

Avaliação de Proteínas com Desordem Intrínseca do Banco de Dados DisProt: Conceitos, Importância Biológica e Características Peculiares

Evaluation of Proteins with Intrinsic Disorder from the DisProt Database: Concepts, Biological Importance and Peculiar Characteristics

Stephanny M. Alves de Souza,^a Theo L. F. de Souza^{a,*} 

^aUniversidade Federal do Rio de Janeiro, Faculdade de Farmácia, Departamento de Fármacos e Medicamentos, Cidade Universitária, CEP 21941-902, Rio de Janeiro-RJ, Brasil.

*E-mail: theo@pharma.ufrj.br

Recebido em: 20 de Fevereiro de 2021

Aceito em: 29 de Abril de 2021

Publicado online: 10 de Janeiro de 2022

The intrinsic disorder (ID) consists of a feature present in an extensive group of proteins in its totality (IDPs) or only in regions (IDRs). Such feature provides a series of peculiar behaviors that differs from those presented for globular proteins. They are characterized by the absence of a well-defined tridimensional structure, owning high flexibility and conformational entropy, which enables a structural plasticity. In this work, we perform an investigation of different issues involving proteins with ID available in the DisProt database, as well as an extensive literature search to gather information that can help to expand the knowledge about this important group of proteins. Proteins with ID are observed in all kingdoms of life from the simplest to the most complex organisms, where they become more present. They participate of the most diverse biological process, acting mainly in process involving signaling, where the ability to adopt different conformations and bind to different partners becomes fundamental. The ID in proteins is also responsible for the development of diseases and its peculiar structural features bring with them the challenge of developing drugs able to act in these proteins. Different techniques can be applied to analyze structural features presented by these proteins, such as the absence or low structural content, possibility of structural gain, large volume, high solvent exposure and flexibility. With this work, we seek to bring a set of information from the most varied topics involving proteins with ID that can help in the theme disclosure.

Keywords: Proteins; intrinsic disorder; promiscuity; structural characterization

1. Introdução

Os avanços nas áreas de biologia molecular e bioinformática têm levado à descoberta de um grupo cada vez mais crescente de proteínas que não apresentam uma estrutura tridimensional definida.¹⁻³ Essa característica, conhecida como desordem intrínseca (ID, *Intrinsic Disorder*), pode ocorrer tanto em regiões intrinsecamente desordenadas (IDRs, *Intrinsically disordered regions*) em uma proteína bem enovelada, como em proteínas intrinsecamente desordenadas (IDPs, *Intrinsically disordered proteins*) em sua totalidade. Ao contrário do que se pensava sobre IDRs funcionarem apenas conectando domínios estruturados, tanto IDRs como IDPs são capazes de mediar diferentes funções.⁴ Isso se deve ao fato de que, por não possuírem uma estrutura rígida, elas são capazes de adotar múltiplas conformações em solução, o que pode permitir a ligação com múltiplos parceiros celulares.^{3,5} Portanto, a identificação e o entendimento de proteínas/regiões intrinsecamente desordenadas (IDPs/IDRs) é importante para o desenho de fármacos e para um melhor entendimento de processos biológicos.

A funcionalidade de IDPs/IDRs pode ocorrer por diferentes mecanismos moleculares, os quais se baseiam principalmente em processos de transição desordem-ordem mediante a presença do parceiro celular, tal como ocorre com os elementos de reconhecimento molecular (MoRFs, "*Molecular Recognition Features*").⁶ Os MoRFs podem compreender regiões de 10 a 70 resíduos e são classificados como α -MoRFs, β -MoRFs ou i-MoRFs, dependendo do tipo de estrutura adotado (α -hélice, folha- β ou irregulares, respectivamente).⁷ As interações de IDPs/IDRs com seus parceiros também podem ser mediadas por motivos lineares, os quais são regiões curtas, possuindo em média 6 resíduos de comprimento, que se ligam geralmente com baixa afinidade aos seus parceiros devido ao número limitado de resíduos que fazem o contato.⁸

As IDPs/IDRs podem atuar por meio da ligação com o DNA e RNA, proteínas e pequenos ligantes, como também sofrendo modificações pós-traducionais.⁵ Elas participam dos mais variados processos biológicos, tais como regulação da tradução e transcrição, transdução de sinal, ciclo celular, estoque de pequenas moléculas, automontagem de complexos multiméricos, mitose,

meiose, apoptose, condensação do DNA, dentre outros.^{3,9} Em um estudo pioneiro, Dunker e colaboradores (2002) identificaram, para um conjunto de mais de 150 proteínas com IDRs de 30 resíduos consecutivos ou mais, um total de 28 diferentes funcionalidades relacionadas à desordem intrínseca.¹⁰ Posteriormente, foi observado que nesse estudo não foram distinguidos processos biológicos de funções moleculares, o que levou à separação em 6 modos funcionais, que são: cadeias entrópicas, sítios de modificação de proteínas, chaperonas, efetores, montadores e sequestradores.^{5,11}

A ID, embora seja presente em todos os reinos da vida, é mais observada em seres vivos mais complexos. Avalia-se que cerca de 2% do proteoma de arqueas, 4,2% de bactérias e 33% de eucariotos contém IDPs/IDRs com mais de 30 resíduos.¹² Acredita-se que a presença da ID em proteínas de organismos mais complexos estaria relacionada principalmente com a participação em vias de sinalização, reconhecimento e regulação, sendo fundamentais para eucariotos, em especial os multicelulares.¹³

Devido a ampla distribuição da ID no proteoma de eucariotos é válido esperar a sua participação em algumas doenças em seres humanos. Muitas dessas doenças podem ser originadas por processos de *splicing* alternativo (dando origem a regiões desordenadas), modificações pós-traducionais em regiões desordenadas, promiscuidade (capacidade de ligação com múltiplos parceiros) e plasticidade estrutural, a qual pode levar ao enovelamento errôneo causando processos de agregação (tal como a formação de fibras), perda de função ou ganho de toxicidade.¹⁴ Exemplos de doenças relacionadas com a ID em proteínas são o câncer (proteínas p53, BRCA-1 e c-Myc),¹⁵⁻¹⁷ doenças cardiovasculares,¹⁸ diabetes tipo 2 (amilina),¹⁹ e doenças neurodegenerativas como a doença do príon (proteína do príon),²⁰ Parkinson (α -sinucleína),²¹ e Alzheimer (proteínas A β e Tau).^{22,23}

A necessidade de desenvolver fármacos capazes de agir em IDPs/IDRs apresenta alguns desafios. Diferente do que ocorre com proteínas estruturadas, proteínas com ID apresentam mais cavidades de ligação disponíveis e superfícies de ligação maiores.²⁴ Além disso, IDPs/IDRs apresentam ligações de baixa afinidade com pequenos ligantes, as quais decorrem do fato dessas proteínas realizarem ligações mais fracas e transientes, e também da grande perda de entropia que ocorre durante a ligação devido à transição desordem-ordem.²⁵

Estudos mostram que, estruturalmente, proteínas com ID não constituem um grupo uniforme.²⁶ Tal constatação foi realizada por meio da análise da compactação e conteúdo estrutural residual dessas proteínas, as quais apresentam em parte um comportamento similar a proteínas num estado *pre-molten globule*, enquanto outras apresentam características de um estado *random-coil*.²⁶ Tais observações levaram a uma subdivisão dessas proteínas em *pre-molten globule-like* (mais compactas e com maior conteúdo estrutural residual) e *coil-like* (menos compactas e sem ou quase nenhum conteúdo estrutural residual).²⁶

A caracterização estrutural de IDPs/IDRs pode se dar avaliando sua flexibilidade, parâmetros hidrodinâmicos, globularidade, conteúdo estrutural residual, degradação proteolítica, migração eletroforética aberrante, alta massa molecular aparente, resistência térmica, resistência ácida e resposta a condições desnaturantes.^{26,27} Técnicas como cromatografia por exclusão de tamanho (SEC), calorimetria diferencial de varredura, calorimetria de titulação isotérmica, imunquímica, viscosimetria, além de uma série de técnicas de espectroscopia como, dicróismo circular (CD), Raman, fluorescência, ressonância magnética nuclear (RMN), espalhamento de raios-X a baixo ângulo (SAXS), espalhamento de luz dinâmico (DLS) e espectrometria de massas (EM), dentre outras, podem ser empregadas para avaliar diferentes aspectos estruturais apresentados por esse grupo de proteínas.²⁷

A composição de aminoácidos tem sido proposta como uma forma de distinção entre proteínas enoveladas e proteínas com ID, uma vez que tem-se observado que essas últimas apresentam uma composição distinta, geralmente ricas em resíduos carregados e com depleção de resíduos hidrofóbicos.²⁸ IDPs/IDRs normalmente apresentam uma maior quantidade de resíduos promotores da desordem (A, R, G, Q, S, P, E, e K) enquanto apresentam uma baixa quantidade de resíduos promotores da ordem (W, C, F, I, Y, V, L, e N).²⁹ A composição de aminoácidos tem sido inclusive utilizada como critério para a predição da ID. Alguns preditores se baseiam nas propriedades físico-químicas dos aminoácidos, como é o caso do GlobPlot,³⁰ FoldUnfold,³¹ e IUPred.³² Outros preditores utilizam redes neurais treinadas a partir de bases de dados, como no caso do PONDR,^{33,34} DisEMBL,³⁵ e ESpritz.³⁶ Existem ainda os preditores que utilizam a combinação da predição da ID por diferentes preditores, como o PONDR-FIT,³⁷ e o metaPrDOS,³⁸ por exemplo.

Considerando o aumento expressivo da descoberta de proteínas contendo ID nos últimos anos e a sua presença em diversos organismos, apresentando as mais variadas funções, torna-se necessário um maior conhecimento acerca dessas proteínas. Devido às suas características peculiares, uma reunião de informações sobre proteínas com desordem intrínseca permite obter um maior conhecimento sobre essa área. Com esse trabalho visamos ainda reunir informações gerais sobre proteínas apresentando ID presentes no banco de dados DisProt, tais como funções, métodos de análises mais utilizados na caracterização, tamanho de IDRs, dentre outras características importantes.

2. Metodologia

Este trabalho consiste na análise de informações sobre características encontradas em proteínas com ID presentes na base de dados DisProt (<http://www.disprot.org/>).³⁹ O DisProt consiste em uma base de dados sobre proteínas com ID, tendo sido utilizada neste trabalho dados da

versão DisProt 7 v0.3 26-09-2016. Essa base dispunha de 803 proteínas com ID, as quais apresentavam um total de 2167 IDRs.

Informações sobre as 803 proteínas presentes no DisProt, tais como o nome da proteína, organismo no qual está presente, reino, função molecular, processos biológicos os quais estão envolvidas, quantidade de parceiros descritos, quantidade de regiões de transição desordem-ordem, quantidade de regiões desordenadas presentes na estrutura, quantidade de resíduos desordenados por região e métodos experimentais de identificação e análise da ID, foram coletadas. Dados sobre função molecular e processos biológicos foram obtidos utilizando os dados disponíveis na base de dados UniProtKB (<https://www.uniprot.org/help/uniprotkb>).

3. Resultados e Discussão

3.1. Proteínas com ID são mais descritas em Organismos Complexos

A partir da análise das informações disponíveis na base de dados DisProt foram reunidas diferentes características apresentadas por proteínas com ID. Inicialmente, foram analisadas as informações sobre a distribuição e presença entre os diferentes reinos da vida. A partir da Figura 1 foi observado que a grande maioria das IDPs/IDRs presentes no DisProt (53%) estão presentes no reino animal, tendo sido observado um total de 426 proteínas. O segundo grupo com maior número observado foram as bactérias (19%), possuindo 154 proteínas. Foi observado também que a ID é bastante presente nos vírus, os quais possuem 75 IDPs/IDRs descritas correspondendo a 9% do total. Também foi observado que a ID é presente de forma considerável entre os fungos e as plantas, os quais possuem 63 e 53 proteínas, correspondendo a 8 e 7% do total, respectivamente. A desordem foi menos observada entre os protozoários e as

arqueas, os quais possuíam 17 e 15 IDPs/IDRs descritas, respectivamente, sendo equivalente a 2% cada do total.

Os presentes resultados confirmaram que a ID está mais descrita, de uma forma geral, entre os organismos mais desenvolvidos tais como os eucariotos, estando de acordo com o sugerido na literatura. Trabalhos utilizando análises de bioinformática para predição da ID em proteomas de espécies de diferentes reinos da vida têm mostrado que a ID é muito mais presente em eucariotos do que em procariotos, sugerindo haver uma relação entre a desordem e a complexidade dos organismos.^{12,40} Embora esses achados tenham sido relatados, é importante destacar que este estudo considerou apenas os dados contidos no banco de dados DisProt. Dessa forma, considerando que muitas proteínas contendo ID podem ainda não serem identificadas e que alguns organismos são mais estudados que outros, os resultados aqui discutidos podem não representar o perfil real de distribuição da ID na natureza.

Dentro dessa base de dados, foi avaliado ainda as espécies com maior presença de ID dentre cada um desses grupos. Foi observado que a espécie humana (*Homo sapiens*) foi a que mais possuía proteínas com ID dentre os animais, com 244 proteínas. Estudos de bioinformática mostram que boa parte do proteoma humano é constituído por proteínas com ID. O estudo de Ward e colaboradores (2004) sugeriu que 35,2% das proteínas humanas possuíam IDRs com mais de 30 resíduos e cerca de 21,9% possuíam IDRs com mais de 50 resíduos.¹² A predição da ID realizada por Pentony e Jones (2010) sugeriu que, para um conjunto de 26011 proteínas humanas, 11477 teriam IDRs de pelo menos 30 resíduos de comprimento.⁴¹ Um estudo realizado por Deiana e colaboradores (2019) considerando como IDPs as proteínas contendo pelo menos 30% de seus resíduos desordenados sugeriram que 32% do proteoma humano é constituído de IDPs.⁴² Em conjunto, esses trabalhos sugerem que por volta de um terço das proteínas humanas contém ID.

Dentre as plantas, a espécie com mais proteínas contendo ID observada foi a *Arabidopsis thaliana*, com

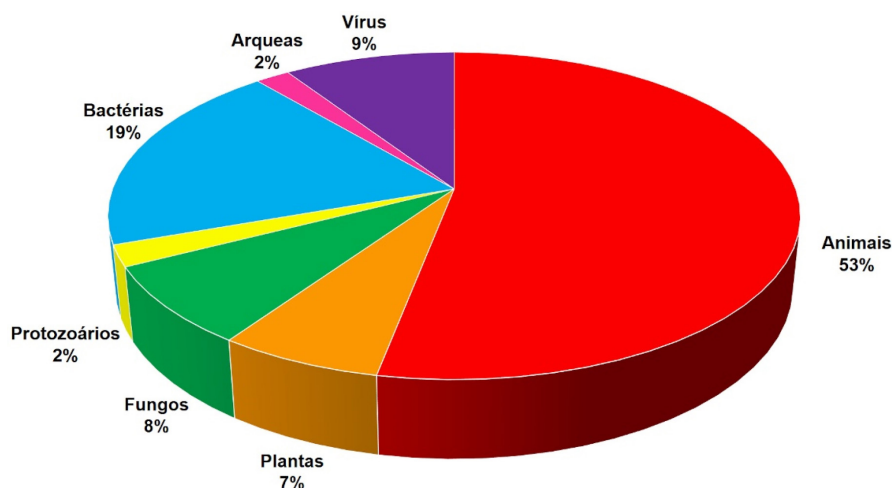


Figura 1. Prevalência de proteínas com ID e distribuição entre os diferentes reinos. A figura apresenta a porcentagem de proteínas com ID descritas na base DisProt pertencendo aos diferentes reinos da vida

20 proteínas. Um estudo realizando a predição da desordem no proteoma de *Arabidopsis thaliana* encontrou que a ID é presente significativamente em processos relacionados com a sinalização, desenvolvimento, ciclo celular e resposta ao estresse.⁴³ Esse mesmo estudo aponta que a ID seria importante em plantas para aumentar a complexidade e versatilidade de respostas a mudanças no meio.⁴³ Como exemplos de proteínas com ID presentes em plantas há as proteínas abundantes da embriogênese tardia (LEA), tal como as deidrininas, as quais são proteínas expressas em condições de estresse como o frio, seca e alta salinidade, por exemplo.⁴⁴ As proteínas GRAS que agem em processos de desenvolvimento e sinalização,⁴⁵ NAC na percepção do estresse e desenvolvimento,⁴⁶ e bZIP na fotomorfogênese são outros exemplos.⁴⁷ Outro exemplo de IDP em plantas é o peptídeo de soja Lunasina, com potencial atividade anticâncer, o qual é parte da proteína 2S albumina e teria participação no processo de estocagem em sementes durante a etapa de expansão celular.^{48,49} Estudos indicam que a Lunasina é uma IDP no estado *pre-molten globule-like* e com alta plasticidade estrutural, como verificado por CD e simulações de dinâmica molecular.⁴⁸

No caso dos fungos, a espécie com maior número de ID foi a *Saccharomyces cerevisiae* com 52 proteínas. Dentre as bactérias, a espécie mais comum foi a *Escherichia coli* com 59 proteínas. Análises de predição de ID nos proteomas do *S. cerevisiae* e *E. coli*, entretanto, têm sugerido que a ID estaria cerca de 3 vezes mais presente na *S. cerevisiae*.⁵⁰ Dentre as arqueas, a espécie contendo mais ID encontrada foi a *Methanocaldococcus jannaschii* com 6 proteínas. Comparando-se as informações contidas nessa base de dados entre os procariotos (bactérias e arqueas), foi observado um número significativamente maior de proteínas com ID em bactérias, tal como se observa na literatura. Estudos comparando a ID no proteoma de procariotos sugerem que as bactérias possuem maior conteúdo de ID do que arqueas.^{51,52} Em arqueas, a ID estaria mais presente naquelas espécies que habitam locais com condições ambientais mais severas.⁵³

Os maiores representantes dentre os protozoários foram o *Plasmodium falciparum* e a *Tetrahymena thermophila* cada um com 4 proteínas apresentando ID. O *Plasmodium falciparum* já tem sido descrito, por meio de análises de predição de ID, possuindo alta taxa de desordem no seu proteoma.^{54,55} Comparativamente, os protozoários apicomplexos possuem a ID como sendo mais comum do que bactérias, arqueas, outros eucariotos unicelulares e organismos maiores.⁵⁴ É proposto que a ID nesses parasitas poderia ser importante para a sua fixação e/ou invasão e sobrevivência no hospedeiro.⁵⁴

Neste trabalho, a presença da ID foi também avaliada nos vírus, sendo o maior representante o vírus da imunodeficiência humana (HIV-1) com 12 proteínas. Os vírus apresentam uma grande variabilidade na quantidade de desordem em seus proteomas, como observado em alguns trabalhos.^{56,57} Entretanto, os vírus, de uma forma

geral, parecem possuir maior conteúdo de ID que os procariotos.^{56,57} Os vírus apresentam uma composição de aminoácidos mais similar a de eucariotos, com um decréscimo de resíduos promotores da ordem, o que em geral proporciona maior flexibilidade e mutabilidade às suas proteínas.⁵⁸ Assim como em eucariotos, muitas IDRs estão presentes em sítios de modificação pós-traducional.⁵⁶ O HIV assim como outros vírus causadores de doenças sexualmente transmissíveis, tal como o da hepatite C e o vírus da herpes simplex tipo-2 possuem em comum a dificuldade no desenvolvimento de vacinas. Alguns autores relacionam essa dificuldade com a grande presença da ID no capsídeo viral, o que levaria a alterações conformacionais dificultando a ligação de anticorpos.⁵⁹ A virulência de vírus da família *Flaviviridae* também tem sido relacionada com a presença da ID em proteínas do capsídeo.⁶⁰

3.2. Principais Funções Moleculares Exercidas por Proteínas com ID

Neste trabalho foi realizada uma avaliação das principais funções moleculares exercidas pelas proteínas presentes na base de dados DisProt. Na Figura 2 são mostradas as funções moleculares comuns a pelo menos 3 proteínas com ID. Foi observado que a grande maioria dessas proteínas estão relacionadas com atividades catalíticas e de ligação com DNA e RNA. Interações entre proteínas com ID e ácidos nucleicos têm sido descritas em diversos trabalhos como uma das principais formas de atuação dessas proteínas.^{9,10,12,54,56} No caso das interações com DNA, a ID tem como papel a modulação da afinidade da interação.⁶¹ Dentre as interações com RNA, tem-se notado principalmente a atuação nos complexos do spliceossomo e do ribossomo.^{62,63}

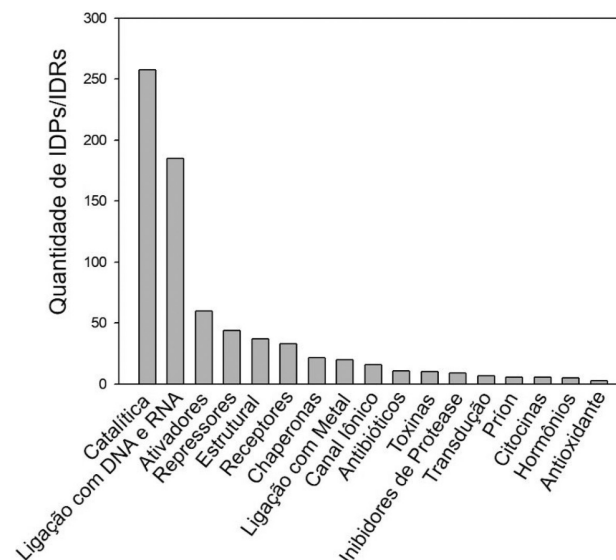


Figura 2. Principais funções moleculares exercidas pelas proteínas com desordem intrínseca. Figura relacionando a quantidade de proteínas contendo ID exercendo as principais funções moleculares observadas para proteínas presentes na base DisProt. As funções moleculares para cada proteína foram obtidas através do UniProtKB.

Dentre as atividades catalíticas observadas, estão as endonucleases, exonucleases, transferases, oxirredutases, descarboxilases, quinases, liases, ligases, sintetases, hidrolases, proteases, isomerases, fosfatase, esterase e helicases. É descrito na literatura que atividades catalíticas estão mais relacionadas com proteínas estruturadas. Entretanto, considerando que na DisProt estão disponíveis tanto IDPs como proteínas contendo IDRs, a ID identificada nessas proteínas pode estar presente em regiões que não necessariamente estão relacionadas com a atividade catalítica. Algumas enzimas com multidomínios são geralmente ricas em ID, os quais atuam como regiões conectoras, sendo consideradas multifuncionais e metabolicamente promiscuas, como exemplo das FADs sintetases e ácidos graxos sintases.⁶⁴

Além da atuação dessas proteínas em processos enzimáticos e de ligação com DNA, também são descritas proteínas com ID atuando como repressores, ativadores, receptores, com funções estruturais, chaperonas, ligação com metais, canais iônicos, antibióticos, toxinas, inibidores de proteases, transdutoras de sinal, príons, citocinas, hormônios e antioxidantes (Figura 2).

3.3. Principais Processos Biológicos com Envolvimento de Proteínas com ID

Os principais processos biológicos dos quais as proteínas contendo ID participam também foram avaliados. A Figura 3A apresenta os processos biológicos mais prevalentes contendo até no mínimo 10 proteínas, enquanto a Figura 3B mostra os processos biológicos os quais essas proteínas estão envolvidas em menor parcela, contendo entre 10 e 3 proteínas com ID. A partir desses dados foi observado que tais proteínas participam em sua grande maioria de

processos de transcrição sendo seguido dos processos de transporte e dano e reparo do DNA (Figura 3A).

Os fatores de transcrição apresentam uma presença significativa da ID, tendo geralmente uma composição rica em resíduos promotores da desordem e uma baixa quantidade de resíduos hidrofóbicos.⁶⁵ A ID também é uma característica mais presente em fatores de transcrição de eucariotos do que em procariotos, o que poderia estar relacionado com o fato de um sistema de transcrição mais desenvolvido ter a capacidade de interagir com múltiplos parceiros celulares e regular a afinidade da ligação.⁶⁵ A fim de que possam exercer a sua função, fatores de transcrição precisam ser transportados para o núcleo das células. O transporte de fatores de transcrição para o núcleo ocorre através do complexo do poro nuclear, por meio de sua interação com nucleoporinas que formam tal complexo.⁶⁶ Nucleoporinas compostas de unidades repetidas de fenilalanina e glicina (FG nucleoporinas) são intrinsecamente desordenadas, e o rápido e seletivo transporte de fatores de transcrição estaria relacionado com a estrutura dinâmica dessas proteínas.⁶⁶ A manutenção da integridade do DNA é essencial para a vida, uma vez que danos podem levar a mutações e, conseqüentemente, o desenvolvimento de doenças.⁶⁷ Por isso, as células possuem diversos mecanismos para detectar e reparar danos no DNA.⁶⁷ Uma importante proteína na via de reparo e excisão de nucleotídeos é a XPA (Xeroderma pigmentosum grupo A) a qual possui cerca de um terço de sua estrutura desordenada.⁶⁸

As proteínas com ID também estão bastante presentes em processos envolvendo o ciclo celular e apoptose (Figura 3A). No ciclo celular, as proteínas p21 e p27 que atuam se ligando a vários complexos Cdk/ciclina e na transmissão de sinal, respectivamente, se utilizam da ID e do envelhecimento induzido por ligação com seus parceiros

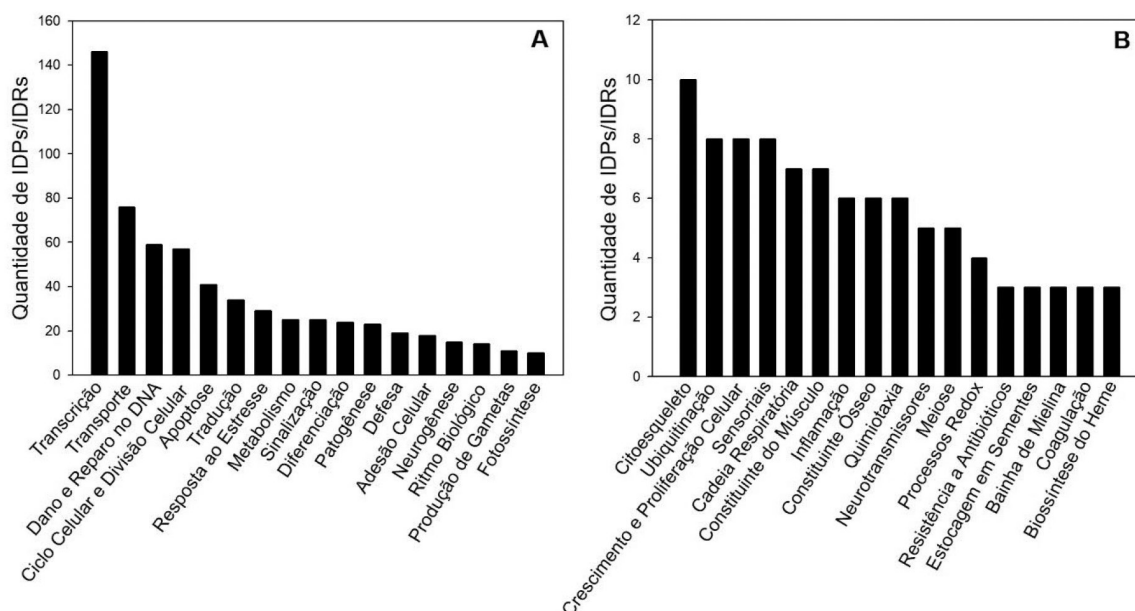


Figura 3. Processos biológicos envolvendo proteínas com ID. (A) São apresentados os processos biológicos mais prevalentes contendo até pelo menos 10 proteínas. (B) Processos biológicos menos prevalentes em proteínas do DisProt, contendo entre 10 e 3 proteínas.

para exercerem as suas múltiplas funções.⁶⁹ As IDPs/IDRs participam não só do processo de apoptose como também de outros tipos de morte celular programada, tal como a autofagia e a necrose.⁷⁰ Importantes IDPs/IDRs que participam de processos de morte celular programada são a p53, BRCA1, proteínas da família Bcl-2, proteína de ligação ao CREB, c-Myc e ARTS, por exemplo.⁷⁰

A análise da base de dados DisProt indicou proteínas contendo ID atuando ainda em processos de controle do ritmo biológico, neurogênese e produção de gametas (principalmente na espermatogênese) em animais (Figura 3A). Em plantas foram encontradas IDPs/IDRs relacionadas com processos de defesa, resposta ao estresse, estocagem em sementes e fotossíntese (Figuras 3A e B). Em vírus e bactérias, proteínas com ID foram encontradas participando de processos envolvidos com a patogênese, tais como invasão, interação com o hospedeiro, e resistência a antibióticos, por exemplo. As proteínas presentes no DisProt foram ainda relacionadas a processos metabólicos e de biossíntese de lipídeos, carboidratos, proteínas, nucleotídeos, vitaminas, neurotransmissores e grupos prostéticos, como o heme, por exemplo.

3.4. ID e Promiscuidade

Uma das principais características de proteínas que possuem ID é a promiscuidade, o que permite a interação com múltiplos parceiros celulares. Dentre as proteínas presentes no DisProt, foi analisado o número de parceiros descritos a fim de avaliar a tendência a promiscuidade dessas proteínas. A Figura 4 apresenta a distribuição de IDPs/IDRs pela quantidade de parceiros descritos. Dentre os dados obtidos, foi observado que grande parte dessas proteínas não possuem parceiro conhecido. Dentre as que possuem parceiros conhecidos, a maior parte possui mais que um parceiro celular. Também foi observado proteínas com alta promiscuidade, apresentando 10, 14 e até 17 parceiros conhecidos. É importante ressaltar que, considerando que

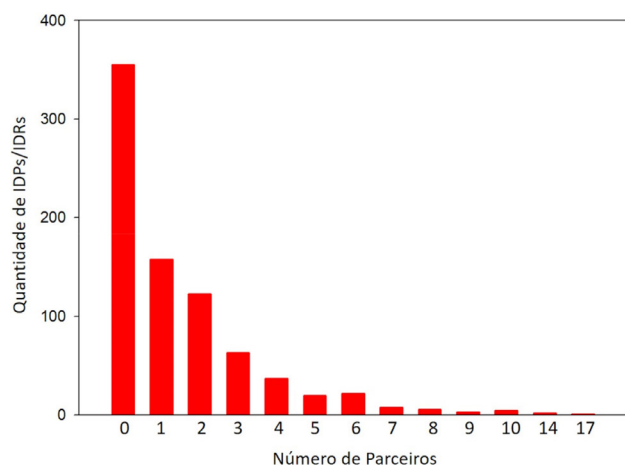


Figura 4. Avaliação da promiscuidade das IDPs/IDRs descritas. Apresenta-se no gráfico a quantidade de proteínas descritas no DisProt possuindo diferentes números de parceiros.

grande quantidade de proteínas com ID não possui parceiro conhecido e mesmo as que possuem podem possuir mais parceiros ainda não relatados, a promiscuidade real dessas proteínas pode ser ainda maior do que o observado por esses dados.

A promiscuidade apresentada por IDPs/IDRs é uma característica intimamente relacionada com as propriedades físico-químicas peculiares apresentadas por essas proteínas. A estrutura adotada por uma proteína é algo determinado pela sua composição de aminoácidos.⁷¹ IDPs/IDRs são ricas em resíduos polares e carregados, enquanto apresentam poucos resíduos hidrofóbicos.²⁸ Essa composição não favorece a formação de um núcleo hidrofóbico e, conseqüentemente de uma proteína enovelada.⁷² Em vez disso, IDPs/IDRs apresentam alta entropia conformacional, o que leva a uma plasticidade estrutural.^{72,73} Dessa forma, a capacidade de adotar múltiplas conformações pode permitir a ligação dessas proteínas com múltiplos parceiros, permitindo assim a promiscuidade.^{3,5}

A promiscuidade é uma característica que apresenta algumas vantagens e desvantagens. A grande vantagem da promiscuidade possuída por proteínas com ID é a sua participação como eixos de conexão em redes, tal como as possuídas por processos celulares de crescimento, geração de energia, divisão celular e sinalização.⁷⁴ A ID é útil para a atuação nessas redes de diferentes formas: 1) como o eixo que se liga com múltiplos parceiros; 2) se ligando com proteínas estruturadas que atuam como o eixo; 3) atuando como regiões flexíveis que conectam domínios funcionais, permitindo a promiscuidade.⁷⁴ O lado negativo da promiscuidade é a sua relação com o surgimento de doenças, principalmente quando tais proteínas são superexpressas.⁷⁵ Devido a essa possibilidade de interação com parceiros indesejados, proteínas com ID são altamente reguladas para estarem numa concentração adequada que permita uma sinalização fidedigna.⁷⁶

3.5. Presença de Regiões de Transição Desordem-Ordem, Quantidade e Tamanho de Regiões Desordenadas em Proteínas com ID

A presença de regiões apresentando transição desordem-ordem foi uma das características de IDPs/IDRs que também foi avaliada neste trabalho. Analisando os resultados obtidos, foi observado que 225 das proteínas analisadas, o equivalente a 28% do total, possuíam regiões de transição desordem-ordem determinadas. As regiões de transição desordem-ordem possuem uma importante participação na interação de IDPs/IDRs com seus parceiros e a sua função.⁶ Os MoRFs constituem regiões curtas presentes em proteínas com ID que sofrem transição desordem-ordem os quais são estabilizados pela ligação com um parceiro.⁷⁷ Uma característica do processo de transição de desordem-ordem é que uma alta especificidade promove uma baixa afinidade de ligação, e vice e versa.⁷⁸ Modelos teóricos sugerem ainda como característica de transições desordem-ordem

a possibilidade dessas regiões se ligarem com múltiplos parceiros devido ao empacotamento alternativo.⁷⁸ A transição desordem-ordem também possui uma contribuição na cinética de ligação e desacoplamento dos complexos.⁷⁸

Adicionalmente, foi avaliada a quantidade de proteínas que possuíam mais de uma IDR na sua estrutura. Dentre as proteínas analisadas, 537 apresentavam apenas uma IDR, o que corresponde a cerca de 67% do total. Por outro lado, 266 proteínas (33%) apresentavam mais de uma IDR na sua estrutura. Outra análise realizada foi a investigação do tamanho das IDRs. Neste trabalho, foi considerado como IDPs/IDRs longas aquelas com 30 resíduos ou mais, e abaixo desse valor como sendo IDRs curtas. Como mostrado na Figura 5, foi verificado que 55,17% delas possuíam IDRs longas, 32,63% possuíam IDRs curtas e 12,20% possuíam IDRs longas e curtas em sua estrutura. Em conjunto, esses dados apontam que a maior parte das proteínas analisadas possuem uma única IDR longa em sua estrutura.

Alguns estudos têm proposto a classificação de proteínas com ID quanto ao percentual de resíduos desordenados e o comprimento da região de desordem mais longa na proteína.^{42,79} Embora não haja um parâmetro definido para classificar uma IDP/IDR quanto ao tamanho, a grande parte dos trabalhos publicados consideram dois tipos de regiões desordenadas: IDRs curtas (abaixo de 30 aminoácidos) e IDRs longas (acima de 30 aminoácidos).⁸⁰⁻⁸³ As IDRs longas e curtas diferem não apenas quanto ao tamanho, mas também em relação à composição de aminoácidos.⁸³ IDRs longas, por exemplo, costumam ter um balanço entre os resíduos de carga positiva e negativa, o que confere uma carga absoluta mais baixa.⁸³ IDRs longas e curtas podem também variar quanto ao tipo de interação com os ligantes. Os motivos lineares são IDRs curtas, e, por conta do tamanho limitado, costumam realizar interações transientes com baixa afinidade.⁸ Por outro lado, IDRs longas possuem uma maior superfície de interação, o que pode ajudar a minimizar o custo entrópico da ligação pelo

ganho em entalpia.⁷³ Tal fato contribui para um ganho em afinidade em comparação com IDRs curtas.

3.6. Principais Métodos de Avaliação da ID em Proteínas

A frequência de utilização de métodos experimentais para análise e identificação de ID em proteínas presentes no DisProt também foi avaliada neste trabalho. Foi observado que RMN, cristalografia e difração de raios-x e CD são os métodos mais utilizados, correspondendo por 75% dos casos. SEC, proteólise, mobilidade em eletroforese em gel de poliacrilamida (PAGE) com dodecil- sulfato de sódio (SDS) (SDS-PAGE), SAXS, ultracentrifugação analítica (UCA), espectroscopia de infravermelho por transformada de Fourier (FTIR), DLS, fluorescência intrínseca e espectrometria de massa com troca de hidrogênio por deutério (Figura 6) também representam métodos bastante utilizados. Outros métodos experimentais utilizados com menor frequência estão apresentados na Tabela 1.

A principal técnica utilizada para a identificação e análise da ID em proteínas observada neste trabalho foi a RMN. Essa técnica consiste em um importante método de determinação estrutural, o qual permite avaliar a proteína em solução, se aproximando melhor das condições fisiológicas e, com isso, permite obter informações sobre a sua dinâmica e enovelamento.⁸⁴ Dessa forma, a RMN constitui uma excelente técnica para avaliação de diferentes propriedades de IDPs/IDRs. Algumas características apresentadas por proteínas com ID por RMN são a baixa dispersão da ressonância dos prótons e a sobreposição de sinais, que dificultam a interpretação dos resultados, enquanto que apresentam linhas estreitas no espectro, devido à baixa taxa de relaxação.⁸⁵ A RMN é útil para avaliar parâmetros hidrodinâmicos (através da utilização de gradientes de campo pulsado), como também para obter informações sobre conteúdos estruturais residuais (utilizando RMN heteronuclear).^{26,27,86}

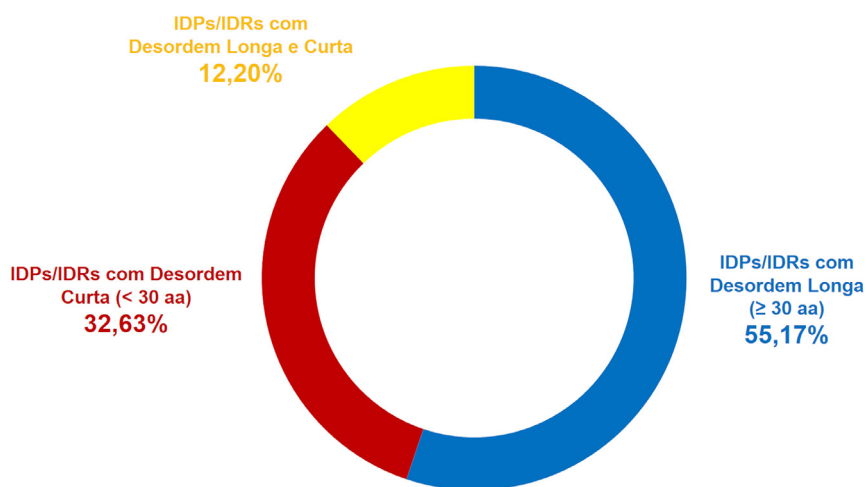


Figura 5. Presença de desordem longa e curta em IDPs/IDRs. As proteínas presentes na base de dados DisProt foram divididas entre as que possuíam desordem longa, curta ou ambos os tipos na mesma sequência. Foi considerado como desordem longa aquelas possuindo regiões com 30 ou mais resíduos desordenados

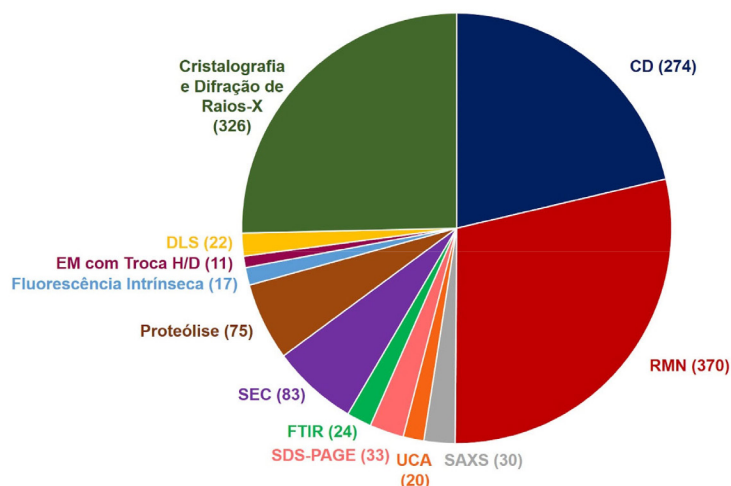


Figura 6. Principais métodos utilizados para identificação e análise da ID em proteínas. Os métodos mais utilizados para a identificação e análise da ID em proteínas presentes no DisProt estão apresentados no gráfico contendo o número de proteínas analisadas por cada método. Apenas os métodos que foram utilizados para identificar e analisar a ID em mais de 10 proteínas foram considerados.

Tabela 1. Métodos Experimentais Utilizados com Menor Frequência para Identificação e Análise da ID em proteínas do DisProt.

Método Utilizado	Número de Proteínas
Calorimetria Diferencial de Varredura (DSC)	9
Polarização de Fluorescência	7
Espectroscopia por Marcação de Spin Sítio Dirigida (SDSL)	5
Imunoquímica	5
Dicroísmo Circular utilizando radiação síncrotron (SRCD)	4
Viscometria	3
Fluorescência Extrínseca	3
Microscopia de Força Atômica	3
Atividade Óptica Raman (ROA)	3
Espectrometria de Massas de Ressonância Ciclotrônica de Íons por Transformada de Fourier com Fonte de Ionização por Eletrospray (ESI-FT-ICR-MS)	3
Espectrometria de Massas de Alta Resolução	3
Microscopia Eletrônica de Sombras Rotativas	3
Supressão Dinâmica da Fluorescência	2
Transferência de Energia por Ressonância de Förster (FRET)	2
Espectroscopia Vibracional de Cisteínas Cianiladas	1
Ensaio de Atividade Enzimática Radio-Marcada	1
Espectroscopia Raman	1
Espalhamento de Luz Estático	1
Espalhamento de Nêutrons a Baixo Ângulo (SANS)	1

A cristalografia de proteínas com difração por raios-x é o principal método de determinação estrutural para estudo de proteínas, embora tenha sido o segundo mais observado pelas análises deste trabalho. Embora seja muito utilizada para estudar proteínas estruturadas, no caso das proteínas com ID (as quais existem num conjunto dinâmico de conformações) a cristalização é dificultada, embora possa ocorrer em alguns casos fornecendo dados pouco representativos.⁸⁶ A alta flexibilidade proporcionada pela ID leva a uma dispersão de raios-x não-coerente, tornando-os

não observáveis por essa técnica, sendo geralmente definidos como as regiões que faltam densidade eletrônica.²⁶

O CD é o nome dado ao fenômeno observado pela diferença de absorção de componentes de luz circularmente polarizada à direita e à esquerda por cromóforos quirais ou situados em ambientes quirais.⁸⁷ Essa técnica pode ser empregada para avaliar diferentes questões em IDPs/IDRs. O CD pode ser utilizado para avaliar o conteúdo de estrutura secundária na faixa do UV-distante (240 – 180 nm), o qual está relacionado com a absorção pela ligação

peptídica e permite distinguir entre os tipos de estrutura secundária.⁸⁷ IDPs geram um espectro característico com um sinal negativo intenso próximo a 200 nm, um sinal quase insignificante a 222 nm e um sinal próximo a zero a 185 nm.²⁷ Essa técnica também pode ser utilizada para observar mudanças conformacionais decorrentes de condições, tal como pH, temperatura e ligação com parceiros, por exemplo.⁸⁷ Ela pode ser ainda utilizada para avaliar a sua estabilidade.⁸⁷ Informações sobre a estrutura terciária também podem ser obtidas avaliando o espectro no UV-próximo (250 – 350 nm), o qual reflete a simetria do ambiente dos resíduos aromáticos.²⁶ A avaliação do espectro de proteínas parcialmente enoveladas no UV-próximo resulta em baixa intensidade e complexidade.²⁷

O FTIR é outra importante técnica espectroscópica para o estudo de proteínas com ID, sendo uma ferramenta poderosa principalmente para obter informações sobre a estrutura secundária e agregação.⁸⁶ A observação do espectro de absorção na região do infravermelho médio (4000 – 1000 cm^{-1}), em especial na banda de absorção amida I (1600 – 1700 cm^{-1}), que corresponde ao alongamento C=O da ligação peptídica, permite identificar os tipos de estrutura secundária.⁸⁶ O interessante dessa técnica é que ela permite analisar tanto a proteína em solução como também no estado sólido, tal como filmes proteicos.⁸⁶ A espectroscopia de fluorescência intrínseca se utiliza de cromóforos presentes na própria molécula, tais como os resíduos aromáticos (triptofano, tirosina e fenilalanina) de proteínas, sendo destes o triptofano o mais utilizado.²⁷ O triptofano tem um máximo de absorção a 280 nm, e o espectro de emissão é sensível ao ambiente e localização desse resíduo na proteína.^{27,86} No caso de IDPs se observa um desvio do espectro de emissão para a região do vermelho, com máximo entre 340 a 353 nm, devido à localização desse resíduo na superfície.⁸⁶ Essa técnica ainda é útil para avaliar transições estruturais.²⁷

O grau de compactação de uma proteína pode ser avaliado por meio da determinação de parâmetros hidrodinâmicos utilizando diferentes técnicas, tal como a SEC, a viscometria, o SAXS, a sedimentação e o espalhamento de luz dinâmico.²⁶ A SEC é uma técnica que permite avaliar o volume de eluição, o qual é dependente das dimensões hidrodinâmicas da proteína.⁸⁶ Construindo-se uma curva de calibração com proteínas de dimensões hidrodinâmicas conhecidas é possível estimar o raio de Stokes de uma proteína, o que pode ser utilizado para classificação conformacional de IDPs, como também para a separação e caracterização de diferentes estados conformacionais.⁸⁶ O SAXS é uma das mais importantes técnicas para avaliar as dimensões e formas de proteínas, estando relacionado com a intensidade do espalhamento.²⁷ Essa técnica permite a determinação do raio de giro, o qual dá uma ideia do seu tamanho, por meio da aproximação de Guinier.^{27,86} O gráfico de Kratky também pode ser utilizado para distinguir IDPs de proteínas globulares, uma vez que proteínas desenoveladas apresentam um platô, enquanto proteínas globulares apresentam um formato de sino com um máximo

definido.^{27,86} A ultracentrifugação analítica pode ser útil por meio de experimentos de velocidade de sedimentação para a determinação dos coeficientes de sedimentação e assim avaliar tamanho e forma de uma proteína.^{27,86} O DLS permite a determinação do coeficiente de difusão translacional, o qual é utilizado através da equação de Stokes-Einstein para a determinação do raio de Stokes da proteína.²⁷

As proteínas com ID também podem ser identificadas pela apresentação de outras características. Devido à maior acessibilidade ao solvente e alta flexibilidade, proteínas com ID sofrem degradação em estudos de proteólise limitada.^{26,27,86} Essas proteínas também apresentam uma mobilidade anormal em SDS-PAGE, o que resulta em uma massa aparente maior do que a esperada.²⁷ Regiões com ID também podem ser identificadas por meio de experimentos de troca de hidrogênio por deutério por meio de espectrometria de massas, por exemplo.⁸⁶ Por conta da acessibilidade dessas regiões observa-se uma maior troca de hidrogênio por deutério em IDPs/IDRs.⁸⁶

4. Considerações Finais

Neste trabalho foram apresentadas e discutidas informações sobre proteínas com ID, tais como conceitos, comportamento, distribuição nos seres vivos, funções exercidas, participação em doenças, além de formas de caracterização dessa propriedade. Utilizando dados obtidos de proteínas presentes na base de dados DisProt bem como uma vasta investigação da literatura, reunimos informações sobre diversos aspectos de ID em proteínas. Mostramos que a ID é uma característica apresentada por um vasto grupo de proteínas, estando presente em todos os reinos da vida, sendo de forma mais significativa em organismos mais complexos. Proteínas com ID possuem as mais variadas funções, entretanto, a sua plasticidade estrutural as tornam úteis principalmente em eventos de sinalização, devido a capacidade de adotar diferentes conformações e se ligar a múltiplos parceiros. A grande presença de proteínas com ID em seres humanos traz consigo a participação em muitas doenças e também o desafio de desenvolver fármacos capazes de ligar em tais proteínas. Por fim, mostramos as principais técnicas utilizadas para a caracterização de IDPs/IDRs bem como as principais características estruturais apresentadas por essas proteínas. Esperamos com esse trabalho ter reunido informações em diferentes níveis sobre a ID em proteínas que possam contribuir para a divulgação desse tema em crescente expansão.

Referências Bibliográficas

1. Radivojac, P.; Iakoucheva, L. M.; Oldfield, C. J.; Obradovic, Z.; Uversky, V. N.; Dunker, A. K.; Intrinsic disorder and functional proteomics. *Biophysical Journal* **2007**, 92, 1439. [[Crossref](#)] [[PubMed](#)]

2. Tompa, P.; Intrinsically disordered proteins: a 10-year recap. *Trends in Biochemical Sciences* **2012**, *37*, 509. [[Crossref](#)] [[PubMed](#)]
3. Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N.; Introducing protein intrinsic disorder. *Chemical Reviews* **2014**, *114*, 6561. [[Crossref](#)] [[PubMed](#)]
4. van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M.; Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews* **2014**, *114*, 6589. [[Crossref](#)] [[PubMed](#)]
5. Tompa, P.; The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters* **2005**, *579*, 3346. [[Crossref](#)] [[PubMed](#)]
6. Uversky, V. N.; Unusual biophysics of intrinsically disordered proteins. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2013**, *1834*, 932. [[Crossref](#)] [[PubMed](#)]
7. Mohan, A.; Oldfield, C. J.; Radivojac, P.; Vacic, V.; Cortese, M. S.; Dunker, A. K.; Uversky, V. N.; Analysis of molecular recognition features (MoRFs). *Journal of Molecular Biology* **2006**, *362*, 1043. [[Crossref](#)] [[PubMed](#)]
8. Davey, N. E.; Roey, K. V.; Weatheritt, R. J.; Toedt, G.; Uyar, B.; Altenberg, B.; Budd, A.; Diella, F.; Dinkel, H.; Gibson, T. J.; Attributes of short linear motifs. *Molecular BioSystems* **2012**, *8*, 268. [[Crossref](#)] [[PubMed](#)]
9. Dyson, H. J.; Wright, P. E.; Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* **2005**, *6*, 197. [[Crossref](#)] [[PubMed](#)]
10. Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradović, Z.; Intrinsic disorder and protein function. *Biochemistry* **2002**, *41*, 6573. [[Crossref](#)] [[PubMed](#)]
11. Tompa, P.; Intrinsically unstructured proteins. *Trends in Biochemical Sciences* **2002**, *27*, 527. [[Crossref](#)] [[PubMed](#)]
12. Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T.; Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology* **2004**, *337*, 635. [[Crossref](#)] [[PubMed](#)]
13. Uversky, V. N.; A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Science* **2013**, *22*, 693. [[Crossref](#)] [[PubMed](#)]
14. Uversky, V. N.; Oldfield, C. J.; Dunker, A. K.; Intrinsically disordered proteins in human diseases: introducing the D² concept. *Annual Review of Biophysics* **2008**, *37*, 215. [[Crossref](#)] [[PubMed](#)]
15. Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradović, Z.; Dunker, A. K.; Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of Molecular Biology* **2002**, *323*, 573. [[Crossref](#)] [[PubMed](#)]
16. Mark, W. Y.; Liao, J. C.; Lu, Y.; Ayed, A.; Laister, R.; Szymczyna, B.; Chakrabarty, A.; Arrowsmith, C. H.; Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein–protein and protein–DNA interactions?. *Journal of Molecular Biology* **2005**, *345*, 275. [[Crossref](#)] [[PubMed](#)]
17. Kumar, D.; Sharma, N.; Giri, R.; Therapeutic interventions of cancers using intrinsically disordered proteins as drug targets: c-Myc as model system. *Cancer Informatics* **2017**, *16*, 1176935117699408. [[Crossref](#)] [[PubMed](#)]
18. Cheng, Y.; LeGall, T.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N.; Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry* **2006**, *45*, 10448. [[Crossref](#)] [[PubMed](#)]
19. Jaikaran, E. T.; Higham, C. E.; Serpell, L. C.; Zurdo, J.; Gross, M.; Clark, A.; Fraser, P. E.; Identification of a novel human islet amyloid polypeptide β -sheet domain and factors influencing fibrillogenesis. *Journal of Molecular Biology* **2001**, *308*, 515. [[Crossref](#)] [[PubMed](#)]
20. Riek, R.; Hornemann, S.; Wider, G.; Glockshuber, R.; Wüthrich, K.; NMR characterization of the full-length recombinant murine prion protein, mPrP (23–231). *FEBS Letters* **1997**, *413*, 282. [[PubMed](#)]
21. Theillet, F. X.; Binolfi, A.; Bekei, B.; Martorana, A.; Rose, H. M.; Stuver, M.; Verzini S.; Lorenz, D.; Rossum, M. V.; Goldfarb, D.; Selenko, P.; Structural disorder of monomeric α -synuclein persists in mammalian cells. *Nature* **2016**, *530*, 45. [[Crossref](#)] [[PubMed](#)]
22. Korsak, M.; Kozyreva, T.; Beta amyloid hallmarks: From intrinsically disordered proteins to alzheimer's disease. *Advances in Experimental Medicine Biology* **2015**, *871*, 401. [[Crossref](#)] [[PubMed](#)]
23. Schweers, O.; Schönbrunn-Hanebeck, E.; Marx, A.; Mandelkow, E.; Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. *Journal of Biological Chemistry* **1994**, *269*, 24290. [[Link](#)] [[PubMed](#)]
24. Zhang, Y.; Cao, H.; Liu, Z.; Binding cavities and druggability of intrinsically disordered proteins. *Protein Science* **2015**, *24*, 688. [[Crossref](#)] [[PubMed](#)]
25. Joshi, P.; Vendruscolo, M.; Druggability of intrinsically disordered proteins. In: Intrinsically disordered proteins studied by NMR spectroscopy. *Advances in Experimental Medicine Biology* **2015**, *870*, 383. [[Crossref](#)] [[PubMed](#)]
26. Uversky, V. N.; Natively unfolded proteins: a point where biology waits for physics. *Protein Science* **2002**, *11*, 739. [[PubMed](#)]
27. Receveur-Bréchet, V.; Bourhis, J. M.; Uversky, V. N.; Canard, B.; Longhi, S.; Assessing protein disorder and induced folding. *Proteins: Structure, Function, and Bioinformatics* **2006**, *62*, 24. [[Crossref](#)] [[PubMed](#)]
28. Uversky, V. N.; Gillespie, J. R.; Fink, A. L.; Why are “natively unfolded” proteins unstructured under physiologic conditions?. *Proteins: Structure, Function, and Bioinformatics* **2000**, *41*, 415. [[Crossref](#)] [[PubMed](#)]
29. Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hippos, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z.; Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling* **2001**, *19*, 26. [[Crossref](#)] [[PubMed](#)]
30. Linding, R.; Russell, R. B.; Neduva, V.; Gibson, T. J.; GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Research* **2003**, *31*, 3701. [[Crossref](#)] [[PubMed](#)]

31. Galzitskaya, O. V.; Garbuzynskiy, S. O.; Lobanov, M. Y.; Prediction of natively unfolded regions in protein chains. *Molecular Biology* **2006**, *40*, 298. [[Crossref](#)] [[PubMed](#)]
32. Dosztányi, Z.; Csizsmok, V.; Tompa, P.; Simon, I.; IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21*, 3433. [[Crossref](#)] [[PubMed](#)]
33. Romero, P.; Obradovic, Z.; Kissinger, C.; Villafranca, J. E.; Dunker, A. K.; Identifying disordered regions in proteins from amino acid sequence. In Proceedings of International Conference on Neural Networks (ICNN'97), *IEEE* **1997**, *1*, 90. [[Crossref](#)]
34. Garner, E.; Romero, P.; Dunker, A. K.; Brown, C.; Obradovic, Z.; Predicting binding regions within disordered proteins. *Genome Informatics* **1999**, *10*, 41. [[PubMed](#)]
35. Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B.; Protein disorder prediction: implications for structural proteomics. *Structure* **2003**, *11*, 1453. [[Crossref](#)] [[PubMed](#)]
36. Walsh, I.; Martin, A. J.; Di Domenico, T.; Tosatto, S. C.; ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* **2012**, *28*, 503. [[Crossref](#)] [[PubMed](#)]
37. Xue, B.; Dunbrack, R. L.; Williams, R. W.; Dunker, A. K.; Uversky, V. N.; PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2010**, *1804*, 996. [[Crossref](#)] [[PubMed](#)]
38. Ishida, T.; Kinoshita, K.; Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* **2008**, *24*, 1344. [[Crossref](#)] [[PubMed](#)]
39. Piovesan, D.; Tabaro, F.; Mičetić, I.; Necci, M.; Quaglia, F.; Oldfield, C. J.; Aspromonte, M. C.; Davey, N. E.; Davidović, R.; Dosztányi, Z.; Elofsson, A.; Gasparini, A.; Hatos, A.; Kajava, A. V.; Kalmar, L.; Leonardi, E.; Lazar, T.; Macedo-Ribeiro, S.; Macossay-Castillo, M.; Meszaros, A.; Minervini, G.; Murvai, N.; Pujols, J.; Roche, D. N.; Salladini, E.; Schad, E.; Schramm, A.; Szabo, B.; Tantos, A.; Tonello, F.; Tsirigos, K. D.; Veljković, N.; Ventura, S.; Vranken, W.; Warholm, P.; Uversky, V. N.; Dunker, A. K.; Longhi, S.; Tompa, P.; Tosatto, S. C. E.; DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Research* **2017**, *45*, D219. [[Crossref](#)] [[PubMed](#)]
40. Pancsa, R.; Tompa, P.; Structural Disorder in Eukaryotes. *PLoS one* **2012**, *7*, e34687. [[Crossref](#)] [[PubMed](#)]
41. Pentony, M. M.; Jones, D. T.; Modularity of intrinsic disorder in the human proteome. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 212. [[Crossref](#)] [[PubMed](#)]
42. Deiana, A.; Forcelloni, S.; Porrello, A.; Giansanti, A.; Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS one* **2019**, *14*, e0217889. [[Crossref](#)] [[PubMed](#)]
43. Pietrosevoli, N.; García-Martín, J. A.; Solano, R.; Pazos, F.; Genome-wide analysis of protein disorder in Arabidopsis thaliana: implications for plant environmental adaptation. *PLoS one* **2013**, *8*, e55524. [[Crossref](#)] [[PubMed](#)]
44. Mouillon, J. M.; Gustafsson, P.; Harryson, P.; Structural investigation of disordered stress proteins. Comparison of full-length dehydrins with isolated peptides of their conserved segments. *Plant Physiology* **2006**, *141*, 638. [[Crossref](#)] [[PubMed](#)]
45. Sun, X.; Jones, W. T.; Rikkerink, E. H.; GRAS proteins: the versatile roles of intrinsically disordered proteins in plant signalling. *Biochemical Journal* **2012**, *442*, 1. [[Crossref](#)] [[PubMed](#)]
46. Jensen, M. K.; Kjaersgaard, T.; Nielsen, M. M.; Galberg, P.; Petersen, K.; O'shea, C.; Skriver, K.; The Arabidopsis thaliana NAC transcription factor family: structure–function relationships and determinants of ANAC019 stress signalling. *Biochemical Journal* **2010**, *426*, 183. [[Crossref](#)] [[PubMed](#)]
47. Yoon, M. K.; Shin, J.; Choi, G.; Choi, B. S.; Intrinsically unstructured N-terminal domain of bZIP transcription factor HY5. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 856. [[Crossref](#)] [[PubMed](#)]
48. Alves de Souza, S. M.; Fernandes, T. V. A.; Kalume, D. E.; Pascutti, P. G.; de Souza, T. L. F.; Physicochemical and structural properties of lunasin revealed by spectroscopic, chromatographic and molecular dynamics approaches. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2020**, *1868*, 140440. [[Crossref](#)]
49. Galvez, A. F.; De Lumen, B. O.; A soybean cDNA encoding a chromatin-binding peptide inhibits mitosis in mammalian cells. *Nature Biotechnology* **1999**, *17*, 495. [[Crossref](#)] [[PubMed](#)]
50. Tompa, P.; Dosztányi, Z.; Simon, I.; Prevalent Structural Disorder in E. coli and S. cerevisiae Proteomes. *Journal of Proteome Research* **2006**, *5*, 1996. [[Crossref](#)] [[PubMed](#)]
51. Pavlović-Lažetić, G. M.; Mitić, N. S.; Kovačević, J. J.; Obradović, Z.; Malkov, S. N.; Beljanski, M. V.; Bioinformatics analysis of disordered proteins in prokaryotes. *BMC Bioinformatics* **2011**, *12*, 66. [[Crossref](#)] [[PubMed](#)]
52. Mitić, N. S.; Malkov, S. N.; Kovačević, J. J.; Pavlović-Lažetić, G. M.; Beljanski, M. V.; Structural disorder of plasmid-encoded proteins in Bacteria and Archaea. *BMC Bioinformatics* **2018**, *19*, 158. [[Crossref](#)] [[PubMed](#)]
53. Xue, B.; Williams, R. W.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N.; Archaic chaos: intrinsically disordered proteins in Archaea. *BMC Systems Biology* **2010**, *4*, S1. [[Crossref](#)] [[PubMed](#)]
54. Feng, Z. P.; Zhang, X.; Han, P.; Arora, N.; Anders, R. F.; Norton, R. S.; Abundance of intrinsically unstructured proteins in *P. falciparum* and other apicomplexan parasite proteomes. *Molecular and Biochemical Parasitology* **2006**, *150*, 256. [[Crossref](#)] [[PubMed](#)]
55. Mohan, A.; Sullivan Jr W. J.; Radivojac, P.; Dunker, A. K.; Uversky, V. N.; Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. *Molecular BioSystems* **2008**, *4*, 328. [[Crossref](#)] [[PubMed](#)]
56. Peng, Z.; Yan, J.; Fan, X.; Mizianty, M. J.; Xue, B.; Wang, K.; Hu, G.; Uversky, V. N.; Kurgan, L.; Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cellular and Molecular Life Sciences* **2015**, *72*, 137. [[Crossref](#)] [[PubMed](#)]
57. Xue, B.; Dunker, A. K.; Uversky, V. N.; Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *Journal of Biomolecular Structure and Dynamics* **2012**, *30*, 137. [[Crossref](#)] [[PubMed](#)]

58. Berezovsky, I. N.; The diversity of physical forces and mechanisms in intermolecular interactions. *Physical Biology* **2011**, *8*, 035002. [[Crossref](#)] [[PubMed](#)]
59. Goh, G. K. M.; Dunker, A. K.; Foster, J. A.; Uversky, V. N.; HIV Vaccine Mystery and Viral Shell Disorder. *Biomolecules* **2019**, *9*, 178. [[Crossref](#)] [[PubMed](#)]
60. Goh, G. K. M.; Dunker, A. K.; Uversky, V. N.; Correlating *Flavivirus* virulence and levels of intrinsic disorder in shell proteins: protective roles vs. immune evasion. *Molecular BioSystems* **2016**, *12*, 1881. [[Crossref](#)] [[PubMed](#)]
61. Oldfield, C. J.; Dunker, A. K.; Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual Review of Biochemistry* **2014**, *83*, 553. [[Crossref](#)] [[PubMed](#)]
62. Korneta, I.; Bujnicki, J. M.; Intrinsic disorder in the human spliceosomal proteome. *PLoS Computational Biology* **2012**, *8*, e1002641. [[Crossref](#)] [[PubMed](#)]
63. Peng, Z.; Oldfield, C. J.; Xue, B.; Mizianty, M. J.; Dunker, A. K.; Kurgan, L.; Uversky, V. N.; A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. *Cellular and Molecular Life Sciences* **2014**, *71*, 1477. [[Crossref](#)] [[PubMed](#)]
64. Niklas, K. J.; Dunker, A. K.; Yruela, I.; The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *Journal of Experimental Botany* **2018**, *69*, 1437. [[Crossref](#)] [[PubMed](#)]
65. Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K.; Intrinsic disorder in transcription factors. *Biochemistry* **2006**, *45*, 6873. [[Crossref](#)] [[PubMed](#)]
66. Hough, L. E.; Dutta, K.; Sparks, S.; Temel, D. B.; Kamal, A.; Tetenbaum-Novatt, J.; Rout, M. P.; Cowburn, D.; The molecular mechanism of nuclear transport revealed by atomic-scale measurements. *Elife* **2015**, *4*, e10027. [[Crossref](#)] [[PubMed](#)]
67. Clancy, S.; DNA damage & repair: mechanisms for maintaining DNA integrity. *Nature Education* **2008**, *1*, 103. [[Link](#)]
68. Iakoucheva, L. M.; Kimzey, A. L.; Masselon, C. D.; Bruce, J. E.; Garner, E. C.; Brown, C. J.; Dunker, A. K.; Smith, R. F.; Ackerman, E. J.; Identification of intrinsic order and disorder in the DNA repair protein XPA. *Protein Science* **2001**, *10*, 560. [[PubMed](#)]
69. Yoon, M. K.; Mitrea, D. M.; Ou, L.; Kriwacki, R. W.; Cell cycle regulation by the intrinsically disordered proteins p21 and p27. *Biochemical Society Transactions* **2012**, *40*, 981. [[Crossref](#)] [[PubMed](#)]
70. Peng, Z.; Xue, B.; Kurgan, L.; Uversky, V. N.; Resilience of death: intrinsic disorder in proteins involved in the programmed cell death. *Cell Death & Differentiation* **2013**, *20*, 1257. [[Crossref](#)] [[PubMed](#)]
71. Anfinsen, C. B.; Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223. [[Crossref](#)] [[PubMed](#)]
72. Forman-Kay, J. D.; Mittag, T.; From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure* **2013**, *21*, 1492. [[Crossref](#)] [[PubMed](#)]
73. Flock, T.; Weatheritt, R. J.; Latysheva, N. S.; Babu, M. M.; Controlling entropy to tune the functions of intrinsically disordered regions. *Current Opinion in Structural Biology* **2014**, *26*, 62. [[Crossref](#)]
74. Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N.; Flexible nets: the roles of intrinsic disorder in protein interaction networks. *The FEBS Journal* **2005**, *272*, 5129. [[Crossref](#)] [[PubMed](#)]
75. Vavouri, T.; Semple, J. I.; Garcia-Verdugo, R.; Lehner, B.; Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* **2009**, *138*, 198. [[Crossref](#)] [[PubMed](#)]
76. Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J.; Intrinsically disordered proteins: regulation and disease. *Current Opinion in Structural Biology* **2011**, *21*, 432. [[Crossref](#)] [[PubMed](#)]
77. Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Romero, P.; Uversky, V. N.; Dunker, A. K.; Coupled folding and binding with α -helix-forming molecular recognition elements. *Biochemistry* **2005**, *44*, 12454. [[Crossref](#)] [[PubMed](#)]
78. Dunker, A. K.; Garner, E.; Guillot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z.; Kissinger, C.; Villafranca, J. E.; Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pacific Symposium on Biocomputing* **1998**, *3*, 473. [[Link](#)] [[PubMed](#)]
79. Deiana, A.; Forcelloni, S.; Porrello, A.; Giansanti, A.; New classification of intrinsic disorder in the Human proteome. *bioRxiv* **2018**, 446351. [[Crossref](#)]
80. Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K.; Predicting intrinsic disorder from amino acid sequence. *Proteins: Structure, Function, and Bioinformatics* **2003**, *53*, 566. [[Crossref](#)] [[PubMed](#)]
81. Peng, K.; Radivojac, P.; Vucetic, S.; Dunker, A. K.; Obradovic, Z.; Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **2006**, *7*, 208. [[Crossref](#)] [[PubMed](#)]
82. Oates, M. E.; Romero, P.; Ishida, T.; Ghalwash, M.; Mizianty, M. J.; Xue, B.; Dosztányi, Z.; Uversky, V. N.; Obradovic, Z.; Kurgan, L.; Dunker, A. K.; Gough, J.; D²P²: database of disordered protein predictions. *Nucleic Acids Research* **2012**, *41*, D508. [[Crossref](#)] [[PubMed](#)]
83. Radivojac, P.; Obradovic, Z.; Smith, D. K.; Zhu, G.; Vucetic, S.; Brown, C. J.; Lawson, J. D.; Dunker, A. K.; Protein flexibility and intrinsic disorder. *Protein Science* **2004**, *13*, 71. [[Crossref](#)] [[PubMed](#)]
84. Wuthrich, K.; Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* **1989**, *243*, 45. [[PubMed](#)]
85. Kosol, S.; Contreras-Martos, S.; Cedeño, C.; Tompa, P.; Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **2013**, *18*, 10802. [[Crossref](#)] [[PubMed](#)]
86. Uversky, V.; Longhi, S.; Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation. *John Wiley & Sons* 2010, 744. [[Crossref](#)]
87. Kelly, S. M.; Price, N. C.; The use of circular dichroism in the investigation of protein structure and function. *Current Protein and Peptide Science* **2000**, *1*, 349. [[Crossref](#)] [[PubMed](#)]