



Modeling of glycosyl hydrolases of the Ustilaginaceae family using the online server SWISS-MODEL

Modelamiento de Glicosil hidrolasas de la familia Ustilaginaceae empleando el servidor en línea SWISS-MODEL

Nazaret Flores-Sánchez², Gerardo Díaz-Godínez¹, Daniel Claudio Martínez-Carrera³, Rubén Díaz^{1*}

¹Laboratorio de Biotecnología, Centro de Investigación en Ciencias Biológicas. Universidad Autónoma de Tlaxcala, Ixtacuixtla, Tlaxcala, CP. 90120, México.

²Maestría en Biotecnología y Manejo de Recursos Naturales, Universidad Autónoma de Tlaxcala, Ixtacuixtla, Tlaxcala, CP. 90120, México

³Biotechnology of Edible, Functional and Medicinal Mushrooms, College of Postgraduates, Puebla Campus, Puebla, México.

*Corresponding author.

E-mail address: rdiazgod@hotmail.com (R. Díaz).

Article history:

Received: 21 November 2017 / Received in revised form: 29 December 2017 / Accepted: 30 December 2017 / Published online: 1 January 2018.

<https://doi.org/10.29267/mxjb.2018.3.1.73>

ABSTRACT

In this work, 20 protein sequences of the Ustilaginaceae family were searched for modeling in the NCBI database. The modeling of these proteins using the online SWISS-MODEL server allowed appreciating the regions conserved in 17 proteins through the presence of secondary structures such as α -helices and β -sheets, however, three of them did not show structural similarity with the rest of proteins. Further studies need to be undertaken on the composition of the amino acids of the active site and on the signal peptide region, as well as carrying out a protein phylogenetic analysis with greater similarity.

Keywords: Glycosyl hydrolases, Homology modeling, Ustilaginaceae.

RESUMEN

En este trabajo se realizó la búsqueda de 20 secuencias proteicas de la familia Ustilaginaceae en la base de datos de NCBI para su posterior modelamiento. El modelamiento de estas proteínas utilizando el servidor en línea SWISS-MODEL permitió apreciar las regiones conservadas en 17 proteínas a través de la presencia de estructuras secundarias como α -hélices y hojas- β , sin embargo tres de ellas no presentaron similitud estructural con el resto de las proteínas. Se sugiere realizar estudios sobre la composición de los aminoácidos del sitio activo y sobre la región del péptido señal, así como llevar a cabo un análisis filogenético de las proteínas con mayor similitud.

Palabras clave: Glicosil hidrolasas, Modelamiento por homología, Ustilaginaceae.

1. INTRODUCCIÓN

La familia Ustilaginaceae cuenta con 17 géneros y 607 especies de hongos, pertenecen al phylum Basidiomycota, destacando los géneros *Ustilago* y *Sporisorium* (Kirk *et al.*, 2008), se caracterizan por ser hongos fitopatógenos secretores de enzimas hidrolíticas, específicamente de glicosil hidrolasas. Las glicosil hidrolasas o glucosidasas (EC 3.2.1) son enzimas que actúan sobre carbohidratos (Carbohydrate Active Enzymes, CAZymes). Éstas enzimas tienen un papel fundamental en la descomposición de la biomasa, compuesta principalmente por celulosa, hemicelulosa y pectina (Lundell *et al.*, 2010). Con ayuda de estas enzimas los hongos infectan cultivos de interés agroalimentario como el sorgo, avena, cebada, mijo, maíz y caña de azúcar, presentando masas de esporas de color marrón oscuro en diferentes partes de la planta hospedera (Cannon & Kirk, 2007; McTaggart *et al.*, 2012), disminuyendo así su producción. Gracias a la bioinformática se han podido clasificar estas enzimas en familias como la GH43, GH51 y GH78 responsables de degradar pectina y hemicelulosa encontradas en *Sporisorium scitamineum* (Zhao *et al.*, 2013; Que *et al.*, 2014), también se han reportado enzimas relacionadas a glicosil hidrolasas como en *Sporisorium reilianum* SRZ2 (Schirawski *et al.*, 2010).

En los últimos años la bioinformática ha contribuido a la predicción de estructuras proteicas de una gran variedad de organismos con la finalidad de conocer características evolutivas que han sido perpetuadas a través del tiempo, como la afinidad por su sustrato, los aminoácidos presentes en el sitio activo y el tipo de familia con la cual presenta una mayor similitud estructural (Sikosek & Chan, 2014). El modelamiento de proteínas por homología es útil para conocer la estructura de una proteína desconocida empleando moldes que han sido generados por muchos años de investigación mediante técnicas como la cristalografía. Esta técnica resulta ser muy buena para conocer la especificidad, función y regiones conservadas de una proteína (Zhang, 2008; Murakami & Mizuguchi, 2014). La generación de una proteína por homología permite visualizar estructuras secundarias como α -hélices y hojas- β . El modelamiento por homología se ha ido desarrollando gracias al aumento de moldes o plantados disponibles en las bases de datos como Protein Data Bank (PDB). El

molde utilizado debe tener una alta similitud de secuencia aminoacídica con la secuencia objetivo o blanco, que es en sí, la que se desea analizar, mientras mayor sea la similitud entre las dos mejor será el resultado, influyendo también la calidad del alineamiento de las secuencias en el modelo obtenido (Leman *et al.*, 2015; Bohórquez *et al.*, 2017).

Uno de los pioneros en esta rama desde 1993 es SWISS-MODEL que es un servidor para el modelamiento comparativo automatizado de estructuras proteicas tridimensionales y ha permanecido en internet por poco más de veinte años. Este servidor tiene la capacidad de realizar la selección del molde o templado, la alineación de las secuencias y la construcción del modelo de manera automática (Schwede, *et al.*, 2003; Bordoli & Schwede, 2012). En el presente trabajo se realizó el modelamiento por homología de veinte secuencias proteicas de diferentes especies de la familia Ustilaginaceae con el objetivo de visualizar las regiones conservadas a través de la similitud presente en sus estructuras secundarias.

2. MATERIALES Y MÉTODOS

Se realizó el modelamiento de proteínas por homología empleando 20 secuencias de glicosil hidrolasas (ETS61490.1, GAC92623.1, ETS61283.1, GAC77333.1, XP_014655537.1, XP_016292886.1, XP_014658579.1, CDU23576.1, GAC74419.1, XP_014657268.1, GAC73010.1, GAC72934.1, XP_014656155.1, CCF51500.1, CDR87383.1, GAC76924.1, XP_014658453.1, KIS67306.1, GAC93679.1, CBQ73989.1) obtenidas de la base de datos NCBI-Protein (<https://www.ncbi.nlm.nih.gov>) de diferentes especies de hongos pertenecientes a la familia Ustilaginaceae. Se utilizó el servidor en línea SWISS-MODEL (<https://swissmodel.expasy.org>) para la generación de las estructuras proteicas empleando las características particulares del software. Posteriormente se buscó el ángulo más apropiado para visualizar las estructuras secundarias y se tomó una foto instantánea de cada una de las proteínas obtenidas.

3. RESULTADOS

Se obtuvo el modelamiento por homología de 20 estructuras proteicas de glicosil hidrolasas, las primeras tres proteínas presentaron una gran similitud en la conformación de estructuras secundarias, especialmente de hojas- β . Se encontró que de las tres proteínas modeladas dos correspondieron a *Moesziomyces aphidis* DSM 70725 (Figs. 1a y 1c) y la restante correspondió a *Pseudozyma hubeiensis* SY62 (Fig. 1b).

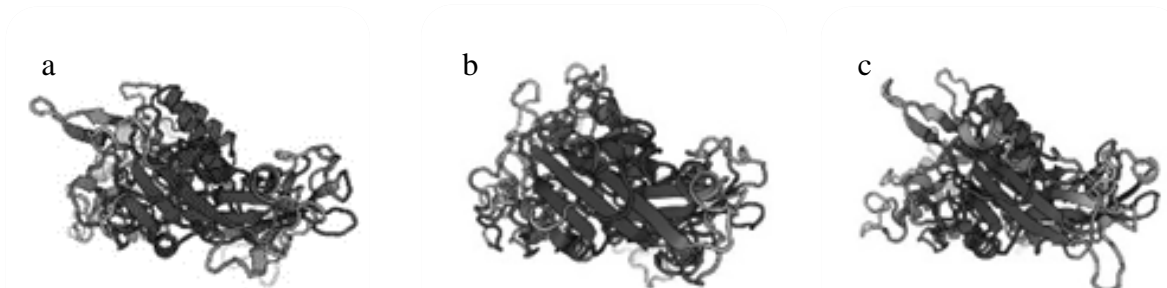


Fig. 1. Modelamiento de proteínas utilizando el servidor en línea SWISS-MODEL a) *Moesziomyces aphidis* DSM 70725 (ETS61490.1), b) *Pseudozyma hubeiensis* SY62 (GAC92623.1), c) *Moesziomyces aphidis* DSM 70725 (ETS61283.1).

En tres glicosil hidrolasas se encontró una mayor presencia de α -hélices y formación de loops con respecto al resto de estructuras modeladas. La alta similitud de estas proteínas es de gran importancia debido a que de las tres, dos provienen de la misma especie y la restante correspondió a una especie diferente, *Moesziomyces antarcticus* T-34, *Moesziomyces antarcticus* y *Kalmanozyma brasiliensis* GHG001 (Figs. 2a, 2b y 2c).

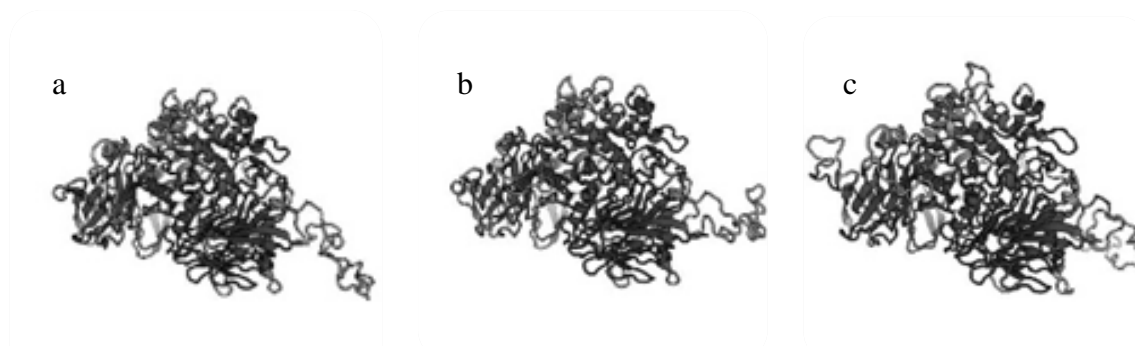


Fig. 2. Modelamiento de proteínas utilizando el servidor en línea SWISS-MODEL a) *Moesziomyces antarcticus* T-34 (GAC77333.1) b) *Moesziomyces antarcticus* (XP_014655537.1) y c) *Kalmanozyma brasiliensis* GHG001 (XP_016292886.1).

Las tres estructuras proteicas provenientes de dos especies diferentes *Moesziomyces antarcticus* T-34, *Moesziomyces antarcticus* y *Ustilago maydis* 521 presentaron una elevada similitud en su disposición estructural. En la región inferior de las tres proteínas destacó la presencia de dominios formados por hojas- β y, en la parte superior, destacaron α -hélices casi en la misma posición (Figs. 3a, 3b y 3c).

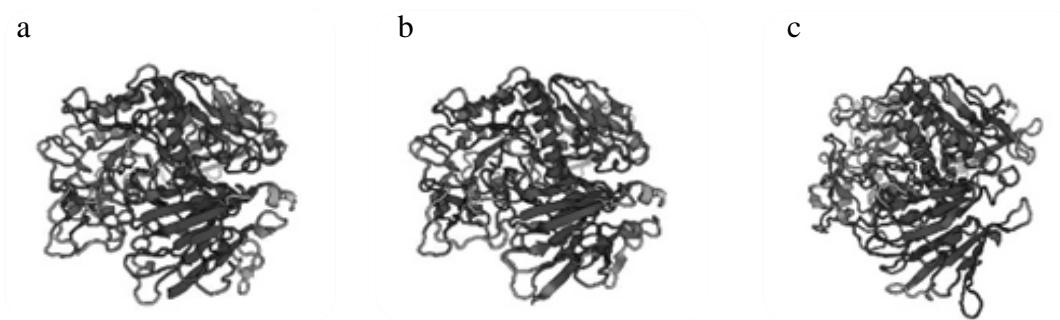


Fig. 3. Modelamiento de proteínas utilizando el servidor en línea SWISS-MODEL a) *Moesziomyces antarcticus* T-34 (GAC76924.1), b) *Moesziomyces antarcticus* (XP_014658453.1) y c) *Ustilago maydis* 521 (KIS67306.1).

Un caso particular se encontró en las proteínas de *Sporisorium scitamineum*, *Moesziomyces antarcticus* y *Sporisorium reilianum* SRZ2 donde ninguna de estas presentó una similitud con las diecisiete estructuras restantes (Figs. 4a, 4b y 4c).

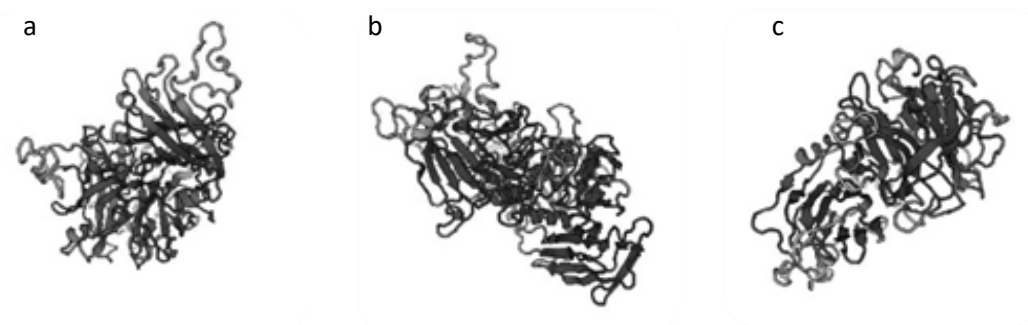


Fig. 4. Modelamiento de proteínas utilizando el servidor en línea SWISS-MODEL a) *Sporisorium scitamineum* (CDR87383.1), b) *Moesziomyces antarcticus* (XP_014657268.1) y c) *Sporisorium reilianum* SRZ2 (CBQ73989.1).

Se encontró alta similitud en dos proteínas que correspondieron a *Moesziomyces antarcticus* y *Moesziomyces antarcticus* T-34, donde los dominios formados por hojas- β y loops predominaron para la formación de ambas estructuras (Figs. 5a y 5b).

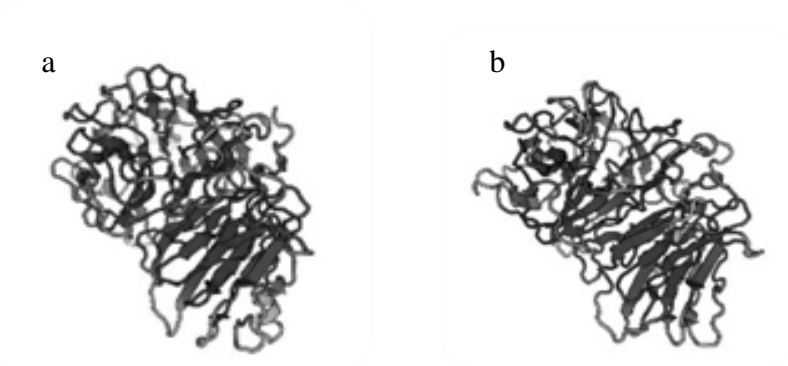


Fig. 5. Modelamiento de proteínas utilizando el servidor en línea SWISS-MODEL a) *Moesziomyces antarcticus* (XP_014658579.1) y b) *Moesziomyces antarcticus* T-34 (GAC74419.1).

Uno de los modelamientos de glicosil hidrolasas con mayor similitud estructural provino de dos organismos diferentes *Sporisorium scitamineum* y *Pseudozyma hubeiensis* SY62 donde destacaron una vez más la presencia de dominios conformados por hojas- β , loops y la ausencia de α -hélices (Figs. 6a y 6b).



Fig. 6. Modelamiento de proteínas utilizando el servidor en línea SWISS-MODEL a) *Sporisorium scitamineum* (CDU23576.1) y b) *Pseudozyma hubeiensis* SY62 (GAC93679.1).

En dos proteínas que correspondieron a *Moesziomyces antarcticus* T34 se halló similitud estructural en los dominios conformados por hojas- β , loops y en la ausencia de α -hélices (Figs. 7a y 7b).

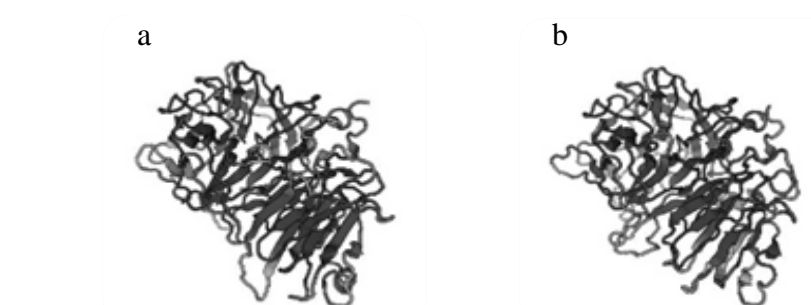


Fig. 7. Modelamiento de proteínas utilizando el servidor en línea SWISS-MODEL a) *Moesziomyces antarcticus* T-34 (GAC73010.1) y b) *Moesziomyces antarcticus* T-34 (GAC72934.1).

Finalmente dos proteínas que provinieron de dos géneros y especies diferentes *Moesziomyces antarcticus* y *Ustilago hordei*, mostraron alta similitud estructural, particularmente en los dominios formados por hojas- β y la presencia de α -hélices en la parte inferior de ambas estructuras (Figs.8a y 8b).

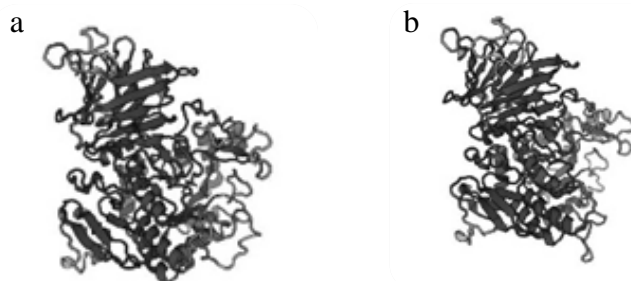


Fig. 8. Modelamiento de proteínas utilizando el servidor en línea SWISS-MODEL a) *Moesziomyces antarcticus* (XP_014656155.1) y b) *Ustilago hordei* (CCF51500.1).

4. DISCUSIÓN

El modelamiento por homología de glicosil hidrolasas utilizando el servidor en línea SWISS-MODEL permitió observar las regiones conservadas representadas a través de la formación de estructuras secundarias, principalmente de hojas- β y loops, además se encontró que las α -hélices presentaron poca frecuencia de aparición, muy posiblemente debido a la ausencia de algunos aminoácidos que determinan este tipo de estructuras secundarias. Por otro lado de las veinte proteínas modeladas diecisiete (ETS61490.1, GAC92623.1, ETS61283.1, GAC77333.1, XP_014655537.1, XP_016292886.1, GAC76924.1, XP_014658453.1, KIS67306.1, XP_014658579.1, GAC74419.1, CDU23576.1, GAC93679.1, GAC73010.1, GAC72934.1, XP_014656155.1, CCF51500.1) presentaron una alta similitud estructural y solo tres de ellas (CDR87383.1, XP_014657268.1, CBQ73989.1) no compartieron características con el resto de las proteínas a pesar de pertenecer a la misma familia Ustilaginaceae, sin embargo no es un factor determinante para no presentar similitud con otros organismos dentro de esta amplia familia de hongos. Se ha utilizado el modelamiento de estructuras tridimensionales por homología para confirmar los análisis realizados con diversos programas bioinformáticos como los de alineamiento múltiple de secuencias aminoacídicas y estudios filogenéticos (Strohmeier *et al.*, 2004; Larroque *et al.*, 2012; Hossain & Roslan, 2014), permitiendo conocer las estructuras secundarias, dominios y regiones conservadas en las proteínas estudiadas.

El estudio realizado sobre el modelamiento de estas proteínas demuestra solo una parte de las características evolutivas con las que cuentan. Para fortalecer los resultados obtenidos se hará una examinación en los aminoácidos que integran las estructuras secundarias y/o dominios más relevantes donde se encontró una alta similitud estructural, siendo las diecisiete proteínas las de mayor interés. Además se efectuarán análisis bioinformáticos del sitio activo y de la secuencia del péptido señal donde a través de las herramientas

bioinformáticas previamente mencionadas se pueden descubrir nuevas familias como lo fue para las β -galactosidasas (Cheng *et al.*, 2017) y un estudio filogenético para ver la historia evolutiva de las diferentes familias a las que pertenecen las glicosil hidrolasas. Intra *et al.* (2008) reportaron secuencias proteicas relacionadas a la familia numero veinte de las glicosil hidrolasas (GH20) de organismos eucariotas y procariotas. El modelamiento de proteínas por homología junto con otras herramientas bioinformáticas permitirá conocer y mejorar las propiedades catalíticas o de termoestabilidad de las glicosil hidrolasas y ser aprovechadas a nivel industrial.

CONFLICTO DE INTERESES

Los autores declaran que no existe conflicto de intereses.

REFERENCIAS

- Bohórquez H. J., Suárez C. F. & Patarroyo M. E. 2017. Mass & secondary structure propensity of amino acids explain their mutability and evolutionary replacements. *Scientific Reports*. 7(7717): 1-12.
- Bordoli L. & Schwede T. 2012. Automated Protein Structure Modeling with SWISS-MODEL Workspace and the Protein Model Portal. *Methods in Molecular Biology*. 857: 107-136.
- Cannon P. F. & Kirk P. M. 2007. *Fungal Families of the World*. Centre for Agriculture and Bioscience International. UK. pp 370-371.
- Cheng J., Romantsov T., Engel K., Doxey A. C., Rose D. R., Neufeld J. D. & Charles T. C. 2017. Functional metagenomics reveals novel β -galactosidasas not predictable from gene sequences. *PLoS ONE*. 12(3): 1-20.
- Hossain M. A. & Roslan H. A. 2014. Molecular Phylogeny and Predicted 3D Structure of Plant beta-D-N-Acetylhexosaminidase. *The Scientific World Journal*. 2014(7): 1-14.
- Intra J., Pavesi G. & Horner D. S. 2008. Phylogenetic analyses suggest multiple changes of substrate specificity within the Glycosyl hydrolase 20 family. *BioMed Central Evolutionary Biology*. 8(214): 1-17.
- Kirk P. M., Ainsworth G. C. & Bisby G. R. 2008: *Ainsworth and Bisby's Dictionary of the Fungi*. Centre for Agriculture and Bioscience International. Wallingford, Connecticut. pp 1-747.
- Larroque M., Barriot R., Bottin A., Barre A., Rougé P., Dumas B. & Gaulin E. 2012. The unique architecture and function of cellulose-interacting proteins in oomycetes revealed by genomic and structural analyses. *BioMed Central Genomics*. 13(605): 1-15.

- Leman J. K., Ulmschneider M. B. & Gray J. J. 2015. Computational modeling of membrane proteins. *Proteins*. 83(1): 1-24.
- Lundell T. K., Mäkelä M. R. & Hildén K. 2010. Lignin-modifying enzymes in filamentous basidiomycetes-ecological, functional and phylogenetic review. *Journal of Basic Microbiology*. 50(1): 5-20.
- McTaggart A. R., Shivas R. G., Geering A. D. W., Vánky K. & Scharaschkin T. 2012. A review of the *Ustilago-Sporisorium-Macalpinomyces* complex. *Persoonia: Molecular Phylogeny and Evolution of Fungi*. 29: 55-62.
- Murakami Y. & Mizuguchi K. 2014. Homology-based prediction of interactions between proteins using Averaged One-Dependence Estimators. *BioMed Central Bioinformatics*. 15(213): 1-11.
- Que Y., Xu L., Wu Q., Liu Y., Ling H., Liu Y. & Zhang C. 2014. Genome sequencing of *Sporisorium scitamineum* provides insights into the pathogenic mechanisms of sugarcane smut. *BioMed Central Genomics*. 15(996): 1-20.
- Schirawski J., Mannhaupt G., Munch K., Brefort T., Schipper K., Doehlemann G., Di Stasio M., Rossel N., Mendoza-Mendoza A., Pester D., Muller O., Winterberg B., Meyer E., Ghareeb H., Wollenberg T., Munsterkotter M., Wong P., Walter M., Stukenbrock E., Guldener U. & Kahmann R. 2010. Pathogenicity determinants in smut fungi revealed by genome comparison. *Science*. 330(6010): 1546-1548.
- Schwede T., Kopp J., Guex N. & Peitsch M. C. 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*. 31(13): 3381-3385.
- Sikosek T. & Chan H. S. 2014. Biophysics of protein evolution and evolutionary protein biophysics. *Journal of the Royal Society Interface*. 11(100): 1-35.
- Strohmeier M., Hrmova M., Fischer M., Harvey A. J., Fincher G. B. & Pleiss J. 2004. Molecular modeling of family GH16 glycoside hydrolases: Potential roles for xyloglucan transglucosylases/hydrolases in cell wall modification in the poaceae. *Protein Science: A Publication of the Protein Society*. 13(12): 3200-3213.
- Zhang Y. 2008. Progress and challenges in protein structure prediction. *Current Opinion in Structural Biology*. 18(3): 342-348.
- Zhao Z., Liu H., Wang C. & Xu J. R. 2013. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BioMed Central Genomics*. 14(274): 1-15.