

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

INSTITUTO DE BIOTECNOLOGIA

CURSO DE BIOTECNOLOGIA

Prospecção e caracterização de desordem intrínseca em proteínas de vírus associados a  
doenças humanas

Millena Almeida Resende

Monografia apresentada à Coordenação  
do Curso de Biotecnologia, da  
Universidade Federal de Uberlândia, para  
obtenção do grau de Bacharel em  
Biotecnologia.

Uberlândia - MG

Dezembro – 2018

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

INSTITUTO DE BIOTECNOLOGIA

CURSO DE BIOTECNOLOGIA

Prospecção e caracterização de desordem intrínseca em proteínas de vírus associados a  
doenças humanas

Millena Almeida Resende

Nilson Nicolau Junior

Orientador

Monografia apresentada à Coordenação  
do Curso de Biotecnologia, da  
Universidade Federal de Uberlândia, para  
obtenção do grau de Bacharel em  
Biotecnologia.

Uberlândia - MG

Dezembro – 2018

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

INSTITUTO DE BIOTECNOLOGIA

CURSO BIOTECNOLOGIA

Prospecção e caracterização de desordem intrínseca em proteínas de vírus associados a  
doenças humanas

Millena Almeida Resende

Nilson Nicolau Junior

Orientador

Homologado pela coordenação do Curso  
de Biotecnologia em \_\_/\_\_/\_\_

Uberlândia - MG

Julho – 2018

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

INSTITUTO DE BIOTECNOLOGIA

CURSO BIOTECNOLOGIA

Prospecção e caracterização de desordem intrínseca em proteínas de vírus associados a  
doenças humanas

Millena Almeida Resende

Aprovado pela Banca Examinadora em:    /    /    Nota: \_\_\_\_

Nome e assinatura do Presidente da Banca Examinadora

Uberlândia,    de    de

Aos meus pais Célio e Edilaine por acreditarem em meu potencial, aos meus avós Orlando e Anália por todo o apoio, à minha madrinha Helena por ser tão especial para mim.

## AGRADECIMENTOS

Agradeço a Deus por sempre me mostrar o caminho mais excelente. Agradeço ao professor Dr. Nilson por toda a orientação e ensinamentos durante o estudo e também pela oportunidade de ter sido sua aluna no laboratório de Modelagem Molecular na Universidade Federal de Uberlândia.

Agradeço também a todos os professores que fizeram parte de minha caminhada compartilhando seus saberes e seu tempo em minha formação. Em especial as professoras Jane, Luciane, Dra. Hebréia e Dra. Veridiana por me despertar a paixão pela ciência e pelo mundo acadêmico.

Agradeço a todos os meus familiares que tornaram meu caminho mais leve oferecendo-me seus valores e momentos de distração, principalmente meus tios e tias: Cenira, Osvaldino, Leontina, Antônio, Anália, Eustáquio e também meus padrinhos Helena e Osvaldino.

Agradeço também meu irmão Pedro e todos os colegas e amigos que tornam meus dias melhores. Agradeço a Universidade Federal de Uberlândia por todo o apoio oferecido a mim e ao ensino de tantos alunos. Agradeço também meus pais Célio e Edilaine, meus avós Orlando e Anália e minha tia Evânia, pessoas singulares em minha vida.

Muito obrigada!

## RESUMO

A desordem intrínseca pode ocorrer em partes de proteínas ou até mesmo na proteína inteira e se caracteriza pela ausência de uma estrutura tridimensional bem definida e alta flexibilidade em condições naturais e fisiológicas. As proteínas que apresentam esta característica parcial ou total são chamadas de proteínas intrinsecamente desestruturadas (IUPs) e na maioria das vezes estão relacionadas com funções de regulação celular. Nesse estudo, mais de 400 proteínas provenientes de diversos vírus associados a doenças em humanos foram caracterizadas quanto à quantidade de desordem na sequência polipeptídica, a sua localização, as funções, o processo biológico que elas fazem parte e predição de regiões de ligação. Os resultados mostraram que, de forma geral, a quantidade de desordem em si não é tão importante para determinada função, mas a localização e a flexibilidade que ela permite à proteína sim. Detectou-se que quanto maior a desordem proteica maior o número predito de regiões de ligação quando comparado a proteínas que possuem menor quantidade de desordem. Além disso, quase 50% delas possuem um nível muito baixo de desordem, sendo normalmente encontrada nas extremidades C e N terminal. Apenas 1% das proteínas possui 0% de desordem, mostrando, dessa forma uma alta frequência de IUPs em vírus associados a doenças em humanos.

**Palavras-chaves: bioinformática; predição; análise.**

## SUMÁRIO

1. INTRODUÇÃO.....	1
2. OBJETIVOS.....	3
2.1. Objetivo Geral.....	3
2.2. Objetivos Específicos.....	3
3. MATERIAIS E MÉTODOS.....	4
3.1. Vírus e sequências proteicas.....	4
3.2. Análise de desordem.....	5
3.2.1. DisCon.....	5
3.2.2. PONDR-FIT.....	6
3.2.3. ANCHOR.....	6
3.3. Anotação de proteínas.....	6
3.3.1. Vírus-Ploc.....	7
3.3.2. Blast2GO.....	7
4. RESULTADOS E DISCUSSÃO .....	7
4.1. Análise de desordem.....	8
4.1.1. Discon.....	8
4.1.2. PONDR-FIT.....	12
4.1.3. Anchor.....	14
4.2. Anotação de proteínas.....	16
4.2.1. Vírus-Ploc.....	16
4.2.2. Blast2GO.....	18
5. CONCLUSÃO.....	22
REFERÊNCIAS.....	24



## 1. INTRODUÇÃO

Proteínas são compostos orgânicos essenciais para a vida e estão relacionadas com inúmeras funções fisiológicas. São formadas através de ligações peptídicas entre aminoácidos e desempenham funções como a catalisação de reações químicas, regeneração tecidual, reações imunológicas e são indispensáveis para o crescimento e homeostasia de qualquer organismo. Com base na química se estabelece que sejam polímeros de alto peso molecular e que possuem suas características determinadas pela quantidade e espécie de cada aminoácido que a compõe, bem como pela ordem desses compostos na cadeia linear da estrutura proteica (BRASIL, 2014).

Por meio de forças intermoleculares formadas entre a sequência de aminoácidos níveis de estruturas proteicas são instituídos podendo ser descritos como estrutura primária, secundária, terciária ou quaternária. Comumente, se afirma haver uma estreita relação entre a estrutura e a função de cada proteína. Entretanto, a posterior descoberta de proteínas que não possuem uma estrutura secundária contrapõe a relação estrutura/função proposta. Tais proteínas são referidas como IUPs (*Intrinsically Unstructured Proteins*) ou IDPs (*Intrinsically Disorderd Proteins*) e são proteínas ou parte de proteínas que não possuem estrutura secundária definida em condições naturais e fisiológicas, porém ainda sim, tais proteínas exercem sua atividade biológica (TOMPA, 2005).

Um fato interessante a cerca das IUPs é que devido à ausência de estrutura secundária, os aminoácidos das IUPs tendem a possuir alta exteriorização, o que facilita atividades moduladoras como fosforilação, acetilação e ubiquitinação em suas cadeias polipeptídicas (TORRIERI, 2010). Ademais, em decorrência da alta exposição dos aminoácidos tem-se maior flexibilidade gerada que se torna um fator favorável no estudo e aperfeiçoamento de novas drogas, uma vez que regiões de desordem dispõem de potenciais sítios de ligações, inclusive sítios com possíveis interações com fármacos (GALEA et al., 2006).

Para se explicar a ocorrência de IUPs pode se apoiar na baixa quantidade de resíduos hidrofóbicos e uma alta presença de resíduos com cargas análogas em regiões desordenadas. Em decorrência disso a sequência passa a ter um caráter hidrofílico. Já em relação às cargas análogas tem-se uma alta repulsão eletrostática entre os resíduos de aminoácidos na

na mesma região. Logo, somando-se ambas as características, se nota uma resistência à formação de estruturas, impedindo, dessa forma a conformação proteica secundária e favorecendo o surgimento de IUPs (FERRON et al., 2006).

Outro fator que merece ser ressaltado devido a sua grande importância se trata da alta promiscuidade de regiões desordenadas. A falta de uma estrutura rígida tridimensional permite que proteínas desordenadas interajam com vários ligantes ao mesmo tempo. Dessa forma, proteínas provenientes de agentes infecciosos adquirem um aumento na capacidade de burlar nosso sistema imune a partir do momento em que proteínas do hospedeiro passam a se ligar a epítomos de proteínas exógenas, como por exemplo, proteínas virais. As ligações feitas tornam as proteínas exógenas difíceis de serem reconhecidas se tornando um facilitador para o desenvolvimento de doenças (UVERSKY; OLDFIELD; DUNKER, 2008).

Por essa razão, as IUPs vêm sendo associadas a doenças humanas como câncer, doenças cardiovasculares, doenças neurodegenerativas, diabetes, entre outras. A partir dessa associação, temos um parecer de que as doenças ocasionadas não ocorrem somente devido ao baixo dobramento das IUPs, mas também em decorrência de erros provenientes do processo de identificação e sinalização de importantes alvos intrínsecos ao desenvolvimento da doença (UVERSKY; OLDFIELD; DUNKER, 2008).

Nesse contexto, sabe-se também que proteínas virais tem a capacidade de interagir com vários componentes celulares do hospedeiro que estão presentes em membranas, citoplasmas ou núcleos de suas células. Todavia, a presença de desordem intrínseca pode ser uma forma facilitadora para as viroses sendo também a forma escolhida para as proteínas virais interagir com proteínas de diversos organismos (ALVES; CUNHA, 2012).

Com base na ressaltada importância das IUPs, torna-se claro, portanto a necessidade de estudos que venham a entender e caracterizar melhor proteínas que apresentam desordem intrínseca, já que esse pode ser um fator alvo no desenvolvimento de novos tratamentos para diversas doenças através, por exemplo, do estudo de interações proteína-ligante. Proposição essa que já fora reforçada por Alves e Cunha (2012): “Espera-se que o crescente conhecimento e dados sobre a caracterização e função de IDP possam-nos permitir desenvolver novas abordagens de tratamento e soluções que levem a um melhor controle de doenças virais em um futuro próximo”.

## 2. OBJETIVOS

### 2.1. Objetivo Geral

Caracterizar e analisar as regiões desordenadas em proteínas virais e estabelecer suas relações com doenças humanas provocadas por vírus através de ferramentas da Bioinformática.

### 2.2. Objetivos específicos

- Revisar a literatura sobre doenças humanas virais, desordem proteica e a aplicação da bioinformática no estudo de proteínas;
- Selecionar as sequências proteicas das proteínas de acordo com seu subtipo viral;
- Quantificar a desordem de cada uma das proteínas;
- Visualizar graficamente o local onde a desordem se encontra na sequência linear proteica;
- Avaliar a quantidade de sítios de ligações presente nas regiões desordenadas;
- Determinar a região subcelular que a proteína se encontra na célula infectada;
- Caracterizar as proteínas com base no Gene Ontology em termos de função e processo biológico;
- Correlacionar as IUPs com as consequências das viroses estudadas.

### 3. MATERIAIS E MÉTODOS

#### 3.1. Vírus e sequências proteicas

Os vírus selecionados para o estudo foram retirados do livro “Microterrores”, de Tony Hart (2009), com o acréscimo de outros tipos virais os quais, apresentam grande potencial para o desenvolvimento de doenças humanas.

As sequências primárias das proteínas foram obtidas no banco de dados UNIPROT situado em [www.uniprot.org](http://www.uniprot.org). As proteínas correspondentes a cada vírus foram retiradas de um único subtipo viral.

Os tipos mais recorrentes de todos os vírus foram escolhidos para a análise e estão listados na tabela a seguir:

Tabela 1. Códigos e linhagens correspondentes a cada vírus estudado.

<b>Vírus</b>	<b>Código</b>	<b>Linhagem</b>
<i>Human astrovirus-1</i>	HAstV-1	-
<i>Rotavirus A</i>	RV-A	Isolate RVA/Human/United Kingdom/A64/1987/G10P11[14]
<i>Dengue virus type 1</i>	DENV-1	Strain Brazil/97-11/1997
<i>Human immunodeficiency virus type 1 group M subtype B</i>	HIV-1	Isolate BRU/LAI
<i>Zaire ebolavirus</i>	ZEBOV	Strain Mayinga-76
<i>Influenza A virus</i>	H1N1	A/lvPR8/34
<i>Human T-cell leukemia virus 1</i>	HTLV-1	Strain Japan ATK-1 subtype A
<i>Hepatitis B virus genotype B2</i>	HBV-B	Isolate Vietnam/16091/1992
<i>Japanese encephalitis virus</i>	JEV	Strain SA-14
<i>Hepatitis E virus genotype 4</i>	HEV-4	Isolate Human/China/T1
<i>Hepatitis C virus genotype 1a</i>	HCV	Isolate H
<i>Chikungunya virus</i>	CHIKV	Strain 37997
<i>West Nile virus</i>	WNV	-

<i>Zika virus</i>	ZIKV	Strain Mr 766
<i>JC polyomavirus type 3A</i>	JCPyV	-
<i>Sapporo virus</i>	-	Strain Human/United Kingdom/Manchester/1993) (Hu/SV/Man/1993/UK
<i>Poliovirus type 1</i>	-	Strain Mahoney
<i>Mayaro virus</i>	MAYV	Strain Brazil
<i>Yellow fever virus</i>	YFV	Strain 17D vaccine
<i>Lassa virus</i>	LASV	Strain GA391
<i>Human parainfluenza 3 virus</i>	HPIV-3	Strain Wash/47885/57
<i>Human rhinovirus 1A</i>	HRV-1A	-
<i>Variola virus</i>	-	-
<i>Human cytomegalovirus</i>	HHV-5	Strain Towne
<i>Mumps virus</i>	MuV	Strain Miyahara vaccine
<i>Rabies virus</i>	RABV	Strain SAD B19
<i>Rubella virus</i>	RUBV	Strain RA27/3 vaccine
<i>Human respiratory syncytial virus B</i>	VSR	Strain B1
<i>Hantaan virus</i>	HTNV	Strain 76-118
<i>O'nyong-nyong virus</i>	ONNV	Strain Gulu
<i>Human papillomavirus type 16</i>	HPV-16	-
<i>Human papillomavirus type 18</i>	HPV-18	-
<i>Merkel cell polyomavirus</i>	MCPyV	-
<i>Epstein-Barr virus</i>	HHV	Strain B95-8

### 3.2. Análise de desordem

#### 3.2.1. DisCon

DisCon é uma ferramenta utilizada para quantificar a desordem em uma dada proteína. Com base na sequência proteica, 29 características complementares típicas de regiões desordenadas e um modelo baseado na

regressão de “ridge” tem-se como resultado a quantidade de desordem presente na biomolécula (MIZIANTY et al., 2011). Para o resultado é necessário que a sequência FASTA seja submetida. O resultado é simples e traz de forma direta a porcentagem de desordem na proteína submetida. A ferramenta está disponível gratuitamente e pode ser acessada em: <http://biomine.cs.vcu.edu/servers/DisCon/>

### 3.2.2. PONDR-FIT

PONDR-FIT é um meta-preditor que foi resultado da combinação de outros seis preditores utilizados individualmente (PONDR® VLXT, PONDR® VL3, PONDR® VSL2, IUPred, FoldIndex, and TopIDP). Sua principal função é fornecer imagens gráficas sobre a localização da desordem dentro da proteína. O PONDR-FIT apresenta acurácia e exatidão em seus resultados que por sua vez são obtidos após a submissão da sequência FASTA proveniente da proteína estudada. A ferramenta pode ser encontrada em [www.disprot.org](http://www.disprot.org) (XUE et al., 2010).

### 3.2.3. ANCHOR

O ANCHOR é uma implementação web-based que traz como resultado a predição de sítios de ligação presente nas regiões desordenadas ao longo da proteína. Para obter um resultado é necessário que uma sequência FASTA seja submetida. A saída do programa consiste em um gráfico que ilustra a posição do sítio de ligação juntamente com o número total de sítios encontrados na sequência. O ANCHOR pode ser encontrado em: <http://anchor.enzim.hu> (DOSZTÁNYI; MÉSZÁROS; SIMON, 2009).

## 3.3. Anotação de proteínas

### 3.3.1. Vírus-PLoc

Vírus-PLoc é um preditor que foi desenvolvido para identificar em escala subcelular a presença de proteínas virais em células infectadas. Apresenta uma boa taxa de sucesso global quando avaliada no ponto de vista estatístico e possui alto potencial, tornando-se dessa forma uma ferramenta de grande utilidade para estudos relacionados a viroses e proteínas virais. O resultado da análise é obtido através da submissão da sequência FASTA da proteína. O output consiste no número de acesso da respectiva proteína, a previsão do local onde a mesma se localiza na célula infectada e o tipo de abordagem estabelecido para que o resultado fosse encontrado. O Vírus-PLoc pode ser encontrado em: <http://www.csbio.sjtu.edu.cn/cgi-bin/VirusPLoc.cgi> (SHEN; CHOU, 2006).

### 3.3.2. Blast2GO

O Blast2GO (B2G) é uma ferramenta que possibilita a mineração de dados com base no Gene Ontology (GO). Funciona através de anotações baseadas em pesquisas de similaridade com análise estatística e visualização feita por meio de gráficos. Possui uma interface interativa, intuitiva e de fácil uso. A aplicação permite a submissão de várias sequências FASTA de uma única vez e traz resultados satisfatórios com destaque para a função molecular (*Molecular Function*), componente celular (*Cellular Component*) e o processo biológico (*Biological Process*) desempenhado pela proteína em estudo. O Blast2GO está disponível para download em <https://www.blast2go.com/> (CONESA et al., 2005).

## 4. RESULTADOS E DISCUSSÃO

No total foram analisadas 419 proteínas que permitiram uma análise da presença de desordem nas proteínas dos 34 vírus causadores de doenças em humanos. Os dados obtidos em cada sessão podem ser vistos a seguir:

### 4.1. Análise de desordem

#### 4.1.1. Discon

Como já mencionado na introdução o Discon fornece a porcentagem de desordem encontrada a partir da sequência proteica inserida. Os resultados obtidos de 0% a 100% foram classificados. Com base nos termos: muito baixa, baixa, média, alta e muito alta foi feita uma classificação em que um intervalo de 20 pontos foi determinado para cada classe, como na Figura 1.

Classe	Intervalo
Nenhuma desordem	$D = 0\%$
Desordem muito baixa	$0\% < D \leq 20\%$
Desordem baixa	$20\% < D \leq 40\%$
Desordem média	$40\% < D \leq 60\%$
Desordem alta	$60\% < D \leq 80\%$
Desordem muito alta	$80\% < D \leq 100\%$

Figura 1: Classes estabelecidas para a classificação quanto à quantidade de desordem encontrada em cada proteína. Cada classe possui um intervalo de 20 pontos. D refere-se à quantidade de desordem em porcentagem encontrada.

O número de proteínas encontrado e classificado em cada classe está disposto na tabela 2. A média de desordem encontrada nas diversas proteínas analisadas de cada vírus também pode ser encontrada na tabela a seguir:



Tabela 2: Número de proteínas em cada classe e a média de desordem encontrada nos vírus.

Vírus	Desordem (%) = D						Média de desordem do vírus
	Nula	Muito Baixa	Baixa	Média	Alto	Muita alta	
<b>HAstV-1</b>	0	9	2	0	0	0	14,3
<b>HTNV</b>	0	2	1	0	0	0	15,6
<b>ZIKV</b>	0	6	4	0	0	0	18,2
<b>LASV</b>	0	3	1	0	0	0	20,7
<b>DENV-1</b>	0	7	4	1	0	0	21,1
<b>HBV-B</b>	0	3	2	1	0	0	21,5
<b>WNV</b>	0	6	3	2	0	0	21,5
<b>JEV</b>	0	6	4	1	0	0	21,6
<b>H1N1</b>	0	6	3	1	0	0	22,2
<b>RUBV</b>	0	3	0	0	1	0	22,8
<b>MCPyV</b>	0	3	0	1	0	0	23,0
<b>JCPyV</b>	0	5	0	0	1	0	23,1
<i>Variola virus</i>	1	4	3	0	1	0	23,6
<b>ONNV</b>	0	5	3	2	0	0	23,8
<b>CHIKV</b>	0	5	3	2	0	0	24,0
<b>HCV</b>	0	6	3	3	0	0	24,2
<b>HHV-5</b>	2	12	6	6	1	0	24,2
<b>HEV-4</b>	0	1	1	0	0	0	24,5
<b>YFV</b>	0	5	5	2	0	0	24,8
<i>Sapporo virus</i>	0	5	1	1	1	0	24,9
<b>HPIV-3</b>	0	3	4	1	0	0	25,1
<b>RABV</b>	0	3	0	1	0	0	25,5
<b>MAYV</b>	0	6	2	2	1	0	25,5
<b>ZEBOV</b>	0	4	10	1	0	0	26,4
<b>RV-A</b>	0	7	3	2	1	0	26,9
<b>HPV-18</b>	0	4	2	3	0	0	27,5

<b>HRV-1A</b>	0	7	6	3	0	1	27,8
<b><i>Poliovirus</i></b>	0	7	6	2	0	1	29,0
<b>MuV</b>	0	5	1	2	1	0	29,2
<b>HHV</b>	2	36	16	9	6	6	29,9
<b>HPV-16</b>	0	2	4	2	1	0	34,4
<b>HTLV-1</b>	0	5	4	4	2	0	35,1
<b>VSR</b>	0	4	3	4	2	1	40,2
<b>HIV-1</b>	0	4	10	3	1	4	42,6

Das 419 proteínas estudadas 414 (98,8%) delas apresentam algum tipo de desordem. Quando comparado a outros organismos os vírus são aqueles que apresentam quantidades de desordem mais variável, neste estudo a média correspondente aos valores individuais de cada proteína parte de 14,3% (*Human astrovirus-1*) a 42,6% (*Human immunodeficiency virus*) como pode ser visualizado na tabela anterior. Em um estudo realizado por Bin Xue, A. (2012) sobre desordem intrínseca em 3500 proteomas de vírus mostra também uma extensa variação de desordem em que o *human coronavirus NL63* possui 7,3% resíduos desordenados em seu proteoma, enquanto o *Avian carcinoma virus* possui 77,3% de resíduos desordenados.

Isso sugere que a grande quantidade de desordem intrínseca em vírus parece estar ligada com a necessidade de maior regulação e sinalização de processos biológicos em que estão envolvidos. Diferentemente dos vírus as bactérias possuem baixa desordem em seu proteoma que na maioria das vezes varia em torno de 18% a 28%, apesar de haver algumas poucas exceções que são superiores a 28% e podem chegar até a 35% de resíduos desordenados em seu proteoma (XUE; DUNKER; UVERSKY, 2012).

Quando se compara o nível de variação de desordem dos vírus com outros organismos como o das bactérias, por exemplo, é que se nota a grande variação de desordem. A partir da classificação dada a cada proteína viral seguindo a Figura 1 observa-se que 198 delas possuem um valor de desordem classificado como muito baixo, possuindo uma desordem que é superior a 0%, mas inferior a 20%, o que indica que as proteínas virais no geral tendem a ter um nível baixíssimo de desordem. Apesar das outras classes também serem contempladas, elas possuem um pequeno número de proteínas. A classe muito alta,

por exemplo, possui apenas 3% de todas as proteínas analisadas. Esse fato pode ser ilustrado pela figura 2:

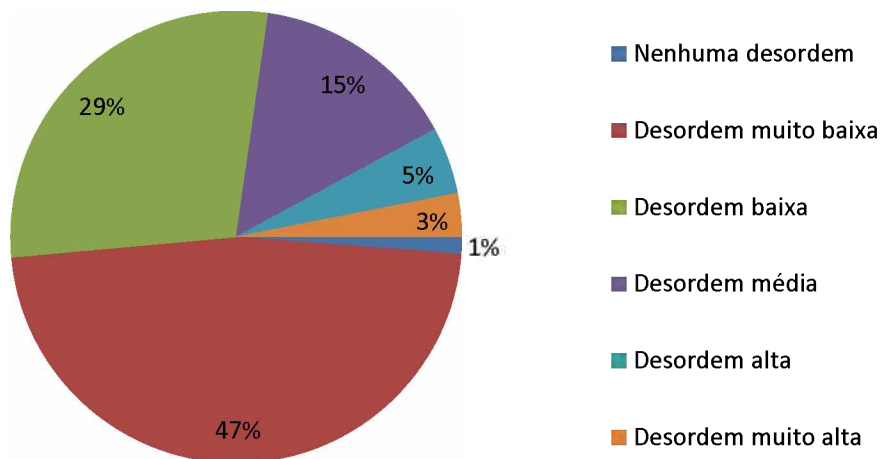


Figura 2: Porcentagem de proteínas virais classificadas em cada classe de desordem.

Assim, com o aumento da desordem nota-se um número cada vez menor de proteínas em classes superiores. Dessa forma, pode-se dizer que o número de proteínas tende a cair em classes que exigem valores mais elevados de desordem. De forma direta pode-se depreender em ordem decrescente uma sequência: n° de proteínas com desordem muito baixa > n° de proteínas com baixa desordem > n° proteínas com média desordem > n° de proteínas com alta desordem > n° de proteínas com desordem muito alta.

Contudo, também é interessante destacar que apenas 1% das proteínas analisadas não possui nenhum tipo de desordem, tal valor corresponde a 5 proteínas das 419 analisadas. Compreender isso é importante, pois ressalta o quão presente está a desordem intrínseca nas diversas proteínas virais e mesmo que as vezes essa desordem se apresente em um nível muito baixo, ela é crucial para que determinadas proteínas exerçam sua atividade normalmente, garantindo, o sucesso da espécie.

Ao destacar apenas as proteínas que são importantes na carcinogênese viral temos um paradigma bastante semelhante às médias gerais de desordem viral, pois, as classes inferiores ainda recebem a maior parte do número de proteínas analisadas. As oncoproteínas virais conhecidas e abordadas na pesquisa são as HBx proveniente do vírus causador da hepatite B, a proteína Core, NS3, NS5A do vírus da Hepatite C, as proteínas EBNA2, EBNA3C, LMP-1

e LMP-2 do Epstein-Barr vírus, as proteínas E5, E6, E7 do papilomavirus humano e a proteína Tax do vírus linfotrófico humano. Tais vírus segundo KEGG (Kyoto Encyclopedia of Genes and Genomes) são responsáveis por cerca de 10% a 15% dos cânceres em todo o mundo. De acordo com o mesmo, essa porcentagem se deve à expressão das oncoproteínas citadas.

Das oncoproteínas estudadas, 63% delas apresentaram uma desordem muito baixa ou baixa em sua maioria. Apenas 6% delas possuem uma desordem muito alta, sendo essa representada pela proteína LMP-1 do vírus Epstein-barr. A classe de proteínas de alta desordem nem sequer foi contemplada. Sendo assim, percebe-se que o padrão das oncoproteínas estudadas é bastante semelhante ao padrão das proteínas virais em geral, pois, em ambas as análises as classes referentes à desordem muito baixa e baixa apresentaram maior representatividade. Possivelmente, para o desenvolvimento do câncer, a quantidade de desordem não é tão importante, mas o local e a forma que esta permite o desempenho da oncoproteína em termos de função e de flexibilidade sim.

Outro ponto interessante também é que os vírus oncogênicos infectam, mas não matam a célula hospedeira, entretanto, conseguem desencadear infecções persistentes que em longo prazo, podem resultar em um câncer. Além disso, uma base genética viral que estimula o desenvolvimento de câncer no hospedeiro é crucial e é encontrada nos oncovírus conhecidos (SMITH; SMITH, 2016). Estudar as estruturas e a desordem das oncoproteínas talvez forneça à humanidade a chave para que muitos cânceres virais sejam evitados.

#### 4.1.2. PONDR-FIT

Através dos resultados do PONDR-FIT pode-se perceber que as proteínas analisadas apresentam a desordem distribuída tanto na extremidade N-terminal quanto na C-terminal ou no meio da cadeia polipeptídica. A grande maioria das proteínas abordadas nesse estudo possui a desordem distribuída nas duas extremidades simultaneamente.

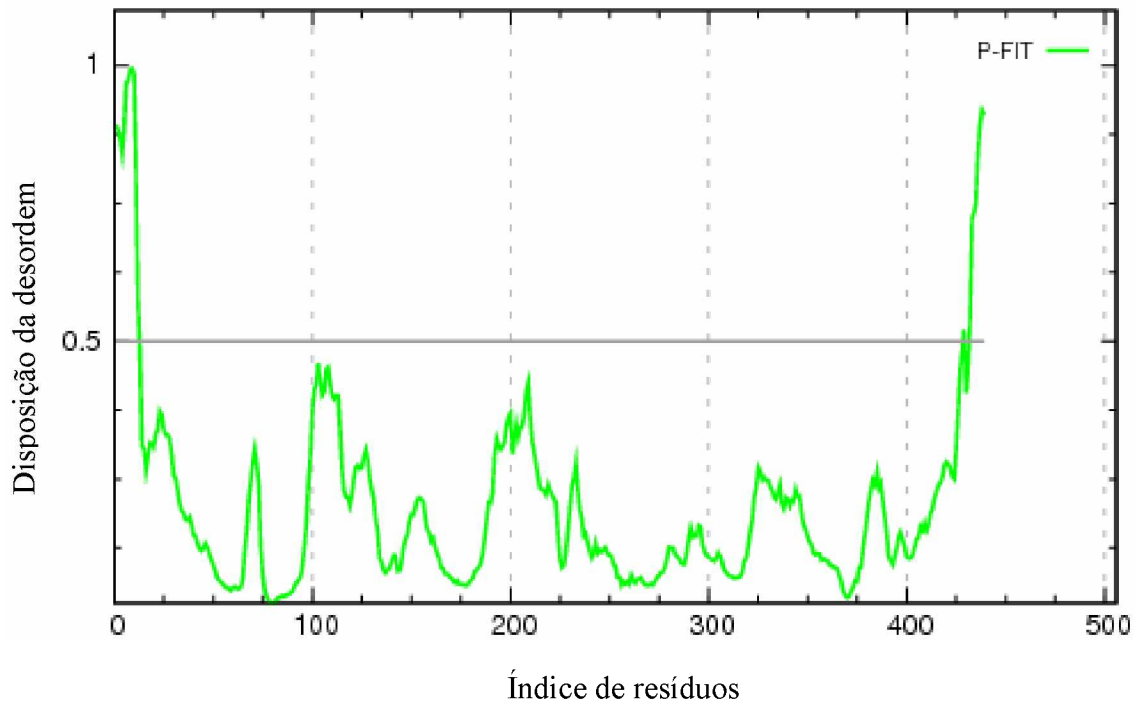


Figura 3: Desordem C-terminal e N-terminal da glicoproteína Q5XXP3, E1 pertencente ao envelope do vírus Chikungunya.

A proteína exposta na Figura 3 possui 13,89% de desordem ao longo de sua estrutura e a mesma é posicionada na região C-terminal e N-terminal, como pode ser visualizado nas extremidades da imagem. Ademais, essa proteína possui 439 aminoácidos e nenhum sítio de ligação segundo a ferramenta Anchor. Segundo a ferramenta “VírusPloc” a região subcelular ocupada na célula do hospedeiro é a membrana plasmática. A linha verde é a responsável por informar a disposição da desordem ao longo da proteína.

A disposição da desordem dentro da sequência é importante para compreender o comportamento da proteína, em estudos que compreendem apenas uma única proteína, essa análise se faz de grande importância, já que pode fornecer informações sobre locais de flexibilidade na molécula e também sobre uma possível região com vários sítios de ligações.

A seguinte proteína proveniente do vírus citomegalovírus (HHV-5) já mostra que além das desordens das extremidades C-terminal e N-terminal, ela também possui a desordem no meio da sequência proteica. Essa proteína identificada no banco de dados Uniprot como B9VXM8 se trata de uma DNA primase possuidora de 946 aminoácidos que atua na replicação do DNA. A desordem corresponde a apenas 0,846% da proteína.

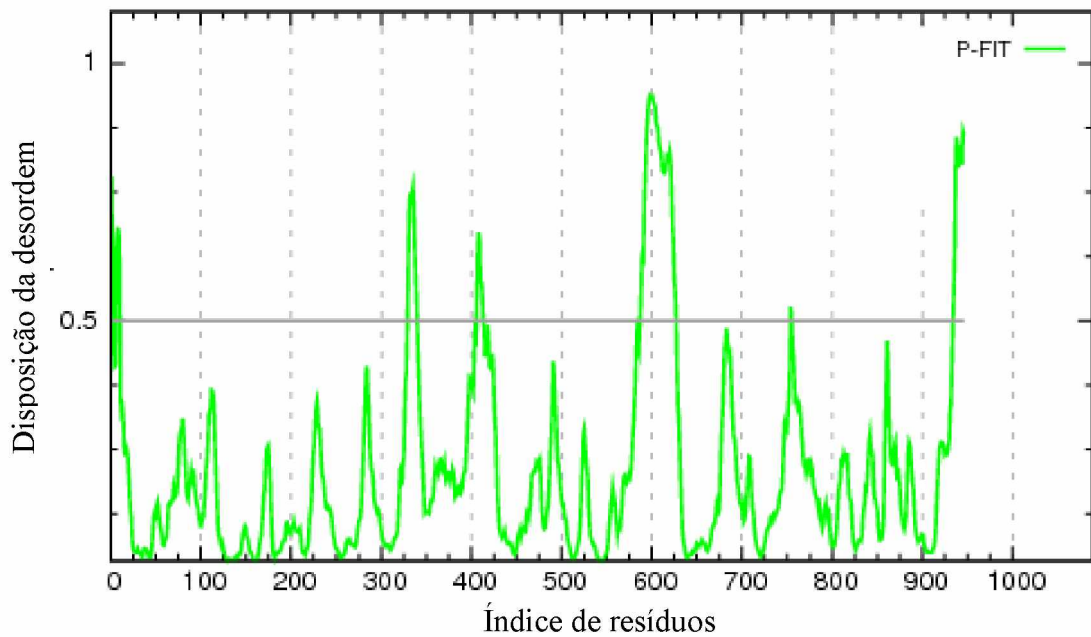
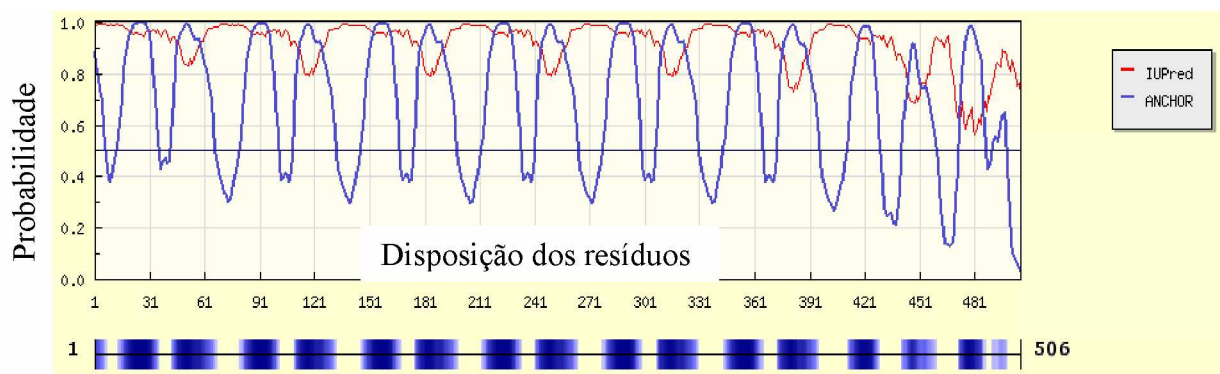


Figura 4: Desordem C-terminal e N-terminal e também no meio da estrutura da proteína do vírus Citomegalovírus.

#### 4.1.3. Anchor

Como exemplo dos dados gerados pelo Anchor temos a seguinte imagem referente à proteína nuclear do vírus Epstein-Barr identificada pelo UniProt como Q8AZK7. A proteína em questão possui 506 aminoácidos e possui 17 possíveis sítios de ligação ao longo de sua estrutura. A porcentagem de desordem corresponde a 88,34%, e pode ser observada a seguir:



#### Regiões de ligação

Figura 5: Possíveis sítios de ligação encontrados ao longo da proteína Q8AZK7 do vírus Epstein-Barr. As marcações em azul na horizontal referem-se aos possíveis sítios de interação encontrados ao longo da cadeia polipeptídica.

As linhas azul e vermelha ao longo do gráfico referem-se à relação entre a posição dos resíduos proteicos e a probabilidade da presença ou não de sítios de ligação, onde a linha que se estende horizontalmente cortando o eixo y no ponto 0,5 representa uma divisória, em que valores superiores a essa linha se tratam de um possível sítio de ligação. No eixo x, nota-se a numeração sequencial da cadeia linear que vai do aminoácido 1 até o aminoácido de número 506. O resultado também traz informações específicas sobre a posição em que o primeiro e último aminoácido ocupam no sítio de ligação. O número de aminoácidos que o compõe também é fornecido (DOSZTÁNYI, 2009).

Ao analisar o número de sítios de ligação nas classes de proteínas estabelecidas e estudadas, nota-se que quantidade de sítios tende a subir com o aumento de desordem da proteína. Isso indica que quanto maior a desordem proteica mais sítios de ligação provavelmente será encontrado na proteína e, conseqüentemente, maior será a facilidade para a mesma interagir com alvos diversos. De acordo com Iakoucheva et al. (2002) é esperado que a quantidade de desordem intrínseca se correlacionasse com o número de possíveis alvos de interação.

Na Figura 6 observa-se claramente a média de sítios encontrada em cada classe. Apesar de haver algumas poucas exceções que demonstra que a proposição feita a cima não é uma regra, mas sim uma tendência, é possível notar que as proteínas que não possuem nenhuma desordem possuem uma média de zero sítios, enquanto que as proteínas de desordem muito alta apresentam quase 4 sítios ao longo de sua extensão.

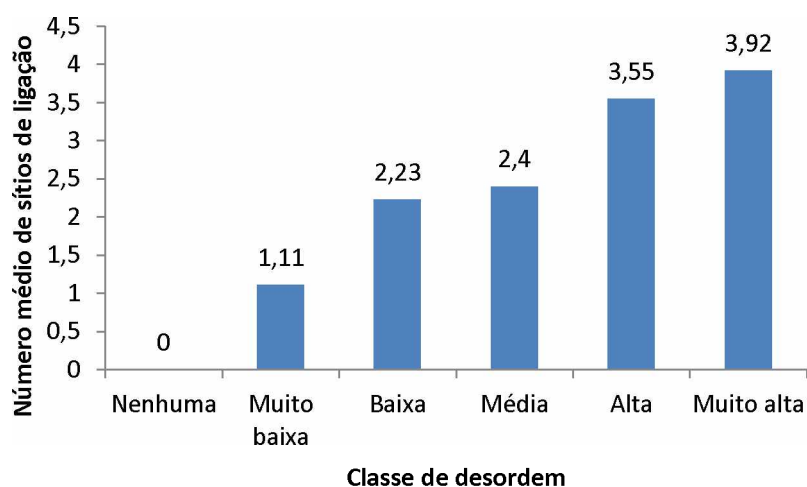


Figura 6: Número médio de regiões de ligação encontrado nas classes determinadas de desordem.

Tais interações com outros alvos são de extrema importância, pois se constitui de uma das razões pela qual, proteínas com desordem intrínseca tenham sido selecionadas ao longo da pressão evolutiva viral. As numerosas interações feitas por proteínas desordenadas com membranas, ácidos nucleicos (DNA e RNA), e até mesmo com outras proteínas que funcionam como alvos, dizem respeito ao comportamento e estilo de vida viral, a partir do momento que permite a participação em processos e funções biológicas cruciais para sua permanência ao longo dos anos. Sendo assim, trata-se de uma forma de adaptação a uma natureza biológica específica que possibilitou o desenvolvimento de infecções no hospedeiro, a replicação do vírus utilizando a maquinaria celular, e rápidas respostas adaptativas a novos hospedeiros (BEREZOVSKY, 2011).

## 4.2. Anotação de proteínas

### 4.2.1. Virus-Ploc

Colocando as proteínas classificadas quanto à desordem em grupos e relacionando com a predição da região subcelular que tais proteínas podem ser encontradas em uma célula infectada podem-se construir os seguintes gráficos dispostos na página seguinte:



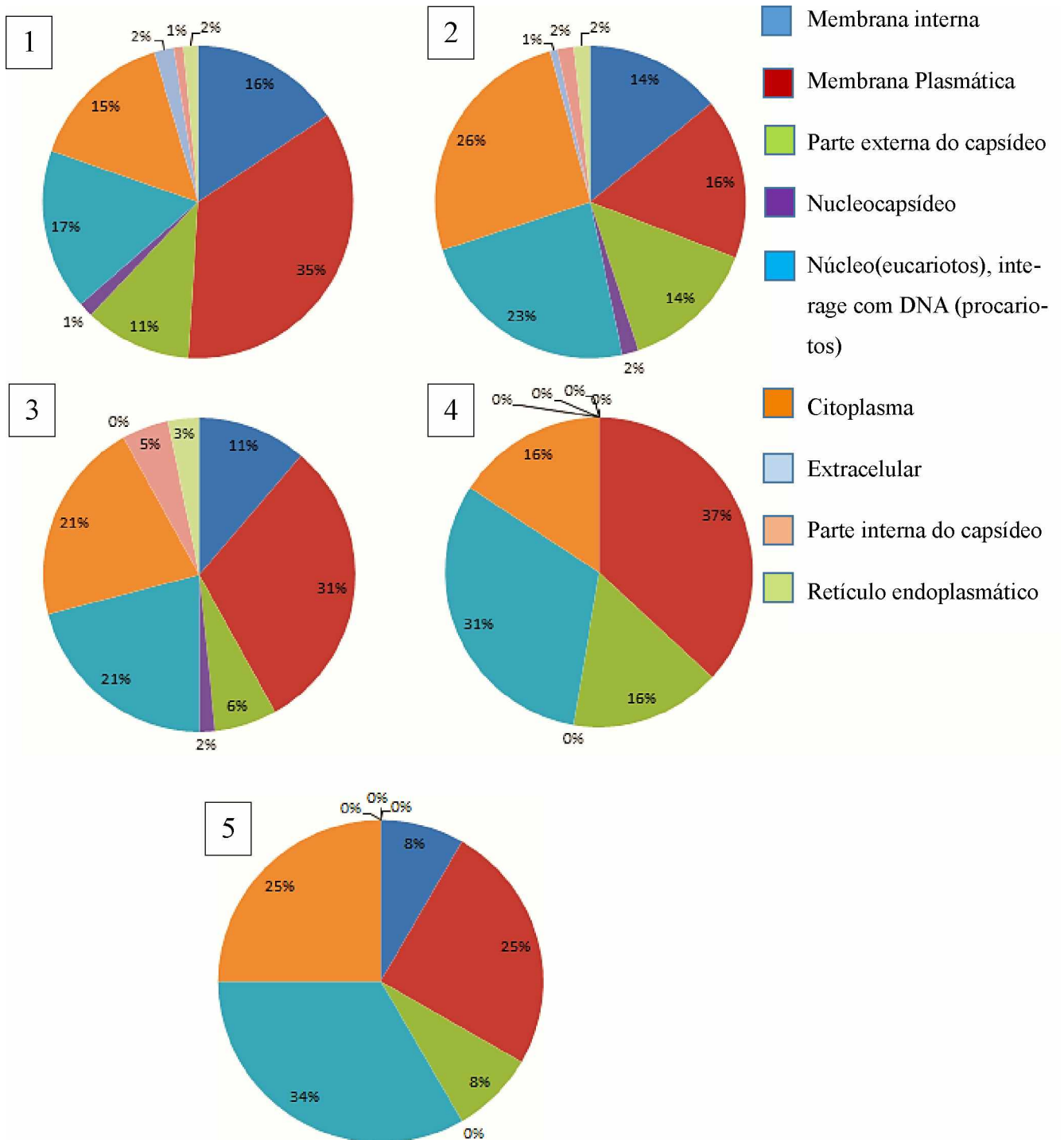


Figura 7: Gráficos da localização subcelular em cada classe de desordem viral. Gráficos: 1 – desordem muito baixa, 2 – desordem baixa, 3 – desordem média, 4 – desordem alta, 5 – desordem muito alta.

Com base nos dados analisados as proteínas de desordem muito baixa e média possuem uma maior probabilidade de serem encontradas na membrana plasmática das células do hospedeiro. Já as proteínas de baixa desordem tendem a ser encontradas no citoplasma em sua maioria. Diferentemente, as proteínas de desordem muito alta se localizam no núcleo de células eucarióticas e interagem com DNA em células procarióticas. Observa-se também que essas proteínas que se localizam no núcleo e interagem com o DNA aumentam em número à medida que a desordem também aumenta, iniciando-se em 17% nas proteínas com desordem muito baixa e terminando em 34% em proteínas com desordem muito alta. As proteínas que não possuíam desordem alguma em sua estrutura não foram representadas pelos gráficos, pois as 5 proteínas que contemplam essa classe apresentaram a predição subcelular correspondente a membrana plasmática unicamente.

Obter informações sobre a localização subcelular das proteínas tem sua importância fundamentada na necessidade de se ter informações de onde tais proteínas exercem suas funções e sobre como e que tipo de ambiente celular elas interagem com alvos diversos. Trata-se de um estudo fundamental para a compreensão de mapas metabólicos e regulação de processos biológicos (CHOU; SHEN, 2008). Ademais, a relação entre o nível de desordem e o ambiente subcelular de interação pode ser interessante, pois podem gerar indícios de como tais vírus conseguem ludibriar o sistema imunológico e assim provocar infecções nas células do hospedeiro. Dessa forma, a junção dessas informações juntamente com o entendimento dos processos biológicos em que as proteínas virais se envolvem gera interesse farmacológico com o objetivo de curar, tratar ou evitar que viroses afetem a humanidade. Assim, como já existem vacinas para o vírus da hepatite B, HPV, dentre outras viroses, novas formas preventivas necessitam ser criadas para os diversos vírus que ainda não possuem tais medidas profiláticas (SMITH; SMITH, 2016).

#### 4.1.4. Blast2GO

A classificação de acordo com a função, processo biológico e o componente celular foi dada pelo programa Blast2GO. O número de proteínas virais correspondentes à mesma função foi computado e expresso nos gráficos que podem ser visualizados adiante. As funções que mais obtiveram proteínas estão representadas:

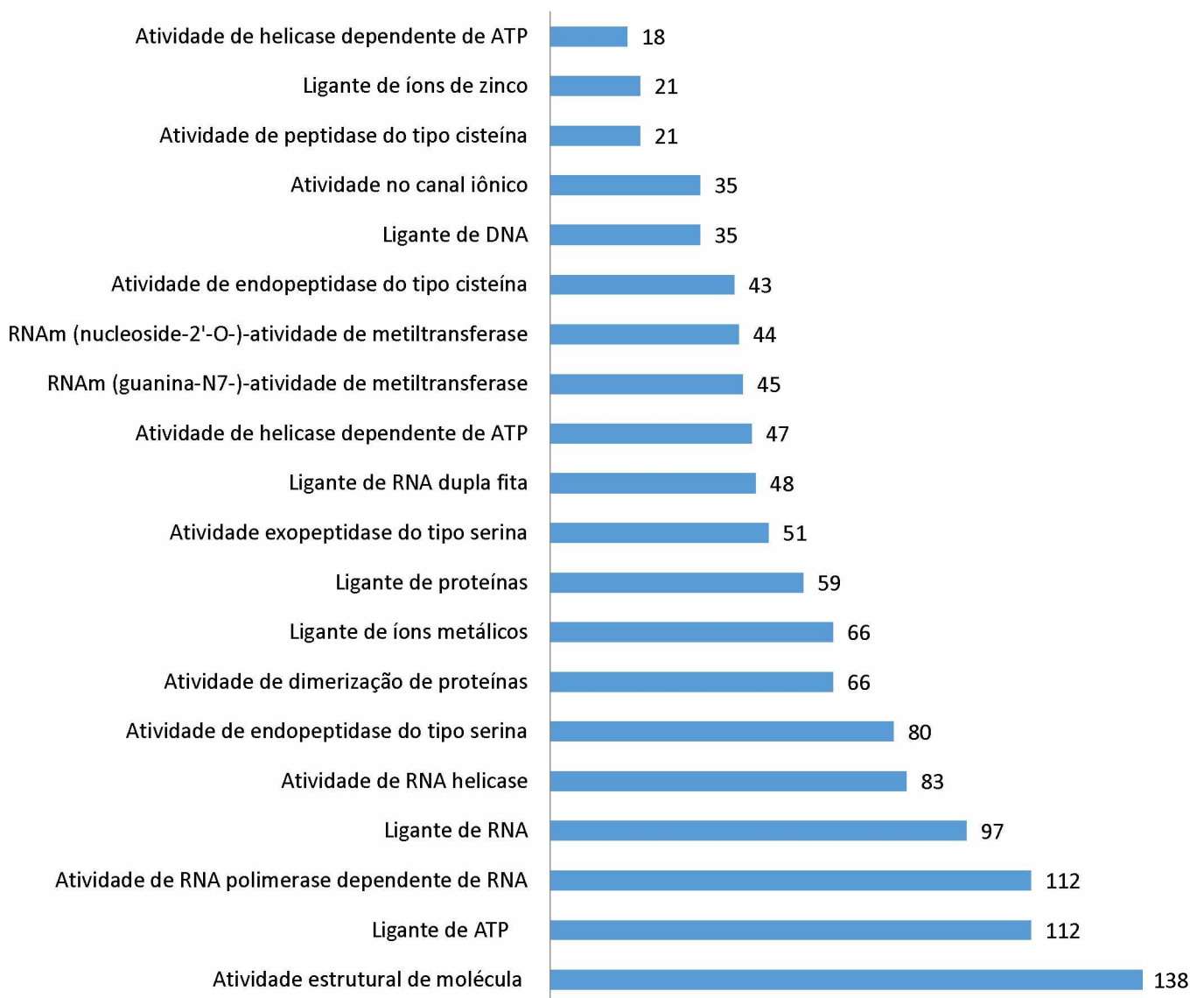


Figura 8: Funções das proteínas mais representadas pelas proteínas estudadas

Dentro dessas classificações quanto à função, uma que chama muita atenção é a “mRNA (nucleoside-2'-O-) metiltransferase”. Selecionando apenas as proteínas que exercem tal função e relacionando à desordem intrínseca dessas proteínas encontra-se que 89% delas possuem desordem classificadas como baixa ou muito baixa. Não tendo nenhuma representação de desordem alta ou muito alta.

Essa função se torna importante no ponto de vista dos estudos relacionados com desordem, pois se trata de uma função relacionada à expressão gênica. A mRNA (nucleoside-2'-O-)-metiltransferase permite a estabilidade do RNA mensageiro e juntamente com outras enzimas garante a ligação deste no ribossomo através da adição de um cap na extremidade 5'

(EGLOFF et al., 2002). Dessa forma, a baixa desordem contida nessa proteína, pode ser de grande importância para o desempenho de sua função.

Como exemplo de proteínas que apresentam uma desordem importante para sua função celular tem-se a proteína supressora de tumor p53 que é formada por uma estrutura definida e também por regiões desordenadas. Essa falta de estrutura em determinadas regiões é o que permite a interação fisiológica da p53 com diversos alvos e conformações (BELL et al., 2002; DYSON; WRIGHT, 2005). Contudo, mesmo o papel da desordem da proteína p53 ser tão importante, ela seria classificada como baixa desordem (38,68%). Da mesma forma, a proteína E7 do vírus HPV, também possui uma taxa de desordem muito baixa (9,52%), mas a mesma possui papel fundamental para a carcinogênese desencadeada pelo papilomavírus humano (NICOLAU-JUNIOR, 2013).

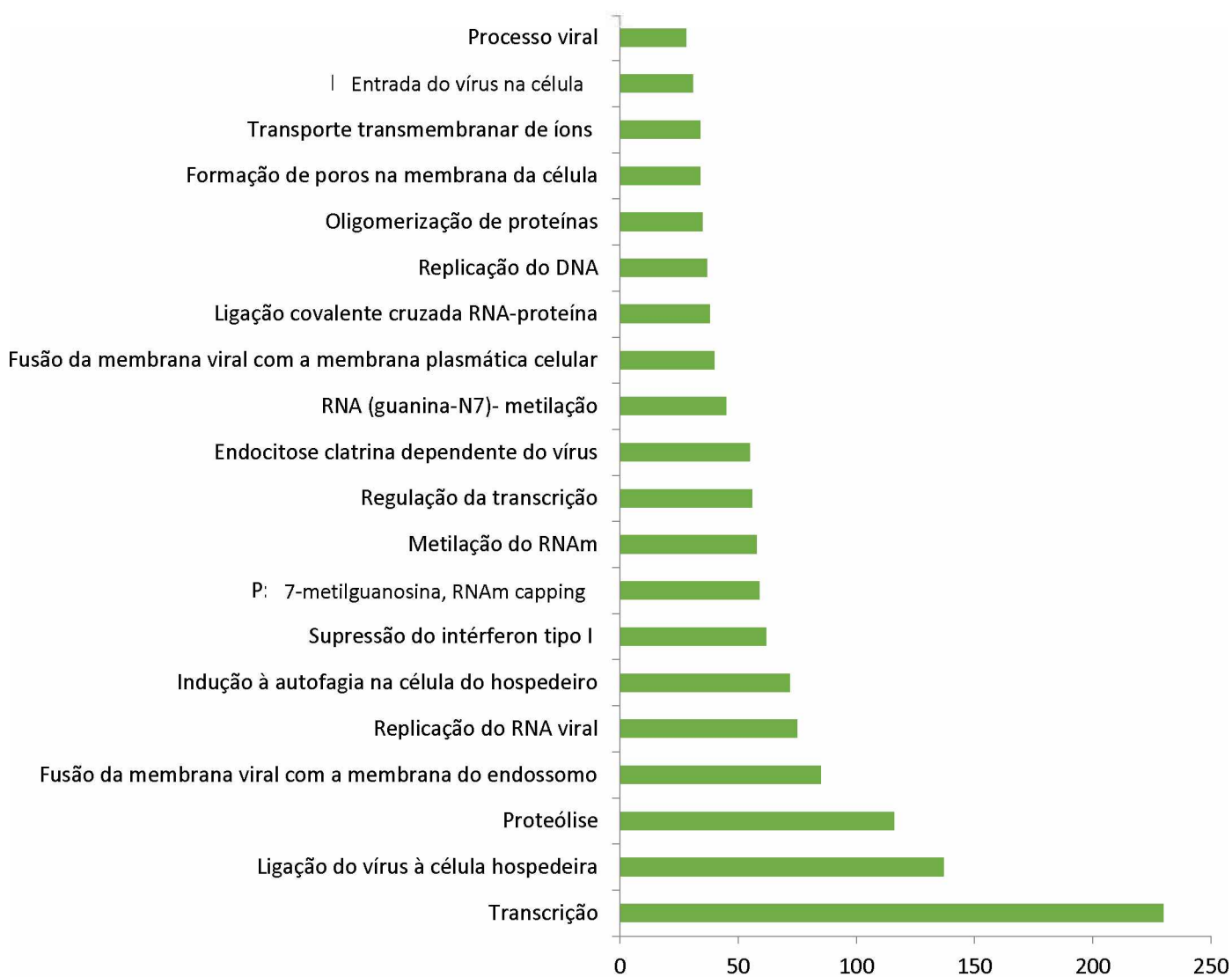


Figura 9: Processos biológicos mais representados pelas proteínas estudadas.

Processos de metilação, acetilação, fosforilação e ubiquitinação são processos fundamentais para a regulação e controle molecular e bioquímico de diversos organismos (Figura 9). Muitas das vezes as sequências desordenadas das proteínas são de grande importância, pois funcionam como estruturas flexíveis que ligam domínios importantes para a adaptação em diversos sítios de ligação constituídos por diferentes conformações de diferentes alvos (DYSON; WRIGHT, 2005).

Em alguns casos, os sítios de fosforilação em proteínas foram encontrados em regiões desordenadas, em outros em regiões com estrutura definida. Entretanto, os efeitos estruturais da fosforilação ocorrem em ambos os casos citados, regiões desordenadas, podem adquirir uma conformação definida, enquanto regiões bem estruturadas passam a ter uma estrutura desordenada. Essas transições de estrutura acarretam consequências que alteram a função proteica (IAKOUCHEVA et al., 2004).

Dois processos biológicos interessantes podem ser relacionados com o nível de desordem (Figura 10):

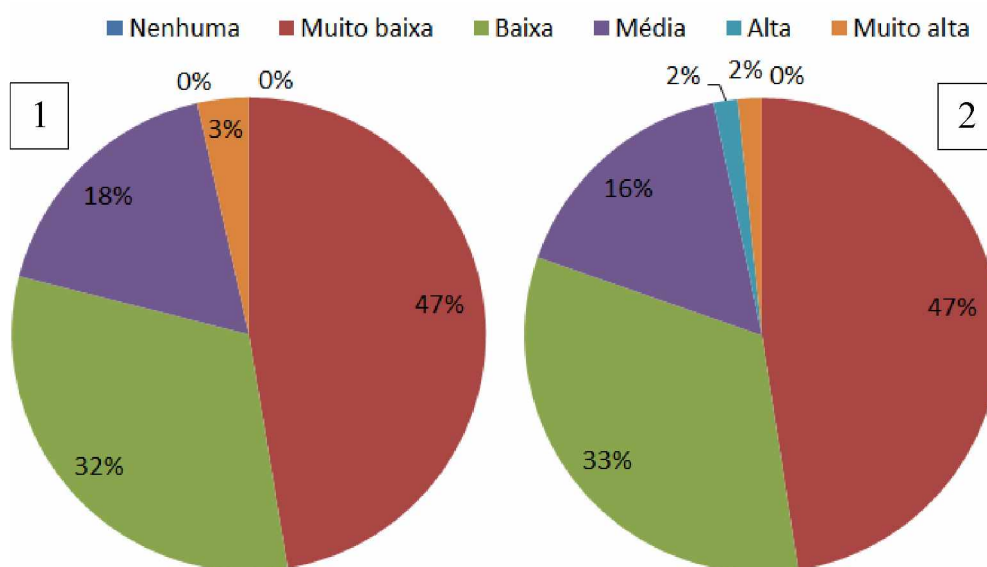


Figura 10: Processo de transcrição (1) e supressão de intérferon pelas proteínas virais (2).

O gráfico de número 1 representa proteínas relacionadas com a regulação do processo de transcrição, já o gráfico de número 2 representa a supressão do intérferon I mediado pelo vírus. Ambos os processos são interessantes para o desenvolvimento da virose, o primeiro porque é através da transcrição em que ocorre a formação do RNA a partir do DNA partici-

pando assim, do dogma central da biologia. O segundo porque a supressão do interferon é uma estratégia viral importante para o sucesso da infecção e replicação viral (GRANDI, 2012). No primeiro gráfico o que se observa é que a desordem presente nessas proteínas relacionadas à transcrição viral é em sua maioria proteínas com desordem muito baixa, valor esse correspondente a 47%, ou seja, quase metade dessas proteínas apresentam uma taxa de desordem inferior a 20% de sua extensão. Cerca de 80% possuem desordem baixa ou muito baixa, restando apenas 3% contendo uma desordem muito alta. Já quando se refere ao gráfico de número 2, nota-se que o panorama não muda muito, pois, aqui, 80% das proteínas também possuem desordem baixa ou muito baixa. Entretanto, a classe de alta desordem é contemplada com cerca de 2% das proteínas. Além disso, a proteína possuidora de desordem muito alta também corresponde a 2 % do total. Outros processos biológicos contemplados pela figura 9, também são interessantes e em futuros estudos haverão de ser analisados.

## 5. CONCLUSÃO

As proteínas virais possuem uma alta taxa de variabilidade da quantidade de desordem, podendo ir de 0% até 100%. Essa alta variação não é observada em outros organismos que por sua vez mantêm uma faixa de variação menor, como, por exemplo, as bactérias. A desordem de uma proteína muitas das vezes está relacionada às funções regulatórias, tais como metilação, fosforilação, acetilação e ubiquitinação dentre outras. Entretanto, a quantidade de desordem que a proteína tem não parece estar relacionada com o desempenho de uma função ou de outra, mas o local em que ela está inserida em uma proteína ou em outro sim. Em outras palavras, a desordem é extremamente importante para a execução de determinada função. O local em que ela está inserida na proteína também é algo a se observar ao estudar uma proteína, pois muitas das vezes ela funciona como um “linker” flexível entre uma cadeia e outra da proteína permitindo um maior número de interação entre a proteína e diversos alvos que a mesma se liga. Sem a desordem, mesmo o nível sendo classificado como muito baixo, talvez fosse impossível o desempenho da p53 ou a carcinogênese da proteína E7 do vírus HPV.

Classificando todas as proteínas estudadas de acordo com os termos “muito baixo”, “baixo”, “médio”, “alto”, e “muito alto”, quase 50% das proteínas virais foram classificadas

como tendo um nível de desordem muito baixo. Quase  $\frac{1}{4}$  delas possuem desordem média, alta ou muito alta. Ademais, aproximadamente 1% possui nenhuma desordem em sua estrutura. Vale ressaltar também que o nível de desordem está ligado com a quantidade de sítios de ligação encontrados na proteína, onde quanto maior o nível de desordem, maior será a tendência a ter um número superior de sítios de ligação em sua cadeia polipeptídica quando comparado a proteínas com um nível inferior de desordem.

Dessa forma, nota-se o quanto a desordem intrínseca é importante e ressalva-se a grande necessidade de mais estudos serem feitos para a compreensão da dinâmica molecular e interação dessas proteínas desordenadas com seus alvos. Uma grande infinidade de conhecimentos ainda está para ser descobertas e talvez, a solução para a cura, tratamento e prevenção de inúmeras doenças estejam exatamente na parte desordenada da proteína.

## REFERÊNCIAS

- ALVES, C.; CUNHA, C. Order and disorder in viral proteins: new insights into an old paradigm. **Future Virology**, v. 7, n. 12, p. 1183–1191, dez. 2012.
- BELL, S. et al. P53 Contains Large Unstructured Regions in Its Native State. **Journal of Molecular Biology**, v. 322, n. 5, p. 917–927, 2002.
- BEREZOVSKY, I. N. The diversity of physical forces and mechanisms in intermolecular interactions. **Physical Biology**, v. 8, n. 3, 2011.
- BRASIL, F. I. Dossiê Proteínas. **FOOD INGREDIENTS BRASIL**, v. 28, p. 30–58, 2014.
- CHOU, K. C.; SHEN, H. BIN. Cell-PLoc: A package of Web servers for predicting subcellular localization of proteins in various organisms. **Nature Protocols**, v. 3, n. 2, p. 153–162, 2008.
- CONESA, A. et al. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, n. 18, p. 3674–3676, 2005.
- DOSZTÁNYI, Z.; MÉSZÁROS, B.; SIMON, I. ANCHOR: Web server for predicting protein binding regions in disordered proteins. **Bioinformatics**, v. 25, n. 20, p. 2745–2746, 2009.
- DYSON, H. J.; WRIGHT, P. E. Intrinsically unstructured proteins and their functions. **Nature Reviews Molecular Cell Biology**, v. 6, n. 3, p. 197–208, 2005.
- EGLOFF, M. P. et al. An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: Crystal structure and functional characterization. **EMBO Journal**, v. 21, n. 11, p. 2757–2768, 2002.
- FERRON, F. et al. A Practical Overview of Protein Disorder Prediction Methods. **PROTEINS: Structure, Function, and Bioinformatics**, v. 70, n. 2, p. 311–319, 2006.
- GALEA, C. A. et al. Proteomic studies of the intrinsically unstructured mammalian proteome. **Journal of Proteome Research**, v. 5, n. 10, p. 2839–2848, 2006.
- GRANDI, T. Estudo de polimorfismos genéticos e resposta ao tratamento com interferon alfa/ribavirina em pacientes com o vírus da hepatite c. (Tese) **Centro de Biotecnologia**, Uni-



versidade Federal do Rio Grande do Sul – UFRGS, 2012.

HART, T. Microterrores: o guia completo das infecções bacterianas, virais e fúngicas que ameaçam a nossa saúde. **Principis**, 2004.

KEEG PATHWAY. Viral carcinogenesis. Disponível em: <[https://www.genome.jp/dbget-bin/www\\_bget?pathway:map05203](https://www.genome.jp/dbget-bin/www_bget?pathway:map05203)> Acesso em: 01 de novembro de 2018

IAKOUICHEVA, L. M. et al. Intrinsic disorder in cell-signaling and cancer-associated proteins. **Journal of Molecular Biology**, v. 323, n. 3, p. 573–584, 2002.

IAKOUICHEVA, L. M. et al. The importance of intrinsic disorder for protein phosphorylation. **Nucleic Acids Research**, v. 32, n. 3, p. 1037–1049, 2004.

MIZIANTY, M. J. et al. In-silico prediction of disorder content using hybrid sequence representation. **BMC bioinformatics**, v. 12, n. 1, p. 245, 2011.

NICOLAU-JUNIOR, N.; GIULIATTI, S. Modeling and molecular dynamics of the intrinsically disordered e7 proteins from high- and low-risk types of human papillomavirus. **Journal of Molecular Modeling**, v. 19, n. 9, p. 4025–4037, 2013.

SHEN, H.-B.; CHOU, K.-C. Virus-PLoc: A Fusion Classifier for Predicting the Subcellular Localization of Viral Proteins Within Host and Virus-Infected Cells. **Wiley InterScience**, v. 85, n. 80, p. 233–240, 2006.

SMITH, A. J.; SMITH, L. A. **Viral Carcinogenesis**. 1. ed. [s.l.] Elsevier Inc., 2016. v. 144

TOMPA, P. The interplay between structure and function in intrinsically unstructured proteins. **FEBS Letters**, v. 579, n. 15, p. 3346–3354, 2005.

TORRIERI, R. Identificação e caracterização computacional de proteínas do tipo IUP no proteoma predito de *Schistosoma mansoni*. **Fundação Oswaldo Cruz**, p. 136, 2010.

UVERSKY, V. N.; OLDFIELD, C. J.; DUNKER, A K. Intrinsically disordered proteins in human diseases: introducing the D2 concept. **Annual review of biophysics**, v. 37, p. 215–46, jan. 2008.

XUE, B. et al. **PONDR-FIT: A meta-predictor of intrinsically disordered amino acids**. [s.l.: s.n.]. v. 1804

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**, v. 30, n. 2, p. 137–149, 2012.