9 DPIN: um dicionário dos nanoambientes internos das proteínas e seu potencial para transformação em ativos para a agricultura

Ivan Mazoni
Goran Neshich

1 Introdução

As proteínas exercem um papel vital na manutenção da vida. Elas são macromoléculas resultantes da combinação, através de ligações peptídicas, destes 20 aminoácidos: alanina, arginina, aspartato, asparagina, cisteína, fenilalanina, glicina, glutamato, glutamina, histidina, isoleucina, leucina, lisina, metionina, prolina, serina, tirosina, treonina, triptofano e valina. Considerando a combinação linear entre esses 20 aminoácidos, o número de possíveis variações é de 20^n, onde "n" é a quantidade de resíduos de aminoácidos da proteína (como os aminoácidos perdem alguns átomos quando da formação da ligação peptídica, é usual denominá-los como resíduos de aminoácidos, uma vez que fazem parte de uma cadeia polipeptídica). Por exemplo, para uma proteína com 100 resíduos de aminoácidos, o número de possíveis combinações será igual a $20^{100} = 1{,}27 \times 10^{130}$. Em comparação, o número total estimado de átomos no Universo é de 9×10^{78} (Villanueva, 2009). Cada organismo, animal ou vegetal, possui milhares de diferentes proteínas. Entre as diversas funções que as proteínas têm, destacam-se, por exemplo: proteínas estruturais, de transporte, proteção e defesa, controle e regulação de expressão, catálise, movimento e

armazenamento. Para um melhor entendimento da relação entre a sequência de aminoácidos de uma proteína, sua estrutura tridimensional e a função desempenhada por ela, foi proposta a análise do nanoambiente proteico, também denominado distrito proteico ou região funcional, onde os elementos da estrutura de hierarquias subordinadas à superior (biologicamente funcional) estão inseridos.

A hipótese que motivou os trabalhos do Grupo de Pesquisa em Biologia Computacional (GPBC) da Embrapa Informática Agropecuária em Campinas (SP), durante a década de 2010-2020, foi uma abordagem que assume a existência de um "sinal", ou seja, uma variação nos valores dos descritores físico-químicos e estruturais que distinguem um local (ou uma subestrutura proteica) específico, onde um determinado elemento de estrutura secundária (ou um sítio ativo, uma interface etc.) está inserido no arcabouço da proteína inteira. Entender como os elementos de estrutura subordinados à estrutura biologicamente funcional são formados e posteriormente mantidos abrirá o caminho para compreendermos como as proteínas assumem sua estrutura final e, consequentemente, sua função. Em nossos trabalhos utilizamos o STING_RDB, uma base de dados única no mundo, produzida e mantida pelo GPBC da Embrapa, que reúne em um único repositório mais de 1300 descritores físico-químicos e estruturais de todos os resíduos de aminoácidos para cada cadeia de todas as estruturas proteicas depositadas no PDB (Protein Data Bank – um repositório público mundial onde foram depositadas todas as estruturas macromoleculares até agora decifradas).

Baseados nos resultados obtidos, concluímos que um determinado nanoembiente pode ser descrito não por um único descritor, mas por um conjunto de descritores, e que esse conjunto de descritores varia de acordo com o elemento da estrutura proteica selecionado a partir de estrutura hierarquicamente superior. Isso diferencia um determinado nanoambiente do restante da proteína e, inclusive, de outros nanoambientes na mesma proteína. O conhecimento adquirido resultante do estudo dos diversos nanoambientes permite que especialistas em diversas áreas, tais como experts em melhoramento de plantas em busca por novos agrodefensivos ou pesquisadores em busca por combustíveis mais sustentáveis, possam avançar nos seus trabalhos com maior introspecção molecular e uso de ferramentas mais precisas e apuradas, trabalhando no nível mais fundamental (molecular-atômico) de todos os processos biologicamente relevantes para medicina, agricultura, pecuária etc.

2 Nanoambientes proteicos e suas características

O ambiente estrutural local das proteínas, aqui denominado nanoambiente (Neshich et al., 2015), caracteriza o propósito funcional de diferentes

distritos proteicos, também conhecidos como "sítios estruturais" nas proteínas. Sugere-se, consequentemente, que o ambiente local em cada ponto e/ou região proteica reflete não somente seu papel estrutural, mas também sua contribuição em providenciar as características necessárias ao objetivo funcional de cada proteína. Por exemplo, a comunicação proteína-proteína é executada via interfaces proteicas: resíduos de aminoácidos em um mesmo sítio possuem algumas características particulares que não somente os diferencia dos demais resíduos da superfície livre da proteína, como também permitem a ligação especifica e seletiva entre as proteínas e a realização de sua função bioquímica (Moraes et al., 2014). De modo similar, a função de uma enzima está normalmente relacionada à atividade de seus resíduos de aminoácidos catalíticos (*Catalytic Site Residues* – CSR). Esses resíduos tão peculiares estão inseridos em um nanoambiente muito específico, definido também pela contribuição dos próprios CSR. Consequentemente, a função enzimática pode ser descrita pelas características dos CSR e sua vizinhança (Salim, 2015). Com base nessas considerações, e assumindo que o nanoambiente local define a função proteica, esse é um conceito que pode ser utilizado para se obterem métricas específicas para quantificar e descrever outros nanoambientes.

A exploração das propriedades dos nanoambientes pode ser feita por meio de um método que é ao mesmo tempo autoexplicativo e intuitivo. Suponha que seja possível inserir uma sonda imaginária em qualquer lugar de uma estrutura proteica e obter como resultado um diagnóstico descrevendo as características do ambiente em que a sonda está inserida. Não se pode realizar esse tipo de intervenção física, e por isso a sonda necessita ser substituída pelo cálculo de valores, métricas e forças que desejamos quantificar em cada sítio/ponto em particular. Essa abordagem assemelha-se ao método de GRID para cálculo de campos de interação molecular no desenvolvimento de drogas (Goodford, 1985; Von Itzstein et al., 1993), mas com um foco diferente. A sua vantagem é que qualquer resíduo de aminoácido, ou qualquer de seus átomos da cadeia principal ou lateral, pode servir como centro para a sonda, e a partir de um ponto selecionado as interações de todas as forças podem ser estimadas, catalogadas e armazenadas em um banco de dados relacional apropriado – em nosso caso, o STING_RDB (Oliveira, 2007). Depois de armazenados, os atributos e seus respectivos valores podem ser mapeados de volta para a estrutura proteica, para a sequência proteica ou até para a sequência de nucleotídeos do gene que codifica essa proteína, e podem ser usados para inspeção visual ou para análises estatísticas e/ou numéricas. Nossa hipótese é de que qualquer ambiente específico (o nanoambiente) possui uma afinação precisa dos escritores físico-químicos e estruturais específicos para o desempenho de sua função e, assim, pode ser identificado e classificado adequadamente. Por exemplo, no caso de interfaces para contatos proteicos, pode-se esperar que áreas específicas da proteína, ocupando parte de sua

superfície, possuam características suficientemente diferentes dos resíduos de aminoácidos encontrados em áreas da superfície livre (Moraes et al., 2014). De fato, nós consideramos que tal suposição é parte dos requisitos biológicos para desempenhar uma função específica: nesse exemplo, a função é na verdade uma espécie de "comunicação" entre os parceiros proteicos bastante específicos. Portanto, um nanoambiente é caracterizado com precisão por seus descritores físico-químicos e/ou estruturais e seus correspondentes valores, tornando possível a sua distinção do restante da estrutura proteica e também a predição, com uso de técnicas computacionais e estatísticas de aprendizado de máquina, das coordenadas desses distritos em outras proteínas (homólogas ou não) que ainda não foram caracterizadas química e funcionalmente.

Dentre os vários nanoambientes proteicos mais estudados, destacam-se 10, os quais são listados a seguir.

i Interfaces entre as proteínas: trata-se de uma intersecção das superfícies proteicas, onde as duas proteínas se aproximam e se tocam, construindo um homo ou héterocomplexo das macromoléculas (Moraes et al., 2014);
ii Interfaces entre anticorpo e antígeno: como no caso i) anterior, porém as duas proteínas em questão são um anticorpo e um antígeno (Viart et al., 2016);
iii Pontos "quentes" na superfície proteica (*hot spots*): localidades delimitadas da área superficial da proteína, obrigatoriamente situadas na sua área de interface, e com aminoácidos hidrofóbicos identificados, propensos para interagir com resíduos similares da interface complementar da outra proteína (Pereira, 2012);
iv Interfaces entre proteínas e DNA: como no caso i) anterior, aqui as duas moléculas em questão são uma proteína e uma molécula da DNA;
v Interfaces entre proteínas e ligantes: como no caso i) anterior, aqui as duas moléculas em questão são uma proteína e um ligante (Borro et al., 2016);
vi Interfaces entre proteínas e membranas;
vii Resíduos de aminoácidos dos sítios catalíticos: identificar os resíduos de aminoácidos que formam o sítio catalíco das enzimas, determinando a sua função (Salim, 2015);
viii Sítios alostéricos: localizados usualmente na superfície proteica; quando ocupados por uma determinada molécula, controlam a velocidade de uma reação química que a proteína executa, usando, por regra, seu conjunto dos CSRs, como parte de sua função;
ix Elementos da estrutura secundária: caracterização físico-química e estrutural das α-hélices (Mazoni et al., 2018), folhas-β e dobras;
x A profundidade de alcance de sensoriamento local entre aminoácidos: uma medida frequentemente usada para delimitar a distância através da qual os átomos, com suas cargas (e outras características), ainda exercem

certa influência em localidades remotas, porém dentro do limite anteriormente mencionado (Silveira et al., 2009).

Os itens: i) a vi) descrevem as interfaces em geral; vii) e viii) descrevem atividade química das proteínas; e ix) e x) descrevem características estruturais das proteínas em geral.

2.1 Lista dos descritores físico-químicos e estruturais que caracterizam os nanoambientes específicos

Atualmente, o Blue Star STING (BSS) (Neshich et al., 2006) apresenta 32 tipos ou classes diferentes e independentes de descritores físico-químicos e estruturais das proteínas (Tabela 1) (Neshich et al., 2005), sendo que um total de 1.307 variações desses descritores estão pré-calculados (utilizando diferentes

Tabela 1
Lista das 32 classes de descritores físico-químicos e estruturas do BSS.

Classes de descritores do Blue Star STING	
1. ResBoxes	17. Hot spots
2. Intra-chain atomic contacts [ITC]	18. Sequence conservation [HSSP]
3. The inter-chain atomic contacts [IFC]	19. Sequence conservation [SH_2Q^s]
4. ITC contacts energy	20. Solvent accessibility
5. IFC contacts energy	21. Dihedral angles
6. Interface area [IF]	22. Pockets/cavities
7. Water contacting [WC]	23. Electrostatic potential
8. Ligand pocket forming [LP]	24. Hydrophobicity
9. Surface forming [SF] residues	25. Curvature
10. Prosite	26. Distance from the N-/C-terminal
11. ProTherm	27. Density
12. Secondary structure indicator [PDB]	28. Sponge
13. Secondary structure indicator [DSSP]	29. Order of cross presence
14. Secondary structure [STRIDE]	30. Order of cross link
15. Multiple occupancy	31. Rotamers
16. Temperature factor	32. Space clash

parametrizações) e armazenados no banco de dados STING_RDB (Oliveira, 2007). Em 18 de maio de 2020, o STING_RDB apresentava 151.711 estruturas, com 467.038 cadeias e 95.148.233 resíduos de aminoácidos: para cada um deles foram pré-calculados os 1.307 parâmetros, totalizando $12x10^9$ registros no banco de dados. Dentre estes, foram escolhidos alguns para utilização na caracterização dos nanoambientes e na composição do seu dicionário, considerando apenas aqueles que apresentam maiores probabilidades de estarem associados com os processos de reconhecimento de padrões nas proteínas selecionadas. Procurando uma definição adequada do nanoambiente dos resíduos catalíticos e que geralmente seja válida também para os outros nanoambientes anteriormente mencionados, baseada nos descritores físico-químicos e estruturais, num primeiro momento descartaram-se descritores referentes à conservação dos aminoácidos, uma vez que esses parâmetros são uma medida de um conjunto de proteínas homólogas e não refletem qualquer característica presente na estrutura proteica (Salim, 2015).

3 Contribuições

3.1 Qual é o significado do conhecimento sobre os nanoambientes proteicos

A estrutura da proteína define a sua funcionalidade. Porém, de que forma isso é feito e quais características das estruturas contribuem crucialmente para a sua função é algo que ainda precisa ser totalmente decifrado. Para responder a essa pergunta, é necessário considerar, preferencialmente, os elementos estruturais (também denominados distritos proteicos ou nanoambientes), em vez de considerar a estrutura como um todo. Esses elementos, por sua vez, devem ser compreendidos com base nas características físico-químicos e estruturais geradas pelas propriedades dos resíduos de aminoácidos, interagindo entre si e criando, efetivamente, um novo elemento estrutural hierárquico. Somente pela consideração desses elementos na hierarquia estrutural é que podemos entender que a funcionalidade das proteínas pode ser deconvoluída em elementos de comunicação, tais como as interfaces em geral, elementos construtivos ou estrutura secundária, e elementos de atividade química. Esses últimos normalmente dão origem à funcionalidade e à especificidade da proteína como um todo. Seguindo esse raciocínio, cada elemento na hierarquia estrutural tem sua característica local distintiva e, por consequência, sua função local. Fica claro que o conhecimento geral e detalhado sobre os nanoambientes proteicos constitui, literalmente, um dicionário com o qual podemos construir complexas expressões que descrevem a relação estrutura-função das proteínas.

3.2 Um dicionário dos descritores dos nanoambientes terá impacto na variedade de pesquisas que visam a inovação em áreas como agricultura, medicina e biologia em geral

Uma compilação dos resultados dos trabalhos feitos desde 1998 – quando a plataforma STING foi lançada nos EUA, como parte integral das ofertas das plataformas para análise estrutural das proteínas no então Laboratório Nacional de Brookhaven, a sede de Banco de Dados das Estruturas Proteicas (PDB) – resultou em um site chamado: "Dictionary of Internal Protein Nanoenvironments" (DIPN)[1].

Nas Figuras 1-3 é possível ver a interface geral dessa nova oferta do GPBC da Embrapa Informática Agropecuária. Trata-se de uma página introdutória, com a descrição geral do objetivo dessa plataforma e dos elementos detalhados, listados em uma ordem funcional.

A Figura 1 apresenta a página de entrada da plataforma Dictionary of Internal Protein Nanoenvironments (DIPN), indicando o intuito desse produto, as opções para acesso, a logística de organização do site e a lista dos dez nanoambientes proteicos mais estudados.

Na Figura 2 temos uma página da plataforma Dictionary of Internal Protein Nanoenvironments (DIPN) mostrando seis dos dez nanoambientes disponíveis, com uma curta descrição e acesso para detalhes da entrada de cada uma das opções: i) interfaces proteína-DNA, ii) interfaces proteína-membrana, iii) elementos de estrutura secundária.

A Figura 3 apresenta uma página da plataforma Dictionary of Internal Protein Nanoenvironments (DIPN) mostrando mais três dos dez nanoambientes disponíveis, com uma curta descrição e acesso para detalhes da entrada

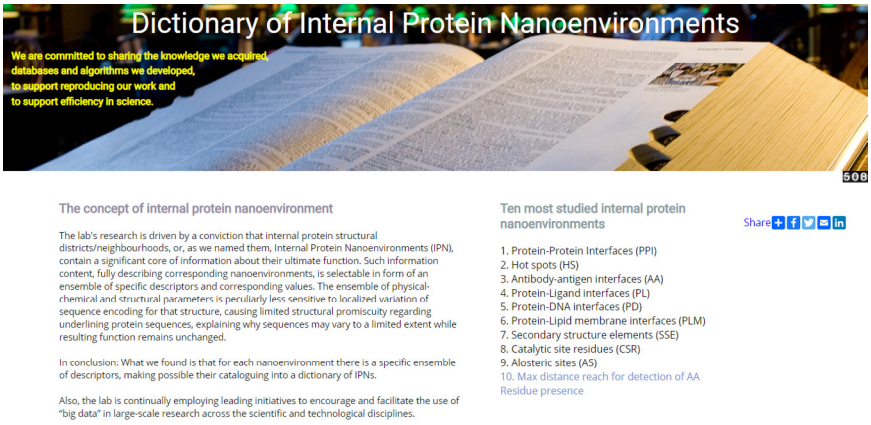

Figura 1
Página de entrada da plataforma Dictionary of Internal Protein Nanoenvironments (DIPN).

Fonte: Embrapa, 2020.

1 Disponível em: https://www.proteinnanoenvironments.cnptia.embrapa.br/index.html.

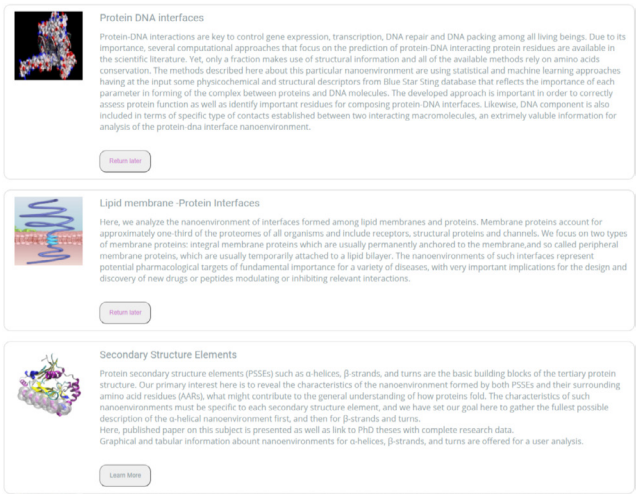

Figura 2
Página da plataforma Dictionary of Internal Protein Nanoenvironments (DIPN).

Fonte: Embrapa, 2020.

Figura 3
Página da plataforma Dictionary of Internal Protein Nanoenvironments (DIPN).

Fonte: Embrapa, 2020.

de cada uma das opções: i) resíduos do sítio catalítico, ii) sítios alostéricos, e iii) profundidade de alcance de sensoriamento local entre aminoácidos.

Na Figura 4, o usuário pode ver os detalhes de apresentação de um dos nanoambientes: interfaces proteicas. O intuito do DIPN é oferecer ao usuário as informações que indicam quais são os descritores mais relevantes que, com a especificidade e a cobertura ampla, descrevem o nanoambiente selecionado para análise. Na parte inferior da Figura 4 pode ser vista uma tabela com os dez descritores das interfaces proteicas mais relevantes. São estes: 1) pontes de

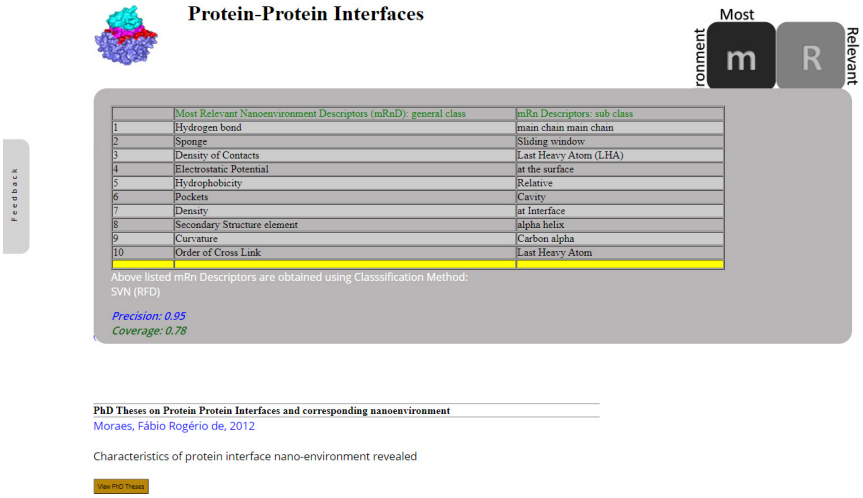

Figura 4
Página da plataforma Dictionary of Internal Protein Nanoenvironments (DIPN) ilustrando as opções do usuário.

Fonte: Embrapa, 2020.

hidrogênio do tipo cadeia principal para cadeia principal; 2) espongicidade (no modo de uma janela deslizante); 3) densidade dos contatos entre aminoácidos (centrados no último átomo pesado da cadeia lateral dos aminoácidos); 4) potencial eletrostático na superfície proteica; 5) hidrofobicidade (na escala relativa); 6) bolsos estruturais (do tipo cavidade); 7) densidade atômica na superfície; 8) elemento da estrutura secundária presente (α-hélice); 9) curvatura a partir do carbono-α; e 10) a ordem de ligação cruzada (a partir de último átomo mais pesado da cadeia lateral). Esses descritores podem ser entendidos como requerimentos principais que exigem sua inclusão para que um conjunto dos aminoácidos, não necessariamente contíguos na sequência primaria, construam um conjunto que poderá ser considerado apto para compor uma interface com outra proteína. Em seguida, a plataforma informa qual o método de classificação estatística foi usado para obter esse ranqueamento da importância dos descritores (neste caso: *Support Vector Machine* e *Random Forest*), e ainda informa com qual precisão e cobertura as conclusões foram atingidas. Neste caso, 0,95 e 0,78, respetivamente. Nessa mesma página, encontra-se uma variedade de informações adicionais, tais como links para a tese de doutorado que gerou os resultados, a publicação que descreve o trabalho pertinente ao assunto em pauta (na Figura 5 estamos ilustrando o *abstract* dessa publicação), e ainda um link para que o usuário possa acessar o software, se ele desejar gerar novos dados para um conjunto de proteínas de interesse biológico.

Na Figura 4, temos a página da plataforma Dictionary of Internal Protein Nanoenvironments (DIPN) ilustrando as opções que o usuário tem, uma vez selecionando o item: interfaces proteicas. Na parte superior dessa figura tem-se a indicação das publicações pertinentes ao assunto em pauta e lista dos softwares. Logo em seguida pode ser visto um abstract da publicação

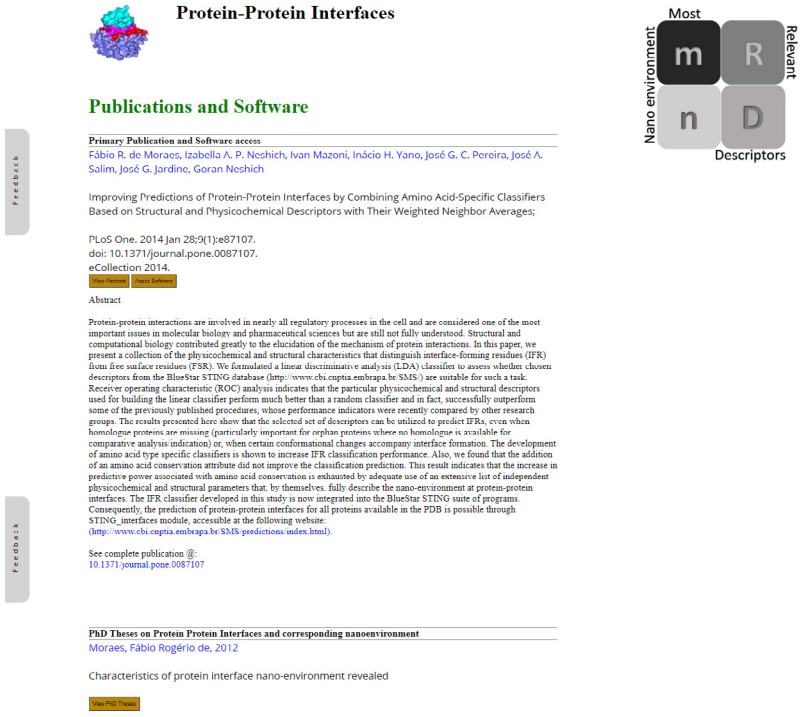

Figura 5
Página da plataforma Dictionary of Internal Protein Nanoenvironments (DIPN), com a opção que oferece ao usuário uma rápida consulta ao abstract da principal publicação, divulgada em revista renomada na área de biologia computacional, sobre o nanoambiente em pauta.

Fonte: Embrapa, 2020.

principal descrevendo nosso trabalho com nanoambiente das interfaces proteicas, com correspondente pointer para a publicação original. Do lado direito superior, existe um ícone com título: mRnD, ou seja, Descritores mais relevantes do nanoambiente. Passando com mouse por cima desse ícone, abre-se uma janela com informação indicada no título do ícone.

Na Figura 6 apresentamos os itens disponíveis para acesso à página do software que pode ajudar o usuário na elaboração de uma lista de descritores para um conjunto de proteínas do interesse dele. Na Figura 7 temos as duas opções principais para elaboração dos dados para interfaces proteicas: metodologia LDA (modelos lineares para inferência da lista dos mais relevantes descritores das interfaces proteicas) e opção SHI, uma metodologia alternativa que determina o índice de hidrofobicidade na superfície proteica, um indicador preciso das interfaces. O usuário pode encontrar um tutorial para se informar sobre detalhes de uso do software, descrição dos *datamarts* para definição dos *benchmarks* e descrição dos complexos usados no treinamento do método, usando tanto homo como heterocomplexos proteicos. Nas Figuras 1-7 mostramos apenas as entradas mais cruciais da plataforma DIPN.

A plataforma é complexa e exige os conhecimentos de um biólogo computacional treinado para que seja possível tratar os dados para um conjunto

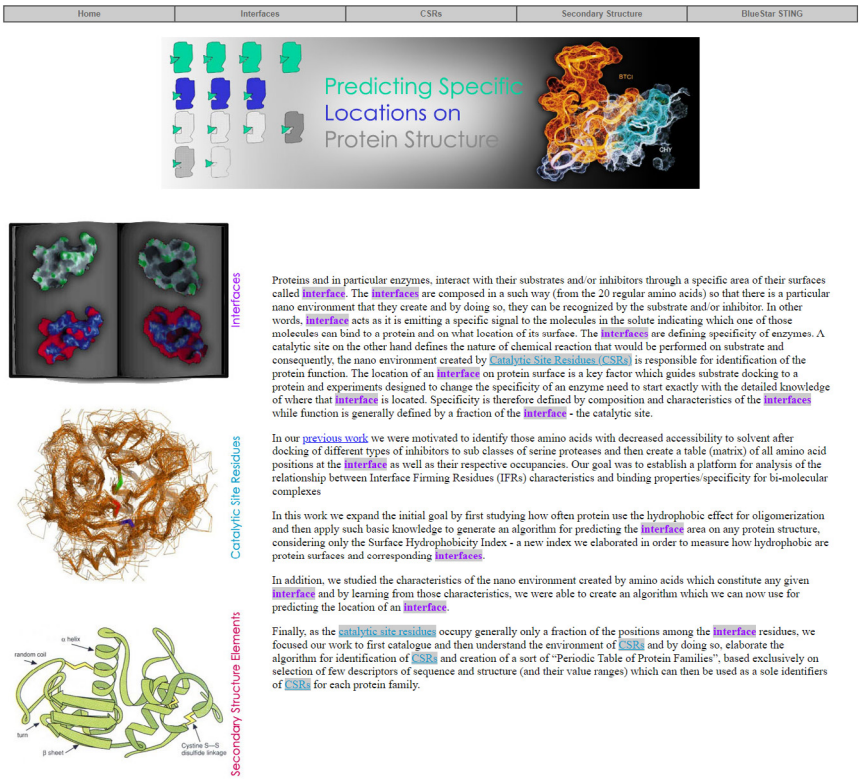

Figura 6
Página da plataforma Dictionary of Internal Protein Nanoenvironments (DIPN), com o acesso aos softwares que elaboram o ranqueamento de descritores para nanoambientes das interfaces proteicas, resíduos catalíticos e elementos da estrutura secundária nas proteínas mais relevantes.

Fonte: Tela captada da plataforma DIPN.[2]

de proteínas selecionadas. Entretanto, especialistas da área de biologia molecular interessados em saber quais são os descritores mais relevantes para cada nanoambiente listado na plataforma DIPN podem fazê-lo em um tempo razoável, com um treinamento mínimo, e saber quais características desses nanoambientes são cruciais. Assim, tem-se os candidatos que não poderão ser objetos de quaisquer modificações, por exemplo nas tentativas que exigem mutações sítio-dirigidas nas proteínas de interesse. As opções para uso dos algoritmos ou até para acesso ao código-fonte são providas com o intuito de oferecer um ambiente completo de trabalho, inclusive para aqueles biólogos computacionais que desejam adaptar os algoritmos aos seus próprios requerimentos, permitindo assim o compartilhamento dos trabalhos já realizados por parte da Embrapa, os quais poderão ser modificados por colegas em outros laboratórios para fins específicos.

2 Disponível em: https://www.proteinnanoenvironments.cnptia.embrapa.br/index.html.

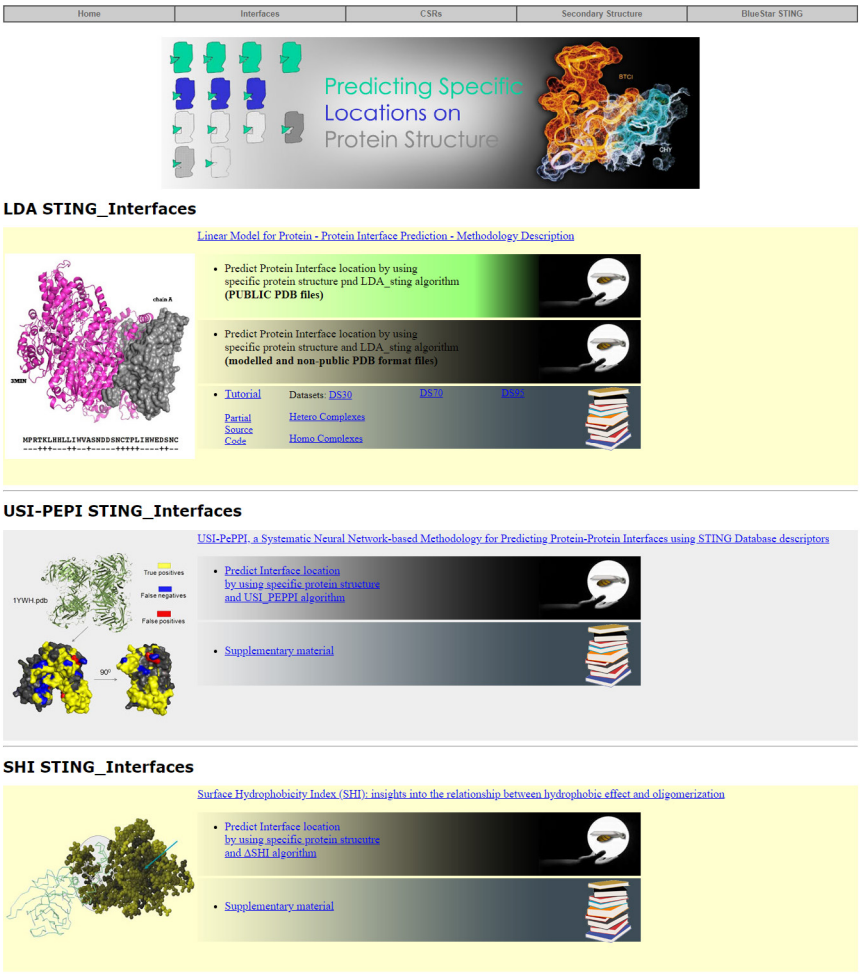

Figura 7
Página da plataforma Dictionary of Internal Protein Nanoenvironments (DIPN), com as opções para o usuário que deseja elaborar o ranqueamento dos descritores mais relevantes para as interfaces proteicas usando um conjunto das proteínas de interesse para problema biológico que exige seu engajamento.

4 Considerações finais

De posse de um dicionário dos descritores dos principais nanoambientes proteicos, constrói-se uma realidade que orienta os pesquisadores e que possibilita o avanço nas áreas que objetivam intensificar a inovação para a agricultura, a medicina e a biologia em geral. Compreende-se que uma compilação dos descritores essenciais dos 10 nanoambientes proteicos mais estudados daria condição otimizada para o desenho mais apurado, eficaz e efetivo de novos fármacos, defensivos agrícolas, vacinas, inibidores, catalisadores e anticorpos. Podemos aqui, a título de exemplo da aplicabilidade do conteúdo apresentado neste capitulo, mencionar algumas das tecnologias, das quais o GPBC da Embrapa Informática Agropecuária conseguiu depositar o pedido

de quatro patentes ao longo dos anos, concentrando-se principalmente no entendimento, na aprendizagem e na análise dos nanoambientes proteicos, que foram cruciais para a solução de algumas demandas biologicamente relevantes e focadas em um caminho para os impactos necessários no campo para o produtor que precisa usar a tecnologia para evitar perdas e aperfeiçoar sua efetividade. São elas:

1) Fungicida: método para o desenho de um novo fungicida por um método para desenhar computacionalmente novos compostos com potencial função inibitória da enzima endopoligalacturonase, envolvida em processos de invasão em células vegetais. (Neshich et al., 2013a)
2) Biodiesel: método para previsão de mutantes que aumentem o índice de hidrofobicidade da superfície das proteínas. (Neshich et al., 2013b).
3) Inseticida: desenho computacional para novos inibidores de alfa-amilases. (Neshich et al., 2013c)
4) Bactericida: identificação de alvos terapêuticos para desenho computacional de drogas contra bactérias dotadas da proteína pilt. (Neshich et al., 2012)

Essas 4 tecnologias refletem sobre a forte interdependência entre as demandas da agricultura moderna e o conhecimento, que pede uma abordagem inovadora, interdisciplinar e, principalmente, molecular, interligada com a matemática, a computação e a estatística, para que se possa fazer um avanço nas cada vez mais complexas necessidades do setor produtivo. O exemplo do GPBC da Embrapa Informática Agropecuária é uma manifestação das possibilidades nacionais para o potencial de desenvolvimento tecnológico no nível mais alto e competitivo. As pesquisas desenvolvidas pelo GPBC da Embrapa Informática Agropecuária chamaram atenção de colaboradores e colegas internacionais das universidades mais renomadas, tais como Oxford, Cambridge, MIT, seguidas pelas companhias de maior impacto digital, tais como Microsoft Research, e companhias do ramo de defensivos agrícolas, tais como Bayer e BASF: uma meia centena de publicações nas revistas científicas com um fator de impacto médio orbitando em volta de valor 3, sendo vários com fator de impacto acima de 11; centenas de palestras e seminários, cursos e workshops internacionais, congressos de mais alto nível organizados aqui no território nacional, com participação inclusive dos vários cientistas com prêmio Nobel; uma meia centena de pacotes de software publicados e disponibilizados para uso da comunidade científica mundial; uma dúzia de bancos de dados relevantes para área de biologia estrutural computacional, incluindo o STING_RDB; 26 projetos aprovados (90%) por fontes externas da Embrapa, com financiamentos beirando 4 milhões de dólares e com total dos entregáveis aproximando-se de 500. Toda essa biblioteca de resultados

e prêmios profissionais foi uma condição *sine qua non* para que possamos, no final, transformar nosso conhecimento adquirido em algo que podemos oferecer para a cadeia produtiva, que agora tem a opção de desenvolver essas soluções em produtos para os mercados nacional e internacional. Portanto, a plataforma chamada "Dicionário dos Nanoambientes Internos das Proteínas" é um produto que desenvolvemos sempre pensando nas aplicações que podem ser geradas a partir do nosso conhecimento, mas tendo paciência e determinação de permanecer no caminho que exige tempo e que requer aprendizagem usando a ciência básica, porque as aplicações científicas não existem sem a ciência para ser aplicada.

5 Referências

BORRO, L.; YANO, I. H.; MAZONI, I.; NESHICH, G. Binding affinity prediction using a nonparametric regression model based on physicochemical and structural descriptors of the nano-environment for protein-ligand interactions. In: STRUCTURAL BIOINFORMATICS AND COMPUTATIONAL BIOPHYSICS, 2016, Orlando. **Proceedings**... Orlando: [s.n.], 2016. p. 116-117.

EMBRAPA. Computational Biology Research Group. **Dictionary of Internal Protein NanoEnvironments**. Disponível em: https://www.proteinnanoenvironments.cnptia.embrapa.br/index.html. Acesso em: 18 maio 2020.

GOODFORD, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. **Journal of Medicinal Chemistry**, v. 28, n. 7, p. 849-857, July 1985. DOI: 10.1021/jm00145a002.

MAZONI, I.; BORRO, L. C.; JARDINE, J. G.; YANO, I. H.; SALIM, J. A.; NESHICH, G. Study of specific nanoenvironments containing **α**-helices in all-**α** and (**α**+ **β**)+(**α**/**β**) proteins. **PLOS One**, v. 13, n. 7, p. 1-25, 2018. Artigo e0200018. https://doi.org/10.1371/journal.pone.0200018.

MORAES, F. R. de; NESHICH, I. A. P.; MAZONI, I.; YANO, I. H.; PEREIRA, J. G. C.; SALIM, J. A.; JARDINE, J. G.; NESHICH, G. Improving predictions of protein-protein interfaces by combining amino acid-specific classifiers based on structural and physicochemical descriptors with their weighted neighbor averages. **PLOS ONE**, v. 9, n. 1, p. 1-15, 2014. DOI: 10.1371/journal.pone.0087107.

NESHICH, G.; JARDINE, J. G.; NESHICH, I. A.; SALIM, J. A.; MAZONI, I. (2013). EUA Patente Nº WO2013/110147Al, 2013c.

NESHICH, G. E. A.; BORRO, L. C.; HIGA, R. H.; KUSER, P. R.; YAMAGISHI, M. E.; FRANCO, E. H.; KRAUCHENCO, J. N.; FILETO, R.; RIBEIRO, A. A.; BEZERRA, G. B.; VELLUDO, T. M.; JIMENEZ, T. S.; FURUKAWA, N.; TESHIMA, H.; KITAJIMA, K.; BAVA, A.; SARAI, A. TOGAWA, R. C.; MANCINI, A. L. The diamond STING server. **Nucleic Acids Research**, v. 33, n. 2, p. W29-W35, July 2005. Supplement. DOI: 10.1093/nar/gki397.

NESHICH, G.; MAZONI, I.; OLIVEIRA, S. R. M.; YAMAGISHI, M. E. B.; KUSER-FALCÃO, P. R.; BORRO, L. C.; MORITA, D. U.; SOUZA, K. R. R.; ALMEIDA, G. V.; RODRIGUES, D. N.; JARDINE, J. G.; TOGAWA, R. C.; MANCINI, A. L.; HIGA, R. H.; CRUZ, S. A. B.; VIEIRA, F. D.; SANTOS, E. H.; MELO, R. C.; SANTORO, M. M. The Star STING server: a multiplatform environment for protein structure analysis **Genetics and Molecular Research**, v. 5, n. 4, p. 717-722, 2006.

NESHICH, G. E. A.; NESHICH, I. A. P.; MORAES, F.; SALIM, J. A.; BORRO, L.; YANO, I. H.; MAZONI, I.; JARDINE, J. G.; ROCCHIA, W. Using structural and physical-chemical parameters to identify, classify, and predict functional districts in proteins – the role of electrostatic potential. In: ROCCHIA, W.; SPAGNUOLO, M. (ed.). **Computational electrostatics for biological applications**: geometric and numerical approaches to the description of electrostatic interaction between macromolecules. Cham: Springer, 2015. p. 227-254. DOI: 10.1007/978-3-319-12211-3_12.

NESHICH, G.; JARDINE, J. G.; NESHICH, I. A.; SALIM, J. A.; MAZONI, I. EUA Patente Nº WO2012/031343A2, 2012.

NESHICH, G.; JARDINE, J. G.; NESHICH, I. A.; SALIM, J. A.; MAZONI. EUA Patente Nº WO2013097012A1, 2013a.

NESHICH, G.; JARDINE, J. G.; NESHICH, I. A.; SALIM, J. A.; MAZONI, I. EUA Patente Nº WO2013/016794Al, 2013b.

OLIVEIRA, S. R. de M.; ALMEIDA, G. V.; SOUZA, K. R. R.; RODRIGUES, D. N.; KUSER-FALCÃO, P. R.; YAMAGISHI, M. E. B.; SANTOS, E. H. dos; VIEIRA, F. D.; JARDINE, J. G.; NESHICH, G. Sting_RDB: a relational database of structural parameters for protein analysis with support for data warehousing and data mining. **Genetics and Molecular Research**, v. 6, n. 4, p. 911-922, 2007.

PEREIRA, J. G. D. C. **Caracterização dos aminoácidos da interface proteína-proteína com maior contribuição na energia de ligação e sua predição a partir dos dados estruturais**. 2012. 106 p. Dissertação (Mestrado) – Programa de Pós-Graduação em Genética e Biologia Molecular, Instituto de Biologia, Universidade Estadual de Campinas, Campinas.

SALIM, J. A. **Aplicação de técnicas de reconhecimento de padrões usando os descritores estruturais de proteínas da base de dados do software STING para discriminação do sítio catalítico de enzimas**. 2015. 214 p. Dissertação (Mestrado) – Programa de Pós-Graduação em Engenharia Elétrica, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas.

SILVEIRA, C. H. da; PIRES, D. E. V.; MINARDI, R.; RIBEIRO, C.; VELOSO, C. J. M.; LOPES, J. C. D.; MEIRA JÚNIOR, W.; NESHICH, G.; RAMOS, C. H. I.; HABESCH, R.; SANTORO, M. M. Protein cutoff scanning: a comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. **Proteins**: structure, function, and bioinformatics, v. 74, n. 3, p. 727-743, Feb 2009. DOI: 10.1002/prot.22187.

VIART, B.; DIAS-LOPES, C.; KOZLOVA, E.; OLIVEIRA, C. F. B.; NGUYEN, C.; NESHICH, G.; CHÁVEZ-OLÓRTEGUI, C.; MOLINA, F.; FELICORI, L. F. EPI-peptide designer: a tool for designing peptide ligand libraries based on epitope–paratope interactions. **Bioinformatics**, v. 32, n. 10, p. 1462-1470, May 2016. DOI: 10.1093/bioinformatics/btw014.

VILLANUEVA, J. C. How many atoms are there in the Universe. **Universe Today**, July 30, 2009. Disponível em: https://www.universetoday.com/36302/atoms-in-the-universe/. Acesso em: 18 maio 2020.

VON ITZSTEIN, M.; WU, W.-Y.; KOK, G. B.; PEGG, M. S.; DYASON, J. C.; JIN, B.; VAN PHAN, T.; SMYTHE, M. L.; WHITE, H. F.; OLIVER, S. W.; COLMAN, P. M.; VARGHESE, J. N.; RYAN, D. M.; WOODS, J. M.; BETHELL, R. C.; HOTHAM, V. J.; CAMERON, J. M.; PENN, C. R. Rational design of potent sialidase-based inhibitors of influenza virus replication. **Nature**, v. 363, p. 418-423, 1993. DOI: 10.1038/363418a0.