



## Modelagem molecular de proteínas: o caso de uma glucuronosiltransferase (GumK) de *Gluconacetobacter diazotrophicus* PAL5

Luanna Maria Filgueiras Dutra<sup>1</sup>; Bruna Regina dos Santos Silva<sup>2</sup>; Luc Felicianus Marie Rouws<sup>3</sup>; Jean Luis Simões Araújo<sup>4</sup>; Marcia Soares Vidal<sup>5</sup>; José Ivo Baldani<sup>6</sup>; Carlos Henrique Salvino Gadelha Meneses<sup>7</sup>

### Resumo

A  $\beta$ -1,2-glucuronosiltransferase (GumK) de *Gluconacetobacter diazotrophicus* é uma proteína de 41 kDa que está envolvida na biossíntese de exopolissacarídeos. GumK é o membro da família de enzimas ativadoras de glicosiltransferases e é responsável pela adição de um resíduo de ácido UDP-glucurônico para um resíduo de manosil-(alfa-1,3)-difosfopoliprenol-celobiose para produzir glucuronil-(beta-1,2)-manosil-(difosfopoliprenol-cellobiose alfa-1,3). Para compreender melhor o mecanismo de ação das glicosiltransferases bacterianas desta família, a modelagem por homologia de GumK foi executado. A estrutura tridimensional da enzima GumK foi construída usando o programa MODELLER 9v8 e validadas com o programa Procheck. Para construir o modelo de GumK, sequência semelhantes foram encontradas no *Protein Data Bank* quando comparada à GumK de *Xanthomonas campestris*. A enzima possui dois domínios *Rossmann* bem definidos com uma fenda catalítica entre eles, que é uma característica típica da superfamília de glicosiltransferase B. O alinhamento de sequências entre GumK e outras glicosiltransferases mostrou várias regiões conservadas incluindo o motivo presente no sítio de ligação do substrato, todos os resíduos que entram em contato com o ligante UDP são conservadas na família GT70, o que indica que, como esperado, a região de ligação com substrato é conservada. Os dados biológicos e estruturais aqui relatados lançam luz sobre a base molecular para a seletividade nesta família de glicosiltransferases. Estas análises enfatizam o modelo tridimensional construído para GumK que representa a primeira informação estrutural para enzimas envolvidas na síntese da goma de *G. diazotrophicus*.

Unitermos: Modelagem por homologia; Bactéria endofítica; Glicosiltransferase família GT70



<sup>1</sup>Programa de Pós-Graduação em Ciências Agrárias – Universidade Estadual da Paraíba, Campina Grande – PB, Brasil, [luannabeserra-uepb@hotmail.com](mailto:luannabeserra-uepb@hotmail.com)

<sup>2</sup>Bacharelado do curso de Agroecologia – Universidade Estadual da Paraíba, Lagoa Seca – PB, Brasil, [bruna-reggina@hotmail.com](mailto:bruna-reggina@hotmail.com)

<sup>3</sup>Laboratório de Genética e Bioquímica, Embrapa Agrobiologia, Rodovia BR 465, km 07, 23890-000, Seropédica, RJ, Brasil, [luc.rouws@embrapa.br](mailto:luc.rouws@embrapa.br)

<sup>4</sup>Laboratório de Genética e Bioquímica, Embrapa Agrobiologia, Rodovia BR 465, km 07, 23890-000, Seropédica, RJ, Brasil, [jean.araujo@embrapa.br](mailto:jean.araujo@embrapa.br)

<sup>5</sup>Laboratório de Genética e Bioquímica, Embrapa Agrobiologia, Rodovia BR 465, km 07, 23890-000, Seropédica, RJ, Brasil, [marcia.vidal@embrapa.br](mailto:marcia.vidal@embrapa.br)

<sup>6</sup>Laboratório de Genética e Bioquímica, Embrapa Agrobiologia, Rodovia BR 465, km 07, 23890-000, Seropédica, RJ, Brasil, [ivo.baldani@embrapa.br](mailto:ivo.baldani@embrapa.br)

<sup>7</sup>Programa de Pós-Graduação em Ciências Agrárias – Universidade Estadual da Paraíba, Campina Grande – PB, Brasil, [carlos@ccaa.uepb.edu.br](mailto:carlos@ccaa.uepb.edu.br)

## Abstract

*Gluconacetobacter diazotrophicus* GumK ( $\beta$ -1,2-glucuronosyltransferase) is a 41-kDa protein that is involved in exopolysaccharide biosynthesis. GumK is the founding member of the glycosyltransferase 70 family of carbohydrate-activating enzymes and is responsible for the addition of a UDP-glucuronic acid residue to mannosyl-( $\alpha$ -1,3)-cellobiose diphosphopolyprenol to produce glucuronyl-( $\beta$ -1,2)-mannosyl-( $\alpha$ -1,3)-cellobiose diphosphopolyprenol. To better understand the mechanism of action of the bacterial glycosyltransferases in this family, the homology modeling structure of GumK was executed. Three-dimensional structure to GumK enzymes we constructed using MODELLER 9v8 and validated with Procheck and Verify 3D programs. To construct the model of GumK, Protein Data Bank found sequence similarities when compared to GumK (from *Xanthomonas campestris*). The enzyme has two well defined Rossmann domains with a catalytic cleft between them, which is a typical feature of the glycosyltransferase B superfamily. The sequence alignment between GumK and another glycosyltransferases showed several preserved regions including the motif present on the substrate binding site, all residues that contact the UDP are conserved in the GT70 family, which indicates that, as expected, binding of the donor substrate is conserved. The biological and structural data reported here shed light on the molecular basis for donor and acceptor selectivity in this glycosyltransferase family. These analyses emphasize



the three-dimensional model constructed for GumK that represents the first structural information for enzymes involved in gum synthesis of *G. diazotrophicus*.

Uniterms: Modeling by homology; Endophytic bacteria; Glycosyltransferase GT70 family

## Introdução

A Fixação Biológica de Nitrogênio (FBN) foi inicialmente descrita em bactérias diazotróficas da rizosfera e do rizoplane de uma grande variedade de plantas não leguminosas (DÖBEREINER, 1992), porém, evidências mostram que algumas bactérias, que colonizam o interior da planta, também são capazes de fixar nitrogênio e, por este fato, são conhecidas como bactérias endofíticas fixadoras de nitrogênio ou bactérias endofíticas diazotróficas (OLIVARES *et al.*, 1996). *Gluconacetobacter diazotrophicus* é uma bactéria diazotrófica, adaptada às condições de pressão osmótica e pH, que prevalecem no interior dos tecidos da cana-de-açúcar (GILLIS *et al.*, 1989), já tendo sido isolada de raízes, colmos e folhas de cana-de-açúcar, capim-elefante e batata-doce (DÖBEREINER, 1992).

Os exopolissacarídeos (EPS), também conhecidos como biopolímeros ou gomas hidrossolúveis, são produzidos por uma grande variedade de microrganismos. Estes possuem um papel importante tanto na sobrevivência da bactéria quanto na associação desta com a planta, auxiliando na adesão da célula a superfícies, na concentração de nutrientes, evitando a dessecação, o contato com o hospedeiro e protegendo contra moléculas tóxicas (SUTHERLAND, 2001).

O papel de moléculas exopolissacarídicas de alto e baixo peso molecular na associação entre bactérias fixadoras de nitrogênio e plantas tem sido muito estudado tanto em nível de caracterização química, como de expressão gênica.

Considerando que o sequenciamento do genoma da *G. diazotrophicus* foi finalizado, e que uma das motivações para este sequenciamento foi o seu potencial biotecnológico, torna-se de extrema importância o estudo comparativo das estruturas terciárias de enzimas relacionadas com a síntese de EPS, com diferentes especificidades pelo substrato, a fim de propor potenciais sítios

para mutação nas enzimas relacionadas com a síntese de EPS de *G. diazotrophicus*, que implicariam na ampliação da utilização de substratos por estas enzimas. Informações cruciais sobre a função de proteínas podem ser obtidas a partir das suas estruturas tridimensionais (3D).

A predição da estrutura 3D de proteínas codificadas no genoma é o objetivo da genômica estrutural (BAKER e ŠALI, 2001), podendo ser realizada por métodos experimentais ou teóricos. As principais técnicas experimentais empregadas na predição estrutural de uma proteína são: a cristalografia por difração de raios-X (RX) e a espectroscopia de ressonância magnética nuclear (RMN). Mesmo sendo considerado o melhor processo para a obtenção da estrutura tridimensional de uma proteína, os métodos de predição experimental além de serem financeiramente custosos, podem levar anos e, em alguns casos, a estrutura final pode não chegar a ser obtida. Além disso, um número elevado de sequências protéicas encontram-se depositadas em bancos de dados como o “Swiss-Prot” (BAIROCH *et al.*, 2004), contudo, poucas estruturas de proteínas foram resolvidas experimentalmente e depositadas em bancos de dados como o PDB (*Protein Data Bank*) (WESTBROOK *et al.*, 2002). Diante disso, o emprego de métodos teóricos para a predição da estrutura tridimensional de proteínas seria uma saída para as limitações e entraves encontrados nos métodos de predição experimental.

Os principais métodos teóricos de predição de estrutura 3D de proteínas são: Modelagem Comparativa; Predição de Enovelamento de Proteínas (*Threading*) e Predição por Primeiros Princípios (*ab initio*). A modelagem comparativa, até pouco tempo conhecida como modelagem por homologia, usa como molde a estrutura 3D de outra proteína de referência, já resolvida experimentalmente e com coordenadas cartesianas depositadas em banco de dados de estruturas. Este tem sido o método mais frequentemente usado e seu limite de predição está intrinsecamente relacionado com o grau de similaridade entre as estruturas alvo e molde. A Predição de Enovelamento de Proteínas tenta ajustar a estrutura da proteína de interesse aos tipos de enovelamentos de proteínas conhecidas, ou seja, registradas e depositados em bibliotecas de enovelamentos. Já a Predição por Primeiros Princípios emprega as propriedades físico-químicas conhecidas de cada aminoácido para a construção de funções de energia, sendo que tais funções



são minimizadas por algoritmos que realizam buscas no espaço de conformações que a proteína de interesse possa assumir (ŠALI e BLUNDELL, 1993; PETSKO e RINGE, 2004).

Diante do exposto acima, o presente trabalho teve como objetivo descrever simplificada a metodologia de modelagem comparativa empregando como modelo a proteína  $\beta$ -1,2-glicuronosiltransferase (GumK) de *Gluconacetobacter diazotrophicus*, que está envolvida na biossíntese de exopolissacarídeos. GumK, é uma proteína de 41 kDa, pertencente a família das glicosiltransferases 70, enzima responsável pela adição de carboidratos ativados (ácido UDP-glicurônico) a um resíduo de difosfoliprenol manosil ( $\alpha$ -1,3) para produzir celobiose glucuronil-( $\beta$ -1,2 difosfoliprenol)-manosil ( $\alpha$ -1,3)-celobiose.

## Materiais e Métodos

A metodologia de modelagem comparativa envolve quatro etapas sucessivas: identificação e seleção da proteína-molde; alinhamento das sequências alvo e molde; construção e otimização do modelo e, validação do modelo.

A sequência de aminoácidos da proteína GumK de *G. diazotrophicus* foi submetida a uma análise comparativa através do programa BLAST (Basic Local Alignment Search Tool) ([www.ncbi.nih.gov/BLAST](http://www.ncbi.nih.gov/BLAST)) (ALTSCHUL *et al.*, 1990) na base de dados PDB (alinhamento local). As sequências de aminoácidos provenientes de alguns dos microrganismos encontrados no resultado da busca foram submetidas ao programa de alinhamento múltiplo de sequências CLUSTALX). Nesta primeira etapa foi identificada uma proteína relacionada à sequência de aminoácidos de GumK de *G. diazotrophicus* (proteína alvo), levando-se em conta durante o processo de escolha da proteína molde, aspectos como: conhecimento estrutural, similaridade de sequência, similaridade de função, expressão pelo mesmo grupo de genes e correlação evolutiva entre as proteínas

Identificada a proteína molde, métodos de alinhamento global entre as sequências foram aplicados, o que consiste na segunda etapa da modelagem comparativa. O alinhamento entre as sequências foi realizado com o emprego do programa CLUSTALW (<http://www2.ebi.ac.uk/clustalw/>). O objetivo desse alinhamento é reconhecer regiões



estruturalmente conservadas e regiões variáveis, observando-se os resíduos estruturalmente equivalentes na sequência primária.

A terceira etapa da modelagem é a construção do modelo, a qual foi baseada na informação contida no alinhamento gerado entre as sequências. A construção do modelo foi realizada com o programa MODELLER (<http://www.salilab.org/modeller/>); no entanto, vale ressaltar aqui a existência de outros programas como por exemplo o SWISS-MOD (ŠALI e BLUNDELL, 1993; GUEX e PEITSCH, 1997; ARNOLD *et al.*, 2006; ESWAR *et al.*, 2007). A modelagem utiliza a satisfação das restrições espaciais, utilizando distâncias geométricas e técnicas de otimização para satisfazer as restrições espaciais das proteínas molde dadas através do alinhamento. A visualização da estrutura tridimensional foi realizada com o programa RASMOL (SAYLE, 1995).

## Resultados e Discussão

A validação do modelo obtido pode ser realizada em três níveis: estereoquímico, conformacional e energético. No presente trabalho, a qualidade do modelo 3D gerado foi validada pelo gráfico de Ramachandran (RAMACHANDRAN *et al.*, 1963) com o programa PROCHECK (<http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html>) (LASKOWSKI *et al.*, 1998), que avalia a estrutura tridimensional da proteína indicando a sua qualidade estereoquímica. A qualidade do modelo está ligada à escolha do molde e à construção do alinhamento que servirá de referência para a construção dos átomos que compõem a proteína.

	10	20	30	40	50	60	70	80	90	100																																																																													
Gluconacetobacter diazotrophicus	RTGS	LVISVHDYYSRREKASVHFITAEAKRGE	TRFFSVGLSRLSEHR	EDTRADAPR	RRR	ETVD	QVECF	WR	A	WHP	FNIR	PPAA																																																																											
Xanthomonas Campestris	RGVSP	AA	PAS	GIR	RPCYL	VLS	SH	DF	TP	RR	NI	HF	ITD	QL	LR	GT	TR	FF	SL	RY	SRL	SR	M	K	D	M	R	L	E	DD	T	NT	V	V	S	H	N	G	Y	L	W	T	T	V	H	P	E	N	T	R	S	W	R																																		
	110	120	130	140	150	160	170	180	190	200																																																																													
Gluconacetobacter diazotrophicus	PVEESH	WLR	RRMI	RRIP	RQL	CD	A	T	F	I	E	S	H	A	P	V	F	D	R	R	L	H	T	R	I	V	L	S	D	D	D	V	G	C	A	E	T	K	R	R	R	A	N	F	S	I	A	A	R	L	P	S	R	L	L	R	D	G	L	P	H	D	R	T	S																						
Xanthomonas Campestris	PVED	A	R	R	W	A	A	H	P	K	Q	L	L	D	H	M	R	E	S	V	V	F	E	S	H	I	V	A	P	E	L	A	K	M	V	H	P	A	K	L	V	R	A	S	D	G	L	S	T	H	V	A	S	Y	E	N	E	D	R	V	A	P	T	L	V	I	A	L	V	S	P	A	M	A	E	V	V	S	R	D	N	V					
	210	220	230	240	250	260	270	280	290	300																																																																													
Gluconacetobacter diazotrophicus	YIPQ	I	D	R	T	A	E	R	S	Y	N	F	G	N	R	R	S	C	S	M	G	S	M	L	F	A	S	E	T	I	A	A	E	F	D	I	D	H	I	G	A	K	A	D	G	L	S	A	P	I	I	H	Q	S	P	F	Q	L	P	F	V	Q	H	A	F	G	V	A	P	L	D	R	T	E	R												
Xanthomonas Campestris	HVGH	V	H	N	L	Q	L	G	D	S	P	Y	A	E	G	I	H	A	V	G	S	M	L	F	D	P	E	F	V	V	A	S	K	A	F	H	V	T	H	V	I	G	S	M	G	R	H	F	G	Y	G	D	N	V	T	V	Y	G	E	K	H	A	T	I	G	Y	I	K	A	R	F	G	I	A	P	A	S	E	Q	V	V	V					
	310	320	330	340	350	360	370	380	390	400																																																																													
Gluconacetobacter diazotrophicus	L	L	T	S	L	K	I	R	F	G	L	F	G	I	P	A	I	C	D	F	A	H	D	Q	P	G	R	Y	G	R	F	G	P	A	S	I	R	Q	A	D	D	A	H	S	D	P	I	R	P	M	G	N	E	V	V	D	R	I	L	T	E	E	A	F	A	D	T	R	I																		
Xanthomonas Campestris	L	A	S	S	M	K	L	L	Y	D	F	G	L	F	G	L	V	A	V	N	A	V	V	P	Y	K	S	R	F	G	T	P	G	N	A	D	V	I	A	T	T	Q	L	E	A	P	R	V	R	Y	R	C	N	N	S	D	T	D	R	V	L	D	R	A	V	P	E	T	R	I	Y	P	H	P	P	T	A	A	Q	L	S	E	A	A	L	S	H
Gluconacetobacter diazotrophicus	.....																																																																																						
Xanthomonas Campestris	HHHHHH																																																																																						

**Figura 1.** Alinhamento entre as sequências de aminoácidos das proteínas GumK de *G. diazotrophicus* com 2HY7 de *X. campestris*. Letras iguais pareadas (sombreadas em vermelho) representam o mesmo aminoácido.

Na busca por proteínas molde foi encontrada uma  $\beta$ -1,2-glucuronosiltransferase (GumK), da super-família das glicosiltransferases do tipo B de *X. campestris*, que foi usada como sequência molde, com identidade de 42%. Vale ressaltar que para uma sequência ser considerada como candidata a estrutura de referência, é necessário que haja no mínimo 25-30% de identidade entre as sequências primárias, o que foi obtido para a proteína de *X. campestris*. A modelagem molecular de GumK de *G. diazotrophicus* foi feita com o programa Modeller (SALI e BLUNDELL, 1993), e a sequência molde escolhida foi a 2HY7 [GumK de *X. campestris* (BARRERAS et al., 2008)], devido ao seu índice maior de identidade (Figura 1). A proteína molde havia sido resolvida por difração de raios X, com resolução de 1,9 Angstroms.

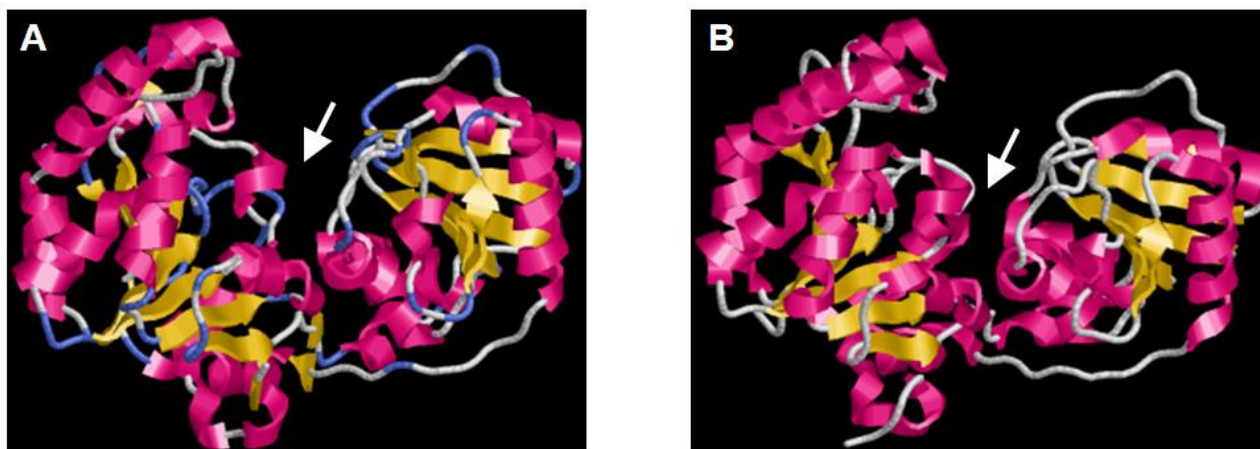
Uma vez escolhida a proteína molde (2HY7), a sequência desta foi alinhada com a sequência de GumK de *G. diazotrophicus*. O objetivo do alinhamento foi justapor resíduos estruturalmente equivalentes, levando em consideração características estruturais comuns como, por exemplo, regiões de aminoácidos conservadas (Figura 1). Este alinhamento foi submetido como um arquivo com extensão **.ali** para o programa de construção do modelo, o MODELLER.

O alinhamento é a entrada para a etapa da modelagem, juntamente com a estrutura da proteína 2HY7. O arquivo gerado pelo programa consiste em um modelo tridimensional para a sequência alvo que contém todos os átomos da cadeia principal e da cadeia lateral da proteína. No alinhamento das sequências da proteína 2HY7 (Molde) e GumK de *G. diazotrophicus* (Alvo) mostrado na Figura 1, é possível verificar quais aminoácidos são idênticos e quais são conservados. Estas características encontradas no alinhamento indicam que a estrutura cristalográfica de 2HY7 pode ser usada como estrutura molde.

Desta modelagem inicial foram obtidos três modelos, que em seguida foram avaliados de acordo com a função objetiva e com os parâmetros estereoquímicos. A Figura 2 mostra o modelo

preliminar obtido, onde o critério inicial para seleção do melhor modelo foi o de menor energia, fornecido pelo próprio programa. Após essa seleção um modelo mais representante foi selecionado levando-se em consideração a porcentagem de identidade e o tamanho do segmento da proteína modelada. Em seguida foram geradas as coordenadas da estrutura através do programa MODELLER, e estas foram visualizadas com o programa de visualização de estruturas RASMOL.

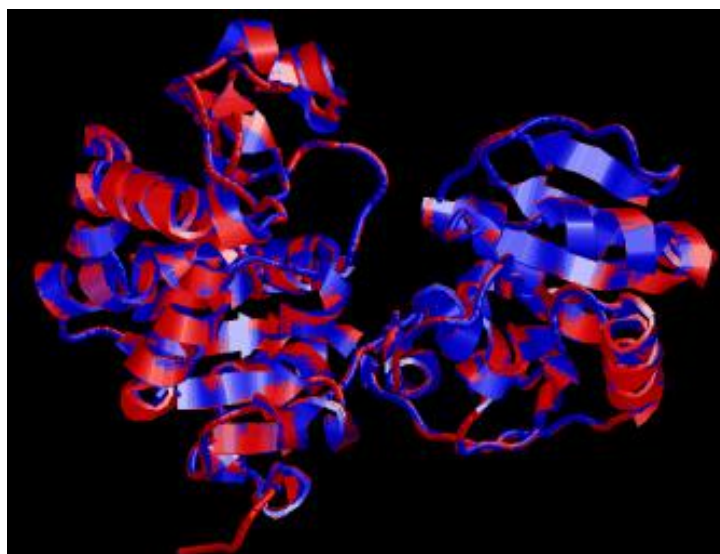
Os resultados mostraram que a proteína possui dois domínios bem definidos, denominados domínios Rossmann, com uma fenda catalítica entre eles (Figura 2), o que é uma característica típica da superfamília das glicosiltransferase do tipo B.



**Figura 2.** Visualização do modelo da estrutura de tridimensional da proteína GumK de *G. diazotrophicus* pelo programa RASMOL. (A) Modelo obtido para GumK de *G. diazotrophicus*. (B) Estrutura para GumK de *X. campestris*, utilizada como molde (BARRERAS *et al.*, 2008). As setas em cada uma das figuras indicam suas respectivas fendas catalíticas.

A figura 3 mostra a sobreposição do modelo gerado com a estrutura da molécula molde (alinhamento estrutural). Nota-se que, além da alta porcentagem de identidade de sequência, a estrutura terciária é bastante conservada devido ao tipo de alinhamento global/estrutural que a sequência de aminoácidos de GumK de *G. diazotrophicus* foi submetida.

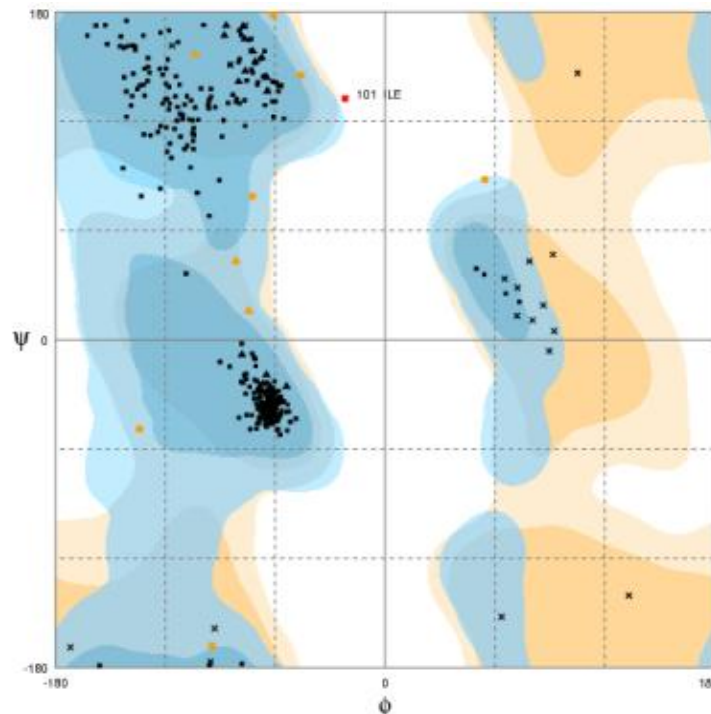




**Figura 3.** Visualização do alinhamento estrutural da proteína GumK de *G. diazotrophicus* (em azul) e GumK de *X. campestris* (em vermelho) pelo programa RASMOL.

O gráfico de Ramachandran pode ser observado na Figura 4. A estrutura deve apresentar acima dos 90% de seus resíduos nas regiões favoráveis (região azul), ou seja, nas regiões A, B e L. Devem-se desconsiderar os resíduos de glicina (apresentados na figura como triângulos), já que eles não possuem cadeia lateral, os de prolina, pois nestes o carbono alfa está ligado à cadeia lateral e os resíduos das extremidades (C-terminal e N-terminal), que apresentam padrões estereoquímicos diferentes dos outros resíduos.

A porcentagem de resíduos encontrados nas regiões favoráveis foi de 95% (em azul escuro), nas regiões adicionalmente permitidas (Regiões a, b, l, p, ou seja, azul claro) de 3,5%, nas regiões generosamente permitidas (Regiões ~a, ~b, ~l, ~p, ou seja, amarelo claro) foi de 1,5% e nas regiões desfavoráveis (Região em branco) foi de 0,0%. Os resíduos que se encontram em regiões desfavoráveis são um reflexo das estruturas usadas como molde, e alguns se encontram nas extremidades da proteína. Sendo assim, os dados do gráfico de Ramachandran suportam o modelo obtido. Em outras palavras, a última etapa do processo de modelagem comparativa, a validação do modelo, foi finalizada com sucesso.



**Figura 4.** Gráfico de Ramachandran do modelo final obtido para GumK de *G. diazotrophicus*, gerado pelo programa Procheck. Os ângulos *Phi* e *Psi* estão em graus. Em azul escuro estão as regiões favoráveis; em azul claro as regiões adicionalmente permitidas; em amarelo claro as regiões generosamente permitidas e em branco as regiões desfavoráveis.

Os dados biológicos e estruturais relatados acima abrem as portas para os estudos de base molecular entre esta proteína e seu respectivo ligante, podendo assim mostrar uma possível seletividade nesta família de glicosiltransferases. Este é o primeiro modelo construído para representar a estrutura tridimensional para GumK de *G. diazotrophicus*, uma das enzimas envolvidas na síntese de EPS em *G. diazotrophicus*.

## REFERENCIAS BIBLIOGRÁFICAS

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MEYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. **Journal of Molecular Biology**. v.215, p.403-410, 1990.



ARNOLD, K.; BORDOLI, L.; KOPP, J.; SCHWEDE, T. The swiss-model workspace: a web-based environment for protein structure homology modelling. **Bioinformatics**, v.22, n.2, p.195-201, 2006.

BAIROCH, A.; BOECKMANN, B.; FERRO, S.; GASTEIGER, E. Swiss-prot: juggling between evolution and stability. **Briefings in Bioinformatics**, v.5, n.1, p.390-55, 2004.

BAKER, D.; ŠALI, A. Protein structure prediction and structural genomics. **Nature**, v.294, p.93-96, 2001.

BARRERAS, M.; SALINAS, S. R.; ABDIAN, P. L.; KAMPEL, M.A. and IELPI, L. Structure and mechanism of GumK, a membrane-associated glucuronosyltransferase. **Journal of Biological Chemistry**, v.283, p.25027-25035, 2008.

DÖBEREINER, J. Recent changes in concepts of plant bacteria interaction: endophytic N<sub>2</sub> fixing bacteria. **Ciência e Cultura**, v.44, p.310-313, 1992.

ESWAR, N.; WEBB, B.; MARTI-RENOM, M.A.; MADHUSUDHAN, M.; ERAMIAN, D.; SHEN, M.-Y.; PIEPER, U.; SALI, A. (2007). **Comparative Protein Structure Modeling Using MODELLER**. In: Current Protocols in Protein Science. volume 50. pages 1–31, John Wiley & Sons, Inc.

GILLIS, M.; KERTERS, B.; HOSTE, DJ.; KROOPPENSTEDT, R.M.; STEPHAN, M.P.; TEIXEIRA, K.R.S.; DÖBEREINER, J.; DE LEY, J. *Acetobacter diazotrophicus* sp. nv. A nitrogen fixing acetic acid bacterium associated with sugar cane. **International Journal of Systematic Bacteriology**, v.39, p.361-364, 1989.

GUEX, N.; PEITSCH, M.C. Swiss-model and swiss-pdb viewer: An environment for comparative protein modeling. **Electrophoresis**, v.18, p.2714, 1997.

LASKOWSKI, R.A.; MACARTHUR, M.W.; MOSS D.S.; THORNTON, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v.26, p.283-291, 1998.

OLIVARES, F.L.; BALDANI, V.L.D.; REIS, V.M.; BALDANI, J.I.; DÖBEREINER, J. Occurrence of the endophytic diazotroph *Herbaspirillum* spp. in roots, stems and leaves predominantly of Gramineae, **Biology and Fertility of Soils**, v.2, p.197-200, 1996.

PETSKO, G.A.; RINGE, D. **Protein Structure and Function**. New Science Press. 2004, 195p.

RAMACHANDRAN, G.N.; RAMAKRISHNAN C.; SASISEKHARAN, V. Stereochemistry of Polypeptide Chain Conformations. **Journal of Molecular Biology**, v.7, p.95-99, 1963.

ŠALI, A.; BLUNDELL, T.L. Comparative protein modeling by satisfaction of spatial restraints. **Journal of Molecular Biology**, v.234, p.779-815, 1993.



SAYLE, R. RASMOL. **Trends in Biochemical Sciences**. v.20, p.374-379, 1995.

SUTHERLAND, I.W. Biofilm exopolysaccharides: a strong and sticky framework. **Microbiology**, v.147, p.3-9, 2001.

WESTBROOK, J.; FENG, Z.; JAIN, S.; BHAT, T.N.; THANKI, N., RAVICHANDRAN, V., GILLILAND, G.L.; BLUHM, W.F.; WEISSIG, H.; GREER, D.S.; BOURNE, P.E.; BERMAN, H. M. The protein data bank: unifying the archive. **Nucleic Acids Research**, v.30, n.1, p.245-248, 2002.