

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE
CURSO DE GRADUAÇÃO EM BIOMEDICINA

ELISA BEATRIZ DE OLIVEIRA JOHN

BIOLOGIA ESTRUTURAL DE PROLIL-4-HIDROXILASES

Porto Alegre
DEZEMBRO/2014

ELISA BEATRIZ DE OLIVEIRA JOHN

BIOLOGIA ESTRUTURAL DE PROLIL-4-HIDROXILASES

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul, como requisito parcial para obtenção do título de Bacharel(a) em Biomedicina.

Orientador: Prof. Dr. Hugo Verli

Porto Alegre

DEZEMBRO/2014

Agradecimentos

Agradeço, primeiramente, ao prof. Hugo Verli, por ter me acolhido no Grupo de Bioinformática Estrutural, por ter oferecido a oportunidade em desenvolver trabalhos na área, depositado confiança para que este trabalho fosse realizado e ter oferecido todo o suporte intelectual para tanto. Sou muito grata pelos conselhos sempre válidos que aumentaram a qualidade do trabalho e ajudaram-me a aprender e mudar muito no período que estive trabalhando no laboratório durante a graduação.

Agradeço aos colegas do Grupo, por todo suporte e convivência. Em especial, agradeço ao Laércio Pol-Fachin, que me ajudou desde o início deste trabalho, ensinando as metodologias e tirando minhas dúvidas sempre com muita paciência.

À UFRGS, aos que trabalham em prol do curso de Biomedicina e ao Centro de Biotecnologia por todas as possibilidades, pelo incentivo à pesquisa e pela grande qualidade de ensino oferecido.

Aos membros da banca, por aceitarem o convite.

Agradeço à minha família e ao meu namorado por todo amor, cuidado (principalmente quando tive problemas de saúde), incentivo constante e apoio emocional. Agradeço especialmente à minha mãe e à minha vó por estarem sempre presentes quando precisei e por todo esforço que fizeram para que eu me mantivesse bem.

“Grau, teurer Freund, ist alle Theorie

Und grün des Lebens goldner Baum.”

Faust I (Mephistopheles) - Johann Wolfgang von Goethe

ÍNDICE GERAL

Lista de Abreviaturas	VI
Resumo	VII
Índice de figuras	VIII
1. Introdução	1
1.1. Paredes celulares de plantas	1
1.1.1. Extensinas	3
1.1.2. Prolil-4-hidroxilases	3
1.2. Caracterização de biomoléculas por modelagem molecular	7
1.2.1. Modelagem comparativa	8
1.2.2. Simulações de dinâmica molecular.....	9
2. Objetivos	11
3. Metodologia	12
4. Resultados	14
5. Conclusões	28
6. Perspectivas	29
7. Agradecimentos	30
8. Bibliografia	31
9. Anexos	37

Lista de Abreviaturas

C-P4H – Prolil-4-hidroxilase relacionada ao colágeno

DM – Dinâmica molecular

EC – Enzyme commission

EXT – Extensina

GFP (*green fluorescent protein*) – Proteína verde fluorescente

HIF-P4H – Prolil-4-hidroxilase relacionada aos fatores de transcrição induzíveis por hipóxia

HRGP (*hydroxiprolin rich glycoproteins*) – Glicoproteínas ricas em hidroxiprolina

Hyp – Hidroxiprolina

PDB – *Protein Data Bank*

Pro – Prolina

P4H – Prolil-4-hidroxilase

RMSD – *root mean square deviation*

RMSF – *root mean square fluctuation*

2OG – 2-oxoglutarato

Resumo

Paredes celulares de plantas são estruturas complexas e dinâmicas, compostas principalmente por polissacarídeos e glicoproteínas. As glicoproteínas ricas em hidroxiprolina (HRGP), por exemplo, são encontradas nas paredes celulares de plantas e algas verdes, relacionando-se ao crescimento e desenvolvimento vegetal. A formação pós-traducional de hidroxiprolina nestas proteínas é catalisada pelas prolil-4-hidroxilases (P4H) que oxidam repetições do tipo PPPP, assim, também assumem papel importante no desenvolvimento das plantas. Em *Arabidopsis thaliana*, existem pelo menos 13 tipos de P4Hs com diferentes perfis de expressão, sendo P4H5, P4H2 e P4H13 os mais expressos em pelos radiculares, estruturas das plantas com um papel importante para a absorção de nutrientes. Neste contexto, o presente trabalho buscou caracterizar a complexação de P4Hs a uma HRPG denominada extensina (EXT), particularmente a preferência sobre os resíduos de prolina nas repetições PPPP, empregando técnicas de modelagem comparativa e dinâmica molecular. Os resultados obtidos indicaram um importante efeito estabilizador da EXT sobre P4H5. Adicionalmente, foram encontradas áreas de maior mobilidade na enzima, como uma alça que cobre o sítio catalítico e parece estar envolvida no acesso do substrato ao sítio ativo. Ainda, certos resíduos de prolina parecem estar em posições preferenciais para a catálise, uma vez que o comportamento conformacional da enzima P4H5 é modificado conforme diferentes resíduos de prolina da repetição PPPP são alinhados ao sítio catalítico da enzima.

Palavras-chave: Prolil-4-hidroxilases; dinâmica molecular; parede celular; modelagem molecular.

Índice de figuras

- Figura 1. Modelo para a estrutura da parede celular primária. As microfibrilas de celulose e os polissacarídeos da matriz tem vias de síntese distintas: a celulose é sintetizada por complexos hexaméricos na membrana plasmática; as hemiceluloses e pectinas são sintetizadas no complexo de Golgi e transportadas por vesículas para a parede celular. As hemiceluloses formam uma rede resistente junto das microfibrilas, enquanto as diversas pectinas estão dispostas dentro desta rede, influenciando assim na porosidade e disposição da celulose, e também criando uma camada adesiva mais externa, a lamela média, que ajuda a unir as células vegetais adjacentes (Adaptado de Cosgrove, 2005; Varner *et al.*, 1989).....2
- Figura 2. Reação catalisada pelas P4Hs. A molécula de 2-oxoglutarato (2OG) coordena o íon ferro através dos átomos de oxigênio C₁-carboxilato e C₂-ceto. O produto da reação é um resíduo trans-4-hidroxi-prolil. Os dois átomos de oxigênio (marcados em vermelho) são usados para a hidroxilação do resíduo de prolina e para a oxidação do 2OG, que é descarboxilado pela liberação do grupo C₁-carboxilato na forma de CO₂ (Reproduzido de Koski *et al.*, 2009).....4
- Figura 3. Estruturas cristalográficas de três diferentes dioxigenases dependentes de 2OG e duas proteínas relacionadas. A primeira estrutura corresponde a uma HIF-P4H de *Homo sapiens* (PDB ID 2G1M), a segunda a uma P4H da alga *Chlamydomonas reinhardtii* (PDB ID 3GZE) e a terceira a uma prolil-3-hidroxilase de *Streptomyces sp* (PDB ID 1E5S). As duas estruturas na parte de baixo da figura (PDB IDs 1H2N e 2FCT) correspondem respectivamente à asparaginil hidroxilase de *Homo sapiens* e à halogenase SyrB2 de *Pseudomonas syringae*, que compartilham com as P4Hs alguns motivos estruturais, como as hélices marcadas em vermelho e folhas marcadas em bege, e os resíduos que coordenam o metal cofator nos centros catalíticos (Reproduzido de Gorres *et al.*, 2010).....5
- Figura 4. Representação da estrutura dos pelos radiculares, projeções da parede celular que auxiliam na absorção de água e nutrientes nas plantas (A) e registros de microscopia de fluorescência evidenciando a expressão de P4Hs, marcadas com GFP, nos pelos radiculares em *A. thaliana* (B). Adaptado de Velasquez *et al.*, 2011.....6

Figura 5. Diferentes fenótipos de pelo radicular em *A. thaliana*. Em (A), o fenótipo selvagem, em (B) o fenótipo de pelos radiculares curtos, resultante de mutações na enzima P4H5. A superexpressão de P4H5 leva à formação de pelos radiculares mais longos (C), enquanto mutações em um tipo de extensina, alvo de P4H5, leva à formação de pelos radiculares curtos (D) (Adaptado de Velasquez *et al.*, 2011).....7

1. Introdução

1.1. Paredes celulares de plantas

Paredes celulares são estruturas resistentes, mas relativamente flexíveis, que envolvem as células de plantas, fungos e da maior parte dos procaríotos, tendo como funções principais o suporte estrutural e proteção, além de serem interfaces de troca seletiva e comunicação intercelular (Popper *et al.*, 2011). Em plantas, as paredes celulares são estruturas complexas e dinâmicas, compostas majoritariamente por polissacarídeos de alto peso molecular e proteínas altamente glicosiladas (Cosgrove, 2005; Sommerville *et al.*, 2004). Eventos durante o crescimento e desenvolvimento das plantas, sinalização em resposta a estímulos do ambiente e exposição a fatores indutores de estresses bióticos e abióticos aumentam a variação de estrutura e composição das paredes celulares (Showalter, 1993).

Quando as células vegetais ainda estão expandindo, uma matriz extracelular flexível é depositada, sendo conhecida como parede celular primária. Em alguns casos, quando a expansão das células cessa, uma camada adicional com organização e composição distinta é depositada junto à parede celular primária, aumentando sua resistência, sendo assim chamada de parede celular secundária (Cosgrove, 2005).

A parede celular primária (Figura 1) tem uma estrutura semelhante à fibra de vidro (Carpita *et al.*, 1993; O'Neill *et al.*, 2003), com microfibrilas de celulose que estão envoltas em uma matriz complexa de polissacarídeos de duas classes, as hemiceluloses e as pectinas. As hemiceluloses são polissacarídeos que ligam diretamente à celulose, formando uma rede muito forte, mas resiliente. Pectinas são polissacarídeos complexos com funções diversas (Willats *et al.*, 2001), que afastam ligeiramente as microfibrilas de celulose facilitando, assim, a expansão lateral durante o crescimento, mas também fixando as fibrilas no local correto quando o crescimento cessa, influenciando na porosidade da parede (Vincken *et al.*, 2003). Ainda, as pectinas são os maiores componentes da lamela média, uma camada adesiva que ajuda a ligar as paredes celulares de células vegetais adjacentes (Varner *et al.*, 1989).

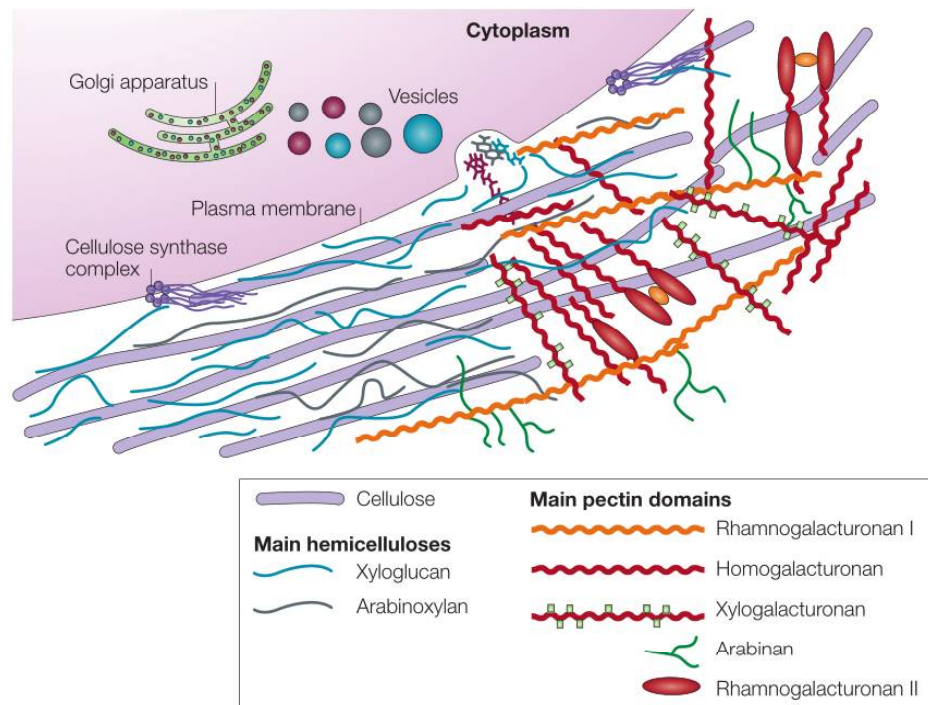


Figura 1. Modelo para a estrutura da parede celular primária. As microfibrilas de celulose e os polissacarídeos da matriz tem vias de síntese distintas: a celulose é sintetizada por complexos hexaméricos na membrana plasmática; as hemiceluloses e pectinas são sintetizadas no complexo de Golgi e transportadas por vesículas para a parede celular. As hemiceluloses formam uma rede resistente junto das microfibrilas, enquanto as diversas pectinas estão dispostas dentro desta rede, influenciando assim na porosidade e disposição da celulose, e também criando uma camada adesiva mais externa, a lamela média, que ajuda a unir as células vegetais adjacentes (Adaptado de Cosgrove, 2005; Varner *et al.*, 1989).

Além dos polissacarídeos de alto peso molecular, as paredes celulares também contém em grande quantidade glicoproteínas ricas em hidroxiprolina (HRGPs), uma superfamília que engloba as extensinas (EXTs), proteínas ricas em prolina (PRPs) e proteínas arabinogalactânicas (AGPs) (Xu *et al.*, 2008; Kieliszewski *et al.*, 1994; Showalter *et al.*, 2010). Estas proteínas sofrem extensas modificações pós-traducionais, destacadamente a modificação de resíduos de prolina (Pro) para hidroxiprolina (Hyp) por prolil-4-hidroxilases ligadas à membrana (Tiainen *et al.*, 2005). As HRGPs nascentes são O-glicosiladas com cadeias ricas em resíduos de arabinose ou galactose por glicosiltransferases no complexo de Golgi e no retículo endoplasmático (Yuasa *et al.*, 2005; Petersen *et al.*, 2011) e sofrem ligações

cruzadas na parede celular, por peroxidases, através de resíduos alternados de tirosina para formar uma rede covalente (Held *et al.*, 2004; Cannon *et al.*, 2008).

1.1.1. Extensinas

As extensinas constituem-se em componentes proteicos majoritários da parede celular primária de plantas. Com estruturas repetitivas e aparência de hastes sinuosas na microscopia eletrônica (Varner *et al.*, 1989), as extensinas são ricas em resíduos de hidroxiprolina e serina, e, em menor grau, lisina (sendo assim básicas, com ponto isoelétrico de cerca de 10 pelo alto conteúdo deste aminoácido), valina, tirosina e histidina, usualmente contendo o motivo pentapeptídico S-Hyp₄ (ou SPPPP, antes da oxidação dos resíduos de prolina por prolyl-4-hidroxilases), que se repete ao longo da proteína (Showalter, 1989). A maior parte dos resíduos de hidroxiprolina são O-glicosilados, com um a quatro resíduos de arabinose, enquanto os resíduos de serina podem ser ligados a uma única unidade de galactose (Stafstrom *et al.*, 1986). Monômeros de extensinas são secretados e depositados na parede celular, através da formação de ligações cruzadas isoditrosina entre as moléculas, gerando uma matriz rígida e altamente insolúvel, que fortalece as paredes celulares (Held *et al.*, 2004).

As extensinas assumem uma estrutura chamada hélice poliprolina II, um tipo de estrutura secundária comumente encontrada em proteínas fibrilares, que não apresenta ligações de hidrogênio intracadeia e é mais flexível que a hélice α (Adzhubei *et al.*, 2013). A hélice poliprolina II é estabilizada pela O-glicosilação (Velasquez *et al.*, 2011), possivelmente por causa das pontes de hidrogênio entre os açúcares que são adicionados (Stafstrom *et al.*, 1986), sendo o processo de glicosilação, assim, essencial para a manutenção da função normal destas proteínas e alongamento da parede celular.

1.1.2. Prolil-4-hidroxilases

Prolil-4-hidroxilases (P4Hs) são enzimas-chave na síntese dos componentes das paredes celulares. As P4Hs são 2-oxoglutarato dioxigenases (EC 1.14.11.2) que catalisam a síntese de 4-hidroxiprolina (Hyp), um aminoácido produzido pela adição de um grupo hidroxila no átomo de carbono C₄ do anel pirrolidina do resíduo de prolina (Figura 2).

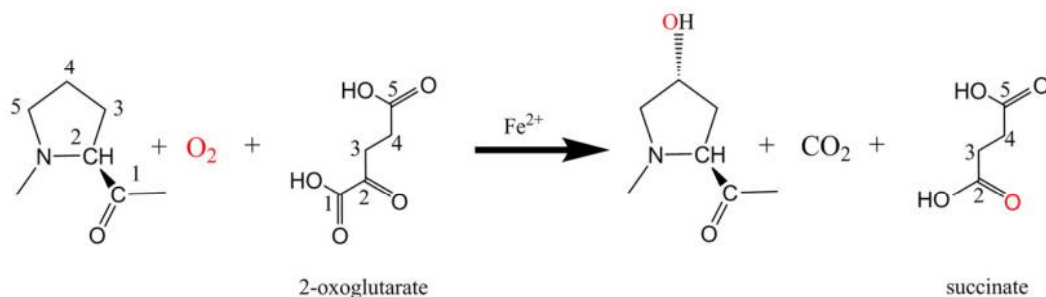


Figura 2. Reação catalisada pelas P4Hs. A molécula de 2-oxoglutarato (2OG) coordena o íon ferro através dos átomos de oxigênio de C₁-carboxilato e de C₂-ceto. O produto da reação é um resíduo trans-4-hidroxirolina. Os dois átomos de oxigênio (marcados em vermelho) são usados para a hidroxilação do resíduo de prolina e para a oxidação do 2OG, que é descarboxilado pela liberação do grupo C₁-carboxilato na forma de CO₂ (Reproduzido de Koski *et al.*, 2009).

Além da relevância no contexto da parede celular de plantas, a formação de Hyp também é importante em outras proteínas de diversos organismos, como no colágeno em vários filos de animais (onde os resíduos de Hyp são necessários para a formação de uma estrutura de hélice tripla estável, mas não são glicosilados) e na família de fatores de transcrição induzíveis por hipóxia (Gorres *et al.*, 2010).

Em geral, as P4Hs de plantas são menores em tamanho, de cerca de 30 a 60 kDa, em comparação com as P4Hs que catalisam a oxidação de Pro no colágeno (C-P4Hs), que são tetrâmeros de subunidades $\alpha_2\beta_2$ (as subunidades α são responsáveis pela hidroxilação e as subunidades β são idênticas à proteína dissulfeto isomerase). Em contraste, as P4Hs de plantas e também as HIF-P4Hs (que hidroxilam os fatores de transcrição induzíveis por hipóxia) são monômeros (Figura 3).

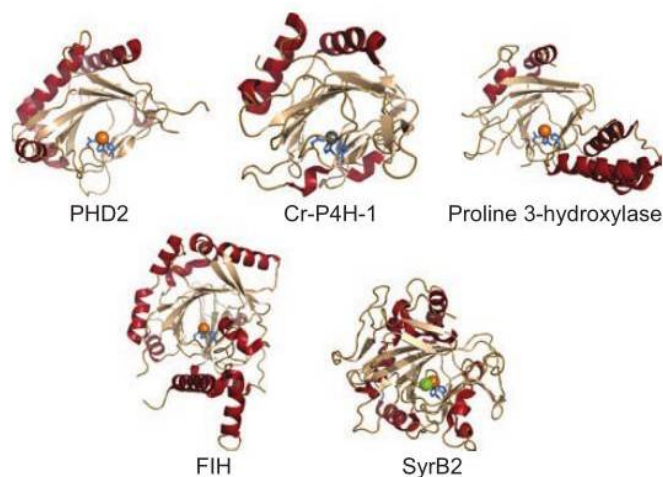


Figura 3. Estruturas cristalográficas de três diferentes dioxigenases dependentes de 2OG e duas proteínas relacionadas. A primeira estrutura corresponde a uma HIF-P4H de *Homo sapiens* (PDB ID 2G1M), a segunda a uma P4H da alga *Chlamydomonas reinhardtii* (PDB ID 3GZE) e a terceira a uma prolil-3-hidroxilase de *Streptomyces sp* (PDB ID 1E5S). As duas estruturas na parte de baixo da figura (PDB IDs 1H2N e 2FCT) correspondem respectivamente à asparaginil hidroxilase de *Homo sapiens* e à halogenase SyrB2 de *Pseudomonas syringae*, que compartilham com as P4Hs alguns motivos estruturais, como as hélices marcadas em vermelho e folhas marcadas em bege, e os resíduos que coordenam o metal cofator nos centros catalíticos (Reproduzido de Gorres *et al.*, 2010).

As P4Hs de plantas tem cerca de 30% de identidade com o domínio catalítico α das C-P4Hs, e, assim como as C-P4Hs, localizam-se no lúmen do retículo endoplasmático, onde hidroxilam polipeptídeos ricos em prolina (como a repetição SPPPP presente em extensinas). Diferentemente das P4Hs de plantas, as HIF-P4Hs são enzimas citoplasmáticas e nucleares, que agem em motivos LXXLAP, não em repetições poliprolina (Gorres *et al.*, 2010). Apesar de certas semelhanças estruturais, como o motivo *jellyroll*, composto por folhas β (Koski *et al.*, 2007), as P4Hs de plantas e as C-P4Hs diferem substancialmente nas suas especificidades aos substratos. P4Hs de plantas conseguem hidroxilar peptídeos poliprolina que são inibidores competitivos das C-P4Hs, mas peptídeos que mimetizam diretamente o colágeno, com sequências $(XPG)_n$ são hidroxilados, em geral, ineficientemente (Jaakkola *et al.*, 2001; Myllyharju *et al.*, 2008).

Em *Arabidopsis thaliana* existem pelo menos 13 tipos diferentes de P4Hs com padrões de expressão distintos, de maneira tecido-específica. P4H5, P4H2 e P4H13 são os tipos mais expressos nos pelos radiculares (Birnbaum *et al.*, 2003) (Figura 4), estruturas unicelulares que apresentam um papel importante para a absorção de nutrientes, e representam um bom modelo para o estudo de fenômenos da parede celular, pois a morfologia do seu crescimento é facilmente observável em microscopia de luz (Velasquez *et al.*, 2011). P4H5 e P4H2 são expressas principalmente na zona morfogênica dos pelos radiculares, em tricoblastos, assim como P4H13, que também é expressa nas raízes, em tricoblastos e atricoblastos (Brady *et al.*, 2007; Velasquez *et al.*, 2011).

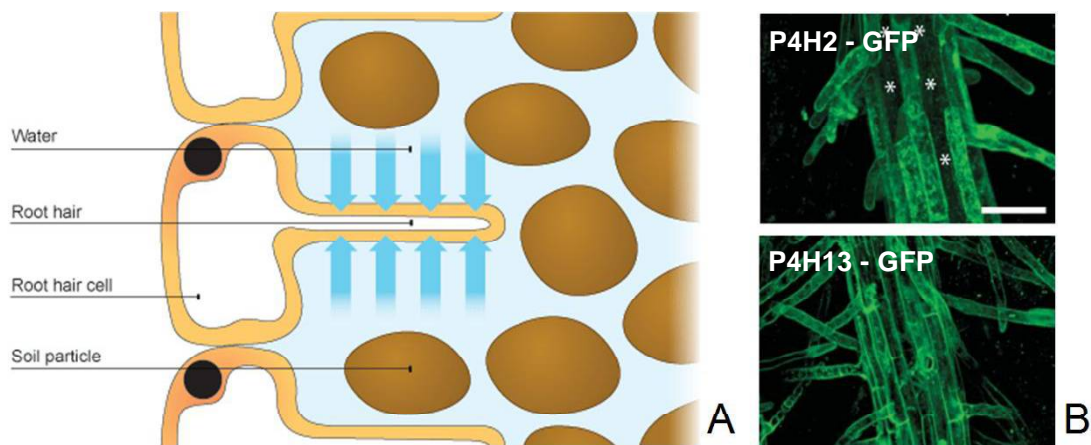


Figura 4. Representação da estrutura dos pelos radiculares, projeções da parede celular que auxiliam na absorção de água e nutrientes nas plantas (A) e registros de microscopia de fluorescência evidenciando a expressão de P4Hs, marcadas com GFP, nos pelos radiculares em *A. thaliana* (B). Adaptado de Velasquez *et al.*, 2011.

A exposição das raízes a inibidores como α -dipiridil e etil-3,4-dihidroxibenzoato bloqueia a hidroxilação dos resíduos de prolina das HRGPs (Chrispeels, 1970; Estévez *et al.*, 2006) e inibe o alongamento dos pelos radiculares (Velasquez *et al.*, 2011). Mutantes de P4H5, P4H2 e P4H13 geram fenótipos de pelos radiculares curtos (Figura 5), evidenciando assim que a correta hidroxilação de prolina, necessária para criar sítios de glicosilação nas HRGPs posteriormente, é essencial para o alongamento dos pelos radiculares (Velasquez *et al.*, 2011).

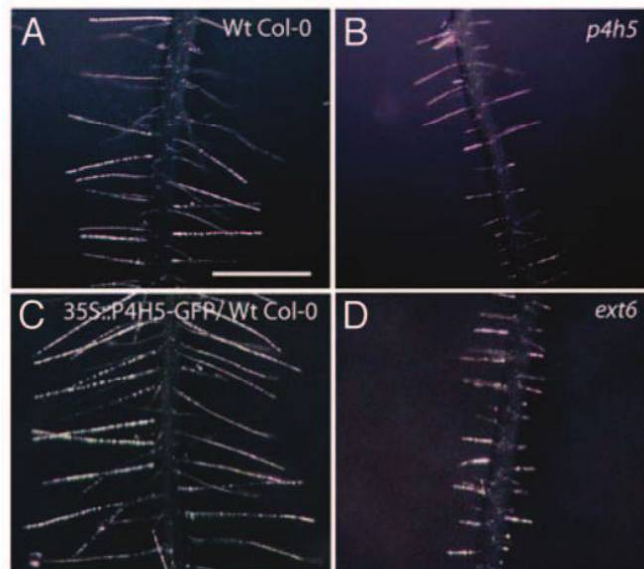


Figura 5. Diferentes fenótipos de pelo radicular em *A. thaliana*. Em (A), o fenótipo selvagem, em (B) o fenótipo de pelos radiculares curtos, resultante de mutações na enzima P4H5. A superexpressão de P4H5 leva à formação de pelos radiculares mais longos (C), enquanto mutações em um tipo de extensina, alvo de P4H5, leva à formação de pelos radiculares curtos (D) (Adaptado de Velasquez *et al.*, 2011).

1.2. Caracterização de biomoléculas por modelagem molecular

Estudos computacionais envolvendo modelagem molecular tem uso muito difundido, não apenas no estudo de sistemas biológicos, mas também em várias outras áreas dentro da química e ciência dos materiais (Kollman, 1996). A modelagem molecular abrange todos os métodos teóricos e técnicas computacionais usadas para mimetizar o comportamento de moléculas, sejam elas sistemas pequenos, de poucos átomos, até moléculas biológicas grandes e complexos de materiais (Karplus *et al.*, 1990). Modelos moleculares podem servir para visualizar hipóteses (como, por exemplo, a orientação de uma molécula de um fármaco em relação a um alvo ou a interação entre duas proteínas), e, conseqüentemente, para elaborar mais testes desta hipótese e, possivelmente, desenvolver novas moléculas a partir destes experimentos ou entender fenômenos complexos, quantificando dados relativos à diversidade estrutural, potencial eletrostático, entre outros (Broughton, 1997).

Dentre as abordagens computacionais para o estudo de biomoléculas está a dinâmica molecular (DM), que oferece informações sobre as sucessivas

configurações de um determinado sistema molecular através de uma trajetória que é gerada baseando-se nas equações de movimento de Newton (Leach *et al.*, 2007). Para que simulações de DM possam ser realizadas, é preciso possuir, primeiramente, um modelo da estrutura tridimensional da molécula de interesse, que pode ser obtido diretamente através de experimentos (fazendo uso de técnicas como a cristalografia de raios X ou RMN) ou por técnicas computacionais, como a modelagem comparativa, que também depende de estruturas obtidas experimentalmente e depositadas em bancos de dados (Capriles *et al.*, 2014).

1.2.1. Modelagem comparativa

No método de modelagem comparativa, também chamada de modelagem por homologia, a proteína de interesse tem sua estrutura tridimensional predita usando como referência a estrutura de outra proteína similar (também chamada de molde, e na maioria das vezes evolutivamente relacionada). Essa proteína similar tem de possuir estrutura tridimensional resolvida experimentalmente, e as coordenadas cartesianas de seus átomos devem estar depositadas em banco de dados de estruturas como o *Protein Data Bank*. A modelagem comparativa é o método empregado mais frequentemente, e seu limite de predição está intrinsecamente relacionado com o grau de similaridade entre as estruturas alvo e molde. Geralmente, considera-se como limites mínimos de aplicabilidade do método valores de 25 a 30% de identidade, obtidos através do alinhamento entre a estrutura primária da proteína alvo e de uma ou mais proteínas molde (Capriles *et al.*, 2014).

Após a construção do modelo, é necessário identificar possíveis erros relacionados aos métodos empregados, à escolha das referências e ao alinhamento entre as sequências alvo e molde. Caso o modelo seja caracterizado como de má qualidade, todo o protocolo anterior deve ser revisto no intuito de se melhorar o alinhamento, escolher outros moldes ou até mesmo decidir-se pelo uso de outros métodos (Capriles *et al.*, 2014). Um dos critérios de validação para modelagem comparativa é o gráfico de Ramachandran, que fornece uma distribuição dos ângulos torcionais de uma proteína, assim como uma visão global das regiões “permitidas” ou “não permitidas” dos valores dos ângulos torcionais, servindo assim como um importante indicador da qualidade do modelo tridimensional (Ramachandran *et al.*, 1963). Além das análises relacionadas aos ângulos torcionais,

também podem ser aplicados para a validação métodos baseados em minimização de funções de energia, que, com a análise de valores normalizados das funções, ajudam a avaliar o quão próximo, pelo menos estatisticamente, o modelo gerado está de proteínas que possuem um mesmo perfil molecular ou até o mesmo tipo de enovelamento (Capriles *et al.*, 2014).

1.2.2. Simulações de dinâmica molecular

Simulações de dinâmica molecular revelam a natureza dos movimentos atômicos das moléculas, resolvendo as equações de movimento de Newton para os átomos do sistema e resultando em uma trajetória que define como as posições destes átomos variam no tempo (Leach *et al.*, 2007). O conjunto de equações descrevendo as forças que agem sobre os átomos da estrutura é chamado de campo de força e, por sua base na mecânica molecular, ignora os elétrons do sistema, considerando somente a posição dos núcleos dos átomos (Kollman, 1987; Broughton, 1997).

Na DM, são inicialmente atribuídos aos átomos do sistema movimentos aleatórios com velocidades que são uma característica coletiva a uma dada temperatura. Então, depois de um primeiro passo de tempo definido (usualmente 1 ou 2 fs), os efeitos somados de várias forças interatômicas que agem no sistema (aquelas devidas à deformação de ligações covalentes, ângulos de ligação e torcionais, assim como interações não covalentes) são calculados de acordo com as equações de Newton para o movimento. Uma vez que todos os átomos no sistema tenham se movido depois desse espaço de tempo (por uma distância que é somente uma pequena fração do comprimento da ligação), as forças interatômicas em cada átomo terão mudado da mesma forma, embora por apenas um pequeno valor. Então, usando estes valores alterados das forças interatômicas, junto com as novas posições e velocidades dos átomos, o cálculo é repetido para qualquer tempo adicional (Karplus *et al.*, 1990). Este processo produz assim um registro de posições e velocidades de todos os átomos do sistema no período de tempo, reproduzindo assim, muitas vezes com elevada fidelidade, o comportamento de moléculas de interesse biológico ou tecnológico.

Para interpretação dos dados gerados pelas simulações de DM, existem vários tipos de análise, que são empregados dependendo da natureza do problema

em estudo. Para análise do comportamento da estrutura proteica ao longo da simulação, as ferramentas mais comumente empregadas incluem o RMSD (*root mean square deviation*), que indica o quanto a estrutura se modifica durante a simulação, comparando com a estrutura inicial), o RMSF (*root mean square fluctuation*), o raio de giro, distâncias entre átomos ou grupamentos e a evolução do conteúdo de estrutura secundária em função do tempo (Verli, 2014).

2. Objetivos

A partir do exposto, o presente trabalho visa a contribuir para o entendimento da biologia estrutural de P4Hs de *A. thaliana*, particularmente sua interação com as EXTs, proteínas ricas em resíduos de prolina. Para tanto, os seguintes objetivos específicos foram traçados:

- Construção e validação de modelos da enzima P4H5 de *A. thaliana*, através de modelagem comparativa;
- Caracterização do comportamento conformacional da enzima P4H5 de *A. thaliana* ligada a um peptídeo com repetições PPPP que representa o substrato EXT, através de simulações por dinâmica molecular;
- Caracterização do comportamento conformacional da enzima livre, através de simulações por dinâmica molecular.

3. Metodologia

Diversas metodologias de modelagem molecular foram utilizadas no presente trabalho, para a construção dos modelos tridimensionais e realização das simulações de dinâmica molecular. Os procedimentos empregados estão descritos em detalhes a seguir.

3.1. Construção dos modelos tridimensionais

Por modelagem comparativa, foram construídos modelos tridimensionais da enzima P4H5, utilizando o programa MODELLER, versão 9.10 (Sali *et al.*, 1993). Foi usada como molde a estrutura da enzima P4H1 de *Chlamydomonas reinhardtii*, depositada no *Protein Data Bank* com o código 3GZE, que teve 54% de identidade de sequência com a enzima P4H5, segundo a busca do algoritmo BLAST (Altschul *et al.*, 1997). O modelo gerado pelo programa MODELLER foi validado usando o servidor PROCHECK (Laskowski *et al.*, 1993), que, entre outras análises, oferece informações sobre a torção dos ângulos de diedros dos resíduos através do gráfico de Ramachandran.

A partir do modelo validado, foram construídos outros quatro modelos, acoplando-se um peptídeo de 25 resíduos com repetições de prolina (de sequência SPPPPYVYSSPPPPYVSPSPKVYYK) que simula a extensina, obtido de trabalhos anteriores (Velasquez *et al.*, 2011). Em cada um destes modelos com acoplamento do peptídeo, um resíduo diferente de prolina do segundo *cluster* foi posicionado no sítio catalítico, utilizando a orientação de um substrato de prolina presente originalmente no cristal e alinhando-se as estruturas no programa SPDBV (Guex *et al.*, 1997).

Para visualização de estruturas e construção de imagens, foram usados os programas Pymol (Delano, 2002) e VMD (Humphrey *et al.*, 1996).

3.2. Simulações por dinâmica molecular

Para as simulações por dinâmica molecular, foi usado o pacote de simulações GROMACS, versão 4.5.4 (Hess *et al.*, 2009). Foram feitas simulações da enzima P4H5 livre e quatro sistemas da enzima P4H5 com o peptídeo EXT acoplado, no campo de força GROMOS96 43A1 (Scott *et al.*, 1999), com replicatas. Para mimetização do ambiente fisiológico onde os processos ocorrem, os sistemas foram

solvatados com água, utilizando o modelo *single point charge* (SPC –Berendsen *et al.*, 1987) e contra-íons (Na^+) foram adicionados para neutralização de cargas. As simulações foram feitas em caixas triclinicas com condições periódicas de contorno, temperatura de 310K e pressão de 1 bar, utilizando constantes de acoplamento de, respectivamente, $\tau = 0,1$ ps e $\tau = 0,5$ (Berendsen *et al.*, 1984). Ainda, o método LINCS (Hess *et al.*, 1997) foi aplicado na restrição de ligações covalentes, de forma a permitir um passo de integração de 2 fs, enquanto que interações eletrostáticas foram calculadas utilizando o método Particle-Mesh Ewald (Darden *et al.*, 1993).

Utilizando ainda o pacote de simulação GROMACS, foram feitas análises diversas das trajetórias geradas pela simulação de dinâmica molecular, a fim de descobrir informações precisas sobre a estabilidade da enzima no tempo de simulação (comparando os dados de modelos com substrato acoplado e da proteína livre), variações de posição de segmentos da proteína e distâncias entre os átomos do sítio ativo.

4. Resultados

A revista escolhida para submissão deste trabalho, sob o formato de *Research Article*, foi a *PROTEINS: Structure, Function, and Bioinformatics*. A revista *Proteins* tem publicação mensal (tendo sua primeira edição em 1986), em língua inglesa, pelo grupo editorial John Wiley & Sons. A revista publica artigos originais de pesquisa com dados experimentais e analíticos significantes para todas as áreas em pesquisa de proteínas, incluindo estrutura, função, computação, genética e *design*. O periódico tem foco na publicação de trabalhos que apresentem novas estratégias experimentais e computacionais para interpretar e entender dados de química biofísica, estudos estruturais de proteínas e complexos macromoleculares, alterações da estrutura e função de proteínas através de técnicas de biologia molecular e genética, análises funcionais sob condições fisiológicas, assim como interações de proteínas com receptores, ácidos nucleicos ou outros substratos e ligantes.

Este periódico teve índice de impacto, em 2014, de 2,921 e esteve nos rankings 32/74 (Biofísica) e 139/291 (Bioquímica e Biologia Molecular) do ISI *Journal Citation Reports* © em 2013. Além dos artigos de pesquisa completos, também são aceitas pequenas comunicações, *prediction reports* e algumas revisões. As normas para formatação neste periódico encontram-se no Anexo 1, página 38.

Structural Biology of Proline-rich Substrates Recognition by *A. thaliana* Prolyl-4-hydroxylase 5

Structural Biology of *A. thaliana* P4H5

Keywords: molecular dynamics, plant cell wall proteins, extensin

Elisa B. O. John¹, Laercio Pol-Fachin^{1,2}, Jose M. Estevez³ and Hugo Verli^{1*}.

¹Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av Bento Gonçalves
9500, CP 15005, Porto Alegre 91500-970, RS, Brazil

²Dep. de Química Fundamental, UFPE, PE, Brazil

³Facultad de Ciencias Exactas y Naturales, UBA, BA, Argentina

*Corresponding Author – Address: Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av Bento Gonçalves 9500, CP 15005, Porto Alegre 91500-970, RS, Brazil. TEL.: +55 51 3308 7770; fax: +55 51 3308 7309. Email address: hverli@cbiot.ufrgs.br

[Abstract]

Plant cell walls are complex and dynamic structures, composed mostly of polysaccharides and glycoproteins. Hydroxyproline-rich glycoproteins are found in cell walls, and are related to plant growth and development. Prolyl-4-hydroxylases (P4H) catalyze the posttranslational formation of hydroxyproline, being essential enzymes to plant growth, as proline hydroxylation defines the subsequent O-glycosylation sites in extensins (EXT), a group of cell wall glycoproteins. In *A. thaliana*, while at least 13 different P4Hs with distinct expression patterns exist, P4H5, P4H2, and P4H13 seem to be the most expressed in plant root hairs, which, in turn, play an important role in nutrient absorption. In spite of their importance for plant development, the recognition, at the atomic level, between these enzymes and their proline substrate is not yet completely understood. Therefore, in order to contribute in the understanding of the structural biology of P4Hs interaction with EXTs, the current work intends to characterize the molecular recognition between *A. thaliana* P4H5 and an EXT repeating peptide. The molecular dynamics calculations and analyses were performed using GROMACS simulation suite and GROMOS43A1 force field. The obtained results indicate that the attachment of EXT to P4H5 stabilizes the enzyme structure, as observed by RMSD values. Further analyses also indicate that some proline residues in the EXT peptide have precedence over others during the catalytic process by P4H5. P4H5 seems to have a differential behavior with the sequential proline residues in extensins. These findings are also expected to contribute in the understanding of the hydroxyproline formation by P4Hs in EXT.

[Introduction]

Cell walls are resistant yet flexible structures that surround the cells of plants, fungi and most prokaryotes, and their major function is to provide structural support and protection, as well as being selective exchange and intercellular communication interfaces¹. In plants, the cell walls are complex and dynamic structures, composed mostly of high-molecular-weight polysaccharides and highly glycosylated proteins^{2,3,4}. Plant cell walls contain a large amount of hydroxyproline-rich glycoproteins (HRGPs), proteins that are related to plant growth and development and constitute a superfamily which includes the extensins (EXTs), proline rich proteins (PRPs) and arabinogalactan proteins (AGPs)⁵. In higher plants, extensins are particularly abundant in the primary cell wall and are generally characterized by being rich in serine and hydroxyproline, and some combination of amino acids lysine (making them basic proteins, possibly interacting with acidic pectic blocks in the cell wall⁶), valine, tyrosine, and histidine^{1,7}. Extensin monomers are secreted and insolubilized in the cell wall, forming a rigid matrix of cross-linked tyrosine residues, which the main function is to strengthen the cell walls^{4,7}. Extensins usually contain the pentapeptide motif Ser-Hyp₄, often within the context of other, larger repeating motifs. Most of hydroxyproline residues are O-glycosylated by glycosyltransferases in the Golgi apparatus and in the endoplasmic reticulum, with one to four arabinosyl residues, and some of the serine residues are glycosylated with a single unit of galactose. O-glycosylation of extensins stabilizes the polyproline II helix structure, because of the hydrogen bonds between the sugars that are added, and, therefore, it is essential for the maintenance of normal function of these proteins during the growth of the cell walls^{4,8}. The HRGPs undergo extensive post-translational modifications which include the modification of proline residues to hydroxyproline by prolyl-4-hydroxylases (P4Hs) bound to the cell membrane. The P4Hs are 2-oxoglutarate dioxygenases (EC 1.14.11.2) that catalyze the addition of a hydroxyl group on the C4 carbon atom of the pyrrolidine ring on the proline

residue⁹. Besides their importance in the cell wall of plants, the P4Hs are relevant enzymes present in various organisms, such as in some animal phyla, which have collagen-P4Hs and HIF-P4Hs. Hyp residues are required for the formation of a stable triple helix structure in collagen, but are not glycosylated, while P4Hs hydroxylate hypoxia-inducible transcription factors^{10,11}. In *Arabidopsis thaliana*, there are at least 13 different types of P4Hs with distinct patterns of expression, in a tissue-specific manner⁴. P4H5, P4H2 and P4H13 are more expressed in the root hairs, which are unicellular structures that have an important role in nutrient and water absorption⁵, and represent a good model for the study of cell wall phenomena, since their growth morphology is easily observed under light microscopy. Mutant P4H5, P4H2 and P4H13 generate short root hairs phenotypes, thus showing that proper hydroxylation of proline, necessary to create glycosylation sites in HRGPs, is essential for the elongation of the root hairs. P4H5 overexpression leads to the formation of longer root hairs⁴. In spite of the importance of P4Hs to the correct growth of root hairs, as evidenced by mutation experiments and inhibition of these enzymes, their interaction at the molecular level with the extensins is not yet completely understood, and molecular modeling studies of these enzymes and their substrates contribute to the structural biology of this interaction, thus revealing more information about the development and growth of plants. Therefore, in order to contribute in the understanding of the structural biology of P4Hs interaction with EXTs, the current work intends to characterize the molecular recognition between *A. thaliana* P4H5 and an EXT repeating peptide.

[Methods]

Homology modeling and system preparation

The input models were based on the crystallographic structure of a prolyl-4-hydroxylase from *Chlamydomonas reinhardtii* (PDB ID code 3GZE), which had 54% of sequence identity with

the P4H5 from *A. thaliana*, by pdb-BLAST¹². The protein modeling was done using MODELLER¹³, version 9.10, and the model was validated using the PROCHECK¹⁴ server tools. Once the first P4H5 model was obtained, other four models were built attaching a 25 residue peptide (sequence SPPPPYVYSSPPPPYSPSPKVYYK), obtained from previous works⁴, in order to simulate the EXT substrate of P4H5. These models were built in SPDBV¹⁵, and in each model a different proline residue from the first cluster was placed in the catalytic site, based on the orientation of a small Ser-Pro substrate that was present in the crystallographic structure.

Molecular Dynamics Simulations

The molecular dynamics calculations and analyses were performed using GROMACS simulation suite, version 4.5.4¹⁶ and GROMOS96 43A1 force field¹⁷. The systems were simulated during 100 ns, within triclinic boxes solvated with water molecules in the SPC model¹⁸ and counter-ions (Na^+) to neutralize the systems, under a constant temperature of 310K and pressure of 1 bar¹⁹ (the EXT peptides were considered as periodic during simulations, in which Ser1 was attached to Lys25 through the specular images). Long range electrostatic interactions were calculated using the reaction field method. Semi-isotropic pressure coupling was performed, thus only in XY axes.

[Results and Discussion]

In order to evaluate the effects of EXT peptide binding to P4H5 and to distinguish these systems from the P4H5 without any substrate complexed, five systems were simulated: four of them with the EXT peptide complexed to P4H5 in sequential positions (each one with one proline residue from the cluster Pro11-Pro12-Pro13-Pro14 aligned to its catalytic site) and one of the noncomplexed enzyme. The free enzyme has a more flexible structure during the

simulations, comparing to the models with the EXT peptide attached, as presented in Figure 1, by RMSD (root mean square deviation) analysis. This indicates that the attachment of the substrate stabilizes the P4H5 structure. In RMSF (root mean square fluctuation) analysis of the models with the EXT peptide attached (Figure 2), two regions of higher flexibility can be distinguished. These regions are loops around the catalytic site, covering the substrate cavity. The regions with higher RMS values are also discriminated in the RMSF analysis of the P4H5 without any peptide attached. However, in this model the loops do not cover the enzyme cavity during the simulation, as presented in Figure 3, indicating that these loops may act as flaps which control the EXT access to P4H5 substrate cavity.

To verify the interaction between P4H5 and the EXT peptide during the MD simulations, several distance measurements were performed. In the measure of the distance between the center of mass of the residues from the catalytic site and the γ -carbon from the aligned proline residue (Figure 4B), within two models this distance was maintained and in the remaining models the distance increased in the course of the simulations. This may indicate that some positions of the EXT peptide have higher affinity with the enzyme cavity residues, and therefore, some proline residues in the EXT peptide may have precedence over others during the catalytic process. The distance between the enzyme cofactor ion and the residues from catalytic site was also measured (Figure 4C), to verify the coordination of the metal ion during the MD simulations. In three of the models, this distance increased in some point of the simulations, and it may be caused by some force field parameters that don't describe well that interaction of metal coordination with the catalytic site.

[Conclusions]

The obtained results indicate that the attachment of EXT to P4H5 stabilizes the enzyme structure, as observed by RMSD values. Further analyses also indicate that some proline residues in the EXT peptide may have precedence over others during the catalytic process by P4H5, once P4H5 seems to have a different behavior depending on which proline residue is aligned in its catalytic site. As observed by RMSF analyses, two groups of residues, located around the enzyme active site, showed high flexibility during the simulations, mainly in the absence of EXT peptide. Subsequently, these regions were identified as the main cause of RMSD differences between free and complexed P4H5, suggesting that EXT interaction has no major effects over P4H5 folded structure. The results show that two loops around EXT catalytic site may act as flaps, controlling substrate access to the enzymes active site. These findings are also expected to contribute in the understanding of the hydroxyproline formation by P4Hs in EXT.

[References]

1. Showalter AM. Structure and Function of Cell Wall Proteins. *The Plant Cell* 1993; 5: 9-23.
2. Cosgrove DJ. Growth of the plant cell wall. *Nature Reviews. Molecular Cell Biology* 2005; 6(11): 850–861.
3. Somerville C *et al.* Toward a systems approach to understanding plant cell walls. *Science* 2004; 306(5705): 2206–2211.
4. Velasquez SM *et al.* O-glycosylated cell wall proteins are essential in root hair growth. *Science* 2011; 332(6036): 1401–1403.
5. Velasquez SM, Iusem ND, Estevez JM. Root hair sweet growth. *Plant Signaling & Behavior* 2011; 6(10): 1600-1602.
6. Keller B. Structural cell wall proteins. *Plant Physiol.* 1993; 101: 1127-1130.
7. Kieliszewski MJ, Lamport DT. Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *The Plant Journal* 1994; 5: 157-172.
8. Stafstrom JP, Staehelin LA. The Role of Carbohydrate in Maintaining Extensin in an Extended Conformation. *Plant Physiol.* 1986; 81: 242-246.
9. Koski MK *et al.* The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *Journal of Biological Chemistry* 2009; 284(37): 25290–25301.
10. Gorres, KL, Raines RT. Prolyl 4-hydroxylase. *Critical Reviews in Biochemistry and Molecular Biology* 2010; 45(2): 106–124.

11. Myllyharju J. Prolyl 4-hydroxylases, key enzymes in the synthesis of collagens and regulation of the response to hypoxia, and their roles as treatment targets. *Annals of Medicine* 2008; 40(6): 402–417.
12. Altschul SF *et al.* Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research* 1997; 25(17): 3389–3402.
13. Sali A, Blundell TL.. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* 1993; 234(3): 779–815.
14. Laskowski RA, MacArthur MW, Moss DS and Thornton JM. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* 1993; 26(2): 283–91.
15. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An Environment for Comparative Protein Modeling. *Electrophoresis* 1997; 18(15): 2714–23.
16. Hess B *et al.* GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* 2008; 4(3): 435–447.
17. Scott WRP *et al.* The GROMOS Biomolecular Simulation Program Package. *The Journal of Physical Chemistry A* 1999, 103(19), 3596–3607.
18. Berendsen, HJC, Grigera JR, Straatsma TP. The Missing Term in Effective Pair Potentials. *Journal of Physical Chemistry* 1987; 91(24): 6269–6271.
19. Pol-Fachin L, Becker CF, Guimarães JA, Verli H. Effects of glycosylation on heparin binding and antithrombin activation by heparin. *Proteins: Structure, Function and Bioinformatics* 2011; 79(9): 2735–2745.

[Figure Legends]

Figure 1. Root mean square deviation (RMSD) analysis. The free enzyme (orange) has the highest RMSD values during the simulations, indicating that the attachment of EXT to P4H5 stabilizes the structure. In spite of the high RMSD, the uncomplexed P4H5 is not denaturing. The conformational modification is located at the regions depicted in Figure 3, and the average RMSD without these regions is 0.437.

Figure 2. In the structures, frames from the MD simulations are superimposed. Different proline residues are positioned to be hydroxylated (Pro11, black; Pro12, red; Pro13, green; Pro14, blue). Two groups of residues with high flexibility are marked on the structures and on the root mean square fluctuation (RMSF) analysis graph.

Figure 3. Frames from the MD simulations of a P4H5 without any peptide complexed. As seen on the bottom view of the structure (excerpt on the right), there is a loop with high flexibility that doesn't cover the enzyme cavity during the simulation. The same loop has high flexibility in the models with peptide attached, but it tends to be closer to the substrate cavity.

Figure 4. **A.** Schematic representation of the distances measured in (B) and (C). **B.** Measure of the distance between the center of mass of the residues from catalytic site and the γ -carbon from the proline residues of each model. **C.** Measure of the distance between the enzyme cofactor ion and the residues from catalytic site (first graph for His181, second for Asp183 and third for His267).

[Figures]

Figure 1

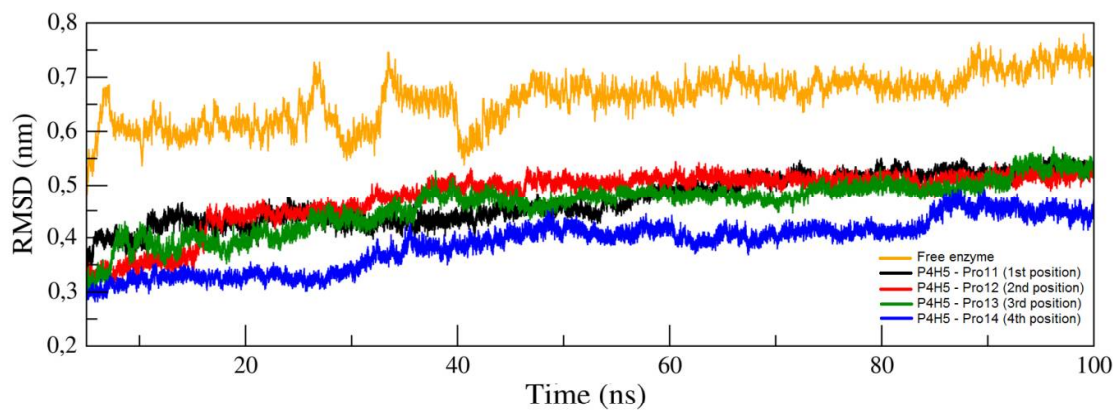


Figure 2

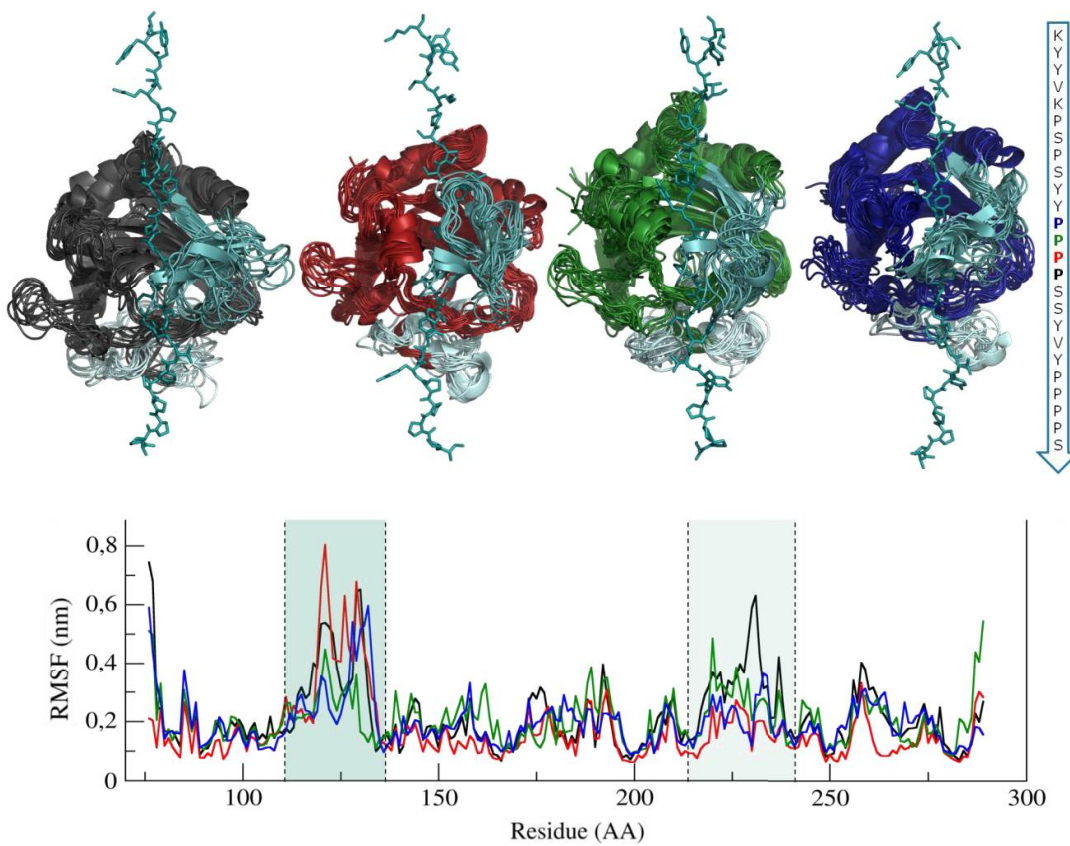


Figure 3

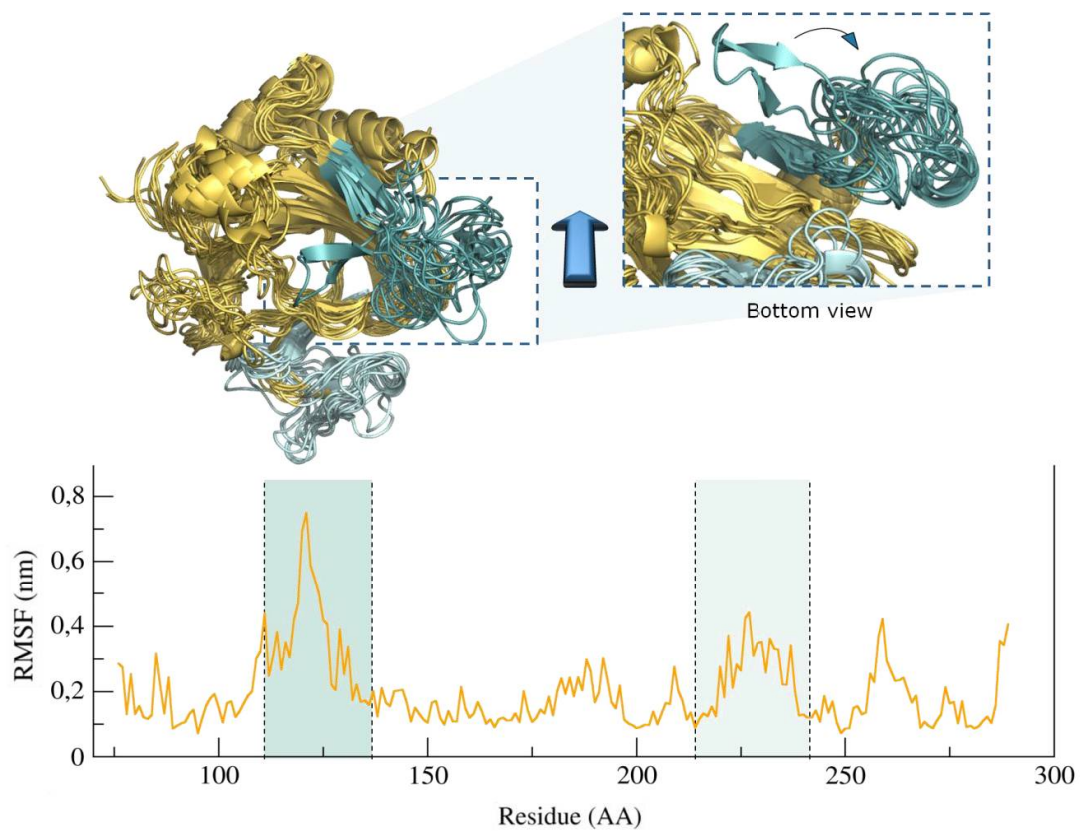
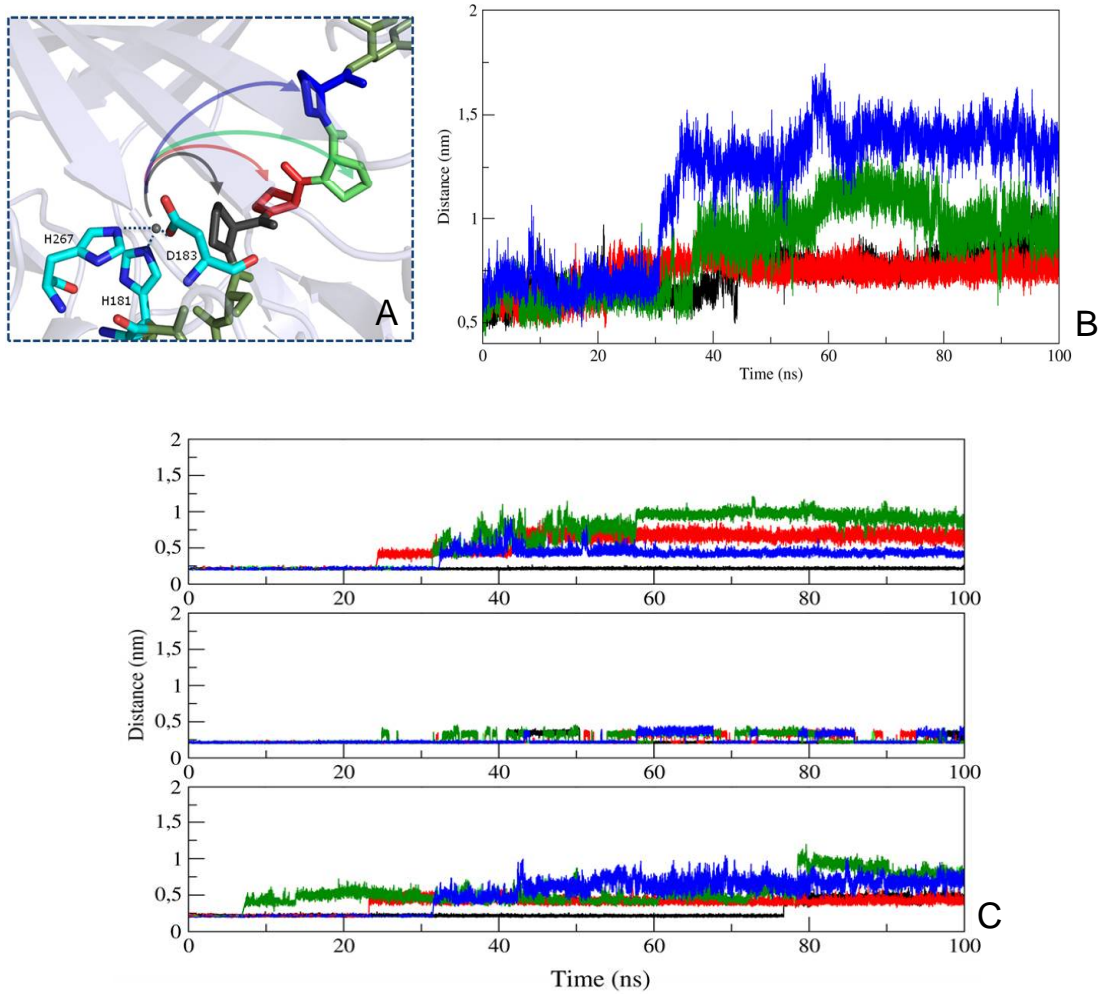


Figure 4



5. Conclusões

A partir da metodologia empregada, foram construídos cinco modelos da enzima P4H5 de *A. thaliana*, um da enzima livre e quatro modelos com um peptídeo que representa a extensina, com diferentes resíduos de prolina posicionados nos sítios catalíticos. Estes modelos foram validados com as ferramentas adequadas e preparados para simulações de DM.

A caracterização conformacional da enzima livre e da P4H5 acoplada com substrato foi feita através de simulações de DM e análises diversas das trajetórias obtidas, como análises de RMSD e RMSF (que indicam o quanto a estrutura da proteína se modifica ao longo da simulação, em relação à estrutura inicial) e várias medidas de distâncias.

Através da DM foi possível, então, observar diferenças entre os sistemas que continham a associação com o substrato e aqueles com a enzima livre, e os resultados obtidos indicam que o acoplamento do peptídeo EXT à P4H5 estabiliza a estrutura da enzima, como foi observado através dos valores da análise de RMSD. Através das análises de RMSF, foi possível destacar algumas regiões com maior mobilidade, que são dois *loops* ao redor do sítio catalítico que parecem agir como *flaps*, controlando o acesso do substrato ao sítio ativo da enzima. Ainda, a P4H5 parece ter um comportamento diferencial com os resíduos de prolina sequenciais nas extensinas, pois há diferença de flexibilidade nas estruturas dos modelos e diferenças entre os sistemas nas medidas de distâncias entre a enzima e peptídeo que foram realizadas.

Assim, espera-se que estes resultados contribuam para o entendimento da biologia estrutural do reconhecimento de extensinas por prolil-4-hidroxilases em *A. thaliana*.

6. Perspectivas

Considerados os resultados obtidos até então, as seguintes perspectivas podem ser traçadas para o trabalho:

- Expandir as análises sobre as simulações já feitas, adicionando dados sobre energia de ligação dos sistemas;
- Fazer simulações de dinâmica molecular com as enzimas P4H2 e P4H13, que também tem expressão relevante nos pelos radiculares;
- Fazer simulações semelhantes com P4Hs e substratos que já tenham Hyp em algumas posições;
- Fazer simulações com oligômeros de P4Hs, pois há evidências experimentais recentes de que estas enzimas formam dímeros.

7. Agradecimentos

Pesquisa desenvolvida junto ao centro Nacional de Supercomputação da Universidade Federal do Rio Grande do Sul.

Esta pesquisa recebeu auxílio da Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul, do Conselho Nacional de Desenvolvimento Científico e Tecnológico, MCT e da Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES), MEC, Brasília, DF, Brasil.

8. Bibliografia

ADZHUBEI, A. A.; STERNBERG, M. J. E.; MAKAROV, A. Polyproline-II helix in proteins: Structure and function. *Journal of Molecular Biology*, v. 425, p. 2100-2132, 2013.

ALTSCHUL, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, n. 25, p. 3389-3402, 1997.

BERENDSEN, H. J. C. *et al.* Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, v. 81, n. 8, p. 3684–3690, 1984.

BERENDSEN, H. J. C.; GRIGERA, J. R.; STRAATSMA, T. P. The Missing Term in Effective Pair Potentials. *Journal of Physical Chemistry*, v. 91, n. 24, p. 6269–6271, 1987.

BIRNBAUM, K. *et al.* A gene expression map of the Arabidopsis root. *Science*, v. 302, n. 5652, p. 1956–1960, 2003.

BRADY, S. M. *et al.* A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science*, v. 318, p. 801, 2007.

BROUGHTON, H. B. Molecular modeling. *Current opinion in chemical biology*, v. 1, n. 3, p. 392–398, 1997.

CANNON, M. C. *et al.* Self-assembly of the plant cell wall requires an extensin scaffold. *Proceedings of the National Academy of Sciences of the United States of America*, v. 105, n. 6, p. 2226–2231, 2008.

CAPRILES, P. V. *et al.* Modelos tridimensionais. In: VERLI, H. (Org.). *Bioinformática da Biologia à Flexibilidade Molecular*. Porto Alegre: 2014. p. 148-171

CARPITA, N. C.; GIBEAUT, D. M. Structural Models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* n. 3, p. 1-30, 1993.

CHRISPEELS, M. J. Synthesis and secretion of hydroxyproline-containing macromolecules in carrot. *Plant Physiol.*, v.45, p. 223-227, 1970.

COSGROVE, D. J. Growth of the plant cell wall. *Nature reviews. Molecular cell biology*, v. 6, n. 11, p. 850–861, 2005.

DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, v. 98, n. 12, p. 10089, 1993.

DELANO, W. L.: The PyMOL Molecular Graphics System. Schrödinger Sales Center, Portland, OR, USA, 2002. <http://www.pymol.org>, 2002.

ESTEVEZ, J. M. *et al.* Characterization of synthetic hydroxyproline-rich proteoglycans with arabinogalactan protein and extensin motifs in *Arabidopsis*. *Plant Physiol.*, v. 142, p. 458-470, 2006.

FRY, S. C. Polysaccharide-Modifying Enzymes in the Plant Cell Wall. *Annual Review of Plant Physiology and Plant Molecular Biology*, n.46, p.497-520, 1995.

GORRES, K. L.; RAINES, R. T. Prolyl 4-hydroxylase. *Critical reviews in biochemistry and molecular biology*, v. 45, n. 2, p. 106–124, 2010.

GUEx, N.; PEITSCH, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, v. 18, n. 15, p. 2714–2723, 1997.

HELD, A. M. *et al.* Di-isodityrosine is the intermolecular cross-link of isodityrosine-rich extensin analogs cross-linked in vitro. *Journal of Biological Chemistry*, n. 279, p. 55474-55482, 2004.

HESS, B. *et al.* GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, v. 4, n. 3, p. 435–447, 2008.

HESS, B. *et al.* LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, v. 18, n. 12, p. 1463–1472, 1997.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, v. 14, n. 1, p. 33–38, 1996.

JAAKKOLA, P. *et al.* Targeting of HIF- α to the von Hippel-Lindau ubiquitylation complex by O₂-regulated prolyl hydroxylation. *Science*, v. 292, n. 5516, p. 468–472, 2001.

KARPLUS, M.; PETSKO, G. A. Molecular dynamics simulations in biology. *Nature*, v. 347, p. 631-639, 1990.

KARR, A. L. Isolation of an enzyme system which will catalyze the glycosylation of extensin. *Plant physiology*, v. 50, n. 2, p. 275–282, 1972.

KELLER, B. Structural cell wall proteins. *Plant Physiol.*, n. 101, p. 1127-1130, 1993.

KIELISZEWSKI, M. J.; LAMPORT, D. T. Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *The Plant Journal*, n. 5, p. 157-172, 1994.

KOLLMAN, P. Molecular modeling. *Ann. Rev. Phys. Chem.*, n. 38, p. 303-316, 1987.

KOLLMAN, P. Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Accounts Chem. Res.* v. 29, p. 461-469, 1996.

KOSKI, M. K. *et al.* The active site of an algal prolyl 4-hydroxylase has a large structural plasticity. *Journal of Biological Chemistry*, v. 282, n. 51, p. 37112–37123, 2007.

KOSKI, M. K. *et al.* The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *Journal of Biological Chemistry*, v. 284, n. 37, p. 25290–25301, 2009.

LASKOWSKI, R. A. *et al.* PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, n. 26, p. 283-291, 1993.

LEACH, A. R.; GILLET, V. J. *An Introduction to Chemoinformatics*. Dordrecht, Springer, 2007.

MACKERELL, D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry. B*, v. 102, n. 18, p. 3586–616, 1998.

MYLLYHARJU, J. Prolyl 4-hydroxylases, key enzymes in the synthesis of collagens and regulation of the response to hypoxia, and their roles as treatment targets. *Annals of medicine*, v. 40, n. 6, p. 402–417, 2008.

O'NEILL, M. A.; YORK, W. S. *The Plant Cell Wall*. Blackwell, Oxford, 2003.

PETERSEN, B. L.; FABER, K.; ULVSKOV, P. Glycosyltransferases of the GT77 Family. *Annual Plant Reviews*, v. 41, p. 305–320, 2010.

POL-FACHIN, L. *et al.* Effects of glycosylation on heparin binding and antithrombin activation by heparin. *Proteins: Structure, Function and Bioinformatics*, v. 79, n. 9, p. 2735–2745, 2011.

POPPER, Z. A. *et al.* Evolution and diversity of plant cell walls: from algae to flowering plants. *Annual Rev. Plant Biol.* n. 62, p. 567-590, 2011.

RAMACHANDRAN, G. N.; RAMAKRISHNAN, C. e SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. *Journal of molecular biology*, v. 7, p. 95–99, 1963.

SALI, A.; BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*, v. 234, n. 3, p. 779–815, 1993.

SCOTT, W. R. P. *et al.* The GROMOS Biomolecular Simulation Program Package. *The Journal of Physical Chemistry A*, v. 103, n. 19, p. 3596–3607, 1999.

SHOWALTER, A. M. Structure and Function of Cell Wall Proteins. *The Plant Cell*, v. 5, p. 9-23, 1993.

- SHOWALTER, A. M. *et al.* A Bioinformatics Approach to the Identification, Classification, and Analysis of Hydroxyproline-Rich Glycoproteins. *Plant Physiol.* n. 153, p. 485, 2010.
- SOMERVILLE, C. *et al.* Toward a systems approach to understanding plant cell walls. *Science*, v. 306, n. 5705, p. 2206–2211, 2004.
- STAFSTROM, J. P.; STAEHELIN, L. A. The Role of Carbohydrate in Maintaining Extensin in an Extended Conformation. *Plant Physiol.* n. 81, p. 242-246, 1986.
- TIAINEN, P.; MYLLYHARJU, J.; KOIVUNEN, P. Characterization of a second *Arabidopsis thaliana* prolyl 4-hydroxylase with distinct substrate specificity. *Journal of Biological Chemistry*, v. 280, n. 2, p. 1142–1148, 2005.
- VARNER, J. E.; LIN, L. Plant cell wall architecture. *Cell*, v. 56, p. 231-239, 1989.
- VELASQUEZ, S. M. *et al.* O-glycosylated cell wall proteins are essential in root hair growth. *Science*, v. 332, n. 6036, p. 1401–1403, 2011.
- VELASQUEZ, S. M.; IUSEM, N. D.; ESTEVEZ, J. M. Root hair sweet growth. *Plant Signaling & Behavior*, v. 6, n. 10, p. 1600-1602, 2011.
- VERLI, H. Dinâmica Molecular. In: VERLI, H. (Org.). *Bioinformática da Biologia à Flexibilidade Molecular*. Porto Alegre: 2014. p. 173-187
- VINCKEN, J. P. *et al.* If homogalacturonan were a side chains of rhamnogalacturonan – Implications for cell wall architecture. *Plant Physiol.* n. 132, p. 1781-1789, 2003.
- WILLATS, W. G. *et al.* Pectin: cell biology and prospects for functional analysis. *Plant Mol. Biol.* n. 47, p. 9-27, 2001.
- XU, J. *et al.* The O-Hyp glycosylation code in tobacco and *Arabidopsis* and a proposed role of Hyp-glycans in secretion. *Phytochemistry*, v. 69, n. 8, p. 1631-1640, 2008.

YUASA, K. *et al.* Membrane-anchored prolyl hydroxylase with an export signal from the endoplasmic reticulum. *Plant Journal*, v. 41, n. 1, p. 81–94, 2005.

9. Anexos

- “Guia para Autores” – Normas de publicação de manuscritos na revista PROTEINS, em inglês.
- Normas para apresentação oral e escrita do trabalho experimental do estágio em pesquisa e monografia.
- Orientações aos examinadores das bancas de TCC.

Proteins: Structure, Function, and Bioinformatics – Author Guidelines

Types of Articles

PROTEINS: Structure, Function, and Bioinformatics accepts six different types of manuscripts, all of which are peer-reviewed:

RESEARCH ARTICLE: This is the main avenue for publication of original research. Most of the articles published in any volume of PROTEINS are in the form of Research Articles. This form of publication is intended for reports of substantial research that make clear contributions that advance a field or a problem. We discourage submission of Research Articles based on material that is more appropriate for an archival journal.

SHORT COMMUNICATION: Research reports can also be submitted in the form of a Short Communication. These are reserved for short reports of highly important and noteworthy findings. A Short Communication cannot be longer than 4 printed pages.

REVIEW: Reviews contribute critical assessment of interesting and important problems or fields. Authors wishing to submit reviews are encouraged to contact the Editor-in-Chief for approval prior to submission. All Reviews in PROTEINS are peer-reviewed.

PERSPECTIVE: Previously known as Research Commentary, the Perspectives section is intended for opinion pieces in a range of styles. The goal of this section is to foment presentation of controversial ideas, critical evaluation of concepts, idiosyncratic viewpoints, and challenges to ideas. Didactic pieces and historical and biographical accounts will also be considered.

STRUCTURE NOTE: This type of report was created to provide an avenue for dissemination of structures determined by x-ray crystallography or solution or solid-state NMR spectroscopy. The goal of a Structure Note is to document a new structure or another interesting structural finding. Some Structure Notes can be followed up with a more comprehensive description of biological impact, published at a later date. Structure Notes are a useful forum for structures determined under structural genomics initiatives and other structural biology projects. They are expected to contribute novel structural or biological insight. Structure Notes cannot be longer than 4 printed pages (3500 words, not including the Abstract and References). Specific instructions for the format of Structure Notes are presented below. We discourage submission of Structure Notes describing structures that are already very well represented in the PDB, unless the structure contributes an exciting new finding, or important biological/biochemical insight.

File Format Instructions for Online Submission

Original submissions: First submissions of Research Articles and Reviews should be submitted as a single PDF document that includes title page, abstract, text, references, figure legends, tables and figures (in that order). The manuscript should conform to the format described in the instructions for

manuscript preparation, below. For original submissions of Short Communications, Perspectives and Structure Notes must be submitted with production-ready files as described in the guidelines for revised submissions below.

Revised submissions: All revised manuscripts, as well as first submissions of Short Communications, Perspectives and Structure Notes, must be submitted with production-ready files. Carefully follow the instructions below to ensure that, if your manuscript is accepted, it will be published as quickly as possible. The manuscripts must also adhere to the format described in the instructions for manuscript preparation, below.

Text

1. Submit your text file in .DOC or .RTF format.
2. The text document should include the title page, abstract, text, references and figure legends. Do not embed figures or tables in this document.
3. Designate this file as "Main Document" in the File Designation field when you upload it to ScholarOne Manuscripts.

Tables

1. Create tables as text files and save in either .DOC or .RTF format.
2. Designate this file as "Table" in the File Designation field when you upload it to ScholarOne Manuscripts.

Figures

1. Submit your figure files in .TIFF or .EPS format. After the files complete the upload process, label the figures by placing the figure number and figure legend in the Caption/Legend box for each figure. This should be done in addition to including a figure legend in your text file. Authors are also encouraged to use the file tags option when uploading image files in order to take advantage of the linked HTML features of ScholarOneManuscripts. All panels in a multi-panel figure (example: Figure 1A, 1B, 1C, etc.) must be submitted within a single file (example: Figure 1).
2. To ensure that your digital graphics are suitable for print purposes, please go to RapidInspector at <http://rapidinspector.cadmus.com/zwi/index.jsp>. This free, stand-alone software application will help you to inspect and verify illustrations right on your computer.
3. Vector-based figures (e.g., figures created in Adobe Illustrator) should be submitted in EPS format.
4. To ensure ease of legibility for reviewers as well as the highest print quality, TIFF and EPS files must be submitted according to the following minimum resolutions:
 - a. 1200 dpi (dots per inch) for black and white line art (simple bar graphs, charts, etc.)

- b. 300 dpi for halftones (black and white photographs)
 - c. 600 dpi for combination halftones (photographs that also contain line art such as labeling or thin lines)
5. In addition to the above resolution guidelines, color figures must be submitted in CMYK colorspace. Do not submit color figures as RGB. This color space does not reproduce well for print production.
6. Designate this file as "Image" in the File Designation field when you upload it to ScholarOne Manuscripts.

Supporting Online Material

Authors may submit Supporting Information for their articles to be posted in the electronic version of the journal. Instructions for preparation of Supporting Information are provided below. These materials must be submitted online using the File Designation "Supplementary Material for Review." Supporting text, tables and figures must be in PDF format. The only exceptions are audio and moving image files, which should be submitted in .mov or .mp3 format. See the Instructions for Preparation of Supporting Material for further instructions.

Instructions for Manuscript Preparation

Manuscripts

1. When submitting via ScholarOne Manuscripts, authors must provide the names, e-mail addresses, and institutions (affiliation) of five possible reviewers and two preferred Editors. Options are also provided for designating names of persons that authors prefer not to review or edit their paper. A brief explanation for exclusion of reviewers is suggested. The Editorial Office will refrain from assigning any person listed as "non-preferred". Authors should note that preferred Editors and Reviewers may not be available nor will these persons necessarily be assigned as Editors or Reviewers.
2. Authors should include a cover letter to the Editor-in-Chief with the manuscript submission. This may be entered as plain text in the field provided, or uploaded as a .DOC, .RTF, or .PDF file.
3. All manuscripts must be in 12-point Times New Roman font, in single column and double spaced format.
 - a. Problems with use of English is increasingly a problem that precludes fair review of manuscripts. Authors that are not fully comfortable with English should have their manuscripts edited by a native English speaker. Authors in Japan may contact Wiley-Japan for a list of recommended services for checking and improving use of English. Please contact Yoko Kobayashi or A. Bocquet in the Wiley-Japan office by facsimile: 81 3-5689-7276 or by E-mail: editorial@wiley.co.jp for more information. Please indicate the name of the journal clearly.
4. Number all pages in sequence, beginning with the title page.

5. Follow the guidelines in CBE Style Manual Committee. CBE style manual: a guide for authors, editors, and publishers in the biological sciences. 5th ed. rev. and expanded. Bethesda, MD: Council of Biology Editors, Inc.; 1983.

Note that formats for Review articles are flexible. The text for all other manuscript types should conform to the following format.

Manuscripts that do not conform to the following format will be returned. The parts of the manuscript must appear in the following order:

Title Page

The title page must contain the following information:

1. The full title of the manuscript. The title should be descriptive and representative of the material in the manuscript.
2. A short title of not more than 45 characters (including spaces).
3. Five to ten key words not used in the title that will adequately index the subject matter of the article.
4. The names and affiliations of all authors.
5. The institution at which the work was performed.
6. Complete contact information for the author responsible for correspondence:
 - a. Name
 - b. Address/Affiliation
 - c. E-mail address

Abstract

This should summarize the purpose, methods, results, and major conclusions of the work. A 250 word limit will be strictly enforced.

Introduction, Materials and Methods, Results, Discussion, and Conclusion

1. The sections should be presented in the following order: Introduction, Materials and Methods, Results, Discussion, and Conclusion (if applicable). The Results and Discussion section can be combined into a single section when appropriate.
2. The manuscript should conform to standard scientific reporting style.
3. Sufficient data and information must be given so that the study can be replicated.

References

1. Cite references to published literature in the text numerically.
2. Provide full titles and complete page numbers for all works cited in the reference section.
3. Refer to the CBE style manual for the style of reference (see below for more information).

Figure Legends

1. Figure legends should be as clear and concise as possible.

Tables

1. Indicate placement of all tables in the text with a citation (e.g., Table1, Table 2, etc.). When uploading tables and figures, use the same tags for file names in order to utilize HTML linking in the manuscript to these elements.
2. Tables must be numbered in order of appearance with Arabic numerals (e.g., Table 1, Table 2, etc.).
 - a. A legend must accompany each illustration, and tables must have titles.
 - b. All abbreviations used must be defined.
 - c. All lettering must meet professional standards and be legible after reduction in size.

Figures

1. Indicate placement of all illustrations in the text with a citation, i.e., (Fig.1). When uploading tables and figures, use the same tags for file names in order to utilize HTML linking in the manuscript to these elements.
2. Label the figures by using the Captions/Legend box to insert the figure number (i.e. FIGURE 1) AND the figure legend for each figure uploaded. This will help reviewers identify each figure in the PDF version of the manuscript.
3. For figures with multiple parts, the figures must be submitted in the assembled form in which they would appear in the published paper.
4. Figures must be submitted in the size (or close to it) in which they will appear in print.
5. Minimize the number of figures.

Color Figure Preparation

1. For best reproduction, bright, clear colors should be used.
2. Dark colors against a dark background do not reproduce well; please place your color images against a white background wherever possible.

3. Unless otherwise arranged with the Editor, stereo pairs will be printed to a scale that yields a separation of 55-60 mm between corresponding points in the left and right images. Parallel-viewing pairs are preferred. Authors who choose to submit cross-eyed pairs must specify this feature in the legend.

4. Follow resolution guidelines listed above under "File Format Instruction for Online Submission."

Color Reproduction Charges

Color printing is free for one color figure and \$250 for each additional color figure. Figures can be printed in black and white and appear online in color (as submitted) at no charge. You will receive a Color Charge form, post-acceptance, to enable you to make your color figure selections.

Cover Illustration

Authors may nominate figures (or portions thereof) from their manuscript to be used as a cover illustration for the print issue of PROTEINS. Include a request in your cover letter and upload your image using the File Designation "Supplementary Material Not for Review. Name your figure file "Cover Nomination".

Instructions for Preparation of Supporting Information

Tables and Figures in Supporting Information should be labeled using the prefix S before the number (e.g. Table 1 in Supplementary Information should be labeled Table S1, Figure 1 in Supporting Information should be Figure S1). References are not allowed in the Supporting Information. All references should be included in the References section of the main text. Preparation of Tables and Figures should follow the same guidelines described above for the main text.

Database Linking

For papers describing structures of biological macromolecules, the atomic coordinates and the related experimental data (structure factor amplitudes/intensities and/or NMR restraints) must be deposited at a member site of the Worldwide Protein Data Bank (www.wwpdb.org): RCSB PDB (www.pdb.org), PDBe (www.ebi.ac.uk/pdbe), PDBj (www.pdbj.org), or BMRB (www.bmrwisc.edu). The PDB ID must be included in the manuscript and entered into the designated field in the ScholarOne manuscripts upon submission. Authors must agree to release the atomic coordinates and experimental data when the associated article is published. Questions relating to depositions should be sent to deposit@wwpdb.org.

References

Journal:

1. King VM, Armstrong DM, Apps R, Trott JR. Numerical aspects of pontine, lateral reticular, and inferior olivary projections to two paravermal cortical zones of the cat cerebellum. *J Comp Neurol* 1998;390:537-551.

Book:

2. Voet D, Voet JG. *Biochemistry*. New York: John Wiley & Sons; 1990. 1223 p.

Book Chapter:

3. Gilmore ML, Rouse ST, Heilman CJ, Nash NR, Levey AI. Receptor fusion proteins and analysis. In: Ariano MA, editor. *Receptor localization*. New York: Wiley-Liss; 1998. p 75-90.

Electronic Media:

4. Bio-Xplor, Version 1.0. New York: Biostructure Inc.; 1991.

All Manuscripts submitted to *PROTEINS: Structure, Function, and Bioinformatics* must be submitted solely to this journal, and may not have been published in another publication of any type, professional or lay. Manuscripts published previously in the proceedings of a scientific meeting, or in a journal such as *Nature Precedings*, should not be submitted to *PROTEINS*.

If your paper is accepted, the author identified as the formal corresponding author for the paper will receive an email prompting them to login into Author Services; where via the Wiley Author Licensing Service (WALS) they will be able to complete the license agreement on behalf of all authors on the paper.

No published material may be reproduced or published elsewhere without the written permission of the publisher and the author. All statements in, or omissions from, published manuscripts are the responsibility of the authors, who will assist the editor by reviewing proofs before publication. Reprint order forms will be sent with the proofs.

Production

Questions regarding production aspects for papers that have been accepted for publication can be addressed to:

Production Editor

E-mail: jrnprodprot@cadmus.com