Structural Biology and Molecular Biophysics

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

Reviewed Preprint
v1 • August 14, 2025
Not revised

Matthew MacAinsh, Sanbo Qin, Huan-Xiang Zhou

Department of Chemistry, University of Illinois Chicago, Chicago, United States • Department of Physics, University of Illinois Chicago, Chicago, United States

https://en.wikipedia.org/wiki/Open_access
Copyright information

**eLife Assessment**

This study presents a sequence-based method for predicting drug-interacting residues in intrinsically disordered proteins (IDPs), addressing an **important** challenge in understanding small-molecule:IDP interactions. The findings have **solid** support in illustrative examples that underscore the role of aromatic interactions. While predicted binding sites remain coarse, validation was done on a total of 10 IDPs, four of which thoroughly and six others less so. The method builds on previous work from the authors, with necessarily ad hoc modifications, and offers a starting point for further exploration in this emerging field.

https://doi.org/10.7554/eLife.107470.1.sa3

## Abstract

Intrinsically disordered proteins (IDPs) are now well-recognized as drug targets. Identifying drug-interacting residues (DIRs) 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 DIRs 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 DIRs identified by NMR chemical shift perturbation and other methods, including residues $L_{22}WK_{24}$ and $Q_{52}WFT_{55}$ in the tumor repressor protein p53. These successes augur well for deciphering the sequence code for IDP-drug binding. DIRseq is available as a web server at https://zhougroup-uic.github.io/DIRseq/ and has many applications, such as virtual screening against IDPs and designing IDP fragments for in-depth NMR and MD studies.

## Introduction

Intrinsically disordered proteins (IDPs) are now recognized as important drug targets [1-4]. For structured protein targets, a crucial step is identifying the drug-binding pocket. Although an IDP can be locked into a specific conformation by a drug molecule in rare cases [5], the prevailing scenario is that the protein remains disordered upon drug binding. 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. Such drug-interacting residues (DIRs), akin to binding pockets in structured proteins, are key to optimizing compounds and elucidating the mechanism of action.

NMR chemical shift perturbation (CSP) is the best experimental method for identifying DIRs and has been applied to many IDPs [6-23]. Aromatic residues are frequently among DIRs [7,10,12,17,18,20,22]. This is understandable as drug molecules typically are rich in aromatic rings (**Figure 1** and Table S1), which can form π-π interactions with aromatic residues. As recently pointed out by Heller et al. [23], 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 DIRs. 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 [11,16,20], so CSPs spread from the directly interacting residues to the rest of the IDP, adding to the difficulty in identifying DIRs.

All-atom molecular dynamics (MD) simulations have presented atomic details in many IDP-drug systems [17,18,20,22,24-29]. These simulation studies have highlighted the frequent engagement of drug molecules with aromatic residues, particularly in π-π interactions [17,18,20,22,26-29]. MD simulations also revealed the emergence of a compact sub-population of the N-terminal disordered region of the tumor repressor protein p53 upon binding the drug epigallocatechin-3-gallate (EGCG) [18] and reduced solvent exposure of hydrophobic residues in the 42-residue amyloid-β peptide (Aβ42) upon binding the drug 10074-G5 [25]. 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 DIRs [30].

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 DIRs seem to overlap with residues exhibiting elevated transverse relaxation rates ($R_2$) [31,32], as exemplified by the C-terminal 20 residues of α-synuclein [20]. Elevated $R_2$ in IDPs is caused by either local inter-residue interactions or residual secondary structure [33]. Based on this understanding, we developed a sequence-based method, SeqDYN, to predict $R_2$ of IDPs [31]. As suggested previously [32], 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 DIRs identified by NMR chemical shift perturbation and other methods, including residues $L_{22}WK_{24}$ and $Q_{52}WFT_{55}$ 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.

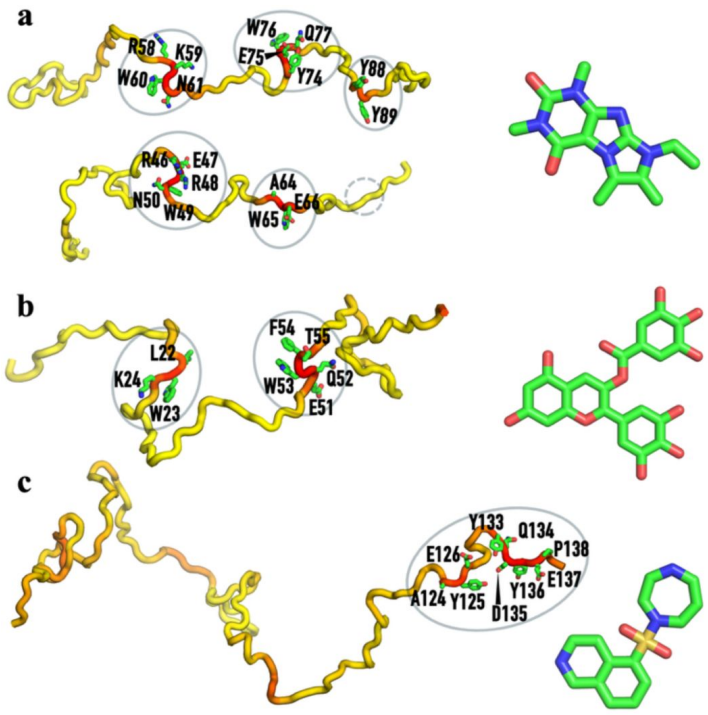


**Figure 1.**

**Four IDPs and drugs that bind to them.**

(a) p27, p21, and SJ403. (b) p53 and EGCG. (c) -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 DIRs are shown with sidechains rendered in stick.

# Results

## Retooling of SeqDYN into DIRseq

SeqDYN [31] predicts the $R_2$ value of residue $n$ by accumulating a contributing factor $f(i; n)$ from each residue $i$:

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

where $N$ is the total number of residues in the IDP and Y 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 $n$:

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

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 S1a) show that aromatic (Trp, Tyr, His, and Phe), Arg, and long aliphatic (Ile and Leu) amino acids are interaction-prone and tend to elevate $R_2$. 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 raised the $q$ value of Asp to be the same as that of Glu, to increase the role of Asp's electrostatic interactions. The modified set of $q$ parameters is displayed in Figure S1b.

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 to 4 residues at a time. Third, the SeqDYN-predicted $R_2$ profile, capturing experimental observations [34], 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 S1b).

Lastly, we converted high and low $R_2$ values into high and low DIR propensities, respectively, using a sigmoid function

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

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

$$R_{2\text{th}} = m + s_1 \text{SD} \tag{3a}$$

$$R_{2\text{wd}} = m/s_2 \tag{4a}$$

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 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 DIR, 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

DIRs 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 DIRs (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 DIRs, for both CSP and DIRseq; this threshold is indicated by a horizontal dashed line in **Figure 2**. Experimentally identified DIRs 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. [10]. 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 DIRs are $R_{58}KWN_{61}$, $Y_{74}EWQ_{77}$, and $Y_{88}Y_{89}$ (**Figure 1a**, 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 S2a).

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 DIRs (**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 S1b). CSP and DIRseq propensity show a strong correlation ($r = 0.66$; Figure S2b).

NMR CSP revealed two motifs, $W_{23}K_{24}$ and $D_{49}IEQWFT_{55}$, in the N-terminal region of p53 as interacting with EGCG [18] (**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 DIR 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 S2c).

Using the *m* + 1.5 SD threshold, CSPs of Robustelli et al. [20] 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 S2d) 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 $Y_{125}$ and $Y_{136}$ by the latter; their correlation coefficient is 0.78 (Figure S3a). Because DIRseq does not consider any information about the drug

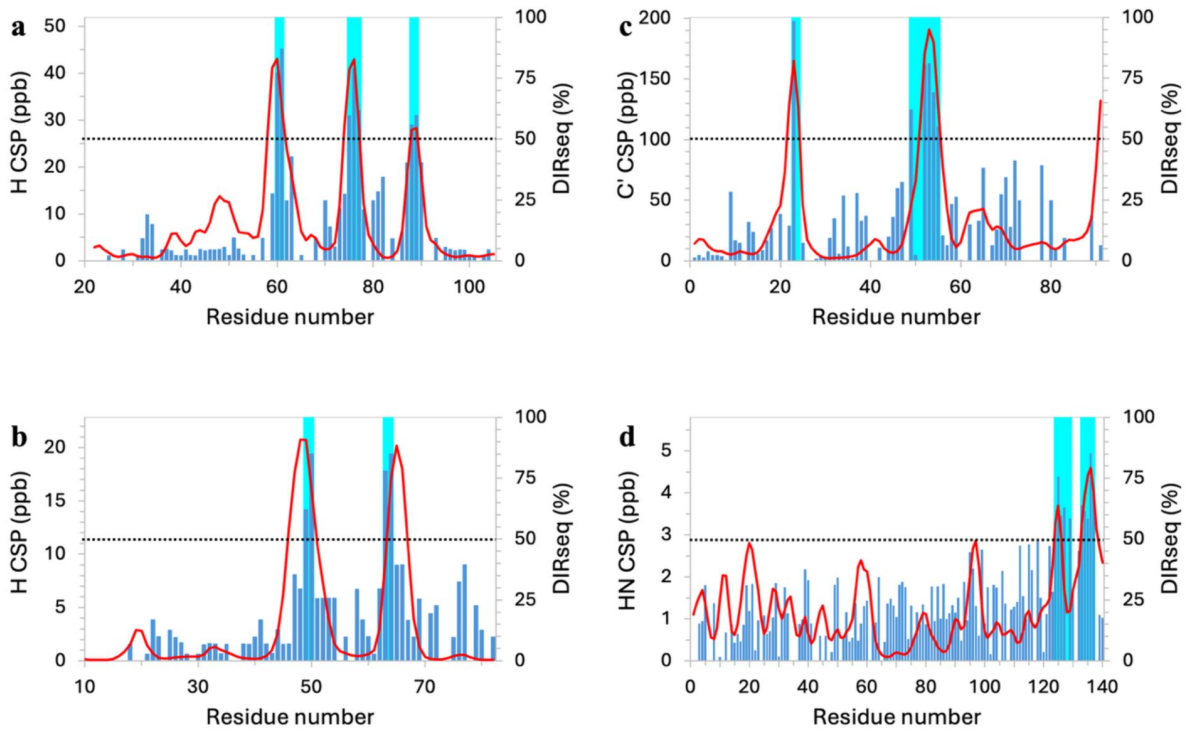


**Figure 2.**

**Comparison of DIRseq propensities with NMR CSPs.**

(a) p27. (b) p21. (c) p53. (d) -synuclein. CSPs are displayed as blue bars and in units of ppb; CSP-identified DIRs 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.

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.

## DIRseq as a complement to CSP for assigning DIRs

After assessing the achievable accuracy of DIRseq, we now consider an application: its combination with CSP to make robust assignments of DIRs. 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 DIRs. However, a consensus identification by CSP and DIRseq may be reliable. MD simulations have played such a complementary role to CSP in several studies [18,20,22,28].

De Mol et al. [11] 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 $^{1}$H NMR spectrum of EPI-001 but the peptides corresponding to R1-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 [28]. Basu et al. [22] 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** and Figure S3b). The contact probabilities of the R2-R3 fragment with 1aa in MD simulations reaffirmed the importance of aromatic residues.

De Mol et al. [11] and Basu et al. [22] were careful not to name any DIRs 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 DIRs in the middle of each of R1-R3: $R_{360}DYY_{363}$, $A_{396}WAA_{399}$, and $W_{433}H_{434}$. In addition, the latter two motifs showed the highest drug-contact probabilities in separate MD simulations of the R2-R3 fragment [22,28]. 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 $A_{396}WAAAAAQ_{403}$ and $W_{433}HTLF_{437}$. The three putative drug-interacting motifs are indicated as cyan shading in **Figure 3a**.

Heller et al. [23] 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 $^{1}$H-$^{15}$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 $P_{310}AWARPD_{316}$ as DIRs. We propose the expanded motif, $A_{308}LPAWARPD_{316}$, as residues interacting with 5-fluorindole (**Figure 3b**, cyan shading). DIRseq also predicts two more motifs, $E_{323}SWRRPDY_{330}$ and $R_{352}RRR_{355}$, as DIRs, which remain to be tested.

The fibrillation of acid-denatured $\beta_2$ microglobulin is inhibited by rifamycin SV [35]. An aromatic-rich motif, $W_{60}SFYLLYYTEF_{70}$, 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 structures) prevented the measurement of CSPs. DIRseq predicts $D_{59}WSFY_{63}$ as DIRs. Combined with the

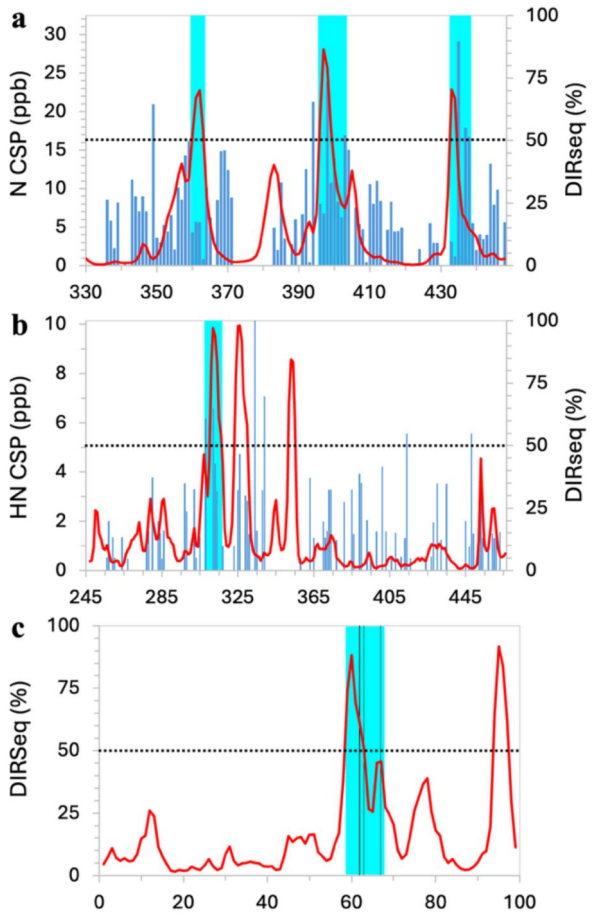


**Figure 3.**

**Drug-binding sites identified by combining 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.

mutational data of Woods et al. [35], we propose the expanded motif, $D_{59}$WSFYLLYY$_{67}$, as the major drug-binding site (**Figure 3c**, cyan shading). DIRseq also predicts an additional motif, $K_{94}$WDR$_{97}$, as DIRs.

The aggregation of human islet amyloid polypeptide (hIAPP; 37 residues) is inhibited by the small molecule YX-I-1 [21]. 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 DIRs to $m$ + 1.0 SD when the number of residues is ≤ 50. With this threshold, three residues are identified by CSP as DIRs: $R_{11}$ and $V_{17}H_{18}$. In comparison, DIRseq identifies $T_9$QRLA$_{13}$ and $F_{15}$ as DIRs, which partially overlap with the CSP identifications. Combining the two types of data, we propose $T_9$QRLANFLVH$_{18}$ as the primary drug-binding site (**Figure 4a**, cyan shading). We note that this motif is also prone to α-helix formation [36].

Ono et al. [8] acquired $^{1}$H-$^{15}$N heteronuclear single quantum coherence spectra of the 42-residue amyloid-β (Aβ42) in the absence and presence of the oligomerization-blocking compound myricetin. CSPs were not calculated, but chemical shift movements were most pronounced at $R_5$, $V_{12}H_{13}$, $K_{16}$LVFFAED$_{23}$, and $I_{31}I_{32}$ (vertical lines in **Figure 4b**). In addition, NMR cross peaks suffered broadening upon ligand binding at four of these residues: $R_5$, $V_{12}$, $K_{16}$, and $V_{18}$ (dark vertical lines in **Figure 4b**), implicating elevated probabilities of ligand interactions. DIRseq predicts $E_3$FRH$_6$ and $H_{13}H_{14}$ as DIRs. Combined with the NMR data, we propose $E_3$FRH$_6$ and $V_{12}$HHQKLV$_{18}$ as the primary ligand-binding sites (**Figure 4b**, cyan shading). Aβ42 CSPs elicited by several other compounds were also widely distributed over the sequence, such that Iwaya et al. [16] "failed to identify" DIRs, implicating a conformational shift.

Hammoudeh et al. [6] identified three distinct drug binding sites in a C-terminal IDR (residues 363-412) of the oncoprotein c-Myc. 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 $R_{366}$RNELKRSFF$_{375}$, $F_{375}$ALRDQIPELE$_{385}$, and $Y_{402}$ILSVQAE$_{409}$ 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 DIRs: $R_{367}$, $L_{370}$, $F_{374}F_{375}$, $A_{399}$, and $S_{405}$. DIRseq predicts two motifs, $E_{363}$RQRR$_{367}$ and $K_{371}$RSFFA$_{376}$, that overlap with the first two sites identified by CSP; a third motif, $K_{397}$KATAY$_{402}$, that corresponds to the third CSP-identified site has moderate DIR propensities. Combining CSP and DIRseq, we revise the three drug-binding sites to be $E_{363}$RQRRNELKRSFF$_{375}$, $S_{373}$FFALRDQI$_{381}$, and $K_{397}$KATAYILS$_{405}$ (**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 [24,37–39].

## 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 DIRs 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 [18] (**Figure 2c**), the polyphenol EGCG also binds to many other IDPs, including α-synuclein [40], hIAPP [41,42], Aβ40 [13] and Aβ42 [40], tau [43], and merozoite surface protein 2 [44]. Likewise,

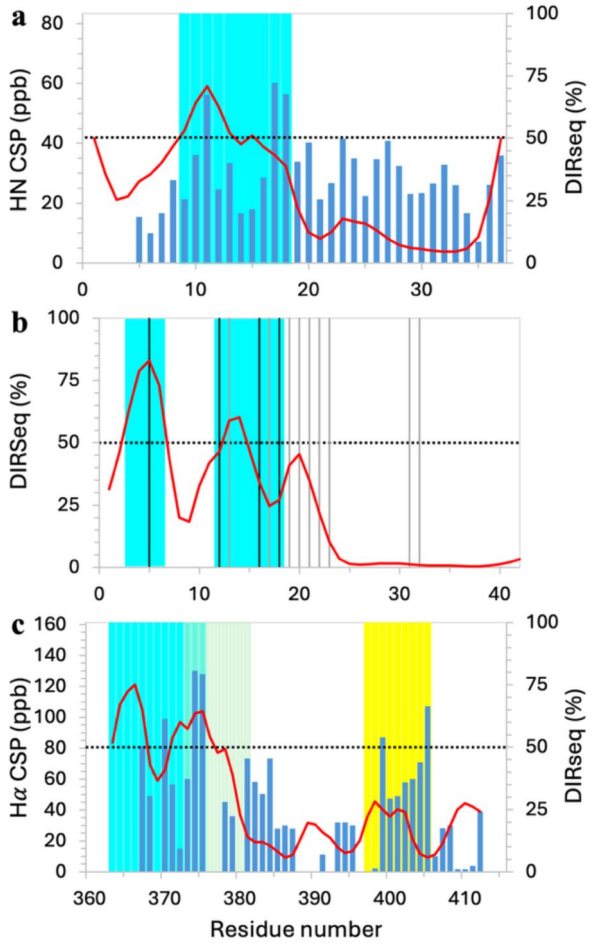


**Figure 4.**

**Drug-binding sites identified by combining CSP or mutation data with DIRseq predictions.**

(a) hIAPP. (b) Aβ42. (c) c-Myc. In panel (b), vertical lines indicate residues with prominent CSPs; those accompanied by NMR peak broadening have their vertical lines in dark color. The threshold for identifying DIRs is lowered to *m* + 1.0 SD.

10074-G5 binds to c-Myc [6] (**Figure 4c**) but also to Aβ42 [25]. On the other hand, c-Myc represents a case where different compounds bind to distinct sites on a single IDP [6]. 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 S4). These features may explain why SJ403 can bind to all three sites whereas SJ710 can bind only to the third site, $F_{87}YYR_{90}$, where three consecutive aromatic residues followed by a basic residue ensure that SJ710 can form both π-π and electrostatic interactions.

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 interactions in forming such sites. DIRseq offers fast speed and a simple, direct link between sequence motif and propensity for drug binding. As more data from CSP and MD simulations become available, DIRseq can be tuned to make even better predictions. For example, it may be feasible to develop DIRseq versions specific to different pharmacophores.

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 fill-length of -synuclein, Robustelli et al. [20] 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. [28] and Basu et al. [22] 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 [17,45]; 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.

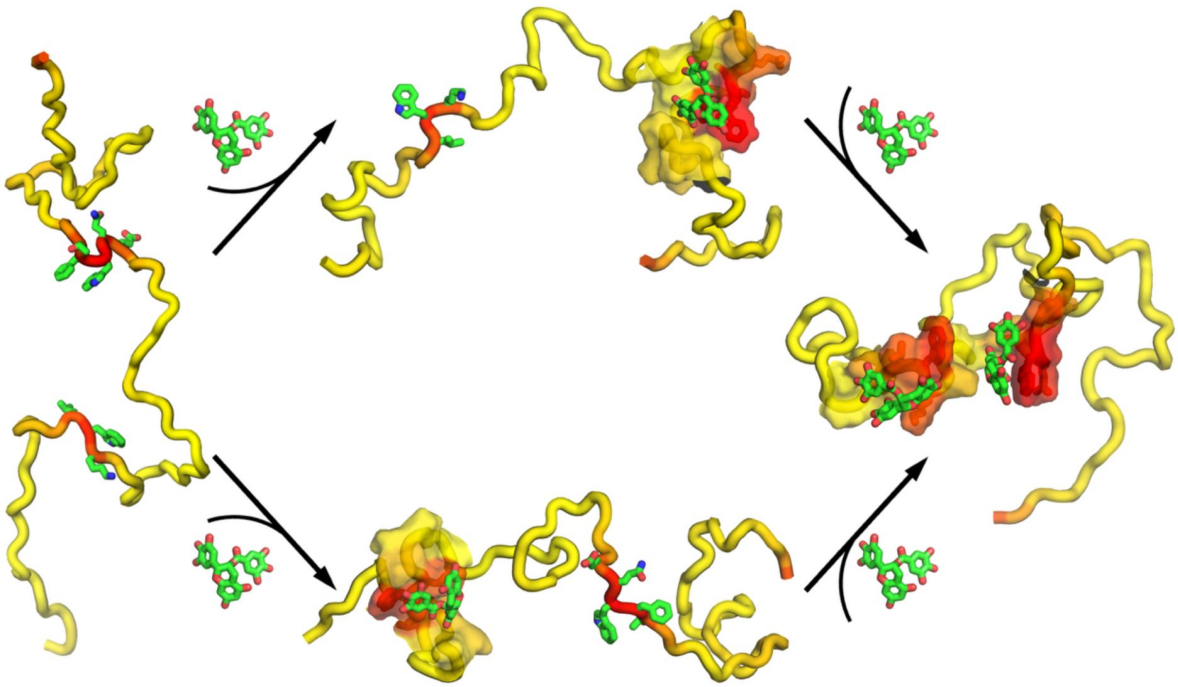

**Figure 5.**

Poses of p53-bound EGCG generated by docking.

# Acknowledgements

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


# Additional information

## Computational Methods

The sequences of the IDP studied here and the drugs that bind to them are listed in Table S1. All DIRseq predictions were obtained using the web server at *https://zhougroup-uic.github.io/DIRseq/*. Conformations of IDPs were generated using the TraDES method [46].

Docking of EGCG onto p53 was performed via the SwissDock web server at *https://www.swissdock.ch/* [47] utilizing the Autodock Vina docking engine [48]. The SMILES string for EGCG from PubChem (*https://pubchem.ncbi.nlm.nih.gov/*; CID 65064) and several conformations of p53 were used as input. A cubic region (13 to 20 Å in side length) around the center of each DIR was selected for docking.


## Funding

**National Institute of General Medical Sciences (GM118091)**


# Additional files

***Supplementary Table and Figures***

# References


(1) Joshi P., Vendruscolo M. (2015) **Druggability of Intrinsically Disordered Proteins** In: Felli I. C., Pierattelli R., editors. *Intrinsically Disordered Proteins Studied by NMR Spectroscopy* Cham: Springer International Publishing https://doi.org/10.1007/978-3-319-20164-1_13 | Google Scholar

(2) Saurabh S., Nadendla K., Purohit S. S., Sivakumar P. M., Cetinel S. (2023) **Fuzzy Drug Targets: Disordered Proteins in the Drug-Discovery Realm** *ACS Omega* **8**:9729–9747 Google Scholar

(3) Wang H., Xiong R., Lai L. (2023) **Rational drug design targeting intrinsically disordered proteins** *Wiley Interdiscip Rev Comput Mol Sci* **13**:e1685 Google Scholar

(4) Uversky V. N. (2024) **How to drug a cloud? Targeting intrinsically disordered proteins** *Pharmacol Rev* **77** https://doi.org/10.1124/pharmrev.124.001113 | Google Scholar

(5) Peterson J. R., Bickford L. C., Morgan D., Kim A. S., Ouerfelli O., Kirschner M. W., Rosen M. K. (2004) **Chemical inhibition of N-WASP by stabilization of a native autoinhibited conformation** *Nat Struct Mol Biol* **11**:747–755 Google Scholar

(6) Hammoudeh D. I., Follis A. V., Prochownik E. V., Metallo S. J. (2009) **Multiple Independent Binding Sites for Small-Molecule Inhibitors on the Oncoprotein c-Myc** *J Am Chem Soc* **131**:7390–7401 Google Scholar

(7) Lamberto G. R., Binolfi A., Orcellet M. L., Bertoncini C. W., Zweckstetter M., Griesinger C., Fernández C. O. (2009) **Structural and mechanistic basis behind the inhibitory interaction of PcTS on α-synuclein amyloid fibril formation** *Proc Natl Acad Sci U S A* **106**:21057–21062 Google Scholar

(8) Ono K., Li L., Takamura Y., Yoshiike Y., Zhu L., Han F., Mao X., Ikeda T., Takasaki J.-i., Nishijo H., et al. (2012) **Phenolic Compounds Prevent Amyloid β-Protein Oligomerization and Synaptic Dysfunction by Site-specific Binding*** *J Biol Chem* **287**:14631–14643 Google Scholar

(9) Krishnan N., Koveal D., Miller D. H., Xue B., Akshinthala S. D., Kragelj J., Jensen M. R., Gauss C.-M., Page R., Blackledge M., et al. (2014) **Targeting the disordered C terminus of PTP1B with an allosteric inhibitor** *Nat Chem Biol* **10**:558–566 Google Scholar

(10) Iconaru L. I., Ban D., Bharatham K., Ramanathan A., Zhang W., Shelat A. A., Zuo J., Kriwacki R. W. (2015) **Discovery of Small Molecules that Inhibit the Disordered Protein, p27Kip1** *Sci Rep* **5**:15686 Google Scholar

(11) De Mol E., Fenwick R. B., Phang C. T., Buzón V., Szulc E., de la Fuente A., Escobedo A., García J., Bertoncini C. W., Estébanez-Perpiñá E., et al. (2016) **EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor** *ACS Chem Biol* **11**:2499–2505 Google Scholar

(12) Tatenhorst L., Eckermann K., Dambeck V., Fonseca-Ornelas L., Walle H., Lopes da Fonseca T., Koch J. C., Becker S., Tönges L., Bähr M., et al. (2016) **Fasudil attenuates aggregation of α-synuclein in models of Parkinson's disease** *Acta Neuropathol Commun* **4**:39 Google Scholar

(13) 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 Aβ Oligomers** *J Am Chem Soc* **139**:13720–13734 Google Scholar

(14) Kocis P., Tolar M., Yu J., Sinko W., Ray S., Blennow K., Fillit H., Hey J. A. (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 Google Scholar

(15) Neira J. L., Bintz J., Arruebo M., Rizzuti B., Bonacci T., Vega S., Lanas A., Velázquez-Campoy A., Iovanna J. L., Abián O. (2017) **Identification of a Drug Targeting an Intrinsically Disordered Protein Involved in Pancreatic Adenocarcinoma** *Sci Rep* **7**:39732 Google Scholar

(16) 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 (15)N labeled amyloid beta (1-42) peptide with osmolytes and phenolic compounds** *Arch Biochem Biophys* **690**:108446 Google Scholar

(17) Ruan H., Yu C., Niu X., Zhang W., Liu H., Chen L., Xiong R., Sun Q., Jin C., Liu Y., et al. (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** *Chem Sci* **12**:3004–3016 Google Scholar

(18) Zhao J., Blayney A., Liu X., Gandy L., Jin W., Yan L., Ha J. H., Canning A. J., Connelly M., Yang C., et al. (2021) **EGCG binds intrinsically disordered N-terminal domain of p53 and disrupts p53-MDM2 interaction** *Nat Commun* **12**:986 Google Scholar

(19) Iruela G., Fernández A., Sagar A., Carvajal F. J., Bernadó P., Pons M. (2022) **A FRET-Based Biosensor for the Src N-Terminal Regulatory Element** *Biosensors* **12**:96 Google Scholar

(20) Robustelli P., Ibanez-de-Opakua A., Campbell-Bezat C., Giordanetto F., Becker S., Zweckstetter M., Pan A. C., Shaw D. E. (2022) **Molecular Basis of Small-Molecule Binding to alpha-Synuclein** *J Am Chem Soc* **144**:2501–2510 Google Scholar

(21) Xu Y., Maya-Martinez R., Guthertz N., Heath G. R., Manfield I. W., Breeze A. L., Sobott F., Foster R., Radford S. E. (2022) **Tuning the rate of aggregation of hIAPP into amyloid using small-molecule modulators of assembly** *Nat Commun* **13**:1040 Google Scholar

(22) Basu S., Martínez-Cristóbal P., Frigolé-Vivas M., Pesarrodona M., Lewis M., Szulc E., Bañuelos C. A., Sánchez-Zarzalejo C., Bielskutė S., Zhu J., et al. (2023) **Rational optimization of a transcription factor activation domain inhibitor** *Nat Struct Mol Biol* **30**:1958–1969 Google Scholar

(23) Heller G. T., Shukla V. K., Figueiredo A. M., Hansen D. F. (2024) **Picosecond Dynamics of a Small Molecule in Its Bound State with an Intrinsically Disordered Protein** *J Am Chem Soc* **146**:2319–2324 Google Scholar

(24) Yu C., Niu X., Jin F., Liu Z., Jin C., Lai L. (2016) **Structure-based Inhibitor Design for the Intrinsically Disordered Protein c-Myc** *Sci Rep* **6**:22298 Google Scholar

(25) Heller G. T., Aprile F. A., Michaels T. C. T., Limbocker R., Perni M., Ruggeri F. S., Mannini B., Löhr T., Bonomi M., Camilloni C., et al. (2020) **Small-molecule sequestration of amyloid-β as a drug discovery strategy for Alzheimer's disease** *Sci Adv* **6** Google Scholar

(26) Herrera-Nieto P., Pérez A., De Fabritiis G. (2020) **Small Molecule Modulation of Intrinsically Disordered Proteins Using Molecular Dynamics Simulations** *J Chem Inf Model* **60**:5003–5010 Google Scholar

(27) 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** *Int J Mol Sci* **23** Google Scholar

(28) 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** *Nat Commun* **13**:6390 Google Scholar

(29) Mehta T., Goyal B. (2024) **Unveiling molecular mechanism underlying inhibition of human islet amyloid polypeptide fibrillation by benzene carboxylic acid-peptide conjugate** *J Mol Liq* **416**:126426 Google Scholar

(30) Lama D., Vosselman T., Sahin C., Liaño-Pons J., Cerrato C. P., Nilsson L., Teilum K., Lane D. P., Landreh M., Arsenian Henriksson M. (2024) **A druggable conformational switch in the c-MYC transactivation domain** *Nat Commun* **15**:1865 Google Scholar

(31) Qin S., Zhou H. X. (2024) **Predicting the sequence-dependent backbone dynamics of intrinsically disordered proteins** *eLife* **12**:RP88958 https://doi.org/10.7554/eLife.88958 | Google Scholar

(32) Muhammedkutty F. N. K., MacAinsh M., Zhou H. X. (2025) **Atomistic molecular dynamics simulations of intrinsically disordered proteins** *Curr Opin Struct Biol* **92**:103029 Google Scholar

(33) Dey S., MacAinsh M., Zhou H. X. (2022) **Sequence-Dependent Backbone Dynamics of Intrinsically Disordered Proteins** *J Chem Theory Comput* **18**:6310–6323 Google Scholar

(34) Klein-Seetharaman J., Oikawa M., Grimshaw S. B., Wirmer J., Duchardt E., Ueda T., Imoto T., Smith L. J., Dobson C. M., Schwalbe H. (2002) **Long-Range Interactions Within a Nonnative Protein** *Science* **295**:1719–1722 Google Scholar

(35) Woods L. A., Platt G. W., Hellewell A. L., Hewitt E. W., Homans S. W., Ashcroft A. E., Radford S. E. (2011) **Ligand binding to distinct states diverts aggregation of an amyloid-forming protein** *Nat Chem Biol* **7**:730–739 Google Scholar

(36) Apostolidou M., Jayasinghe S. A., Langen R. (2008) **Structure of alpha-helical membrane-bound human islet amyloid polypeptide and its implications for membrane-mediated misfolding** *J Biol Chem* **283**:17205–17210 Google Scholar

(37) Han H., Jain A. D., Truica M. I., Izquierdo-Ferrer J., Anker J. F., Lysy B., Sagar V., Luan Y., Chalmers Z. R., Unno K., et al. (2019) **Small-Molecule MYC Inhibitors Suppress Tumor Growth and Enhance Immunotherapy** *Cancer Cell* **36**:483–497 Google Scholar

(38) Shirey R. J., Hart J. R., Sridharan B., Novick S. J., Turner L. D., Zhou B., Nielsen A. L., Eubanks L. M., Ueno L., Hixon M. S., et al. (2021) **Synthetic fluorescent MYC probe: Inhibitor binding site elucidation and development of a high-throughput screening assay** *Bioorg Med Chem* **42**:116246 Google Scholar

(39) Li Z., Huang Y., Hung T. I., Sun J., Aispuro D., Chen B., Guevara N., Ji F., Cong X., Zhu L., et al. (2024) **MYC-Targeting Inhibitors Generated from a Stereodiversified Bicyclic Peptide Library** *J Am Chem Soc* **146**:1356–1363 Google Scholar

(40) Ehrnhoefer D. E., Bieschke J., Boeddrich A., Herbst M., Masino L., Lurz R., Engemann S., Pastore A., Wanker E. E. (2008) **EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers** *Nat Struct Mol Biol* **15**:558–566 Google Scholar

(41) Young L. M., Cao P., Raleigh D. P., Ashcroft A. E., Radford S. E. (2014) **Ion Mobility Spectrometry–Mass Spectrometry Defines the Oligomeric Intermediates in Amylin Amyloid Formation and the Mode of Action of Inhibitors** *J Am Chem Soc* **136**:660–670 Google Scholar

(42) Meng F., Abedini A., Plesner A., Verchere C. B., Raleigh D. P. (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 Google Scholar

(43) Wobst H. J., Sharma A., Diamond M. I., Wanker E. E., Bieschke J. (2015) **The green tea polyphenol (−)-epigallocatechin gallate prevents the aggregation of tau protein into toxic oligomers at substoichiometric ratios** *FEBS Lett* **589**:77–83 Google Scholar

(44) Chandrashekaran I. R., Adda C. G., MacRaild C. A., Anders R. F., Norton R. S. (2010) **Inhibition by Flavonoids of Amyloid-like Fibril Formation by Plasmodium falciparum Merozoite Surface Protein 2** *Biochemistry* **49**:5899–5908 Google Scholar

(45) Dhar A., Sisk T. R., Robustelli P. (2025) **Ensemble docking for intrinsically disordered proteins** *bioRxiv* https://doi.org/10.1101/2025.01.23.634614 | Google Scholar

(46) Feldman H. J., Hogue C. W. (2002) **Probabilistic sampling of protein conformations: new hope for brute force?** *Proteins* **46**:8–23 Google Scholar

(47) Bugnon M., Röhrig U. F., Goullieux M., Perez Marta A. S., Daina A., Michielin O., Zoete V. (2024) **SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina** *Nucl Acids Res* **52**:W324–W332 Google Scholar

(48) Eberhardt J., Santos-Martins D., Tillack A. F., Forli S. (2021) **AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings** *J Chem Inf Model* **61**:3891–3898 Google Scholar

## Author information

**Matthew MacAinsh**
Department of Chemistry, University of Illinois Chicago, Chicago, United States

**Sanbo Qin**
Department of Chemistry, University of Illinois Chicago, Chicago, United States

**Huan-Xiang Zhou**
Department of Chemistry, University of Illinois Chicago, Chicago, United States, Department of Physics, University of Illinois Chicago, Chicago, United States
ORCID iD: 0000-0001-9020-0302

**For correspondence:** hzhou43@uic.edu

## Editors

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

Senior Editor
**Qiang Cui**
Boston University, Boston, United States of America

**Reviewer #1 (Public review):**

Summary:

The authors developed a sequence-based method to predict drug-interacting residues in IDP, based on their recent work, to predict the transverse relaxation rates (R2) of IDP trained on 45 IDP sequences and their corresponding R2 values. The discovery is that the IDPs interact with drugs mostly using aromatic residues that are easy to understand, as most drugs contain aromatic rings. They validated the method using several case studies, and the predictions are in accordance with chemical shift perturbations and MD simulations. The location of the predicted residues serves as a starting point for ligand optimization.

Strengths:

This work provides the first sequence-based prediction method to identify potential drug-interacting residues in IDP. The validity of the method is supported by case studies. It is easy to use, and no time-consuming MD simulations and NMR studies are needed.

Weaknesses:

The method does not depend on the information of binding compounds, which may give general features of IDP-drug binding. However, due to the size and chemical structures of the compounds (for example, how many aromatic rings), the number of interacting residues varies, which is not considered in this work. Lacking specific information may restrict its application in compound optimization, aiming to derive specific and potent binding compounds.

https://doi.org/10.7554/eLife.107470.1.sa2

**Reviewer #2 (Public review):**

Summary:

In this work, the authors introduce DIRseq, a fast, sequence-based method that predicts drug-interacting residues (DIRs) in IDPs without requiring structural or drug information. DIRseq builds on the authors' prior work looking at NMR relaxation rates, and presumes that those residues that show enhanced R2 values are the residues that will interact with drugs, allowing these residues to be nominated from the sequence directly. By making small modifications to their prior tool, DIRseq enables the prediction of residues seen to interact with small molecules in vivo.

Strengths:

The preprint is well written and easy to follow

Weaknesses:

(1) The DIRseq method is based on SeqDYN, which itself is a simple (which I do not mean as a negative - simple is good!) statistical predictor for R2 relaxation rates. The challenge here is that R2 rates cover a range of timescales, so the physical intuition as to what exactly elevated R2 values mean is not necessarily consistent with "drug interacting". Presumably, the authors are not using the helix boost component of SeqDYN here (it would be good to explicitly state this). This is not necessarily a weakness, but I think it would behove the authors to compare a few alternative models before settling on the DIRseq method, given the somewhat ad hoc modifications to SeqDYN to get DIRseq.

Specifically, the authors previously showed good correlation between the stickiness parameter of Tesei et al and the inferred "q" parameter for SeqDYN; as such, I am left wondering if comparable accuracy would be obtained simply by taking the stickiness parameters directly and using these to predict "drug interacting residues", at which point I'd argue we're not really predicting "drug interacting residues" as much as we're predicting "sticky" residues, using the stickiness parameters. It would, I think, be worth the authors comparing the predictive power obtained from DIRseq with the predictive power obtained by using the lambda coefficients from Tesei et al in the model, local density of aromatic residues, local hydrophobicity (note that Tesei at al have tabulated a large set of hydrophobicity scores!) and the raw SeqDYN predictions. In the absence of lots of data to compare against, this is another way to convince readers that DIRseq offers reasonable predictive power.

(2) Second, the DIRseq is essentially SeqDYN with some changes to it, but those changes appear somewhat ad hoc. I recognize that there is very limited data, but the tweaking of parameters based on physical intuition feels a bit stochastic in developing a method; presumably (while not explicitly spelt out) those tweaks were chosen to give better agreement with the very limited experimental data (otherwise why make the changes?), which does raise the question of if the DIRseq implementation of SeqDYN is rather over-parameterized to the (very limited) data available now? I want to be clear, the authors should not be critiqued for attempting to develop a model despite a paucity of data, and I'm not necessarily saying this is a problem, but I think it would be really important for the authors to acknowledge to the reader the fact that with such limited data it's possible the model is over-fit to specific sequences studied previously, and generalization will be seen as more data are collected.

(3) Third, perhaps my biggest concern here is that - implicit in the author's assumptions - is that all "drugs" interact with IDPs in the same way and all drugs are "small" (motivating the change in correlation length). Prescribing a specific lengthscale and chemistry to all drugs seems broadly inconsistent with a world in which we presume drugs offer some degree of specificity. While it is perhaps not unexpected that aromatic-rich small molecules tend to interact with aromatic residues, the logical conclusion from this work, if one assumes DIRseq has utility, is that all IDRs bind drugs with similar chemical biases. This, at the very least, deserves some discussion.

(4) Fourth, the authors make some general claims in the introduction regarding the state of the art, which appear to lack sufficient data to be made. I don't necessarily disagree with the author's points, but I'm not sure the claims (as stated) can be made absent strong data to support them. For example, the authors state: "Although an IDP can be locked into a specific conformation by a drug molecule in rare cases, the prevailing scenario is that the protein remains disordered upon drug binding." But is this true? The authors should provide evidence to support this assertion, both examples in which this happens, and evidence to support the idea that it's the "prevailing view" and specific examples where these types of interactions have been biophysically characterized.

Similarly, they go on to say:

"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." But again, where is the data to support this assertion? I don't necessarily disagree, but we need specific empirical studies to justify declarative claims like this; otherwise, we propagate lore into the scientific literature. The use of "typically" here is a strong claim, implying most IDP complexes behave in a certain way, yet how can the authors make such a claim?

Finally, they continue to claim:

"Such drug interacting residues (DIRs), akin to binding pockets in structured proteins, are key to optimizing compounds and elucidating the mechanism of action." But again, is this a fact or a hypothesis? If the latter, it must be stated as such; if the former, we need data and evidence to support the claim.

https://doi.org/10.7554/eLife.107470.1.sa1

**Author response:**

> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *The authors developed a sequence-based method to predict drug-interacting residues in IDP, based on their recent work, to predict the transverse relaxation rates (R2) of IDP trained on 45 IDP sequences and their corresponding R2 values. The discovery is that the IDPs interact with drugs mostly using aromatic residues that are easy to understand, as most drugs contain aromatic rings. They validated the method using several case studies, and the predictions are in accordance with chemical shift perturbations and MD simulations. The location of the predicted residues serves as a starting point for ligand optimization.*
>
> *Strengths:*
>
> *This work provides the first sequence-based prediction method to identify potential drug-interacting residues in IDP. The validity of the method is supported by case studies. It is easy to use, and no time-consuming MD simulations and NMR studies are needed.*
>
> *Weaknesses:*

The method does not depend on the information of binding compounds, which may give general features of IDP-drug binding. However, due to the size and chemical structures of the compounds (for example, how many aromatic rings), the number of interacting residues

varies, which is not considered in this work. Lacking specific information may restrict its application in compound optimization, aiming to derive specific and potent binding compounds.

We fully recognize that different compounds may have different interaction propensity profiles along the IDP sequence. In future studies, we will investigate compound-specific parameter values. The limiting factor is training data, but such data are beginning to be available.

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *In this work, the authors introduce DIRseq, a fast, sequence-based method that predicts drug-interacting residues (DIRs) in IDPs without requiring structural or drug information. DIRseq builds on the authors' prior work looking at NMR relaxation rates, and presumes that those residues that show enhanced R2 values are the residues that will interact with drugs, allowing these residues to be nominated from the sequence directly. By making small modifications to their prior tool, DIRseq enables the prediction of residues seen to interact with small molecules in vivo.*
>
> *Strengths:*
>
> *The preprint is well written and easy to follow*
>
> *Weaknesses:*
>
> *(1) The DIRseq method is based on SeqDYN, which itself is a simple (which I do not mean as a negative - simple is good!) statistical predictor for R2 relaxation rates. The challenge here is that R2 rates cover a range of timescales, so the physical intuition as to what exactly elevated R2 values mean is not necessarily consistent with "drug interacting". Presumably, the authors are not using the helix boost component of SeqDYN here (it would be good to explicitly state this). This is not necessarily a weakness, but I think it would behove the authors to compare a few alternative models before settling on the DIRseq method, given the somewhat ad hoc modifications to SeqDYN to get DIRseq.*

Actually, the factors that elevate R2 are well-established. These are local interactions and residual secondary structures (if any). The basic assumption of our method is that intra-IDP interactions that elevate R2 convert to IDP-drug interactions. This assumption was supported by our initial observation that the drug interaction propensity profiles predicted using the original SeqDYN parameters already showed good agreement with CSP profiles. We only made relatively small adjustments to the parameters to improve the agreement. Indeed we did not apply the helix boost portion of SeqDYN to DIRseq, and will state as such. We will also compare DIRseq with several alternative models.

> *Specifically, the authors previously showed good correlation between the stickiness parameter of Tesei et al and the inferred "q" parameter for SeqDYN; as such, I am left wondering if comparable accuracy would be obtained simply by taking the stickiness parameters directly and using these to predict "drug interacting residues", at which point I'd argue we're not really predicting "drug interacting residues" as much as we're predicting "sticky" residues, using the stickiness parameters. It would, I think, be worth the authors comparing the predictive power obtained from DIRseq with the predictive power obtained by using the lambda coefficients from Tesei et al in the model, local density of aromatic residues, local hydrophobicity (note that Tesei at al have tabulated a large set of hydrophobicity scores!) and the raw SeqDYN predictions. In the absence of*

> *lots of data to compare against, this is another way to convince readers that DIRseq offers reasonable predictive power.*

We will compare predictions of these various parameter sets, and summarize the results in a table.

> *(2) Second, the DIRseq is essentially SeqDYN with some changes to it, but those changes appear somewhat ad hoc. I recognize that there is very limited data, but the tweaking of parameters based on physical intuition feels a bit stochastic in developing a method; presumably (while not explicitly spelt out) those tweaks were chosen to give better agreement with the very limited experimental data (otherwise why make the changes?), which does raise the question of if the DIRseq implementation of SeqDYN is rather over-parameterized to the (very limited) data available now? I want to be clear, the authors should not be critiqued for attempting to develop a model despite a paucity of data, and I'm not necessarily saying this is a problem, but I think it would be really important for the authors to acknowledge to the reader the fact that with such limited data it's possible the model is over-fit to specific sequences studied previously, and generalization will be seen as more data are collected.*

We have explained the rationale for the parameter tweaks, which were limited to q values for four amino-acid types, i.e., to deemphasize hydrophobic interactions and slightly enhance electrostatic interactions (p. 4-5). We will add that these tweaks were motivated by observations from MD simulations of drug interactions with a-syn (ref 20). As already noted in the response to the preceding comment, we will also present results for the original parameter values as well as for when the four q values are changed one at a time.

> *(3) Third, perhaps my biggest concern here is that - implicit in the author's assumptions - is that all "drugs" interact with IDPs in the same way and all drugs are "small" (motivating the change in correlation length). Prescribing a specific lengthscale and chemistry to all drugs seems broadly inconsistent with a world in which we presume drugs offer some degree of specificity. While it is perhaps not unexpected that aromatic-rich small molecules tend to interact with aromatic residues, the logical conclusion from this work, if one assumes DIRseq has utility, is that all IDRs bind drugs with similar chemical biases. This, at the very least, deserves some discussion.*

The reviewer raises a very important point. In Discussion, we will add that it is important to further develop DIRseq to include drug-specific parameters when data for training become available.

> *(4) Fourth, the authors make some general claims in the introduction regarding the state of the art, which appear to lack sufficient data to be made. I don't necessarily disagree with the author's points, but I'm not sure the claims (as stated) can be made absent strong data to support them. For example, the authors state: "Although an IDP can be locked into a specific conformation by a drug molecule in rare cases, the prevailing scenario is that the protein remains disordered upon drug binding." But is this true? The authors should provide evidence to support this assertion, both examples in which this happens, and evidence to support the idea that it's the "prevailing view" and specific examples where these types of interactions have been biophysically characterized.*

We will cite several studies showing that IDPs remain disordered upon drug binding.

> *Similarly, they go on to say:*

> *"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." But again, where is the data to support this assertion? I don't necessarily disagree, but we need specific empirical studies to justify declarative claims like this; otherwise, we propagate lore into the scientific literature. The use of "typically" here is a strong claim, implying most IDP complexes behave in a certain way, yet how can the authors make such a claim?*

Here again we will add citations to support the statement.

> *Finally, they continue to claim:*
>
> *"Such drug interacting residues (DIRs), akin to binding pockets in structured proteins, are key to optimizing compounds and elucidating the mechanism of action." But again, is this a fact or a hypothesis? If the latter, it must be stated as such; if the former, we need data and evidence to support the claim.*

We will add citations to both compound optimization and mechanism of action.

https://doi.org/10.7554/eLife.107470.1.sa0