



## SÀNG LỌC *In silico* CÁC MỤC TIÊU THUỐC TIỀM NĂNG Ở VI KHUẨN *Staphylococcus aureus* KHÁNG METHICILLIN (MRSA) BẰNG CÁC PHƯƠNG PHÁP PHÂN TÍCH DỮ LIỆU PROTEIN

Lê Anh Vũ<sup>1\*</sup>, Phan Thị Cẩm Quyên<sup>2</sup> và Nguyễn Thúy Hương<sup>1</sup>

<sup>1</sup>Khoa Kỹ thuật Hóa học, Đại học Bách Khoa Thành phố Hồ Chí Minh

<sup>2</sup>Phòng Công nghệ Sinh học, Trung tâm Giống Kiên Giang

\*Người chịu trách nhiệm về bài viết: Lê Anh Vũ (email: [lavu68@gmail.com](mailto:lavu68@gmail.com))

### Thông tin chung:

Ngày nhận bài: 13/11/2018

Ngày nhận bài sửa: 22/01/2019

Ngày duyệt đăng: 12/04/2019

### Title:

*In silico* screening of potential drug targets in Methicillin-resistant *Staphylococcus aureus* (MRSA) using protein database analysis methods

### Từ khóa:

MRSA, phân tích CSDL, phương pháp tiếp cận *in silico*, protein không tương đồng, protein thiết yếu

### Keywords:

Database analysis, essential proteins, *in silico*-based approach, MRSA, non-homologous proteins

### ABSTRACT

Due to the emergence of multidrug-resistant strains of Methicillin resistant *Staphylococcus aureus* (MRSA), there is an urgent need for the identification of new targets for drug development. In this study, an *in silico*-based approach was used to investigate several protein/gene databases to find potential target proteins in MRSA. The results show that 158 proteins are non-homologous essential proteins in which 49 proteins take part in 11 unique metabolic pathways. According to DrugBank database, two proteins namely respiratory nitrate reductase alpha chain (NarG) and tubulin homolog protein (FtsZ) was suggested as best putative targets against MRSA. The other proteins were considered as putative novel target. The identified drug targets are expected to be of great potential for discovery of novel therapeutic compounds against MRSA.

### TÓM TẮT

Với sự xuất hiện của nhiều chủng *Staphylococcus aureus* kháng Methicillin (MRSA), có một nhu cầu cấp thiết cho sự xác định các mục tiêu mới phục vụ quá trình phát triển thuốc. Trong nghiên cứu này, phương pháp tiếp cận *in silico* được sử dụng trong sàng lọc một số cơ sở dữ liệu (CSDL) protein/gene để tìm các protein mục tiêu tiềm năng ở MRSA. Kết quả cho thấy 158 protein thiết yếu không tương đồng trong đó có 49 protein tham gia vào 11 con đường trao đổi chất riêng biệt. Theo CSDL DrugBank, hai protein là tiểu đơn vị alpha enzyme hô hấp khử nitrat (NarG) và protein tương đồng tubulin (FtsZ) được xác định là các mục tiêu tốt nhất. Các protein còn lại được đề xuất là các mục tiêu giả định mới ở MRSA. Những mục tiêu thuốc được xác định dự kiến sẽ có tiềm năng lớn cho việc khám phá các hợp chất trị liệu mới chống lại MRSA.

Trích dẫn: Lê Anh Vũ, Phan Thị Cẩm Quyên và Nguyễn Thúy Hương, 2019. Sàng lọc *In silico* các mục tiêu thuốc tiềm năng ở vi khuẩn *Staphylococcus aureus* kháng Methicillin (MRSA) bằng các phương pháp phân tích dữ liệu protein. Tạp chí Khoa học Trường Đại học Cần Thơ. 55(Số chuyên đề: Công nghệ Sinh học)(1): 29-39.

## 1 GIỚI THIỆU

*Staphylococcus aureus* (*S. aureus*) là vi khuẩn Gram dương, yếm khí tùy ý, sống hội sinh và là mầm

bệnh cơ hội. Tụ cầu này có thể tồn tại trong các điều kiện cực đoan và bất lợi như nhiệt độ 7 - 48°C, pH 4.0 - 10.0, nồng độ muối cao, hoạt độ nước thấp (aw)

và stress thâm thấu. Tất cả những yếu tố này cho phép nó tồn tại và cư trú ở nhiều vị trí trên cơ thể người như: khoang mũi, đường tiêu hóa và hệ xương (Patnala and Zaveri, 2016). Sự lây lan chủ yếu của *S. aureus* là do ngộ độc thực phẩm, nhiễm trùng bệnh viện và từ cộng đồng. Loài này có khả năng gây ra nhiều vấn đề sức khỏe, từ nhiễm trùng da đơn giản đến các biến chứng nghiêm trọng dẫn đến tử vong (Stapleton and Taylor, 2002). Các bệnh nhiễm trùng này thường được điều trị bằng các kháng sinh nhóm  $\beta$ -lactam như: methicillin, penicillin, cephalosporin và oxacillin, chủ yếu hoạt động trên các protein gắn với penicillin. Tuy nhiên theo thời gian, vi khuẩn này đã có được sự đề kháng với methicillin và *S. aureus* kháng methicillin (MRSA) đã được báo cáo lần đầu vào năm 1961 (Patnala and Zaveri, 2016). Hiện nay, MRSA được điều trị bằng các kháng sinh phổ rộng đơn trị hoặc phác đồ kết hợp, bao gồm glycopeptide như: vancomycin và teicoplanin, sulfamid và daptomycin (Cordwell *et al.*, 2002). Những loại kháng sinh này vẫn chưa đạt đến mức hoàn toàn chữa bệnh và *S. aureus* cũng đang phát triển đề kháng ngay cả đối với những liệu pháp này. Ngoài ra, việc sử dụng nhiều loại kháng sinh khác nhau trong những năm qua đã góp phần vào sự xuất hiện của các chủng MRSA đa kháng thuốc (Stapleton and Taylor, 2002). Các báo cáo gần đây về chủng *S. aureus* có tính kháng trung gian hoặc hoàn toàn đối với vancomycin đã làm cho thời kỳ hóa trị liệu trong đó các kháng sinh diệt khuẩn có hiệu quả chống lại vi khuẩn này có thể sẽ không còn kéo dài (Cordwell *et al.*, 2002). Do đó, việc tìm kiếm các hợp chất mới chống tụ cầu là một vấn đề cấp bách.

Trong khám phá thuốc, một nhiệm vụ quan trọng là xác định được các mục tiêu tiềm năng, từ đó có thể sử dụng để thiết kế thuốc dựa trên mục tiêu. Các phương pháp thí nghiệm cổ điển dùng để xác định các mục tiêu thuốc thường yêu cầu nhiều nhân lực, thời gian và tốn kém về chi phí. Gần đây, với sự phát triển của các ngành khoa học omic (genomics, transcriptomics, proteomics và metabolomics) và các phương pháp tin sinh học (hay còn gọi là phương pháp *in silico*) đã đem đến hướng tiếp cận thay thế và cung cấp thông tin chi tiết trong việc tìm kiếm các mục tiêu thuốc thay thế với ưu thế tiết kiệm thời gian và chi phí. Loại trừ proteomic là một trong những cách tiếp cận *in silico* được sử dụng để xác định mục tiêu thuốc dựa trên việc xác định các protein thiết yếu và không tương đồng với vật chủ ở sinh vật gây bệnh (Hasan *et al.*, 2016). Hơn nữa, các protein tương đồng với vật chủ được sàng lọc còn có thể giúp tránh các phản ứng có hại của thuốc trong quá trình phát triển thuốc. Hướng tiếp cận này đã được sử dụng thành công để xác định mục tiêu thuốc mới của các tác nhân gây bệnh như *Pseudomonas*

*aeruginosa*, *Helicobacter pylori*, *Listeria monocytogenes*, *Candida tropicalis* (Morya *et al.*, 2010), *S. aureus* kháng vancomycin (Hasan *et al.*, 2016) và *Salmonella enterica* (Hossain *et al.*, 2017).

Nghiên cứu này nhằm mục đích tìm kiếm các mục tiêu thuốc tiềm năng ở MRSA theo hướng tiếp cận loại trừ proteomic sử dụng các công cụ và cơ sở dữ liệu (CSDL) tin sinh học để xác định, mô tả và phân tích các protein thiết yếu của MRSA. Những protein này được sàng lọc dựa trên các tiêu chí bao gồm sự không tương đồng của protein mục tiêu đối với bộ proteome vật chủ và vai trò của protein trong quá trình phát triển của vi khuẩn. Tìm kiếm các hợp chất điều trị mới chống lại các mục tiêu như vậy sẽ có nhiều khả năng làm giảm tính đề kháng của vi khuẩn với các phương pháp điều trị hiện có, đồng thời cũng hạn chế các tác động bất lợi lên tế bào vật chủ. Mặc dù nỗ lực về chiến lược tương tự với MRSA đã được báo cáo trong một số nghiên cứu (Haag *et al.*, 2012; Hossain *et al.*, 2013) nhưng trong nghiên cứu này chiến lược trên được áp dụng trên 15 chủng MRSA cùng lúc. Việc phân tích như vậy cho phép xác định các protein có tiềm năng sử dụng làm mục tiêu cho sự phát triển của các tác nhân trị liệu trên một loạt các chủng MRSA.

## 2 PHƯƠNG TIỆN VÀ PHƯƠNG PHÁP

### 2.1 Phương tiện

Tất cả các dữ liệu, công cụ và chương trình sử dụng trong nghiên cứu này đều có sẵn trên internet và được truy cập/tải về theo như mô tả trong mục phương pháp.

### 2.2 Phương pháp nghiên cứu

Quy trình chi tiết để xác định các protein quan trọng ở MRSA nhằm xác định các mục tiêu thuốc tiềm năng được sửa đổi từ quy trình được mô tả bởi Hasan *et al.* (2016).

Bộ proteome của 15 chủng MRSA được trích xuất từ CSDL UniProtKB (UniProt Consortium, 2018) dưới định dạng FASTA. Tiếp theo đó, các protein này được sàng lọc với chương trình USEARCH với lệnh `-cluster_fast` để tìm các protein paralog có “sequence identity cut off value” ở mức 0,6 (tức 60% mức độ đồng nhất). Các nhóm protein còn lại được sàng lọc thêm để tìm ra các chuỗi riêng biệt (còn gọi là quá trình dereplication) bằng cách sử dụng lệnh `-fastx_uniques`. Các bản sao được loại bỏ và lệnh `-sortbylength` được sử dụng để sàng lọc các protein có hơn 100 acid amin (Edgar, 2010). Các protein kết quả được nhóm lại. Phân tích BLASTp đã được thực hiện cho các protein không lặp lại (non-paralog) đối với bộ proteome của người (*Homo sapiens*) sử dụng CSDL NCBI (NCBI BLASTp) (Boratyn *et al.*, 2013). Protein có giá trị dưới ngưỡng

kỳ vọng (E-value)  $10^{-4}$  bị loại bỏ, giả sử rằng chúng có một mức độ tương đồng nhất định với bộ proteome vật chủ (Zhang *et al.*, 2004). Các protein không tương đồng sau đó đã được tách ra để phân tích sâu hơn.

Tiếp theo các protein không tương đồng được phân tích với CSDL DEG (Zhang *et al.*, 2004). Giá trị ngưỡng kỳ vọng (E-value) được thiết lập ở mức  $10^{-4}$  và giá trị “minimum bit-score cut-off value 100”, ma trận BLOSUM62 và chế độ căn chỉnh “gapped” được chọn để xác định các gen thiết yếu (Rathi *et al.*, 2009). Các protein được lựa chọn của MRSA có thể được coi là mục tiêu thuốc, bởi vì chúng không có mặt trong vật chủ nhưng chỉ có ở MRSA. Phân tích sâu hơn được thực hiện trên các protein thiết yếu này để xác định vị trí và chức năng trong tế bào của chúng. Máy chủ KAAS của CSDL Kyoto Encyclopedia of Genes and Genomes (KEGG) (Moriya *et al.*, 2007) được sử dụng để xác định chức năng của các protein này.

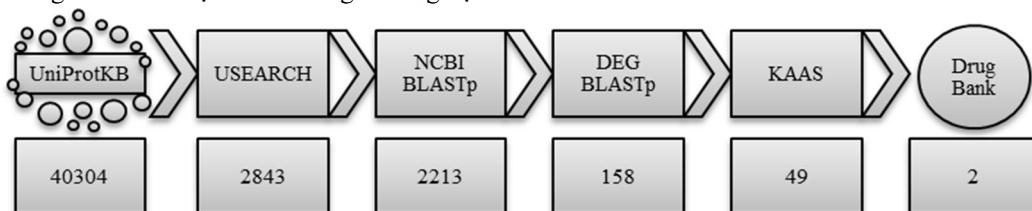
Máy chủ KAAS cung cấp chú thích chức năng của các protein bằng các so sánh BLASTp với CSDL KEGG. Kết quả có chứa các mã số KO (KEGG Orthology) và các đường dẫn KEGG được tạo tự động (Moriya *et al.*, 2007). Nghiên cứu KEGG pathway cũng được tiến hành để phân tích con đường chuyển hóa mà các protein được chọn làm mục tiêu tiềm năng tham gia vào. Điều này cho phép dự đoán chức năng của các protein và dẫn đến việc xác định các mục tiêu tiềm năng. Việc sàng lọc các mục tiêu thuốc tiềm năng được thực hiện bằng cách thực hiện BLASTp chuỗi protein của tất cả các protein tiềm năng đối với CSDL DrugBank (Knox *et al.*, 2011) để xác định các mục tiêu thuốc hợp lý. DrugBank là một CSDL có chứa thông tin chi tiết về thuốc, mục tiêu thuốc, tác động của thuốc và tương tác thuốc về các loại thuốc được Cơ quan Quản lý Thực phẩm và Thuốc Hoa Kỳ (FDA) chấp thuận cũng như các loại thuốc đang thử nghiệm

thông qua quy trình phê duyệt của FDA. Để thực hiện phân tích này, trình tự acid amin của các protein được sàng lọc ở các bước trên đã được tải lên và các tham số mặc định được chọn với giá trị ngưỡng kỳ vọng là  $10^{-5}$  (Hasan *et al.*, 2016).

### 3 KẾT QUẢ VÀ THẢO LUẬN

Vì sự phát triển đề kháng ngày càng tăng ở các vi khuẩn gây bệnh nên ngày càng có nhiều nghiên cứu trong lĩnh vực được phẩm nhằm khám phá các loại thuốc kháng khuẩn mới. Xác định các protein đặc hiệu ở vi khuẩn thông qua hướng tiếp cận phát hiện thuốc và thiết kế các loại thuốc mới nhằm vào các mục tiêu đã biết là hai phương tiện phổ biến để chống lại sự kháng thuốc. Mục tiêu của nghiên cứu này là xác định các mục tiêu thuốc giả định tiềm năng ở MRSA bằng cách sử dụng phương pháp tiếp cận loại trừ proteome mà có thể xác định và phân loại các mục tiêu theo các tiêu chí đã lựa chọn. Hình 1 và Bảng 1 mô tả số lượng các protein đầu ra được sàng lọc ở mỗi bước.

Bộ proteome của 15 chủng MRSA sử dụng trong nghiên cứu này có tổng cộng 40304 chuỗi protein (Bảng 2). Trong số này, 26171 chuỗi được xác định là các protein paralog. Các protein paralog, protein lặp lại (replication) và các protein có độ dài chuỗi ít hơn 100 acid amin, như được phân tích bởi chương trình USEARCH được loại ra khỏi quá trình phân tích tiếp theo để giảm số lượng chuỗi phân tích và tránh kết quả lặp lại. Điều này dẫn đến 2843 protein ứng viên cho các bước sàng lọc tiếp theo. Phân tích BLASTp của các protein này đối với bộ proteome người (*Homo sapiens*) đã cho thấy có 2213 protein không tương đồng với bộ proteome vật chủ. Sự tương đồng giữa protein vật chủ và mầm bệnh nếu được chọn làm mục tiêu thiết kế thuốc có thể dẫn đến phản ứng chéo và gây độc cho tế bào vật chủ. Do đó, chỉ có các protein không tương đồng được chọn để xác định các protein thiết yếu.



Hình 1: Sơ đồ các bước sàng lọc bộ proteome MRSA và số lượng protein đầu ra ở mỗi bước

**Bảng 1: Kết quả phân tích bộ proteome vi khuẩn MRSA**

Các bước sàng lọc	Số lượng protein tìm được
Protein từ CSDL UniProtKB	40304
Protein paralog (>60% mức độ tương đồng) tìm thấy bởi USEARCH	26171
Protein non-paralog	14133
Protein non-paralog/không lặp lại tìm thấy bởi USEARCH	3815
Protein non-paralog/không lặp lại có độ dài chuỗi lớn hơn 100 amino acid (USEARCH)	2843
Protein không tương đồng với proteome người (E-value $10^{-4}$ ) (NCBI BLASTp)	2213
Protein không tương đồng/Protein thiết yếu tìm thấy trong DEG (E-value $10^{-100}$ ) (DEG BLASTp)	158
Các con đường chuyển hóa riêng biệt được tìm thấy trên máy chủ KAAS	27
Protein thiết yếu trong các con đường chuyển hóa riêng biệt tìm được trên máy chủ KEGG (KAAS KEGG)	49
Protein thiết yếu được lựa chọn làm mục tiêu thuốc qua DrugBank	2

Các protein thiết yếu là những protein quan trọng để hỗ trợ sự sống của tế bào. Bởi vì hầu hết các kháng sinh nhắm vào các quá trình tế bào quan trọng trong các vi sinh vật gây bệnh, gen thiết yếu của vi khuẩn và các sản phẩm dịch mã của nó là những mục tiêu mới đầy tiềm năng cho các loại thuốc kháng khuẩn (Zhang *et al.*, 2004). Các protein thiết yếu đặc trưng cho một sinh vật có thể được xem là các mục tiêu thuốc cụ thể cho loài đó (Judson and Mekalanos, 2000). Bằng cách sử dụng CSDL DEG, 158 protein thiết yếu có ở các chủng MRSA đã được xác định. Bên cạnh việc sàng lọc các protein thiết yếu, nghiên cứu của chúng tôi cũng đồng thời xác định các con đường chuyển hóa riêng biệt ở MRSA mà các protein này tham gia. Phân tích các con đường chuyển hóa có một ý nghĩa quan trọng trong

việc xác định mục tiêu thuốc tiềm năng vì nó có thể là chu trình cần thiết cho sự tồn tại và khả năng gây bệnh của vi khuẩn (Hasan *et al.*, 2016). Sau khi so sánh các con đường chuyển hóa giữa MRSA và *Homo sapiens* tại máy chủ KEGG, 27 con đường chuyển hóa riêng biệt của MRSA đã được tìm thấy (Bảng 3). Tất cả 158 protein thiết yếu sau đó được phân tích với máy chủ KAAS. Phân tích chi tiết con đường chuyển hóa cho thấy sự tham gia của 134 protein trong số 158 protein vào các con đường chuyển hóa khác nhau ở MRSA, 24 protein còn lại chưa được mô tả rõ hoặc chưa được xác định (Hình 2). Trong số 134 protein này, 49 protein được tìm thấy rất quan trọng cho sự tồn tại của sinh vật và tham gia vào 11 con đường chuyển hóa riêng biệt khác nhau (Bảng 4 và Hình 3).

**Bảng 2: Các chủng MRSA sử dụng trong nghiên cứu và số lượng protein tương ứng**

Mã số proteome tại CSDL UniProtKB	Tên chủng	Mã số chủng tại CSDL UniProtKB	Số lượng protein
UP000024517	<i>Staphylococcus aureus</i> MRSA-136	1413497	2688
UP000000596	<i>Staphylococcus aureus</i> (strain MRSA252)	282458	2635
UP000051781	<i>Staphylococcus aureus</i> MRSA_CVMN25720PS	1387326	2546
UP000003093	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> DR10	1155079	2684
UP000031632	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> ST772-MRSA-V (Strain: DAR4145)	1343064	2630
UP000025090	<i>Staphylococcus aureus</i> MRSA-118	1413495	2661
UP000003572	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> MRSA177	754026	2773
UP000005003	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> 71193	1155084	2620
UP000008696	<i>Staphylococcus aureus</i> (strain MRSA ST398/isolate S0385)	523796	2648
UP000003548	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> MRSA131	754025	2770
UP000050970	<i>Staphylococcus aureus</i> MRSA_CVMN27231PS	1387328	2743
UP000050900	<i>Staphylococcus aureus</i> MRSA_CVMN26035PS	1387327	2719
UP000054253	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> MRSA S4	1337995	2886
UP000186671	<i>Staphylococcus aureus</i> (strain MRSA ST398/isolate S0385) (Strain: 55488)	523796	2650
UP000031939	<i>Staphylococcus aureus</i> MRSA_CVM43477	1448137	2651

Cuối cùng, sàng lọc các mục tiêu thuốc tiềm năng được thực hiện bằng cách sử dụng CSDL DrugBank cho 49 protein thiết yếu được xác định bởi DEG và KAAS. Các protein này đã được so sánh BLASTp với CSDL DrugBank với giá trị ngưỡng

kỳ vọng mặc định (E value  $10^{-5}$ ) nhằm đánh giá tiềm năng sử dụng làm mục tiêu thuốc. Trong số 49 protein, có 2 protein là tiểu đơn vị alpha của enzyme hô hấp khử nitrat (NarG) và protein tương đồng tubulin (FtsZ) đã được xác định là mục tiêu được phê duyệt bởi FDA (Bảng 5).

**Bảng 3: Các con đường chuyển hóa riêng biệt ở MRSA**

<b>Đường chuyển hóa chính</b>	<b>Các đường chuyển hóa tương ứng</b>	<b>Các đường chuyển hóa riêng biệt</b>	<b>Mã số KEGG</b>
Sinh tổng hợp	Chuyển hóa lipid	Sinh tổng hợp bile acid thứ cấp	sau00121
Sinh tổng hợp	Sinh tổng hợp các chất chuyển hóa thứ cấp khác	Sinh tổng hợp monobactam	sau00261
Sinh tổng hợp	Chuyển hóa các terpenoid và polyketide	Thoái hóa geraniol	sau00281
Sinh tổng hợp	Chuyển hóa amino acid	Sinh tổng hợp lysine	sau00300
Sinh tổng hợp	Thoái hóa các xenobiotic và chuyển hóa	Thoái hóa benzoate	sau00362
Sinh tổng hợp	Sinh tổng hợp các chất chuyển hóa thứ cấp khác	Sinh tổng hợp novobiocin	sau00401
Sinh tổng hợp	Chuyển hóa của các amino acid	Chuyển hóa cyanoamino acid	sau00460
Sinh tổng hợp	Chuyển hóa của các amino acid	Chuyển hóa D-alanine	sau00473
Sinh tổng hợp	Sinh tổng hợp các chất chuyển hóa thứ cấp khác	Sinh tổng hợp streptomycin	sau00521
Sinh tổng hợp	Sinh tổng hợp glycan và chuyển hóa	Sinh tổng hợp peptidoglycan	sau00550
Sinh tổng hợp	Thoái hóa các xenobiotic và chuyển hóa	Thoái hóa xylene	sau00622
Sinh tổng hợp	Thoái hóa các xenobiotic và chuyển hóa	Thoái hóa chloroalkane and chloroalkene	sau00625
Sinh tổng hợp	Thoái hóa các xenobiotic và chuyển hóa	Thoái hóa naphthalene	sau00626
Sinh tổng hợp	Chuyển hóa carbohydrate	Chuyển hóa C5-Branched dibasic acid	sau00660
Sinh tổng hợp	Chuyển hóa năng lượng	Chuyển hóa methane	sau00680
Sinh tổng hợp	Chuyển hóa các terpenoid và polyketide	Sinh tổng hợp carotenoid	sau00906
		Sinh tổng hợp chất chuyển hóa thứ cấp	sau01110
		Chuyển hóa vi khuẩn trong các môi trường đa dạng	sau01120
		Sinh tổng hợp các kháng sinh	sau01130
		Thoái hóa các hợp chất vòng thơm	sau01220
Yếu tố gây bệnh	Đề kháng thuốc	Kháng beta-lactam	sau01501
Yếu tố gây bệnh	Đề kháng thuốc	Kháng vancomycin	sau01502
Yếu tố gây bệnh	Đề kháng thuốc	Kháng cationic antimicrobial peptide (CAMP)	sau01503
Yếu tố trao đổi thông tin với môi trường	Truyền tín hiệu	Hệ thống hai thành phần	sau02020
Yếu tố trao đổi thông tin với môi trường	Vận chuyển qua màng	Hệ thống phosphotransferase (PTS)	sau02060
Yếu tố trao đổi thông tin với môi trường	Vận chuyển qua màng	Hệ thống tiết của vi khuẩn	sau03070
Chu trình tế bào	Cộng đồng tế bào prokaryote	Quorum sensing	sau02024

**Bảng 4: Các protein liên quan đến các con đường chuyển hóa riêng biệt**

Mã số protein tại CSDL UniProtKB	Mã số KO (KEGG Orthology)	Tên protein; Chức năng; Mã số EC (Enzyme Commission)	Chu trình
UPI000005480D	K01428	ureC; urease subunit alpha [EC:3.5.1.5]	Chuyển hóa của vi sinh vật trong môi trường đa dạng
UPI00003B18C5	K01429	ureB; urease subunit beta [EC:3.5.1.5]	Chuyển hóa của vi sinh vật trong môi trường đa dạng
UPI00000D77A0	K01430	ureA; urease subunit gamma [EC:3.5.1.5]	Chuyển hóa của vi sinh vật trong môi trường đa dạng
UPI0001C1200D	K00370	narG, narZ, nxrA; nitrate reductase / nitrite oxidoreductase, alpha subunit [EC:1.7.5.1 1.7.99.-]	Chuyển hóa của vi sinh vật trong môi trường đa dạng, hệ thống hai thành phần
UPI00003B14DA	K02769	PTS-Fru-EIIB, fruA; PTS system, fructose-specific IIB component [EC:2.7.1.202]	Chuyển hóa của vi sinh vật trong môi trường đa dạng, hệ thống phosphotransferase (PTS)
UPI0001C120B6	K02769	PTS-Fru-EIIB, fruA; PTS system, fructose-specific IIB component [EC:2.7.1.202]	Chuyển hóa của vi sinh vật trong môi trường đa dạng, hệ thống phosphotransferase (PTS)
UPI0001C1203D	K01921	ddl; D-alanine-D-alanine ligase [EC:6.3.2.4]	Chuyển hóa D-alanine, sinh tổng hợp peptidoglycan, kháng vancomycin
UPI000005480C	K00625	E2.3.1.8, pta; phosphate acetyltransferase [EC:2.3.1.8]	Chuyển hóa methane, chuyển hóa của vi sinh vật trong môi trường đa dạng
UPI00003B175E	K00925	ackA; acetate kinase [EC:2.7.2.1]	Chuyển hóa methane, chuyển hóa của vi sinh vật trong môi trường đa dạng
UPI00000D7773	K01624	FBA, fbaA; fructose-bisphosphate aldolase, class II [EC:4.1.2.13]	Chuyển hóa methane, sinh tổng hợp các chất chuyển hóa thứ cấp, chuyển hóa của vi sinh vật trong môi trường đa dạng, sinh tổng hợp kháng sinh
UPI00003B184D	K01546	kdpA; K+-transporting ATPase ATPase A chain [EC:3.6.3.12]	Hệ thống hai thành phần
UPI00003B12E0	K02313	dnaA; chromosomal replication initiator protein	Hệ thống hai thành phần
UPI000005449B	K07681	vraS; two-component system, NarL family, vancomycin resistance sensor histidine kinase VraS [EC:2.7.13.3]	Hệ thống hai thành phần
UPI0000054B87	K07683	nreB; two-component system, NarL family, sensor histidine kinase NreB [EC:2.7.13.3]	Hệ thống hai thành phần
UPI00003B1989	K20118	PTS-Glc1-EIIC, ptsG, glcA, glcB; PTS system, glucose-specific IIC component	Hệ thống phosphotransferase (PTS)
UPI00019CD777	K20118	PTS-Glc1-EIIC, ptsG, glcA, glcB; PTS system, glucose-specific IIC component	Hệ thống phosphotransferase (PTS)
UPI0000054B5C	K08483	PTS-El.PTSI, ptsI; phosphotransferase system, enzyme I, PtsI [EC:2.7.3.9]	Hệ thống phosphotransferase (PTS)
UPI00003B18DE	K02750	PTS-Glv-EIIC, glvC, malP, aglA; PTS system, alpha-glucoside-specific IIC component	Hệ thống phosphotransferase (PTS)

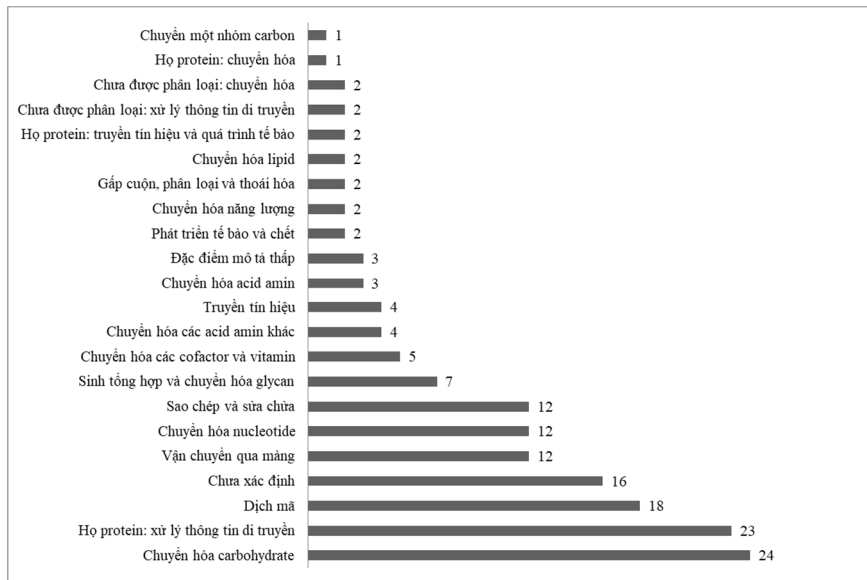
<b>Mã số protein tại CSDL UniProtKB</b>	<b>Mã số KO (KEGG Orthology)</b>	<b>Tên protein; Chức năng; Mã số EC (Enzyme Commission)</b>	<b>Chu trình</b>
UPI00000DCBDB	K02791	PTS-MalGlc-EIIC, malX; PTS system, maltose/glucose-specific IIC component	Hệ thống phosphotransferase (PTS)
UPI0000696C65	K02804	PTS-Nag-EIIC, nagE; PTS system, N-acetylglucosamine-specific IIC component	Hệ thống phosphotransferase (PTS)
UPI00000B9AC7	K02755	PTS-Bgl-EIIA, bglF, bglP; PTS system, beta-glucoside-specific IIA component [EC:2.7.1.-]	Hệ thống phosphotransferase (PTS)
UPI0001AE8711	K02777	PTS-Glc-EIIA, crr; PTS system, sugar-specific IIA component [EC:2.7.1.-]	Hệ thống phosphotransferase (PTS)
UPI00003B187B	K02798	PTS-Mtl-EIIA, mtlA, cmtB; PTS system, mannitol-specific IIA component [EC:2.7.1.197]	Hệ thống phosphotransferase (PTS)
UPI0000054BFA	K03076	secY; preprotein translocase subunit SecY	Quorum sensing, hệ thống tiết của vi khuẩn
UPI00003B14FC	K00370	secA; preprotein translocase subunit SecA	Quorum sensing, hệ thống tiết của vi khuẩn
UPI0001C120BC	K00370	PTS-GlcI-EIIC, ptsG, glcA, glcB; PTS system, glucose-specific IIC component	Quorum sensing, hệ thống tiết của vi khuẩn
UPI0005ADADD6	K03076	secY; preprotein translocase subunit SecY	Quorum sensing, hệ thống tiết của vi khuẩn
UPI0001AB1796	K03621	plsX; glycerol-3-phosphate acyltransferase PlsX [EC:2.3.1.15]	Sinh tổng hợp các chất chuyển hóa thứ cấp
UPI00000601EC	K21064	ycsE, yitU, ywtE; 5-amino-6-(5-phospho-D-ribitylamino)uracil phosphatase [EC:3.1.3.104]	Sinh tổng hợp các chất chuyển hóa thứ cấp
UPI00000549F0	K08591	plsY; glycerol-3-phosphate acyltransferase PlsY [EC:2.3.1.15]	Sinh tổng hợp các chất chuyển hóa thứ cấp
UPI000012AF6D	K01810	GPI, pgi; glucose-6-phosphate isomerase [EC:5.3.1.9]	Sinh tổng hợp các chất chuyển hóa thứ cấp, chuyển hóa của vi sinh vật trong môi trường đa dạng, sinh tổng hợp kháng sinh
UPI00003B173F	K01845	hemL; glutamate-1-semialdehyde 2,1-aminomutase [EC:5.4.3.8]	Sinh tổng hợp các chất chuyển hóa thứ cấp, chuyển hóa của vi sinh vật trong môi trường đa dạng
UPI00003B1663	K01696	trpB; tryptophan synthase beta chain [EC:4.2.1.2]	Sinh tổng hợp các chất chuyển hóa thứ cấp, sinh tổng hợp kháng sinh
UPI0001AE9B80	K01809	manA, MPI; mannose-6-phosphate isomerase [EC:5.3.1.8]	Sinh tổng hợp các chất chuyển hóa thứ cấp, sinh tổng hợp kháng sinh
UPI00000D7844	K01809	manA, MPI; mannose-6-phosphate isomerase [EC:5.3.1.8]	Sinh tổng hợp các chất chuyển hóa thứ cấp, sinh tổng hợp kháng sinh
UPI00003B1660	K01658	trpG; anthranilate synthase component II [EC:4.1.3.27]	Sinh tổng hợp các chất chuyển hóa thứ cấp, sinh tổng hợp kháng sinh, quorum sensing
UPI00003B1436	K04042	glmU; bifunctional UDP-N-acetylglucosamine pyrophosphorylase / Glucosamine-1-phosphate N-acetyltransferase [EC:2.7.7.23 2.3.1.157]	Sinh tổng hợp kháng sinh

Mã số protein tại CSDL UniProtKB	Mã số KO (KEGG Orthology)	Tên protein; Chức năng; Mã số EC (Enzyme Commission)	Chu trình
UPI00000DCBDB	K05362	murE; UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-L-lysine ligase [EC:6.3.2.7]	Sinh tổng hợp peptidoglycan
UPI000012F989	K01925	murD; UDP-N-acetylmuramoylalanine--D-glutamate ligase [EC:6.3.2.9]	Sinh tổng hợp peptidoglycan
UPI0000167807	K11694	femA; peptidoglycan pentaglycine glycine transferase (the second and third glycine) [EC:2.3.2.17]	Sinh tổng hợp peptidoglycan
UPI0002550F93	K01924	murC; UDP-N-acetylmuramate--alanine ligase [EC:6.3.2.8]	Sinh tổng hợp peptidoglycan
UPI00003B18AF	K11693	femX, fmhB; peptidoglycan pentaglycine glycine transferase (the first glycine) [EC:2.3.2.16]	Sinh tổng hợp peptidoglycan
UPI00003B185E	K00790	murA; UDP-N-acetylglucosamine 1-carboxyvinyltransferase [EC:2.5.1.7]	Sinh tổng hợp peptidoglycan
UPI00000549D5	K11695	femB; peptidoglycan pentaglycine glycine transferase (the fourth and fifth glycine) [EC:2.3.2.18]	Sinh tổng hợp peptidoglycan
UPI0001AE8448	K00790	murA; UDP-N-acetylglucosamine 1-carboxyvinyltransferase [EC:2.5.1.7]	Sinh tổng hợp peptidoglycan
UPI00003B14CB	K06153	bacA; undecaprenyl-diphosphatase [EC:3.6.1.27]	Sinh tổng hợp peptidoglycan
UPI00003B1851	K01929	murF; UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase [EC:6.3.2.10]	Sinh tổng hợp peptidoglycan, kháng vancomycin
UPI0001C11E0E	K02563	murG; UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase [EC:2.4.1.227]	Sinh tổng hợp peptidoglycan, kháng vancomycin
UPI00000543CB	K01000	mraY; phospho-N-acetylmuramoyl-pentapeptide-transferase [EC:2.7.8.13]	Sinh tổng hợp peptidoglycan, kháng vancomycin

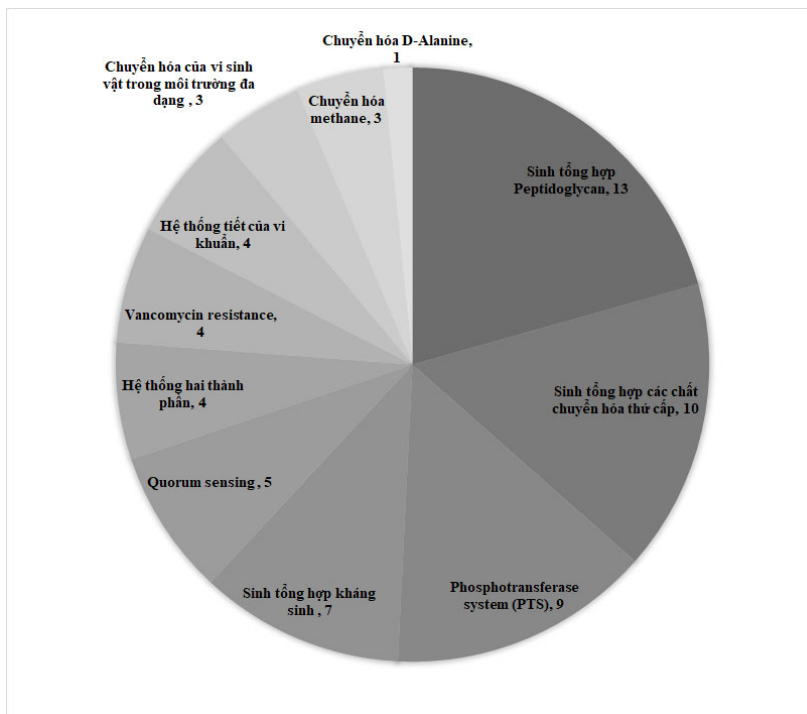
**Bảng 5: Các mục tiêu thuốc giả định tốt nhất ở MRSA xác định bởi CSDL DrugBank**

Mã số protein tại CSDL UniProtKB	Tên protein	Tên thuốc (tiềm năng)	Mã số DrugBank	Nhóm thuốc	Tác động dược lý
UPI0001C1200D	Chuỗi alpha enzyme hô hấp khử nitrat (NarG)	[(2S)-2,3-dihydroxypropyl] [(2R)-2-octanoyloxy-3-pentanoyloxypropyl] phosphate	DB07349	Thử nghiệm	Chưa biết
		N-Formylmethionine	DB04464	Thử nghiệm	Chưa biết
UPI00000543C7	Protein phân chia tế bào FtsZ	Phosphomethylphosphonic Acid Guanylate Ester	DB03532	Thử nghiệm	Chưa biết
		Guanosine-5'-Diphosphate	DB04315		
		5'-Guanosine-Diphosphate-Monothiophosphate	DB01864	Thử nghiệm	Chưa biết
		Citric Acid	DB04272	Cho phép sử dụng trong dược và thú y	Chưa biết





Hình 2: Sự tham gia của 158 protein thiết yếu vào các con đường chuyển hóa ở MRSA



Hình 3: Các protein thiết yếu tham gia vào các con đường chuyển hóa riêng biệt ở MRSA (một protein có thể tham gia vào nhiều hơn một con đường)

*Tiểu đơn vị alpha của enzyme hô hấp khử nitrat (NarG)*

Những thành tựu trong nghiên cứu đặc điểm sinh hóa và di truyền của khử nitrat đã cho thấy sự phức tạp của quá trình này và ý nghĩa đặc biệt của nó đối với chu trình nitơ. Cấu trúc của các enzyme khử nitrat cũng khác nhau giữa tế bào prokaryote và tế bào eukaryote. Ở tế bào prokaryote đã có ba hệ

thống khử nitrat khác nhau ở vị trí, cấu trúc, đặc tính sinh hóa, sự điều hòa và tổ chức gen được mô tả bao gồm: đồng hóa tế bào chất (Nas), hô hấp liên kết màng (Nar) và dị hóa tế bào chất (Nap). Các enzyme hô hấp khử nitrat liên kết màng cho phép tổng hợp ATP bằng cách sử dụng nitrat như chất nhận điện tử thay thế trong điều kiện yếm khí (Moreno-Vivián *et al.*, 1999). Ở *S. aureus*, enzyme hô hấp khử nitrat liên kết màng được mã hóa bởi các gen *narGHJ*.

Hệ thống Nar được cảm ứng bởi sự có mặt của nitrat và bị ức chế bởi oxy. NarG là tiểu đơn vị alpha có chức năng xúc tác, NarH là tiểu đơn vị beta với bốn trung tâm sắt – lưu huỳnh và NarI là tiểu đơn vị gamma quinol oxy hóa. NarJ không phải là một phần của phức hợp enzyme, nhưng nó tham gia vào việc lắp ráp hai tiểu đơn vị alpha - beta. Sự tạo thành nitric oxide (NO) là một cơ chế quan trọng trong cơ chế đề kháng của vật chủ đối với tác nhân gây bệnh. Sự hiện diện của *S. aureus* kích thích sự sản xuất NO bởi thực bào ở người, mà đã được quan sát thấy ở những người bị nhiễm tụ cầu. Tuy nhiên, *S. aureus* có khả năng kháng lại tác dụng kháng khuẩn của NO thông qua phản ứng nitro hóa, trong đó bao gồm sự chuyển đổi từ chuyển hóa hiếu khí sang kỵ khí. Khả năng này của *S. aureus* khi có mặt NO là một đặc điểm phân biệt so với nhiều loài gây bệnh khác (Holman, 2011). Ngoài ra, enzyme này đôi khi có thể có nhiều vai trò khác nhau dưới những điều kiện trao đổi chất khác nhau và các con đường đồng hóa, hô hấp và dị hóa có thể được kết nối với nhau để tạo điều kiện thích nghi nhanh chóng với sự thay đổi điều kiện nitơ và/hoặc oxy, giúp tăng sự linh hoạt của vi khuẩn để sống sót trong môi trường tự nhiên (Moreno-Vivián *et al.*, 1999). Do đó, quá trình hô hấp khử nitrat và vai trò của NarG trong con đường này có thể là mục tiêu tiềm năng cho việc thiết kế thuốc dựa trên mục tiêu vì ở vi khuẩn, đường chuyển hóa và enzyme này là duy nhất.

#### Protein tương đồng tubulin (FtsZ)

Protein tương đồng tubulin (FtsZ) là một protein cần thiết cho quá trình phân chia tế bào ở tế bào prokaryote. FtsZ được xem là protein đầu tiên xuất hiện tại điểm phân chia và thúc đẩy sự hoạt động của các protein khác liên quan đến sự phân chia tế bào vi khuẩn. Bên cạnh việc hoạt động như một giá đỡ cho các protein phân chia tế bào khác, bản thân FtsZ cũng có thể gây ra lực cytokinetic dẫn đến sự phân chia tế bào (Erickson *et al.*, 1996). FtsZ có hoạt tính GTPase polymer hóa phụ thuộc nucleotide, hoạt động theo hướng đầu-đuôi để tạo thành sợi đơn mà sau đó được lắp ráp hoàn chỉnh thành một vòng gọi là Z-ring. Vòng này giúp đánh dấu vị trí tương lai của vách ngăn tế bào vi khuẩn khi phân chia và được duy trì trong suốt quá trình phân chia tế bào bởi sự đảo chiều liên tục của các polymer do FtsZ tạo ra (Mingorance *et al.*, 2010). Vì vai trò thiết yếu của FtsZ trong quá trình phân chia tế bào vi khuẩn, gần đây protein này đã trở thành một mục tiêu tiềm năng cho việc thiết kế và tìm kiếm các kháng sinh mới và các tác nhân chống vi khuẩn (Schaffner-Barbero *et al.*, 2011). Sự gia tăng của các mầm bệnh kháng kháng sinh và thiếu sự phát triển kháng sinh mới đã làm nổi bật yêu cầu tìm kiếm các hoạt chất chống lại các mục tiêu mới như protein phân chia tế bào vi

khẩn (Kumar *et al.*, 2010). Do tính mới của nó khi được sử dụng làm mục tiêu thuốc, các hợp chất nhắm mục tiêu FtsZ sẽ ít có khả năng hoặc không bị ảnh hưởng bởi các cơ chế đề kháng hiện tại của vi khuẩn. Sự sàng lọc hợp chất tiềm năng với protein mục tiêu FtsZ ban đầu tập trung vào các loại thuốc ức chế trùng hợp tubulin do sự giống nhau về chức năng và cấu trúc giữa tubulin và FtsZ (de Pereda *et al.*, 1996). Tuy nhiên, có rất ít nguy cơ rằng một hợp chất chống FtsZ cũng sẽ chống tubulin vì FtsZ và tubulin chỉ chia sẻ 10 - 18% mức độ tương đồng trình tự (Huang *et al.*, 2007).

Các protein còn lại chưa được báo cáo để sử dụng làm mục tiêu thuốc chống lại MRSA. Tuy nhiên, vì sự tham gia của các protein này vào những chu trình tế bào quan trọng và không tương đồng với protein vật chủ, sự ức chế của 47 protein này cũng có thể được sử dụng để chống lại MRSA. Đã có những nghiên cứu cho thấy phân tử thuốc không nhất thiết phải liên kết với toàn bộ protein để thể hiện hoạt tính của nó. Sự gắn kết của một phân tử thuốc với một protein đôi khi chỉ cần vùng trình tự acid amin của vị trí gắn với thụ thể (Hossain *et al.*, 2017). Vì vậy có thể thực hiện những biến đổi nhỏ ở phân tử thuốc đã có để mang lại hoạt tính với các protein đã lựa chọn. Điều này đem lại lợi thế là có thể sử dụng các phân tử thuốc đã được chứng minh hiệu quả cho những vi khuẩn khác có protein mục tiêu tương đồng về số lượng, vị trí acid amin và cấu trúc không gian với các protein mục tiêu đã được xác định trong nghiên cứu này. Do đó, 47 protein trên nên được xem như các mục tiêu thuốc giả định tiềm năng cho những nghiên cứu trong tương lai.

#### 4 KẾT LUẬN

Trong nghiên cứu này, phương pháp tiếp cận so sánh bộ proteome giữa mầm bệnh và vật chủ dựa trên các công cụ tin sinh học nhằm xác định mục tiêu thuốc hợp lý là một chiến lược hiệu quả cho việc phát triển thuốc, với lợi thế tiết kiệm thời gian và chi phí. Bằng cách áp dụng phương pháp trên, hai mục tiêu thuốc tốt nhất đã được xác định ở 15 chủng MRSA, cụ thể là tiểu đơn vị alpha của enzyme hô hấp khử nitrat (NarG) và protein tương đồng tubulin (FtsZ), trong khi đó 47 protein được tìm thấy là những mục tiêu thuốc giả định tiềm năng. Các nghiên cứu sâu hơn có thể được thực hiện để xác định các hợp chất có khả năng ức chế các protein mục tiêu này, từ đó có thể tăng tốc quá trình phát triển của các tác nhân trị liệu mới chống lại MRSA.

#### LỜI CẢM ƠN

Nghiên cứu được tài trợ bởi Trường Đại học Bách khoa – Đại học Quốc gia TP. HCM trong khuôn khổ đề tài mã số TNCS-KTHH-2017-12.

## TÀI LIỆU THAM KHẢO

- Boratyn, G.M., Camacho, C., Cooper, P.S., et al., 2013. BLAST: a more efficient report with usability improvements. *Nucleic Acids Research*. 41(W1): W29-W33.
- Cordwell, S.J., Larsen, M.R., Cole, R.T. and Walsh, B.J., 2002. Comparative proteomics of *Staphylococcus aureus* and the response of methicillin-resistant and methicillin-sensitive strains to Triton X-100. *Microbiology*. 148(9): 2765-2781.
- de Pereda, J.M., Leynadier, D., Evangelio, J.A., Chacón, P. and Andreu, J.M., 1996. Tubulin secondary structure analysis, limited proteolysis sites, and homology to FtsZ. *Biochemistry*. 35(45): 14203-14215.
- Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 26(19): 2460-2461.
- Erickson, H.P., Taylor, D.W., Taylor, K.A. and Bramhill, D., 1996. Bacterial cell division protein FtsZ assembles into protofilament sheets and minirings, structural homologs of tubulin polymers. *Proceedings of the National Academy of Sciences of the United States of America*. 93(1): 519-523.
- Haag, N.L., Velk, K.K. and Wu, C., 2012. Potential antibacterial targets in bacterial central metabolism. *International Journal on Advances in Life Sciences*. 4(1-2): 21–32.
- Hasan, M.A., Alauddin, S.M., Al-Amin, M., Nur, S.M. and Mannan, A., 2016. Identification of putative drug targets in vancomycin-resistant *Staphylococcus aureus* (VRSA) using computer aided protein data analysis. *Gene*. 575(1): 132-143.
- Holman, S.E., 2011. Effect of the nitrate reductase operon on *Staphylococcus aureus* biofilm formation. MSc dissertation. Graduate School, University of Florida, Florida, United States of America, 77 pages.
- Hossain, T., Kamruzzaman, M., Choudhury, T.Z., et al., 2017. Application of the subtractive genomics and molecular docking analysis for the identification of novel putative drug targets against *Salmonella enterica* subsp. *enterica* serovar Poona. *BioMed Research International*. Vol. 2017, Article ID 3783714, 9 pages.
- Hossain, M., Chowdhury, D.U.S., Farhana, J., Akbar, M.T., Chakraborty, A., Islam, S. and Mannan, A., 2013. Identification of potential targets in *Staphylococcus aureus* N315 using computer aided protein data analysis. *Bioinformation*. 9(4): 187–192.
- Huang, Q., Tonge, P.J., Slayden, R.A., Kirikae, T. and Ojima, I., 2007. FtsZ: A novel target for tuberculosis drug discovery. *Current Topics in Medicinal Chemistry*. 7(5): 527-543.
- Judson, N. and Mekalanos, J.J., 2000. TnAraOut, A transposon-based approach to identify and characterize essential bacterial genes. *Nature Biotechnology*. 18(7): 740-745.
- Knox, C., Law, V., Jewison, T., et al., 2011. DrugBank 3.0: a comprehensive resource for “Omics” research on drugs. *Nucleic Acids Research*. 39(Database issue): D1035–D1041.
- Kumar, K., Awasthi D., Berger, W.T., Tonge, P.J., Slayden, R.A. and Ojima, I., 2010. Discovery of anti-TB agents that target the cell-division protein FtsZ. *Future Medicinal Chemistry*. 2(8): 1305-1323.
- Mingorance, J., Rivas, G., Vélez, M., Gómez-Puertas, P. and Vicente, M., 2010. Strong FtsZ is with the force: mechanisms to constrict bacteria. *Trends in Microbiology*. 18(8): 348–356.
- Moreno-Vivián, C., Cabello, P., Martínez-Luque, M., Blasco, R. and Castillo, F., 1999. Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *Journal of Bacteriology*. 181(21): 6573-6584.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C. and Kanehisa, M., 2007. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*. 35(Web Server issue): W182-W185.
- Morya, V.K., Dewaker, V., Mccarty, S.D. and Singh, R., 2010. In silico analysis metabolic pathways for identification of putative drug targets for *Staphylococcus aureus*. *Journal of Computer Science and Systems Biology*. 3(3): 062-069.
- Patnala, K. and Zaveri, K., 2016. Screening of putative therapeutic candidates in superbug (*Staphylococcus aureus*): a systematic in silico approach. *Asian Journal of Pharmaceutical and Clinical Research*. 9(14): 283-291.
- Rathi, B., Sarangi, A.N. and Trivedi, N., 2009. Genome subtraction for novel target definition in *Salmonella typhi*. *Bioinformation*. 4(4): 143-150.
- Schaffner-Barbero, C., Martín-Fontecha, M., Chacón, P. and Andrew, J.M., 2011. Targeting the assembly of bacterial cell division protein FtsZ with small molecules. *ACS Chemical Biology*. 7(2): 269-277.
- Stapleton, P.D. and Taylor, P.W., 2002. Methicillin resistance in *Staphylococcus aureus*: mechanisms and modulation. *Science Progress*. 85(Pt1): 57–72.
- UniProt Consortium, T., 2018. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*. 46(5): 2699.
- Zhang, R., Ou, H.Y. and Zhang, C.T., 2004. DEG: A database of essential genes. *Nucleic Acids Research*. 32(Database issue): D271-D272.