

DIRseq as a method for predicting drug-interacting residues of intrinsically disordered proteins from sequences

Matt MacAinsh¹, Sanbo Qin¹, Huan-Xiang Zhou^{1,2*}

¹Department of Chemistry, University of Illinois Chicago, Chicago, United States;

²Department of Physics, University of Illinois Chicago, Chicago, United States

eLife Assessment

This **important** study presents a sequence-based method for predicting drug-interacting residues in intrinsically disordered proteins (IDPs), addressing a significant challenge in understanding small-molecule:IDP interactions. The findings have **solid** support through examples underscoring the role of aromatic interactions. While predicted binding sites remain coarse, validation was done on a total of 10 IDPs at varying depths. The method builds on the authors' previous work and, with ad hoc modifications, is poised to benefit this emerging field.

Abstract Intrinsically disordered proteins (IDPs) are now well-recognized as drug targets. Identifying drug-interacting residues is valuable for both optimizing compounds and elucidating the mechanism of action. Currently, NMR chemical shift perturbation and all-atom molecular dynamics (MD) simulations are the primary tools for this purpose. Here, we present DIRseq, a fast method for predicting drug-interacting residues from the amino-acid sequence. All residues contribute to the propensity of a particular residue to be drug-interacting; the contributing factor of each residue has an amplitude that is determined by its amino-acid type and attenuates with increasing sequence distance from the particular residue. DIRseq predictions match well with drug-interacting residues identified by NMR chemical shift perturbation and other methods, including residues L₂₂WK₂₄ and Q₅₂WFT₅₅ in the tumor suppressor protein p53. These successes augur well for deciphering the sequence code for IDP-drug binding. DIRseq is available as a web server at <https://zhougrou-pu.github.io/DIRseq/> and has many applications, such as virtual screening against IDPs and designing IDP fragments for in-depth NMR and MD studies.

*For correspondence:

hzhou43@uic.edu

Competing interest: The authors declare that no competing interests exist.

Funding: See page 12

Sent for Review

02 May 2025

Preprint posted

29 July 2025

Reviewed preprint posted

14 August 2025

Reviewed preprint revised

16 September 2025

Version of Record published

07 October 2025

Reviewing Editor: Martin Graña, Institut Pasteur de Montevideo, Uruguay

© Copyright MacAinsh et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

Intrinsically disordered proteins (IDPs) are now recognized as important drug targets (Joshi and Vendruscolo, 2015; Saurabh et al., 2023; Wang et al., 2023; Uversky, 2024). For structured protein targets, a crucial step in characterizing drug binding is identifying the drug-binding pocket. Although an IDP can be locked into a specific conformation by a drug molecule in rare cases (Peterson et al., 2004), the prevailing scenario is that the protein remains disordered upon drug binding (Woods et al., 2011; Ono et al., 2012; Jin et al., 2013; De Mol et al., 2016; Heller et al., 2020; Iwaya et al., 2020; Zhao et al., 2021; Robustelli et al., 2022; Heller et al., 2024). Consequently, the IDP-drug complex typically samples a vast conformational space, and the drug molecule only exhibits preferences, rather than exclusiveness, for interacting with subsets of residues (Jin et al., 2013; Zhao et al., 2021; Robustelli et al., 2022). Such drug-interacting residues, akin to binding pockets in structured proteins, are key to optimizing compounds (Robustelli et al., 2022; Iconaru et al., 2015; Basu et al.,

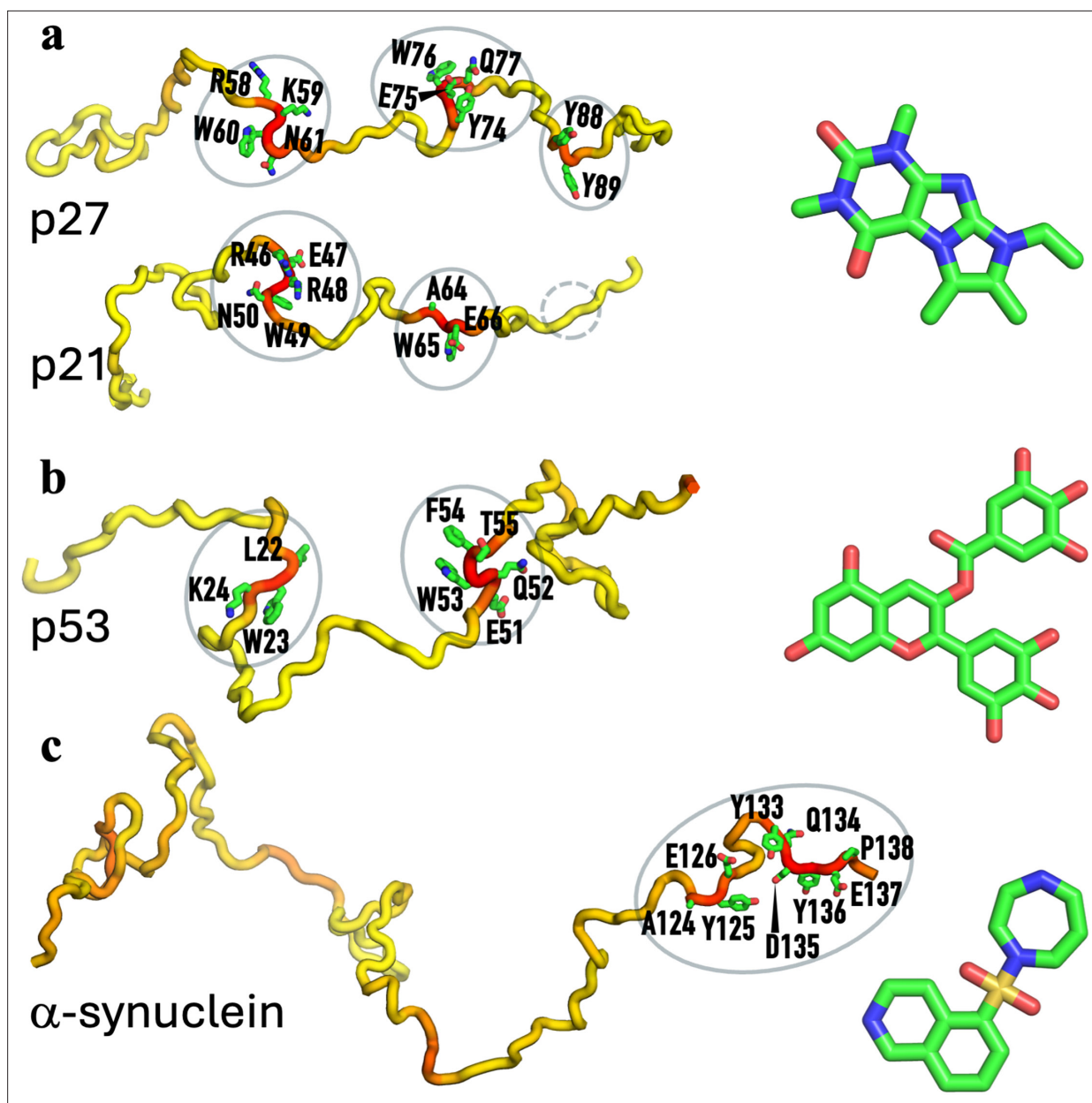


Figure 1. Four intrinsically disordered proteins (IDPs) and drugs that bind to them. (a) p27, p21, and SJ403. (b) p53 and epigallocatechin-3-gallate (EGCG). (c) α-Synuclein and Fasudil. For each IDP, the DIRseq propensities are rendered by a color spectrum from yellow for low values to red for high values. Predicted drug-interacting residues are shown with sidechains rendered in stick.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. *q* parameters.

2023) and elucidating the mechanism of action (Woods et al., 2011; De Mol et al., 2016; Zhao et al., 2021; Iconaru et al., 2015; Basu et al., 2023).

NMR chemical shift perturbation (CSP) is the best experimental method for identifying drug-interacting residues and has been applied to many IDPs (Ono et al., 2012; De Mol et al., 2016; Iwaya et al., 2020; Zhao et al., 2021; Robustelli et al., 2022; Heller et al., 2024; Iconaru et al., 2015; Basu et al., 2023; Hammoudeh et al., 2009; Lamberto et al., 2009; Krishnan et al., 2014; Tatenhorst et al., 2016; Ahmed et al., 2017; Kocis et al., 2017; Neira et al., 2017; Ruan et al., 2021; Iruela et al., 2022; Xu et al., 2022). Aromatic residues are frequently among drug-interacting residues (Zhao et al., 2021; Robustelli et al., 2022; Iconaru et al., 2015; Basu et al., 2023; Lamberto et al.,

2009; Tatenhorst et al., 2016; Ruan et al., 2021). This is understandable as drug molecules typically are rich in aromatic rings (Figure 1 and Supplementary file 1A), which can form π - π interactions with aromatic residues. As recently pointed out by Heller et al., 2024, CSPs of IDPs elicited by drug binding can be small enough to fall within the spectral resolution of NMR spectroscopy, therefore making it difficult to unequivocally identify drug-interacting residues. The small magnitude of CSPs may arise because the drug does not have a strong preference for interacting with any residues. Another scenario may be that drug binding induces a conformational shift such as secondary structure formation or even partial folding (De Mol et al., 2016; Iwaya et al., 2020; Robustelli et al., 2022), so CSPs spread from the directly interacting residues to the rest of the IDP, adding to the difficulty in identifying drug-interacting residues.

All-atom molecular dynamics (MD) simulations have presented atomic details in many IDP-drug systems (Jin et al., 2013; Heller et al., 2020; Zhao et al., 2021; Robustelli et al., 2022; Basu et al., 2023; Ruan et al., 2021; Yu et al., 2016; Herrera-Nieto et al., 2020; Wang et al., 2022; Zhu et al., 2022; Mehta and Goyal, 2024). These simulation studies have highlighted the frequent engagement of drug molecules with aromatic residues, particularly in π - π interactions (Zhao et al., 2021; Robustelli et al., 2022; Basu et al., 2023; Ruan et al., 2021; Herrera-Nieto et al., 2020; Wang et al., 2022; Zhu et al., 2022; Mehta and Goyal, 2024). MD simulations also revealed the emergence of a compact subpopulation of the N-terminal disordered region of the tumor suppressor protein p53 upon binding the drug epigallocatechin-3-gallate (EGCG) (Zhao et al., 2021) and reduced solvent exposure of hydrophobic residues in the 42-residue amyloid- β peptide (A β 42) upon binding the drug 10074-G5 (Heller et al., 2020). However, MD simulations of IDPs still suffer from the perennial issues of inaccurate force fields and insufficient sampling, which are exacerbated when drug molecules are included. For example, two replicate simulations may identify different drug-interacting residues (Lama et al., 2024).

Here, we present a sequence-based method, DIRseq, as a complement to NMR CSP and MD simulations. This method was motivated by our observation that drug-interacting residues seem to overlap with residues exhibiting elevated transverse relaxation rates (R_2) (Qin and Zhou, 2024; Muhammedkutty et al., 2025), as exemplified by the C-terminal 20 residues of α -synuclein (Robustelli et al., 2022). Elevated R_2 in IDPs is caused by either local inter-residue interactions or residual secondary structures (Dey et al., 2022). Based on this understanding, we developed a sequence-based method, SeqDYN, to predict R_2 of IDPs (Qin and Zhou, 2024). As suggested previously (Muhammedkutty et al., 2025), the propensities of residues to form intramolecular interactions and therefore elevate R_2 should be similar to those for forming intermolecular interactions with drug molecules. DIRseq is therefore an adaptation of SeqDYN. DIRseq predictions match well with drug-interacting residues identified by NMR CSP and other methods, including residues L₂₂WK₂₄ and Q₅₂WFT₅₅ in p53 and C-terminal residues in α -synuclein. DIRseq is available as a web server at <https://zhougroup-uic.github.io/DIRseq/> and has many applications.

Results

Retooling of SeqDYN into DIRseq

SeqDYN (Qin and Zhou, 2024) predicts the R_2 value of residue n by accumulating a contributing factor $f(i; n)$ from each residue i :

$$R_2(n) = \Upsilon \prod_{i=1}^N f(i; n)$$

where N is the total number of residues in the IDP and Υ is a uniform scale parameter. The contributing factor $f(i; n)$ has an amplitude $q(i)$ that is determined by the amino-acid type of residue i and attenuates with increasing sequence distance, $s = |i - n|$, from residue n :

$$f(i; n) = 1 + \frac{q(i) - 1}{1 + bs^2}$$

The 20 q parameters (one for each amino-acid type) and b were trained on the measured R_2 values for 45 IDPs. The original q values (Figure 1—figure supplement 1a) show that aromatic (Trp, Tyr, His, and Phe), Arg, and long aliphatic (Ile and Leu) amino acids are interaction-prone and tend to elevate

R_2 . (The original SeqDYN also has an option of applying a helix boost; we did not apply this boost here.) We transformed SeqDYN into DIRseq by implementing four changes.

First, we reassigned four q parameters. We lowered the q values of the three long aliphatic amino acids, Leu, Ile, and Met, to the q value of a short aliphatic amino acid, Val, because long aliphatic amino acids primarily participate in hydrophobic interactions, which may be less important for stabilizing the binding of a small molecule in sites largely exposed to water. At the same time, we increased the q value of Asp to be the same as that of Glu, to increase the role of Asp's electrostatic interactions. Both the downgrade of hydrophobic interactions and the upgrade of electrostatic interactions were motivated by observations on drug binding to α -synuclein in MD simulations by **Robustelli et al., 2022**. The modified set of q parameters is displayed in **Figure 1—figure supplement 1b**.

Second, the original b value, 0.0316, corresponds to a correlation half-length, $b^{-1/2}$, of 5.6 residues. Given the small size of drug molecules, we increased the b value to 0.3, corresponding to a correlation half-length of 1.8 residues. That is, we expect a drug molecule to interact with 3–4 residues at a time. Third, the SeqDYN-predicted R_2 profile, capturing experimental observations (**Klein-Seetharaman et al., 2002**), falls off at both termini, because no residues beyond the termini are present to provide a contributing factor to R_2 . However, whereas intra-IDP interactions experience such a terminal effect, IDP-drug interactions do not. To eliminate the terminal effect, we padded the original IDP sequence by a stretch of 12 Gln residues at each terminus. Gln was selected because its q value is at the middle of the 20 q values (**Figure 1—figure supplement 1b**).

Lastly, we converted high and low R_2 values into high and low drug-interacting propensities, respectively, using a sigmoid function

$$P = \frac{100}{1 + e^{-\frac{R_2 - R_{2th}}{R_{2wd}}}}$$

where the midpoint R_{2th} and width R_{2wd} of the transition region are determined by the mean (m) and standard deviation (SD) of the R_2 values over the entire sequence. Specifically,

$$R_{2th} = m + s_1 SD$$

$$R_{2wd} = m/s_2$$

We chose s_1 to be 1.5 and s_2 to be 14.0. On the DIRseq web server, users can either keep these default values for b , s_1 , and s_2 , or enter values of their choice. The source code has been deposited on GitHub under the file name DIRseq.js (<https://github.com/hzhou43/DIRseq/>; **Zhou, 2025**). The values of P range from 0 to 100; $P = 50$ when R_2 is at the 'threshold' value R_{2th} . Residues with $P \geq 50$ are predicted to be drug-interacting, but an isolated residue with $P \geq 50$ is discarded, as it can be reasonably expected that interactions with at least two consecutive (or nearby) residues are needed to generate sufficient drug binding stability.

Detailed assessment of DIRseq on four IDPs

Drug-interacting residues in four IDPs or intrinsically disordered regions (IDRs) have been thoroughly characterized by NMR CSP. In **Figure 1** (left images), we show these IDPs in a single conformation, with DIRseq propensities rendered in a color spectrum (from low in yellow to high in red); predicted drug-interacting residues (i.e. those with $P \geq 50$) are shown with sidechains in stick. Drug molecules that bind to these IDPs are also displayed in **Figure 1** (right images). We display NMR CSPs as blue bars and DIRseq propensities as a red curve in **Figure 2**. Unless otherwise noted, we use $m+1.5$ SD as the threshold for identifying drug-interacting residues, for both CSP and DIRseq; this threshold is indicated by a horizontal dashed line in **Figure 2**. Experimentally identified drug-interacting residues are further indicated by cyan shading. Below, we present a detailed comparison between CSP and DIRseq.

The kinase inhibitory domain (residues 22–105) of the cell cycle regulator p27 harbors three aromatic-centered motifs that interact with a group of compounds represented by SJ403 (**Figure 1a**), as found by **Iconaru et al., 2015**. These motifs, $W_{60}N_{61}$, $E_{75}WQ_{77}$, and $Y_{88}Y_{89}$ (**Figure 2a**, cyan shading), are correctly predicted by DIRseq (**Figure 2a**, portions of the red curve above the 50% threshold). Specifically, the predicted drug-interacting residues are $R_{58}KWN_{61}$, $Y_{74}EWQ_{77}$, and $Y_{88}Y_{89}$ (**Figure 1a**,

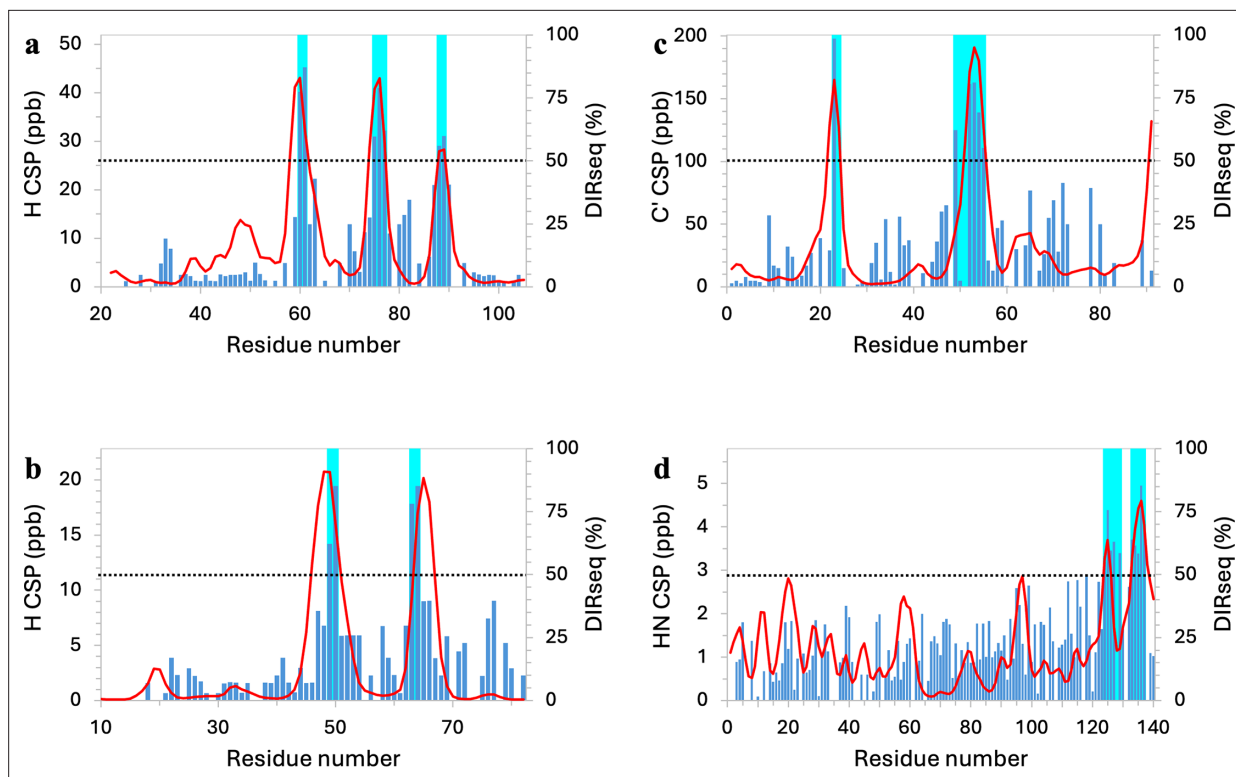


Figure 2. Comparison of DIRseq propensities with NMR chemical shift perturbations (CSPs). (a) p27. (b) p21. (c) p53. (d) α -Synuclein. CSPs are displayed as blue bars and in units of ppb; CSP-identified drug-interacting residues are indicated by cyan shading. DIRseq predictions are shown as red curves. The ordinate scales are chosen so that the $m+1.5$ SD threshold for CSP is at the same height as the 50% threshold for DIRseq, indicated by a horizontal dashed line.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Source data for **Figure 2**.

Figure supplement 1. Correlation between NMR chemical shift perturbation (CSP) and DIRseq propensity.

Figure supplement 2. Correlation of chemical shift perturbations (CSPs) elicited by two different compounds.

Figure supplement 3. Two compounds that bind to p27 with different characteristics.

left image). Moreover, the DIRseq propensity profile matches well the CSP profile along the sequence (**Figure 2a**). The two parameters show a very strong correlation, with a Pearson correlation coefficient (r) of 0.82 (**Figure 2—figure supplement 1a**).

Iconaru et al. also studied SJ403 binding to a related cell cycle regulator p21. Their CSP data identified the first two corresponding motifs, $W_{49}N_{50}$ and $F_{63}A_{64}$, but not the third as drug-interacting residues (**Figure 2b**, cyan shading). DIRseq correctly predicts both the first two motifs above the 50% threshold and the third motif below the threshold (**Figure 2b**, red curve; **Figure 1b**, left image). As noted by Iconaru et al., the p21 counterpart to the p27 third motif $F_{87}YY_{89}$ is $L_{76}YL_{78}$; the two Leu residues in p21 are no match for F_{87} and Y_{89} in p27. For the DIRseq method, Leu has a lower q value than Tyr and Phe (**Figure 1—figure supplement 1b**). CSP and DIRseq propensity show a strong correlation ($r=0.66$; **Figure 2—figure supplement 1b**).

NMR CSP revealed two motifs, $W_{23}K_{24}$ and $D_{49}LEQWFT_{55}$, in the N-terminal region of p53 as interacting with EGCG (**Zhao et al., 2021**; **Figure 2c**, cyan shading). MD simulations supported these drug-binding sites. DIRseq again correctly predicts these two motifs (**Figure 2c**, red curve). Specifically, the predicted drug-interacting motifs are $L_{22}WK_{24}$ and $E_{51}QWFT_{55}$ (**Figure 1c**, left image). Here also CSP and DIRseq propensity show a strong correlation ($r=0.67$; **Figure 2—figure supplement 1c**).

Using the $m+1.5$ SD threshold, CSPs of **Robustelli et al., 2022** select residues $A_{124}YEMPS_{129}$ and $Y_{133}QDYE_{137}$ of α -synuclein as interacting with the drug Fasudil (**Figure 2d**, cyan shading). These two C-terminal motifs are also identified by DIRseq, with residues $A_{124}YE_{126}$ and $Y_{133}QDYEP_{138}$ above the 50% threshold (**Figure 2d**, red curve; **Figure 1d**, left image). There is a moderate correlation ($r=0.51$;

Figure 2—figure supplement 1d) between CSP and DIRseq propensity. Robustelli et al. further characterized the IDP-drug interactions by all-atom simulations, using both full-length α -synuclein and a C-terminal fragment (residues 121–140). Further simulations of the C-terminal fragment binding with additional compounds led to a compound known as Ligand 47 with a somewhat higher affinity than Fasudil. The CSP profiles elicited by Fasudil and Ligand 47 are similar, though with larger amplitudes around Y₁₂₅ and Y₁₃₆ by the latter; their correlation coefficient is 0.78 (**Figure 2—figure supplement 2a**). Because DIRseq does not consider any information about the drug molecule, this correlation coefficient between CSPs elicited by two different compounds might be viewed as an upper bound of what can be achieved for the correlation between CSP and DIRseq propensity.

Dependences of DIRseq prediction accuracies on model parameters

We now use the CSP data of the above four IDPs to assess the dependences of DIRseq prediction accuracies on model parameters (**Supplementary file 1B**). We use two complementary measures for accuracy: the Pearson correlation between CSP and drug-interacting propensity and the difference between true positive (TP) and false positive (FP). The four IDPs have a total of 31 CSP-identified drug-interacting residues. Recall that DIRseq uses a sigmoid function to convert the R_2 output of SeqDYN into drug-interacting propensities. Using the original 20 q parameters along with the present values for b , s_1 , and s_2 , the predictions are already pretty reasonable: the correlations for the four IDPs range from 0.46 to 0.79 (sum = 2.43) and TP outnumbers FP by 7 residues. Indeed, this initial success was anticipated (**Qin and Zhou, 2024; Muhammedkutty et al., 2025**) and validated the premise of DIRseq, i.e., intra-IDP interactions that elevate R_2 also tend to mediate drug interactions. As stated above, we tweaked 4 of the 20 q parameters to arrive at the final DIRseq; now the r sum increases to 2.66 and TP – FP increases to 15.

Examining the four q parameter changes one at a time, the downgrade of q for Leu increases both r sum (to 2.53) and TP – FP (to 16) relative to the counterparts with the original SeqDYN q parameters. In comparison, the downgrade of q for Ile does not affect r sum and yields a small increase in TP – FP (to 10), the upgrade of q for Asp yields small increases in both r sum (to 2.52) and TP – FP (to 8), but the downgrade of q for Met actually decreases both r sum (to 2.37) and TP – FP (to 6). The latter result shows that we made these parameter tweaks not solely for increasing accuracy but for physical reasons, i.e., to reduce the role of hydrophobic interactions but elevate the role of electrostatic interactions in IDP-drug binding, as suggested by MD simulations.

We also tested alternative models for q parameters. Given the prominence of aromatic amino acids in drug binding as revealed by CSP, we wondered how a model that solely emphasizes aromatic amino acids would perform. To that end, we tested a model with only two different q values for the 20 amino acids: a high value (same as that of Trp in SeqDYN) for all three aromatic amino acids (Trp, Tyr, and Phe) and a low value (same as that of Gly in SeqDYN) for all other amino acids. This is similar to a sticker-spacer model for simulating liquid-liquid phase separation (**Martin et al., 2020**). This aromatic model achieves the same r sum, 2.66, as DIRseq, but its FP outnumbers its TP, such that TP – FP = –1. **Tesei and Lindorff-Larsen, 2022** parameterized a coarse-grained force field to simulate liquid-liquid phase separation. Their λ ‘stickiness’ parameters have a good correlation with the SeqDYN original q parameters (**Zhou et al., 2024**). We replaced the q parameters by these λ parameters (after scaling to the same numerical range as the q parameters); the resulting model has both a low r sum (1.21) and a low TP – FP (–21). This outcome suggests that, unlike liquid-liquid phase separation where stickiness is the main drive, drug binding is more selective in the type of intermolecular interactions. Lastly, we tested a model where an average hydropathy scale (compiled by **Tesei et al., 2021**) was used in place of the q parameters; the resulting model has very little predictive value (r sum = 0.44 and TP – FP = –20). This last outcome is in line with our downgrade of hydrophobic interactions in DIRseq. The test results from all these alternative models indicate that the q parameters in the final DIRseq model capture the appropriate balance among aromatic, electrostatic, and hydrophobic interactions in IDP-drug binding.

In addition to the 20 q parameters, DIRseq has 3 other parameters. The b parameter determines the number of residues that simultaneously interact with a drug molecule. In **Supplementary file 1**, we list the performance measures when b is varied. These results show that our final choice, 0.3, for b is optimal for both r sum and TP – FP. The corresponding number of residues, 3–4, that simultaneously interact with a drug molecule fits with the narrow range of drug molecule sizes in the present

study (**Supplementary file 1A**; molecular weights: 360 ± 130 Da). The last two parameters, s_1 and s_2 , are in the sigmoid function that converts the R_2 output into drug-interacting propensities. s_1 sets the R_2 threshold for labeling a residue as drug-interacting. Again, our final choice, corresponding to $m+1.5$ SD for the threshold, achieves an optimum for r sum and TP – FP. A lower threshold leads to a high FP and also a slight deterioration in r sum, whereas a higher threshold leads to a low FP and possibly a minuscule improvement in r sum. s_2 controls the sharpness of the sigmoid function in the transition region and affects r sum but not TP – FP. r sum increases with increasing s_2 ; we chose an s_2 value, where r sum is nearly at the saturation level.

DIRseq as a complement to CSP for assigning drug-interacting residues

After assessing the achievable accuracy of DIRseq, we now consider an application: its combination with CSP to make robust assignments of drug-interacting residues. We note that the above CSP-identified drug-interacting motifs are all anchored on one or more aromatic residues, and this feature likely contributes to the good performance of DIRseq. When a clear ‘aromatic signal’ is not present, CSP or DIRseq alone may not be able to conclusively identify drug-interacting residues. However, a consensus identification by CSP and DIRseq may be reliable. MD simulations have played such a complementary role to CSP in several studies (**Zhao et al., 2021**; **Robustelli et al., 2022**; **Basu et al., 2023**; **Zhu et al., 2022**).

De Mol et al., 2016 reported the CSPs of an IDR called AF-1* (residues 142–448) in the androgen receptor elicited by EPI-001 and its stereoisomers, including EPI-002; small but reproducible CSPs were found in three subregions, R1 (residues 341–371), R2 (residues 391–414), and R3 (residues 426–446). Conversely, AF-1* caused changes in the ^1H NMR spectrum of EPI-001, but the individual peptides corresponding to R1, R2, and R3 did not, suggesting that, rather than separately interacting with the three subregions, EPI-001 induces a partial folding of the three subregions. MD simulations captured the partial folding of the R2-R3 fragment (residues 391–446) induced by EPI-002 and underscored the importance of aromatic residues in drug binding (**Zhu et al., 2022**). **Basu et al., 2023** reported the CSPs of the transactivation unit 5 (Tau-5*, residues 336–448) in AF-1* by EPI-001 and a more potent variant, 1aa (**Figure 3a**, **Figure 2—figure supplement 2b**). The contact probabilities of the R2-R3 fragment with 1aa in MD simulations reaffirmed the importance of aromatic residues.

De Mol et al., 2016 and **Basu et al., 2023** were careful not to name any drug-interacting residues based on CSPs. In **Figure 3a**, we compare the 1aa-elicited CSPs and DIRseq propensities of Tau-5*. In agreement with both NMR studies, DIRseq identifies drug-interacting residues in the middle of each of R1-R3: $R_{360}\text{DYY}_{363}$, $A_{396}\text{WAA}_{399}$, and $W_{433}\text{H}_{434}$. In addition, the latter two motifs showed the highest drug-contact probabilities in separate MD simulations of the R2-R3 fragment (**Basu et al., 2023**; **Zhu et al., 2022**). For the R2 and R3 subregions, CSPs above the $m+1.5$ SD threshold are observed at residues downstream of the DIRseq identifications, so we propose to expand the drug-interacting motifs to $A_{396}\text{WAAAAAQ}_{403}$ and $W_{433}\text{HTLF}_{437}$. The three putative drug-interacting motifs are indicated as cyan shading in **Figure 3a**.

Heller et al., 2024 used ^{19}F transverse relaxation measurements to determine the binding affinity of the disordered domains 2 and 3 of the hepatitis C virus NS5A protein (NS5A-D2D3, residues 247–466) for 5-fluorindole. They also measured ^1H - ^{15}N CSPs at two ligand concentrations but described them as ‘nearly undetectable’ (**Figure 3b**). We speculate that the small CSPs may be due to the small size of the ligand, making it difficult to interact with multiple residues simultaneously and thus achieve sufficient binding stability. In any event, CSPs above the $m+1.5$ SD threshold were largely isolated (and thus appeared to be random), and there was not much overlap between residues having these above-the-threshold CSPs at the two ligand concentrations. The one exception is a motif around W_{312} , which had above-the-threshold CSPs at both ligand concentrations, and nearby residues A_{308} and A_{313} also had above-the-threshold CSPs at one of the ligand concentrations. DIRseq predicts the motif $P_{310}\text{AWARPD}_{316}$ as drug-interacting residues. We propose the expanded motif, $A_{308}\text{LPAWARPD}_{316}$, as residues interacting with 5-fluorindole (**Figure 3b**, cyan shading). DIRseq also predicts two more motifs, $E_{323}\text{SWRRPDY}_{330}$ and $R_{352}\text{RRR}_{355}$, as drug-interacting residues, which remain to be tested.

The fibrillation of acid-denatured β_2 microglobulin is inhibited by rifamycin SV (**Woods et al., 2011**). An aromatic-rich motif, $W_{60}\text{SFYLLYYTEF}_{70}$, was implicated in the nucleation of fibrillation and also involved in ligand binding, as a triple mutation, F62A/Y63A/Y67A, significantly weakened binding. Low intensities of NMR peaks from residues 58–79 (possibly due to the formation of residual

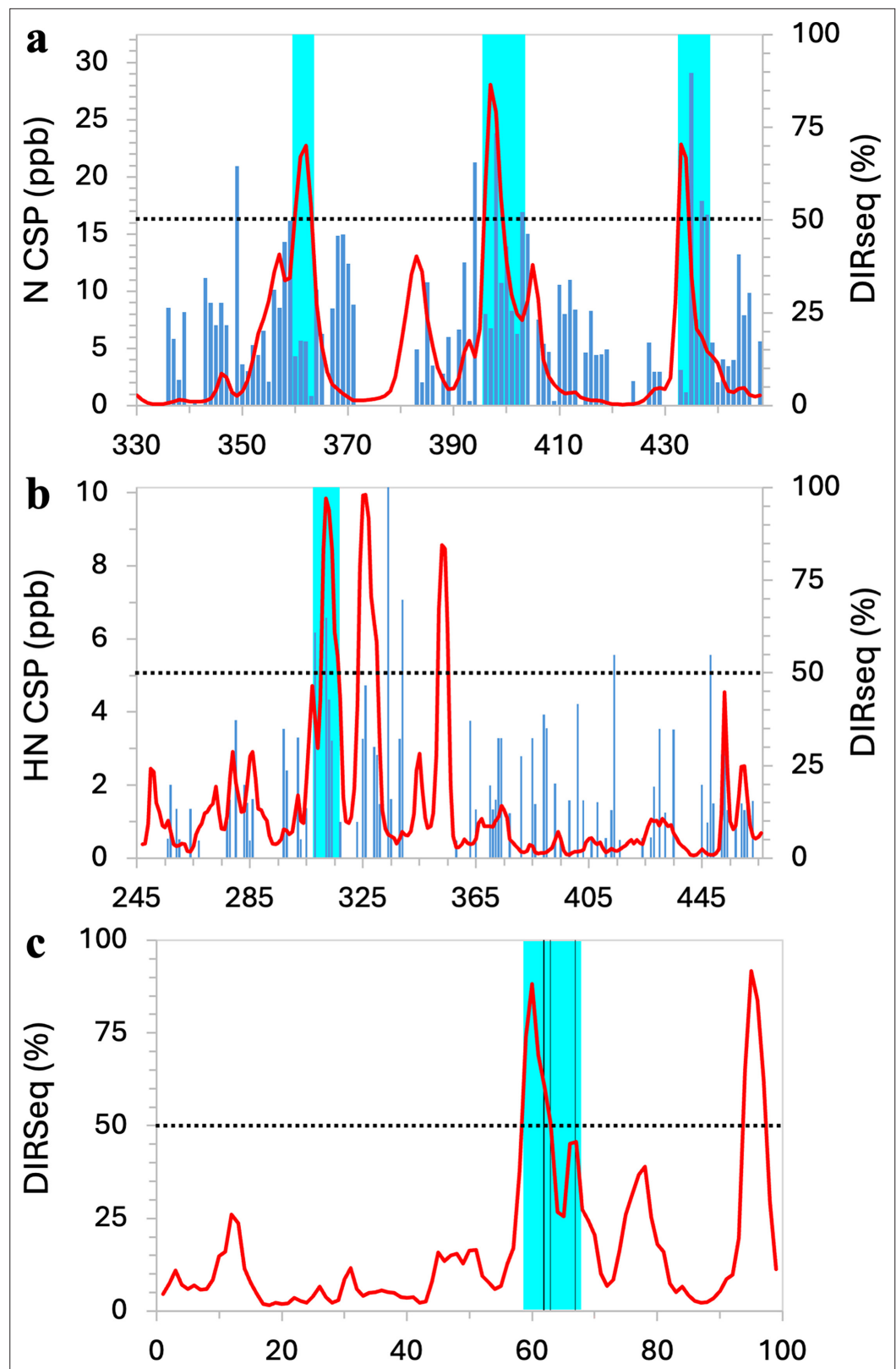


Figure 3. Drug-binding sites identified by combining chemical shift perturbation (CSP) or mutation data with DIRseq predictions. (a) Tau-5*. (b) NS5A-D2D3. (c) β2 microglobulin. Display items in panels (a, b) have the same meanings as in **Figure 2**, except that cyan shading indicates consensus identification; in panel (c), vertical lines indicate mutation sites.

Figure 3 continued on next page

Figure 3 continued

The online version of this article includes the following source data for figure 3:

Source data 1. Source data for **Figure 3**.

structures) prevented the measurement of CSPs. DIRseq predicts D₅₉WSFY₆₃ as drug-interacting residues. Combined with the mutational data of [Woods et al., 2011](#), we propose the expanded motif, D₅₉WSFYLLYY₆₇, as the major drug-binding site (**Figure 3c**, cyan shading). DIRseq also predicts an additional motif, K₉₄WDR₉₇, as drug-interacting residues.

The aggregation of human islet amyloid polypeptide (hIAPP; 37 residues) is inhibited by the small molecule YX-I-1 ([Xu et al., 2022](#)). CSPs elicited by this molecule were small (**Figure 4a**). In addition, CSPs of short IDPs may not exhibit strong disparities because amino acids may be too well mixed along the sequence or drug binding may induce a conformational shift. We thus reduce the threshold for identifying drug-interacting residues to $m+1.0$ SD when the number of residues is ≤ 50 . With this threshold, three residues are identified by CSP as drug-interacting residues: R₁₁ and V₁₇H₁₈. In comparison, DIRseq identifies T₉QLA₁₃ and F₁₅ as drug-interacting residues, which partially overlap with the CSP identifications. Combining the two types of data, we propose T₉QLANFLVH₁₈ as the primary drug-binding site (**Figure 4a**, cyan shading). We note that this motif is also prone to α -helix formation ([Apostolidou et al., 2008](#)).

[Ono et al., 2012](#) acquired ¹H-¹⁵N heteronuclear single quantum coherence spectra of the 42-residue amyloid- β (A β 42) in the absence and presence of the oligomerization-blocking compound myricetin. CSPs were not calculated, but chemical shift movements were most pronounced at R₅, V₁₂H₁₃, K₁₆LVF-FAED₂₃, and I₃₁I₃₂ (vertical lines in **Figure 4b**). In addition, NMR cross peaks suffered broadening upon ligand binding at four of these residues: R₅, V₁₂, K₁₆, and V₁₈ (dark vertical lines in **Figure 4b**), implicating elevated probabilities of ligand interactions. DIRseq predicts E₃FRH₆ and H₁₃H₁₄ as drug-interacting residues. Combined with the NMR data, we propose E₃FRH₆ and V₁₂HHQKLV₁₈ as the primary ligand-binding sites (**Figure 4b**, cyan shading). A β 42 CSPs elicited by several other compounds were also widely distributed over the sequence, such that [Iwaya et al., 2020](#) ‘failed to identify’ drug-interacting residues, implicating a conformational shift.

[Hammoudeh et al., 2009](#) identified three distinct drug binding sites in a C-terminal IDR (residues 363–412) of the oncoprotein c-Myc. This region is disordered on its own but forms a helix-loop-helix structure upon heterodimerization with Max. These authors measured the CSPs of three overlapping fragments (residues 363–381, 370–409, and 402–412), each in the presence of a single compound and of the full IDR in the presence of all three compounds. Using 20 parts per billion (ppb) as the threshold, Hammoudeh et al. named residues R₃₆₆RNELKRSFF₃₇₅, F₃₇₅ALRDIPELE₃₈₅, and Y₄₀₂ILSVQAE₄₀₉ as the binding sites for 10074-G5, 10074-A4, and 10058-F4, respectively. We present their CSP data for the full IDR in **Figure 4c**. Using the $m+1.0$ SD threshold, only six residues are identified as drug-interacting residues: R₃₆₇, L₃₇₀, F₃₇₄F₃₇₅, A₃₉₉, and S₄₀₅. DIRseq predicts two motifs, E₃₆₃RQRR₃₆₇ and K₃₇₁RSFFA₃₇₆, that overlap with the first two sites identified by CSP; a third motif, K₃₉₇KATAY₄₀₂, that corresponds to the third CSP-identified site has moderate drug-interacting propensities. Combining CSP and DIRseq, we revise the three drug-binding sites to be E₃₆₃RQRRNELKRSFF₃₇₅, S₃₇₃FFALRDI₃₈₁, and K₃₉₇KATAYILS₄₀₅ (**Figure 4c**, cyan, olive, and yellow shading, respectively). In addition to 10074-G5, 10074-A4, and 10058-F4, many compounds bind to these three sites ([Yu et al., 2016](#); [Han et al., 2019](#); [Shirey et al., 2021](#); [Li et al., 2024](#)).

Discussion

We have presented the first sequence-based method, DIRseq, to predict drug-interacting residues of IDPs. Assessment against NMR CSP demonstrates the accuracy of DIRseq. Drug-binding motifs are anchored on one or more aromatic residues for forming π - π interactions with drug-like molecules that are rich in aromatic groups. The success of DIRseq comes without any specific information on the drug molecules, suggesting that IDPs may have a relatively simple sequence code for drug binding.

The notion that drug-interacting residues may be agnostic to the molecular details of drug compounds is supported by the fact that the same drug can bind to different IDPs. For example, in addition to p53 ([Zhao et al., 2021](#); **Figure 2c**), the polyphenol EGCG also binds to many other IDPs, including α -synuclein ([Ehrnhoefer et al., 2008](#)), hIAPP ([Young et al., 2014](#); [Meng et al., 2010](#)), A β 40

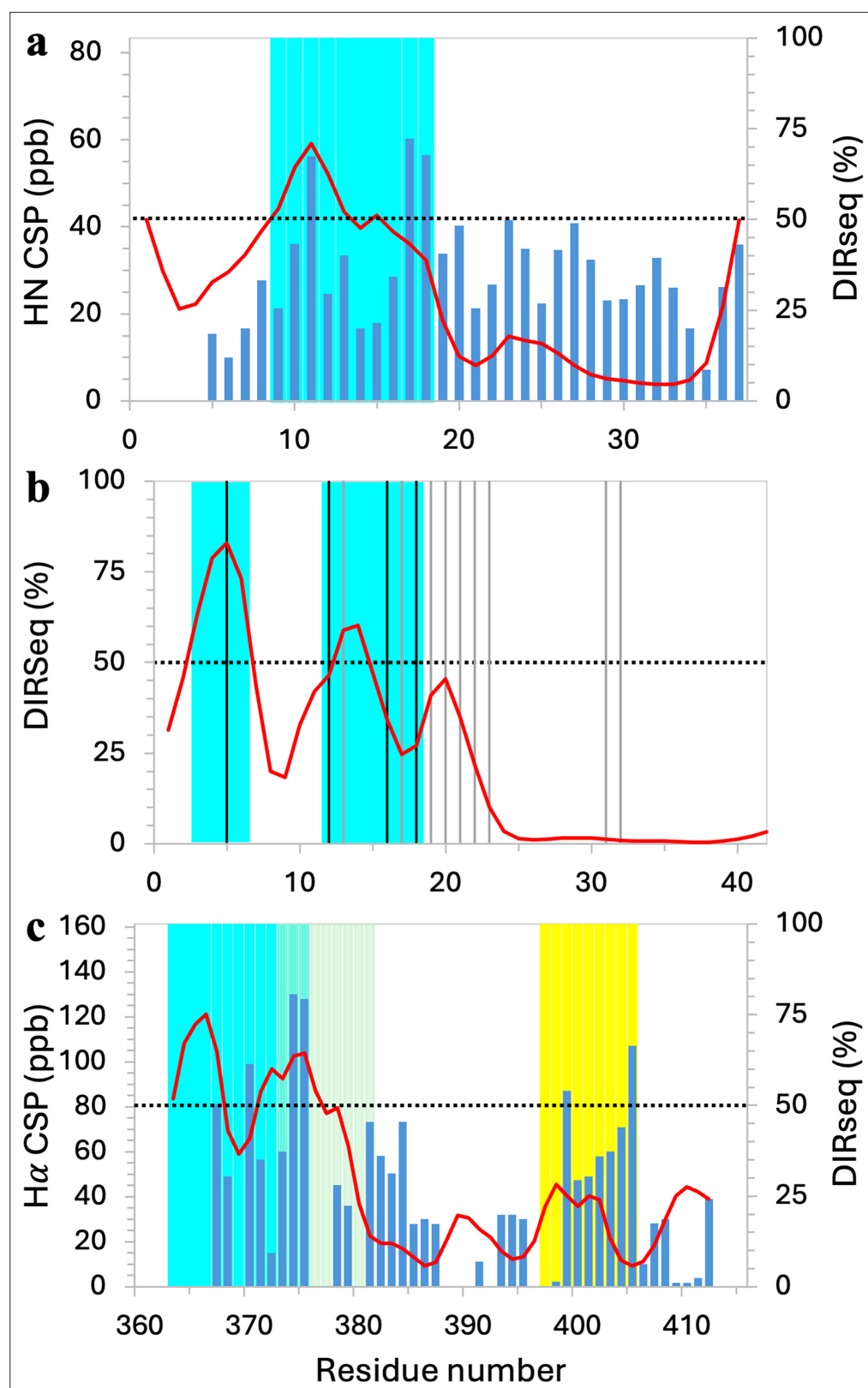


Figure 4. Drug-binding sites identified by combining chemical shift perturbation (CSP) or mutation data with DIRseq predictions. (a) hIAPP. (b) Aβ42. (c) c-Myc. Display items have the same meanings as in **Figure 2**, but with the following exceptions. (1) In panel (b), vertical lines indicate residues with prominent CSPs; those accompanied by NMR peak broadening have their vertical lines in dark color. (2) In panel (c), three CSP-identified

Figure 4 continued on next page

Figure 4 continued

drug-interacting regions are indicated by cyan, olive, and yellow shading. (3) The threshold for identifying drug-interacting residues is lowered to $m+1.0$ SD.

The online version of this article includes the following source data for figure 4:

Source data 1. Source data for **Figure 4**.

(Ahmed et al., 2017) and A β 42 (Ehrnhoefer et al., 2008), tau (Wobst et al., 2015), and merozoite surface protein 2 (Chandrashekar et al., 2010). Likewise, 10074-G5 binds to c-Myc (Hammoudeh et al., 2009; Figure 4c) but also to A β 42 (Heller et al., 2020). On the other hand, c-Myc represents a case where different compounds bind to distinct sites on a single IDP (Hammoudeh et al., 2009). A related example is presented by p27, where SJ403 typifies a group of compounds that share the same three binding sites (Figure 2a). Another group of compounds, typified by SJ710, binds only to the third site. Chemically, the presence of nitrogen atoms in the rings of SJ403 enhances its aromaticity and thus strengthens π - π interactions; in addition, the electronegative groups of SJ403 project into different directions, making it less restricted when forming electrostatic interactions (Figure 2—figure supplement 3). These features may explain why SJ403 can bind to all three sites, whereas SJ710 can bind only to the third site, F₈₇YYR₉₀, where three consecutive aromatic residues followed by a basic residue ensure that SJ710 can form both π - π and electrostatic interactions. When more data for multiple drugs binding to a single IDP becomes available, it will be important to use such data to train the next generation of DIRseq where the parameters are drug-specific. As a simple example, the number of residues that can simultaneously bind a drug molecule may grow with the latter's size; this dependence can be modeled by making the parameter b dependent on drug molecule size. The drug molecules studied in the present work have molecular weights of 360 ± 130 Da and thus span a relatively narrow size range.

We have illustrated the combination of DIRseq with NMR CSP to make robust identifications of drug-binding sites in IDPs. Indeed, CSP, MD simulations, and DIRseq are three orthogonal approaches that have great potential in complementing each other, not only for identifying drug-binding sites but also for elucidating the roles of amino acids, their sequence context, and different types of noncovalent

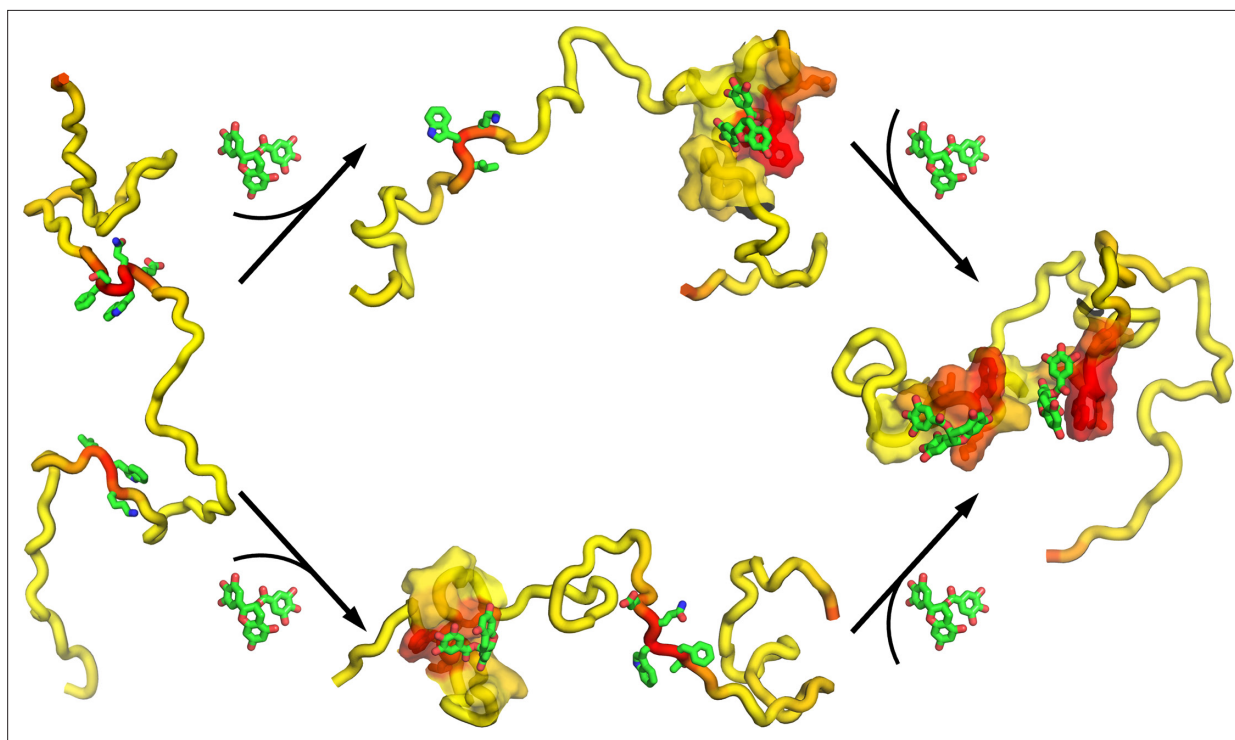


Figure 5. Poses of p53-bound EGCG generated by docking.

interactions in forming such sites. DIRseq offers fast speed and a simple, direct link between sequence motif and propensity for drug binding.

Another application of DIRseq is to define IDP fragments for in-depth study by MD simulations, as shorter constructs both enable the use of a smaller simulation box and reduce the size of the conformational space. For example, based on CSP data and initial MD simulations of full-length α -synuclein, [Robustelli et al., 2022](#) chose a 20-residue C-terminal fragment for simulations of binding with additional compounds, leading to the identification of Ligand 47 as a stronger binder than the original Fasudil. Similarly, based on CSP data from longer constructs of the androgen receptor, both [Zhu et al., 2022](#) and [Basu et al., 2023](#) chose the 56-residue R2-R3 fragment for MD simulations of drug binding. DIRseq can now play a similar role in selecting fragments for MD simulations when CSP data are unavailable. Longer constructs may also present challenges such as resonance assignments to NMR experiments, so well-chosen fragments guided by DIRseq can also benefit NMR studies.

Lastly, virtual screening has been conducted against conformational ensembles of IDPs ([Ruan et al., 2021](#); [Dhar et al., 2025](#)); drug-binding sites predicted by DIRseq can be used to guide such screening. As a simple illustration, we present poses of EGCG generated by screening against the two DIRseq-predicted binding sites in p53 in [Figure 5](#). As IDPs sample a vast conformational space, knowledge of the binding site can drastically reduce the computational cost. The subset of conformations that generate high docking scores for a given drug at the known site can also provide insight into the mechanism of drug action.

Methods

The sequences of the IDP studied here and the drugs that bind to them are listed in [Supplementary file 1A](#). All DIRseq predictions were obtained using the web server at <https://zhougroupp-uic.github.io/DIRseq/>. Conformations of IDPs were generated using the TraDES method ([Feldman and Hogue, 2002](#)).

Docking of EGCG onto p53 was performed via the SwissDock web server at <https://www.swiss-dock.ch/> ([Bugnon et al., 2024](#)) utilizing the Autodock Vina docking engine ([Eberhardt et al., 2021](#)). The SMILES string for EGCG from [PubChem](#) (CID 65064) and several conformations of p53 were used as input. A cubic region (13–20 Å in side length) around the center of each drug-interacting residue was selected for docking.

Acknowledgements

This work was supported by National Institutes of Health Grant GM118091.

Additional information

Funding

Funder	Grant reference number	Author
National Institute of General Medical Sciences	GM118091	Huan-Xiang Zhou

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Matt MacAinsh, Data curation, Formal analysis, Validation, Investigation, Methodology; Sanbo Qin, Software, Visualization; Huan-Xiang Zhou, Conceptualization, Data curation, Formal analysis, Supervision, Funding acquisition, Investigation, Visualization, Methodology, Writing – original draft, Project administration, Writing – review and editing

Author ORCIDs

Huan-Xiang Zhou  <https://orcid.org/0000-0001-9020-0302>

Peer review material

Reviewer #1 (Public review): <https://doi.org/10.7554/eLife.107470.3.sa1>

Reviewer #2 (Public review): <https://doi.org/10.7554/eLife.107470.3.sa2>

Author response <https://doi.org/10.7554/eLife.107470.3.sa3>

Additional files

Supplementary files

Supplementary file 1. Supplementary tables.

MDAR checklist

Data availability

Figure 2-source data 1, Figure 3-source data 1, Figure 4-source data 1 contain the numerical data used to generate the figures. The source code for DIRseq can be downloaded at <https://github.com/hzhou43/DIRseq/> with file name DIRseq.js (copy archived at [Zhou, 2025](#)).

References

- Ahmed R, VanSchouwen B, Jafari N, Ni X, Ortega J, Melacini G. 2017. Molecular mechanism for the (-)-Epigallocatechin gallate-induced toxic to nontoxic remodeling of A β oligomers. *Journal of the American Chemical Society* **139**:13720–13734. DOI: <https://doi.org/10.1021/jacs.7b05012>, PMID: 28841302
- Apostolidou M, Jayasinghe SA, Langen R. 2008. Structure of alpha-helical membrane-bound human islet amyloid polypeptide and its implications for membrane-mediated misfolding. *The Journal of Biological Chemistry* **283**:17205–17210. DOI: <https://doi.org/10.1074/jbc.M801383200>, PMID: 18442979
- Basu S, Martínez-Cristóbal P, Frigolé-Vivas M, Pesarrodonna M, Lewis M, Szulc E, Bañuelos CA, Sánchez-Zarzalejo C, Bielskutė S, Zhu J, Pombo-García K, Garcia-Cabau C, Zodi L, Dockx H, Smak J, Kaur H, Batlle C, Mateos B, Biesaga M, Escobedo A, et al. 2023. Rational optimization of a transcription factor activation domain inhibitor. *Nature Structural & Molecular Biology* **30**:1958–1969. DOI: <https://doi.org/10.1038/s41594-023-01159-5>, PMID: 38049566
- Bugnon M, Röhrig UF, Goullieux M, Perez MAS, Daina A, Michielin O, Zoete V. 2024. SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. *Nucleic Acids Research* **52**:W324–W332. DOI: <https://doi.org/10.1093/nar/gkac300>, PMID: 38686803
- Chandrashekar IR, Adda CG, MacRaild CA, Anders RF, Norton RS. 2010. Inhibition by flavonoids of amyloid-like fibril formation by *Plasmodium falciparum* merozoite surface protein 2. *Biochemistry* **49**:5899–5908. DOI: <https://doi.org/10.1021/bi902197x>, PMID: 20545323
- De Mol E, Fenwick RB, Phang CTW, Buzón V, Szulc E, de la Fuente A, Escobedo A, García J, Bertoncini CW, Estébanez-Perpiñá E, McEwan IJ, Riera A, Salvatella X. 2016. EPI-001, a compound active against castration-resistant prostate cancer, targets transactivation unit 5 of the androgen receptor. *ACS Chemical Biology* **11**:2499–2505. DOI: <https://doi.org/10.1021/acschembio.6b00182>, PMID: 27356095
- Dey S, MacAinsh M, Zhou HX. 2022. Sequence-dependent backbone dynamics of intrinsically disordered proteins. *Journal of Chemical Theory and Computation* **18**:6310–6323. DOI: <https://doi.org/10.1021/acs.jctc.2c00328>, PMID: 36084347
- Dhar A, Sisk TR, Robustelli P. 2025. Ensemble docking for intrinsically disordered proteins. *bioRxiv*. DOI: <https://doi.org/10.1101/2025.1101.1123.634614>
- Eberhardt J, Santos-Martins D, Tillack AF, Forli S. 2021. AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. *Journal of Chemical Information and Modeling* **61**:3891–3898. DOI: <https://doi.org/10.1021/acs.jcim.1c00203>, PMID: 34278794
- Ehrnhoefer DE, Bieschke J, Boeddrich A, Herbst M, Masino L, Lurz R, Engemann S, Pastore A, Wanker EE. 2008. EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nature Structural & Molecular Biology* **15**:558–566. DOI: <https://doi.org/10.1038/nsmb.1437>, PMID: 18511942
- Feldman HJ, Hogue CWV. 2002. Probabilistic sampling of protein conformations: new hope for brute force? *Proteins* **46**:8–23. PMID: 11746699.
- Hammoudeh DI, Follis AV, Prochownik EV, Metallo SJ. 2009. Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *Journal of the American Chemical Society* **131**:7390–7401. DOI: <https://doi.org/10.1021/ja900616b>, PMID: 19432426
- Han H, Jain AD, Truica MI, Izquierdo-Ferrer J, Anker JF, Lysy B, Sagar V, Luan Y, Chalmers ZR, Unno K, Mok H, VataPELLI R, Yoo YA, Rodríguez Y, Kandela I, Parker JB, Chakravarti D, Mishra RK, Schiltz GE, Abdulkadir SA. 2019. Small-molecule MYC inhibitors suppress tumor growth and enhance immunotherapy. *Cancer Cell* **36**:483–497. DOI: <https://doi.org/10.1016/j.ccell.2019.10.001>, PMID: 31679823
- Heller GT, Aprile FA, Michaels TCT, Limbocker R, Perni M, Ruggeri FS, Mannini B, Löhr T, Bonomi M, Camilloni C, De Simone A, Felli IC, Pierattelli R, Knowles TPJ, Dobson CM, Vendruscolo M. 2020. Small-molecule sequestration of amyloid- β as a drug discovery strategy for Alzheimer's disease. *Science Advances* **6**:eabb5924. DOI: <https://doi.org/10.1126/sciadv.abb5924>, PMID: 33148639

- Heller GT**, Shukla VK, Figueiredo AM, Hansen DF. 2024. Picosecond dynamics of a small molecule in its bound state with an intrinsically disordered protein. *Journal of the American Chemical Society* **146**:2319–2324. DOI: <https://doi.org/10.1021/jacs.3c11614>, PMID: 38251829
- Herrera-Nieto P**, Pérez A, De Fabritiis G. 2020. Small molecule modulation of intrinsically disordered proteins using molecular dynamics simulations. *Journal of Chemical Information and Modeling* **60**:5003–5010. DOI: <https://doi.org/10.1021/acs.jcim.0c00381>, PMID: 32786705
- Iconaru LI**, Ban D, Bharatham K, Ramanathan A, Zhang W, Shelat AA, Zuo J, Kriwacki RW. 2015. Discovery of small molecules that inhibit the disordered protein, p27(Kip1). *Scientific Reports* **5**:15686. DOI: <https://doi.org/10.1038/srep15686>, PMID: 26507530
- Iruela G**, Fernández A, Sagar A, Carvajal FJ, Bernadó P, Pons M. 2022. A FRET-based biosensor for the Src N-terminal regulatory element. *Biosensors* **12**:96. DOI: <https://doi.org/10.3390/bios12020096>, PMID: 35200356
- Iwaya N**, Goda N, Matsuzaki M, Narita A, Shigemitsu Y, Tenno T, Abe Y, Hoshi M, Hiroaki H. 2020. Principal component analysis of data from NMR titration experiment of uniformly ¹⁵N labeled amyloid beta (1–42) peptide with osmolytes and phenolic compounds. *Archives of Biochemistry and Biophysics* **690**:108446. DOI: <https://doi.org/10.1016/j.abb.2020.108446>, PMID: 32593678
- Jin F**, Yu C, Lai L, Liu Z. 2013. Ligand clouds around protein clouds: a scenario of ligand binding with intrinsically disordered proteins. *PLOS Computational Biology* **9**:e1003249. DOI: <https://doi.org/10.1371/journal.pcbi.1003249>, PMID: 24098099
- Joshi P**, Vendruscolo M. 2015. In *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-20164-1_13
- Klein-Seetharaman J**, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H. 2002. Long-range interactions within a nonnative protein. *Science* **295**:1719–1722. DOI: <https://doi.org/10.1126/science.1067680>, PMID: 11872841
- Kocis P**, Tolar M, Yu J, Sinko W, Ray S, Blennow K, Fillit H, Hey JA. 2017. Elucidating the Aβ42 anti-aggregation mechanism of action of tramiprosate in Alzheimer's Disease: integrating molecular analytical methods, pharmacokinetic and clinical data. *CNS Drugs* **31**:495–509. DOI: <https://doi.org/10.1007/s40263-017-0434-z>, PMID: 28435985
- Krishnan N**, Koveal D, Miller DH, Xue B, Akshinthala SD, Kragelj J, Jensen MR, Gauss C-M, Page R, Blackledge M, Muthuswamy SK, Peti W, Tonks NK. 2014. Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nature Chemical Biology* **10**:558–566. DOI: <https://doi.org/10.1038/nchembio.1528>, PMID: 24845231
- Lama D**, Vosselman T, Sahin C, Liaño-Pons J, Cerrato CP, Nilsson L, Teilum K, Lane DP, Landreh M, Arsenian Henriksson M. 2024. A druggable conformational switch in the c-MYC transactivation domain. *Nature Communications* **15**:1865. DOI: <https://doi.org/10.1038/s41467-024-45826-7>, PMID: 38424045
- Lamberto GR**, Binolfi A, Orcellet ML, Bertocini CW, Zweckstetter M, Griesinger C, Fernández CO. 2009. Structural and mechanistic basis behind the inhibitory interaction of PcTS on alpha-synuclein amyloid fibril formation. *PNAS* **106**:21057–21062. DOI: <https://doi.org/10.1073/pnas.0902603106>, PMID: 19948969
- Li Z**, Huang Y, Hung TI, Sun J, Aispuro D, Chen B, Guevara N, Ji F, Cong X, Zhu L, Wang S, Guo Z, Chang C, Xue M. 2024. MYC-targeting inhibitors generated from a stereodiversified bicyclic peptide library. *Journal of the American Chemical Society* **146**:1356–1363. DOI: <https://doi.org/10.1021/jacs.3c09615>
- Martin EW**, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, Grace CR, Soranno A, Pappu RV, Mittag T. 2020. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**:694–699. DOI: <https://doi.org/10.1126/science.aaw8653>, PMID: 32029630
- Mehta T**, Goyal B. 2024. Unveiling molecular mechanism underlying inhibition of human islet amyloid polypeptide fibrillation by benzene carboxylic acid-peptide conjugate. *Journal of Molecular Liquids* **416**:126426. DOI: <https://doi.org/10.1016/j.molliq.2024.126426>
- Meng F**, Abedini A, Plesner A, Verchere CB, Raleigh DP. 2010. The flavanol (-)-epigallocatechin 3-gallate inhibits amyloid formation by islet amyloid polypeptide, disaggregates amyloid fibrils, and protects cultured cells against IAPP-induced toxicity. *Biochemistry* **49**:8127–8133. DOI: <https://doi.org/10.1021/bi100939a>, PMID: 20707388
- Muhammedkutty FNK**, MacAinsh M, Zhou HX. 2025. Atomistic molecular dynamics simulations of intrinsically disordered proteins. *Current Opinion in Structural Biology* **92**:103029. DOI: <https://doi.org/10.1016/j.sbi.2025.103029>, PMID: 40068541
- Neira JL**, Bintz J, Arruebo M, Rizzuti B, Bonacci T, Vega S, Lanás A, Velázquez-Campoy A, Iovanna JL, Abián O. 2017. Identification of a drug targeting an intrinsically disordered protein involved in pancreatic adenocarcinoma. *Scientific Reports* **7**:39732. DOI: <https://doi.org/10.1038/srep39732>, PMID: 28054562
- Ono K**, Li L, Takamura Y, Yoshiike Y, Zhu L, Han F, Mao X, Ikeda T, Takasaki J, Nishijo H, Takashima A, Teplow DB, Zagorski MG, Yamada M. 2012. Phenolic compounds prevent amyloid β-protein oligomerization and synaptic dysfunction by site-specific binding. *The Journal of Biological Chemistry* **287**:14631–14643. DOI: <https://doi.org/10.1074/jbc.M111.325456>, PMID: 22393064
- Peterson JR**, Bickford LC, Morgan D, Kim AS, Ouerfelli O, Kirschner MW, Rosen MK. 2004. Chemical inhibition of N-WASP by stabilization of a native autoinhibited conformation. *Nature Structural & Molecular Biology* **11**:747–755. DOI: <https://doi.org/10.1038/nsmb796>, PMID: 15235593
- Qin S**, Zhou HX. 2024. Predicting the sequence-dependent backbone dynamics of intrinsically disordered proteins. *eLife* **12**:RP88958. DOI: <https://doi.org/10.7554/eLife.88958>, PMID: 39475380

- Robustelli P**, Ibanez-de-Opakua A, Campbell-Bezat C, Giordanetto F, Becker S, Zweckstetter M, Pan AC, Shaw DE. 2022. Molecular basis of small-molecule binding to α -synuclein. *Journal of the American Chemical Society* **144**:2501–2510. DOI: <https://doi.org/10.1021/jacs.1c07591>, PMID: 35130691
- Ruan H**, Yu C, Niu X, Zhang W, Liu H, Chen L, Xiong R, Sun Q, Jin C, Liu Y, Lai L. 2021. Computational strategy for intrinsically disordered protein ligand design leads to the discovery of p53 transactivation domain I binding compounds that activate the p53 pathway. *Chemical Science* **12**:3004–3016. DOI: <https://doi.org/10.1039/D0SC04670A>
- Saurabh S**, Nadendla K, Purohit SS, Sivakumar PM, Cetinel S. 2023. Fuzzy drug targets: disordered proteins in the drug-discovery realm. *ACS Omega* **8**:9729–9747. DOI: <https://doi.org/10.1021/acsomega.2c07708>, PMID: 36969402
- Shirey RJ**, Hart JR, Sridharan B, Novick SJ, Turner LD, Zhou B, Nielsen AL, Eubanks LM, Ueno L, Hixon MS, Lairson LL, Spicer TP, Scampavia LD, Griffin PR, Vogt PK, Janda KD. 2021. Synthetic fluorescent MYC probe: Inhibitor binding site elucidation and development of a high-throughput screening assay. *Bioorganic & Medicinal Chemistry* **42**:116246. DOI: <https://doi.org/10.1016/j.bmc.2021.116246>, PMID: 34130216
- Tatenhorst L**, Eckermann K, Dambeck V, Fonseca-Ornelas L, Walle H, Lopes da Fonseca T, Koch JC, Becker S, Tönges L, Bähr M, Outeiro TF, Zweckstetter M, Lingor P. 2016. Fasudil attenuates aggregation of α -synuclein in models of Parkinson's disease. *Acta Neuropathologica Communications* **4**:39. DOI: <https://doi.org/10.1186/s40478-016-0310-y>, PMID: 27101974
- Tesei G**, Schulze TK, Crehuet R, Lindorff-Larsen K. 2021. Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *PNAS* **118**:e2111696118. DOI: <https://doi.org/10.1073/pnas.2111696118>, PMID: 34716273
- Tesei G**, Lindorff-Larsen K. 2022. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Research Europe* **2**:94. DOI: <https://doi.org/10.12688/openreseurope.14967.2>, PMID: 37645312
- Uversky VN**. 2024. How to drug a cloud? Targeting intrinsically disordered proteins. *Pharmacological Reviews* **1**:PHARMREV-AR-2023-001113. DOI: <https://doi.org/10.1124/pharmrev.124.001113>, PMID: 39433443
- Wang G**, Zhu X, Song X, Zhang Q, Qian Z. 2022. Melatonin inhibits hIAPP oligomerization by preventing β -sheet and hydrogen bond formation of the amyloidogenic region revealed by replica-exchange molecular dynamics simulation. *International Journal of Molecular Sciences* **23**:10264. DOI: <https://doi.org/10.3390/ijms231810264>
- Wang H**, Xiong R, Lai L. 2023. Rational drug design targeting intrinsically disordered proteins. *WIREs Computational Molecular Science* **13**:e1685. DOI: <https://doi.org/10.1002/wcms.1685>
- Wobst HJ**, Sharma A, Diamond MI, Wanker EE, Bieschke J. 2015. The green tea polyphenol (-)-epigallocatechin gallate prevents the aggregation of tau protein into toxic oligomers at substoichiometric ratios. *FEBS Letters* **589**:77–83. DOI: <https://doi.org/10.1016/j.febslet.2014.11.026>, PMID: 25436420
- Woods LA**, Platt GW, Hellewell AL, Hewitt EW, Homans SW, Ashcroft AE, Radford SE. 2011. Ligand binding to distinct states diverts aggregation of an amyloid-forming protein. *Nature Chemical Biology* **7**:730–739. DOI: <https://doi.org/10.1038/nchembio.635>, PMID: 21873994
- Xu Y**, Maya-Martinez R, Guthertz N, Heath GR, Manfield IW, Breeze AL, Sobott F, Foster R, Radford SE. 2022. Tuning the rate of aggregation of hIAPP into amyloid using small-molecule modulators of assembly. *Nature Communications* **13**:1040. DOI: <https://doi.org/10.1038/s41467-022-28660-7>
- Young LM**, Cao P, Raleigh DP, Ashcroft AE, Radford SE. 2014. Ion mobility spectrometry-mass spectrometry defines the oligomeric intermediates in amylin amyloid formation and the mode of action of inhibitors. *Journal of the American Chemical Society* **136**:660–670. DOI: <https://doi.org/10.1021/ja406831n>, PMID: 24372466
- Yu C**, Niu X, Jin F, Liu Z, Jin C, Lai L. 2016. Structure-based inhibitor design for the intrinsically disordered protein c-Myc. *Scientific Reports* **6**:22298. DOI: <https://doi.org/10.1038/srep22298>
- Zhao J**, Blayney A, Liu X, Gandy L, Jin W, Yan L, Ha JH, Canning AJ, Connelly M, Yang C, Liu X, Xiao Y, Cosgrove MS, Solmaz SR, Zhang Y, Ban D, Chen J, Loh SN, Wang C. 2021. EGCG binds intrinsically disordered N-terminal domain of p53 and disrupts p53-MDM2 interaction. *Nature Communications* **12**:986. DOI: <https://doi.org/10.1038/s41467-021-21258-5>
- Zhou HX**, Kota D, Qin S, Prasad R. 2024. Fundamental aspects of phase-separated biomolecular condensates. *Chemical Reviews* **124**:8550–8595. DOI: <https://doi.org/10.1021/acs.chemrev.4c00138>, PMID: 38885177
- Zhou HX**. 2025. DIRseq. swlh:1:rev:85fd0bff563eab074f7e2ea9b72812bd8867dacf. Software Heritage. <https://archive.softwareheritage.org/swlh:1:dir:6f314b8037f16621af9b15b4d4fe6b43f7f305e7;origin=https://github.com/hzhou43/DIRseq;visit=swlh:1:snp:fa4506d032107aa7950d2e1ec595a0e29b3f0f8e;anchor=swlh:1:rev:85fd0bff563eab074f7e2ea9b72812bd8867dacf>
- Zhu J**, Salvatella X, Robustelli P. 2022. Small molecules targeting the disordered transactivation domain of the androgen receptor induce the formation of collapsed helical states. *Nature Communications* **13**:6390. DOI: <https://doi.org/10.1038/s41467-022-34077-z>