



HAL
open science

Assessment of machine-learning predictions for the Mediator complex subunit MED25 ACID domain interactions with transactivation domains

Didier Monté, Zoé Lens, Frédérique Dewitte, Vincent Villeret, Alexis Verger

► To cite this version:

Didier Monté, Zoé Lens, Frédérique Dewitte, Vincent Villeret, Alexis Verger. Assessment of machine-learning predictions for the Mediator complex subunit MED25 ACID domain interactions with transactivation domains. *FEBS Letters*, 2024, 598 (7), pp.758-773. 10.1002/1873-3468.14837 . hal-04676267

HAL Id: hal-04676267

<https://hal.science/hal-04676267v1>

Submitted on 27 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Assessment of machine-learning predictions for MED25 ACID domain interactions with transactivation domains

Didier Monté¹, Zoé Lens¹, Frédérique Dewitte¹, Vincent Villeret¹ & Alexis Verger^{1,*}

¹CNRS EMR 9002 Integrative Structural Biology, Inserm U 1167 – RID-AGE, Univ. Lille, CHU Lille, Institut Pasteur de Lille, F-59000 Lille, France

*Correspondence should be addressed to AV (Alexis.Verger@univ-lille.fr)

Author contributions: DM, VV and AV conceived and supervised the study. AV performed the structure predictions and analyzed data with DM and VV. ZL and FD performed the biochemical experiments. AV wrote the manuscript with contributions from DM and VV. All authors contributed to the reviewing and editing of the manuscript.

Abstract

Human Mediator complex subunit MED25 binds transactivation domains (TADs) present in various cellular and viral proteins using two binding interfaces found on opposite sides of its ACID domain, and referenced as H1 and H2. Here, we use and compare deep learning methods to characterize Human MED25-TADs interfaces and assess the predicted models to published experimental data. For the H1 interface, AlphaFold produces predictions with high reliability scores that agree well with experimental data, while the H2 interface predictions appear inconsistent, preventing reliable binding modes. Despite these limitations, we experimentally assess the validity of Lana-1 and IE62 MED25 interface predictions. AlphaFold predictions also suggest the existence of a unique hydrophobic pocket for Arabidopsis MED25 ACID domain.

Introduction

MED25 is a subunit specific to higher eukaryote Mediator complex, an essential component of the RNA polymerase II general transcriptional machinery (1). Many transcriptional activators have been reported to interact with human MED25 through its central activator interacting domain (ACID), also known as prostate tumor overexpressed (PTOV)/activator binding domain (ABD) (2, 3). This included the archetypal acidic transcriptional activation domain of the herpes virus activator VP16 (4–6), the Kaposi's sarcoma-associated herpesvirus (KSHV) Lana-1 TAD (7), the varicella-zoster virus (VZV) major transactivator IE62 (8), the Respiratory Syncytial Virus Nonstructural Protein 1 (RSV NS1) (9–11) and the Ets related human transcription factors PEA3, ER81 and ERM (3, 12–14). In Arabidopsis, MED25 (AtMED25, also known as PFT1 (15)) also interacts with several transcriptional regulators and is considered as an integrator of multiple signaling pathways (16). Although dozens of partners have been described for Arabidopsis MED25 (17), there are no biophysical studies available besides those with AtDREB2a (18), VP16 (19) and AtMYC3 (20).

While the structure of human MED25 ACID domain has been solved by NMR by four different groups (2, 21–23), structures of MED25 ACID domain in complex with a target protein are still unknown. The human ACID domain contains a seven-stranded β -barrel flanked with three α helices and dynamic loops (Figure 1). Specifically, the human ACID domain coordinates its interactome using two largely hydrophobic interfaces named H1 et H2 and located on opposite faces of the β -barrel, surrounded by patches of positively charged residues. As defined by NMR chemical shift perturbation (CSP) and mutagenesis studies of ACID domain in complex with VP16 and ERM TADs (2, 13, 21, 22), the H1 binding site is delineated by strands β 1- β 3- β 5, while the H2 binding site involves helix α 1 and strands β 6- β 7. These ACID surfaces appear adequately designed to accommodate the specific patterns of bulky hydrophobic and negatively charged residues found in different acidic transactivation domains (2, 3, 10, 12, 13, 19, 21, 22, 24–26) (Figure 1).

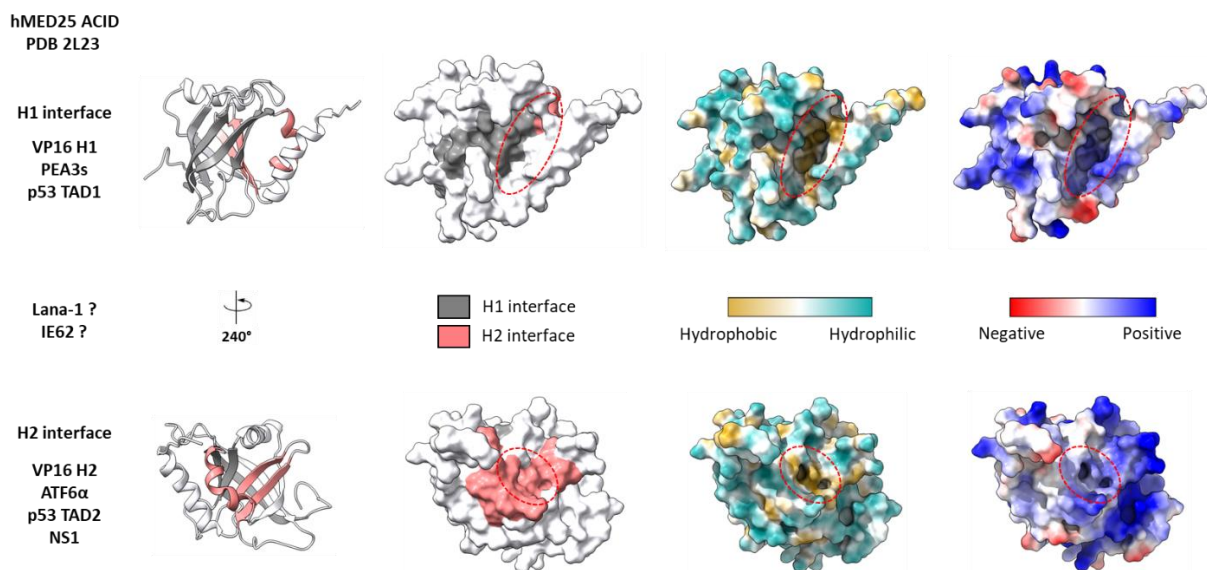


Figure 1: The human MED25 ACID domain. Cartoon and surface representation of human MED25 ACID domain (PDB 2L23 in white). The H1 binding site is formed mainly by strands β 1- β 3- β 5 as defined by NMR chemical shift perturbation (21, 22) and is shown in gray. The H2 binding site is formed mainly by helix α 1 and

strands $\beta 6$ - $\beta 7$ as defined by NMR chemical shift perturbation (2, 21, 22) and is shown in light coral. The various views of H1 and H2 are related by a 240° rotation along the y axis. PEA3s (ETV1/ER81, ETV4/PEA3 and ETV5/ERM) TADs, VP16 H1 subdomain and p53 TAD1 interact with the H1 face while VP16 H2 subdomain, ATF6 α TAD, p53 TAD2 and the $\alpha 3$ helix of the respiratory syncytial virus NS1 protein interact with the H2 face. See main text for details and Supplementary Figure 1 for a summary of human MED25 ACID domain-TADs interactions. The binding interfaces of Lana-1 and IE62 are much less characterized as indicated by the question mark. The ACID domain surface is also colored by the molecular lipophilicity potential, where cyan denotes hydrophilic residues and gold denotes hydrophobic ones, and by electrostatic potential (blue and red indicate positively and negatively charged residues, respectively) as calculated with Coulombic surface coloring in ChimeraX (27). The two hydrophobic and positively charged pockets are indicated as a red dotted circle.

Numerous studies have helped to discriminate between MED25 partners according to the interface they contact. More specifically, the VP16 TAD subdomain H1, p53 subdomain TAD1 and PEA3s TADs directly interact with the H1 face of MED25 ACID (3, 12–14, 21, 22, 24, 25) while the VP16 TAD subdomain H2, p53 subdomain TAD2, ATF6 α TAD and the C-terminal $\alpha 3$ helix of NS1 primarily bind to the MED25 H2 interface (2, 10, 21, 22, 24, 25). In contrast, the MED25 interaction surface and the minimal interaction domain for Lana-1 (7) and IE62 (8) remain largely unknown (Figure 1). A summary of the domain architecture of these MED25 interacting proteins with available experimental binding data is provided in Supplementary Figure 1.

In the absence of experimental structures of MED25-TADs complexes, computational modeling provides a valuable alternative. Machine learning (ML) approaches for *de novo* protein structure prediction such as neural networks AlphaFold2 (28) and RosettaFold (29) and large language model (LLM) OmegaFold (30) and ESMFold (31) have recently revolutionized structural biology by predicting highly accurate structures of proteins and their complexes.

In this study we report a systematic assessment of deep learning methods using the publicly available ColabFold version (32) to predict Human and Arabidopsis MED25 ACID domain-TADs interfaces and evaluate the accuracy of the models through comparison with published experimental data. We modelled 9 different human MED25 ACID-TAD complexes starting with the best-characterized experimentally (VP16, p53, ATF6 α , ERM) and ending with the lesser-known ones (Lana-1 and IE62), enabling us to precisely predict and validate their interaction interfaces. Although AlphaFold was unable to discriminate between H1 and H2 interfaces, it remains a formidable hypothesis generator. The models correctly predict the minimal MED25 ACID interacting sequences of its binding partners that fold into α -helices in the vast majority of cases. We also reveal a new interaction surface unique to plants by predicting 3 different AtMED25 complexes (AtAP2/ERF, AtDREB2a and VP16). Finally, we compare the four main ML protein folding methods (AlphaFold, RoseTTAFold, OmegaFold and ESMFold) and discuss usability and limitations of their predictions.

Materials and Methods

Structural predictions

Protein sequences used in this study were extracted from the UniProt (33) database (Supplementary file 1). The predicted structures from AlphaFold, RoseTTAFold, OmegaFold and ESMFold were obtained (accessed on May and June 2023) from the python notebook

available through the ColabFold interface (<https://github.com/sokrypton/ColabFold>) (32). An experimental notebook developed by Sergey Ovchinnikov (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/omegafold_hacks.ipynb) was also used to incorporate Multiple Sequence Alignment (MSA) input into OmegaFold structure prediction (Figure 1). This notebook only supported monomeric and homo-oligomeric predictions and was therefore only used for the isolated human and Arabidopsis MED25 ACID domain (Figure 2). We also used a local installation of AlphaFold-multimer (version 2.3.0) or accessed directly from ChimeraX (version 1.6rc202303310103 (2023-03-31)) (27).

For each prediction, five initial models were produced with default settings within multimer mode. To note, the python notebook does not use PDB templates (no template), thereby providing a totally naive structure prediction. Based on template modelling scores and predicted alignment error (PAE) values for each model, we then picked the top-ranked predicted structure of the complex as the final model for subsequent analysis. Visualization of the PAE for predicted complex structures was done with the webserver PAE Viewer (34) and figures were prepared using ChimeraX (27).

The protein-protein interactions modeled by AlphaFold-multimer distribute over the full pTM and ipTM range and multimer scores for the best ranked structures are shown in each figure and in Supplementary Figure 8. pTM, ipTM and mean pLDDT scores > 0.8 were considered highly accurate. For RoseTTAFold (29), mean interface predicted aligned error (pAE) < 5 indicated high confidence. The pLDDT (predicted local-distance difference test) is a confidence measure for the per-residue accuracy of the structure, the pTM (predicted template modeling score) is a metric for the similarity between protein structures (accuracy of prediction within each protein chain) and the ipTM (Interface pTM) scores interactions between residues of different chain to estimate the accuracy of interfaces (accuracy of the complex between chain) (28, 35, 36).

Plasmids, protein expression and purification

The pET24d vector encoding human MED25 ACID domain (391-548)-6xHis has been previously described (2). BamHI/XhoI primers corresponding to Lana-1 (280-297) were annealed and cloned into pGEX 6P1 plasmid. For IE62 1-200, we applied the synthetic gene approach to synthesize (GenBank X04370.1) a codon-optimized pGEX 6P1 expression vector from GenScript. IE62 1-86 and 87-200 were amplified by PCR (template IE62 1-200), digested by BamHI/NotI and subcloned into the pGEX 6P1 vector. The bacterial expression vectors pGEX 6P1, pGEX 6P1 Lana-1 280-297, IE62 1-200, IE62 1-86 and IE62 87-200 (ampicillin) were co-transformed with pET24d hMED25 ACID-6xHis (kanamycin) in competent *E. coli* BL21 (DE3) strain and grown at 37°C in LB medium (150 ml) to an optical density of 0.8 at 600 nm and expression was induced with 0,1 mM IPTG for 18 h at 18°C. Cells were harvested by centrifugation and store frozen at -20°C. Pellets were resuspended in 12 ml cold lysis buffer (50 mM Tris-HCl pH 8,0, 100 mM NaCl, 10% glycerol and 0.1% NP-40 supplemented with 0.1 mg of lysozyme/ml and complete inhibitor cocktail (Roche)) and sonicated to disrupt the cells. The lysates were cleared by centrifugation at 10,000 × g at 4°C for 45 min to collect soluble proteins. The lysates were incubated with glutathione Sepharose 4B beads or TALON® Superflow™ resin for 2h at 4°C and the beads were pelleted and washed three times with 10

volumes of lysis buffer. Bound proteins were analysed by mPAGE™ 4-20% Bis-Tris gels (Merck) and Coomassie staining. Gels were scanned with Amersham ImageQuant 800.

Results

Benchmarking machine learning models for human and Arabidopsis MED25 ACID domain structure

We first assessed the accuracy of ML-based prediction algorithms AlphaFold (28), RoseTTAFold (29), OmegaFold (30) and ESMFold (31) to correctly predict the structure of hMED25 ACID domain alone. Unlike AlphaFold and RoseTTAFold, OmegaFold and ESMFold use large protein language model as a backbone and do not need a Multiple Sequence Alignment (MSA) step to generate prediction. The ColabFold version (32) was used with default settings (no template) to predict a set of five structural models of hMED25 ACID domain and the models with the highest overall pLDDT scores were picked (Figure 2A). Predictions showed an excellent agreement (with an average pLDDT score > 0.8 and a RMSD below 1Å) with the NMR structures (PDB 2L23, 2L6U, 2XNF and 2KY6) (2, 21–23) with the exception of OmegaFold, which only becomes accurate by incorporating MSA as additional input (Figure 2A).

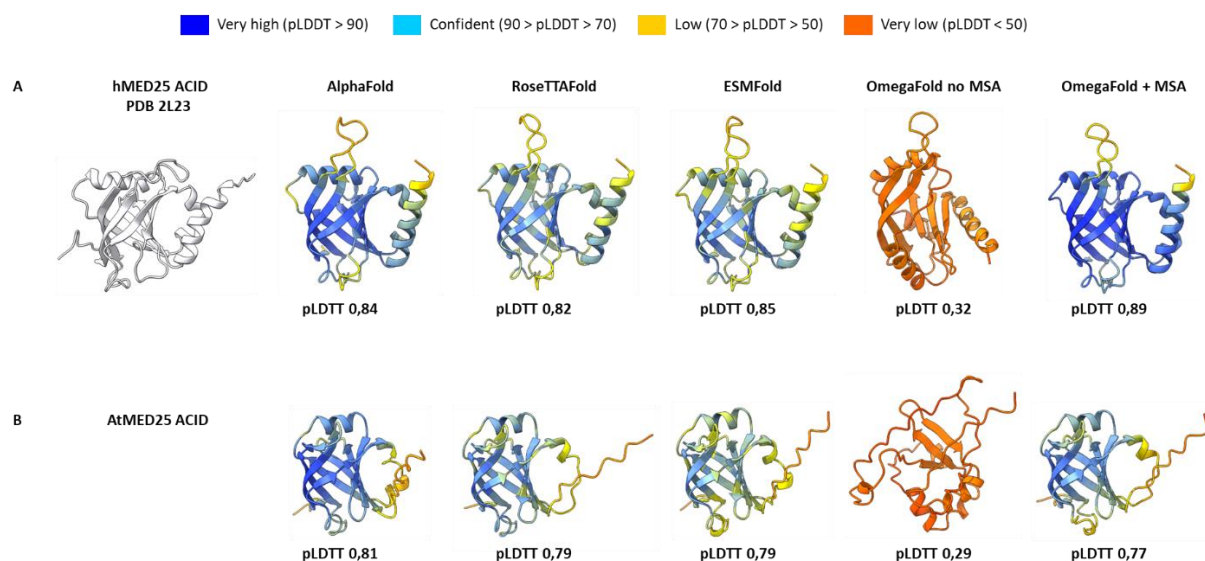


Figure 2: ML-based structural prediction of Human and Arabidopsis MED25 ACID domain. (A) Cartoon representation of human (PDB 2L23) and (B) Arabidopsis thaliana MED25 ACID domain and comparison of the initial structures obtained with machine learning protein folding methods AlphaFold, RoseTTAFold, ESMFold and OmegaFold. Each residue in human (hACID) and Arabidopsis (AtACID) MED25 ACID domain is color-coded based on the model confidence score, pLDDT.

The structure of AtMED25 ACID domain has not been experimentally determined and despite a low sequence identity with the human ACID domain, structure prediction indicates that the β -barrel architecture is topologically conserved (Figure 2B). AtMED25 ACID predicted β -strands β 1- β 4- β 5- β 7 and α helix H2 overlap quite well with their human counterpart with an

RMSD value of 1.12 Å (Supplementary Figure 2). The major difference being a missing H3 helix in Arabidopsis in line with previous results (18). As with human ACID, OmegaFold prediction of AtACID appears only accurate with an additional MSA input (Figure 2B). Thus, ML-based structural prediction of MED25 ACID is in good agreement with NMR structures, indicating that the generated models are suitable for further analysis.

Case study of human MED25 ACID domain structure prediction in complex with well-characterized interacting protein partners

We next tested if AlphaFold can predict the structure of human MED25 ACID domain in complex with partner proteins that were well characterized biochemically, but for which the structure of the complexes has not been experimentally resolved to date. We focused on the herpes virus activator VP16 (21, 22), the three Ets transcription factors ERM, PEA3 and ER81, (3, 12–14), p53 (25) and ATF6α (24, 37). These six MED25 partners represent an excellent starting point for our study since they are distributed across the entire spectrum of all possible interactions: bivalent MED25 interaction (H1 and H2 interfaces) for VP16 H1H2 and p53 TAD1 TAD2, MED25 H1 interface interaction for VP16 H1 subdomain, PEA3s TADs and p53 TAD1 and MED25 H2 interface interaction for VP16 H2 subdomain, p53 TAD2 and ATF6α TAD (Figure 1 and Supplementary Figure 1). In all cases, AlphaFold predicted an interface which is mediated by alpha helices formed by the acidic TADs, although with varying accuracy (Figure 3).

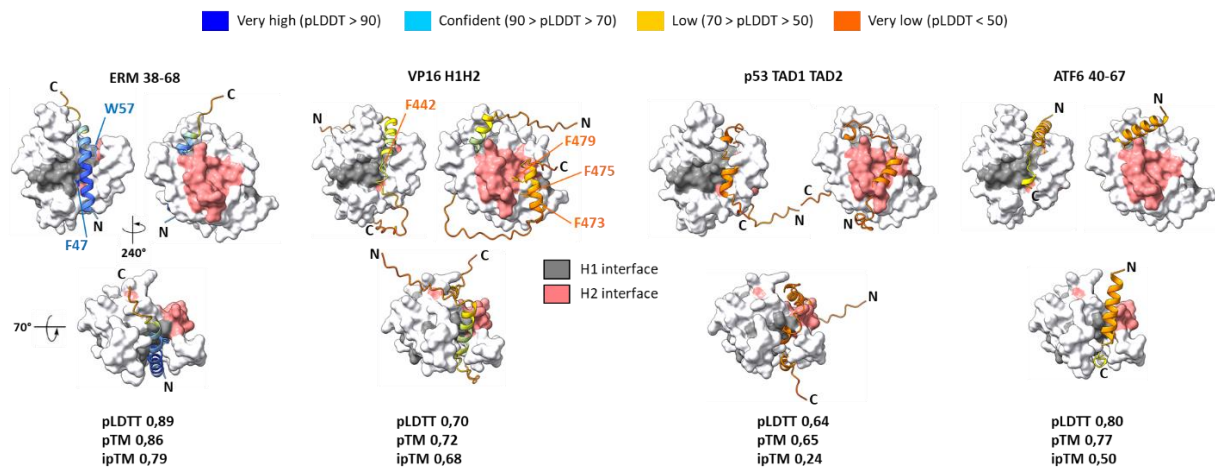


Figure 3: AlphaFold structural predictions of human MED25 ACID domain in complex with ERM, VP16, ATF6α and p53 transactivation domains. Surface representation of human MED25 ACID domain (white) with interfaces H1 (gray) and H2 (light coral) in complex with ERM, VP16, ATF6α and p53 transactivation domains (cartoon representation color-coded based on the model confidence score, pLDDT). N-terminal (N) and C-terminal (C) regions of the transactivation domains are indicated. Cartoon representation related by 240° rotation along the y axis and related by 70° rotation along the x axis. For each prediction, the pLDDT (predicted local-distance difference test), pTM (predicted template modeling score) and ipTM (Interface pTM) scores are indicated.

PEA3s. ERM (38-68) TAD has the highest ipTM (0,79) and pTM (0,86) scores supported by high pLDDT (0,89), indicating very high confidence in both the interface and in the prediction of the overall structure. The ERM TAD helical structure predicted by AlphaFold fitted into the

MED25 H1 interface with residues Y47 (pLDDT 0,95) and W57 (pLDDT 0,90) that pointed toward the concave hydrophobic cavity (Figure 3), in agreement with experimental data (12, 13) (Supplementary Figure 1). Highly accurate predictions were also obtained with ER81/ETV1 (38-69) and PEA3/ETV4 (45-76) TADs (Supplementary Figure 3A and 3B). The sequence variations across the PEA3 subfamily TADs resulted in a slightly shorter helix for PEA3 which could explain the unique PEA3 MED25 engagement mode as compared to ERM and ER81 (3). In addition to the PEA3s TADs, MED25 ACID domain was also reported to interact with the α -helix H4 that is specific to ETV1/ETV4/ETV5 DNA binding domain (DBD) (14). Three distinct sites on MED25 bind the ETS DBD, two of which correspond roughly to the H1 and H2 interfaces (site 1 and site 2 in (14)). The MED25 ACID-PEA3 DBD (337-436) model has a low pTM score (0,57) and a very low iTM score (0,18) (Supplementary Figure 3C), indicating poor confidence in the interface that roughly corresponds to the third site (site 3) identified in (14). Nevertheless, using full length PEA3 (1-484) as template (Supplementary Figure 4), AlphaFold correctly predicts two distinct binding surfaces for PEA3 on MED25 ACID.

VP16. AlphaFold successfully predicted the extended binding interface of VP16 with the H1 and H2 subdomains wrapping around the β -barrel and adopting a partially folded conformation when bound (Figure 3). Three separate regions with α -helical propensity are apparent spanning residues 428 to 441, 444 to 447 (VP16 H1 subdomain) and 469 to 482 (VP16 H2 subdomain) which included F442, F473, F475 and F479 that have been shown to be critical for VP16 transcriptional activity and MED25 recruitment (5, 21, 22, 38).

p53. In contrast to VP16, AlphaFold did not accurately predict the relative orientation of the MED25 ACID/p53 TAD1 TAD2 complex (Figure 3 and Supplementary Figure 1). The model correctly predicts two distinct binding surfaces but with the binding site for p53 TAD1 on the MED25 H2 interface and the binding site for p53 TAD2 on the H1 interface. The confidence scores are low (pLDDT 0,64, pTM 0,65, ipTM 0,24), suggesting that the MED25/p53 TAD model is unreliable.

ATF6 α . We next modelled MED25/ATF6 α TAD 40-67 complex that was reported as an H2 binding site specific partner (24) (Supplementary Figure 1). AlphaFold prediction suggested a preferential occupation of MED25 ACID by both a very short α -helix in the C-terminus of ATF6 α (64-LDL-66) in the H1 interface and a second predicted long N-terminal α -helix (41-TDELQLEAANETYENNFDN-59) arranged in a near perpendicular fashion with the H1 α -helix of the ACID β -barrel (Figure 3). However, the global quality metrics are good but not very high for the ipTM score (pLDDT 0,8, pTM 0,77, ipTM 0,50), indicating limited reliability of the predicted binding interface. We also explored how predictions would be impacted by queries in which the putative bound ATF6 α motif is embedded in larger fragments. Using ATF6 α 1-80 (Supplementary Figure 5), 1-150 and 38-75 (data not shown) as templates (24, 37), AlphaFold predictions that were obtained are quite different from the initial ATF6 α 40-67 prediction (Figure 3). In all cases, AlphaFold was unable to accurately predict the binding mode and to discriminate between the H1 and H2 ACID interfaces with many different predicted ATF6 α short α -helixes fragments embedded in long disordered regions.

p53 TAD1, TAD2 and VP16 H1 and H2 subdomains. By comparing p53 TAD1, p53 TAD2, VP16 H1 and VP16 H2 subTADs boundaries, we also found variation in predictive performance among models (Figure 3 and Supplementary Figure 6). In all cases, higher

accuracy models were generated but predicted interaction interfaces do not always overlap. While the VP16 H1 subdomain model is almost indistinguishable from the one with VP16 H1H2, the VP16 H2 subdomain prediction gives now 2 different solutions with very similar confidence scores, either VP16 H2 in the ACID H1 interface or in the H2 interface. For p53 TAD1 subdomain, AlphaFold now predicts the correct H1 interface but suggests the same solution for p53 TAD2 (Figure 3 and Supplementary Figure 6).

Comparison of machine-learning models. Likewise, comparing AlphaFold, RoseTTAFold and ESMFold, we found that ERM, VP16, p53 and ATF6 α TADs models do not always converge (Figure 3, Supplementary Figures 5 and 7). For example, none of the tested ML protocols converged to an identical model for ATF6 α (40-67) and ESMFold fitted incorrectly the ERM TAD into the H2 interface.

Case study of respiratory syncytial virus NS1 protein

We next assessed the accuracy of AlphaFold, RoseTTAFold, and ESMFold at correctly predicting the structure of hMED25 ACID domain in complex with the respiratory syncytial virus NS1 protein. MED25 was recently identified as an NS1 interacting protein (9, 11) and NMR experiments indicate that the NS1 C-terminal α 3 helix contacts directly the ACID H2 interface (10) (Supplementary Figure 1). RoseTTAFold and ESMFold predictions converged to an identical model for NS1 (118-139) bound to the H2 ACID domain interface that matches experimental data while AlphaFold fitted the NS1 C-terminal helix into the H1 interface (Figure 4). However, AlphaFold prediction has the highest pTM (0,80) score supported by high confidence pLDDT (0,82). This is intriguing in light of the NMR data, which showed that a peptide consisting of the sequence of the NS1 α 3 helix primarily bound to MED25 ACID H2 with 17 μ M affinity. However, the H1 interface was also detected as a secondary binding site with an apparent affinity of 500 μ M (10).

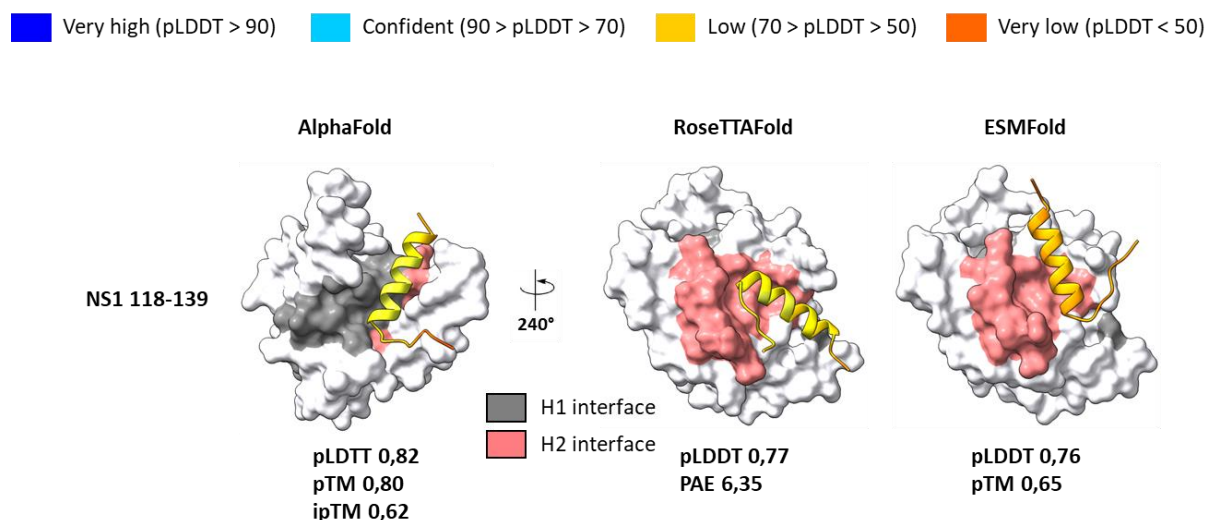


Figure 4: ML-based structural prediction of human MED25 ACID domain in complex with respiratory syncytial virus NS1 C-terminal α 3 helix. Surface representation of human MED25 ACID domain (white) with the two opposite binding surfaces H1 (gray) and H2 (light coral) in complex with NS1 C-terminal helix (118-139) (cartoon representation color-coded based on the model confidence score, pLDDT). For each prediction, the

pLDDT (predicted local-distance difference test), pTM (predicted template modeling score), PAE (predicted alignment error matrix) and ipTM (Interface pTM) scores are indicated.

Case study of human MED25 ACID domain structure prediction in complex with poorly characterized interacting protein partners

We next used AlphaFold to map precisely the binding region of KSHV Lana-1 and VZV IE62 to MED25. Previous reports have shown an interaction between the N-terminal region of Lana-1 (1-340) (7) and the N-terminal region of IE62 (1-226 and 1-86) (8, 39) and MED25 ACID domain but the minimal domains of interaction have not yet been identified (Supplementary Figure 1).

Lana-1. The Lana-1 segment (1-340) is predicted to be almost entirely disordered by AlphaFold except for a long α -helix at the C-terminus (residues 279-306) fitted in the ACID H2 interface but the confidence scores are very low (pLDDT 0,43, pTM 0,39, ipTM 0,36) (Figure 5A). Delineating the input sequence into a shorter fragment (Lana-1 279-308) clearly increases the confidence metrics (pLDDT 0,82, pTM 0,79, ipTM 0,63) (Figure 5B). However, the predicted interaction interfaces are opposite from each other between Lana-1 (1-340) and Lana-1 (279-308), making the identification of the most plausible interface very challenging. Looking more closely, we noted that the ACID H1 interface is in fact occupied by Lana-1 (1-340) disordered region centered on residues Y235 and W238, probably rendering it inaccessible to the C-terminal predicted α -helix (Figure 5A). To experimentally verify the AlphaFold predictions and establish the contribution of this Lana-1 (279-308) putative interface, we co-expressed GST and GST Lana-1 (280-297) with human MED25 ACID domain in bacteria. Bacterial lysates were incubated with glutathione or cobalt beads, washed extensively and bound complexes were analysed by SDS-PAGE and Coomassie blue staining. As shown in Figure 5C, MED25 ACID was effectively and specifically pulled down by GST Lana-1 (280-297), thus validating the AlphaFold model.

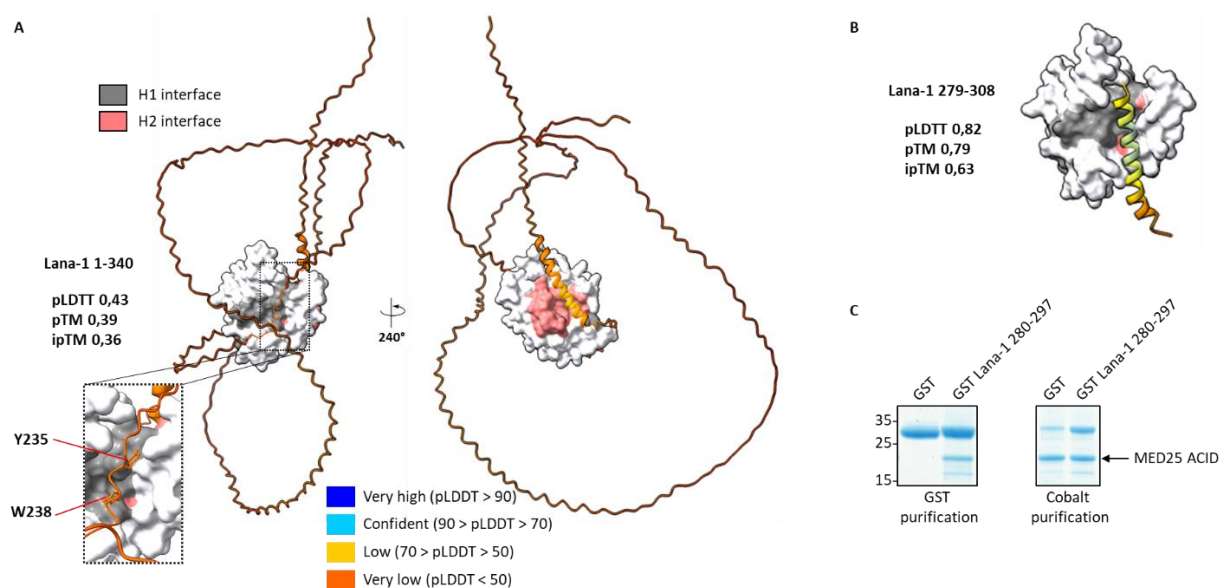


Figure 5: AlphaFold structural prediction of human MED25 ACID domain in complex with the Kaposi's sarcoma-associated herpesvirus (KSHV) Lana-1 protein. (A) Surface representation of human MED25 ACID

domain (white) with the two opposite binding surfaces H1 (gray) and H2 (light coral) in complex with (A) Lana-1 (1-340) and (B) Lana-1 (279-308) (cartoon representation color-coded based on the model confidence score, pLDDT). (A) Inset, close-up view of Lana-1 (1-340)-ACID H1 interface with Lana-1 residues Y235 and W238 highlighted. Each structure in H1 and H2 view related by 240° rotation along the y axis. For each prediction, the pLDDT (predicted local-distance difference test), pTM (predicted template modeling score) and ipTM (Interface pTM) scores are indicated. (C) Copurification of MED25 ACID with GST Lana-1 (280-297). GST Lana-1 (280-297) or GST were co-expressed with MED25 ACID-6xHis in bacteria and purified with glutathione Sepharose 4B beads (left) or cobalt resin (right). Fractions were analysed by SDS-PAGE and coomassie blue staining. As indicated with cobalt resin, MED25 ACID domain is expressed to similar level. Molecular weight marker (in kDa) is indicated.

IE62. A similar approach was used with the varicella-zoster virus (VZV) major transactivator IE62 (Figure 6). We first used IE62 (1-200) as template (Figure 6A) which roughly corresponds to the (1-226) domain initially described as interacting with MED25 (39). AlphaFold predicted an extended binding interface reminiscent of VP16 H1H2 and p53 TAD1 TAD2 with two partially folded α -helix wrapping around MED25 ACID β -barrel, while the remaining part of IE62 (1-200) appeared mostly unstructured (Figure 6A). The predicted IE62 N-terminal α -helix (residues 17-43) fitted in the ACID H1 interface (Figure 6A) is in good agreement with previous experimental data that identified a potential extended α -helix involving residues 19-35 of IE62 TAD (8). The second predicted IE62 α -helix (residues 106-117) bound to the H2 interface (Figure 6A) has never been considered. The confidence scores are low (pLDDT 0,53, pTM 0,53, ipTM 0,68) but given the fact that disordered regions lack proper structure, the model quality of those regions will be always low, hence lowering the overall model quality (40). Using IE62 1-86, 99-200 (Figure 6A), 17-41 and 106-120 (Figure 6B) as templates, AlphaFold predictions that were obtained are quite different from the initial IE62 1-200 prediction. In all cases, AlphaFold predicted an H2 interface binding mode. To establish the relative contribution of these IE62 putative interfaces, we co-expressed GST, GST IE62 (1-200), (1-86) and (87-200) with human MED25 ACID domain in bacteria. Bacterial lysates were incubated with glutathione or cobalt beads, washed extensively and bound complexes were analysed by SDS-PAGE and Coomassie blue staining. As shown in Figure 6C, MED25 ACID was effectively and specifically pulled down only by GST IE62 (1-200) and (1-86), indicating that the IE62 99-200 and IE62 106-120 MED25 ACID models are incorrectly classified as interacting pairs by AlphaFold.

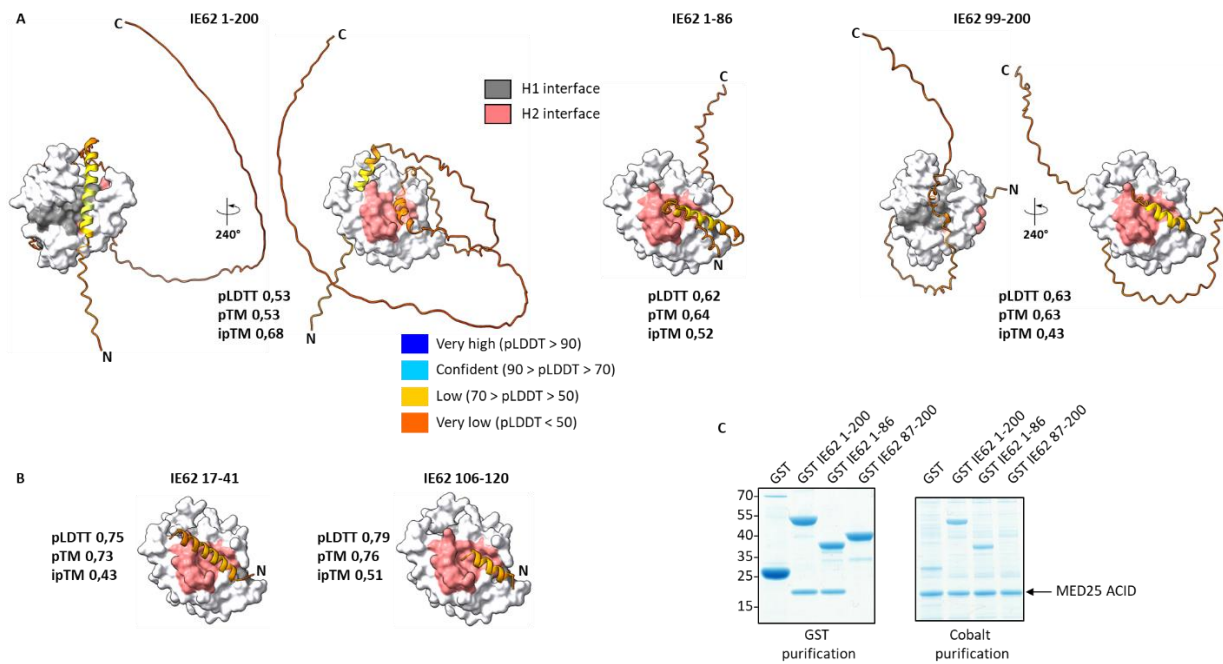


Figure 6: AlphaFold structural prediction of human MED25 ACID domain in complex with the varicella-zoster virus (VZV) major transactivator IE62. Surface representation of human MED25 ACID domain (white) with the two opposite binding surfaces H1 (gray) and H2 (light coral) in complex with various fragments of IE62 (cartoon representation color-coded based on the model confidence score, pLDDT). **(A)** IE62 (1-200), (1-86) and (99-200) predictions. **(B)** IE62 (17-41) and (106-120) predictions. N-terminal (N) and C-terminal (C) regions of the transactivation domains are indicated. Each structure in H1 and H2 view related by 240° rotation along the y axis. For each prediction, the pLDDT (predicted local-distance difference test), pTM (predicted template modeling score) and ipTM (Interface pTM) scores are indicated. **(C)** IE62-MED25 ACID interaction by GST pull-down assay. GST IE62 (1-200), (1-86) and (87-200) or GST were co-expressed with MED25 ACID-6xHis in bacteria and purified with glutathione Sepharose 4B beads (left) or cobalt resin (right). Fractions were analysed by SDS-PAGE and coomassie blue staining. As indicated with cobalt resin, MED25 ACID domain is expressed to similar level. Molecular weight marker (in kDa) is indicated.

Case study of Arabidopsis MED25 ACID domain structure prediction in complex with interacting protein partners

We finally tested if AlphaFold can predict the structure of Arabidopsis MED25 ACID domain in complex with selected interacting proteins. We focused on the Ethylene Response factors (AP2/ERF) family (16), the transcription factor AtDREB2a (18, 41) and the herpes virus activator VP16 (19).

In all cases, AlphaFold predicted an interface which is mediated predominantly by α helices, although with varying accuracy (Figure 7A). AtERF98 (117-139) TAD has the highest ipTM (0,82) and pTM (0,86) scores supported by high pLDDT (0,88), indicating very high confidence in both the interface and in the prediction of the overall structure. A C-terminal AtERF98 helical structure predicted by AlphaFold (DKVLEELLDDSEERK) occupied a hydrophobic pocket defined by the AtMED25 ACID H2 α -helix and strands β 1- β 2 loop located at the top of the β -barrel core (Figures 7A and 7C and Supplementary Figure 2). This interface differs greatly from those determined for human MED25, likely due to the fact that there is no predicted H3 α -helix in AtMED25 and therefore no equivalent of the human H1 interface (Figure 1 and Supplementary Figure 2). The AlphaFold model suggested a second potential interface

(Supplementary Figure 9) between AtACID and AtERF98 with a short N-terminal AtERF98 β -strand (FEFEY) that associates with an extra AtACID β -strand ($\beta 8$ VVFKP) not predicted in AtMED25 ACID structure alone (Figure 1 and Supplementary Figure 2), reminiscent of a two-stranded β zipper (42). Nevertheless, the predicted AtMED25 ACID/AtERF98 complex matched experimental data closely with the conserved EDLL motif previously implicated in transcriptional activation and interaction with AtMED25 (16, 43) adopting a partially folded conformation when bound (Figure 7A and Supplementary Figure 9). In contrast, confidence scores are very low for VP16 H1H2 TAD, VP16 H1 and H2 subdomains (Figure 7A), making the identification of the most plausible interface very challenging.

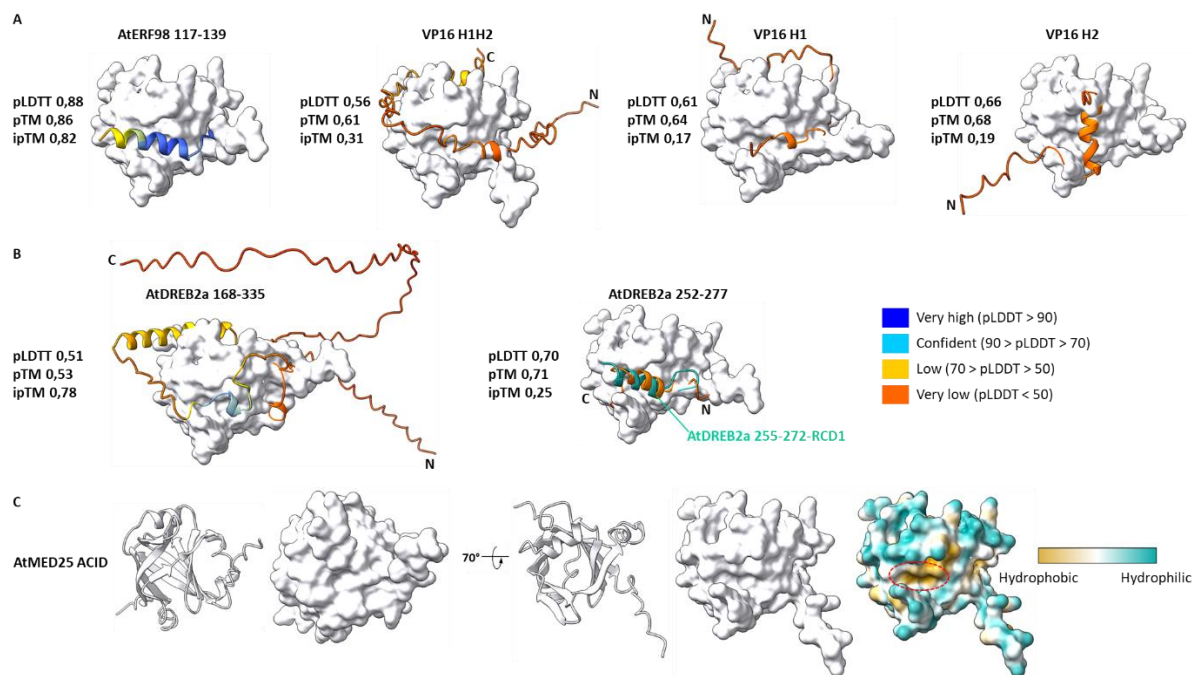


Figure 7: AlphaFold structural prediction of *Arabidopsis thaliana* MED25 ACID domain in complex with selected partners. (A) Surface representation of AtMED25 ACID domain (white) in top view in complex with AtERF98 (117-139), VP16 H1H2, VP16 H1 and VP16 H2 and (B) with AtDREB2a (168-335) and AtDREB2a (252-277) (cartoon representation color-coded based on the model confidence score, pLDDT). N-terminal (N) and C-terminal (C) regions of the transactivation domains are indicated. For each prediction, the pLDDT (predicted local-distance difference test), pTM (predicted template modeling score) and ipTM (Interface pTM) scores are indicated. The AtMED25 ACID-AtDREB2a (252-277) AlphaFold model was superimposed with the HADDOCK model of the RCD1-RST-DREB2A (255-272) complex. RCD1-RST was omitted for simplicity and AtDREB2a (255-272) α -helix is colored in light sea green. (C) Cartoon and surface representation of AtMED25 ACID domain as in Figure 1. Each structure in front and top view related by 70° rotation along the x axis. AtACID domain surface is also colored by the molecular lipophilicity potential, where cyan denotes hydrophilic residues and gold denotes hydrophobic ones with ChimeraX (27). The top hydrophobic pocket is indicated as a red dotted circle.

We next modelled AtMED25/AtDREB2a (168-335) complex (Figure 7B) that was reported to interact with a K_d in micromolar range by SPR (18) and ITC (19). This C-terminal AtDREB2a fragment comprised both the BD (168-253) and AD (254-335) domains but only the extended (168-335) fragment interacted with AtMED25 ACID with significant affinity (18, 19). AlphaFold was unable to accurately predict the binding mode with many different predicted AtDREB2a (168-335) α -helices embedded in long disordered regions bound to multiple AtMED25 ACID surfaces (Figure 7B). Interestingly, VP16 and AtDREB2a that both bind

AtMED25, display very low sequence identity apart from a short positively charged stretch of amino acids in DREB2a (252-277 HLDSSDMFDVDELLRDLNGDDVFAGL) which lies exactly at the junction between the (168-253) and (254-335) domains (18, 19). Delineating the input sequence into this shorter 252-277 fragment clearly increases pLDTT and pTM confidence metrics (Figure 7B). Importantly, the 3 different complexes (ERF98, DREB2a and VP16) predicted here occupied the same AtMED25 hydrophobic pocket (Figure 7). Interestingly, AtDREB2a has also been reported to interact with the plant Radical-Induced Cell Death1 (RCD1) protein through the same C-terminal region that partially adopt an α -helical conformation upon association with RCD1 (44, 45). When our AlphaFold AtMED25-DREB2a (252-277) model was superimposed with the HADDOCK model of AtRCD1 in complex with DREB2a (255-272) (44), the two AtDREB2a α -helices aligns well (Figure 7B). Taken together, our results predicted a unique hydrophobic pocket located at the top edge of the AtMED25 β -barrel core, not conserved with hMED25 ACID, that is targeted by AtAP2/ERF family, AtDREB2a and VP16 (Figure 7C). The observation that AtDREB2a and VP16 interacted with overlapping regions of AtMED25 ACID (19) and that the MED25 ACID- and RCD1-binding regions of DREB2a partly embraced (44, 45) supported our AlphaFold modeled structures.

Discussion

In this paper, we compared machine-learning algorithms for the structure prediction of several acidic transactivation domains in complex with Mediator complex subunit MED25. This case study is particularly well suited for testing the current limits of artificial intelligence, as it combines a number of challenges: (i) The MED25 ACID domain is structurally unique among activator binding domains (21) with its seven-stranded β barrel core flanked by 3 α -helices, and consequently might be underrepresented in the training set of AlphaFold, (ii) TADs are intrinsically disordered, poorly conserved and fold into multiple orientation and conformation when bound (46–49), (iii) TADs acquired structure might vary with context such as the nature of the binding partner (50) and (iv) AlphaFold models have limited conformational variance for now. To address these limitations, different approaches are currently under development to increase the structural heterogeneity of the predicted ensemble (51).

An overview of the MED25 ACID/TADs structural models produced here shows that the large majority of them form between a defined putative short α -helix residing in the intrinsically disorder TAD and the structured MED25 ACID β core domain. However, all the protein–protein interactions modeled by AlphaFold distribute over the full pTM and ipTM range, with only a subpopulation of highly confident predictions with ipTM and pTM > 0.80 (Supplementary Figure 8). This suggests that some complexes can't be correctly predicted regardless of the input conditions, most notably all the partners who are supposed to interact with the MED25 H2 interface. Several hypotheses can explain the low confidence scores associated with the folded ML model segments. First, the observed differences in modeling performance can be due to the intrinsic characteristics of each AI model in particular with the language model ESMFold (31) and OmegaFold (30) that do not require multiple sequence alignment and templates. Second, structure predictions may be unreliable due to the lack of representation in the training set or an insufficient amount of homologous sequences to validate the predicted contacts. In particular, it has been shown recently (52) that the percentage of the predicted disordered region in the sequences of unassigned domains had an inverse correlation coefficient with model quality for AlphaFold and RoseTTAFold models, indicating that model

quality are lower because the sequences might be disordered. Third, as previously observed with the yeast acidic transcription activator Gcn4-MED15 complex (46, 53, 54) and human MED25-PEA3s (3, 13, 14, 26), TADs likely bind MED25 ACID in multiple conformations and orientations that prevents accurate prediction of their correct binding mode. It is also possible that AlphaFold training procedure may have been biased by overrepresented experimental structures in the PDB. On closer inspection, we noticed an almost systematic presence of an aromatic residue in the MED25 interacting partners that may constrain the machine learning algorithms to propose a ‘preferred’ solution in interaction with the H1 interface (Supplementary Figure 10). Superimposition of the MED25 ACID/TADs complexes shows that in the majority of cases, tyrosine or tryptophan residues have their side chains pointing in the same direction with ERM (38-68) F47 residue (Supplementary Figure 10). Nevertheless, the signature of TAD folding is there and thus, given that some proteins such as Lana-1, IE62 and the majority of Arabidopsis MED25 protein partners remain largely uncharacterized, AlphaFold predictions constitute interesting working hypotheses. As prediction of complexes is, on average, less accurate than that of folds of individual proteins, it is particularly important to always verify interface predictions experimentally as we have done here with IE62 and Lana-1. Our results also underscore the interest of fragment-based searching to identify interaction motifs when one of the protein partners is predicted highly disordered. In case where the binding region is unknown, delineating the interaction region into fragments of decreasing size may increase the overall success rate in agreement with a recent analysis (55).

The observation that VP16 H1 subdomain, PEA3s TADs and NS1 C-terminal α 3-helix bind the same MED25 ACID H1 interface despite low sequence identity suggests that conformational plasticity within the 7-strands β -barrel core domain could play a role in mediating partner recognition. Transient kinetics experiments and molecular dynamics simulations clearly support that MED25 ACID-TAD complexes are highly dynamic (24). In particular, strands β 1- β 2 loop (residues 409-424) and α -helix H3 (residues 529-543) flanking the H1 binding interface are the most dynamic region of the ACID domain and show the greatest stabilization upon binding (24). AlphaFold predictions of ERM TAD (Figure 3) and NS1 C-terminal α 3-helix (Figure 4) in complex with MED25 ACID support the structural model of binding seen from molecular dynamics. Superimposition of VP16 H2-, ERM (38-68)- and NS1 (118-139)-bound MED25 models suggest opening of the H1 hydrophobic groove where ACID α -helix H3 shifts away to accommodate the different relative positions of the TAD α -helical conformation (Supplementary Figure 11). Our results also support the recently proposed acidic exposure model for TAD function in which acidic residues and intrinsic disorder keep hydrophobic motifs exposed to solvent where they are available to bind coactivators (56–59).

In this work, we also predicted the structure of Arabidopsis MED25 ACID domain in complex with three different interacting proteins. The predicted shared binding interface located at the top of the β -barrel core, led us to hypothesize that hMED25 and AtMED25 ACID are structurally conserved (Figure 1) but that AtMED25-TADs interactions are specific to plants (Figure 7). During the preparation of our manuscript, an independent biophysical analysis of AtMED25 ACID domain in complex with AtDREB2a became available (60). Our results with AtDREB2a (252-277) are in good agreement with the NMR, ITC, MD and AlphaFold experiments performed in this study with a longer AtDREB2a fragment (234-276). In particular, the authors identified two AtMED25 ACID binding motifs in AtDREB2a (234-276) (60), one of which corresponds to the one we have just highlighted in this study. Additional AlphaFold Arabidopsis MED25 ACID domain/TADs models to guide experimental studies

will help elucidate how sequences divergence between orthologs may play a role in dictating molecular recognition.

Acknowledgements

We thank Christina Sizun for critically reading the manuscript and Sergey Ovchinnikov for the OmegaFold_hacks experimental notebook.

References

1. Soutourina, J. (2018) Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol.* **19**, 262–274
2. Bontems, F., Verger, A., Dewitte, F., Lens, Z., Baert, J.-L., Ferreira, E., de Launoit, Y., Sizun, C., Guittet, E., Villeret, V., and Monté, D. (2011) NMR structure of the human Mediator MED25 ACID domain. *J Struct Biol.* **174**, 245–251
3. Henley, M. J., Linhares, B. M., Morgan, B. S., Cierpicki, T., Fierke, C. A., and Mapp, A. K. (2020) Unexpected specificity within dynamic transcriptional protein-protein complexes. *Proc Natl Acad Sci U S A.* **117**, 27346–27353
4. Lee, H.-K., Park, U.-H., Kim, E.-J., and Um, S.-J. (2007) MED25 is distinct from TRAP220/MED1 in cooperating with CBP for retinoid receptor activation. *EMBO J.* **26**, 3545–3557
5. Mittler, G., Stühler, T., Santolin, L., Uhlmann, T., Kremmer, E., Lottspeich, F., Berti, L., and Meisterernst, M. (2003) A novel docking site on Mediator is critical for activation by VP16 in mammalian cells. *EMBO J.* **22**, 6494–6504
6. Yang, F., DeBeaumont, R., Zhou, S., and Näär, A. M. (2004) The activator-recruited cofactor/Mediator coactivator subunit ARC92 is a functionally important target of the VP16 transcriptional activator. *Proc Natl Acad Sci U S A.* **101**, 2339–2344
7. Roupelieva, M., Griffiths, S. J., Kremmer, E., Meisterernst, M., Viejo-Borbolla, A., Schulz, T., and Haas, J. (2010) Kaposi's sarcoma-associated herpesvirus Lana-1 is a major activator of the serum response element and mitogen-activated protein kinase pathways via interactions with the Mediator complex. *J Gen Virol.* **91**, 1138–1149
8. Yamamoto, S., Eletsky, A., Szyperski, T., Hay, J., and Ruyechan, W. T. (2009) Analysis of the varicella-zoster virus IE62 N-terminal acidic transactivating domain and its interaction with the human mediator complex. *J Virol.* **83**, 6300–6305
9. Pei, J., Beri, N. R., Zou, A. J., Hubel, P., Dorando, H. K., Bergant, V., Andrews, R. D., Pan, J., Andrews, J. M., Sheehan, K. C. F., Pichlmair, A., Amarasinghe, G. K., Brody, S. L., Payton, J. E., and Leung, D. W. (2021) Nuclear-localized human respiratory syncytial virus NS1 protein modulates host gene transcription. *Cell Rep.* **37**, 109803
10. Dong, J., Basse, V., Bierre, M., Peres de Oliveira, A., Vidalain, P.-O., Sibille, P., Tangy, F., Galloux, M., Eleouet, J.-F., Sizun, C., and Bajorek, M. (2022) Respiratory Syncytial Virus NS1 Protein Targets the Transactivator Binding Domain of MED25. *Journal of Molecular Biology.* **434**, 167763
11. Van Royen, T., Sedeyn, K., Moschonas, G. D., Toussaint, W., Vuylsteke, M., Van Haver, D., Impens, F., Eyckerman, S., Lemmens, I., Tavernier, J., Schepens, B., and Saelens, X. (2022) An Unexpected Encounter: Respiratory Syncytial Virus Nonstructural Protein 1 Interacts with Mediator Subunit MED25. *J Virol.* **96**, e01297-22
12. Verger, A., Baert, J.-L., Verreman, K., Dewitte, F., Ferreira, E., Lens, Z., de Launoit, Y., Villeret, V., and Monté, D. (2013) The Mediator complex subunit MED25 is targeted by the N-terminal transactivation domain of the PEA3 group members. *Nucleic Acids Res.* **41**, 4847–4859
13. Landrieu, I., Verger, A., Baert, J.-L., Rucktooa, P., Cantrelle, F.-X., Dewitte, F., Ferreira, E., Lens, Z., Villeret, V., and Monté, D. (2015) Characterization of ERM transactivation domain binding to the ACID/PTOV domain of the Mediator subunit MED25. *Nucleic Acids Res.* **43**, 7110–7121
14. Currie, S. L., Doane, J. J., Evans, K. S., Bhachech, N., Madison, B. J., Lau, D. K. W., McIntosh, L. P., Skalicky, J. J., Clark, K. A., and Graves, B. J. (2017) ETV4 and AP1 Transcription Factors Form Multivalent Interactions with three Sites on the MED25 Activator-Interacting Domain. *J Mol Biol.* **429**, 2975–2995
15. Cerdán, P. D., and Chory, J. (2003) Regulation of flowering time by light quality. *Nature.* **423**, 881–885

16. Çevik, V., Kidd, B. N., Zhang, P., Hill, C., Kiddle, S., Denby, K. J., Holub, E. B., Cahill, D. M., Manners, J. M., Schenk, P. M., Beynon, J., and Kazan, K. (2012) MEDIATOR25 acts as an integrative hub for the regulation of jasmonate-responsive gene expression in Arabidopsis. *Plant Physiol.* **160**, 541–555
17. Zhai, Q., Deng, L., and Li, C. (2020) Mediator subunit MED25: at the nexus of jasmonate signaling. *Curr Opin Plant Biol.* **57**, 78–86
18. Blomberg, J., Aguilar, X., Brännström, K., Rautio, L., Olofsson, A., Wittung-Stafshede, P., and Björklund, S. (2012) Interactions between DNA, transcriptional regulator Dreb2a and the Med25 mediator subunit from Arabidopsis thaliana involve conformational changes. *Nucleic Acids Res.* **40**, 5938–5950
19. Aguilar, X., Blomberg, J., Brännström, K., Olofsson, A., Schleucher, J., and Björklund, S. (2014) Interaction studies of the human and Arabidopsis thaliana Med25-ACID proteins with the herpes simplex virus VP16- and plant-specific Dreb2a transcription factors. *PLoS One.* **9**, e98575
20. Takaoka, Y., Suzuki, K., Nozawa, A., Takahashi, H., Sawasaki, T., and Ueda, M. (2022) Protein–protein interactions between jasmonate-related master regulator MYC and transcriptional mediator MED25 depend on a short binding domain. *Journal of Biological Chemistry.* **298**, 101504
21. Vojnic, E., Mourão, A., Seizl, M., Simon, B., Wenzek, L., Larivière, L., Baumli, S., Baumgart, K., Meisterernst, M., Sattler, M., and Cramer, P. (2011) Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nat Struct Mol Biol.* **18**, 404–409
22. Milbradt, A. G., Kulkarni, M., Yi, T., Takeuchi, K., Sun, Z.-Y. J., Luna, R. E., Selenko, P., Näär, A. M., and Wagner, G. (2011) Structure of the VP16 transactivator target in the Mediator. *Nat Struct Mol Biol.* **18**, 410–415
23. Eletsky, A., Ruyechan, W. T., Xiao, R., Acton, T. B., Montelione, G. T., and Szyperski, T. (2011) Solution NMR structure of MED25(391-543) comprising the activator-interacting domain (ACID) of human mediator subunit 25. *J Struct Funct Genomics.* **12**, 159–166
24. Henderson, A. R., Henley, M. J., Foster, N. J., Peiffer, A. L., Beyersdorf, M. S., Stanford, K. D., Sturlis, S. M., Linhares, B. M., Hill, Z. B., Wells, J. A., Cierpicki, T., Brooks, C. L., Fierke, C. A., and Mapp, A. K. (2018) Conservation of coactivator engagement mechanism enables small-molecule allosteric modulators. *Proc Natl Acad Sci U S A.* **115**, 8960–8965
25. Lee, M.-S., Lim, K., Lee, M.-K., and Chi, S.-W. (2018) Structural Basis for the Interaction between p53 Transactivation Domain and the Mediator Subunit MED25. *Molecules.* **23**, 2726
26. Jeffery, H. M., and Weinzierl, R. O. J. (2020) Multivalent and Bidirectional Binding of Transcriptional Transactivation Domains to the MED25 Coactivator. *Biomolecules.* **10**, 1205
27. Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., and Ferrin, T. E. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science.* **30**, 70–82
28. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature.* **596**, 583–589
29. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., Van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhllheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J., and Baker, D. (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science.* **373**, 871–876
30. Wu, R., Ding, F., Wang, R., Shen, R., Zhang, X., Luo, S., Su, C., Wu, Z., Xie, Q., Berger, B., Ma, J., and Peng, J. (2022) High-resolution de novo structure prediction from primary sequence. *bioRxiv: The Preprint Server for Biology.* 10.1101/2022.07.21.500999
31. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., Dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., and Rives, A. (2023) Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science.* **379**, 1123–1130
32. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nat Methods.* **19**, 679–682
33. The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., Da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Joshi, V., Jyothi, D., Kandasamy, S., Lock, A., Luciani, A., Lugaric, M., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Mishra, A., Moulang, K., Nightingale, A., Pundir, S., Qi, G., Raj, S., Raposo, P., Rice, D. L., Saidi, R., Santos, R., Speretta, E., Stephenson, J., Tootoo, P., Turner, E., Tyagi, N.,

- Vasudev, P., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A. J., Aimo, L., Argoud-Puy, G., Auchincloss, A. H., Axelsen, K. B., Bansal, P., Baratin, D., Batista Neto, T. M., Blatter, M.-C., Bolleman, J. T., Boutet, E., Breuza, L., Gil, B. C., Casals-Casas, C., Echioukh, K. C., Coudert, E., Cuche, B., De Castro, E., Estreicher, A., Famiglietti, M. L., Feuermann, M., Gasteiger, E., Gaudet, P., Gehant, S., Gerritsen, V., Gos, A., Gruaz, N., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Kerhornou, A., Le Mercier, P., Lieberherr, D., Masson, P., Morgat, A., Muthukrishnan, V., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Poux, S., Pozzato, M., Pruess, M., Redaschi, N., Rivoire, C., Sigrist, C. J. A., Sonesson, K., Sundaram, S., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Huang, H., Laiho, K., McGarvey, P., Natale, D. A., Ross, K., Vinayaka, C. R., Wang, Q., Wang, Y., and Zhang, J. (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*. **51**, D523–D531
34. Elfmann, C., and Stülke, J. (2023) PAE viewer: a webserver for the interactive visualization of the predicted aligned error for multimer structure predictions and crosslinks. *Nucleic Acids Research*. **51**, W404–W410
35. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Židek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., Potapenko, A., Ballard, A. J., Romera-Paredes, B., Nikolov, S., Jain, R., Clancy, E., Reiman, D., Petersen, S., Senior, A. W., Kavukcuoglu, K., Birney, E., Kohli, P., Jumper, J., and Hassabis, D. (2021) Highly accurate protein structure prediction for the human proteome. *Nature*. **596**, 590–596
36. Evans, R., O’Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Židek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., and Hassabis, D. (2021) Protein complex prediction with AlphaFold-Multimer. *bioRxiv: The Preprint Server for Biology*. 10.1101/2021.10.04.463034
37. Sela, D., Conkright, J. J., Chen, L., Gilmore, J., Washburn, M. P., Florens, L., Conaway, R. C., and Conaway, J. W. (2013) Role for human mediator subunit MED25 in recruitment of mediator to promoters by endoplasmic reticulum stress-responsive transcription factor ATF6 α . *J Biol Chem*. **288**, 26179–26187
38. Sullivan, S. M., Horn, P. J., Olson, V. A., Koop, A. H., Niu, W., Ebright, R. H., and Triezenberg, S. J. (1998) Mutational analysis of a transcriptional activation region of the VP16 protein of herpes simplex virus. *Nucleic Acids Res*. **26**, 4487–4496
39. Yang, M., Hay, J., and Ruyechan, W. T. (2008) Varicella-zoster virus IE62 protein utilizes the human mediator complex in promoter activation. *J Virol*. **82**, 12154–12163
40. Bruley, A., Mornon, J.-P., Duprat, E., and Callebaut, I. (2022) Digging into the 3D Structure Predictions of AlphaFold2 with Low Confidence: Disorder and Beyond. *Biomolecules*. **12**, 1467
41. Elfving, N., Davoine, C., Benloch, R., Blomberg, J., Brännström, K., Müller, D., Nilsson, A., Ulfstedt, M., Ronne, H., Wingsle, G., Nilsson, O., and Björklund, S. (2011) The Arabidopsis thaliana Med25 mediator subunit integrates environmental cues to control plant development. *Proc Natl Acad Sci U S A*. **108**, 8245–8250
42. Remaut, H., and Waksman, G. (2006) Protein–protein interaction through β -strand addition. *Trends in Biochemical Sciences*. **31**, 436–444
43. Tiwari, S. B., Belachew, A., Ma, S. F., Young, M., Ade, J., Shen, Y., Marion, C. M., Holtan, H. E., Bailey, A., Stone, J. K., Edwards, L., Wallace, A. D., Canales, R. D., Adam, L., Ratcliffe, O. J., and Repetti, P. P. (2012) The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant J*. **70**, 855–865
44. Bugge, K., Staby, L., Kemplen, K. R., O’Shea, C., Bendsen, S. K., Jensen, M. K., Olsen, J. G., Skriver, K., and Kragelund, B. B. (2018) Structure of Radical-Induced Cell Death1 Hub Domain Reveals a Common $\alpha\alpha$ -Scaffold for Disorder in Transcriptional Networks. *Structure*. **26**, 734–746.e7
45. Theisen, F. F., Staby, L., Tidemand, F. G., O’Shea, C., Prestel, A., Willemoës, M., Kragelund, B. B., and Skriver, K. (2021) Quantification of Conformational Entropy Unravels Effect of Disordered Flanking Region in Coupled Folding and Binding. *J. Am. Chem. Soc.* **143**, 14540–14550
46. Tuttle, L. M., Pacheco, D., Warfield, L., Luo, J., Ranish, J., Hahn, S., and Klevit, R. E. (2018) Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex. *Cell Rep*. **22**, 3251–3264
47. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*. **6**, 197–208
48. Berlow, R. B., Dyson, H. J., and Wright, P. E. (2022) Multivalency enables unidirectional switch-like competition between intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117338119
49. Radhakrishnan, I., Perez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., and Wright, P. E. (1997) Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell*. **91**, 741–752

50. Ruff, K. M., and Pappu, R. V. (2021) AlphaFold and Implications for Intrinsically Disordered Proteins. *Journal of Molecular Biology*. **433**, 167208
51. Sala, D., Engelberger, F., Mchaourab, H. S., and Meiler, J. (2023) Modeling conformational states of proteins with AlphaFold. *Current Opinion in Structural Biology*. **81**, 102645
52. Sen, N., Anishchenko, I., Bordin, N., Sillitoe, I., Velankar, S., Baker, D., and Orengo, C. (2022) Characterizing and explaining the impact of disease-associated mutations in proteins without known structures or structural homologs. *Briefings in Bioinformatics*. **23**, bbac187
53. Tuttle, L. M., Pacheco, D., Warfield, L., Wilburn, D. B., Hahn, S., and Klevit, R. E. (2021) Mediator subunit Med15 dictates the conserved “fuzzy” binding mechanism of yeast transcription activators Gal4 and Gcn4. *Nat Commun*. **12**, 2220
54. Brzovic, P. S., Heikaus, C. C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Klevit, R. E., and Hahn, S. (2011) The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. *Mol Cell*. **44**, 942–953
55. Bret, H., Andreani, J., and Guerois, R. (2023) From interaction networks to interfaces: Scanning intrinsically disordered regions using AlphaFold2. *bioRxiv: The Preprint Server for Biology*. 10.1101/2023.05.25.542287
56. Staller, M. V., Holehouse, A. S., Swain-Lenz, D., Das, R. K., Pappu, R. V., and Cohen, B. A. (2018) A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. *Cell Syst*. **6**, 444-455.e6
57. Staller, M. V., Ramirez, E., Kotha, S. R., Holehouse, A. S., Pappu, R. V., and Cohen, B. A. (2022) Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains. *Cell Systems*. **13**, 334-345.e5
58. Erijman, A., Kozlowski, L., Sohrabi-Jahromi, S., Fishburn, J., Warfield, L., Schreiber, J., Noble, W. S., Söding, J., and Hahn, S. (2020) A High-Throughput Screen for Transcription Activation Domains Reveals Their Sequence Features and Permits Prediction by Deep Learning. *Mol Cell*. **78**, 890-902.e6
59. Sanborn, A. L., Yeh, B. T., Feigerle, J. T., Hao, C. V., Townshend, R. J., Lieberman Aiden, E., Dror, R. O., and Kornberg, R. D. (2021) Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. *Elife*. **10**, e68068
60. Theisen, F., Prestel, A., Elkjær, S., Leurs, Y., Morffy, N., Strader, L., O’Shea, C., Teilum, K., Kragelund, B., and Skriver, K. (2023) Molecular switching in transcription through splicing and proline-isomerization regulates stress responses in plants. *Research Square*. 10.21203/rs.3.rs-3034274/v1