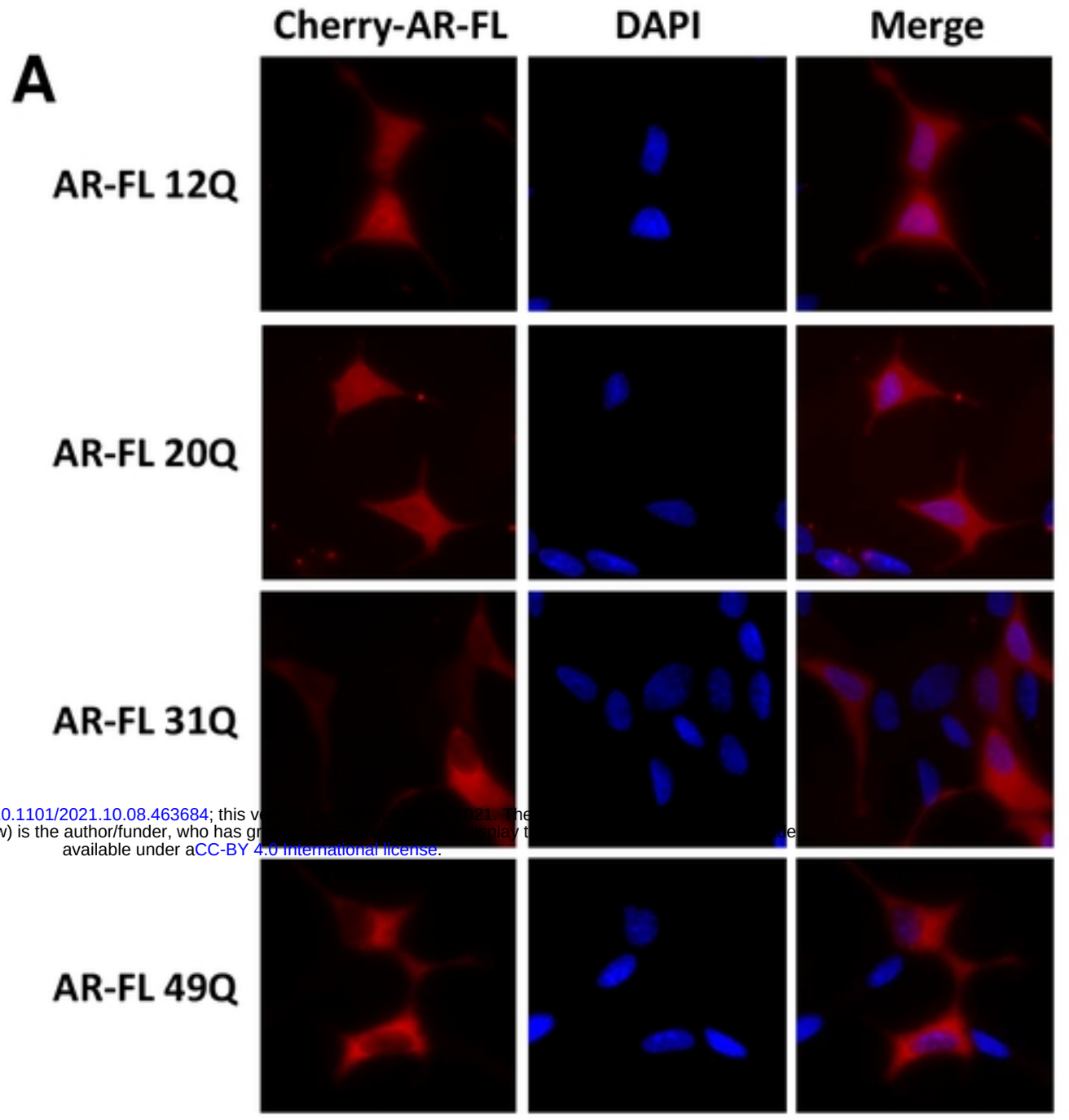
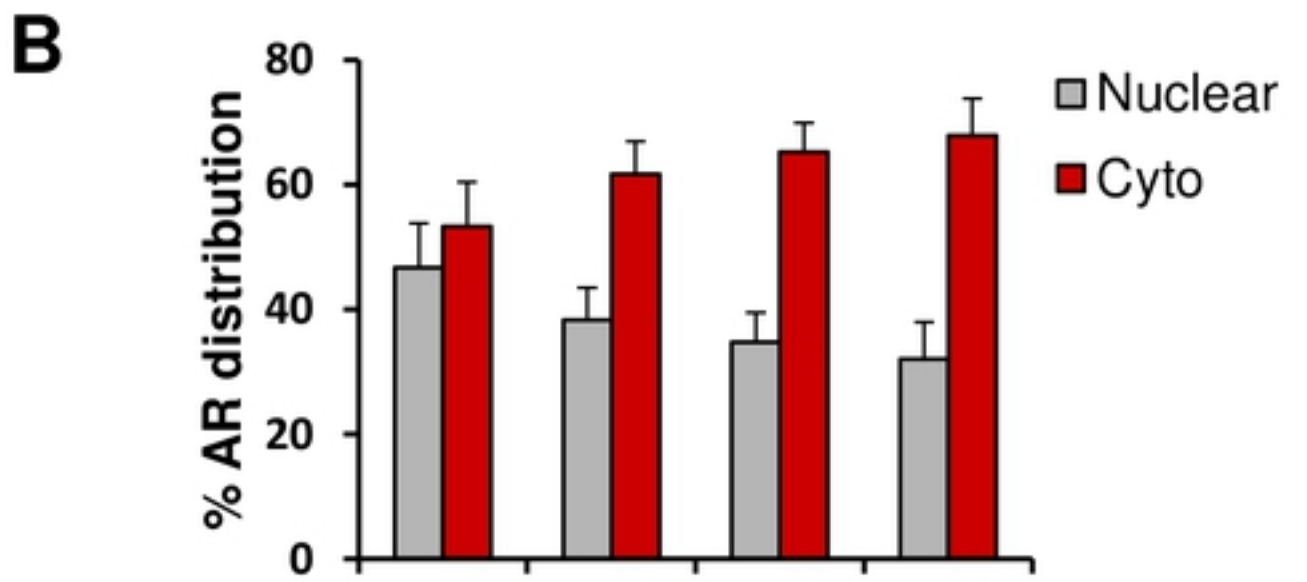


Figure 4
Roggero et al.

Figure 4





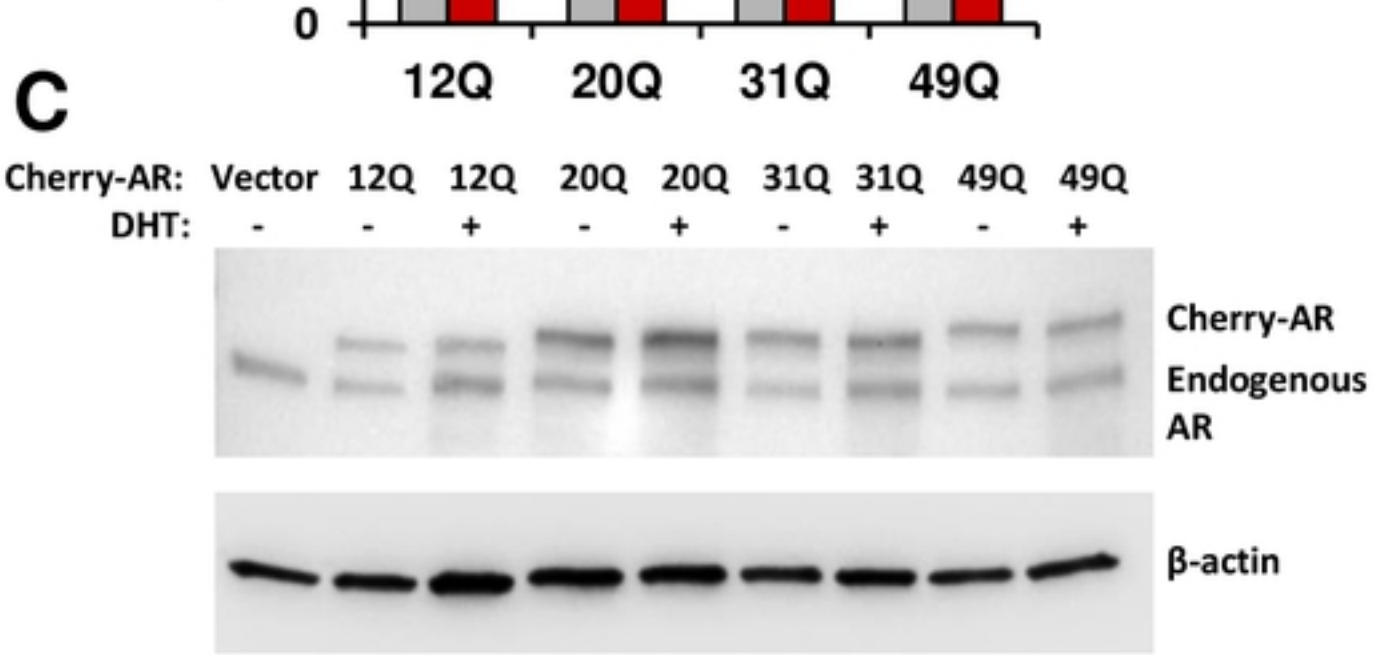
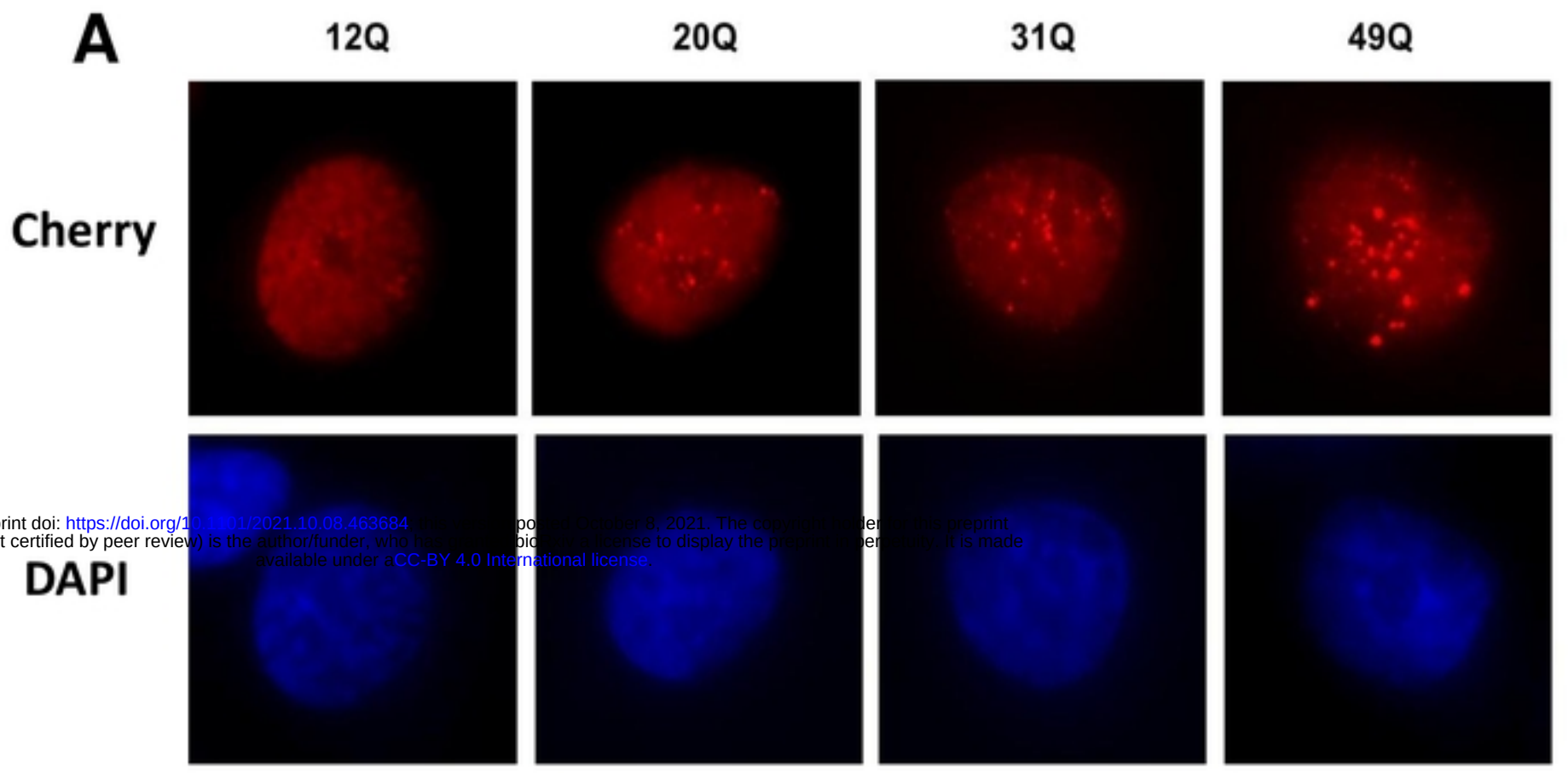


Figure 5
Roggero et al.

Figure 5



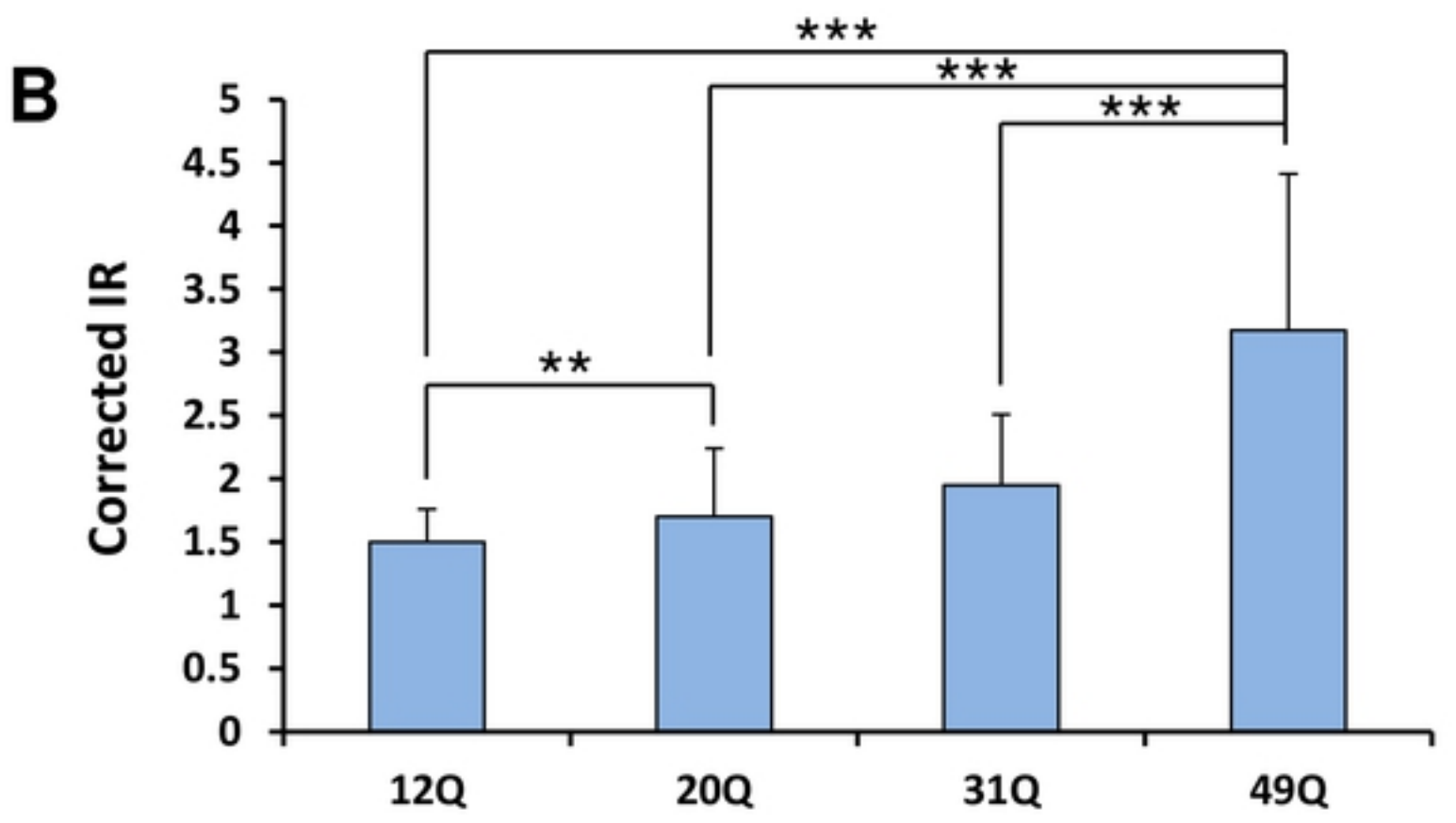


Figure 6
Roggero et al.

Figure 6

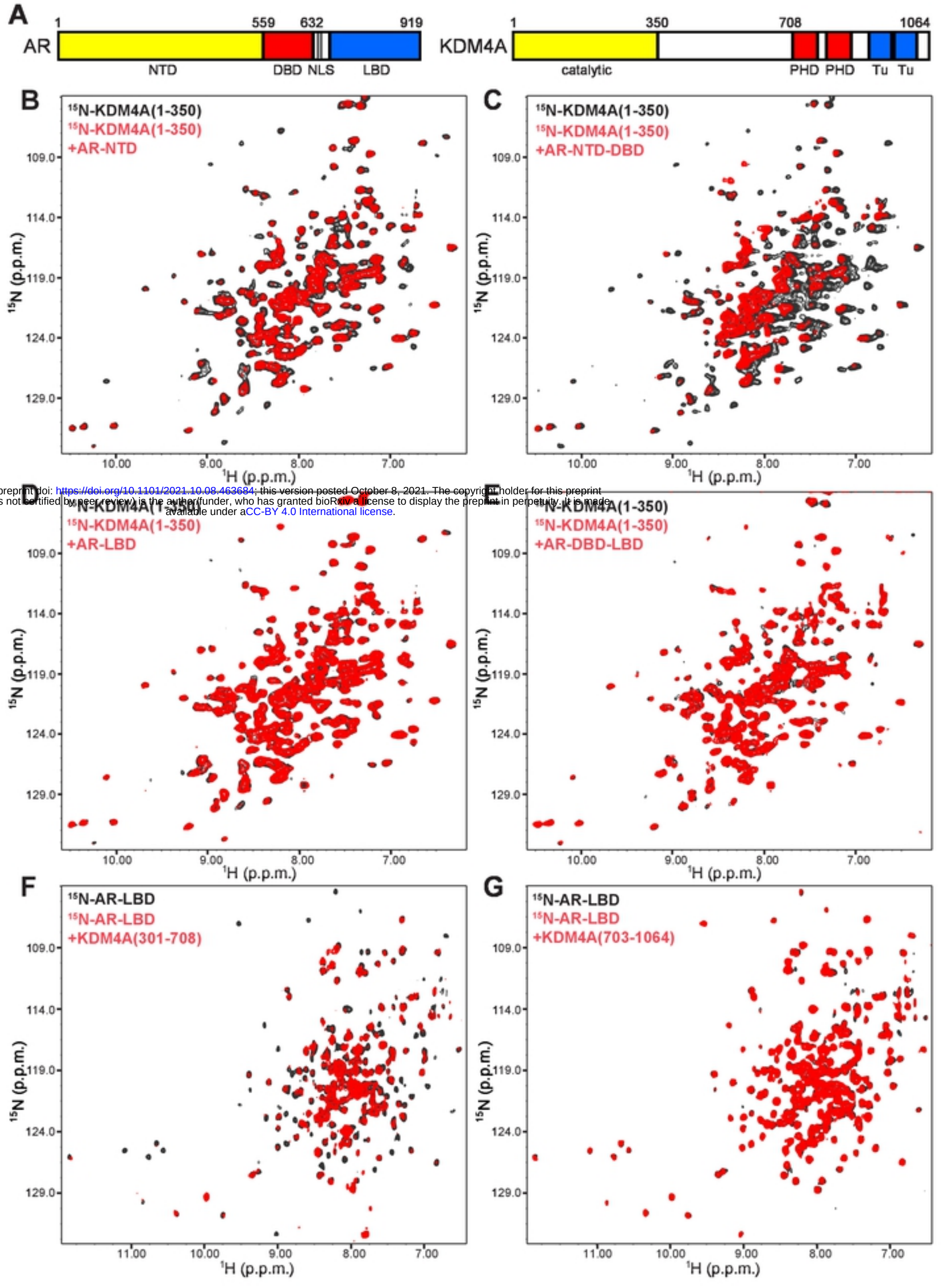


Figure 1
Roggero et al.

Figure 1

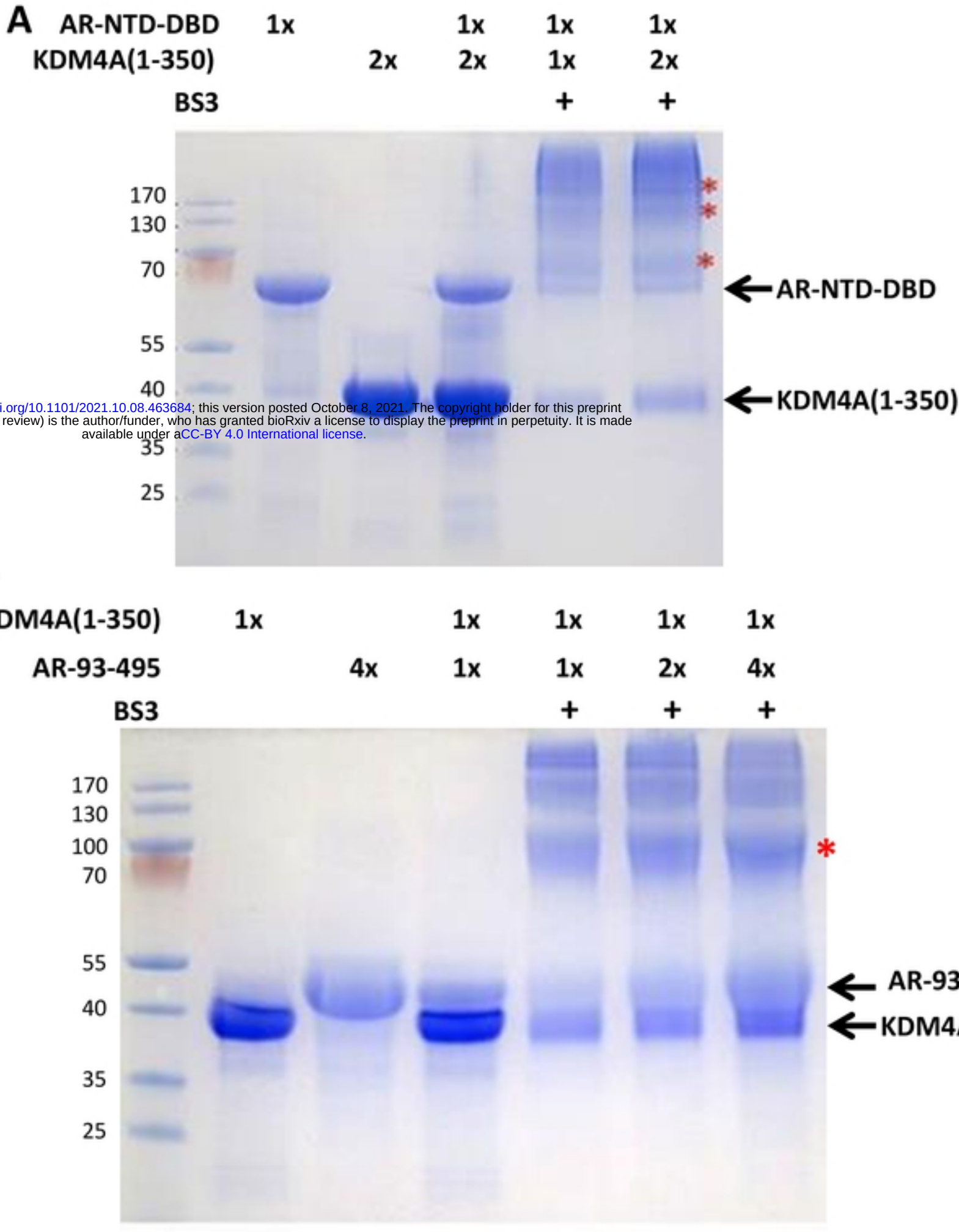


Figure 2
Roggero et al.

Figure 2

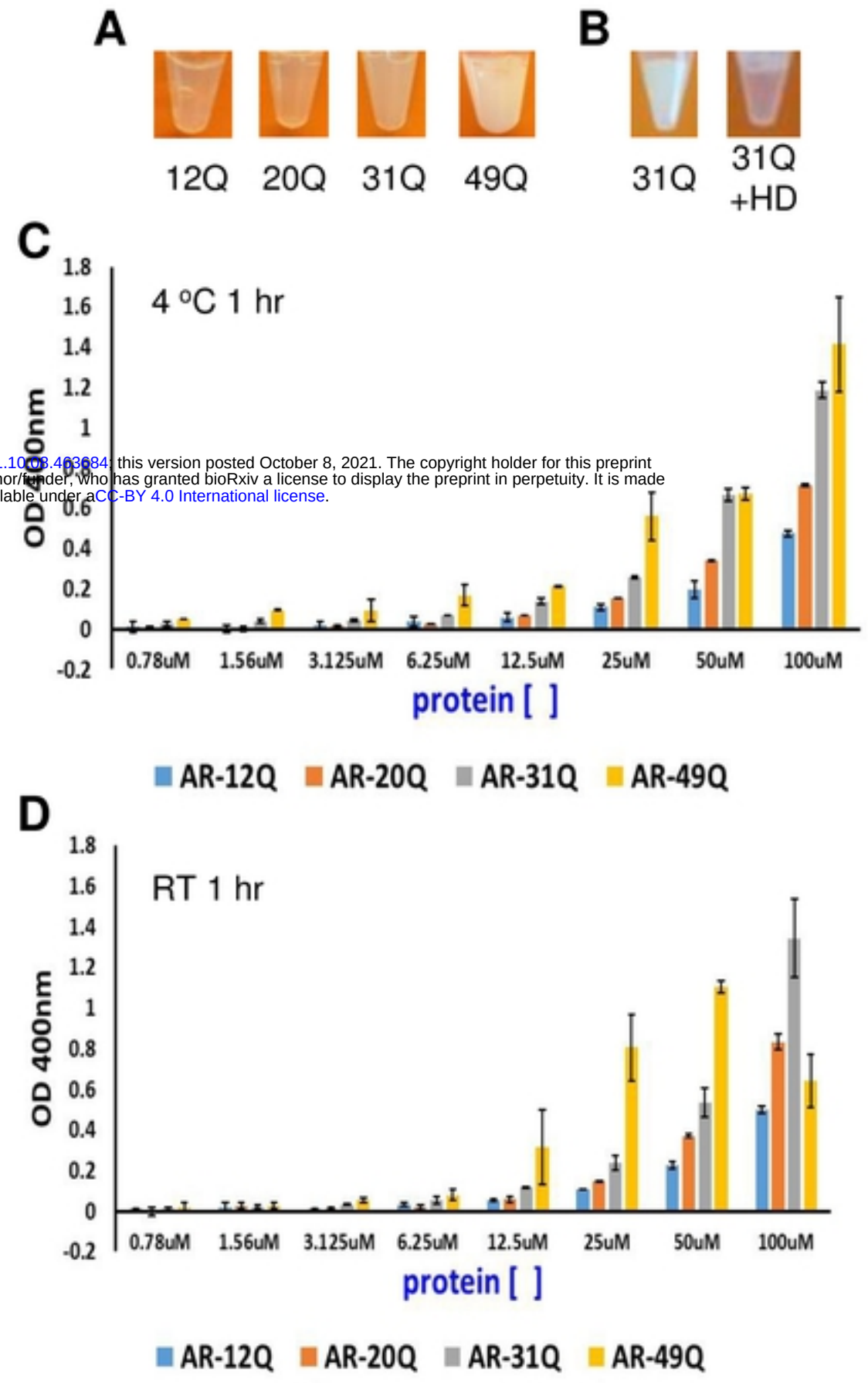


Figure 3
Roggero et al.

Figure 3