**O-ORT-05**

# Computational approaches to predict human protein targets of Spirulina Compounds for the Treatment of Systemic Lupus Erythematosus

**Han Ping[1], Teeraphan Laomettachit[1], Marasri Ruengjitchatchawalya[1,2]**

[1] *Bioinformatics and Systems Biology program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkok, 10150, Thailand*

[2] *Biotechnology program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkok, 10150, Thailand*

**Corresponding author. E-mail: marasri.rue@kmutt.ac.th*
**DOI:**

## ABSTRACT

Systemic lupus erythematosus (SLE) is a heterogeneous autoimmune disease characterized by autoantigen exposure, autoantibody production, chronic inflammation and tissue damage. The patients can be affected by SLE in heterogeneous severity ranging from skin and joint pain to heart attack, neuro-disorder, and even death. Unfortunately, there is no specific treatment for SLE at present. Therefore, it is necessary to identify more potential targets. From a previous study, potential bioactive compounds from *Arthrospira Platensis* C1 were proposed to target SLE proteins. These compounds are natural supplement and not associated with toxicity.

In the present study, we utilized multiple computational approaches with the aim to identify potential SLE-associated protein targets of those Spirulina compounds. First, 471 known SLE genes were used as an input to construct an SLE association protein network. A random walk with restart algorithm was then applied to the network to prioritize proteins that are highly associated with the disease, resulting in 44 protein candidates. The 44 candidates were then BLASTPed against SLE protein targets of Spirulina compounds previously reported. Based on the concept that targets with similar sequences tend to bind to the same compounds, we predicted eight SLE proteins as a target of three Spirulina compounds. This prediction was then preliminarily validated by molecular docking technique. Our results

propose phycocyanobilin, a highly available compound in Spirulina, and other compounds as potential therapeutic agents that target SLE proteins, including LYN, LCK, FYN and EGFR.

Keywords: Bioactive Compounds, Spirulina, Systemic Lupus Erythematosus, Random walk with restart, Drug target prediction

## INTRODCUTION

Systemic lupus erythematosus (SLE) is a heterogeneous autoimmune disease characterized by antinuclear autoantibodies. This disease can cause severe organ damage, and it companies with autoantigen exposure, autoantibody production, chronic inflammation and tissue damage. The prevalence of SLE is around 20 to 150 out of 100,000 [1], and the incidence is in the range of one to 10 out of 100,000. Both prevalence and incidence have been risen very quickly in recent years. SLE mainly affects women rather than men (ratio of 9:1) [2]. Associated genes of SLE patients show polymorphic features, which leads to great functional diversity of proteins. The high variety of associated genes also account for the complicated pathogenesis and clinical heterogeneity of SLE. The patients can suffer from skin and joint symptoms, heart attack and neuro disorder to other long-term complications [3]. Unfortunately, there is no specific treatment for SLE at present. The immunosuppressive drugs are used to control the symptoms of patients, but they are associated with drug toxicities and side effects. Therefore, it is very urgent to find other potential drugs to alleviate the disease.

In a previous study, Chaiprasert A. *et al*. (manuscript in preparation) identified several bioactive compounds from Spirulina (*Arthrospira platensis* C1) that show potential benefit to human immunologic and SLE conditions. The authors also predicted some target proteins of the compounds for SLE immunologic treatment using network-based methods.

In the current work, we combined multiple computational approaches, including disease-associated protein network reconstruction, random walk with restart to prioritize proteins, target similarity method for drug-target prediction, and molecular docking, to identify potential SLE protein targets of Spirulina compounds. Predicted interactions can aid experimental design and; therefore, reduce time consumption and cost of the experiments. Our results propose phycocyanobilin, a highly available compound in Spirulina, and other compounds as potential therapeutic agents that target SLE proteins, including LYN, LCK, FYN and EGFR**.**

## MATERIAL AND METHODS

### Materials

Known 471 SLE genes were retrieved from VisANT 4.0 (http://visant.bu.edu/visantnet.html), which is a tool for drug-target network construction based on the Predictome database. Protein sequences of 471 SLE genes were retrieved from the Uniprot (450 sequences) and NCBI (6 sequences) database. The other 15 protein sequences were not available from both databases.

### Spirulina bioactive compounds predicted from structure similarity based method

In a previous work, Chaiprasert A. *et al*. (manuscript in preparation) used structure similarity based method to search for bioactive compounds from Spirulina (*Arthrospira platensis* C1) that have similar chemical structures to SLE immunosuppressive drugs. The identified Spirulina compounds were assumed to share the same targets with the SLE drugs. Six predicted Spirulina compounds and eight predicted SLE targets are shown in Table 1.

**Table 1.** Six predicted Spirulina compounds and eight predicted SLE targets from a previous work

| Spirulina bioactive compounds | SLE targets |
|---|---|
| 1. Cytidine | NCOA3, AR |
| 2. Dl-Dethiobiotin | RIN1 |
| 3. Anthranilic acid | JUN |
| 4. Agmatine | TDP1 |
| 5. Lomustine | STMN4, IL4I1 |
| 6. Phycocyanobilin | MAPK14 |

### Method

In the present work, we used target based similarity method to identify the targets of Spirulina compounds. The workflow of this work is illustrated in Figure 1.

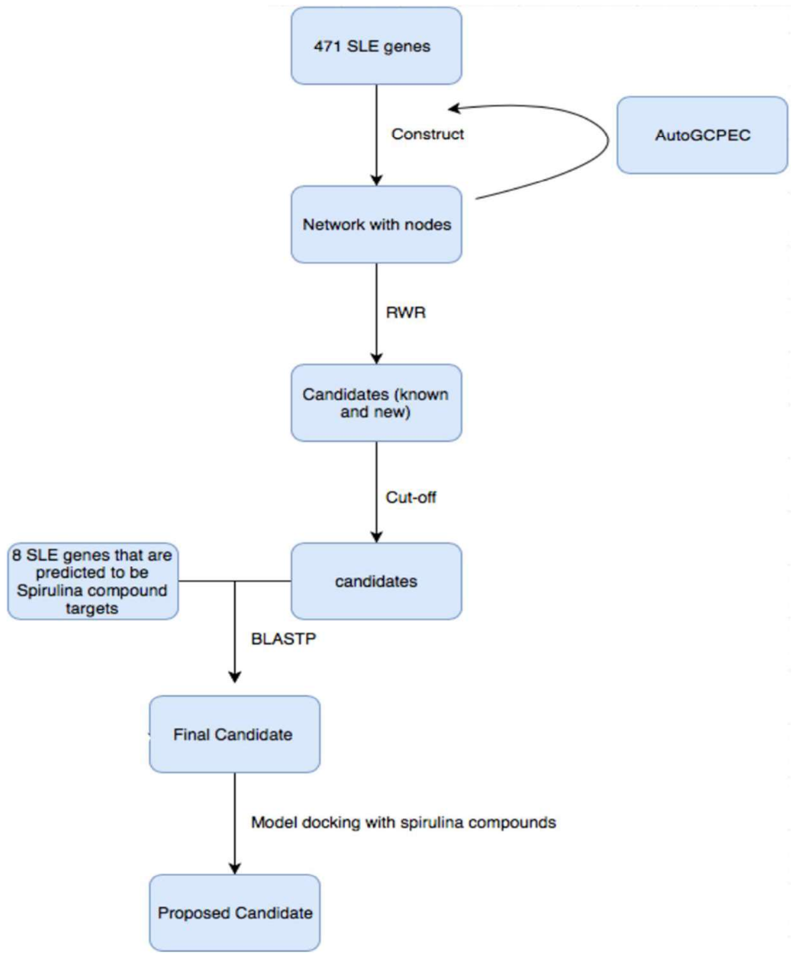


**Figure 1.** Workflow of predicting targets for Spirulina compounds in SLE treatment.

## Data preparation

Protein sequences of 471 known SLE genes were retrieved from Uniprot and NCBI databases in FASTA format.

## AutoHGPEC Network reconstruction

After the sequences of SLE proteins were retrieved, AutoHGPEC was used to perform protein prioritization (ranking). As a Cytoscape plugin, AutoHGPEC is a tool for prediction of the disease-associated genes/proteins by using random walk with restart algorithm. First, the tool constructed a protein-protein interaction network associated with SLE disease based on our 471 SLE input proteins.

Second, a random walk with restart algorithm was performed to calculate a score for each protein in the network. The scores were used to prioritize proteins highly associated with SLE disease.

After the prioritization, the list of protein ranks was collected and only proteins with the score above 5E-4 (Lei Chen, 2013) were selected and assigned as SLE protein targets. This enabled us to narrow down the size of candidate targets to use in the further steps.

**Target based similarity based identification of Spirulina compound targets**

The selected SLE protein targets were BLASTPed against those SLE protein targets identified in a previous work (see Table 1). The resulting hits with high similarity were considered as the final potential SLE targets of Spirulina compounds for SLE treatment.

**Molecular Docking**

To validate our approach, we performed molecular docking to check the binding affinity score between Spirulina compounds (ligand) and final candidate targets (receptor). The goal of the docking was to find the best binding model of ligand with the receptor. Both MGLTool and PyRx were used for molecular docking. Both of the tools applied AutoDOCK in the docking process. The input of MGLTool and PyRx is the 3D structures of compounds, which were retrieved from PubChems, and the 3D structures of receptor proteins, which were retrieved from RCSB PD (https://www.rcsb.org).

## RESULTS AND DISSCUSION

**Visualization of SLE protein association network**

The SLE protein association network constructed by AutoHGPEC contains 10,486 nodes and is shown in Figure 2.

**Figure 2.** Visualization of the SLE protein association network constructed by AutoHGPEC

**Random walk with restart for protein ranking**

AutoHGPEC performed random walk with restart algorithm to prioritize proteins in the network from the training set of known 471 SLE proteins provided by the authors. We set a cutoff of 5E-4 to filter only proteins with high association with SLE to be considered as SLE protein targets. In total there are 44 proteins that passed the cutoff. Table 2 shows top 10 proteins that have highest SLE association scores. The symbol represents names of that candidate proteins. The alternate symbol represents alternate names. The Boolean value of the training column represents whether the proteins were included in the training set (the 471 known SLE proteins). The proteins labeled with FASLE in this column might indicate novel SLE protein targets. There are 17 such novel candidate targets identified from our analysis.

**Table 2.** The top 10 SLE proteins with highest SLE association scores reported by AutoHGPEC. (Training column, TRUE represents the gene is including in training set, FALSE represents that gene is not including in training set. Candidate column, TRUE represents the gene is from the 471 input genes, FALSE represents that gene is from the network.)

| Rank | Symbol | Alternate Symbol | Training | Candidate | Score |
|---|---|---|---|---|---|
| 1 | C2 | DADB-122G4.1, CO2, DKFZp779M0311 | TRUE | TRUE | 0.00306315 |
| 2 | C1R | | FALSE | FALSE | 0.00225846 |
| 3 | CYBB | CGD, GP91-1, GP91-PHOX, GP91PHOX, NOX2, p91-PHOX | FALSE | FALSE | 0.00188202 |
| 4 | PDCD1 | CD279, PD1, SLEB2, hPD-1, hPD-l | TRUE | TRUE | 0.00072992 |
| 5 | C3 | AHUS5, ARMD9, ASP, CPAMD1 | TRUE | TRUE | 0.00067582 |
| 6 | HLA-C | XXbac-BCX101P6.2, D6S204, FLJ27082, HLA-Cw, HLA-JY3, HLC-C, PSORS1 | TRUE | TRUE | 0.00065285 |
| 7 | IL10 | RP11-262N9.1, CSIF, IL-10, IL10A, MGC126450, MGC126451, TGIF | TRUE | TRUE | 0.00064924 |
| 8 | TP53 | FLJ92943, LFS1, P53, TRP53 | TRUE | TRUE | 0.00063529 |
| 9 | HLA-DRB1 | XXbac-BPG161M6.1, DRB1, DRw10, FLJ75017, FLJ76359, HLA-DR1B, HLA-DRB, HLA-DRB1*, SS1 | TRUE | TRUE | 0.00062342 |
| 10 | TRAF6 | MGC:3310, RNF85 | TRUE | FALSE | 0.00061036 |

**Target based similarity identification of Spirulina compound targets**

After the 44 SLE highly related proteins were collected, we used BLASTP to compare sequence similarity of the 44 SLE proteins with the eight SLE proteins previously reported as targets of Spirulina compounds (see Table 1). If the candidates from AutoHGPEC have sequences which are highly similar with one of the eight reference targets, the candidates are assumed to potentially be a target of the Spirulina compound as well. We used the BLAST E-value cutoff of 0.01, resulting in eight SLE targets matching with three Spirulina compounds (Table 3). The eight protein targets (HLA-B, TP53, LCK, LYN, BLK, EGFR, FYN and ESR1) are considered as the final SLE protein candidates. Among the eight candidates, six of them potentially bind to phycocyanobilin, a highly available compound in Spirulina.

**Table 3.** BLASTP results showing final SLE protein candidates and their Spirulina compound ligands

| GENE | Bit Score | E value | Identity | Rank score | Reference | Spirulina Compound |
|---|---|---|---|---|---|---|
| HLA-B | 23.9 | 0.002 | 15/46(33%) | 0.00060461 | STMN4 | Lomustine |
| TP53 | 24.6 | 0.004 | 12/35(34%) | 0.00063529 | MAPK14 | Phycocyanobilin |
| LCK | 84.7 | 3E-22 | 70/235(30%) | 0.0004677 | MAPK14 | Phycocyanobilin |
| LYN | 84.3 | 3E-22 | 68/226(30%) | 0.00046506 | MAPK14 | Phycocyanobilin |
| BLK | 89.0 | 9E-24 | 70/254(28%) | 0.00046022 | MAPK14 | Phycocyanobilin |
| EGFR | 57.4 | 6E-13 | 60/212(28%) | 0.0004584 | MAPK14 | Phycocyanobilin |
| FYN | 82.5 | 2E-21 | 69/234(29%) | 0.0004529 | MAPK14 | Phycocyanobilin |
| ESR1 | 156 | 3E-44 | 99/357(28%) | 0.00056875 | AR | Cytidine |

**Molecular Docking**

The last step is to preliminarily validate our predicted interactions between the eight final protein candidates and their potential ligands from Spirulina compounds using molecular docking. Figure 3 shows relative docking energy between the targets and the compounds provided by AutoDOCK. It can be seen that EGFR, FYN, LCK, and LYN have good docking energy scores with Phycocyanobilin. HLA-B also has a good docking energy score with Lomustine. More details of docking results between LCK, LYN, EGFR and Phycocyanobilin are shown in Tables 4, 5, and 6, respectively.

To reconfirm our docking results, two more docking tools (iGEMDOCK and FlexX) were used to calculate the binding affinity. The results showed similar patterns as reported above.

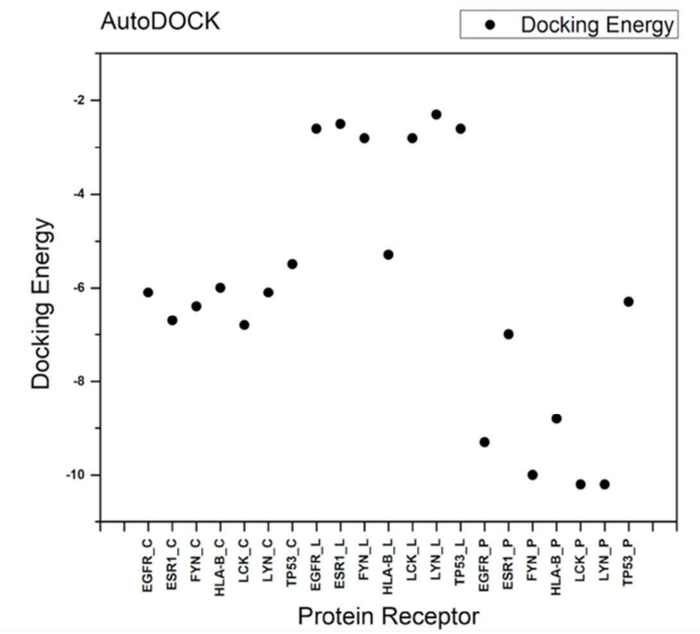


**Figure 3.** AutoDock results **(**the x**-**axis presents the protein receptor and sprulina compound short name**.** C represents Cytidne; L represents Lomustine; P represents Phycocyanobilin**)**

**Table 4.** Docking results between Phycocyanobilin and LCK by MGLTool

| Mode | Affinity (kcal/mol) | Dist from rmsd l.b. | Best mode rmsd u.b. |
|---|---|---|---|
| 1 | -9.3 | 0.000 | 0.000 |
| 2 | -9.1 | 2.654 | 4.498 |
| 3 | -9.1 | 2.575 | 4.862 |
| 4 | -9.0 | 3.189 | 5.439 |
| 5 | -9.0 | 2.341 | 4.861 |
| 6 | -8.8 | 2.989 | 4.889 |
| 7 | -8.8 | 2.655 | 3.612 |
| 8 | -8.8 | 2.613 | 5.051 |
| 9 | -8.7 | 11.778 | 14.748 |

**Table 5.** Docking results between Phycocyanobilin and LYN by MGLTool

| Mode | Affinity (kcal/mol) | Dist from rmsd l.b. | Best mode rmsd u.b. |
|---|---|---|---|
| 1 | -10.0 | 0.000 | 0.000 |
| 2 | -10.0 | 1.846 | 2.591 |
| 3 | -9.8 | 1.456 | 2.410 |
| 4 | -9.8 | 3.221 | 9.998 |
| 5 | -9.8 | 3.503 | 6.501 |
| 6 | -9.7 | 2.494 | 9.708 |
| 7 | -9.7 | 1.523 | 10.312 |
| 8 | -9.4 | 1.394 | 1.812 |
| 9 | -9.3 | 3.144 | 6.435 |

**Table 6.** Docking results between Phycocyanobilin and EGFR by MGLTool

| Mode | Affinity (kcal/mol) | Dist from rmsd l.b. | Best mode rmsd u.b. |
|---|---|---|---|
| 1 | -9.3 | 0.000 | 0.000 |
| 2 | -9.0 | 4.186 | 10.165 |
| 3 | -8.8 | 3.572 | 8.991 |
| 4 | -8.8 | 4.795 | 10.913 |
| 5 | -8.8 | 2.891 | 9.689 |
| 6 | -8.6 | 4.634 | 9.317 |
| 7 | -8.6 | 9.999 | 16.380 |
| 8 | -8.6 | 3.916 | 9.814 |
| 9 | -8.6 | 16.922 | 21.852 |

## CONCLUSION

The predicted therapeutic targets ESR1, HLA-B, FYN, LCK, LYN, EGFR and TP53 have the potential to become the targets of bioactive compounds form *Arthrospira platensis* C1. From the molecular docking results, the LYN, LCK, FYN and EGFR achieved good docking scores with Phycocyanobilin. Phycocyanobilin has the potential biological action on SLE. The Bilirubin is homologue phycocyanobilin and the low serum bilirubin level is associated with SLE disease. SLE is accompanied by severe oxidative stress while the bilirubin has a key role in regulation of oxidative stress. Therefore, the high serum bilirubin level might mitigate the development of SLE. As the homologue with bilirubin, phycocyanobilin might have the high potential to become the therapeutic drug for SLE targets. The docking score of ESR1 and Cytidine also indicates relatively good interaction. ESR1 also has been reported to be activated in response to 5-Aza-2′-deoxycytidineis (a Cytidine analog).

Our docking results indicate that our approach potentially identifies SLE targets of the Spirulina compounds and supports the potential use of Spirulina compounds as therapeutic agents. Our predicted interactions between SLE proteins and the compounds, however, require further investigation using, for example, molecular dynamics simulation or laboratory validation.

## REFERENCES

Oon, S., Wilson, N. J., & Wicks, I. (2016). Targeted therapeutics in SLE: emerging strategies to modulate the interferon pathway. *Clinical & Translational Immunology, 5*(5), e79.

Harley, J. B., Alarcón-Riquelme, M. E., Criswell, L. A., Jacob, C. O., Kimberly, R. P., Moser, K. L., . . . Langefeld, C. D. (2008). Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nature Genetics, 40*(2), 204-210.

Yu, C., Gershwin, M. E., & Chang, C. (2014). Diagnostic criteria for systemic lupus erythematosus: A critical review. *Journal of Autoimmunity, 48-49*, 10-13.

Rask-Andersen, M., Almén, M. S., & Schiöth, H. B. (2011). Trends in the exploitation of novel drug targets. *Nature Reviews Drug Discovery, 10*(8), 579-590.

Cheng, F., Liu, C., Jiang, J., Lu, W., Li, W., Liu, G., . . . Tang, Y. (2012). Prediction of Drug-Target Interactions and Drug Repositioning via Network-Based Inference. *PLoS Computational Biology, 8*(5), e1002503.

Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., . . . Zeng, J. (2017). A Network Integration Approach for Drug-Target Interaction Prediction and Computational Drug Repositioning from Heterogeneous Information, *Nature Communications*, *8*(1), 573.

Cheng, F., Zhou, Y., Li, W., Liu, G., & Tang, Y. (2012). Prediction of Chemical-Protein Interactions Network with Weighted Network-Based Inference Method. *PLoS ONE, 7*(7), e41064.

Barabasi AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. *Nature Reviews Genetics, 12*(1), 56–68.

Chen, L., Liu, T., & Zhao, X. (2018). Inferring anatomical therapeutic chemical (ATC) class of drugs using shortest path and random walk with restart algorithms. *Biochimica Et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1864*(6), 2228–2240.

[HLA-B gene - Genetics Home Reference - NIH. Retrieved from https://ghr.nlm.nih.gov/gene/HLA-B#conditions

Zepecki, J. P., Snyder, K. M., Moreno, M. M., Fajardo, E., Fiser, A., Ness, J., … Tapinos, N. (2018). Regulation of human glioma cell migration, tumor growth, and stemness gene expression using a Lck targeted inhibitor. *Oncogene*, *38*(10), 1734–1750.

Y. Wang and J. Zeng, predicting drug-target interactions using restricted Boltzmann machines. Bioinformatics, 2013. 29(13): p. i126.

Xie, L., Zhang, Z., He, S., Bo, X., & Song, X. (2017). Drug—target interaction prediction with a deep-learning-based model. *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 469-470.

5-Aza-2′-deoxycytidine A3656. (n.d.). Retrieved from https://www.sigmaaldrich.com/catalog/product/sigma/a3656?lang=en®ion=TH

Ferguson AT, Vertino PM, Spitzner JR, Baylin SB, Muller MT, Davidson NE. Role of estrogen receptor gene demethylation and DNA methyltransferase.DNA adduct formation in 5-aza-2deoxy-cytidine-induced cytotoxicity in human breast cancer cells. *J BiolChem*.1997;272:32260 –32266.

Database, G. (n.d.). LYN Gene(Protein Coding). Retrieved from https://www.genecards.org/cgi-bin/carddisp.pl?gene=LYN

Goldenberg-Furmanov, M., Stein, I., Pikarsky, E., Rubin, H., Kasem, S., Wygoda, M., . . . Ben-Sasson, S. A. (2004). LynIs a Target Gene for Prostate Cancer. *Cancer Research,64*(3), 1058-1066,1058-1060.

Vítek L, Muchová L, Jančová E, Pešičková S, Tegzová D, Peterová V, et al. Association of systemic lupus erythematosus with low serum bilirubin levels. Scand J Rheumatol. 2010 Nov;39(6):480–4.