



**UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA**



**JOSÉ JORGE SILVA JÚNIOR**

**DESCOBERTA DE NOVOS INIBIDORES PARA A UDP-N-  
ACETILGLICOSAMINA PIROFOSFORILASE DO  
*Moniliophthora perniciosa* POR TRIAGEM VIRTUAL**

Feira de Santana, BA  
2014

**JOSÉ JORGE SILVA JÚNIOR**

**DESCOBERTA DE NOVOS INIBIDORES PARA A UDP-N-  
ACETILGLICOSAMINA PIROFOSFORILASE DO  
*Moniliophthora perniciosa* POR TRIAGEM VIRTUAL**

Dissertação apresentada ao Programa de Pós-graduação em Biotecnologia, da Universidade Estadual de Feira de Santana como requisito parcial para obtenção do título de Mestre em Biotecnologia.

Orientador: Prof. Dr. Manoelito Coelho Santos Junior  
Co-orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Sandra Aparecida de Assis

Feira de Santana, BA  
2014

**JOSÉ JORGE SILVA JÚNIOR**

**DESCOBERTA DE NOVOS INIBIDORES PARA A UDP-N-  
ACETILGLICOSAMINA PIROFOSFORILASE DO  
*Moniliophthora perniciosa* POR TRIAGEM VIRTUAL**

Dissertação apresentada ao Programa de Pós-graduação em Biotecnologia UEFS/FIOCRUZ,  
como requisito parcial para obtenção do título de Mestre.

**Banca Examinadora**

---

Prof. Dr. Manoelito Coelho Santos Junior (Orientador)  
Universidade Estadual de Feira de Santana

---

Prof. Dr. Alex Gutterres Taranto  
Universidade Federal de São João Del-Rei

---

Prof. Dr. Samuel Silva da Rocha Pita  
Universidade Federal da Bahia

Feira de Santana – BA

2014

Dedico esta obra que mudou totalmente o rumo da minha vida às minhas três mulheres, minha mãe Ana Maria, minha filha Maria Eduarda, minha esposa Paloma Poliana, retribuindo com muita modéstia o amor, a dedicação, o carinho e a paciência.

## AGRADECIMENTOS

Gostaria de agradecer inicialmente a minha família e amigos, que realmente acreditaram nos meus sonhos e me dispuseram o apoio, incentivo para superar os obstáculos, em especial a minha mãe **Ana Maria Mota Silva**, que está sempre vibrando a cada conquista minha e, acima de tudo, estendendo a mão em cada momento difícil.

À **Paloma Poliana Souza Santana** por sempre estar ao meu lado, me incentivando e compreendendo as minhas ausências.

À minha filha **Maria Eduarda Souza Mota Silva** que apesar de não entender ainda a transformação em nossas vidas me forneceu diversas vezes o alívio necessário em momentos de tensão, com a sua graça, carinho, inocência e carisma.

Ao meu orientador e grande Mestre, **Dr. Manoelito Coelho Santos Junior**, por acreditar na minha capacidade, conceder a oportunidade, ter a paciência, a compreensão necessária e contribuir muito para meu crescimento acadêmico, científico, pessoal e profissional.

À **Dra. Sandra Assis** minha co-orientadora por ter aceitado o desafio e ter acreditado em meu potencial.

Aos **Mestres, Franco Henrique, Odailson Paz, André Teles** e à **Dra. Catiane Sacramento Souza**, pelas contribuições e disponibilidade em ajudar na minha jornada contribuindo para a minha formação científica durante o mestrado.

A todos os meus amigos do **LMM**, o nobre **Vinicius Pinto, Genilson Costa** e os demais colegas que ajudaram na construção do saber e aperfeiçoamento do trabalho. Também os colegas e professores da Pós- graduação em Biotecnologia.

À **Universidade Estadual de Feira de Santana**, por ter me proporcionado um ensino de qualidade e ter ampliado os meus horizontes de atuação profissional e científica.

A **CAPES** pela concessão da bolsa, imprescindível para a realização deste trabalho.

“Seja quem você for, seja qualquer posição que você tenha na vida, do nível altíssimo ao mais baixo, tenha sempre como meta muita força, muita determinação e sempre faça tudo com muito amor e com muita fé em Deus. Que um dia você chega lá, de alguma maneira você chega lá.”

*Ayrton Senna da Silva.*

## RESUMO

Pragas são responsáveis por elevadas perdas na produção de cacau no Brasil e no mundo, dentre elas a vassoura-de-bruxa (VB) é uma das mais importantes e destrutivas para o cacauero, chegando a causar perdas de até 95% da produção. Essa praga disseminou-se muito facilmente no estado da Bahia devido a condições ambientais que proporcionaram a propagação da VB, causada pelo fungo *Moniliophthora perniciosa*. Diversos compostos químicos vêm sendo testados com o objetivo de prevenir ou erradicar a VB, porém não foram obtidos bons resultados, portanto, o presente trabalho teve como principal objetivo realizar ensaios *in silico*, a fim de identificar inibidores da UDP-N-acetilglicosamina pirofosforilase (UNAcP) do *M. perniciosa*. Para tanto, foram empregados métodos computacionais na busca de novos inibidores para a UNAcP, onde foram realizadas diferentes etapas de busca e avaliação. A etapa inicial da triagem virtual consistiu na escolha da função de pontuação, assim, foram avaliadas as seguintes funções de pontuação: Broyden-Fletcher-Goldfarb-Shanno (BFGS) presente no AUTODOCK VINA 1.1.2; Grid Score e Grid Score+Hawkins GB/SA ambos presentes no DOCK 6.5 e o cálculo do escore de consenso. Os resultados foram analisados através do cálculo de Fator de Enriquecimento (FE), análise da curva ROC e sua respectiva Área Sobre a Curva (AUC). O Grid Score apresentou  $FE_{(5)}=7,85$ . A análise da curva ROC permitiu observar que a função Grid Score consegue identificar quase 40% das moléculas ativas com menos de 10% do banco de dados (moléculas ativas e falso positivos), a análise da AUC demonstrou que o Grid Score tem maior exatidão ( $AUC=0,87$ ). Assim os resultados da avaliação apontaram o Grid Score como melhor função de pontuação para esse sistema. Foi utilizado um banco de dados composto por moléculas oriundas de fontes naturais. Os dez melhores resultados da triagem virtual feita no DOCK6.5 foram submetidos à plataforma *on line ChemGPS-NP*, para o cálculo dos descritores químicos. Assim, as moléculas foram recategorizadas, baseando-se nos valores do Grid Score do DOCK6.5 e descritores químicos do *ChemGPS-NP*. Os resultados apontaram a molécula ZINC68592326 com a melhor pontuação, a análise das interações intermoleculares indica que esta molécula apresenta interações hidrofóbicas com os resíduos Ala380, Gln113, Gli112, Gli381, Ser168, Arg383, Pro221 e ligação de hidrogênio do tipo aceptora com distância de 3,32Å com o resíduo Asn224. A triagem virtual em banco de dados de moléculas oriundas de produtos naturais permitiu a investigação com um universo de estruturas com características muito diversas. Foi possível obter moléculas com grande diversidade estrutural entre os primeiros do ranking, porém, foram encontradas também moléculas muito similares às moléculas de referência. A utilização de métodos quimiométricos, é considerada muito útil e permitem uma escolha sistemática e consistente das estruturas, principalmente por levar em consideração, descritores químicos e características moleculares, permitindo uma avaliação mais criteriosa.

**Palavras chaves:** *Moniliophthora perniciosa*, UDP-N-acetilglicosamina pirofosforilase, triagem virtual, quimiometria.

## ABSTRACT

Pests are responsible for high losses in cocoa production in Brazil and other countries, among them the witches' broom (WB) is one of the most important and destructive to the cocoa, even causing losses of up to 95% of production. This plague has spread very easily in the state of Bahia due to environmental conditions that provided the spread of WB, caused by the fungus *Moniliophthora perniciosa*. Several chemical compounds have been tested in order to prevent or eradicate WB, however it has not showed good results, so the present study aimed to perform *in silico* assays were obtained in order to identify inhibitors of UDP-N-acetylglucosamine pyrophosphorylase (UNAcP) of *M. perniciosa*. For achieve this goal, computational methods have been employed in the search for new inhibitors for UNAcP where different stages of research and evaluation were performed. The initial stage of virtual screening consisted of the choice of the scoring function, thus the following scoring functions were evaluated: Broyden Fletcher Goldfarb Shanno (BFGS) present in AutoDock VINA 1.1.2; Grid and Grid Score Score+Hawkins GB/SA both present in DOCK 6.5 and calculate the consensus score. The results were analyzed by calculating Enrichment Factor (EF) analysis of the ROC curve and its respective Area under an ROC curve (AUC). The Grid Score presented  $EF_{(5)}=7.85$ . The ROC curve analysis allowed us to observe that the Grid Score function can identify almost 40% of active molecules with less than 10% of the database (false positive and active molecules), AUC analysis demonstrated that the Grid Score has greater accuracy (AUC=0.87). Thus, these results showed what the Grid Score was the best scoring function for this system. A database composed of molecules derived from natural sources was also used. The top ten results of virtual screening of the DOCK6.5 underwent online platform ChemGPS-NP, for the calculation of chemical descriptors. Thus, the molecules were re-categorized, based on the values of the Grid Score DOCK6.5 and chemical descriptors ChemGPS-NP. The results indicate the ZINC68592326 molecule his the best score and the analysis indicates that this has hydrophobic interactions with Ala380, Gln113, Gly112, Gly381, Ser168, Arg383, Pro221 and hydrogen bond interaction (3.32Å) with Asn224. The virtual screening database of molecules derived from natural products research allowed with a universe of structures with very different characteristics. It was possible to obtain molecules with great structural diversity between the top ranking, however, also found very similar to the reference molecules. The use of chemometric methods is considered very useful and allows a systematic and consistent choice of structures, mainly by taking into account chemical descriptors and molecular characteristics, allowing for a more detailed evaluation.

**Keywords:** *Moniliophthora perniciosa*, UDP-N-acetylglucosamine pyrophosphorylase, virtual screening, chemometric.



## LISTA DE FIGURAS

- Figura 1** – Área isenta de contaminação (A); área de plantio afetada por VB (B); vassouras em lançamentos foliares (C); Fruto com área afetada (D) e (E).
- Figura 2** – Basidiomas de *Moniliophthora perniciosa* em fruto de cacau (A), em vassouras (B).
- Figura 3** – Esquema da via metabólica da síntese da quitina.
- Figura 4** – Etapas da reação catalisada pela UNAcP em procariotos
- Figura 5** – Reação de formação da UDP-GlcNAc Mecanismo  $S_N2$
- Figura 6** – Modelo de descritores que permitem a localização das moléculas no espaço químico
- Figura 7** – Posicionamento das moléculas no espaço químico: betacianinas em violeta, betaxantinas em amarelo, e os muscaflavinas em preto.
- Figura 8** – Modelos das principais propriedades traduzidos em descritores químicos para as três primeiras dimensões: A) *ChemGPS* e B) *ChemGPS-NP*.
- Figura 9** – Modelo da estrutura tridimensional construído para a UNAcP do fungo *Moniliophthora perniciosa*.
- Figura 10** – Cluster gerado pelo DUD-E a partir dos estados de protonação da molécula UDP-N-acetilglicosamina.
- Figura 11** – Fator de enriquecimento para as funções de pontuação utilizadas em 5, 10, 25, 50, 75 e 100% de amostragem de dados.
- Figura 12** – Avaliação das funções de pontuação quanto à capacidade de recuperação de compostos ativos frente a compostos falsos positivos.
- Figura 13** – Representação em 3D da localização espacial baseados nos componentes principais gerados pela plataforma *on line ChemGPS-NP*. Em rosa UMP; em vermelho a Uridina; em azul e verde o *Cluster* de UNAcGln; em laranja UNAcGal; em preto as 10 melhores classificados na triagem virtual.
- Figura 14** – Interação no sitio ortostérico entre o modelo da UNAcP e as moléculas selecionadas após a recategorização de variáveis em (A) ZINC68592326, em (B) ZINC68592201 e em (C) ZINC68600011.

## LISTA DE TABELAS

- Tabela 1 -** Ligantes selecionados com valores de  $K_i$  estabelecidos.
- Tabela 2 -** Comparativo do fator de enriquecimento entre as funções e o escore por consenso.
- Tabela 3 -** Classificação dos intervalos das AUC.
- Tabela 4 -** Compostos selecionados na triagem virtual utilizando a função de pontuação Grid Score do DOCK6.5.
- Tabela 5 -** Distância Euclidiana entre os compostos selecionados e os ligantes.
- Tabela 6 -** Recategorização de variáveis ligantes selecionados relacionando Grid Score e Distância Euclidiana.

## SUMÁRIO

1	<b>INTRODUÇÃO</b>	10
2	<b>REVISÃO DA LITERATURA</b>	12
2.1	A vassoura-de-bruxa	12
2.2	O fungo <i>Moniliophthora perniciosa</i>	15
2.3	Parede celular	17
2.4	UDP-N-acetilglicosamina pirofosforilase	19
2.5	Triagem Virtual	23
2.6	Métodos quimiométricos	31
3	<b>MATERIAIS E MÉTODOS</b>	36
3.1	Obtenção do alvo	36
3.2	Avaliação das funções de pontuação	36
3.3	Triagem virtual em banco de dados	39
3.4	Análise das interações	40
4	<b>RESULTADOS E DISCUSSÕES</b>	41
4.1	Moléculas bioativas selecionadas	41
4.2	Avaliação das funções de pontuação	42
4.3	Etapa de Triagem virtual	47
4.4	Recategorização de variáveis	51
4.5	Análise das interações	52
5	<b>CONCLUSÃO</b>	56
	<b>REFERÊNCIAS</b>	58

## 1 INTRODUÇÃO

Pragas são responsáveis por elevadas perdas na produção de cacau no Brasil e no mundo, dentre elas a vassoura-de-bruxa (VB) é uma das mais importantes e destrutivas para o cacauzeiro, chegando a causar perdas de até 95% da produção (OLIVEIRA; LUZ, 2005). A cacauicultura é a principal cultura agrícola do estado da Bahia que sofreu com a VB, pois, à alta densidade cacauzeira, aliada à falta de um período seco na região costeira da Mata Atlântica da Bahia, proporcionaram melhores condições para o estabelecimento e propagação desta praga (PURDY, 2005), causada pelo fungo *Moniliophthora perniciosa* (AIME & PHILLIPS-MORA, 2005).

Grandes impactos na cultura cacauzeira da Bahia foram provocados pela vassoura-de-bruxa ocasionando queda de produção, aumento dos custos e gastos diretos em função do uso de medidas de controle, afetando diretamente os produtores, aumentando o preço do produto, e atingindo de forma indireta o consumidor. Além de ter provocado problemas socioeconômico resultantes da menor produção de cacau, outras mudanças ocorreram na região produtora da Bahia, como: uso da terra, venda de propriedades, nível de emprego e danos ao meio ambiente (MEDEIROS et al., 2010).

Com objetivo de estudar novas formas de controle da praga, pesquisadores de várias instituições iniciaram o projeto genoma do *M. perniciosa*. Este projeto identifica genes muito importantes relacionados ao metabolismo do fungo, e/ou ao seu mecanismo de infecção. Muitos desses genes codificam enzimas e receptores essenciais para *M. perniciosa* e são considerados alvos moleculares para o desenvolvimento de fungicidas mais potentes e seletivos contra esse fungo. Adicionalmente, o entendimento do funcionamento do metabolismo do fitopatógeno a partir dos dados moleculares poderá levar ao desenvolvimento de diversas estratégias de controle da praga (PEREIRA, 2005).

Na aplicação de estratégias de planejamento de moléculas bioativas, os estudos dos processos evolutivos de reconhecimento molecular em sistemas biológicos assumem importância, pois constituem as bases fundamentais para o entendimento de propriedades como potência, afinidade e seletividade (LIMA, 2007). A seleção do alvo biológico determina a priorização de estratégias de planejamento com base em seu papel na fisiopatologia que se pretende tratar (GUIDO, 2010).

Dentre os alvos moleculares do fungo *M. pernicioso*, as enzimas associadas ao metabolismo dos principais carboidratos da parede celular (quitina e glicanos) são atrativos alvos moleculares (PIROVANI et al., 2005). A interferência nestas enzimas promove a instabilidade da parede celular do fungo, levando à destruição das hifas. Como resultado, ocorre à inibição do crescimento do patógeno (GEORGOPAPADAKOU; TKACZ, 1995). Na busca por um controle efetivo da vassoura-de-bruxa, escolheu-se a rota metabólica que leva a síntese da quitina. O alvo escolhido nessa via metabólica foi a enzima UDP-N-acetilglicosamina pirofosforilase (UNAcP) que é a responsável pela realização da reação que forma UDP-N-acetilglicosamina (UNAcGln). Desta forma, a inibição da formação de UNAcGln afetará a formação de quitina e assim a síntese da parede celular, um componente crucial para o desenvolvimento do fungo (OLSEN, RODERICK, 2001; MARUYAMA et al, 2007).

As moléculas bioativas podem ser identificadas a partir de triagens virtuais em bancos de estruturas químicas originárias de produtos naturais ou até mesmo formados por compostos sintéticos ou coleções combinatórias (GUIDO, 2010). Os produtos naturais em particular os metabólitos secundários, são fontes essenciais para a descoberta de compostos bioativos e desenvolvimento de novos compostos promissores (ROSÉN, 2009).

Com o auxílio de métodos *in silico*, é possível explorar o imenso espaço químico delineando o trabalho de identificação, seleção e otimização de moléculas capazes de interagir com alta afinidade e seletividade com o alvo molecular selecionado, o qual representa o espaço biológico. Para tanto, diversas estratégias para a investigação desse espaço químico-biológico podem ser empregadas, incluindo a organização dessas bases de dados identificadas, a aplicação de filtros moleculares no conjunto selecionado e o emprego propriamente dito de triagens biológicas (GUIDO, 2010).

Desta forma, este trabalho teve como objetivo geral, realizar ensaios *in silico*, a fim de identificar potenciais inibidores da UNAcP do *M. pernicioso* e como objetivos específicos: realizar triagem virtual utilizando a função de pontuação mais adequada ao sistema em estudo; utilizar banco de estruturas oriundas de produtos naturais; selecionar as moléculas conforme suas características quimiométricas.

## 2 REVISÃO DE LITERATURA

### 2.1 A vassoura-de-bruxa

A VB é uma das mais importantes e destrutivas pragas associadas ao cultivo do cacau, chegando a causar perdas de até 95% na produção cacaueteira. A praga foi descoberta pela primeira vez no Suriname no final do século XVIII e embora ocorresse de forma endêmica na região Amazônica, constatada a partir do século XIX, está presente em diversos países da América do Sul e Central ao exemplo de Bolívia, Colômbia, Equador, Guiana, Peru, Suriname, Venezuela, Trinidad e Tobago (OLIVEIRA; LUZ, 2005).

Na região sul da Bahia, considerada a principal região produtora de cacau do Brasil, essa praga foi constatada em 1989, o primeiro registro foi no município de Uruçuca-Ba (PEREIRA et al., 1989) e logo em seguida em Camacã-Ba e pouco depois na cidade de Ilhéus-Ba, uma grande área de plantações de cacau foi infectada estando disseminada, em toda a região cacaueteira do estado da Bahia, já tendo sido detectada, inclusive, no estado do Espírito Santo (OLIVEIRA; LUZ, 2005).

No cenário mundial o Brasil se destacava como um dos principais produtores mundiais de cacau e atingiu o apogeu na década de 80 com uma produção de 400.000 toneladas de amêndoas secas, sendo o primeiro produtor do continente americano e terceiro no mundo, com ênfase na produtividade (750 kg/ha), sendo esta considerada como a maior no contexto mundial. Sempre com predomínio absoluto do estado da Bahia, representando mais de 80% total, seguido do estado do Pará, Rondônia e Espírito Santo. A partir da década de 90 houve o declínio acentuado chegando ao número mínimo em torno de 95 toneladas (PEREIRA, 2005).

A queda na produção de cacau resultou em um drástico processo de empobrecimento na região sul da Bahia, devido à redução de até 100% da produção de cacau em diversas propriedades rurais, aliado aos baixos preços do cacau no mercado internacional, à baixa produtividade das lavouras devido à VB e aos fatores climáticos adversos. Fato que resultou na falência de inúmeros produtores, no desemprego de milhares de trabalhadores rurais, migração, erradicação de lavouras em declínio, depreciação da infraestrutura e desvalorização das propriedades rurais e na exploração descontrolada de espécies arbóreas de alto valor ecológico e econômico (DANTAS NETO, 2000).

As condições ambientais favoráveis propiciaram a disseminação rápida desta praga, provocando um colapso na economia regional (OLIVEIRA; LUZ, 2005). As condições de

umidade relativa do ar, a frequência e a duração das chuvas das regiões afetadas, favorecem a produção dos basidiomas. Este último libera de forma ativa milhares de esporos disseminando-os nas primeiras horas do dia através das correntes de ar e contaminando tecidos meristemáticos. Podem ocorrer deposições de esporos a quilômetros da fonte de inoculação em condições atmosféricas favoráveis (BETTIOL, 2009).

Muito embora estejam registradas diversas estratégias com o intuito de controlar a vassoura-de-bruxa, estas ainda são consideradas incipientes. Nenhuma das formas registradas de controle da praga, sejam técnicas de controle químico ou ainda de controle biológico, são capazes de restaurar a posição que o Brasil ocupava na produção de cacau. Isto motiva novas pesquisas com o intuito de encontrar alternativas de controle que possam efetivamente combater a VB (SANTOS JUNIOR, 2012).

Diversos esforços são necessários na busca de novas metodologias a fim de ampliar a eficácia do controle da VB. Neste contexto, a proteção das plantas com o uso de produtos naturais à base de extratos vegetais, bem como a indução de resistência com produtos de natureza biótica ou abiótica, são estratégias importantes que podem apresentar menor impacto ambiental. Esses produtos podem ser capazes de propiciar melhora na produtividade agrícola, com a redução de perdas e, até mesmo, incrementos no desenvolvimento vegetativo (RESENDE et al., 2006).

O controle da vassoura-de-bruxa, através dos métodos químico e cultural como a poda fitossanitária mostrou-se oneroso e pouco eficiente na proteção das plantas, além de causar grande prejuízo ao ambiente. Métodos isolados de controle genético e biológico não ofereceram resultados satisfatórios (BETTIOL, 2009).

A aplicação de fungicidas é um complemento à poda fitossanitária e visa impedir a formação de basidiomas nas partes vegetais removidas. O uso de fungicidas representa custo adicional expressivo. O controle genético é medida de médio e longo prazo, duradoura, porém, não definitiva, devido à variabilidade do fungo, que acarreta a necessidade permanente do desenvolvimento de novas variedades resistentes. A indução de resistência baseada na ativação de mecanismos de defesa latentes no cacauzeiro, em resposta ao tratamento com agentes bióticos ou abióticos, pode se tornar uma estratégia interessante de controle alternativo, porém ainda carente de estudos em campo. Até o presente momento, a estratégia ideal tem sido a adoção conjunta de todas essas técnicas no manejo integrado da doença (BETTIOL, 2009, p. 251).

O aumento dos custos de fungicidas à base de cobre ao longo dos últimos anos e baixo rendimento (300kg/ha) tornou a prática ainda mais onerosa como forma de intervenção para a grande maioria dos agricultores (MEDEIROS et al., 2010).

O controle biológico utilizando antagonistas e competidores como o fungo *Hypocrea stromatica* Bezerra, Costa & Bastos, *Trichoderma Stromaticum* Samuels & Pardo-schulth, *T. viride* Pers, capazes de suprimir a formação ou destruir os basidiomas de *M. Perniciosa* em vassouras secas ou até reduzir a incidência de frutos infectados, revelaram-se alternativas promissoras utilizadas no controle de fitopatógenos (OLIVEIRA; LUZ, 2005).

Alguns fungicidas sistêmicos, com atividade contra fungos da classe dos basidiomicetos têm criado novas perspectivas para o controle da VB, não só em frutos, mas também em lançamentos foliares e almofadas florais. Alguns destes fungicidas têm mostrado atividade *M. perniciosa*, entretanto, os primeiros grupos que demonstraram alguma eficácia *in vivo* foram derivados triazóis e as benzanilidas. Em uma série de experimentos realizados na região cacauieira da Bahia, o fungicida tebuconazole foi que apresentou os melhores resultados (OLIVEIRA; LUZ, 2005).

A sintomatologia da VB foi descrita por diversos autores. A infecção se dá em tecidos em crescimento, como brotos vegetativos, almofadas florais e frutos variado seus sintomas e envolvendo hipertrofia e outras anormalidades. Nas figuras 1A e 1B podem ser observadas, uma área isenta de sintomas da VB do cacauieiro e uma área afetada pela enfermidade respectivamente (BETTIOL, 2009).

Como a infecção ocorrem em tecidos meristemáticos os principais sintomas são apresentados de maneira diferente para cada tipo de tecido. Nos lançamentos foliares ocorre hipertrofia com brotação intensa das gemas laterais que aparentam uma vassoura (Figura 1C), já as almofadas florais se agrupam em um arranjo anormal, hipertrofiado, de pedicelo alongado e inchado, originando frutos deformados (Figura 1D) onde pode ser observada a formação da vassoura vegetativa. Quando são infectados frutos jovens esses se tornam inchados e deformados, amadurecendo precocemente, as sementes tornam-se impróprias para o consumo (Figura1E) (BETTIOL, 2009). De início, o desenvolvimento da praga é rápido, porém depois de 5 a 6 semanas o tecido infectado começa a secar, podendo cair ou ficar aderente à árvore (PEREIRA, 1990).



**Figura 1** – Área isenta de contaminação (A); área de plantio afetada por VB (B); vassouras em lançamentos foliares (C); Fruto com área afetada (D) e (E).



Fonte: SAUNDERS, 2010.

## 2.2 O fungo *Moniliophthora perniciosa*

O fungo *Moniliophthora perniciosa* é o agente etiológico, causador da VB em (*Theobroma cacao*). Ele foi inicialmente descrito como *Marasmius pernicius* posteriormente, *Crinipellis perniciosa* e finalmente reclassificado, com base em características moleculares, como *Moniliophthora perniciosa* (Stahel) Aime & Phillips-Mora (AIME; PHILLIPS-MORA, 2005 *apud* BETTIOL, 2009).

O patógeno é um parasita hemibiotrófico que apresenta duas fases fisiológicas e morfológicas em seu ciclo no desenvolvimento da doença, a fase biotrófica considerada parasítica, monocariótica, de crescimento intercelular, com ausência de grampos de conexão, encontrada em tecidos vivos e a fase saprofítica, dicariótica, também com crescimento intracelular porem com a presença de grampos de conexão que são encontrados apenas em tecidos mortos (LUZ et al., 1997; MARELLI et al., 2009).

Este fungo representa um dos principais fatores limitantes na produção de cacau na América do Sul e nas Ilhas do Caribe, tendo assumido o papel de mais importante patógeno do cacauzeiro. Além disso, sua disseminação acompanhou a do *T. cacao* na Bacia Amazônica, atualmente ambos são endêmicos nesta região (GRIFFITH et al, 1994).

Originário da bacia Amazônica, este fungo infecta plantas da família *Malvaceae*, *Solanaceae*, *Bignoniaceae*, *Bixaceae* e *Malpighiaceae* (RINCONES et al., 2006). É um patógeno pertencente à classe *Basidiomycetes*, à ordem *Agricales* e à família *Marasmiaceae* é, sobretudo um hemibiotrófico nos quais os basidiósporos são as únicas estruturas do fungo em condições naturais capazes de infectar os tecidos meristemáticos do *Theobroma cacao* e várias outras espécies do gênero *Theobroma* e *Herrania* (todos os membros da família *Sterculiaceae*), (BETTIOL, 2009).

De acordo com Luz e colaboradores (1997), os basidiósporos germinados através da penetração diretamente ou pelos estômatos em regiões meristemáticas (almofadas florais ou frutos jovens) promovem intensa hipertrofia no tecido seguida, após algumas semanas, pela morte celular e necrose geral dos tecidos. Após certo tempo, período de latência, intervalo onde ocorre o crescimento do micélio saprofítico e é formada a reserva energia, surgem os basidiomas ou frutificações do patógeno.

Os basidiomas são produzidos em todos os tecidos afetados depois de mortos e mumificado em fruto (Figura 2A) em vassouras como mostrado na (Figura 2B), o micélio secundário do fungo, não é infectivo, somente os basidiósporos, incolores e com dimensões de 10 a 14mm x 4 a 5mm que são produzidos no interior dos basidiomas, são capazes de induzir a doença. Ainda na figura 2B é possível observar o himenóforo do patógeno com lamelas e lamélulas esparsas visível o estirpe excêntrico do basidiomas (BETTIOL, 2009).

**Figura 2** – Basidiomas de *Moniliophthora perniciosa* em fruto de cacau (A), em vassouras (B).



**Fonte:** (BOWERS; BAILEY; HEBBAR; SANOGO; LUMSDEN, 2001).

### 2.3 Parede celular

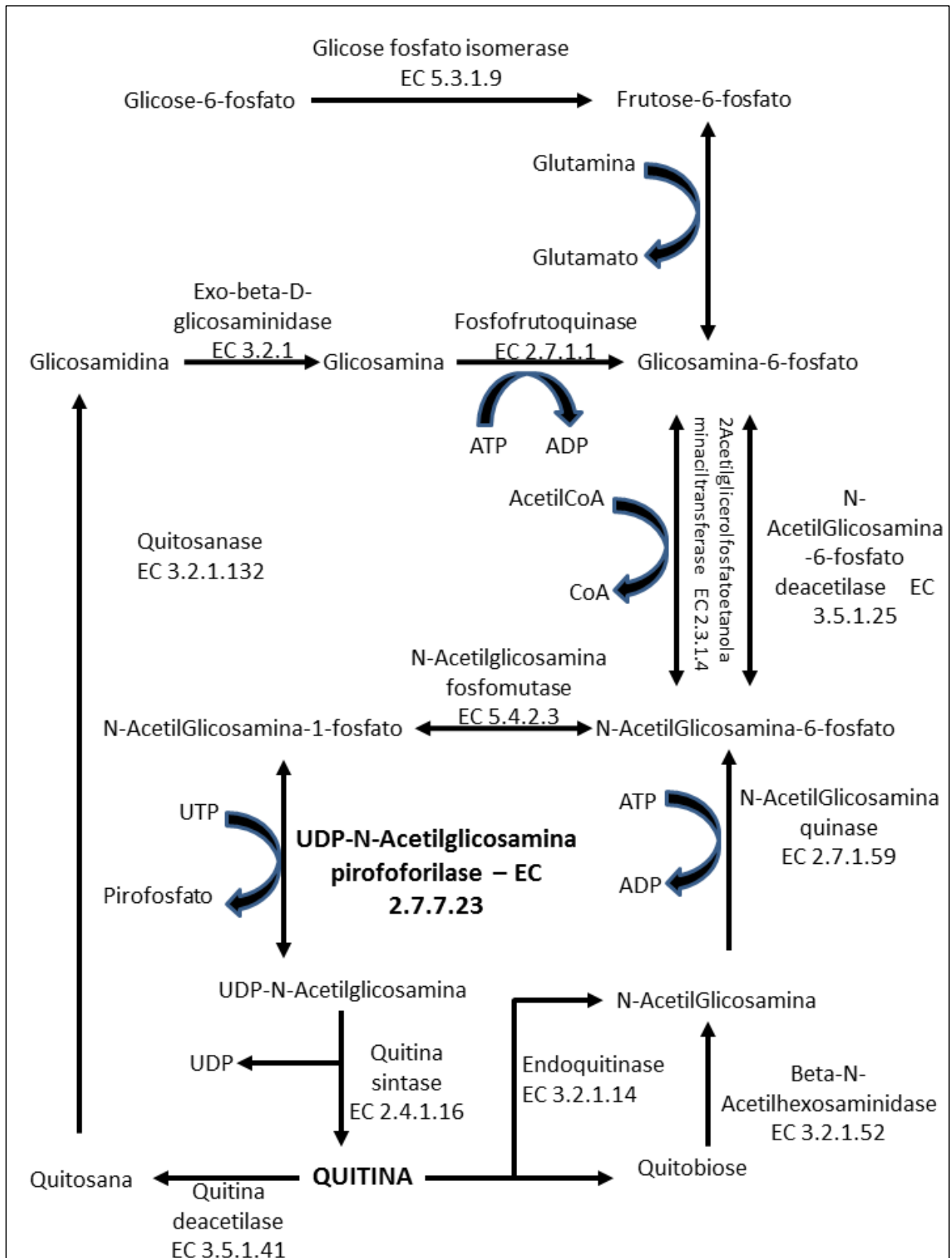
As células fúngicas ou as hifas, para a manutenção da sua integridade, dependem de um envelope externo à membrana plasmática, a parede celular. Esta confere resistência mecânica, permitindo que as células suportem a pressão de turgor, além de lhes fornecer proteção contra condições ambientais extremas e ao mesmo tempo em que interage com o ambiente (RONCERO, 2002).

A parede celular, portanto, protege a célula fúngica contra variações osmóticas, químicas e biológicas, e está relacionada com diversas funções incluindo a morfogênese, expressão antigênica, desempenhando um papel fundamental no crescimento, desenvolvimento e interações dos fungos com o ambiente e com outras células. Sua arquitetura e composição são reguladas de forma coordenada com o crescimento da célula, tendo polissacarídeos (quitina, glicanos e mananos) e glicoproteínas como seus principais componentes. Sua estrutura é altamente dinâmica e está sujeita às constantes mudanças, como, por exemplo, durante a expansão e divisão celular nas leveduras, e durante a germinação de esporos e formação de septos e crescimento apical de hifas em fungos filamentosos (BOWMAN; FREE, 2006).

O estudo e a compreensão complexa dos detalhes estruturais, bem como o entendimento dos mecanismos metabólicos que envolvem os principais componentes da parede celular do *M. perniciosa* resultaram na identificação de pontos potencialmente sensíveis a infecção. No entanto, sabendo que enzimas específicas ao longo desta via

metabólica ainda não são totalmente caracterizadas, o desenvolvimento de estratégias para a desestabilização da parede celular fúngica dependem do conhecimento destes sistemas biológicos, dentre estes se destaca a UNAcP (EC: 2.7.7.23) presente em *M. pernicioso*. Esta enzima pode ser considerado um alvo importante para o desenvolvimento de antifúngicos devido ao fato de que é a única na rota metabólica de produzir UNAcGln (Figura 3) componente essencial para a formação da quitina. Tendo em vista que outros substratos na rota metabólica para a formação da quitina podem apresentar rotas alternativas de formação (KURTZ, REX, 2001; PIROVANI et al., 2005; SANTOS JÚNIOR, 2009).

**Figura 3** – Esquema da via metabólica da síntese da quitina.



Fonte: Santos Junior, 2012; Pirovani et al., 2004 adaptado.

A quitina é uma estrutura rígida amplamente distribuída nos invertebrados, fungos e algas. Nos fungos, a quitina é o maior componente da parede celular atuando de maneira importante na manutenção da integridade celular e ainda sendo um dos mais importantes biopolímeros da natureza (BOWMAN; FREE, 2006).

Nos fungos, a quitina é sintetizada por uma sequência de cinco reações sucessivas. Inicialmente ocorre a conversão da frutose-6-fosfato em glicosamina-6-fosfato pela ação da frutose-6-fosfato aminotransferase (EC 2.6.1.16); em seguida ocorre a acetilação da glicosamina-6-fosfato formando N-acetilglicosamina-6-fosfato catalisada pela 2acetilglicerolfosfatoetanolaminaciltransferase (EC 2.3.1.4); logo após ocorre a interconversão de N-acetilglicosamina-6-fosfato em N-acetilglicosamina-1-fosfato pela ação da N-acetilglicosamina fosfomutase (EC 5.4.2.3); ocorre então a conversão da N-acetilglicosamina-1-fosfato (NAcGln-1-P) em uridina difosfato-N-acetilglicosamina ativada (UNAcGln) UNAcGln catalisada pela ação da UDP-N-acetilglicosamina pirofosforilase (UNAcP) (EC: 2.7.7.23) e por fim segue a conversão irreversível da UDP-N-acetilglicosamina em quitina através da sintase da quitina (EC: 2.4.1.16) (MIO et al, 1998; POMPEO et al, 2001; YAMADA-OKABE et al., 2001; LAGORCE et al., 2002).

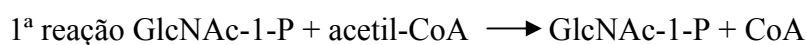
#### **2.4 UDP-N-acetilglicosamina pirofosforilase**

A UDP-N-acetilglicosamina pirofosforilase (UNAcP) (EC 2.7.7.23) é um enzima essencial que pertence à superfamília nucleotídeos-difosfo-açúcar transferase. Ela está presente em procariotos e eucariotos, intimamente relacionado a processos metabólicos que envolvem a formação de componentes extremamente necessários a sobrevivências das células como a formação da quitina, lipopolissacarídeos, peptidoglicanos, glicoproteínas e proteoglicanos (PENEFF, 2001; MARUYAMA et al, 2007).

A isoforma da UNAcP presente em eucariotos catalisa somente a pirofosforilação (MURYAMA et al, 2006). Nos eucariotos, esta enzima é dimérica em mamíferos ao contrário do que ocorre em não mamíferos, onde, para a realização da sua atividade biológica, esta apresenta-se na forma monomérica. Apesar de conservar semelhança na sequência de aminoácidos, a diferença observada numa região de *loop* de cerca de 10 aminoácidos determina esta diferenciação fenotípica (MARUYAMA et al, 2007).

Em procariotos além de catalisar a reação que forma UDP-N-acetilglicosamina, a UNAcP realiza também a acetilação da glicosamina-1-fosfato em NAcGlc-1-P (Figura 4) em duas etapas a primeira a transferência do grupo acetilo a partir da acetil-CoA para a glicosamina-1-fosfato produzindo NAcGlc-1-P. Na segunda reação, o monofosfato de uridina é transferido do UTP para NAcGlc-1-P para produzir UNAcGln e pirofosfato (PPi) (ANDERSON; RAETZ, 1987; MARUYAMA et al, 2007; MOCHALKIN et al, 2007).

**Figura 4** – Etapas da reação catalisada pela UNAcP em procariotos

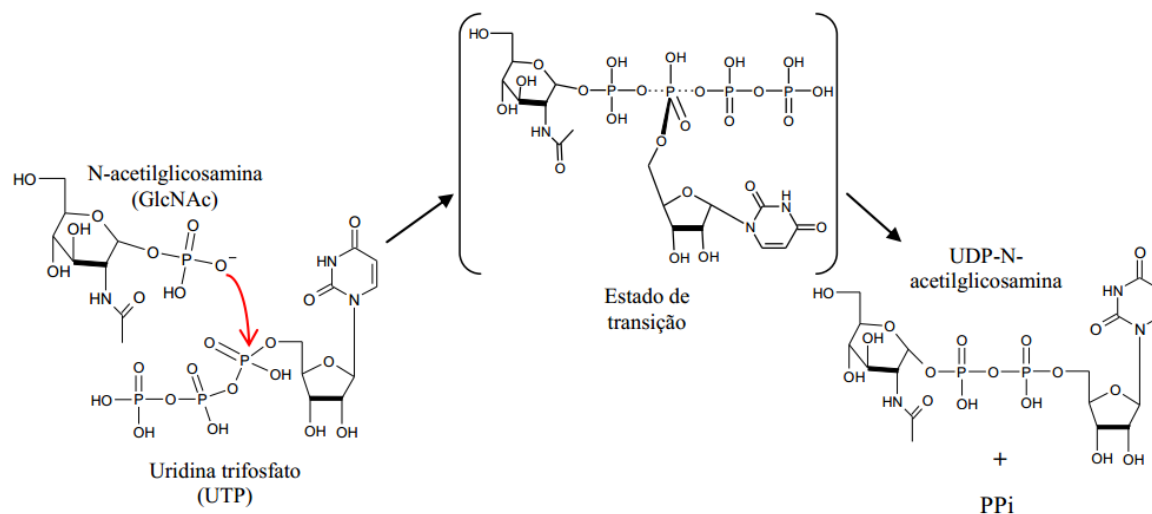


**Fonte:** MOCHALKIN et al., 2007 adaptado.

A reação inversa é também catalisada pela própria pirofosforilase, essa reação ocorre por mecanismo de substituição nucleofílica bimolecular (Figura 4), sendo que o oxigênio não esterificado do grupamento fosfato da N-acetilglicosamina-1-fosfato ataca o fosfato da posição  $\alpha$  da molécula de UTP (SHEU, 1978; KOSTREWA, 2001 *apud*. MARUYAMA et al, 2007). Os fosfatos das posições  $\beta$  e  $\gamma$  são então liberados na forma de pirofosfato (PPi). A presença de cátions divalentes, como íons  $\text{Mg}^{+2}$ , que atuam como cofator aumentando a atividade enzimática devido á sua ação estabilizadora das cargas negativas dos grupos fosfatos (OLSEN; RODERICK, 2001; MARUYAMA et al, 2007).



**Figura 5** – Reação de formação da UDP-GlcNAc Mecanismo  $S_N2$



**Fonte:** (TELES, 2011; MARUYAMA et al, 2007) adaptado.

Alguns estudos apontam interações importantes da enzima com o substrato e outros relatam o mecanismo de catálise (Teles, 2011; Maruyama et al, 2007), porém existem poucos estudos envolvendo ensaios cinéticos em escala de inibição utilizando a UNAcP. Um estudo realizado por Yamamoto e colaboradores (1980), observou que a UNAcP é inibida pela uridina, sendo esta última o núcleo da porção lateral da UNAcGln. Neste trabalho a inibição foi demonstrada em ambas as direções da reação. As análises cinéticas revelaram que a uridina demonstra um tipo de inibição não competitiva frente à UNAcGln e um tipo de inibição competitiva com PPi.

De acordo com Yamamoto e colaboradores (1980), a uridina não se liga de modo irreversível a enzima e vale ressaltar que alguns derivados da uridina, tal como 5-hidroxiuridina; 5, 6-dihidroxiuridina, pseudouridina, também a 5-D-ribosiluracil apresentaram inibição, porém, deoxiuridina apresentou pouca capacidade de inibição.

A interação do nucleotídeo no sítio ortostérico da pirofosforilase é similar em todas as isoformas da enzima. O complexo açúcar-nucleotídeo acopla-se na região central da enzima, estabelecendo contato com a primeira metade desta região (resíduos 68-260) e em particular com o *loop* composto pelos resíduos Asp221-Leu226. O açúcar é estabilizado principalmente por ligações de hidrogênio formadas, com os resíduos da segunda metade do *loop* central (resíduos 261- 417) (PENEFF et al, 2001).



A disposição dos átomos na acetilglicosamina tem importância estabelecida para o reconhecimento no sítio ortostérico da enzima. O grupamento hidroxila, ligado ao C4, liga-se em conformação equatorial por meio de ligações de hidrogênio a Gli290 e Asn327. A porção N-acetil estabelece numerosas ligações de hidrogênio com os aminoácidos Glu303, His331 e Asn223, e uma interação hidrofóbica com Phe381 e Phe383. Estes contatos sugerem uma especificidade da enzima por hexosaminas acetiladas. Ao contrário do açúcar e do nucleotídeo, o grupamento fosfato é bem menos estabilizado (PENEFF et al, 2001).

Os estudos iniciais de purificação da pirofosforilase ocorreram primeiramente no final da década de 50 e foram publicados no *Journal of Biological Chemistry*, tendo obtida a proteína purificada a partir de fígado de bezerros e cepas de *Staphylococcus aureus* e a sequência gênica que codifica esta enzima foi elucidada em 1993 a partir de genes de *Escherichia coli* (DE LUCA et al. 1996).

Os estudos realizados por Santos Junior (2012) apontam para certas particularidades que devem ser observadas no processo de purificação e caracterização da UNAcP. Fatores como estabilidade, temperatura e pH ótimo sofrem variações dependendo do tipo de célula ou tecido no qual a enzima será extraída. Portanto, as diversas variáveis do alvo no patógeno que se deseja atingir devem ser levadas em consideração em ensaios de inibição da UNAcP.

Pelo menos 50 estruturas cristalográficas da UNAcP foram determinadas e encontram-se disponíveis no *Protein Data Bank* – PDB (BERMAN, 1999), entretanto esta enzima tem sido estudada especialmente em sistemas eucarióticos (BULIK et al., 2003).

Os estudos de modelagem comparativa de Santos Junior e colaboradores (2011) a partir de estruturas cristalográficas determinaram um modelo tridimensional da UNAcP do *M. perniciosus*. O molde utilizado para a elucidação estrutural da enzima foi descrita por Peneff et al, (2001). E depositada no PDB com código 1JV1. O modelo definitivo apresentou Desvio Médio Quadrado (RMSD) de 1,2 Å em relação ao molde, sendo possível concluir que a UNAcP do fungo *M. perniciosus* não apresenta um bolsão hidrofóbico muito comum em UNAcP de outros organismos (SANTOS JUNIOR, 2011).

As UNAcP de diferentes organismos apresentam um domínio central na forma alfa/beta/alfa. Este pequeno domínio carboxi terminal extra é formado por 68 resíduos de aminoácidos. Em contraste, a região amino terminal é formada por segmentos semiconservados (PENEFF et al., 2001). No entanto, mesmo semiconservada, esta região apresenta certo grau de identidade quando comparada com enzimas de bactérias e de outros

organismos superiores. A sequência semiconservada inclui Leu-X<sub>2</sub>-Gly-X-Gly-Thr-X-Met-X<sub>4</sub>-Pro-Lys. Esta região é a responsável pela reação de transferência do UTP para a acetilglicosamina (Figura 4 e 5), sendo que esta reação é estimulada por vários cátions divalentes, incluindo Mg<sup>+2</sup>, Co<sup>+2</sup> e Mn<sup>+2</sup> (OLSEN; RODERICK, 2001).

De acordo com Pennef e colaboradores (2001) as interações hidrofóbicas que ocorrem no sítio ativo da pirofosforilase derivam principalmente de interações hidrofóbicas formadas por anéis aromáticos dos resíduos de fenilalanina localizado nas posições 381 e 383. As interações hidrofóbicas são importantes para o reconhecimento molecular, é reconhecido que algumas enzimas apresentam um bolsão hidrofóbico formado por duas fenilalaninas. No entanto, no modelo contruído por Santos Junior e colaboradores (2011) nas posições 381 e 383 são encontrados uma glicina e leucina, respectivamente. Vale resaltar que esses aminoácidos têm cadeias laterais apolares fato que mantém a característica das contribuições em interações hidrofóbicas.

## 2.5 Triagem Virtual

A busca crescente pela descoberta e desenvolvimento de novos fármacos promoveu o crescimento contante de investimentos em novas tecnologias avançadas em quimio e bioinformática para o planejamento racional de farmacos (AMARAL; MONTANARI, 2002). Essa busca, no entanto, geralmente envolvendo triagens biológicas, exigem investimentos expressivos de capital, processos de alta complexidade e longos períodos de estudo, tornando-se um grande desafio para as pesquisas e para o planejamento racional de fármacos. (GUIDO, OLIVA; ANDRICOPULO, 2008).

Os programas computacionais e as abordagens metodológicas baseadas no receptor ou no ligante que buscam otimizar a seleção de novas moléculas para ensaios *in vitro* aumentam a cada dia, visto que cada vez mais evoluem apresentando ótimos resultados em várias fases do processo de planejamento racional (KALYAANAMOORTHY; CHEN, 2011). Nesse contexto as estratégias de triagens biológicas *in silico*, denominada triagem virtual ou ensaio virtual tornaram-se parte integrante no planejamento racional, apresentando novas perspectivas de análise e gerenciamento eficaz da informação, além de reduzir consideravelmente os custos operacionais (GUIDO, OLIVA, ANDRICOPULO, 2008; MUNOS, 2009). A aplicação dessa técnica auxilia desde a fase de seleção de moléculas

bioativas até a otimização das propriedades físico-químicas de compostos promissores (KITCHEN et al., 2004).

Para o planejamento de novos compostos com atividade biológica por meio de triagem virtual é precedido da organização de base de dados de bibliotecas virtuais enriquecidas e o emprego de filtros de seleção moleculares (THOMAS, 2003). Para tanto, existem diversas bibliotecas virtuais de compostos e esses podem possuir parâmetros estruturais e/ou características físicoquímicas que permitam a interação com um provável receptor com o objetivo de estimar a afinidade de ligação de moléculas no sítio ativo do alvo terapêutico (IRWIN; SHOICHET, 2005). A aplicação de filtros moleculares pode proporcionar uma melhor seleção para identificação dos compostos dentre as centenas que compõem as bibliotecas virtuais disponibilizadas no começo do projeto de planejamento racional de fármacos (HAJDUK; GREER, 2007).

As bases de dados podem conter milhares ou milhões de compostos comercialmente disponíveis ou acessíveis por síntese orgânica, podem também ser oriundos de produtos naturais, ou coleções combinatórias que poderão ser obtidos para direcionar ensaios biológicos futuros (LYNE, 2002; RODRIGUES et al., 2012). Uma base de dados que merece destaque é o ZINC (<http://zinc.docking.org>), pois, consiste em um banco de dados de acesso gratuito pela rede mundial de computadores. Sendo desenvolvida especialmente para aplicação de metodologias de triagem virtual sua coleção atualmente conta com cerca de 35 milhões de compostos que podem ser disponibilizados para aquisição, organizados em vários formatos e subdivididos em diversas bases de dados menores de acordo com alguns critérios incluindo fornecedores (IRWIN et al., 2012).

Ao estabelecer o direcionamento de pesquisas envolvendo métodos de triagem virtual, a classificação dos compostos deve ser analisada criteriosamente, pois, esta é realizada de acordo com funções de pontuação. Além disso, deve ser levada em consideração: a estimativa das interações de ligação, diferenciando os acoplamentos corretos dos incorretos; os compostos químicos priorizados devem estar disponíveis para utilização em ensaios biológicos posteriores e os métodos experimentais devem ser padronizados para que esses compostos possam ser avaliados biologicamente de forma reprodutível (JORGENSEN, 2004; TERAMOTO; FUKUNISSEHI, 2008).

Diante dessas observações se faz necessário obter o maior número de informações possíveis sobre o sistema, a fim de determinar quais métodos poderão ser utilizados na triagem virtual, buscando, por exemplo, informações sobre o alvo biológico estudado,

observando, se este já apresenta estrutura tridimensional elucidada ou informações que permitam a obtenção da mesma. Sendo assim as triagens podem ser baseadas na estrutura do receptor (TERAMOTO, FUKUNISSEHI, 2008; DOUGUET, 2008; SONG; LIM; TONG, 2009).

A execução da TVBR requer obrigatoriamente o conhecimento preliminar da estrutura tridimensional do alvo biológico. Para que esta seja bem sucedida, essas estruturas podem ser obtidas a partir de estudos experimentais de Ressonância Magnética Nuclear (RMN), cristalografias de raios X ou modelagem comparativa (KITCHEN et al., 2004). Em sua maioria as estruturas oriundas de cristalografia com alta resolução ( $<1,5 \text{ \AA}$ ) são as preferidas em ensaios que envolvem acoplamento com a predição de afinidade (OSHIRO, et al., 2004). Porém, em função do aumento do número de alvos de interesses para pesquisas associado à dificuldade em se determinar estruturas tridimensionais pelas técnicas convencionais de RMN e cristalografia de raios X, a modelagem comparativa tem sido empregada com sucesso para estudos de triagem virtual (RHODES, 1993; DAVIS, TEAGUE, KLEYWEGT, 2004).

Os estudos de Tang e colaboradores (2012), que compararam o desempenho de alguns sistemas em triagens baseadas no alvo biológico utilizando modelos teóricos frente a estruturas cristalográficas, demonstram que os modelos podem ter resultados comparáveis apesar de serem notadas diferenças entre os sistemas. Em alguns casos os modelos apresentaram melhor desempenho no que se refere ao acoplamento e ao enriquecimento frente aos agonista e antagonistas conhecidos. Eles sugerem ainda a aplicação de modelos em determinados sistemas em detrimento das estruturas cristalográficas, porém em sua conclusão fica claro que esses estudos não têm a intenção de minimizar a importância crítica de estruturas experimentais para compreensão das relações estrutura atividade de proteínas, bem como o papel que as estruturas cristalográficas têm como referência crítica na avaliação da precisão das interações previstas entre ligante-proteína.

As triagens virtuais baseadas na estrutura do receptor e dos ligantes são responsáveis por muitos avanços na seleção de candidatos a fármacos, lembrando que o emprego de uma estratégia não descarta, necessariamente, o uso da outra. (HAJDUK; GREER, 2007; TERAMOTO; FUKUNISSEHI, 2008). Porém, considerando-se a quantidade de alvos biológicos conhecidos e disponíveis em bancos de dados de moléculas como o *Protein Data Bank* (PDB), a triagem virtual baseada na estrutura do receptor destaca-se na identificação de moléculas biologicamente ativas (FERREIRA; OLIVA; ANDRICOPULO, 2011; BRITO, 2012).

A triagem virtual baseada na estrutura tridimensional do receptor (TVBR) lança mão de informações importantes relacionadas ao sítio de ação desses alvos como estratégia para simular as interações com ligantes oriundos de bases de dados. Portanto, estudos de acoplamento molecular podem ser empregados para avaliação da capacidade de interação de moléculas com o sítio ortostático (BROOIJMANS; KUNTZ, 2003; SHOICHET, 2004).

O acoplamento molecular envolve a predição de uma conformação e orientação de um ligante dentro de um sítio ortostérico (DIAS, AZEVEDO JR., 2008), portanto, são aplicados nas abordagens baseadas no alvo, os resultados de cálculos de acoplamento molecular permitem investigar diferentes modos de interação para inibidores, substratos naturais ou de ligantes distintos oriundos, por exemplo, de bibliotecas virtuais, esse fato contribui na determinação de compostos com maior probabilidade de interação com o receptor (TANG; MARSHALL, 2011).

No acoplamento molecular, a geração dos modelos de interação pode ser dividida em duas fases de crescente complexidade. Primeiro é estabelecida a aplicação de algoritmos para posicionar os ligantes no sítio ortostérico do receptor, nessa etapa o grande desafio está paltado na liberdade conformacional que devem ser amostrados com precisão suficiente para identificar a melhor conformação de acoplamento. Em seguida, são aplicadas funções de pontuação, para avaliar as interações entre as moléculas de interesse e os alvos em potencial, no intuito de prever atividade biológica (KITCHEN et al., 2004; COUPEZ; LEWIS, 2006; DIAS; DE AZEVEDO JR., 2008). Entretanto a identificação de características moleculares responsáveis pelo reconhecimento biológico ou as modificações estruturais que determinam aumento de potencia são frequentemente difíceis de compreender e simular (DIAS, AZEVEDO JR., 2008).

Para o acoplamento molecular os graus de liberdade devem ser considerados para a seleção da conformação em um processo rápido, permitindo avaliar milhares de moléculas em tempo real. Para utilização de base de dados, diversos algoritmos aplicados em métodos diferentes, estão à disposição levando em consideração a flexibilidade do ligante e em alguns casos, alguma flexibilidade do alvo (MONTANARI, 2011). Os métodos empregados para tratar da flexibilidade do ligante podem ser divididos em três categorias: métodos sistemáticos (construção incremental, busca conformacional, banco de dados); métodos estocásticos ou randômicos (Monte Carlo, algoritmo genético, método tabu); métodos de simulação (dinâmica molecular e minimização energética). A Flexibilidade da proteína geralmente é tratada com abordagens com simulações de Dinâmica Molecular e Monte Carlo, Bibliotecas

Rotaméricas e médias das grades de conjuntos de proteínas (BROOIJMANS, KUNTZ, 2003; DIAS; DE AZEVEDO JR., 2008).

Um aspecto estratégico que deve ser levado em consideração na aplicação de métodos de triagem virtual é a aplicação de filtros capazes de diminuir o número de moléculas. Deve ser levada em consideração também a avaliação criteriosa das propriedades reais que as moléculas pertencentes às bases de dados apresentam, observando inclusive a origem dessas moléculas (MONTANARI, 2011). Não são somente filtros com descritores unidimensionais (1D) que geralmente são escalares representados em números simples, e descritores bidimensionais (2D) com base em matrizes 2D, podem ser utilizados apesar de serem os mais amplamente empregados. Filtros tridimensionais (3D) tensores 3D desdobrado em matrizes 2D, também podem ser usados para reduzir o número de compostos em um dado conjunto. A massa molar a conectividade e o volume molecular são propriedades com características unidimensional, bidimensional e tridimensional respectivamente (FERREIRA, 2002).

Métodos de análise baseadas nos descritores, índices, variáveis independentes podem ser aplicados para seleção de moléculas com esqueletos estruturais determinados. Esses filtros podem ser utilizados também quando na metodologia lança-se mão de subestruturas ou grupos funcionais com interações conhecidas, podendo auxiliar na identificação ou eliminação, utilizando ferramentas da quimioinformática, e cálculos de quimiometria (STAHURA, BAJORATH, 2005; STEINDL, LANGER, 2005).

A triagem virtual é um processo de enriquecimento, no qual a predição exata de afinidade não é necessária. É importante, no entanto, que compostos ativos sejam selecionados para um grupo de moléculas que será testado em laboratório, não importa se o mais ativo está no topo do *ranking* ou é o primeiro da lista. Tão importante quanto selecionar ativos é excluir moléculas inativas (DE AZEVEDO JR., 2008; MYSINGER, et. al., 2012).

Após o acoplamento, deve ser aplicada uma função de pontuação capaz de analisar a qualidade do mesmo, onde serão consideradas a forma e complementaridade eletrostáticas do complexo ligante-receptor. O resultado gerado será utilizado para a classificação em relação aos demais compostos no universo de interesse (MONTANARI, 2011). Algumas vezes funções de pontuação mais complexas são aplicadas em compostos pré-selecionados, funções com menor grau para aproximações que levam em consideração interações eletrostáticas e de van der Waals. Uma vez que a interação ligante-receptor é regida por efeitos entálpicos e

entrópicos, uma determinada interação pode ser mais bem prevista por uma ou outra função (GOHLKE, KLEBE, 2001).

Atualmente estão disponíveis diversos programas que apresentam diferentes abordagens metodológicas e podem ser aplicados a fim de atender os parâmetros determinados na pesquisa, como exemplo temos o DOCK (EWING et al., 2001) e o AutoDock Vina (TROTT; OLSON, 2010) ambos os programas têm código aberto.

Uma função de pontuação confiável deve ser capaz de classificar melhor ligantes conhecidos de acordo com a sua pontuação durante o processo de triagem do banco de dados. Dado o custo elevado de técnicas de rastreamento e técnicas experimentais (HUANG et al, 2010).

As funções podem ser baseadas no campo de força aplicada com base na mecânica molecular, podem ser empíricas, quando ajustadas para reproduzir as energias de ligação e conformação, as baseadas no conhecimento, quando derivadas de análises estatísticas das informações da estrutura. Estas não levam em consideração dados de energia de afinidade obtidos experimentalmente (GOHLKE, & KLEBE, 2001; WARREN et al, 2006).

As funções de pontuação baseadas no campo de força são desenvolvidas com base em interações físicas atômicas, incluindo: interações de van der Waals, eletrostáticas e de forças de torção. Os parâmetros são geralmente obtidos, a partir de dados experimentais e cálculos de mecânica quântica, de acordo com os princípios da física. Apesar de seu significado físico, um grande desafio para as funções de pontuação de campo de força é a forma de tratar o solvente em contato com o ligante. Uma função de pontuação bastante utilizada é o Grid Score presente no DOCK, onde os parâmetros de energia são baseados no campo de força presente no Amber (*Assisted Model Building with Energy Refinement*) (EWING et al., 2001; HUANG et al, 2010; TROTT; OLSON, 2010).

Apesar da eficiência computacional atribuída a função de pontuação Grid Score presente no DOCK, o fator dielétrico dependente da distância não pode explicar seu efeito na ausência do solvente, um importante efeito do solvente a ser observado seria a presença ou a ausência de grupos polares e não polares em solventes aquosos ou não aquosos (EWING et al., 2001; HUANG et al, 2010).

Uma maneira rigorosa para explicar o efeito do solvente é tratar moléculas de água de forma explícita. Técnicas como a perturbação da energia livre (*free energy perturbation-FEP*) e integração termodinâmica (*thermodynamic integration-TI*) usam uma representação explícita de moléculas de água. No entanto, esses métodos, são computacionalmente

dispendiosos para serem usados em triagens virtuais. Além disso, enquanto simulações com as moléculas de água explícitas são teoricamente mais rigorosas, a precisão dos métodos é limitado pela amostragem e pela precisão do campo de força. Este por sua vez, depende da formula matemática e da parametrização. Com o intuito de reduzir o custo computacional, alguns modelos de campo de força acelerados foram desenvolvidos para uso de pontuação em acoplamento molecular tratando água como um meio dielétrico constante (HUANG et al, 2010).

Exemplos típicos de tais modelos de solventes implícitos incluem o modelo de área da superfície de Poisson-Boltzmann (*PB/SA*) modelo e área da superfície do modelo generalizado de Born (*GB/SA*). Ambos são geralmente utilizados em reacoplamentos após a aplicação de uma determinada função de pontuação. Embora seja possível encontrar coeficientes de aproximação apropriados para uma determinada proteína ou família de proteínas, é difícil obter um conjunto universal para os diversos complexos proteína-ligante (ROCCHIA et al, 2002; QIU et al, 1997).

Outra função de pontuação amplamente empregada é a função de pontuação empírica, que estima a afinidade de ligação de um complexo com base em um conjunto de termos energéticos ponderados individuais, como energia van der Waals, energia eletrostática, ligações de hidrogênio, dessolvatação, entropia de ligação, hidrofobicidade, entre outros. Ou seja, emprega a suposição de que a energia livre de interação pode ser decomposta em uma série de interações (HUANG et al, 2010; WANG; LAI; WANG, 2002). Estes coeficientes também são calculados através da determinação de dados relacionados à afinidade de ligação da uma formação de um conjunto de complexos ligante-proteína com estruturas tridimensionais conhecidas, onde se observa a limitação do método, determinado pela dependência das contribuições individuais apresentadas por conjuntos de dados específicos. Em comparação com a pontuação baseada no campo de força, as funções de pontuação empíricas são muito mais rápidas devido aos seus termos de energia simples (WANG; LAI; WANG, 2002; SOTRIFER; SANSCHAGRIN; MATTER; KLEBE, 2008).

Além das funções baseadas no campo de força e empíricas, outras funções de pontuação podem ser utilizadas como as que são baseadas na estatística potencial. Estas empregam potenciais de energia derivadas das informações estruturais obtidas isoladamente a partir da análise de dados estatísticos de estruturas atômicas sem referência às afinidades de ligação determinados experimentalmente. A grande vantagem de se utilizar essas funções baseadas na estatística potencial está no processo computacional se apresentar de maneira



muito simplificada o que permite a triagem eficiente de grandes bases de dados (GOHLKE; HENDLICH; KLEBE, 2000).

O princípio que norteia as funções de pontuação baseadas no conhecimento é simples, os complexos entre proteína e ligantes são descritos de acordo com o ambiente molecular usando potenciais atômicos. No entanto a função de energia relaciona a frequência de ocorrência de contatos individuais entre a proteína e os tipos de átomos do ligante como uma medida de suas contribuições energéticas para a ligação. Se um contato ocorre com mais frequência do que aleatoriamente, uma interação atrativa é designada (SIPPL, 1990; GOHLKE; KLEBE, 2001).

As limitações das funções de pontuação são conhecidas e em alguns casos várias funções podem ser combinadas para obtenção da pontuação de consenso (*consensus scoring*). Neste caso, as moléculas que apresentam alta afinidade devem ser classificadas por mais de um tipo de função de pontuação e devem ser priorizadas de forma consensual (ALVAREZ; SHOICHET, 2005). Este processo pode conduzir a um decréscimo no número de falsos positivos, ou seja, moléculas classificadas como potenciais ligantes, mas que nos ensaios biológicos *in vitro* são inativas. Essa metodologia é apropriada quando não se tem conhecimento suficiente para definir qual a função mais adequada para o sistema em questão (KUKOL, 2011).

Os ensaios do acoplamento molecular devem passar por métodos de enriquecimento para contornar as limitações que geralmente os métodos teóricos apresentam. Nesse caso a utilização de bancos de dados de compostos falso positivos como o DUD-E (*Database of Useful Decoys: Enhanced*), pode ser utilizado como uma ferramenta para avaliação e otimização do acoplamento molecular. O DUD-E fornece 50 moléculas para serem utilizadas como conjunto de falsos positivos, para cada ligante ativo do banco de dados. Os falsos positivos apresentam propriedades físicas semelhantes as dos ligantes, mas são inativos do ponto de vista biológico e por isso não é um desafio para os acoplamentos, mas, ao mesmo tempo, são topologicamente desiguais para minimizar a probabilidade de ligação real (MYSINGER, et. al., 2012).

A relação entre compostos falsos positivos e os ligantes conhecidos é de todo importante na avaliação das funções de pontuação. O fator enriquecimento (FE) do acoplamento molecular reflete a capacidade dos cálculos de acoplamento encontrar

verdadeiros positivos em uma base de dados selecionada em detrimento de uma seleção aleatória (HUANG; SHOICHET; IRWIN, 2006).

O sucesso do método de acoplamento molecular é avaliado através de dois critérios: A) Capacidade em enriquecer um pequeno número de compostos ativos conhecidos frente a um grande conjunto de falsos positivos, obtidos em banco de dados; B) a fidelidade geométrica da disposição do composto no sítio ortostérico em comparação com das estruturas experimentais (HUANG; SHOICHET; IRWIN, 2006).

## 2.6 Métodos quimiométricos

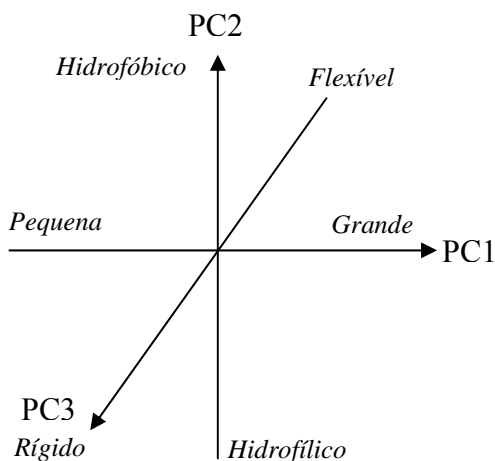
Na busca por novos fármacos, enormes quantidades de dados são produzidos e as informações essenciais, muitas vezes, são difíceis de extrair sem a aplicação de métodos quimiométricos. Uma missão importante para os químicos medicinais é identificar estruturas químicas interessantes, que têm o potencial para se tornar fármacos (ROSÉN, 2009). Apesar de ferramentas quimiométricas inicialmente terem sido usadas principalmente para resolução de problemas químicos, sabe-se que são hoje amplamente utilizados em muitos outros campos da ciência, por exemplo, ciências biológicas (ERIKSSON, et al. 2004) e ciências farmacêutica (GABRIELSSON; LINDBERG & LUNDSTEDT, 2002).

As ferramentas quimiométricas são fundamentadas em conceitos importantes, como o espaço químico, que pode ser descrito como uma região multidimensional, em um sistema de coordenadas com múltiplos eixos que representa um número de propriedades físico-químicas (DOBSON, 2004). Essas ferramentas levam em consideração que as estruturas químicas estão posicionadas neste sistema de coordenadas com base nos diferentes valores de seus respectivos eixos. De maneira inteligente é possível então navegar de forma eficiente através do espaço químico, a fim de, extrair informações significativas a partir da grande quantidade de dados obtidos na busca por novos fármacos ou para selecionar e priorizar quais compostos testar uma determinada atividade fora das infinitas possibilidades de compostos (LINUSSON, et al. 2000).

Dentro dessa perspectiva, é importante citar o conceito de quimiografia que denota "a arte de navegar no espaço químico", ela se assemelha a geografia e, basicamente, significa o mapeamento de objetos usando descritores químicos. A idéia é construir um mapa sobre o espaço químico usando diferentes direções, ou seja, eixos diferentes, no mapa. Algumas

dessas regras que compõem os descritores incluem flexibilidade, tamanho da molécula, capacidade de fazer ligação de hidrogênio e lipofilicidade. Um exemplo pode ser observado na figura 6 (ROSÉN, 2009).

**Figura 6** – Modelo de descritores que permitem a localização das moléculas no espaço químico.



**Fonte:** adaptado de ROSÉN, 2009.

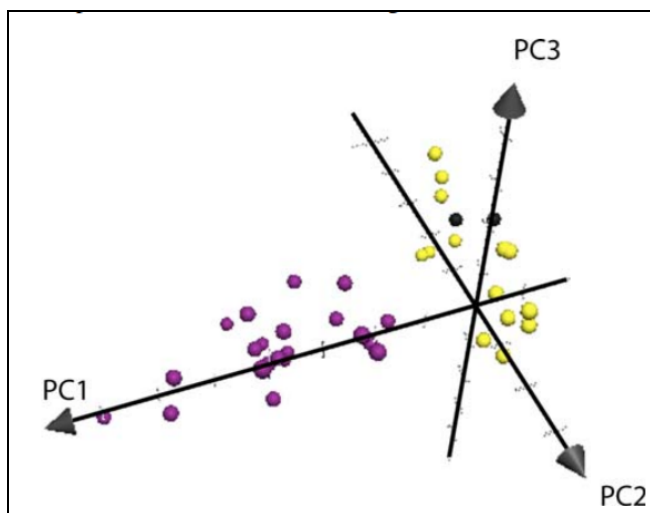
Nesse contexto a quimiometria se apresenta como aplicação de métodos matemáticos e estatísticos aos dados químicos e dentre as metodologias quimiométricas de análises multivariadas incluem a utilização da análise de componentes principais (PCA) (WOLD, 1987 *apud*. ROSÉN, 2009).

A PCA é um método matemático, que tem sido amplamente utilizada na descoberta de fármacos. Tem sido considerada como um ponto de partida para a análise multivariada de dados, pois, fornece rapidamente uma visão geral das informações onde se espera que as informações mais importantes e relevantes se tornem mais visíveis e fáceis de serem interpretadas. A PCA fornece uma visão geral de todas as observações das amostras ou tabela de dados onde os padrões, como agrupamentos, tendências, desvios, valores discrepantes, podem ser encontrados (TEÓFILO, et al. 2009).

A distância relativa entre os objetos no espaço químico é uma estimativa de como eles são semelhantes em relação às variáveis selecionadas. Os objetos semelhantes estão situados próximos uns dos outros ou serão agrupados nas parcelas de pontuação. Portanto, variáveis que contêm informações semelhantes estão situadas próximas umas as outras e se

correlacionam positivamente ou diagonalmente opostas umas as outras, se elas são negativamente correlacionadas (Figura 7) (ROSÉN, 2009).

**Figura 7** – Posicionamento das moléculas no espaço químico: betacianinas em violeta, betaxantinas em amarelo, e os muscaflavinas em preto.



Fonte: ROSÉN, et al, 2008.

As buscas por similaridade têm sua base fundamentada no chamado princípio semelhança (JOHNSON, MAGGIORA, 1990, *apud*. ROSÉN, 2009) afirmando que moléculas semelhantes são susceptíveis de ter propriedades físico-químicas semelhantes e, portanto, pode ter atividade biológica semelhante. Na busca por novos fármacos muitas vezes é de interesse saber como ou quanto uma molécula é semelhante à outra. Pesquisas de similaridade são muito importantes para triagem aplicadas a grandes bibliotecas de compostos (BENDER, GLEN, 2004).

Se o objetivo é identificar novos compostos promissores é um desperdício de recursos à execução de ensaios com moléculas muito semelhantes, no entanto, quando se tem conhecimento sobre compostos com certa atividade, a metodologia é eficiente para testar compostos que são semelhantes a um composto ativo com atividade conhecida. Nesse caso a expectativa é de encontrar estruturas moleculares alternativas, que preservem as propriedades necessárias, por exemplo, para permitir patente ou aperfeiçoar perfis farmacocinéticos (ROSÉN, 2009).

Modelos de representação de estruturas químicas, impressões digitais estruturais e descritores moleculares são frequentemente utilizados nos cálculos de similaridade e nas

buscas por similaridade. Essas representações são comparadas com um índice de similaridade, onde as medidas mais usadas são medidas de distância simples tais como a Distância Euclidiana e coeficientes de associação como os coeficientes de Hamming e Tanimoto (ROSÉN, 2009).

Nesse sentido uma estratégia que pode ser aplicada é o *ChemGPS* uma técnica computacional para investigação global dos compostos baseados em propriedades espaciais. A sigla pode ser interpretada como Sistema Químico de Posicionamento Global, do inglês *Chemical Global Positioning System* (LARSSON, et al., 2005).

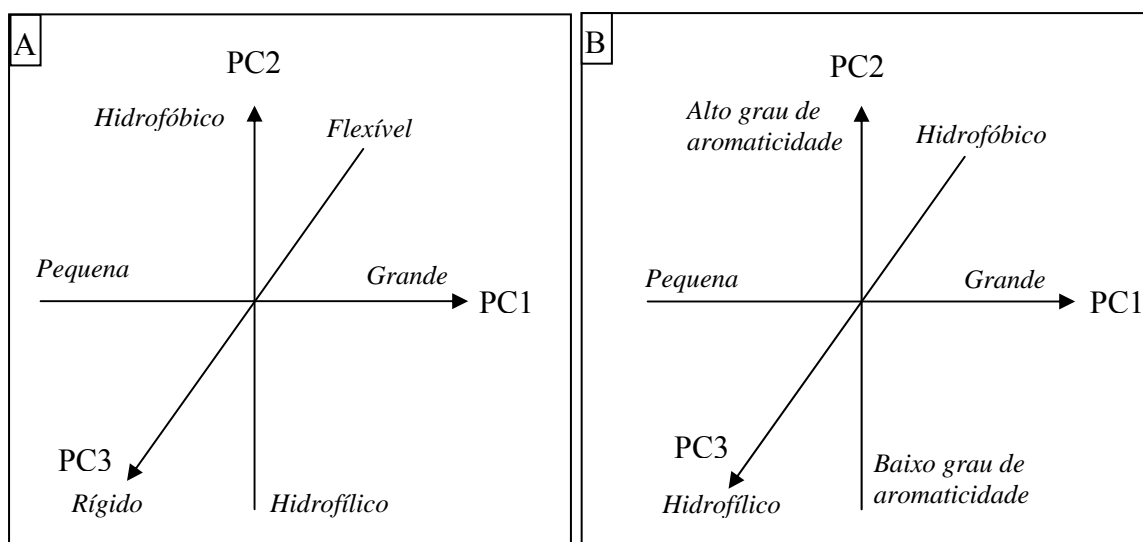
Uma grande porção do espaço químico, onde se localizam compostos químicos considerados de relevância biológica é ocupado por moléculas oriundas de produtos naturais, ou seja, entidades químicas produzidas por organismos vivos. Esses compostos têm sido a fonte de inspiração para estudiosos por milênios. Esse grupo provou ser de longe a fonte mais rica de classes de compostos novos, e uma fonte essencial de novas espécies ativas (HARVEY, 2008).

Compostos oriundos de produtos naturais podem ser considerados como pré- validado pela Natureza em termos de interação biológica, pois, eles apresentam grande diversidade química e foram otimizados para interações ideais com macromoléculas biológicas por meio de seleção evolutiva. Praticamente todos os compostos biossintetizados possuem uma atividade biológica com propósito benéfico para o organismo que a produz, cumprindo assim a exigência de relevância biológica. Com base nessas afirmações, a classe dos compostos naturais torna-se excepcionalmente interessante, como recurso em projetos de descoberta de novos compostos promissores (HARVEY, 2008).

A ferramenta *ChemGPS* foi aprimorada pelo trabalho de Larsson e colaboradores (2007) na tentativa de melhor representar todo o espaço químico incrementando informações biologicamente relevantes que incluem compostos promissores e produtos naturais bioativos. Então foi desenvolvida uma nova ferramenta, o *ChemGPS-NP*, que é uma análise de componentes principais com base em um sistema de posicionamento químico global, mas com um novo conjunto de referência satélite de estruturas principais, e uma nova matriz descritor escolhido e avaliados a partir de uma série de critérios que melhor representam os descritores estruturais e que tem maior prevalência e relevância em termos químicos em moléculas oriundas de produtos naturais (Figura 8) (LARSSON, et al., 2007).

Essa ferramenta está parametrizada para a exploração das regiões do espaço químico, mais propensos a incluir compostos com funções e atividades biologicamente relevantes. O *ChemGPS-NP* pode ser aplicado em diversos tipos de metodologias relacionadas à descoberta de drogas. Em termos gerais, ele pode ajudar na priorização e na seleção de compostos promissores candidatos a fármaco. Com referência ao princípio da similaridade (JOHNSON & MAGGIORA, 1990 *apud*. ROSÉN, 2009) inibidores conhecidos de um determinado alvo poderiam ser mapeados em conjunto com um número de compostos disponíveis a partir dos quais os que se situam mais próximos aos inibidores conhecidos podem ser selecionados para testes adicionais, aumentando assim as possibilidades (ROSÉN, et al., 2010).

**Figura 8** – Modelos das principais propriedades traduzidos em descritores químicos para as três primeiras dimensões: A) *ChemGPS* e B) *ChemGPS-NP*.



**Fonte:** adaptado de ROSÉN, 2009.

Um procedimento de seleção alternativo como um tratamento estatístico, pode ser realizado, se, por exemplo, apenas um pequeno número de compostos deve ser selecionado a partir de um conjunto inicial grande para testes posteriores (ROSÉN, 2009).

O *ChemGPS-NP* apresenta oito componentes principais ou seja, oito dimensões que descrevem propriedades físico-químicas de um conjunto de compostos de referência. Diferente da análise realizada pelo *ChemGPS* (Figura 8A), a interpretação básica das três primeiras dimensões da *ChemGPS-NP* (Figura 8B) são: Tamanho da molécula que aumenta na direção positiva da componente principal 1 (PC1); compostos que são cada vez mais aromáticos no sentido positivo da componente principal 2 (PC2); compostos lipofílicos que

ficam situados no sentido positivo da componente principal 3 (PC3) e, predominantemente, compostos polares estão localizados na direção negativo da PC3 (ROSÉN, 2009).

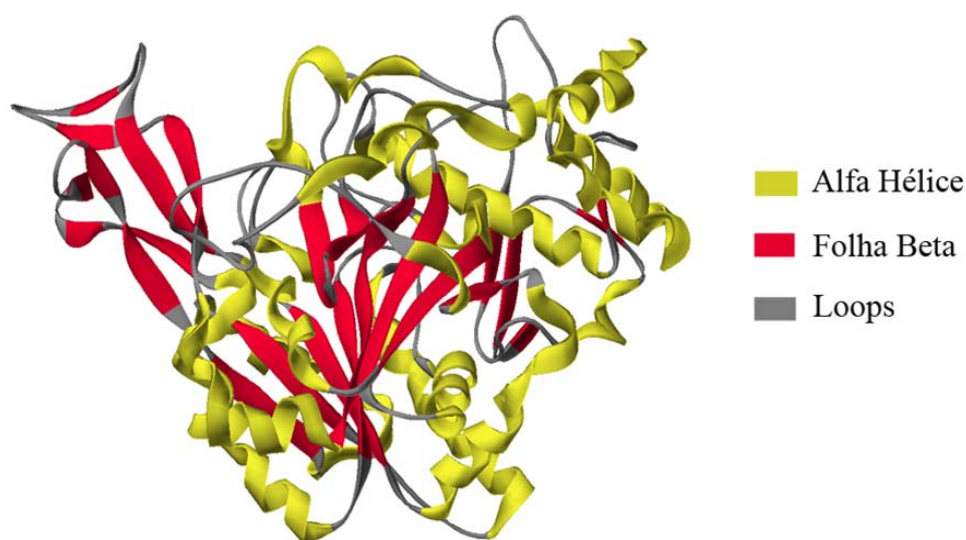
### 3 MATERIAIS E MÉTODOS

Os estudos utilizando métodos computacionais na identificação de inibidores da UNAcP do *M. pernicioso* foram realizados em diferentes etapas de busca e validação. Todos os cálculos e simulações foram realizados no Laboratório de Modelagem Molecular (LMM) da Universidade Estadual de Feira de Santana (UEFS).

#### 3.1 Obtenção do alvo

O alvo utilizado para os estudos de acoplamento molecular foi o modelo construído para a UNAcP do fungo *Moniliophthora pernicioso* obtido por modelagem comparativa nos estudos de Santos Junior e colaboradores (2011). O molde utilizado para a elucidação estrutural foi a estrutura de UNAcP de humanos descrita por PENEFF e colaboradores (2001).

**Figura 9** – Modelo da estrutura tridimensional construído para a UNAcP do fungo *Moniliophthora pernicioso*.



Fonte: SANTOS JUNIOR, M. C.

#### 3.2 Avaliação das funções de pontuação

Para a realização da avaliação das funções de pontuação empregadas nos ensaios virtuais, foram escolhidas moléculas ativas com base em informações de atividade em estudos realizados sobre a UNAcP. Para isso foi realizado uma busca na literatura por moléculas que passaram por algum tipo de ensaio de inibição sobre a UNAcP.



As moléculas selecionadas foram construídas no programa Marvin Sketch 6.1.0, o *Simplified Molecular-Input Line-Entry System* (SMILES) foi gerado neste mesmo programa. Os SMILES foram então submetidos ao banco de dados DUD-E (<http://dude.docking.org>) (MYSINGER, et al., 2012), no intuito de obter os compostos falsos positivos com similaridades físicas, mas, sem aproximações das funções químicas, fato que permite a diferenciação entre os ligantes e os falsos positivos gerados para avaliação da taxa de recuperação das funções de pontuação utilizadas no estudo de acoplamento e para selecionar a função de pontuação com maior grau de precisão para o sistema em questão.

Para os estudos de acoplamento foram empregados métodos diferentes no intuito de encontrar o que conseguisse melhor taxa de recuperação para o sistema proposto, e para tanto foram utilizadas ferramentas que aplicam o método estocástico e método sistemático ambos com função de pontuação baseados no campo de força de mecânica molecular que quantifica a soma das energias de interação receptor-ligante e energia interna do ligante envolvido no processo de reconhecimento molecular (MONTANARI, 2011).

Para o método estocástico o programa utilizado foi o AUTODOCK VINA 1.1.2 que utiliza o algoritmo *Iterated Local Search Global Optimizer*, que aplica uma otimização local e global a partir do método Broyden-Fletcher-Goldfarb-Shanno (BFGS) que é baseado no método quasi-Newton (TROTT, OLSON, 2010). A caixa de acoplamento foi a mesma utilizada nos estudos de Santos Junior et al., (2013) sendo definida então como um cubo com centro geométrico na região central do sítio ortostérico do modelo da UNAcP, possuindo dimensões de 18 x 14 x 16Å, espaçamento de pontos de 1Å e coordenadas X, Y e Z de -58.323, -18.926 e 79.633, respectivamente.

Para o método sistemático foi utilizado o programa DOCK 6.5 e seus programas acessórios (EWING et al., 2001). O DMS (*Display Midas System*) foi utilizado para criar uma superfície molecular do receptor (FERRIN, et al., 1988). O programa SPHGEN (KUNTZ, et al, 1982) foi usado para criar um conjunto de esferas com base na superfície molecular gerada pelo DMS, e o arquivo resultante foi editado para incluir apenas esferas no perímetro de 8,0Å do ligante complexado com o modelo. Esta distância foi suficiente para incluir todo o sítio ortostérico do modelo.

O programa interativo Showbox foi usado para definir a localização e o tamanho da caixa onde será calculado o GRID. A caixa construída automaticamente permite a inclusão do conjunto de esferas com margem extra de 5Å. O programa GRID foi utilizado para gerar a grade de pontuação e a grade de relevo, os seguintes parâmetros foram utilizados: os escores de contato e energia foram calculados utilizando distância atrativa de energia de 6,0Å e

distância repulsiva de van der Waals de 9,0Å, tamanho mínimo de âncora de 5, energia interna de repulsão de 9,0Å e choque de sobreposição de 0,25Å, os parâmetros restantes foram mantidos nas recomendações padrões do programa (SHOICHET; KUNTZ; BODIAN, 1992. MENG; SHOICHET; KUNTZ, 1992).

Após a definição dos parâmetros utilizados na caixa de acoplamento, definida pelos programas acima, o programa DOCK 6.5 foi utilizado para os estudos de acoplamento molecular. Inicialmente, a função de pontuação Grid Score foi utilizada para o ranqueamento das moléculas. Essa função de pontuação tem seu método de ranqueamento baseado no escore de energia calculado a partir de parâmetros da física, tipicamente determinados usando um campo de força baseado na mecânica molecular simplificada (MENG; SHOICHET; KUNTZ 1992).

Adicionalmente foi realizada uma reclassificação utilizando a pontuação Hawkins GB/SA que é uma implementação da Mecânica Molecular aplicada a área de superfície generalizada (MM-GB/SA) que usa o modelo solvatação GB relatado nos trabalhos de Hawkins e colaboradores (1995 e 1996) a partir dos resultados do Grid Score.

Neste estudo foi adotada também a estratégia de classificação número por número com a aplicação da pontuação por consenso, que combina os resultados das funções de pontuação utilizadas. Para a estratégia de consenso é ideal utilizar a partir de três funções de pontuação para ser suficiente e significativa com o intuito de possibilitar o aperfeiçoamento dos resultados (LIU et al., 2012). Para que fosse possível a aplicação do consenso, tendo em vista, as diferentes unidades obtidas a partir das funções aplicadas, foi aplicado o tratamento estatístico utilizando a fórmula da equação:

Equação para aplicação do consenso Onde: *ZScore* é o novo valor de classificação para a molécula, *f<sub>i</sub>* é o valor de escore de determinada função de pontuação,  $\mu$  é o valor da média aritmética dos valores apresentados de cada molécula em cada função e  $\sigma$  o desvio padrão das funções de pontuação aplicada.

$$ZScore = \frac{f_i - \mu}{\sigma}$$

Fonte: LIU, S. et. al., 2012

A qualidade das funções de pontuação foi avaliada isoladamente utilizando duas metodologias diferentes. A primeira foi à realização do cálculo do fator de enriquecimento baseando-se na seguinte equação:

Equação utilizada para calcular fator de enriquecimento: Onde:  $FE_{(sub)}$ : Fator de enriquecimento para um subconjunto; VPr: Verdadeiros positivos recuperados; VPt: Número total de verdadeiros positivos; NT: Número total de moléculas do banco; Nsub: Número de moléculas do subconjunto

$$FE_{(sub)} = \frac{VPr}{VPt} \cdot \frac{NT}{N_{(sub)}}$$

**Fonte:** (HUANG; SHOICHET; IRWIN, 2006) adaptado.

A segunda metodologia permitiu a análise dos resultados através da comparação das respectivas curvas ROC (*Receiver Operating Characteristic*), uma técnica visual para avaliação, organização e seleção de classificadores com o objetivo de avaliar e comparar algoritmos de aprendizado (FAWCETT, 2006). Sendo então o eixo y da Curva ROC (Sensibilidade) que representa a taxa acertos ou verdadeiros positivos recuperados, enquanto que o eixo x (Especificidade), expressa à taxa de falsos positivos. As análises foram realizadas a partir do cálculo da área sob a curva (AUC) que determina a exatidão do método (BAKER; CANNOR; KESSLER, 1998).

### 3.3 Triagem virtual em banco de dados

Para a realização do acoplamento utilizando a função selecionada na etapa anterior, foi utilizada uma biblioteca de produtos naturais (*ZINC Natural Products*) onde está incluída a base de dados com compostos extraídos de plantas do semi-árido baiano. Os compostos foram obtidos através da plataforma <http://zinc.dock.org> (IRWIN; SHOICHET, 2005), com total de 166.210 moléculas, obtidas do Zinc, distribuídas em 14 catálogos de plataformas de formação de banco de dados diferentes que estão subdivididas e 25 subconjuntos diferentes.

As moléculas foram submetidas ao programa Open Babel 2.3.1 (O'BOYLE, 2011), para a adição de hidrogênios polares e das cargas atômicas. Após o acoplamento molecular, as moléculas foram organizados por ordem da afinidade estimada com o sítio ortostérico do modelo da UNAcP. Foram selecionadas 10 moléculas que apresentaram melhor posição no ranqueamento baseado nos resultados do acoplamento molecular.

Adicionalmente as estruturas químicas dessas moléculas, juntamente com os ligantes utilizados para a construção dos falsos positivos, foram submetidas à plataforma *on line ChemGPS-NP* (LARSSON et al, 2007) para a geração dos componentes principais que permitiram a realização das análises quimiográficas, localizando as moléculas de acordo com os descritores químicos específicos que mais representam compostos oriundos de produtos naturais no espaço químico.

A similaridade química entre as moléculas pode ser avaliada através da localização espacial de cada molécula no mapa construído pelo *ChemGPS-NP*, já que os descritores químicos para as componentes são os mesmos aplicados a todas as moléculas. Assim, a similaridade/dissimilaridade pode ser mensurada pela Distância Euclidiana (DE), pois, quanto menor for essa distância maior será a similaridade e menor a dissimilaridade química. A DE foi calculada conforme a seguinte equação:

Equação aplicada no cálculo da Distância Euclidiana entre dois pontos em um gráfico 3D sendo P<sub>1</sub>, P<sub>2</sub> e P<sub>3</sub> os eixos x, y e z respectivamente de determinado ponto e q<sub>1</sub>, q<sub>2</sub> e q<sub>3</sub> os eixos x, y e z respectivamente, do outro ponto que se deseja determinar a distância:

$$DE_{x,y} = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}$$

Fonte: (ROSÉN, 2009) adaptado.

A partir das Distâncias Euclidianas estabelecidas entre as 10 moléculas selecionadas na etapa da triagem virtual e os ligantes, foi possível analisar a similaridade e dissimilaridade entre as estruturas. Com base nesses resultados foi realizado um escalonamento utilizando as variáveis originadas da Função de Pontuação, ou seja, as moléculas selecionadas que apresentaram maior afinidade com o modelo da UNAcP obtidas a partir do método de acoplamento selecionado, essa variável foi considerada de maior importância e, portanto os valores atribuídos a ela foram maiores que os valores atribuídos a segunda variável Escalonada (DE).

Para as 10 moléculas do ranking por Função de Pontuação foram atribuídas uma pontuação de 1 a 10, ou seja, a primeira recebeu pontuação 10 enquanto que a última recebeu pontuação 1. Para o ranking obtido por distância euclidiana foram atribuídas pontuações

variando entre 0,5 a 5 com isto a primeira molécula do escalonamento recebeu pontuação 0,5 enquanto que a última recebeu pontuação 5, assim foi possível pontuar com prioridade maior por afinidade e em contrapartida selecionar moléculas dentro do conjunto com maior diversidade estrutural visto que as moléculas mais distantes apresentam valores melhores no escalonamento por distância.

### **3.4 Análise das interações**

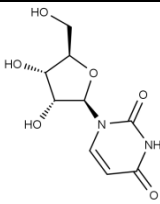
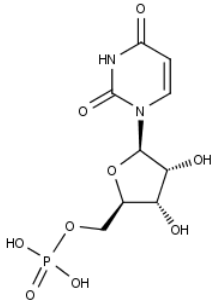
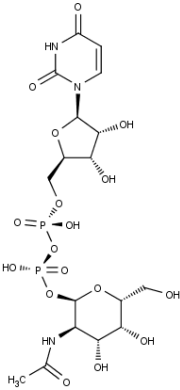
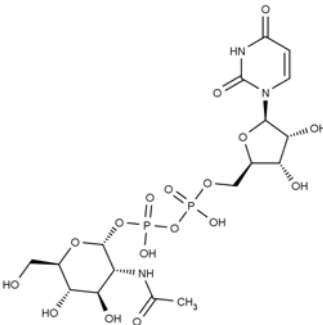
No intuito de analisar as interações, entre as moléculas selecionadas na triagem virtual e os resíduos do sitio ortostérico do modelo da UNAcP, os complexos gerados a partir do posicionamento fornecido pela função de pontuação escolhida foram submetidos ao LigPlot<sup>+</sup> v. 1.4 (LASKOWSKI; SWINDELLS, 2011).

## 4 RESULTADOS E DISCUSSÕES

### 4.1 Moléculas bioativas selecionadas

Após busca na literatura foram identificadas 4 moléculas com constante de inibição (K<sub>i</sub>) para a UNAcP, as estruturas das mesmas e seus respectivos K<sub>i</sub> encontram-se na tabela 1.

**Tabela 1-** Ligantes selecionados com valores de K<sub>i</sub> estabelecidos.

Ligantes	Estrutura	Valor de K <sub>i</sub>
Uridina		1,8μM
Monofosfato de uridina(UMP)		0,25μM
UDP-N-acetilgalactosamina		10μM
UDP-N-acetilglicosamina		10μM

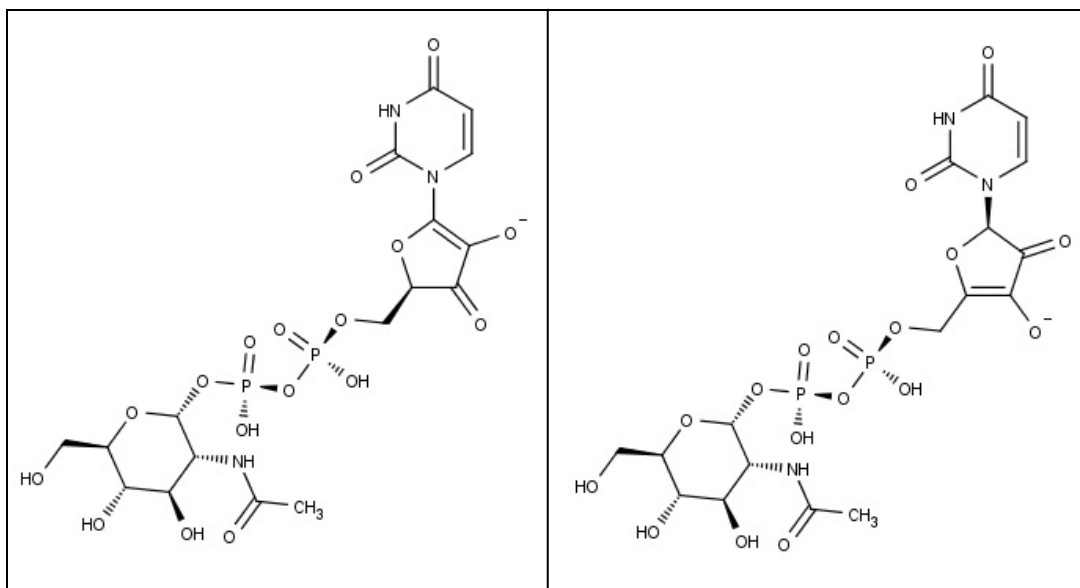
A tabela 1 apresenta os dados referentes à uridina, inibidor não competitivo para UNAcP de *Saccharomyces cerevisiae* com  $K_i=1,8\mu\text{M}$  e inibição competitiva com o PPI (YAMAMOTO et al., 1980). O monofosfato de uridina (UMP) possui  $K_i=0,25\mu\text{M}$  e inibe competitivamente a atividade catabólica da UNAcP de *Giardia intestinalis*, sendo considerado dentre os ligantes selecionados o mais potente. As inibições parecem ser reversíveis tendo em vista que após diálise a atividade enzimática foi retomada (BULIK; LINDMARK; JARROLL, 1998; STROMIGER, SMITH, 1959).

Os estudos de Zeng, e colaboradores (1996) da síntese de derivados da UDP-N-acetilglicosamina e UDP-N-acetilgalactosamina apresentaram inibição eficaz frente à preparação homogênea de UNAcP de fígado de porco sendo os valores de  $K_i=10\mu\text{M}$  para ambas UNAcGln e UNAcGal.

#### **4.2 Avaliação das funções de pontuação**

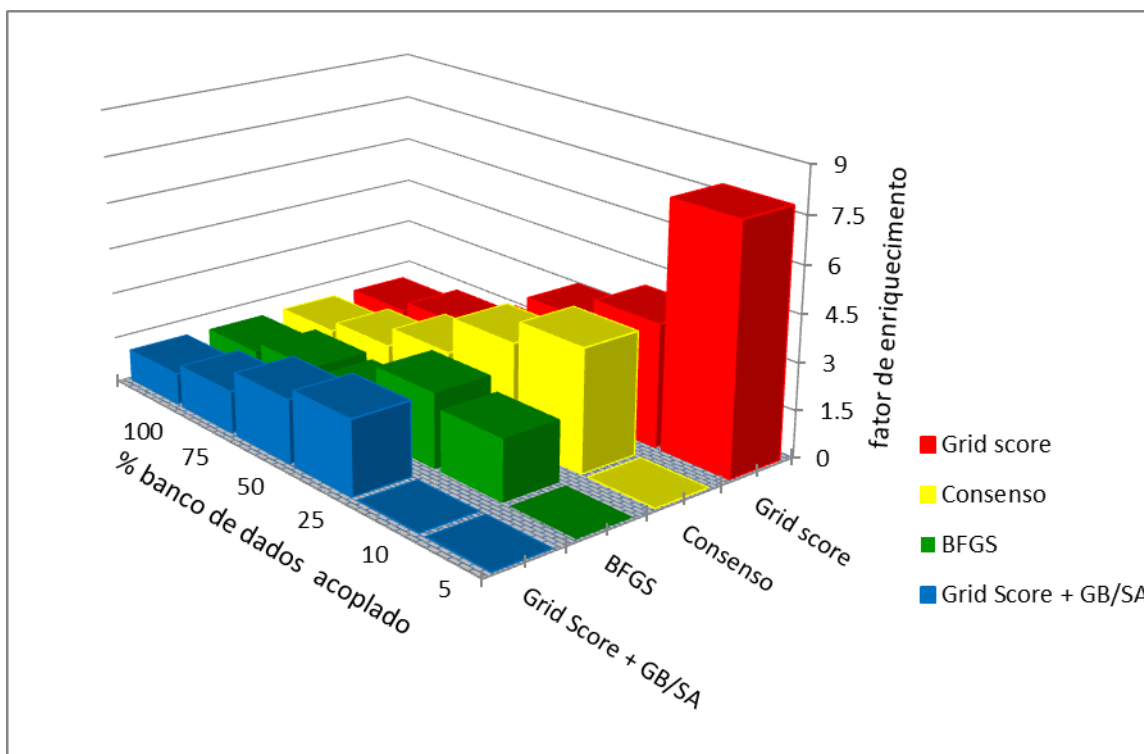
Para avaliação das funções de pontuação foi construído uma biblioteca com 255 moléculas, destas 4 foram os ligantes encontrados na literatura e 250 que representa 98% do conjunto de falso positivos fornecidos a partir da ferramenta de pesquisa DUD-E divididos entre 5 grupos de 50 moléculas para cada ativo submetido à plataforma. Deve-se atentar que o DUD-E considerou os dois estados de protonação da molécula UDP-N-acetilglicosamina (Figura 10) gerando a partir daí um *cluster*, ou seja, ao invés de 200 falsos positivos como seria esperado, foram gerados mais 50 compostos falsos positivos para o segundo estado de protonação formando então uma biblioteca com os 250 falsos positivos. Portanto, apesar de inicialmente termos selecionado 4 ligantes, por conta *do cluster*, tem-se então 5 compostos ativos representando 1,96% do banco de dados, com valores de constante de inibição variando de  $0,25\mu\text{M}$  a  $10\mu\text{M}$  coletados a partir da literatura.

**Figura 10** – Cluster gerado pelo DUD-E a partir dos estados de protonação da molécula UDP-N-acetilglicosamina.



Para a primeira avaliação do Fator de Enriquecimento (FE), os resultados obtidos para cada metodologia estão apresentados na Figura 11, com o respectivo FE para cada função.

**Figura 11** - Fator de enriquecimento para as funções de pontuação utilizadas em 5, 10, 25, 50, 75 e 100% de amostragem de dados.





Entre as funções de pontuação testadas apenas a função Grid Score foi capaz de identificar o maior número de compostos ativos no nível inicial de amostragem da base de dados (5%) (Figura 11) quando comparada com as outras funções de pontuação. Assim, o Grid Score apresentou 7,85 vezes maior chance de identificar os compostos ativos frente a inativos em apenas 5% dos compostos do banco de dados.

Observando a Tabela 2 nota-se que a aplicação do escore por consenso teve um melhor desempenho em relação ao algoritmo BFGS do Autodock Vina 1.1.2 e a repontuação utilizando o algoritmo Hawkins GB/SA, o que denota que sua aplicação também pode ser uma opção passível de ser aplicada. Porém essa metodologia de escore por consenso apenas não conseguiu recuperar nos 5% iniciais do banco de dados, no entanto, a partir de 10% do banco ela permaneceu com desempenho superior e manteve-se a frente dos métodos BGFS e Grid Score + Hawkins GB/SA em termos de recuperação nas faixas percentuais definidas (Figura 11). Seu declínio na taxa de recuperação apresenta uma maior estabilidade e isso pode ser observado na proporção dos 50% do banco de dados (Tabela 2). Nesse ponto, o escore por consenso consegue recuperação superior ao Grid Score, mas logo em seguida a taxa de recuperação torna-se igual.

**Tabela 2** – comparativo do fator de enriquecimento entre as funções e o escore por consenso.

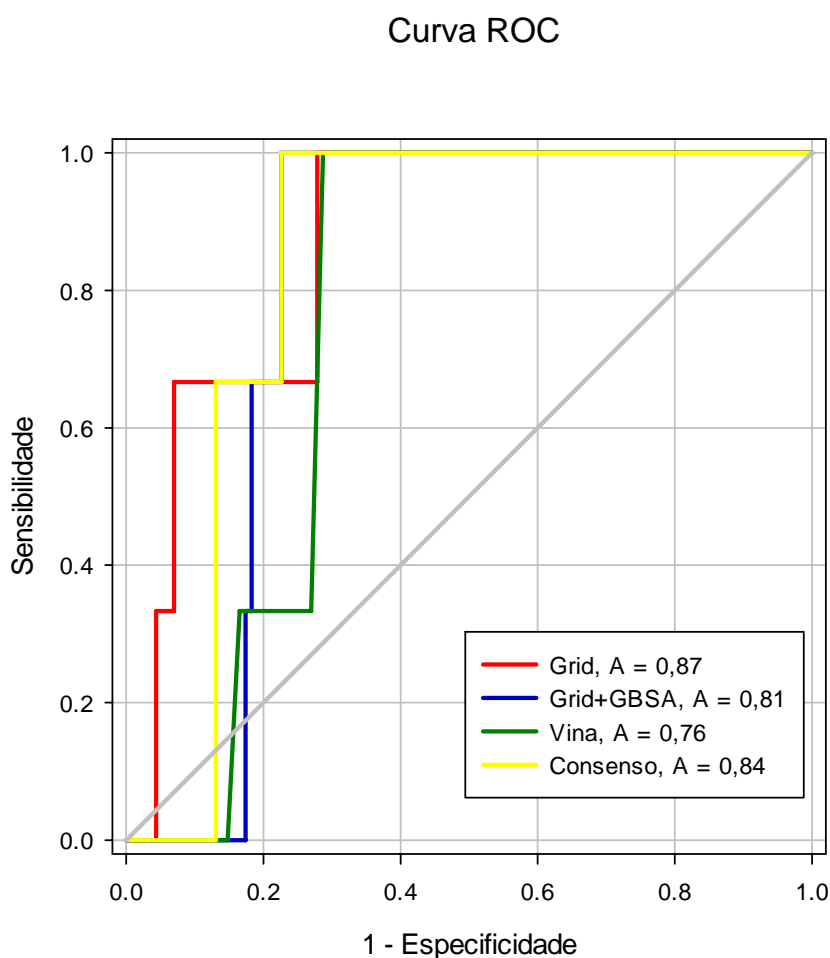
Função de pontuação	5%	10%	25%	50%	75%	100%
Grid score	7,85	3,92	3,18	1,59	1,34	1
Consenso	0	3,92	3,18	1,99	1,34	1
BFGS	0	1,96	2,39	1,19	1,34	1
Grid Score + GB/SA	0	0	2,39	2,03	1,34	1

O método BGFS não consegue recuperação no início do banco de dados e de maneira irregular apresenta taxa de recuperação inferior às demais funções aplicadas. A função aplicada utilizando o Grid Score repontuado com Hawkins GB/SA não conseguiu determinar recuperação em até 10% (Tabela 2) do banco de dados o que denota que os compostos ativos só irão aparecer para essa metodologia ao final da triagem e seria contestável sua utilização para o modelo tendo em vista que o ideal é que os compostos ativos sejam detectados pela triagem logo no início dos acoplamentos.

Adicionalmente foi avaliada em cada uma das metodologias à capacidade de recuperar as moléculas ativas através da análise da curva ROC e a exatidão dos métodos é analisada pelos resultados das AUC. O Gráfico da curva ROC é uma ferramenta muito útil para visualização e avaliação de classificadores. Ele tem a capacidade de fornecer informação mais rica do desempenho das medidas de classificação escalares como especificidade e sensibilidade. De acordo com os resultados apresentados na Figura 12 a função *Grid Score* apresenta uma curva ROC ligeiramente superior às demais metodologias, portanto, pode-se observar que esta função recupera maior quantidade de compostos ativos entre as primeiras do *ranking* e sua precisão é superior quando comparada as outras funções. A análise da curva ROC permite observar que a função *Grid Score* consegue identificar quase 40% das moléculas ativas com menos de 10% do banco de dados de moléculas ativas e falso positivas.

Analisando as curvas pode-se observar que o escore por consenso apresenta uma recuperação superior a partir de 0,65, porém, geralmente são levados em consideração os valores entre 0,1-0,5 de sensibilidade (ligantes recuperados), valores que indicam que o classificador teria então maior precisão (FAWCETT, 2006). De acordo com as análises de Fawcett (2006) isso equivale a dizer que o classificador é melhor para identificar aspectos positivos prováveis do que em identificar prováveis negativos, e nesse sentido, a opção mais indicada para realização da triagem por acoplamento molecular seria a Função *Grid Score*.

**Figura 12** - Avaliação das funções de pontuação quanto à capacidade de recuperação de compostos ativos frente a compostos falsos positivos.



Considerando que a AUC é uma porção da área da unidade quadrada, o seu valor será sempre entre 0 e 1,0, portanto, uma seleção aleatória seria representada pela linha diagonal entre (0,0) e (1,1). A AUC é uma maneira prática de medir o desempenho geral dos testes. Representando uma área de 0,5, os valores menores indicariam seleções inferiores a uma escolha aleatória de baixo desempenho ou sem importância. Os valores maiores que 0,5 denotam propriedade estatística e qualifica a seleção como classificadora. O que significa dizer, neste caso, que o método aplicado que apresenta maior valor de área, tem maior probabilidade de classificar melhor um composto verdadeiro positivo no início dos acoplamentos do que um falso positivo. Nos casos de AUC igual a 1 corresponderia a um caso ideal (TRIBALLEAU et al., 2005; MATSUBARA, 2008).

De acordo com essa análise, quanto maior for a AUC, mais eficaz será a exatidão do método de triagem virtual em discriminar ativos entre moléculas inativas. Dentro dessa faixa

que varia de 0,5 a 1,0 os estudos de Matsubara (2008) sugerem uma classificação em relação aos valores da AUC que auxiliam na interpretação da qualidade da metodologia avaliada, no qual são considerados como bons valores da AUC aqueles acima de 0,8 (Tabela 3).

**Tabela 3** – Classificação dos intervalos das AUC.

<b>Intervalo da AUC</b>	<b>Classificação</b>
<b>0,5 a 0,6</b>	Insignificante
<b>0,6 a 0,7</b>	Ruim
<b>0,7 a 0,8</b>	Regular
<b>0,8 a 0,9</b>	Bom
<b>0,9 a 1,0</b>	Excelente

**Fonte:** Adaptado de MATSUBARA (2008).

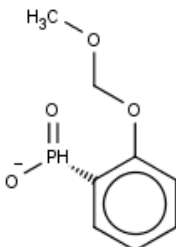
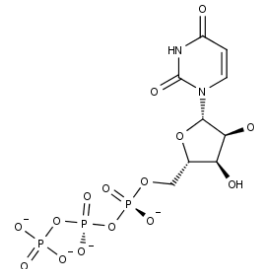
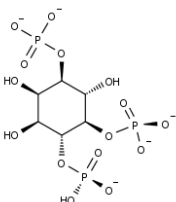
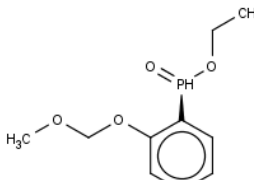
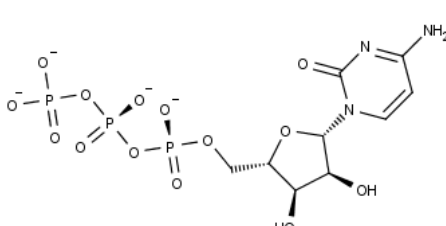
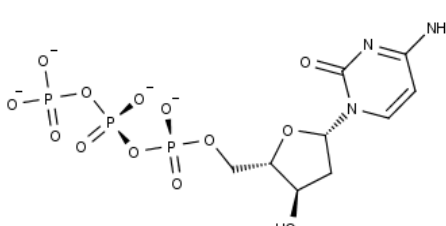
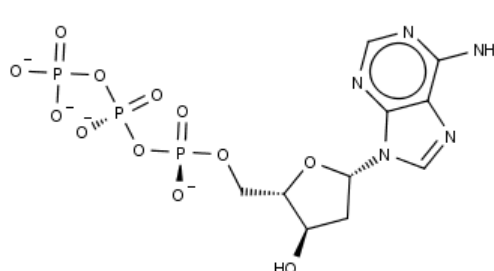
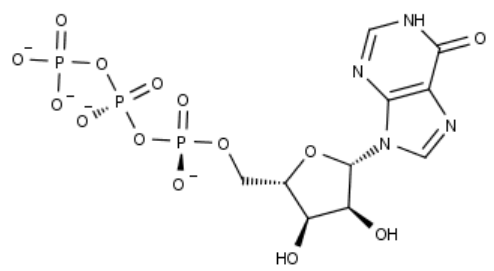
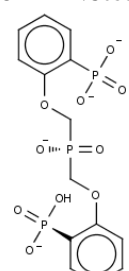
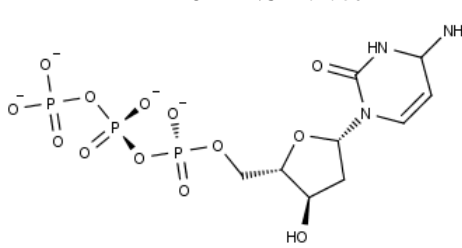
Essa avaliação foi realizada com base na análise da (AUC) da curva ROC (Figura 12). A AUC obtida para a função *Grid Score* apresentou o maior valor (0,87), apresentando-se como a metodologia que classifica melhor, portanto, apresenta maior exatidão em relação às outras funções.

A função *Grid Score* apresenta (AUC=0,87), o escore por Consenso (AUC=0,84) e a função *Grid Score + Hawkins GB/SA* (AUC=0,81) foram consideradas metodologias com boa exatidão (Figura 12), pois, seus valores de AUC apresentam-se acima de 0,8, porém, a exatidão do algoritmo BFGS do Autodock Vina 1.1.2 foi considerada como regular (AUC=0,76) (Figura 12). Nesse caso uma menor exatidão do BFGS pode ser atribuída ao tratamento simplificado do componente entrópico da energia livre de ligação calculada pelo programa, o que pode prejudicar a seleção da conformação bioativa (TROT; OLSON, 2010).

### 4.3 Etapa de Triagem virtual

Uma vez selecionado o programa e a respectiva função de pontuação, realizou-se a triagem virtual por acoplamento molecular utilizando moléculas oriundas de produtos naturais. Ao final, os compostos foram organizados por ordem da afinidade estimada com o sítio ortostérico do modelo da UNAcP. Os 10 compostos melhores ranqueados de acordo com os resultados do *Grid Score* encontram-se dispostos na Tabela 4:

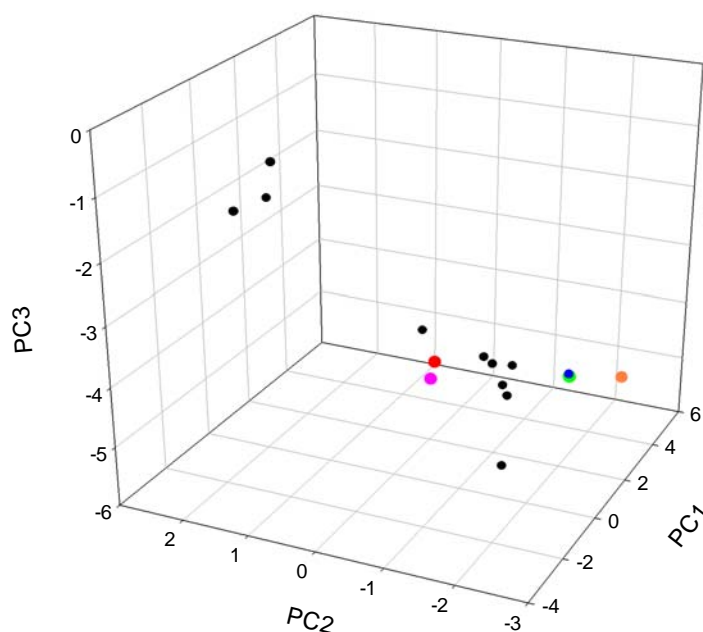
**Tabela 4** – Compostos selecionados na triagem virtual utilizando a função de pontuação Grid Score do DOCK6.5.

<p><b>1- ZINC68592326</b></p>  <p><b>Grid Score: -145.49 Kcal/mol</b></p>	<p><b>2- ZINC19796107</b></p>  <p><b>Grid Score: -141.28 Kcal/mol</b></p>
<p><b>3- ZINC24433941</b></p>  <p><b>Grid Score: -137.38 Kcal/mol</b></p>	<p><b>4- ZINC68592201</b></p>  <p><b>Grid Score: -137.17 Kcal/mol</b></p>
<p><b>5- ZINC19850119</b></p>  <p><b>Grid Score: -136.84 Kcal/mol</b></p>	<p><b>6- ZINC19850122</b></p>  <p><b>Grid Score: -135.53 Kcal/mol</b></p>
<p><b>7- ZINC08215662</b></p>  <p><b>Grid Score: -134.45 Kcal/mol</b></p>	<p><b>8- ZINC08215530</b></p>  <p><b>Grid Score: -134.29 Kcal/mol</b></p>
<p><b>9- ZINC68600011</b></p>  <p><b>Grid Score: -131.36 Kcal/mol</b></p>	<p><b>10- ZINC12494956</b></p>  <p><b>Grid Score: -131.34 Kcal/mol</b></p>

Na tentativa de reduzir a dimensionalidade do espaço químico e evitar problemas de multicolinearidades, os descritores químicos das moléculas selecionadas foram submetidos à plataforma *on line ChemGPS-NP* (LARSSON et al, 2007) para conversão em componentes principais relacionadas à suas propriedades químicas, de acordo com o procedimento disponibilizado no servidor *ChemGPS-NP*.

Com o intuito de maximizar a diversidade química dos potenciais inibidores que poderão ser adquiridos, os descritores topológicos foram utilizados para descrever o espaço químico das moléculas selecionadas a fim de identificar dentre os compostos oriundos de produtos naturais selecionados os vizinhos mais distantes dentre os selecionados frente aos compostos com atividade para o modelo da UNAcP. Os resultados encontram-se na figura 13.

**Figura 13** - Representação em 3D da localização espacial baseados nos componentes principais gerados pela plataforma *on line ChemGPS-NP*. Em rosa UMP; em vermelho a Uridina; em azul e verde o *Cluster* de UNAcGln; em laranja UNAcGal; em preto as 10 moléculas melhores classificados na triagem virtual.



Representações estruturais químicas, impressões digitais estruturais e descritores moleculares são parâmetros comumente utilizados nos cálculos de similaridade e dissimilaridade. Essas entradas são comparadas usando um índice de similaridade, as métricas mais amplamente utilizadas são medidas de distância, simples, tais como as distâncias Euclidianas (ROSÉN, 2009).

Os compostos em estudo foram posicionados no mapa usando interpolação em termos de previsão de pontuação das componentes principais (PCA). As propriedades dos compostos em conjunto com as tendências e os agrupamentos podem ser facilmente interpretados a partir dessas projeções resultantes. Nesse caso a PC1 representa o tamanho, forma e a capacidade de polarização da molécula, a PC2 corresponde aos constituintes aromáticos e propriedades relacionadas conjugação e a PC3 descreve a lipofilicidade, polaridade e capacidade ligação de hidrogênio. Os valores atribuídos aos descritores químicos foram empregados para fazer uma análise dos agrupamentos utilizando as distâncias euclidianas observando os vizinhos mais distantes (Tabela 5). Essa estratégia permite observar o grau de identidade entre as moléculas.

**Tabela 5** - Distância Euclidiana entre os compostos selecionados e os ligantes.

	Moléculas	Distância Euclidiana (DE)					
		Média	Uridina	Ump	UNAcGln1	UNAcGln2	UNAcGal
1	ZINC68592326	6,37	3,45	3,89	7,85	7,83	8,82
2	ZINC68592201	2,63	3,58	4,05	7,76	7,73	8,68
3	ZINC68600011	3,13	4,12	3,86	5,57	5,54	6,49
4	ZINC24433941	6,36	2,84	2,16	3,25	3,25	4,14
5	ZINC08215662	2,60	3,38	2,69	2,71	2,68	3,72
6	ZINC19796107	2,69	3,02	2,32	2,27	2,26	3,26
7	ZINC08215530	3,04	3,92	3,23	1,41	1,39	2,43
8	ZINC19850119	2,48	3,28	2,57	2,06	2,05	3,05
9	ZINC19850122	5,12	2,68	2,01	2,59	2,58	3,57
10	ZINC12494956	2,74	2,46	1,81	2,82	2,80	3,80

Na análise quantitativa da similaridade molecular tabela 5 podem-se observar as distâncias médias entre os produtos naturais e os ligantes variando de 2,48 a 6,37, no entanto as DE absolutas entre as moléculas selecionadas e ranqueadas de acordo com a afinidade baseada no *Grid Score* variam de 1,39 a 8,82.

Nos estudos relacionados a pesquisas de similaridade/dissimilaridade baseadas em descritores químicos, utilizados para identificar moléculas de origem natural que sejam vizinhas mais próximas de medicamentos aprovados, através de cálculos das distâncias euclidianas utilizando as coordenadas do *ChemGPS-NP*, Rosén e colaboradores (2010) encontraram 99,5% de todas os fármacos com pelo menos 1 vizinho próximo sendo produto natural quando a DE = 10 e 85,0% desses fármacos tinham pelo menos um vizinho mais próximo quando a DE = 1. Como os resultados observados na Tabela 5 apresentam DE absolutas abaixo de 10, é possível que essas moléculas apresentem atividade biológica frente ao alvo.

Ainda de acordo com Rosén e colaboradores (2010) moléculas oriundas de produtos naturais dissimilares frente a pequenas moléculas com atividades conhecidas e aprovados como medicamentos desreguladores endócrinos, foram propostas para testes adicionais como compostos líderes com potencial atividade contra esses mesmos alvos. Vários produtos naturais foram revelados por este método e foram confirmadas que diversas moléculas exibem a mesma atividade que os fármacos vizinhos mais próximos.

#### 4.4 Recategorização de variáveis

Após a recategorização com base nas variáveis, *Grid Score* e Distâncias Euclidianas, as pontuações obtidas estão dispostas na Tabela 6.

**Tabela 6** – Recategorização de variáveis ligantes selecionados relacionando *Grid Score* e Distância Euclidiana.

	Moléculas	Pontuação	Grid Score Kcal/mol	Distância Euclidiana				
				Uridina	Ump	UNAcGln1	UNAcGln2	UNAcGal
1	ZINC68592326	33,00	-145,49	3,45	3,89	7,85	7,83	8,82
2	ZINC68592201	29,50	-137,17	3,58	4,05	7,76	7,73	8,68
3	ZINC68600011	23,00	-131,36	4,12	3,86	5,57	5,54	6,49
4	ZINC24433941	21,50	-137,38	2,84	2,16	3,25	3,25	4,14
5	ZINC08215662	17,50	-134,45	3,38	2,69	2,71	2,68	3,72
6	ZINC19796107	17,00	-141,28	3,02	2,32	2,27	2,26	3,26
7	ZINC08215530	16,50	-134,29	3,92	3,23	1,41	1,39	2,43
8	ZINC19850119	14,00	-136,84	3,28	2,57	2,06	2,05	3,05
9	ZINC19850122	13,00	-135,53	2,68	2,01	2,59	2,58	3,57
10	ZINC12494956	11,00	-131,34	2,46	1,81	2,82	2,80	3,80

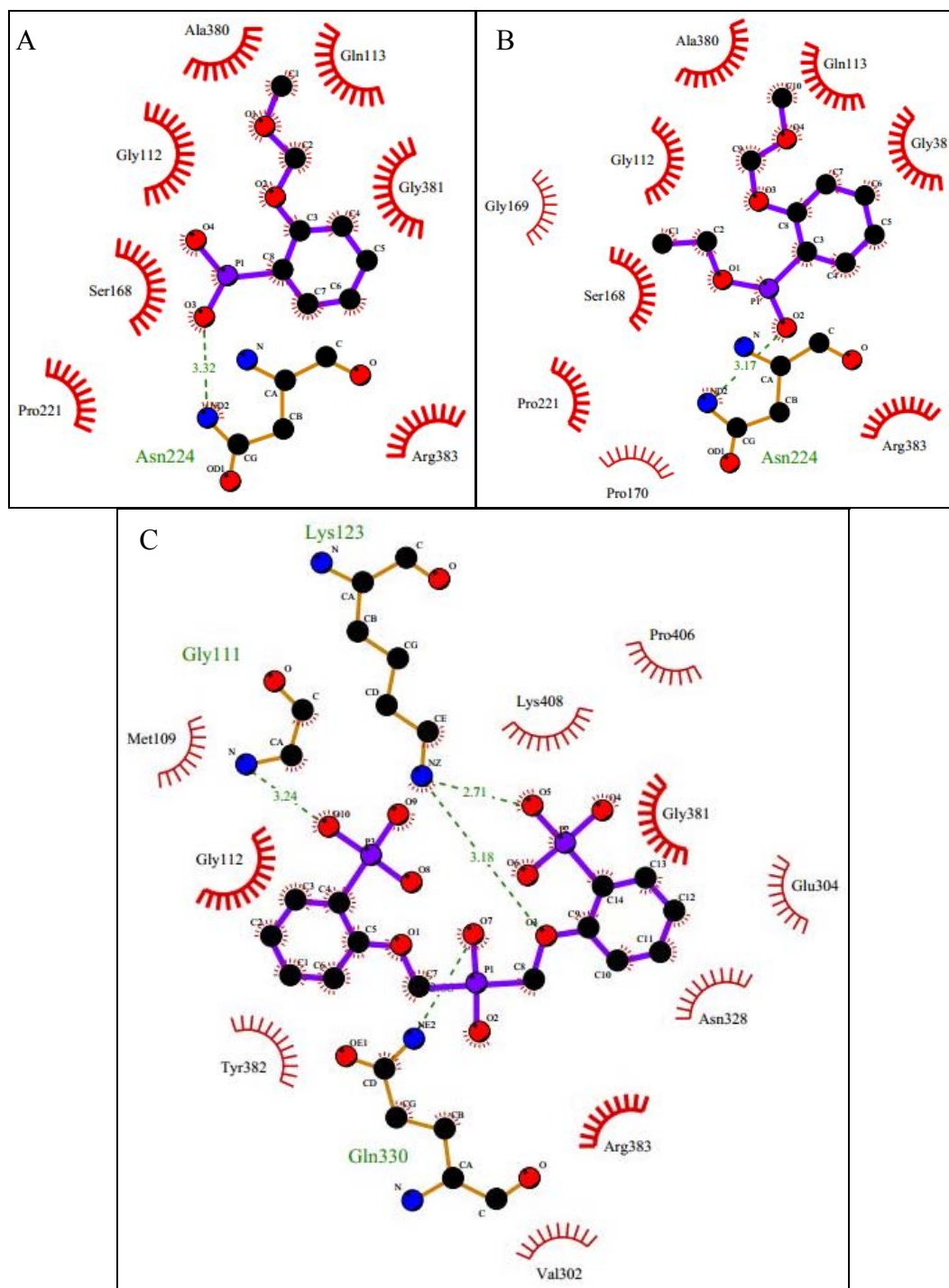
A pontuação baseada na recategorização variou entre 11,00 a 33,00 pontos e a nova classificação está apresentada na Tabela 6. Pode-se observar que dentre as moléculas melhores classificadas, por exemplo, entre a molécula 1 e a molécula 4 há uma diferença de mais de 10 pontos. Levando em consideração que a pontuação dada pela função *Grid Score* teve maior peso para a recategorização e as moléculas com maior DE também obtiveram maior pontuação, a fim de possibilitar a escolha valorizando a função de pontuação e a diversidade estrutural, ficam definidas como estruturas de trabalho as moléculas 1, 2 e 3.



## 4.5 Análise das interações

As interações obtidas a partir do acoplamento das 3 moléculas selecionadas após a recategorização frente ao modelo construído por Santos Junior e colaboradores (2011) estão representadas na Figura 14.

**Figura 14** – Interação no sítio ortostérico entre o modelo da UNAcP e as moléculas selecionadas após a recategorização de variáveis em (A) ZINC68592326, em (B) ZINC68592201 e em (C) ZINC68600011.



De maneira geral dentre os aminoácidos presentes nas interações intermoleculares observadas na Figura 14, a maioria apresentam cadeias laterais apolares são eles, alanina, a cisteína, a glicina, a valina, a leucina, a prolina, e metionina. Vale ressaltar que alanina e valina apresentam na cadeia lateral hidrocarboneto alifático. A prolina tem uma estrutura cíclica alifática e seu nitrogênio está ligado a dois átomos de carbono, o que caracteriza o sítio ortostérico como predominantemente hidrofóbico (LEHNINGER; NELSON; COX, 2000).

Entretanto, podem ser observadas também interações com resíduos de cadeias laterais polares considerados eletricamente neutros em pH neutro, como serina, tirosina, glutamina e asparagina, Vale lembrar que na serina e na treonina, o grupo polar é uma hidroxila (-OH) ligadas a hidrocarbonetos alifáticos e o grupo hidroxila na tirosina é ligado a um grupo hidrocarboneto aromático, o qual eventualmente perde um próton em pH mais altos (LEHNINGER; NELSON; COX, 2000).

O complexo formado entre o modelo da UNAcP e a molécula ZINC68592326 (Figura 14A) apresentam interações hidrofóbicas com os resíduos Ala380, Gln113, Gly112, Gly381, Ser168, Arg383, Pro221 e ligação de hidrogênio do tipo aceptora com distância de 3,32Å com o resíduo Asn224.

Na figura 14B as interações entre o sítio ortostérico do modelo da UNAcP e a molécula ZINC68592201 são observadas interações hidrofóbicas com os resíduos Gly169, Gly112, Gly381, Gln113, Ala380, Pro170, Pro221, Ser168, Arg383, e ligação de hidrogênio do tipo aceptora com distancia de 3,17Å com o resíduo Asn224.

No complexo formado entre o modelo da UNAcP e a molécula ZINC68600011 (Figura 14C) são observadas interações hidrofóbicas com os resíduos Met109, Gly381, Gly112, Val382, Tyr382, Arg383, Glu304, Asn328, Lys408, Pro406 e duas ligações de hidrogênio do tipo aceptora com o resíduos Lis123 em O3 e O5 do ligantes com distancias de 3,18 e 2,71Å respectivamente. São observadas também ligações de hidrogênio do tipo aceptora nos resíduos, Gln330 com O7 do ligante, na distância de 2,60Å e Gly111 com O10 distância 3,24Å.

De acordo com Peneff et al., (2001) os seguintes resíduos são cruciais para a reação feita para UNAcP: Leu108, Gly111, Gly112, Gln197, Gly223, Asn224, Asp254, Gly291, Glu304, Tyr305, Asn328, Hys331, Phe381, Phe383 e Lys408, a porção osídica da UNAcGln

estabelece numerosas ligações de hidrogénio com a Glu304, Hys331 e o Asn224, e também uma interação hidrofóbica com a Phe381 e Phe383.

Santos Junior e colaboradores (2013) em estudos de otimização do complexo formado entre o modelo da UNAcP e o produto UNAcGln demonstraram respectivamente, que teve a formação de ligações de hidrogênio entre os resíduos Asn224 e Leu383 com oxigênio 22 e 29 do ligante, Interações de ligação de hidrogênio também foram constatadas para os composto ZINC68592326 e ZINC68592201 com o resíduo Asn224, Este resíduo é considerado importante para a ligação no sítio ortotérico da UNAcP e, portanto, é essencial para o acoplamento do ligante, levando à formação de interações como ligação de hidrogênio e interações hidrofóbicas (PENEFF et al, 2001; POMPEO, 2001).

Os trabalhos de Santos Junior e colaboradores (2013) sugerem a ocorrência ligações hidrogênio principalmente entre ligantes e os resíduos Arg116, Arg383, Gly381 e Lys408 e interações hidrofóbicas com resíduos Tyr382 e Lys 123.

Os estudos realizados por Teles, (2012) para elucidação teórica do mecanismo catalítico da UNAcP do *M. pernicioso* baseados na geometria da estrutura cristalográfica da UNAcP do fungo *Candida albicans*, descrita por Maruyama e colaboradores (2007) envolvendo simulações de Dinâmica molecular por métodos de QM/MM elucidou a geometria do intermediário da reação e revelou interações para estabilização e orientação molecular com os resíduos de Arg116, Lis123, Gli113 e 225, Esses estudos sugeriram um papel importante posicionamento do fosfato  $\alpha$  da molécula de UTP pelo resíduo Gli112, estes resíduos, em especial a Arg116, seriam então essenciais para a atividade enzimática e devem ser considerados no desenvolvimento de potenciais inibidores.

Estruturas cristalográficas de UNAcP de *Candida albicans* determinadas por com Maruyama e colaboradores (2007) demonstraram que os resíduos de Asn227, Gly294, Glu309, Tyr310, Asn335, e Lys421 interagem ligando-se com o substrato NAcGlc-1-P e também porção NAcGlc-1-P do produto após a conclusão da reação, A porção uridina do produto UNAcGln ligam-se aos resíduos de Met109, Gly111, Gly112, Gln199, e Gly226.

## 5 CONCLUSÃO

Os estudos de triagem virtual apresentaram ferramentas que possibilitaram maior rapidez e menor custo na identificação de moléculas promissoras e permitiram a seleção de moléculas em meio a bancos de dados com um número relevante de estruturas. Porém, as limitações dos métodos de acoplamento demonstraram a necessidade de avaliação para a escolha da ferramenta que melhor representou o sistema em estudo.

A triagem virtual em banco de dados de moléculas oriundas de produtos naturais permitiu a investigação com um universo de estruturas com características muito diversas. Levando em consideração a diversidade estrutural dos ligantes selecionados para a etapa de avaliação da triagem, foi possível obter moléculas também com diversidade estrutural entre os primeiros do ranking, porém, foram encontradas também moléculas muito similares às moléculas de referência.

A utilização de métodos quimiométricos e das ferramentas da quimioinformática para auxiliar na escolha dos compostos sugeridos na etapa de triagem são considerados, nesse estudo como opções muito úteis, pois, permitem uma escolha mais sistemática e consistente das estruturas do que simplesmente uma análise visual, principalmente por levar em consideração descritores químicos e características moleculares das estruturas, permitindo uma avaliação mais criteriosa.

Dentre as interações relatadas vale destacar para a molécula ZINC68592326 as interações hidrofóbicas com os resíduos, Gln113, Gly112, e ligação de hidrogênio do tipo acceptora com distância de 3,32Å com o resíduo Asn224, para a molécula ZINC68592201 as interações hidrofóbicas com os resíduos, Gly112, Gln113, e ligação de hidrogênio do tipo acceptora com distância de 3,32Å com o resíduo Asn224 e para a molécula ZINC68600011 as interações hidrofóbicas com os resíduos Met109, Gly112 e duas ligações de hidrogênio do tipo acceptora com os resíduos, Lys123 em O3 e O5 do ligante, com distâncias de 3,18 e 2,71Å respectivamente, são observadas também ligações de hidrogênio do tipo acceptora no resíduo Gli111 com O10 do ligante distância 3,24Å. Esses resíduos são relatados por diversos autores como essenciais para a atividade da UNAcP.

Através de bancos de moléculas oriundas de produtos naturais é possível identificar novas estruturas com potencial de interagir com a UNAcP permitindo selecionar compostos promissores para serem utilizados em estudos de síntese orgânica, em simulações de Dinâmica Molecular ou serem empregadas em ensaios *in vitro* com o intuito de testar sua

inibição frente a UNAcP do *M. pernicioso*, a fim de, erradicar a VB ou obter um controle mais eficaz dessa praga.

## REFERÊNCIAS

- AIME, M. C. & PHILLIPS-MORA, W. The Causal agents of witche's broom and frosty pod of cacao (chocolate, *Theobroma cacao*) form a new lineage of Marasmiaceae. **Mycol.** 97(5): 1012-1022, 2005.
- ALVAREZ, J. SHOICHET, B. virtual screening in Drug Discovery, London, **CRC Press**, 2005.
- AMARAL, A. T. & MONTANARI, C. A. "Química Medicinal: 25 Anos De Planejamento Racional De Farmacos", **Química Nova**, 25: 39-44, 2002.
- ANDERSON, M. S. & RAETZ, C. R. H. Biosynthesis of lipid A precursors in *Escherichia coli*. A cytoplasmic acetyltransferase that converts UDP-N-acetylglucosamine to UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucos-amine. **J. Biol. Chem.** v.262 p. 5159–5169. 1987.
- BAKER, S. G. CANNOR, R. J. KESSLER, L. G. The partial test design: a less costly way to test equivalence for sensitivity and specificity, **Statist. Med.** v. 17, p. 2219-2232, 1998.
- BENDER, A. AND GLEN, R.C. Molecular similarity: a key technique in molecular informatics. **Org. Biomol. Chem.** v. 2, p. 3204-3218, 2004.
- BERMAN, H. M. et al. The Protein Data Bank. **Oxford Journals.** v. 28, p. 235-242, 1999.
- BETTIOL, W. **Biocontrole de doenças de plantas: uso e perspectivas.** Jaguariúna SP: Embrapa Meio Ambiente, 2009. 341p.
- BOYLE, N. M. O.; BANCK, M.; JAMES, C. A; MORLEY, C.; VANDERMEERSCH, T. AND HUTCHISON, G. R. Open Babel: An open chemical toolbox. **Journal of Chem. informatics.** v. 3, p. 33, 2011.
- BOWMAN, S. M.; FREE, S. J. The structure and synthesis of the fungal cell wall. **Bio Essays**, v. 28, p. 799-808, 2006.
- BOWERS, J. H., BAILEY, B. A., HEBBAR, P. K., SANOGO, S., LUMSDEN, R. D.. The impact of plant diseases on world chocolate production. **Online. Plant Health Progress.** 2001.<<http://www.apsnet.org/edcenter/intropp/HungryPlanet/Chapter14/Pages/ImageGallery.aspx>> acesso em: 27/06/2014.
- BROOIJMANS, N & KUNTZ, I. D. Molecular recognition and docking algorithms. **Annual Reviews of Biophysics and Biomolecular Structure**, v. 32, p. 335-373, 2003.
- BULIK, D. A. et al. Chitin Synthesis in *Saccharomyces cerevisiae* in response to supplementation of growth medium with glucosamine and cell wall stress. **Eukaryotic Cell**, v. 2, n. 5, p. 886-900, 2003.
- CASE, D. A. et al. An overview of the Amber biomolecular simulation package. **Wires Comput. Mol. Sci.** v. 3, p. 198-210, 2013.

CASE, D.A. et al. The Amber biomolecular simulation programs. *J. Computat. Chem.* v. 26, p. 1668-1688, 2005.

COMPANHIA DAS DOCAS DO ESTADO DA BAHIA, 2002. **Apresentam em estatística os principais produtos movimentados.** Disponível em: <http://www.cobeda.com.br/portoilheus>. Acesso em: 25 jan. 2005.

COUPEZ, B; LEWIS, R. A. Docking and scoring – theoretically easy, practically impossible? *Current. Medicinal Chemistry*, v. 13, n. 25, p. 2995-3003, 2006.

DANTAS NETO, A. Perspectivas da Biotecnologia na Cacaucultura. 2000. Disponível em: <http://www.ceplac.gov.br/radar/semfaz/cacaucultura.htm> Acesso em. 02/01/2014.

DAVIS, A. M., TEAGUE, S.J. & KLEYWEGT, G. J. Application and Limitations of X-Ray Crystallographic Data in Structure-Based Ligand and Drug Design, *Aust. J. Chem.*, v. 57, p. 1029-1037, 2004.

DE LUCA, C. et al. Overexpression, one-step purification and characterization of UDP-glucose dehydrogenase and UDP-N-acetylglucosamine pyrophosphorylase. *Biorg. Med. Chem.* v. 4, n.1, p. 131-142, 1996.

DIAS, R. & DE AZEVEDO JR., W. F. Molecular Docking Algorithms. *Current Drug Targets*, v. 9, p. 1040-1047, 2008.

DOBSON, C.M. Chemical space and biology. *Nature*. v. 432, p. 824-828, 2004.

DOUGUET, D. Ligand-based approaches in virtual screening. *Current Computer-Aided Drug Desing*, v. 4, n. 3, p. 180-190, 2008.

ERIKSSON, L. et al. Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). *Anal Bioanal. Chem.* v. 380, p. 419-429, 2004.

EWING, T. J. A. et al. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.*, v. 15, n. 5, p.411-428, maio 2001.

FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*. n. 27, p. 861-874, 2006.

FERRIN, T.E., et al. The Midas display system, *J. Mol. Graf.* v. 6 p. 13–27, 1988.

FERREIRA, M. M. C. Multivariate QSAR, *J. Braz. Chem. Soc.*, v.13 742-753, 2002.

FRIESNER, R. A. et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, v. 47, n. 7, p.1739-1749, 2004.

GABRIELSSON, J., LINDBERG, N. O. & LUNDSTEDT, T. Multivariate methods in pharmaceutical applications. *J. Chemo*. v. 16, p. 141-160, 2002.

GEORGOPAPADAKOU, N. H., TKACZ, J. S. The fungal cell wall as a drug target *Trends Microbiol.* v3, n3, p98-104, 1995.

GOHLKE, H. & KLEBE, G. Statistical Potential and Scoring Functions Applied To Protein-Ligand Binding, **Curr. Op. Struct. Biol.** v. 11 p. 231-235, 2001.

GOHLKE, H. HENDLICH, M. KLEBE, G. Knowledge-based Scoring Function to predict Protein-Ligand Interactions. **J. Mol. Biol.**, v. 295, p. 337-356, 2000.

GUIDO, R. C. V., OLIVA, G. & ANDRICOPULO, A.D. ‘Virtual Screening and Its Integration with modern Drug Design Technologies’, **Curr. Med. Chem.**, v. 15. p. 37-46, 2008.

GUIDO, R. C. V., OLIVA, G. & ANDRICOPULO, A.D. Planejamento de fármacos, biotecnologia e química medicinal: aplicações em doenças infecciosas. **Estud. av.** v. 24. n. 70. 2010.

HAJDUK, P. J.; GREER, J. A decade of fragment-based drug design: strategic advances and lessons learned. **Nature Reviews Drug Discovery**, USA, v. 6, p. 211-219, 2007.

HAWKINS, G. D.; CRAMER, C. J.; TRUHLAR, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. **Chem. Phys. Lett.** v. 246 p. 122-129, 1995.

HAWKINS, G. D.; CRAMER, C. J.; TRUHLAR, D. G. Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. **J. Phys. Chem.** v. 100, p. 19824-19839, 1996.

HUANG, N., SHOICHET, B. K., IRWIN, J. J. Benchmarking Sets for Molecular Docking. **American Chemical Society**, 2006.

HUANG, SHENG-YOU. GRINTER, SAM Z. ZOU, XIAOQIN. Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. **Phys. Chem. Chem. Phys.**, v. 12, 2010.

IRWIN, J. J., et al. ZINC: A Free Tool to Discover Chemistry for Biology. **Journal Chemical of Information Modeling**, v. 52, p. 1757-1768, 2012.

JAIN, A. N. Surflex: Fully automatic flexible molecular docking using based search engine. **J. Med. Chem.**, v. 46, p. 2287-2303, 2003.

JOHNSON, M. MAGGIORA, G. M. Concepts and Applications of Molecular Similarity, John Wiley & Sons, **New York**, 1990. *apud*. Rosén, J. **ChemGPS-NP and the Exploration of Biologically Relevant Chemical Space**. Dissertation (Doctor of Philosophy Faculty of Pharmacy). Universities Upsaliensis, Uppsala, 2009.

JORGENSEN, W.L. ‘The Many Roles of Computation in Drug Discovery’, **Science**, v. 303 p. 1813-1818, 2004.

KALYAANAMOORTHY, S.; CHEN, Y. P. Structure-based drug design to augment hit discovery. **Drug. Discov. Today**, v. 16, p. 831-839, 2011.

KITCHEN et al. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and applications. **Nat. Rev. Drug Discov.**, v. 3, p. 935-949, 2004.



KRAMER, B.; RAREY, M.; LENGAUER, T. Evaluation of the FLEXX Incremental Construction Algorithm for Protein–Ligand Docking. **Proteins: Structure, Function, and Genetics**, v. 37, p. 228-241, 1999.

KUKOL, A. Consensus virtual screening approaches to predict protein. **Ligands European Journal of Medicinal Chemistry**. v. 46, p. 4661-4664, 2011.

KUNTZ, I. D. et al. A geometric approach to macromolecule–ligand interactions, **J. Mol. Biol.** v.161 p. 269–288, 1982.

KURTZ, M. B. ; REX, J. H. Glycan synthase inhibitors as antifungal agents. In: RICHARDS F. M. ; EISENBERG, D. S. ; KIM, P. S. (Org.). **Advances in protein chemistry: drug discovery and design**. Pennsylvania: Academic Press. v. 56, p. 424-475. 2001.

LAGORCE, A., BERRE-ANTON, V., AGUILAR-USCANGA, B., MARTIN-YKEN, H., DAGKESSAMANSKAIA, A., FRANÇOIS, J. Involvement of GFA1, which encodes glutamine–fructose-6-phosphate aminotransferase, in the activation of the chitin synthesis pathway in response to cell-wall defects in *Saccharomyces cerevisiae*. **Eur. J. Biochem.** v. 269, p. 1697-1707, 2002.

LARSSON, J., GOTTFRIES, J., BOHLIN, L., BACKLUND, A. Expanding the ChemGPS chemical space with natural products. **Journal of Natural Products**. v. 68, p. 985-991, 2005.

LARSSON, J., GOTTFRIES, J., MURESAN, S., BACKLUND, A. ChemGPS-NP: Tuned for navigation in biologically relevant chemical space. **Journal of Natural Products**. v. 70, p. 789-794, 2007.

LASKOWSKI R A, SWINDELLS M B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. **J. Chem. Inf. Model.** v. 51, p. 2778-2786, 2011.

LEHNINGER, A. L.; NELSON, D. L.; COX, M. M. **Princípios de Bioquímica**. 3. ed. São Paulo: Reimpressão, 2000.

LINUSSON, A. et al. Statistical molecular design of building blocks for combinatorial chemistry. **J. Med. Chem.** v. 43, p.1320-1328, 2000.

LIU, S. et al. Application of Consensus Scoring and Principal Component Analysis for Virtual Screening againstb-Secretase (BACE-1). **PLoS ONE**. 6 ed., v. 7, 2012.

LUZ, E. D. M. N. et al. M. L. de Cacau (*Theobroma cacao L.*) – controle de doenças. In: Valle, F. X. R. & Zambolim, L. (Eds.) **Controle de doenças de plantas: grandes culturas**. v. 2. Visconde do Rio Branco. Suprema. p. 611-655. 1997.

LYNE, P. D. “STRUCTURE-BASED Virtual Screening: No overview”, **Drug Discov. Today**, v. 7, p. 1047-1055, 2002.

MACEDO, V.U.M., et al. Aspectos gerais do *Moniliophthora perniciosa* (stahel) aime& phillips-mora, o agente etiológico da vassoura-de-bruxa. **Sitientibus Série Ciências Biológicas**. v. 9 p. 57-65, 2009.

MARELLI, J. P., MAXIMOVAS. N., GRAMACHO K. P., KANG, S. JOHN, GUILTINAN M. J. Infection Biology of *Moniliophthora perniciosa* on *Theobroma cacao* and Alternate Solanaceous Hosts. **Tropical Plant Biol.** v. 2. p. 149–160. 2009.

MARUYAMA, D.; et al. Purification, crystallization and preliminary X-ray diffraction studies of UDP-N-acetylglucosamine pyrophosphorylase from *Candida albicans*. **Acta Crystallographica**, v. 62, p. 1206-1208, 2006.

MARUYAMA, D. et al. Crystal Structure of Uridine-diphospho-N-acetylglucosamine Pyrophosphorylase from *Candida albicans* and Catalytic Reaction Mechanism. *J. Bio. Chem.* v. 282, n. 23, p. 17221-17230, 2007.

MATSUBARA, E. T. **Relações entre Ranking, Análise ROC e Calibração em Aprendizado de Máquina.** Tese (Doutorado em ciências de computação e matemática computacional) USP – São Carlos 2008.

MEDEIROS, F.H.V. et al. A novel, integrated method for management of witches' broom disease in Cacao. *Crop Protection.* v. 29, p. 704-711, 2010.

MENG, E. C., SHOICHET, B. K. KUNTZ, I. D. Automated docking with grid-based energy evaluation. **J. Comp. Chem.** v. 13: p. 505-524, 1992.

MIO, T. et al. The eukaryotic UDP-N-Acetylglucosamine pyrophosphorylases. **The Journal of Biological Chemistry.** v. 273, n. 23, p. 14392-14397, 1998.

MIO, T. YAMADA-OKABE, T. ARISAWA, M. YAMADA-OKABE, H. Functional cloning and mutational analysis of the human cDNA for phosphoacetylglucosamine mutase: identification of the amino acid residues essential for the catalysis. **Biochem. Biophys. Acta.** v. 1492 p. 369–376 2000.

MOCHALKIN, I. et al. Characterization of substrate binding and catalysis in the potential antibacterial target N-acetylglucosamine-1-phosphate uridylyltransferase (GlmU). **Protein. Science.** v. 16. p. 2657–2666. 2007.

MONTANARI, C. A., **Química Medicinal: métodos e fundamentos em planejamentos de fármacos**, EDUSP. São Paulo, 2011.

MUNOS, B. Lessons from 60 years of pharmaceutical innovation. **Drug Discovery**, v. 8, p. 959-968, 2009.

MYSINGER, M. M. et al. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. **Journal of Medicinal Chemistry.** v. 55, p. 6582–6594, 2012.

O'BOYLE, N. M., et al. Open Babel: An open chemical toolbox. **Journal of Cheminformatics** v.3 p. 33, 2011.

OLIVEIRA, M. L.; LUZ, E.D.M.N. **Identificação e manejo das principais doenças do cacaueteiro no Brasil.** Ilhéus, CEPLAC/CEPEC/SEFIT. p. 132. 2005.

OLSEN, L. R.; RODERICK, S. L. Structure of *Escherichia coli* GlmU pyrophosphorylase and acetyltransferase active sites. **Biochem.** n. 40, p. 1913-1921, 2001.

OSHIRO, C. et al. Performance of 3D-Database Molecular Docking Studies into Homology Model, **J. Med. Chem.**, v. 47, p. 764-767, 2004.

PATRICK, G. L. **An introduction to medicinal chemistry**. 2ed. Oxford University, 2001.

PENEFF, C. et al. Crystal structures of two human pyrophosphorylase isoforms in complexes with UDPGlc(Gal)NAc: role of the alternatively spliced insert in the enzyme oligomeric assembly and active site architecture. **The EMBO J.** v. 20, n. 22, p. 6191-6202, 2001.

PEREIRA, J. L. et al. First occurrence of witches 'broom disease in the principal cocoa-growers region of Brazil. **Trop. Agric.** v. 67, n. 2, p. 188-189, 1990.

PEREIRA, R. G. **Relação da qualidade do cacau no mercado atual e no mundo**: Interação produção de cacau, preço versus qualidade, 2005. Disponível em: <http://www.ceplac.gov.br/radar/semfaz/mercadoatual.htm> Acesso em: 02/01/2014.

PETTERSEN, E. F., et al. UCSF Chimera—a visualization system for exploratory research and analysis, **J. Comput. Chem.** v. 25 p. 1605–1612, 2004.

PIROVANI, C. P et al., Knowledge Discovery in genome database: the chitin metabolic pathway in *Crinipellis Perniciosa*. Proceedings of International Symposium on Mathematical and Computational Biology. 2005.

POMPEO, F. et al. Dissection of the bifunctional *Escherichia coli* N-Acetylglucosamine –1-phosphate uridyltransferase enzyme into autonomously functional domains and evidence that trimerization is absolutely required for glucosamine-1-phosphate acetyltransferase activity and cell growth. **J. Bio. Chem.** v. 276, p. 3833-3839, 2001.

PURDY, L. H. Fungal disease of cacao. 2005. Disponível em: <<http://www.cabi-commodities.org/Acc/ACCrc/PDFFiles/W-BPD/Ch1.pdf>>. Acesso em: 16 jun. 2014.

QIU, D. et al. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. **J. Phys. Chem.** 101 (16), p. 3005–3014, 1997.

RESENDE, M. L. V.; ARAUJO, D. V; COSTA, J. C. B.; DEUNER, C. C.; FERREIRA, J. B.; MUNIZ, M. F. S.; REIS, S. N.; SANTOS, F. S.; CAVALCANTI, L. S.; NOJOSA, G. B. A. Produtos comerciais à base de bioindutores de resistência em plantas. **Revisão Anual de Patologia de Plantas**. v. 14. p. 361-380, 2006.

RHODES, G. Crystallography Made Crystal Clear: A Guide for users of Macromolecular Models, New York, **Academic Press**, 1993.

RIZZO, R. C.; AYNECHI, T.; CASE, D. A.; KUNTZ, I. D. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. **J. Chem. Theory Comput.** v. 2: p. 128-139, 2006.

ROCCHIA, W. et al., Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects **J. Comput. Chem.** v. 23, p. 128–137, 2002.

RODRIGUES, R. P. et al. Estratégias de Triagem Virtual no Planejamento de Fármacos. **Revista Virtual de Química**, São Paulo, v. 4, n. 6, p.739-776, 2012.

ROSÉN, J. et al. Novel Chemical Space Exploration via Natural Products. **J. Med. Chem.** v. 52 (7), p. 1953–1962, 2010.

ROSÉN, J. **ChemGPS-NP and the Exploration of Biologically Relevant Chemical Space.** Dissertation (Doctor of Philosophy Faculty of Pharmacy). Universities Upsaliensis, Uppsala, 2009.

ROSÉN, J. et al. ChemGPS-NP Web: chemical space navigation online. *J. Comput. Aided. Mol. Des.* 2008.

SANTOS JUNIOR, M. C. **Purificação parcial, caracterização e estudos de ancoragem molecular da pirofosforilase do *Moniliophthora perniciosa*.** Tese (Doutorado em Biotecnologia). Universidade Estadual de Feira de Santana, Feira de Santana, 2012.

SANTOS JUNIOR, M. C. Purification, characterization and structural determination of UDP-N-Acetylglucosamine Pyrophosphorylase Produced by *Moniliophthora perniciosa*. **J. Braz. Chem. Soc.**, Vol. 22, No. 6, 1015-1023, 2011.

SANTOS JUNIOR, M. C. et al. A. Homology modeling of pyrophosphorylase, enzyme involved in chitin pathway of *Moniliophthora perniciosa*. **Int. J. Bioinf. Res. App.** v.5, n.2, 133-153, 2009.

SANTOS JUNIOR, M. C. DE ASSIS, S. A. GÓES-NETO, A. DUARTE, A. A. ALVES, R. J. COMAR JUNIOR, M. TARANTO, A. G. Structure-based drug design studies of UDP-N-acetylglucosamine pyrophosphorylase, a keyenzyme for the control of witches' broom disease. **Chemistry Central Journal.** v. 7, p. 48, 2013.

SHOICHET, B. K. KUNTZ, I.D. BODIAN, D.L. Molecular docking using shape descriptors, **J. Comput. Chem.** v. 13 p. 380–397, 1992.

SHOICHET, B.K. Virtual Screening of Chemical Libraries, **Nature**, v. 432 p. 862-865, 2004.

SONG, M. S.; LIM, S. J.; TONG J. C. Recent advances in computer-aided drug desing. **Briefings in Bioinformatics**, v. 10, n. 5, p. 579-591, 2009.

SOTRIFFER, C. A. SANSCHAGRIN, P. MATTER, H. KLEBE, G. "SFCscore: scoring functions for affinity prediction of protein- ligand complexes," **Proteins**, v. 73, p. 395-419, 2008.

SAUNDERS, J. A. **Witches broom fungus *Crinipellis perniciosa* (Stahel) Singer.** disponível em: <<http://www.invasive.org/browse/subinfo.cfm?sub=4881#sthash.AiJB99ls.dpuf>> May 04, 2010. Acesso em 20/06/2014.

SOUZA, C.S. **Sintase da quitina de *Moniliophthora perniciosa* (Sthael) (Singer) Phillips-Mora:** caracterização gênica e modelagem do provável sítio catalítico. Dissertação (Mestrado em Biotecnologia). Universidade Estadual de Feira de Santana, Feira de Santana 2007.

SIPPL, M. J. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. **J. Mol. Biol.**, v. 213, p. 859–883. 1990.

STAHURA, F.L. & BAJORATH, New Methodologies for Ligand-Based Virtual Screening, **Curr. Pharm. Des.**, v. 11 p. 1189-1202, 2005.

STEINDL, T. & LANGER, T. Docking versus Pharmacophore Model Geration: A comparison of High-Throughput the search of human Rhinovirus Virtual Screening Strategies for Coat Protein Inhibitors. **QSAR Comb. Sci.**, v. 24, p 470-479, 2005.

STROMINGER, J. L. & SMITH, M. S. Uridine Diphosphoacetylglucosamine Pyrophosphorylase. **The Journal of Biological Chemistry**. v. 234, nº 7, 1959.

SZUMILO, T.; et al. Purification to homogeneity and properties of UDP-GlcNAc (GalNAc) pyrophosphorylase. **The Journal of Biological Chemistry**, v. 271, n. 22, p. 13147-13154, 1996.

TANG, H. et al. Do crystal structures obviate the need for theoretical models of GPCRs for structure-based virtual screening?. **Proteins**, v. 80, p. 1503–1521 2012.

TANG, Y. T.; MARSHALL, G. R. Virtual screening for lead Discovery. **Methods Molecular Biology**, v.716, p.1-22, 2011.

TELES, A. L. B. **Estudo do mecanismo catalítico da enzima pirofosforilase do fungo *Moniliophthora perniciosa* por de métodos QM/MM**. Dissertação (Mestrado em Biotecnologia). Universidade Estadual de Feira de Santana, Feira de Santana, 2011.

TEÓFILO, R. F. et al. Sorting Variables by Using Informative Vectors as a Strategy for Feature Selection in Multivariate Regression. **J. Chemom.** v. 23, p. 32-48, 2009.

TERAMOTO, R. & FUKUNISHI, H. Consensus Scoring with Feature Selection for Struture-Based Virtual Screening, **J. Chem. Inf. Model.**, v. 48 n. 288-295, 2008.

THOMAS, G. **Fundamentals of medicinal chemistry**. West Sussex: Wiley, 2003.

TRIBALLEAU, N. et al. Virtual Screening Workflow Development Guided by the “Receiver Operating Characteristic” Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. **J. Med. Chem.** v. 48, p. 2534-2547. 2005.

TROTT, O., OLSON, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **J. Comp. Chem.** v. 31, n.2, p. 455-461, 2010

TSUI, V. AND CASE, D. A. Theory and applications of the generalized solvation model in macromolecular simulations. **Biopolymers**. v. 56, p.275-291, 2001.

WANG, R., LAI, L., WANG, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. **J. Comput.-Aided Mol. Des.** v. 16, p. 11–26, 2002.

WARREN, G. L. et al, A critical assessment of Docking Programs and Scoring Functions, **J. Med.Chem.** v. 49 p. 5012-5931, 2006.

WILLETT, P., BARNARD, J.M. AND DOWNS, G.M. Chemical similarity searching. **J. Chem. Inf. Comput. Sci.** v. 38, p. 983-996 1998.

WOLD, S., ESBENSEN, K. AND GELADI, P. Principal component analysis. **Chemometr. Intell Lab. Syst.** v. 2, p. 37-52, 1987. *apud.* Rosén, J. **ChemGPS-NP and the Exploration of Biologically Relevant Chemical Space.** Dissertation (Doctor of Philosophy Faculty of Pharmacy). Universities Upsaliensis, Uppsala, 2009.

VERDONK, M. L. et al. Improved Protein–Ligand Docking Using GOLD. **Proteins: Structure, Function, and Genetics**, v. 52, p.609-623, 2003.

YAMADA-OKABE T, YSAKARNORIL, TMIO, H YAMADA-OKABE. 2001. Identification and characterization of the genes for N-Acetylglucosamine kinase and N-Acetylglucosamine-phosphate deacetylase in the pathogenic fungus *Candida albicans*. **Eur. J. Biochem.** 268: 2498-2505.

YAMAMOTO, K. et al. Inhibition of UDP-N-acetylglucosamine pyrophosphorylase by uridine. **Biochimica et Biophysica.** n. 614, p. 367-372, 1980.