PRO-Simat: Protein network simulation and design tool

Rana Salihoglu [a], Mugdha Srivastava [a,b], Chunguang Liang [a], Klaus Schilling [c], Aladar Szalay [d,e,f], Elena Bencurova [a,*,1], Thomas Dandekar [a,g,**,1]

[a] Department of Bioinformatics, University of Würzburg, Würzburg, Germany
[b] Core Unit Systems Medicine, University of Würzburg, 97080 Würzburg, Germany
[c] Informatics VII, Robotics and Telematics, Department of Mathematics and Informatics, Am Hubland, University of Würzburg, D-97074 Würzburg, Germany
[d] Dept. of Biochemistry, Biocenter, Am Hubland, University of Würzburg, D-97074 Würzburg, Germany
[e] Department of Radiation Medicine and Applied Sciences, Rebecca & John Moores Comprehensive Cancer Center, University of California, San Diego, CA, USA
[f] Dept. of Pathology, Center of Immune technologies, Stanford University School of Medicine, Stanford, CA 94305, USA
[g] Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg, Germany

ARTICLE INFO

ABSTRACT

PRO-Simat is a simulation tool for analysing protein interaction networks, their dynamic change and pathway engineering. It provides GO enrichment, KEGG pathway analyses, and network visualisation from an integrated database of more than 8 million protein-protein interactions across 32 model organisms and the human proteome. We integrated dynamical network simulation using the Jimena framework, which quickly and efficiently simulates Boolean genetic regulatory networks. It enables simulation outputs with in-depth analysis of the type, strength, duration and pathway of the protein interactions on the website. Furthermore, the user can efficiently edit and analyse the effect of network modifications and engineering experiments. In case studies, applications of PRO-Simat are demonstrated: (i) understanding mutually exclusive differentiation pathways in *Bacillus subtilis*, (ii) making Vaccinia virus oncolytic by switching on its viral replication mainly in cancer cells and triggering cancer cell apoptosis and (iii) optogenetic control of nucleotide processing protein networks to operate DNA storage. Multilevel communication between components is critical for efficient network switching, as demonstrated by a general census on prokaryotic and eukaryotic networks and comparing design with synthetic networks using PRO-Simat. The tool is available at https://prosimat.heinzelab.de/ as a web-based query server.

1. Introduction

Gene ontology (GO) enrichment analysis [1] is commonly conducted to identify statistically overrepresented GO terms in omics datasets (e.g., a set of differentially expressed genes in RNA-seq datasets). Various tools are offering such analyses (e.g., EasyGO, [2]; NaviGO, [3]; and FLAME, [4]). Furthermore, several innovative protein-protein detection and analysis methods were developed in the last decade (reviewed in [5]), including clustering and network analysis [6,7], so the complex biological datasets and their related interpretations can be handled.

Recently, web applications and platforms have become popular in bioinformatics, enabling simple processing, storage, analysis and visualization of biological data using open-source programming languages (e.g. Python, R) and their developed packages. The primary motivation is generally to enable the non-expert general user to evaluate their data directly on the platforms, where several analyses and comprehensive datasets/databases are integrated. Some popular tools include GeneTrail [8]; FungiFun2 [9]; ShinyGO [10]; NASQAR [11]; PlacentaCellEnrich [12]; and CPA [13]. Centrality analysis tools like CentiServer also contribute to the field [14].

Such capabilities are combined in our tool with dynamic protein network simulations and pathway design tools. Moreover, we make key molecular constructs available for the user, such as from oncolytic viruses or light-gated nucleotide processing enzymes.

* Corresponding author.
** Corresponding author at: Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg, Germany.
 E-mail addresses: elena.bencurova@uni-wuerzburg.de (E. Bencurova), dandekar@biozentrum.uni-wuerzburg.de (T. Dandekar).
1 Equally contributing corresponding authors

Users ranging from students doing simple network biology to bioinformaticians conducting sophisticated network simulations will benefit from PRO-Simat's different functionalities. A fully trained bioinformaticians will be able to exploit the tool's functionalities in-depth. As many operations of PRO-Simat still need some user interaction, and hence work only semi-automatically, i.e., switching from one analysis step to the next, some training will enhance the ease of use. To help with this, a tutorial (Supplementary file 1) has been created which explains all analysis steps for the user. Moreover, there are online hints and help for every analysis step while doing the actual analysis.

After mastering the simple analysis steps, we want the user to go on to also use the different refined analysis tools of PRO-Simat, as this is a powerful analysis package combining static and dynamic network analysis with network design and engineering.

While network analyses as well as GO enrichment can easily be done elsewhere, it is not so easy to perform simulations starting from a graph that is actually created by the user. PRO-Simat is powerful for performing modelling of cellular differentiation or cancer pathways, for example. This tool also allows hypothesis generation, for instance that pathway activation is mutually exclusive for *B. subtilis* biofilm formation. Moreover, it allows the actual generation of novel networks, e.g., for pathogen-host interactions.

PRO-Simat was developed using the Shiny package of the R programming language and includes the following features:

  (i) A database containing more than 8 million protein-protein interactions and GO information of proteins belonging to 32 organisms,
 (ii) A user-friendly protein-protein interaction (PPI) network analysis tool that allows visualising, examining editing and design of protein-protein interactions,
(iii) The GO and KEGG enrichment analyses using either selected proteins available within the PRO-Simat tool or can use user-uploaded transcriptome and differential gene expression datasets,
 (iv) Boolean genetic regulatory network analysis via the Jimena simulation framework [15], [16],
  (v) Visualisation and comparison of the data obtained from the protein network simulations via the R SQUADD package.

PRO-Simat allows also processing of user data, such as GO and KEGG analysis of the protein, pathway or network of interest. When the networks obtained with visNetwork are downloaded, our tool offers interactive editing of HTML files. The PRO-Simat interface allows all graphics obtained from GO and KEGG analyses to be saved in SVG format and tables in PDF, CSV, and EXCEL formats.

Synergies and advancements compared to the above tools include an improved graphical user interface and integrated searches in different databases. Based on this, PRO-Simat has new capabilities as it integrates dynamic network simulation with protein network analysis, editing and pathway exploration capabilities to specifically accelerate pathway analysis and design in synthetic and systems biology. The combination of topological, structural and dynamic analysis of the protein network by PRO-Simat allows the user to understand control and information processing in the network, involved cascades and feed-back / feed-forward loops. Dynamic simulations of protein networks allow monitoring of each protein in the network and its activity course over time while plotting selected ones. Powerful application examples demonstrate this: we investigate (i) how and why bacterial cellular differentiation pathways are mutual exclusive in *B. subtilis*, (ii) how to engineer a Vaccinia virus for best oncolytic action on cancer cells and (iii) the design of light-gated nucleotide processing circuits such that they can operate DNA storage. The application allows GO analysis of any biological data sources resulting in gene or protein lists. PRO-Simat is

particularly useful for experimentalists and computational biologists working on a protein network of interest and its dynamics. This includes smooth analysis of connected omics data sets and validation data. Moreover, one can investigate intended operation and dynamics of designed or natural protein networks.

2. Materials and methods

2.1. R packages used to build the PRO-Simat tool

The PRO-Simat web tool with a user-friendly graphical user interface was developed using the following R [17] packages: Shiny [18]**,** fresh [19], shinyalert [20], shinycssloaders version [21], shinydashboard [22], shinyjs version [23], shinythemes [24], shinyWidgets [25], gridSVG [26], svglite [27], and svgPanZoom [28]. The following Bioconductor and data visualisation packages were also used for development: org.Hs.eg.db [29], org.At.tair.db [30], org.Ce.eg.db [31], org.Dr.eg.db [32], org.Dm.eg.db [33], org.EcK12.eg.db [34], org.Gg.eg.db [35], org.Mm.eg.db [36], org.Rn.eg.db [37], org.Bt.eg.db [38], org.Cf.eg.db [39], org.Ss.eg.db [40], org.Mmu.eg.db [41], org.Xl.eg.db [42], org.Pt.eg.db [43], and org.EcSakai.eg.db [44], AnnotationDbi [45], BiocManager [46], DOSE [47], enrichplot [48], pathview [49], ReactomePA [50], visNetwork [51], dplyr [52], DT [53], RmySQL [54], ggplot2 [55], htmlwidgets [56], igraph [57], and SQUADD [58]. To allow efficient operation of large-scale data, the maximum table size in the Shiny app is at 60 Mb. The server permissions for this large data size were changed accordingly.

2.2. The PRO-Simat framework

The PRO-Simat framework consists of four analytic modules: protein-protein interaction (PPI) network, Gene Ontology (GO) enrichment, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, and genetic regulatory network. The user manual further describes each analysis step in detail and is found in Supplementary file **1**. The graphical overview of each module is depicted in Fig. 1, and detailed information is given in the following sub-sections. The menu box for each module opens by default to show the user all available options.

2.2.1. Protein database and PPI network analysis

The comprehensive database includes almost 8 million data points of protein-protein interaction data on the following organisms: *Arabidopsis thaliana*, *Bacillus subtilis* (1423), *Bacillus subtilis* (strain 168), *Bacillus anthracis*, *Bos taurus*, *Caenorhabditis elegans*, *Cavia porcellus*, *Danio rerio*, *Drosophila melanogaster*, *Escherichia coli*, *Escherichia coli* (strain K12)*, Escherichia coli O157:H7* (including strain EDL933)*, Escherichia coli O127:H6* (strain E2348/69 / EPEC)*, Gallus gallus*, *Homo sapiens*, *Helicobacter pylori*, *Listeria monocytogenes*, *Listeria monocytogenes serovar 1/2a* (including strains ATCC BAA-679 / EGD-e and EGD / Mackaness)*, Listeria monocytogenes serotype 4b* (strains LL195 and F2365)*, Mus musculus*, *Oryzias latipes*, *Ovis aries*, *Plasmodium falciparum*, *Rattus norvegicus*, *Saccharomyces cerevisiae*, *Saccharomyces cerevisiae* (strain ATCC 204508 / S288c)*, Schizosaccharomyces pombe* (strain 972 / ATCC 24843)*, Staphylococcus aureus*, *Treponema pallidum*, *Vaccinia virus* (including strains: Copenhagen, WR, ʟ-IPV, Ankara, GLV-1h68, and VVΔTKΔN1L)*, and *Xenopus laevis.* Data were collected from databases IntAct [59], MINT [60], StringDB [61], and Viruses.String-DB [62]. The human-human physical interaction data was obtained from the database StringDB; other organisms' interactions were collected from the specified databases. The database of PRO-Simat is created using MySQL workbench and then linked to the web tool via the RMySQL package [54]. The gene ontology (GO) information of proteins is obtained with the help of biomaRt [63], QuickGO [64], and UniProt.ws [65] packages. In addition, the signalling network (activation/inhibition) information
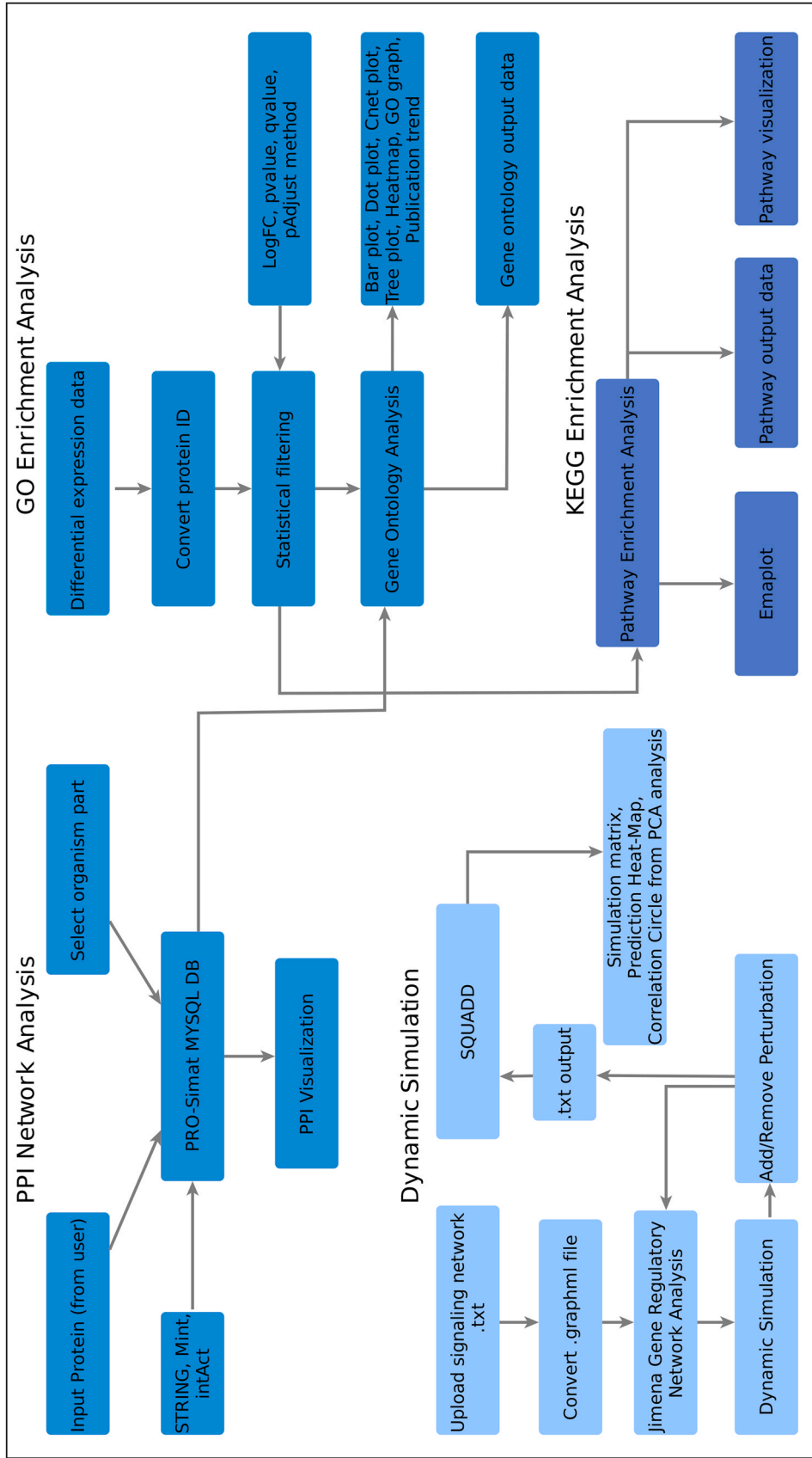
Fig. 1. The workflow of the PRO-Simat web tool. Upper left panel: The protein data input and visualisation of the PPI network. Right panel: The GO enrichment and KEGG pathway analyses (data upload for differential expression gene, ID converter, statistical filtering). Lower left panel: Dynamic simulation with the integration of Jimena software and visualisation (e.g., simulation matrix, prediction heat-map) via SQUADD.

for 40,000 human proteins was added to the tool using the Omnipath R package [66].

In order for the network visualization to work efficiently, not all of the networks are presented to the user in a single image. For instance, there are nearly 7 million interactions for *Homo sapiens*, precluding single image depiction. There are many databases that offer a maximum of two thousand interactions, such as IntAct. The number of interactions you see in a network differs from the one given in the detailed table below because these databases store the two-sided network information in the form of organism 1 → organism 2 and organism 2 → organism 1. Therefore, if there is a direct interaction between protein A and protein B, and at the same time there is a direct interaction between protein B and protein A, both of them appear in the table.

2.2.2. The GO enrichment and KEGG pathway analyses

The second module was developed using the clusterProfiler R package (enrichGO and enrichKEGG functions), which enables the analysis and visualisation of functional profiles of genomic coordinates, genes, and gene clusters, as well as performs over-representation and gene-set enrichment analyses on the GO terms and KEGG pathways [67]. For the GO enrichment module, the user can submit their own set of differential gene expression data (DGE), and subsequently, the loaded gene ID types will be converted to the Entrez gene ID (ENTREZIDs). For the GO enrichment, several additional species are available for analysis, including *Arabidopsis thaliana, Anopheles gambiae, Bos taurus, Canis familiaris, Caenorhabditis elegans, Danio rerio, Drosophila melanogaster, Escherichia coli, Escherichia coli* strain Sakai, *Gallus gallus, Mus musculus, Rattus norvegicus, Homo sapiens, Macaca mulatta, Myxococcus xanthus, Pan troglodytes, Saccharomyces cerevisiae, Sus scrofa,* and *Xenopus laevis*.

Regarding KEGG enrichment analysis, the following organisms were implemented: *Anopheles gambiae*, *Arabidopsis thaliana*, *Bos taurus*, *Candida albicans*, *Canis familiaris*, *Caenorhabditis elegans*, *Danio rerio*, *Drosophila melanogaster*, *Escherichia coli*, *Escherichia coli O157:H7* (strains Sakai and EDL933), *Gallus gallus*, *Homo sapiens*, *Listeria monocytogenes*, *Mus musculus*, *Macaca mulatta*, *Pan troglodytes*, *Myxococcus xanthus*, *Rattus norvegicus*, *Saccharomyces cerevisiae*, *Sus scrofa*, and *Xenopus laevis*. User-defined statistical filtering and the modification of parameters (e.g., p-value, q-value, and p-adjust) required for the GO enrichment and KEGG pathway analyses are available as the input data, including "Gene", "Log$_2$FC" and "pAdjust" columns, which can be adapted to the user's requirements. Typically, the user will upload his/her own data or use the existing data prepared for the user to test the "GO enrichment analysis" section using the "use example data" option in the lower left corner. Then one can visualize the KEGG pathway results in the "KEGG Pathway" section. The results of both GO and KEGG pathway enrichment analyses are presented as tables and graphs (barplot, dotplot, treeplot, heatmap, cnetplot, GO graph, and emapplot). A warning message alerts the user if they are not visiting the KEGG pathway section first before selecting/uploading data and running them in the GO enrichment analysis section.

2.2.3. Jimena dynamic simulation and SQUADD visualization

At its most basic level, protein-protein interactions tell us that there is a link between two proteins. For example, there is a physical interaction between the Vaccinia virus K3L protein and the IL32 human protein. However, in order to use a dynamic simulation of the involved signalling, more information is needed, such as which protein activates or inhibits another. Hence the Jimena module works independently of the others as a user tests his/her own topology of a protein network of interest by uploading this network in a "txt" format. Nevertheless, any topology or protein network analysed in the steps before or the example networks we offer can be

used for this dynamic simulation module. The software alerts the user to upload the starting topology of a network first.

To dynamically simulate the resulting signalling behaviour of the network, the Java-based Jimena framework is integrated into the PRO-Simat tool ([15]; tutorial and code at https://www.biozentrum.uni-wuerzburg.de/en/bioinfo/computing/jimenae/).

Jimena interpolates between Boolean network states by exponential and other functions. Our PRO-Simat tool did not develop Jimena further but uses the output of Jimena in order to visualise and analyse dynamic outcomes of the network with attached explanatory notes directly on its website. Using the melt function [68] of the reshape2 R package, the results from the dynamic simulation are adapted for animation visualisation by converting the data frame to a long format and expanding it. The modified data are depicted on the web tool with the plotly R package [69].

The dynamic simulation by SQUAD software [70] is also included as a simulation method for the Jimena framework. The SQUADD R package [58] is used to visualise the outputs obtained from the Jimena framework. The "plotSimMatrix" and "plotCC" functions in this package are utilised to simplify the visualisation of the simulation output and to evaluate the consistency of a network model (via principal component analysis), respectively.

2.2.3.1. Engineering light-gated cooperative changes in proteins. To turn a protein to be light-responsive (active only after stimulation with light of specific wavelength), several natural domains can be added to the protein, such a BLUF domain (sensors of blue light using FAD; flavin adenine dinucleotide). To optimise engineered cooperativity by using our light-gated domains, we fused an optimized BLUF domain to the protein using an approach pioneered and introduced by Rama Ranganathan [71]. Detailed methods and constructs achieved are found in [72]. For one example, we used a BLUF-fused polymerase from *E. coli* (KEGG pathway ID eco03030) and compared activity with and without the presence of blue light.

2.2.3.2. Network setup to model B. subtilis cellular differentiation. For the first and basic network, information was collected from databases STRING [61] and SubtiWiki [73]. Further interactions in the network, its size, master, and hub nodes need to be defined and collected, and conflicts in interactions, as well as the relevance of biologically important nodes had to be determined by comparing different databanks and literature resources together with expert curation. Well-established interactions (52 of them) from the central differentiation network are studied here (detailed listing in supplementary Table S1; excel file). After the construction and visualisation of the network using yEd, the result is a Boolean network with the Boolean states (either 0 = off or 1 = on) for each node (network structure files: Table S2 xgmml and Table S3 graphml). Alternatively, the network can also be created by using another editor. The Graph editor can process several file formats like *.xlsx, *.graphml, *.xml, and *.xgml. Individual nodes can also be added, deleted, or represented in different pictorial forms in the editor platform.

The xml file can also be saved using CellDesigner or Cytoscape, and the.net file can be created manually. We provide our *B. subtilis* network file as a supplement with all interactions as an excel file (Supplementary Table S1) and the resulting network graph as an xgmml file (Table S2) and as a graphml file (Table S3).

2.2.3.3. Manipulation and engineering of cooperative pathway changes. To understand the behaviour of the *B. subtilis* network during activation (example: Spo0A) and inhibition (examples: knockouts for ComA, DegU and cannibalism, Table 1) of biological processes, dynamic simulations were performed. The protocol file can be edited to include the perturbation at different time points and

Table 1

Master regulators of cooperative change modelled and tested in *B. subtilis*[a].

Impaired regulators	Effect	Genetic test described in
Spo0A	Sporulation	[74]
Spo0A (*Bacillus polymyxa*)	Sporulation	[75]
ComA	Surfactin production, competence development and spore formation	[76]
DegS	Altered phosphorylation of DegU	[77]
AbrB[2]	Biofilm development	[78]

[a] mutated or activated regulators are represented along with their biological effects on the cell fate determination. The simulated network accurately confirms the behavior in the presence of the corresponding genetic knockouts and tests. [2]The node AbrB represses 39 operons in *Bacillus subtilis*. To test and validate the pathway behavior of all the other nodes, the AbrB repressor was removed from the network (validation experiments and data: see references).

duration. The node is perturbed for a particular time point ("single-pulse action") while "range-pulse" adds the perturbation for a time period. For all the simulation calculations, we introduced the perturbation after five time-units at a range pulse of up to 15 time units. Each simulation indicates the effect of perturbations on the other proteins (nodes in the network), and the cell fate (activity of nodes). Furthermore, the resulting functional output is normalised in the simulation between 0 - minimal and 1 - maximal activity. PRO-Simat simulates all network nodes, but for clarity only key trajectories of interest are shown and plotted.

2.2.3.4. Dynamics of cooperative pathway changes. To investigate the network dynamics and to analyse and calculate the number of steady states in protein-protein interaction networks, JimenaE was used to provide output data for the dynamic simulations. Briefly, Jimena [15] allows a selection of different functions (exponential, Bool cube and others) and improvements to alternatives such as SQUAD [70] to interpolate between full on and off states of network nodes, considering the logical network connectivity. The resulting model is next visualised and analysed by PRO-Simat. It is first semiquantitative, i.e. arbitrary time units are given, and activation can dynamically change between complete (100%) or none (0%). Such a model allows only order predictions (rank, order and relative strengths of protein activation or inhibition). However, with more experimental data available, time courses can be normalised, and the receptor activation or subsequent protein activation quantified (e.g. phosphorylation data used for data-driven modelling of network parameters using PRO-Simat).

2.2.3.5. Cooperative change and synthetic design experiments using the oncolytic vaccinia virus. The experiments and data, described in detail in the publications [79] and [80], are used here as examples for engineering cooperative pathway changes in synthetic biology. To establish Vaccinia virus vector GL-ONC-1, foreign gene expression cassettes were inserted into the *F 14.5 L, J2R* and *A56R* loci of the Vaccinia virus (wt-LIVP):

(i) a Renilla luciferase-*Aequorea* green fluorescent protein (*RUC-GFP*) fusion cDNA under the control of the Vaccinia synthetic early/late promoter was inserted into the *F 14.5 L* locus of wt-LIVP;

(ii) a β-galactosidase (*lacZ*) encoding cDNA, under the control of p7.5 promoter, was inserted into the *TK* (Thymidine kinase) locus of GLV-1d27M and

(iii) a β-glucuronidase (*gusA*) encoding cDNA under the control of p11 promoter was inserted into the *HA* locus.

PRO-Simat models exactly how the mitigated virus with the gene cassette induces a coordinated cellular network change in cancer cells with improved oncolysis and enhanced apoptotic pathway activation. However, this is only one of many designs we tested and investigated. The different engineering of oncolytic virus strains is described in the United States patent US 10,238,700 B2 for all

oncolytic Vaccinia strains we (coauthor Szalay of this paper and his group) [81] have engineered so far.

2.2.3.6. Methods for comparison of cooperative action in protein networks and technical networks. In Supplementary file 2, we give data comparing cooperative change in simple systems and complex organisms and we analyze proteins using the established GoSynthetic database nomenclature [82]. Updated functional information on interactions and cooperativity in biological and technical terms used in engineering and robotics can be viewed in Fig. S1. Sequence searches, process classification, pathway and interactome analysis are demonstrated for design of new pathways. Category information was added after analyzing of all protein functions and functional terms. Supplementary file 2 further illustrates the enhanced capabilities of PRO-Simat, e.g. engineering knockouts in bacteria, application for biotechnology (for example: carbon harvesting in plants) and planning new innovative nanotechnology by pathway and network control using light-gated proteins.

3. Results

First, PRO-Simat web tool was designed using the R programming language and the four analysis modules implemented: (i) PPI network visualisation, (ii) GO enrichment, (iii) KEGG pathway analysis and visualisation, and (iv) genetic regulatory network analysis and visualisation. The web-based tool PRO-Simat is publicly accessible at http://prosimat.heinzlab.de, and the source code is available on https://github.com/salihoglu/PRO-Simat.

The PRO-Simat is a versatile tool containing several types of analyses and an extensive protein database containing approximately 8 million protein and protein interaction data and about 40,000 signalling networks. It includes two modes – analysis of the single organism network (organism one and organism two are identical in the sidebar menu) or a network between two organisms can be examined. This can be selected in the sidebar menu; another option is to use the multiple protein box to search the PRO-Simat database by entering custom UniProtKB IDs identifiers.

As further gene expression data we federated into PRO-Simat over 4000 entries from the GEPIA2 cancer database regarding lung adenocarcinoma gene expression data including survival correlations.

Default exemplary data for each example in the paper is embedded in PRO-Simat. The *B. subtilis* network data is input as default data for system biological analysis. All example data given in the results are accessible both from within the application as well as from github. This allows the learner to test the different Jimena options available. Head and Neck data and all programming source code is available on github. Moreover, tutorial guide for the case studies appear in the form of easy tutorials on github.

For all four modules the menu box opens by default. Using the tutorial, the user is taken step-by-step through each analysis step for each tab and option. In addition, pop-up comments guide the user step-by-step to each analysis feature and application. Moreover, a

Fig. 2. Sample GO enrichment output generated by PRO-Simat. **(a)** *Homo sapiens*-Vaccinia virus WR strain PPI network overview. **(b)** *Homo sapiens*-Vaccinia virus WR strain PPI zoomed out network overview. **(c)** *Homo sapiens*-Vaccinia virus PPI in detail, zoomed into the network. **(c)** Functional grouping tree diagram of the result of GO over-representation analysis of lung adenocarcinoma data. **(d)** Cnet plot image depicting five category connections of genes and GO cellular component biological concepts as a network. **(e)** Barplot displaying 15 sub-categories of enriched GO terms belonging to the main category of the biological process based on the p-adjust values as enrichment score and gene number. **(f)** The proliferation category of the Cnet plot showing the enriched GO terms of PRO-Simat and the associated genes.

"skip" button allows users to fast-forward if they know this feature or application already. Under exception conditions, an alert button tells the user what is missing and should be provided before continuing.

3.1. Module network

The interactions between the proteins of the selected organisms are visualised in the form of the interactive network (Fig. 2**a** overview, Fig. 2**b** detailed view) as well as in table form below the network. Default settings for the visualisation are set to provide a quick overview for the user. Therefore, up to 100 interactions in total are listed. However, the user can modify this setting using the plus and minus buttons in the left lower corner, resulting in more or fewer interactions being visualised. In addition, the multi-protein box allows the UniProtKB IDs to be searched in the PRO-Simat database. For each protein entered here, ten interactions are visualised. With the help of the "Multiple Protein" box, all interactions of any desired gene in the PRO-Simat database can be accessed. The network module also contains a search bar, which allows fast access to any protein in the network. By clicking on the selected node, gene name, UniProtKB ID, species name and taxonomy information are displayed as a popup, as well as gene ontology fact sheets with search boxes for each protein (Supplementary file 1). Mouse hovering over the selected node displays the gene name, UniProtKB ID, species name, and taxonomy information displayed as a popup. In addition, GO fact sheets with search boxes for each protein appear in a modal window (**user manual**, Supplementary file 1). Gene names or UniProt IDs are shown on the web-page/network itself for ease of visual interpretation.

In addition, the modification of nodes in the interactive graph can be changed directly, and the nodes and edges can be added/or removed, as depicted in Supplementary file 1. Each new node must be in UniprotKB or SymbolID format. Subsequently, new nodes are actively used while performing GO and KEGG analysis.

3.2. Module GO enrichment

GO analysis facilitates the biological classification of selected genes and is a valuable tool for identifying significantly differentially regulated genes or functional groups. Traditionally, three main ontologies are used to describe the nature of the genes, biological process, molecular function and location in the cell. In the PRO-Simat, there are three options for performing such analysis: analysis of the protein list obtained from the PPI network, analysis of Lung adenocarcinoma (LUAD) dataset obtained from the GEPIA2 database [83] (implemented in PRO-Simat), or uploading of the user's differentially expressed genes (DEGs) dataset.

In case the user's DEG data are not available, GO enrichment analysis is performed based on the protein information that can be gathered from the integrated database. In this case, the log_2FC values of the Pancreatic adenocarcinoma (PDAAD) data in the background are used in the analysis for human genes. The pre-processing analysis can be carried out when necessary, including the conversion of protein IDs and statistical filtering (see below). The primary analysis is focused on the three main ontologies, where the user can determine the size of each category. The outputs are interactive graphs and tables, such as treeplot, dotplot and barplot (Fig. 2**c-f**). The treeplot (Fig. 2**c**) depicts hierarchical clustering of sub-categories obtained by the GO enrichment results, including all ontologies with the default configuration. Other visualisation options include cnetplot and barplot where the users can change the sub-category sizes (Figs. 2**d and** 2**e**). Moreover, the genes can be examined based on the five clusters with different biological and functional roles (apoptosis, positive and negative regulation, kinases, and proliferation) (Fig. 2**f**).

3.3. Module KEGG pathway

PRO-Simat has integrated mapping of protein-protein interaction networks to pathways, allowing visualisation of metabolic pathways or highlighting differential regulation of involved genes or both. In this analysis, KEGG pathway diagrams are downloaded directly from the KEGG servers using the Bioconductor packages and link the pathway with the GO analysis from the previous module. To use the KEGG pathway analysis, module 2 (GO enrichment) must be already activated. KEGG pathway analysis allows us to visualise the data in the form of an enrichment map (emap plot, Fig. 3**a**) or as a more complex metabolic pathway (Fig. 3**b**). The results are also displayed under the figures in a table including statistical data (Fig. 3**c**). In the KEGG analysis using the protein list obtained from the PPI network, PRO-Simat calculates whether colored KEGG pathways are switched on or off. To determine the coloring PRO-Simat uses gene expression values, in this example the log_2FC values of the Pancreatic adenocarcinoma (PDAAD) data in the PRO-Simat database.

Census and large-scale comparisons were performed using the large-scale data of the PRO-Simat database and available protein interaction data (Fig. S1 in file Supplementary file 2). We selected 18 organisms and the synthetic oncolytic virus. An organism census on protein interaction networks and key implementation routes for network control finds streamlined, direct communication in bacteria. In contrast, eukaryotic systems implement checks and balances by many levels of sensors, actuators, feedback and feed-forward loops for controlling cooperative pathway change in protein interaction networks. This identifies four factors essential to engineering cooperation between network components, pathways or subnetworks: Cooperative planning, efficient task distribution, cooperative perception and information exchange between network components (Supplementary file 2; section general principles; and Table S4). These insights can next be applied for synthetic biology and future applications, exemplified for DNA storage operated by nucleotide processing enzymes and protein network control by light-gated enzymes in nanofactories (Supplementary file 2; section future applications) and for the protein and network analysis and design (Supplementary file 2; section further application examples), e.g. regulatory networks and genetics in bacteria and engineering carbon harvesting in plants (metabolic pathway example).

3.4. Dynamic protein network simulation module

The last feature of the PRO-Simat is the analysis of the signaling network of selected genes. For this, we implemented Java-based packages JIMENA and SQUADD. Jimena allows dynamic simulation of protein networks using the Boolean state of nodes and a continuous hill function and can be applied to describe the semiquantitative models in infection, cancer or synthetic biology. Using PRO-Simat, we can engineer dynamic pathway change, including iterative model refinement with improved convergence and comparison to experimental data. To run the analysis, the signalling network data must be uploaded and converted to a Graphml file in PRO-Simat (the button is located directly under the upload line). The text file in a ".txt" format must include node1, label, and node2 columns. In addition, the user can modify the configurations of perturbation processes, such as determining the time interval and changing the activation/ inhibition states of proteins. The output graph (Fig. 3**d**) shows the time-dependent activation of each protein in the network; these will be visible by clicking on the specific gene/s. Fig. 3**e** illustrates the output of the dynamic modelling of selected nodes from the Jimena analysis using SQUADD. The estimation of the changes in the activation states between two experiments can also be illustrated as heatmaps (Fig. 3**f**). The consistency of a regulatory network model obtained by Jimena can be tested by principal component analysis (PCA) as illustrated in the correlation circle (CC) plot (Fig. 3**g**).
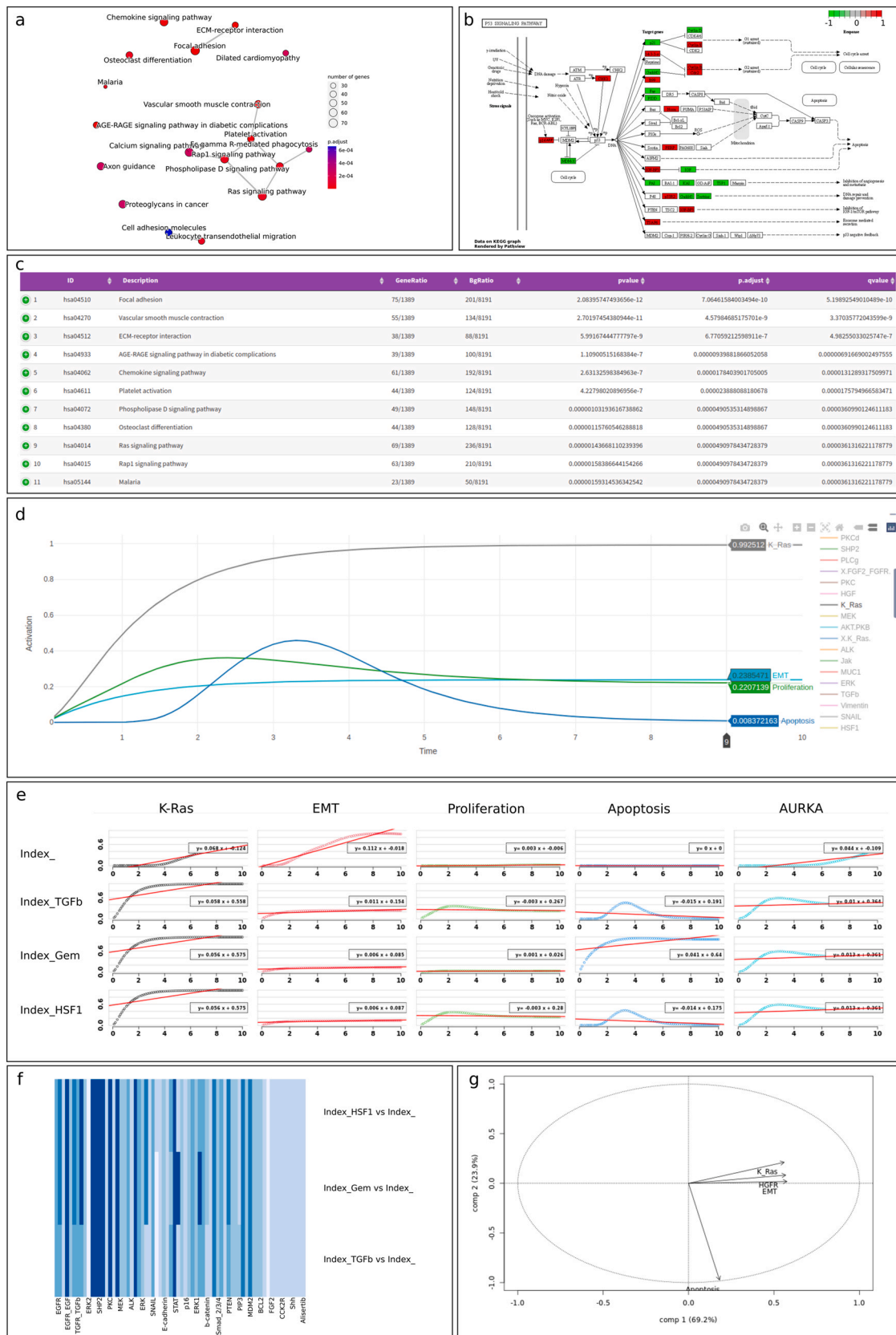
Fig. 3. Visualisation of KEGG pathway output and Jimena dynamic simulation embedded in the PRO-Simat tool. **(a)** Emap plot as a result of KEGG enrichment analysis using lung adenocarcinoma data. **(b)** The pathview plot of KEGG pathway analysis. **(c)** The results of enrichment analysis given as a list in the table appearing in the PRO-Simat web tool. **(d)** Dynamic modelling of protein network using Jimena. **(e)** The simulation graphs selected human proteins using the SQUADD analysis. **(f)** Heatmap illustrates the analysis results obtained by Jimena. **(g)** PCA analysis of Jimena analysis.

3.5. Case studies

To demonstrate that PRO-Simat by its focus on structural and dynamic network analysis can be very broadly applied and model essentially any protein network of interest, we give first an application example from bacteria (prokaryotic domain of life) and study networks and pathways around cellular specialisation demonstrated in biofilm formation. Our second example looks at human cells (eukaryotic domain of life) and their interactions with Vaccinia virus and show how processes in human cells (apoptosis, proliferation) are naturally changed by the interaction with the virus and how this can be exploited for efficient removal of cancer cells (oncolysis). Finally, the third example starts from DNA replication by DNA polymerase (eukaryotic domain of life) but takes this to synthetic biology for usage in DNA storage (synthetic domain of life).

3.5.1. Natural example: bacillus subtilis

To model pathway change, we selected *B. subtilis* and its differentiation pathways. They are mutually exclusive and PRO-Simat allows us to study how the respective on and off switching is achieved and where control loops and key genes are, as well as the differentiation signal evolution over time. Fig. 4 in the results explains the different subnetworks pertaining to the differentiation state of *B. subtilis* (network setup: see Materials and Methods). As only one cellular differentiation type can be simultaneously active in a bacterial cell, the pathways must change in a coordinated manner, and communication signals are required. Our model here is semiquantitative, so the arbitrary time units are given, and activation can dynamically change between complete (100%) or none (0%). Such a model allows only order predictions (degree and order of protein activation or inhibition). In this example, we were focused on the dynamic simulation of activation. With more data, data-driven modelling adapts network topology and ground state activation to the actual situation according to the data measured. PRO-Simat thus allowed us to efficiently model activation by Spo0A protein (Fig. 4a) and show the active protein necessary for differentiation of *B. subtilis* cells for biofilm formation. Similarly, inhibitory stimuli can be studied, for example the effect of the protein DegU. This leads to the differentiation of *B. subtilis* into cannibalistic cells (Fig. 4b, red curve, miner activation signal).

The modified network and the modified dynamics (output from JimenaE, see methods) are again efficiently visualised and analysed using PRO-Simat.

3.5.2. Design of new protein-interaction networks example: oncolytic virus

PRO-Simat enables design of synthetic protein-interaction networks such as engineering oncolytic viruses and artificial system circuits. The oncolytic virus Vaccinia is successfully used as a tumour-targeting agent in animal models, however, the virus-host interaction dynamics are still unclear. While there are millions of curated protein interactions available from PRO-Simat, editing and federating further protein interactions is necessary in typical use cases. For instance, one adds or removes specific proteins in a cancer model to account for model-specific cancer drivers or gain of function / loss of function mutations in the cancer cells.

For active network design, modification and consideration of protein interactions are critical. For this, PRO-Simat has a focus on human and viral interactions allowing now (i) direct editing of protein networks by desired new interactions, (ii) analysing the resulting pathways formed and (iii) dynamic simulations of such networks. Together this allows design, network modification and revision of the resulting effects of cooperative change.

For instance, a virus can be modified to replicate with preference in cancer cells ("oncolytic virus"). Engineering a Vaccinia virus to become oncolytic thus involves mitigating the wild type virus (for

vaccinia the Lister strain) by replacing three reading frames important for its rapid replication (*F 14.5 L, J2R* and *A56R* [79]), with foreign gene expression cassettes. Fig. 4c gives an overview of the involved host-virus interactions when the oncolytic virus is in the host cell, as established using PRO-Simat. The three specific modifications to obtain the oncolytic virus are:

(1) F 14.5 L by insertion of RUC_GFP (*Renilla* luciferase-*Aequorea* green fluorescent protein) expression cassette,
(2) Gene J2R by *lacZ* and
(3) Gene A56R by *gusA* (β-glucuronidase) so that the virus can be better monitored and at the same time is attenuated.

The three specific pathway interaction networks mitigating virus replication and enhancing oncolysis are shown in detail in Fig. 4d and can be examined in detail using PRO-Simat.

To understand involved pathways, GO-Annotation analysis can provide the information in three different directions according to gene ontology and cellular compartment, biological process and molecular function. For the dynamic changes, the interaction network can be dynamically simulated. In particular, in presence of the virus in the cancer cells (GLV-1h68 fully activated over time) apoptosis increases while antiapoptotic proteins such as BCL2 stay low (Fig. 4e; only few protein activation trajectories shown but all for the whole network are calculated). The PRO-Simat analysis shows hence that these three modifications cooperate with efficient task distribution to improve virus oncolysis and apoptosis allowing efficient removal of tumours in animal models. Therefore, synthetically prepared strains become an optimized tool for cancer treatment (oncolysis), far better than native strains, for instance in pancreatic cancer (e.g. LOAd703 Oncolytic Virus Therapy for Pancreatic Cancer; Trial NCT02705196).

Moreover, optimizing oncolytic action against cancer cells by the vaccinia virus involves redesigning the normal replication cycle of the vaccinia virus, such that it replicates with high performance and preference in cancer cells. Information exchange couples different required cooperative pathway changes. Sensing exploits viral and hosts receptors, but other tumor features can also be sensed and inhibited. Thus, a newly designed VACV strain GLV-5b451, expresses the anti-vascular endothelial growth factor (VEGF) single-chain antibody (scAb) GLAF-2 to suppress VEGF action, improving oncolytic therapy as tested in different canine cancers. The oncolytic vaccinia virus mediates an oncolytic system switch in cancer cells best if virus replication progresses several hours before apoptosis or oncolysis starts. This requires a mitigated system switch to oncolysis. To achieve this, it is better to deliver the virus by adipose stromal vascular fraction cells, as currently tested in phase I trials in patients with advanced solid tumours or acute myeloid leukemia (AML; [84]).

Other modifications, which the designer of an oncolytic virus may want to explore and can model with PRO-Simat are an expression of an apoptosis enhancing protein or a chemokine attractant protein (e.g. IL2) to attract T-cells to the tumour for even more efficient oncolysis.

Specific advantages of PRO-Simat include its tools for optimal pathway design, translation of technology into biological language and general analysis capabilities of multi-component interaction networks to allow also designs for oncolysis. Vaccinia virus proved to be a robust synthetic biology tool to implement different oncolysis designs. This includes new therapeutic options even for challenging cancers such as pancreatic cancer [85] or lung cancer [86]. To better analyse the latter, PRO-Simat has included lung adenocarcinoma gene expression data.

3.5.3. Controlling enzyme activity by light-activating domains

Controlling enzyme activity is crucial for specific applications, such as optogenetic studies. PRO-Simat offers a toolbox to fuse light-gated
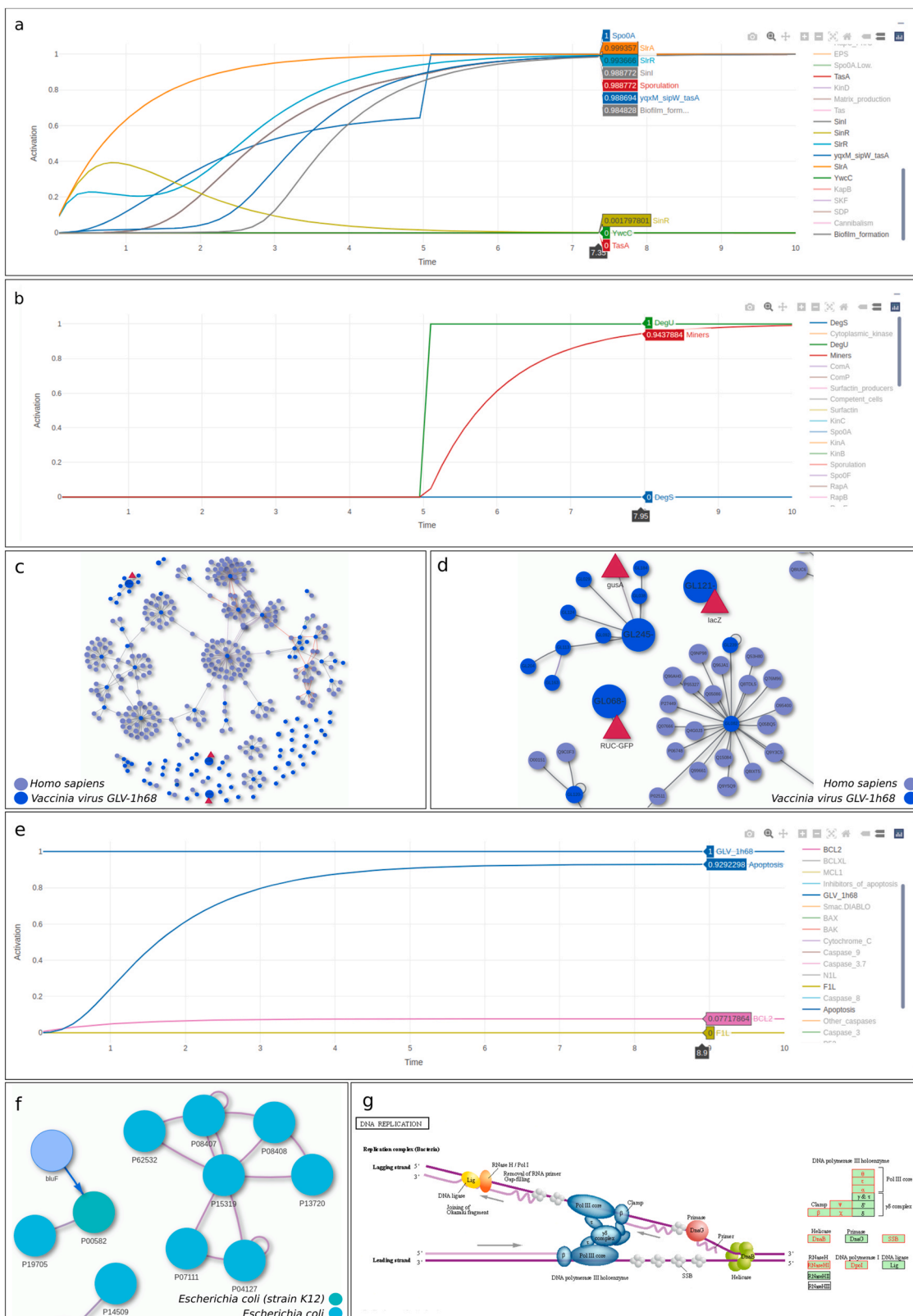
(caption on next page)

Fig. 4. Dynamic modeling of *B. subtilis* differentiation network, oncolytic Vaccinia virus GLV-1h68 and synthetic light-controlled nucleotide processing networks by PRO-Simat. The individual protein activities and trajectories in the *B. subtilis* regulatory network compared different differentiation statuses of the network and different pathway activities d as follows: **(a)** Simulation of spo0A activity. Activation of the Spo0A protein downregulates the repressor SinR and activates the genes involved in sporulation, resulting in biofilm formation. **(b)** Simulation of DegU activity. DegU is a crucial gene regulator, which controls competence, motility and degradative enzyme secretion. DegS Kinase phosphorylates master regulator DegU and is activated by starvation or salt stress. Activation of DegU promotes the subpopulation of cells called 'miners', which promote degradation of large biopolymers into nutritive peptides (cannibalism). **(c)** Overview on the *Homo sapiens*-Vaccinia virus GLV-1h68 protein-protein interaction network. **(d)** Detail of the *Homo sapiens*-Vaccinia virus GLV-1h68 protein-protein interaction network: The genome modifications (red triangles) inhibit virulence pathways and reduce vaccinia replication. **(e)** Increased apoptosis with Vaccinia virus GLV-1h68 activation analysed by Jimena. **(f)** Attachment of BLUF light-gating protein domain to polA polymerase. This protein engineering modification is modelled in PRO-Simat by rearranging the PPI network such that the BLUF protein domain (left, light blue) activates pol A polymerase. **(g)** DNA replication pathway as a result of KEGG pathway analysis including the added light-gated polymerase and shining of blue light on the BLUF domain to activate the DNA polymerase: active DNA replication is the result.

domains to a selected enzyme (Fig. S2). Using the BLUF domain [87] allows the researcher to trigger the enzymatic activity by illumination of blue light into the fusion construct [72]. The enzymes tested are not affected in their catalysis; however, after fusion with this domain, they are controlled in their activity and respond in about a minute. We modify the BLUF domain using the Ranganathan approach to allow a more rapid reaction (a few seconds) [72]. After fusion with this domain, various light-induced domains with chromophores can render the protein light-gated so that the enzyme activity starts only if the light of specific wavelength hits the light-gating domain.

Alternative approaches include optogenetics or neurogenetics and inducible promotor systems [88] for protein expression. However, our specific system applies to any protein of choice, rendering the protein light-gated and light-controlled. At the same time, the switching behavior can be fine-tuned and different domains and light-gating wavelengths are available.

Applications of this toolbox to design *in silico* and engineer in vitro molecular switches into proteins are many, including in biological systems to study cooperative changes in signaling proteins such as two-component systems, in biotechnology to control fluorescence in GFP, for use in optogenetics or enzyme activities in general by light-mediated cooperative change in activity. PRO-Simat allows to identify best places to insert light-gated control for a protein in the network of interest by examining resulting dynamics and modified interactions. Most visionary applications involve intelligent enzyme networks for information processing or smart nano-factories [72] where the start, stop, substrate change or speed of the enzyme cascade is controlled by light. By this, for technical DNA storage, the BLUF domain-guided polymerase does not feed into the cell cycle to change information processing, but rather synthesizes synthetic DNA while light-guided exonucleases and fluorescent oligonucleotides allow read-out from the artificial DNA storage. Fig. 4f and Fig. 4g shows both design steps analysed and implemented with the PRO-Simat tool. The protein network engineering with inserting the BLUF domain is shown in Fig. 4f and the resulting active process, now light-controlled, is examined using PRO-Simat in Fig. 4g.

In practice, it turns out that a nanocellulose chassis works best as support for the storage. Based on this support, more rapid processing of the stored information is achieved step-by-step: Conventional electronic parts such as light-responsive diodes and light-emitting diodes are easily integrated into the nanocellulose as well as semi-conducting "green" nanocellulose paper with semiconductors [72,89]. Next, one can integrate transistors for even faster processing or even DNA aptamer transistors [90]. Our point is here to demonstrate that these very modern design steps are also easily implemented and analyzed using PRO-Simat. More details on the design steps shown in Fig. 4 are given in Supplementary file 2. Supplemental file 2 demonstrates this by additional examples from bacteria (engineering of knock-outs), plants (carbon harvesting plants) and further pathway engineering in synthetic biology.

4. Discussion

PRO-Simat combines unique capabilities for network set-up, structural network analysis and dynamic network simulation. We

demonstrate rapid network set-up for host-pathogen interactions up to designing an oncolytic virus or modifying a DNA storage device. Structural network analysis rapidly identifies involved protein interactions but even allows identification of targets in cellular differentiation and cancer pathways. Dynamic network analysis describes mutually exclusive bacterial differentiation pathways in biofilm formation, optimizes oncolysis in cancer cells while the provided optogenetic enzymes transfer designed networks into synthetic biology constructs.

While database size is one of the typical limitations of such a software, we strived to have here a strong starting backbone of 8 million protein-protein interactions across 32 model organisms and the human proteome while specific further data sets of interest, in particular omics data, can easily be loaded up and incorporated by the user into PRO-Simat.

Network set-up and structural analysis is supported by a number of scripts, features and R shiny apps. Here the key limitation is that PRO-Simat works semi-automatically, in a dialog and requiring edits and additions by the user. However, dialogs, pop-up menus and warnings allow any level of user to learn and train PRO-Simat for optimal results.

Dynamic network analysis using R SQUADD package allows the user to monitor, print and follow the dynamical behaviour of each and all proteins involved in the network. A clever dynamic interpolation routine allows rapid semi-quantitative drawing of trajectories and dynamical behaviour. However, ground state values and network topology need to be fine-tuned in light of available experimental data in an iterative process of model refinement and comparison with experimental data. This process requires an experienced user, but then the model delivers even quantitative results for the whole dynamic network simulation (data-driven modelling).

Researchers not familiar with bioinformatics tools often struggle to find the right tool or platform to study their biological protein network or designing new circuits in synthetic biology or study medical pathways and diseases. Several tools are already available to analyse structural network features, for example, ShinyGO application, which provides a graphical illustration of enrichment, pathway, gene characteristics, and PPI networks [10]. In ShinyGO, however, the PPI data are retrieved directly from the STRING database. In contrast, the PRO-Simat tool offers an embedded database compiled from different sources (e.g., STRING, Mint, and IntAct) so that the users can search proteins via UniProtKB IDs. Another interesting example is the NASQAR tool that integrates several analyses such as differential gene expression, and the enrichment analysis of GO terms and the KEGG pathway [11]. However, PRO-Simat allows the user now to combine powerful dynamic network simulation with in depth pathway analysis and network design options.

Our work is filling a gap between computing and biological systems (broadly defined, but mainly on the side of PPIN and GRN analysis) and even gives useful suggestions for operations research or computer science in particular considering more computing-based approaches for protein-protein interaction (PPI) detection [5] as well as clustering and network analysis methods [6,7].

Compared to available web tools, PRO-Simat contains powerful advances, such as a more interactive user interface (e.g., closer look

at the network and editable PPI network graphics), an embedded comprehensive database (8 million protein data, 40000 signalling network information), and the integration of Jimena dynamic simulation as well as synthetic biology network editing. PRO-Simat provides an efficient and fast PPI network visualization compared to similar tools (e.g., ShinyGO) as the PPI network data are directly retrieved from the embedded database. In addition, the obtained network graphs can be modified with the changes in layout and node shape options. Both graphical and tabular outputs of gene ontology information and enrichment analysis can be easily assessed in the PRO-Simat. As for the graphical outputs, the PRO-Simat wraps almost all types of plotting options for the visualization of enrichment and Jimena dynamic simulation analyses.

Unlike other databases, PRO-Simat allows the user to always filter information at the beginning of the PPI analysis. Therefore, the interactions of the selected two organisms are shown directly (those with different taxonomy IDs are omitted). This allows the creation of a protein-protein interaction network more quickly, and the first 100 interactions of the determined species are presented. With the help of the "More" button, the number of interactions is increased, and 150 new interactions are added to the network. The "Less" button allows removal of 140 interactions at a time if there are more than 100 interactions shown, but only 10 interactions at a time if there are less than 100 interactions shown. If a negative number is reached by this feature, then an additional 100 interactions are added to the negative number and the resulting number of interactions is shown. Finally, for a single protein query only 10 interactions are initially shown. The interaction information of all the proteins in the network is presented in the form of a table and a visualized network with a search feature. Interactive nodes are in clickable form, and gene ontology information of each protein is displayed as a modal table.

With the help of the integrated Jimena framework, PRO-Simat provides the user with the opportunity to perform online Boolean genetic regulatory network analysis. The activity degrees and simulation times of the proteins are changed in the perturbation section and the outputs are shown both visually and in tabular form. The simulation data given in the form of a table is suitable for simulation visualization with commonly used programming languages such as R and Python. Our three use case examples stress how powerful and rapid design can be implemented by the PRO-Simat tool. The different constructs essential for the specific implementation are made available and should stimulate research in the biological application areas examined. Importantly, the application of the PRO-Simat tool is broader and not limited to only these example areas.

5. Conclusion

The tool PRO-Simat provides free access to study protein networks for all researchers with or without computational experience. It offers a comprehensive toolbox of interactive bioinformatics and visualization applications for PPI network, GO and KEGG enrichment, and Jimena dynamic simulation analyses. For PRO-Simat´s efficient implementation of pathway design, we also offer to the community the light-gated constructs established and the vaccinia virus strains examined here.

CRediT authorship contribution statement

R.S. programmed, tested and implemented PRO-Simat, supervised by E.B. and T.D. M.S. modelled pathway changes in *B.subtilis*. C.L. provided programming and database expertise for PRO-Simat. K.S. provided insights and analysis on cooperative behavior in technical multi-component systems. A.S. provided synthetic oncolytic virus and vaccine validation data. E.B. engineered light-gated proteins. E.B. and T.D.

led and guided the study, supervised and analyzed the data and prepared the manuscript. All authors edited and polished the manuscript together and approved the final version.

Declaration of Competing Interest

All authors (RS, MS, CL, KL, AS, EB and TD) declare that they have no competing interests. This includes no financial interests.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.04.023.

References

[1] Harris MA, et al. (57 authors). Gene Ontology Consortium. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 2004;32:D258–61.
[2] Zhou X, Su Z. EasyGO: gene Ontology-based annotation and functional enrichment analysis tool for agronomical species. BMC Genom 2007;8(1):1–4.
[3] Wei Q, Khan IK, Ding Z, Yerneni S, Kihara D. NaviGO: interactive tool for visualization and functional similarity and coherence analysis with gene ontology. BMC Bioinforma 2017;18(1):1–13.
[4] Thanati F, Karatzas E, Baltoumas FA, Stravopodis DJ, Eliopoulos AG, Pavlopoulos GA. Flame: a web tool for functional and literature enrichment analysis of multiple gene lists. Biol (Basel) 2021;10(7):665.
[5] Rao VSrinivasa, et al. Protein-protein interaction detection: methods and analysis. Int J Proteom 2014.
[6] Bhowmick Sourav S, Seah Boon Siew. Clustering and summarizing protein-protein interaction networks: a survey. IEEE Trans Knowl Data Eng 2015;28.3:638–58.
[7] Rasti Saeid, Vogiatzis Chrysafis. A survey of computational methods in protein-protein interaction networks. Ann Oper Res 2019;276:35–87.
[8] Backes C, Keller A, Kuentzer J, Kneissl B, Comtesse N, Elnakady YA, et al. GeneTrail-advanced gene set enrichment analysis. Nucleic Acids Res 2007;35(SUPPL.2).
[9] Priebe S, Kreisel C, Horn F, Guthke R, Linde J. FungiFun2: a comprehensive online resource for systematic analysis of gene lists from fungal species. Bioinformatics 2015;31(3):445–6.
[10] Ge SX, Jung D, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics 2020;36(8):2628–9.
[11] Yousif A, Drou N, Rowe J, Khalfan M, Gunsalus KC, Gunsalus KC. NASQAR: a web-based platform for high-throughput sequencing data analysis and visualization. BMC Bioinforma 2020;21(1).
[12] Jain A, Tuteja G. PlacentaCellEnrich: a tool to characterize gene sets using placenta cell-specific gene enrichment analysis. Placenta 2021;103:164–71.
[13] Nguyen H, Tran D, Galazka JM, Costes SV, Beheshti A, Petereit J, et al. CPA: a web-based platform for consensus pathway analysis and interactive visualization. Nucleic Acids Res 2021;49(W1):W114–24.
[14] Jalili Mahdi, et al. CentiServer: a comprehensive resource, web-based application and R package for centrality analysis. PloS One 2015;10(11):e0143111.
[15] Osmanoglu Ö, Shams S, Dandekar T, Naseem M. Modeling immune dynamics in plants using JIMENA-package. Methods Mol Biol 2021:183–9.
[16] Kaltdorf M, Breitenbach T, Karl S, Fuchs M, Kessie DK, Psota E, Prelog M, Sarukhanyan E, Ebert R, Jakob F, Dandekar G, Naseem M, Liang C, Dandekar T. Software JimenaE allows efficient dynamic simulations of Boolean networks, centrality and system state analysis. Sci Rep 2023. (in press).

[17] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. R Foundation for Statistical Computing. R Found Stat Comput Vienna, Austria 2012;

[18] Chang W., Cheng J., Allaire J., … YXS http://citeseerx. ist, 2015 U. Package "shiny." cran.microsoft.com 2015;

[19] Perrier V. and Meyer F. (2020). fresh: Create Custom 'Bootstrap' Themes to Use in 'Shiny'. R package version 0.2.0 https://CRAN-project.org/package=fresh.

[20] Attali D. shinyalert: Easily Create Pretty Popup Messages (Modals) in "Shiny". R package version 2;2021.

[21] Sali A., Attali D. shinycssloaders: Add Loading Animations to a "shiny" Output While It's Recalculating. R package version 1.0.0. 2020.

[22] Chang W. shinydashboard: Create Dashboards with "Shiny". https://CRAN.R-project.org/package=shinydashboard. 2015.

[23] Attali D. shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds. R package version 1.1. 2020.

[24] Chang W., Park T., Dziedzic L., Willis N., McInerney M. Themes for Shiny. R Package Version 1.2.;2018.

[25] Perrier V., Meyer F., Granjon D. shinyWidgets: Custom Inputs Widgets for Shiny. 2022.

[26] Murrell P., Potter S. gridSVG: Export "grid" Graphics as SVG. 2022.

[27] Wickham H., Henry L., Pedersen T.L., Luciani T.J., Decorde M., Lise V. svglite: An "SVG" Graphics Device. 2022.

[28] Riutta et. al. A, Tangelder J, Russell K. svgPanZoom: R "Htmlwidget" to Add Pan and Zoom to Almost any R Graphic. 2020.

[29] Carlson M. org.Hs.eg.db: Genome wide annotation for Human. 2022.

[30] Carlson M. org.At.tair.db: Genome wide annotation for Arabidopsis. 2022.

[31] Carlson M. org.Ce.eg.db: Genome wide annotation for Worm. 2022.

[32] Carlson M. org.Dr.eg.db: Genome wide annotation for Zebrafish. 2022.

[33] Carlson M. org.Dm.eg.db: Genome wide annotation for Fly. 2022.

[34] Carlson M. Org EcK12 eg db: Genome wide Annot E coli strain K12 2022.

[35] Carlson M. org.Gg.eg.db: Genome wide annotation for Chicken. 2022.

[36] Carlson M. org.Mm.eg.db: Genome wide annotation for Mouse. 2022.

[37] Carlson M. org.Rn.eg.db: Genome wide annotation for Rat. 2022.

[38] Carlson M. org.Bt.eg.db: Genome wide annotation for Bovine. 2022.

[39] Carlson M. org.Cf.eg.db: Genome wide annotation for Canine. 2022.

[40] Carlson M. org.Ss.eg.db: Genome wide annotation for Pig. 2022.

[41] Carlson M. org.Mmu.eg.db: Genome wide annotation for Rhesus. 2022.

[42] Carlson M. org.Xl.eg.db: Genome wide annotation for Xenopus. 2022.

[43] Carlson M. org.Pt.eg.db: Genome wide annotation for Chimp. 2022.

[44] Carlson M. org.EcSakai.eg.db: Genome wide annotation for E coli strain Sakai. 2022.

[45] Pagès H., Carlson M., Falcon S., Li N. AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor. 2022.

[46] Morgan M. BiocManager: Access the Bioconductor Project Package Repository. 2022.

[47] Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics 2015;31(4):608–9.

[48] Yu G. enrichplot: Visualization of Functional Enrichment Result. 2022.

[49] Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics 2013;29(14):1830–1.

[50] Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst 2016;12(2):477–9.

[51] Almende B.V. and Contributors, Thieurmel B. visNetwork: Network Visualization using "vis.js" Library. 2022.

[52] Wickham H., François R., Henry L., Müller K. dplyr: A Grammar of Data Manipulation. 2022.

[53] Xie Y., Cheng J., Tan X. DT: A Wrapper of the JavaScript Library "DataTables". 2022.

[54] Ooms J., James D., DebRoy S., Wickham H., Horner J. RMySQL: Database Interface and "MySQL" Driver for R. 2022.

[55] Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.

[56] Vaidyanathan R., Xie Y., Allaire J.J., Cheng J., Sievert C., Russell K. htmlwidgets: HTML Widgets for R. 2021.

[57] Csardi G, Nepusz T, Luo W, Brouwer C, Yu G, He QY, et al. The igraph software package for complex network research. Bioinformatics 2015;29(14):1830–1.

[58] Sankar M., supervised by Christian Hardtke, Xenarios I. SQUADD: Add-on of the SQUAD Software. 2022.

[59] Kerrien S, Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, et al. The IntAct molecular interaction database in 2012. Nucleic Acids Res 2012;38((SUPPL.1).

[60] Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, et al. MINT, the molecular interaction database: 2012 Update. Nucleic Acids Res 2012;40(D1).

[61] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43(D1):D447–52.

[62] Cook H.V., Doncheva N.T., Szklarczyk D.: Viruses. STRING: a virus-host protein-protein interaction database. Viruses. 2018 Sep 23;10(10):519.

[63] Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. Nat Protoc 2009;4(8):1184–91.

[64] Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics 2009;25(22):3045–6.

[65] Carlson M. UniProt. ws: a package for retrieving data from the UniProt web service. R Packag Version 2012;2180:1–3.

[66] Türei D, Korcsmáros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat Methods 2016;13(12):966–7.

[67] Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Yu G. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation 2021;2(3):100141.

[68] Wickham H. Reshaping data with the reshape package. J Stat Softw 2007;21(12):1–20.

[69] Sievert C. Interactive Web-Based Data Visualization with R, plotly, and shiny. Interact Web-Based Data Vis with R, plotly, shiny 2020;

[70] Di Cara A, Garg A, De Micheli G, Xenarios I, Mendoza L. Dynamic simulation of regulatory networks using SQUAD. BMC Bioinforma 2007:8.

[71] Salinas VH, Ranganathan R. Coevolution-based inference of amino acid interactions underlying protein function. Elife 2018:7.

[72] Bencurova E, Shityakov S, Schaack D, Kaltdorf M, Sarukhanyan E, Hilgarth A, et al. Nanocellulose composites as smart devices with chassis, light-directed DNA storage, engineered electronic properties, and chip integration. Front Bioeng Biotechnol 2022;10:1243.

[73] Zhu B, Stülke J. SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism Bacillus subtilis. Nucleic Acids Res 2018;46(D1):D743–8.

[74] Widderich N, Rodrigues CDA, Commichau FM, Fischer KE, Ramirez-Guadiana FH, Rudner DZ, et al. Salt-sensitivity of σH and Spo0A prevents sporulation of Bacillus subtilis at high osmolarity avoiding death during cellular differentiation. Mol Microbiol 2016;100(1):108–24.

[75] Hou X, Yu X, Du B, Liu K, Yao L, Zhang S, et al. A single amino acid mutation in Spo0A results in sporulation deficiency of Paenibacillus polymyxa SC2. Res Microbiol 2016;167(6):472–9.

[76] Wang X, Luo C, Liu Y, Nie Y, Liu Y, Zhang R, et al. Three non-aspartate amino acid mutations in the ComA response regulator receiver motif severely decrease surfactin production, competence development, and spore formation in Bacillus subtilis. J Microbiol Biotechnol 2010;20(2):301–10.

[77] Tanaka T, Kawata M, Mukai K. Altered phosphorylation of Bacillus subtilis DegU caused by single amino acid changes in DegS. J Bacteriol 1991;173(17):5507–15.

[78] López D, Vlamakis H, Losick R, Kolter R. Cannibalism enhances biofilm development in bacillus subtilis. Mol Microbiol 2009;74(3):609–18.

[79] Zhang Q, Liang C, Yu YA, Chen N, Dandekar T, Szalay AA. The highly attenuated oncolytic recombinant vaccinia virus GLV-1h68: Comparative genomic features and the contribution of F14.5L inactivation. Mol Genet Genom 2009;282(4):417–35.

[80] Adelfinger M, Bessler S, Frentzen A, Cecil A, Langbein-Laugwitz J, Gentschev I, et al. Preclinical testing oncolytic vaccinia virus strain GLV-5b451 expressing an anti-VEGF single-chain antibody for canine cancer therapy. Viruses 2015;7(7):4075–92.

[81] Cecil A, Gentschev I, Adelfinger M, Dandekar T, Szalay AA. Vaccinia virus injected human tumors: oncolytic virus efficiency predicted by antigen profiling analysis fitted boolean models. Bioengineered 2019;10:190–6.

[82] Liang C, Krüger B, Dandekar T. GoSynthetic database tool to analyse natural and engineered molecular processes. Database 2013:2013.

[83] Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res 2019;47(W1):W556–60.

[84] Minev BR, Lander E, Feller JF, Berman M, Greenwood BM, Minev I, et al. First-in-human study of TK-positive oncolytic vaccinia virus delivered by adipose stromal vascular fraction cells. J Transl Med 2019;17(1).

[85] Inoue M, Kim M, Inoue T, Tait M, Byrne T, Nitschké M, Murer P, Cha H, Subramanian A, De Silva N, Chiaverotti T, McDonald DM. Oncolytic vaccinia virus injected intravenously sensitizes pancreatic neuroendocrine tumors and metastases to immune checkpoint blockade. Dec21 Mol Ther Oncolytics 2021;24:299–318. https://doi.org/10.1016/j.omto.2021.12.016

[86] Malhotra J, Kim ESOncolytic Viruses. and Cancer Immunotherapy. (Jan). Curr Oncol Rep 2023;25(1):19–28. https://doi.org/10.1007/s11912-022-01341-w

[87] Banerjee S, Mitra D. Structural basis of design and engineering for advanced plant optogenetics. Trends Plant Sci 2020;Vol. 25:35–65.

[88] Das T, Tenenbaum A, Berkhout L, Tet-On Systems B. For doxycycline-inducible gene expression. Curr Gene Ther 2016;16(3):156–67.

[89] Jung YH, Chang TH, Zhang H, Yao C, Zheng Q, Yang VW, et al. High-performance green flexible electronics based on biodegradable cellulose nanofibril paper. Nat Commun 2015:6.

[90] Nakatsuka N. Aptamer–Field-Effect Transistors for Small-Molecule Sensing in Complex Environments. Methods in Molecular Biology Humana Press Inc.; 2023. p. 187–96.