

BLUESTAR STING

O QUE É?

A API BlueStar Sting é mais que uma camada de aplicação responsável pela interface entre o banco de dados STING_RDB e o front-end, uma vez que através dela usuários do mundo todo podem acessar os dados dos descritores físico-químicos e estruturais de qualquer estrutura proteica depositada no Protein Data Bank (PDB).

O QUE A API OFERECE?

Acesso aos Descritores com informações sobre as propriedades físicas, químicas, físico-químicas, geométricas e espaciais dos aminoácidos (e.g. energia de contatos, potencial eletrostático, hidrofobicidade, densidade, etc.) presentes em cada estrutura proteica depositada no PDB.

- ✓ **Air:** Área da superfície acessível que cada resíduo fornece para a superfície proteica. É calculado através de uma sonda esférica com raio igual ao da molécula de água (1.4Å) que percorre toda a superfície de van der Waals. A superfície gerada pelo centro geométrico da esfera sonda é a superfície acessível ao solvente.
- ✓ **Atom:** ATOM é usado para identificar proteínas ou átomos de ácidos nucleicos. Cada átomo é identificado por um número sequencial, um nome de átomo específico, o nome e o número do resíduo ao qual pertence, um código de uma letra para especificar a cadeia, seus x, y e z coordenadas e um fator de ocupação e temperatura. Esta informação lhe dá muito controle ao explorar a estrutura. Por exemplo, a maioria dos programas de gráficos moleculares permite colorir seletivamente porções identificadas da molécula – por exemplo, selecionar todos os átomos de carbono e colorir-los de verde, ou selecionar um determinado aminoácido e destacá-lo.
- ✓ **Atom Asa:** Nem todos os aminoácidos que constituem as proteínas estão acessíveis ao solvente. Alguns estão enterrados no núcleo das estruturas proteicas e, normalmente, estão associados à estabilidade de tais estruturas, promovendo contatos que não são estabelecidos com o solvente, como por exemplo, os contatos hidrofóbicos entre átomos de carbono dos resíduos de aminoácidos. Esse descritor calcula a área acessível que cada resíduo.

- ✓ **Chain:** Uma proteína consiste numa (ou várias) cadeias longas obtidas por reação de condensação de aminoácidos. Cada cadeia é identificada por uma letra.
- ✓ **Contact:** A maior parte dos resíduos de aminoácidos estabelece diferentes contatos com outros resíduos de aminoácidos, além da ligação peptídica que une a sequência primária de organização das proteínas. A plataforma Blue Star STING diferencia 14 tipos de contatos intra e intercadeias e mais 12 tipos de contatos entre cadeias proteicas e moléculas de DNA/RNA, que são armazenados em sua base de dados em nível atômico, ou seja, possui informações sobre quais pares de átomos estão estabelecendo contato e a energia equivalente ao tipo de contato.
- ✓ **Cross Link Order:** Devido ao dobramento da sequência de aminoácidos na estrutura tridimensional, os resíduos distantes na sequência primária são colocados próximos na conformação assumida pela proteína funcional, e podem, portanto, interagir entre si. A palavra “ordem” no nome do descritor é devido à relação do número de contatos estabelecidos entre segmentos de aminoácidos, separados por um número mínimo de aminoácidos na estrutura primária da proteína, e que foram encontrados dentro de uma sonda esférica de raio r centrada nos carbonos α e β e no LHA de qualquer resíduo de uma proteína.
- ✓ **Cross Presence Order:** O descritor Cross Presence contabiliza a presença dos aminoácidos dentro de sondas esféricas de raios r , centradas nos carbono- α , carbono- β e LHA, que estão distantes na estrutura primária da proteína por no mínimo l aminoácidos (tamanho do segmento), mesmo que os resíduos não estejam estabelecendo nenhum contato entre si. É contabilizado somente um resíduo para cada segmento.
- ✓ **Curvate:** A curvatura é definida em nível atômico, ou seja, a cada átomo da proteína é atribuído um valor de curvatura correspondente à região onde este se localiza na superfície da molécula, sendo atribuído um valor negativo àqueles átomos em regiões côncavas, positivo para átomos em regiões convexas e o valor zero aos átomos enterrados (presentes no interior da proteína). Para atribuir valores aos resíduos de aminoácidos de cada proteína, realiza-se o cálculo da curvatura média considerando-se apenas os resíduos na superfície da proteína (curvatura > 0). Considera-se ainda cada cadeia isoladamente e em complexo, sendo que cada resíduo possui dois valores de curvatura.
- ✓ **Curvate Atom:** Curvatura é definida em nível atômico, ou seja, a cada átomo da proteína é atribuído um valor de curvatura correspondente à região onde este se localiza na superfície da molécula, sendo atribuído um valor negativo àqueles átomos em regiões côncavas, positivo para átomos em regiões convexas e o valor zero aos átomos enterrados (presentes no interior da proteína).

- ✓ **Density Sponge:** A densidade local de cada aminoácido é calculada utilizando uma abordagem de sonda esférica. Para cada resíduo, uma sonda esférica de raio r é centrada no carbono- α ou no LHA. As massas dos átomos internos à sonda esférica são somadas e divididas pelo volume da sonda esférica. São calculadas 20 variações deste descritor, utilizando diferentes raios para a sonda esférica (de 3 a 7Å, com incremento de 1Å), centrada em dois átomos diferentes, e distinguindo entre cadeias isoladas e em complexos proteicos. Esponjicidade soma o volume ocupado por cada átomo. Esse volume é então subtraído e normalizado pelo volume da sonda esférica, resultando em uma medida do espaço vazio no nanoambiente de cada resíduo.
- ✓ **Distances:** Para cada resíduo de aminoácido de uma proteína, são calculados três descritores de distância. Em cada cadeia proteica, tem-se definidos os aminoácidos N-terminal e C-terminal, que correspondem as duas extremidades da sequência proteica. As distâncias euclidianas no espaço tridimensional entre um resíduo qualquer da cadeia proteica e os resíduos nas duas extremidades da sequência são calculadas, e esses dois valores são tomados como descritores. Utiliza-se também a distância entre o resíduo de aminoácido e o centro de massa da cadeia proteica.
- ✓ **DSSP:** Diferentemente da estrutura terciária, a estrutura secundária não define posições tridimensionais específicas dos resíduos de aminoácidos, mas sim conformações locais de um grupo de resíduos que interagem entre si para formar diferentes combinações estruturais. Entre as mais encontradas, têm-se hélices alfa, folhas beta e laços (loops). O DSSP (Define Secondary Structure of Proteins) é um algoritmo que define as estruturas secundárias.
- ✓ **Electrostatic Potential:** Para cada resíduo de aminoácido, quatro tipos de descritores de potencial eletrostático são gerados: potencial eletrostático no carbono- α , no último átomo pesado da cadeia lateral (LHA da sigla em inglês Last Heavy Atom), valor médio dos átomos do resíduo, e soma dos potenciais dos átomos na superfície da molécula para cada resíduo.
- ✓ **Energy Density:** O descritor definido como Densidade de Energia de Contatos (ou CED – Contacts Energy Density, em inglês) utiliza uma sonda esférica centrada no carbono- α ou LHA de cada resíduo. Todos os contatos entre resíduos de aminoácidos dentro da sonda esférica têm suas energias somadas e divididas pelo volume da sonda esférica. Dessa forma, para cada resíduo, considerando apenas uma esfera sonda, há 4 atributos para este descritor, pois se consideram de forma separada os contatos internos (entre resíduos de uma mesma cadeia) e externos (entre resíduos de cadeias diferentes). Para este trabalho, utilizaram-se somente os descritores de contatos internos e esfera de raio 3 a 7Å, com incremento de 1Å

(valores pré-calculados presentes no STING_RDB), totalizando de 10 atributos para a Densidade de Energia de Contatos.

- ✓ **Epatom:** Os átomos que compõem os aminoácidos das proteínas podem, em determinadas condições, apresentar carga elétrica, que interagem com outras regiões carregadas da própria proteína ou ainda com outras moléculas e/ou íons de seu ambiente. Portanto, em um determinado ponto do espaço é possível calcular o potencial eletrostático devido às cargas presentes nas macromoléculas ao redor deste ponto.
- ✓ **Evolutionary Pressure:** Diferentes aminoácidos em uma proteína estão sujeitos a diferentes taxas de variação, ou biologicamente falando, diferentes pressões evolutivas. Um aminoácido em uma posição da sequência que evolui lentamente é dito conservado, enquanto que aqueles em posições que evoluem rapidamente são ditos "variáveis".
- ✓ **FPocket:** Descreve a área e o volume ocupados pelos resíduos de aminoácidos que estão presentes em diferentes cavidades das proteínas, calculados pelo programa FPocket.
- ✓ **FPocket Atom:** Descreve a área e o volume ocupados pelos átomos de resíduos de aminoácidos que estão presentes em diferentes cavidades das proteínas, calculados pelo programa FPocket.
- ✓ **Graph Descriptor:** Cadeias proteicas podem ser representadas como grafos não direcionados onde o conjunto de vértices é composto pelos resíduos de aminoácidos ou os átomos de uma cadeia proteica, enquanto que as arestas do grafo representam interações entre estes resíduos ou átomos. Utiliza-se como conjunto de vértices os resíduos de aminoácidos da proteína, sendo o conjunto de arestas definido de duas formas: utilizando-se contatos interatômicos previamente calculados ou considerando-se resíduos de aminoácidos que se encontram próximos, respeitando uma distância máxima pré-estabelecida entre seus carbonos- α .
- ✓ **HSSP:** HSSP é um banco de dados derivado que mescla informações de estrutura (3-D) e sequência (1-D), que fornece um alinhamento de múltiplas sequências de homólogos putativos e um perfil de sequência característico da família de proteínas, centrado na estrutura conhecida.
- ✓ **Hydrophobic Patch:** Diferentes regiões da superfície das proteínas apresentam diferentes características físico-químicas. Dessa forma, estas regiões podem ser fragmentadas (patches) em regiões majoritariamente hidrofóbicas, polares, apolares, etc. O descritor de patches hidrofóbicos encontra tais fragmentos na superfície das proteínas, que se configuram tipicamente como hidrofóbicas. Tais regiões são também denominadas de hot-spots, devido à sua importância nas

interações entre macromoléculas, uma vez que estas regiões não estabelecem ligações com a moléculas de água, sendo predominante locais de alta energia de contatos entre a proteína e seu ligante e proteína-proteína.

- ✓ **Hydrophobicity:** Aminoácidos são classificados como hidrofóbicos ou hidrofílicos, dependendo de suas afinidades para estabelecerem ligações de hidrogênio com as moléculas de água. Existem diversas escalas que definem a hidrofobicidade relativa dos resíduos de aminoácidos: quanto mais positivo é seu valor, mais hidrofóbicos são os aminoácidos localizados em dada região da molécula.
- ✓ **Ligand Pocket and Water Contact:** Proteínas podem interagir com outras moléculas de diferentes naturezas, como por exemplo moléculas de água. Os descritores Ligand Pocket Residue (LPR) e Water Contact Residue (WCR) descrevem tais interações. Resíduos de aminoácidos que se encontram próximos de átomos de um ligante, considerando um valor máximo entre suas distâncias, são considerados como LPR's, enquanto resíduos de aminoácidos que estão próximos de moléculas de água cocrystalizadas com a estrutura são considerados WCR's.
- ✓ **Mouth:** Descreve a área e o volume ocupados pelos de resíduos de aminoácidos que estão presentes em diferentes cavidades das proteínas. É usado o programa proshape.
- ✓ **Mouth Atom:** Descreve a área e o volume ocupados pelos átomos dos resíduos de aminoácidos que estão presentes em diferentes cavidades das proteínas. É usado o programa proshape.
- ✓ **Pocket:** Descreve a área e o volume ocupados pelos de resíduos de aminoácidos que estão presentes em diferentes cavidades das proteínas.
- ✓ **Pocket Atom:** Descreve a área e o volume ocupados pelos átomos de resíduos de aminoácidos que estão presentes em diferentes cavidades das proteínas.
- ✓ **PROSITE:** PROSITE é um banco de dados de famílias e domínios de proteínas. Baseia-se na observação de que, embora haja um grande número de proteínas diferentes, a maioria delas pode ser agrupada, com base em semelhanças em suas sequências, em um número limitado de famílias. Proteínas ou domínios de proteínas pertencentes a uma família particular geralmente compartilham atributos funcionais e são derivados de um ancestral comum.
- ✓ **Residue:** Informações sobre os resíduos de aminoácidos extraídos do PDB (por exemplo, número do resíduo na sequência primária, tipo de aminoácido, coordenadas x, y, z do carbono- α).
- ✓ **Residue Contacts:** A plataforma Blue Star STING diferencia 14 tipos de contatos intra e intercadeias e mais 12 tipos de contatos entre cadeias proteicas e moléculas

de DNA/RNA, que são armazenados em sua base de dados em nível atômico, ou seja, possui informações sobre quais pares de átomos estão estabelecendo contato e a energia equivalente ao tipo de contato. Residue Contacts é um sumário desses contatos, do nível atômico para o nível dos resíduos de aminoácidos formados por esses átomos.

- ✓ **Rotamer:** Os átomos dos resíduos de aminoácidos se organizam conformacionalmente formando ângulos entre si. Para os átomos da cadeia lateral, podem haver até cinco ângulos, $\chi-1$ até $\chi-5$ (CHI). Uma vez que as cadeias laterais dos aminoácidos possuem comprimentos diferentes, nem todos possuem os mesmos grupos de ângulos.
- ✓ **Side Chain Orientation:** A orientação da cadeia lateral é calculada para cada resíduo de aminoácido em uma cadeia proteica como um ângulo formado entre dois vetores: $C\alpha$ -CENTROID e $C\alpha$ -LHA. O $C\alpha$ -CENTROID é o vetor do átomo $C\alpha$ de um resíduo de aminoácido para o centro de massa de uma região específica (esfera sonda), e o $C\alpha$ -LHA é o vetor do átomo $C\alpha$ ao Último Átomo Pesado desse mesmo resíduo de aminoácido.
- ✓ **Solvation:** A energia de solvatação, corresponde à energia entre ligações de átomo do soluto e do solvente. A energia livre de hidratação do i-ésimo átomo de um resíduo de aminoácido é calculada como sendo o produto entre seu parâmetro de solvatação atômico, determinado experimentalmente, e a área relativa acessível ao solvente (ASArrelativa). O somatório das energias livre de hidratação dos átomos que compõem o resíduo r e os átomos na vizinhança de r , normalizado pela soma das áreas relativas acessíveis ao solvente de todos os átomos considerados, fornece a energia de solvatação do resíduo r .
- ✓ **Space Clash:** Space clash é a medida do choque estérico que ocorre entre os aminoácidos. Muitas vezes, a medida para o choque estérico é um fator de sobreposição, que é calculado como a razão entre as distâncias entre os dois centros dos átomos para a soma dos seus raios de Van der Waals. Esse fator de sobreposição será menor que 1 se as duas esferas se penetram mutuamente. Isso acontece, por exemplo, em um arranjo de ligações de hidrogênio. Fatores de sobreposição com valores entre 0,8 e 0,75 são comuns nas estruturas obtidas por difração de raios X. Fatores de sobreposição igual a 0,7 seria um choque estérico relativamente menor, enquanto valores iguais ou menores que 0,65 seriam mais graves, indicando possível erro no modelo estrutural.
- ✓ **Sting HSSP:** Fornece um alinhamento de múltiplas sequências de homólogos putativos e um perfil de sequência característico da família de proteínas, centrado na estrutura conhecida, baseados em um algoritmo próprio.

- ✓ **Stride:** Diferentemente da estrutura terciária, a estrutura secundária não define posições tridimensionais específicas dos resíduos de aminoácidos, mas sim conformações locais de um grupo de resíduos que interagem entre si para formar diferentes combinações estruturais. Entre as mais encontradas, têm-se hélices alfa, folhas beta e laços (loops). O Stride (Structural Identification) é um algoritmo que define as estruturas secundárias.
- ✓ **Structure:** Informações sobre a estrutura proteica extraídas do arquivo PDB.
- ✓ **Unused Contacts:** Comparando os valores do número de contatos estabelecidos por cada aminoácido em todas as proteínas presente no PDB, é possível calcular quantos contatos não foram estabelecidos em relação ao valor máximo encontrado para os mesmos tipos de aminoácido e contato. O número de contatos não utilizados estabelece um potencial de contatos, que não leva em conta o meio ambiente em que cada aminoácido está inserido.
- ✓ **Weighted Contact Number:** O Número de Contato Ponderado (WCN) é uma medida da flexibilidade da espinha dorsal dos resíduos de aminoácidos.

PRECIFICAÇÃO

Plano	Número máximo de requisições	Valor mensal (R\$)
Gratuito100KPorMes	100 mil requisições por mês	Gratuito

Para utilização de mais de 100 mil requisições por mês entre em contato conosco pelo e-mail agroapi@embrapa.br.