Published online 29 April 2016 *Nucleic Acids Research, 2016, Vol. 44, Web Server issue* ***W507–W513***
doi: 10.1093/nar/gkw335

systemsDock: a web server for network pharmacology-based prediction and analysis

Kun-Yi Hsin[1,*], Yukiko Matsuoka[2], Yoshiyuki Asai[1], Kyota Kamiyoshi[1], Tokiko Watanabe[3], Yoshihiro Kawaoka[3,4] and Hiroaki Kitano[1,2,5,*]

[1]Integrated Open Systems Unit, Okinawa Institute of Science and Technology Graduate University, Onna-son, Okinawa 904-0412, Japan, [2]The Systems Biology Institute, Minato, Tokyo 108-0071, Japan, [3]Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, The University of Tokyo, Minato, Tokyo 108-8639, Japan, [4]Department of Pathobiological Sciences, Influenza Research Institute, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, Wisconsin 53711, USA and [5]Laboratory for Disease Systems Modeling, RIKEN Center for Integrative Medical Sciences, Yokohama City, Kanagawa 230-0045, Japan

Received January 27, 2016; Revised April 8, 2016; Accepted April 15, 2016

ABSTRACT

We present systemsDock, a web server for network pharmacology-based prediction and analysis, which permits docking simulation and molecular pathway map for comprehensive characterization of ligand selectivity and interpretation of ligand action on a complex molecular network. It incorporates an elaborately designed scoring function for molecular docking to assess protein–ligand binding potential. For large-scale screening and ease of investigation, systemsDock has a user-friendly GUI interface for molecule preparation, parameter specification and result inspection. Ligand binding potentials against individual proteins can be directly displayed on an uploaded molecular interaction map, allowing users to systemically investigate network-dependent effects of a drug or drug candidate. A case study is given to demonstrate how systemsDock can be used to discover a test compound's multi-target activity. systemsDock is freely accessible at http://systemsdock.unit.oist.jp/.

INTRODUCTION

Drugs may interact with multiple molecules in the human body, and such drug action, known as polypharmacology, may be efficacious or deleterious for the treatment of disease. For example, β-lactams exhibit antibacterial action principally by targeting multiple penicillin-binding proteins (1). Similarly, a multi-target strategy has advanced the treatment of neurodegenerative diseases (2). On the other hand, poor drug selectivity may increase therapeutic risks and negatively impact drug development because of unintended drug-target interactions. An example of this is the cardiotoxicity of kinase inhibitor Sunitinib (3). Identification of drug targets is therefore a critical stage of drug development (4,5).

It is only relatively recently that interactive web interfaces for prediction of drug-target interactions have become available to non-commercial research groups. One such resource is DINIES (6), which is based on supervised machine learning methods to predict unknown drug–target interaction networks. SwissTargetPrediction (7) predicts targets on the basis of molecular similarity. Results from these resources mainly focus on predicting interaction targets, but network pharmacology information is limited. Advances in systems biology are creating opportunities for the application of network pharmacology to identify drug targets over molecular pathway maps (8).

A well-curated pathway map explicitly describes molecular interactions. Deeply curated maps, such as Toll-Like Receptor pathways (9), mTOR pathways (10), EGFR pathways (11), AlzPathway (signaling pathways of Alzheimer's disease) (12) and FluMap (pathways describing the influenza A virus replication cycle) (13), allow systematic understanding of specific disease-relevant processes. Those maps were composed in machine-readable standard Systems Biology Markup Language (SBML) format and visualized as a Systems Biology Graphical Notation diagram, which enables computational modeling and simulation. Well-annotated databases like KEGG (14), Reactome (15) and PANTHER (16) with many pathway entries provide information on various molecular interactions and reaction networks. Structured diagram editors such as CellDesigner (17) enable users to draw, browse or modify biological networks of interest. Payao (18) allows map sharing and community comments through a social network service.

*To whom correspondence should be addressed. Tel: +81 98 966 8711; Fax: +81 98 966 2152; Email: kunyi.hsin@oist.jp
Correspondence may also be addressed to Hiroaki Kitano. Tel: +81 98 966 8711; Fax: +81 98 966 2152; Email: kitano@sbi.jp

These open resources allow a deeper understanding of how small molecules acts within molecular networks.

In this article, we describe systemsDock, a web service specifically designed to make network-based screening efficient and practical. It incorporates an elaborately designed scoring function called docK-IN (combining docking with intelligence) for molecular docking to assess protein–ligand binding potential. docK-IN was initially developed in our previous work (19) to address the critical issue of commonly used docking programs that were too inaccurate for reliable prediction (20). Another unique feature of systemsDock is that it allows screening of a large number of proteins with ease. This is distinguished from other web resources by having a series of efficient stepwise methods for molecule preparation, parameter specification, and result inspection. Predictions can be displayed in an interactive table, histogram or projected on a pathway map for intuitive inspection.

THE systemsDock METHOD AND INTERFACE

To carry out a network pharmacology-based prediction and analysis using systemsDock, the required processes are divided into four main steps, including (i) specification of target proteins by various efficient options, or by direct upload of a molecular pathway map in SBML format, (ii) definition of binding sites through an interactive molecular visualizer by clicking on the displayed structure model or residues listed, (iii) preparation of small molecules for test either by uploading structure files or by composing a compound of interest using the provided web-based molecule editor, (iv) run docking simulation and proceed to the inspection of the result.

Molecule preparation

For network-based screening, systemsDock provides highly efficient flexible options that allow users to test a large number of proteins by specifying (i) protein names or gene symbols, (ii) protein PDB IDs or (iii) to upload a pathway map file in SBML format. Selected protein structures with the best resolution are then retrieved from an in-house protein identity-to-structure mapping system for docking simulation. Mapping data are collected mainly from UniProt (21), GenBank (22), ChEMBL (23) and PhosphoSitePlus (24). Through the system, it is possible to rapidly select small, related families of molecules for testing. An extension of this selection procedure provides a very efficient way to specify protein structure and binding site. Considering the data instantaneity, structure files are retrieved from a synced copy of the PDB archive on the serve, or dynamically downloaded from the RCSB PDB database whilst data is unavailable locally. A binding site for each protein is automatically identified by exploring the position where the biggest native ligand is bound. Alternatively, binding sites can be defined through an interactive molecular visualizer by clicking on the displayed structure model or amino acids listed in the sequence table (Figure 1A).

systemsDock accepts compounds in commonly used formats, including 2D/3D SDF, Mol2 or SMILES. Users can also employ an integrated web-based molecule editor (JSME; http://peter-ertl.com/jsme/) to compose compounds of interest. Compounds will be displayed in various representations with calculated molecular properties. There are also convenient links for compounds in external databases including PubChem (25), DrugBank (26), BindingDB (27) and KEGG, as well as over 130 chemical vendors, allowing easy access to biological data and identifying commercially available molecules (Figure 1B).

Docking simulation and performance

systemsDock applies AutoDock VINA (28) to perform docking simulation. Based on the characterized binding interaction and molecular properties, all of the binding modes are then rescored by docK-IN and ranked accordingly. docK-IN utilizes a machine learning algorithm (Random Forest) together with a series of characterized binding interactions and test compound molecular properties (Supplementary Data S1) to perform the rescoring. The best binding mode is selected for each test compound. Unlike other scoring methods (29,30), the score reported by docK-IN is a negative logarithm of the experimental dissociation/inhibition constant (pK_d/pKi), usually ranging from 0 to 10 (i.e. from weak to strong binding), allowing a straightforward indication of binding strength. Depending upon molecule specifications, docking simulations through systemsDock may take a long time. Users can check the progress or retrieve results simply using a given web link or session ID. An email notification is sent to the user when simulation is completed.

The scoring function of docK-IN is a machine learning model based upon the Random Forest algorithm. Training data were mainly composed of a set of quality co-crystallized complexes with their corresponding binding affinity data obtained from PDBbind (31). We have applied the newly released PDBbind collection to increase the training data of machine learning models and to extend learning assumptions. Davis *et al*. (32) proposed a quantitative analysis of the selectivity of numerous kinase inhibitors against diverse kinases using an *in vitro* competition binding assay. It was of great interest to assess the consistency between bioassay outcomes and docK-IN predictions. Taking protein structure availability and binding site certainty into consideration, we tested 120 different kinases against 63 inhibitors. Docking results of those molecules are listed in Supplementary Data S2, and parameters and settings for docking simulations are summarized in Supplementary Data S3. In order to evaluate and visualize the screening performance, we applied Receiver operating characteristic (ROC) (33) and ROCR (34) package in R which estimates the sensitivity and specificity based on the measured rates of classification accuracy. Figure 2 presents the results of ROC analysis showing the performance of docK-IN in classifying test compound activity. The true positive rate (vertical axis of Figure 2A) represents the fraction of measurements in which the activity is predicted correctly. The false positive rate (horizontal axis) represents the fraction in which inactive compounds are incorrectly classified as active ones. The calculated AUC (Area Under the ROC Curve) measured by interpolating a series of docking score cutoffs is 0.84. AUC is a measure of prediction accuracy ranging from 0–1. A conventional rating of binary classifiers according to AUC

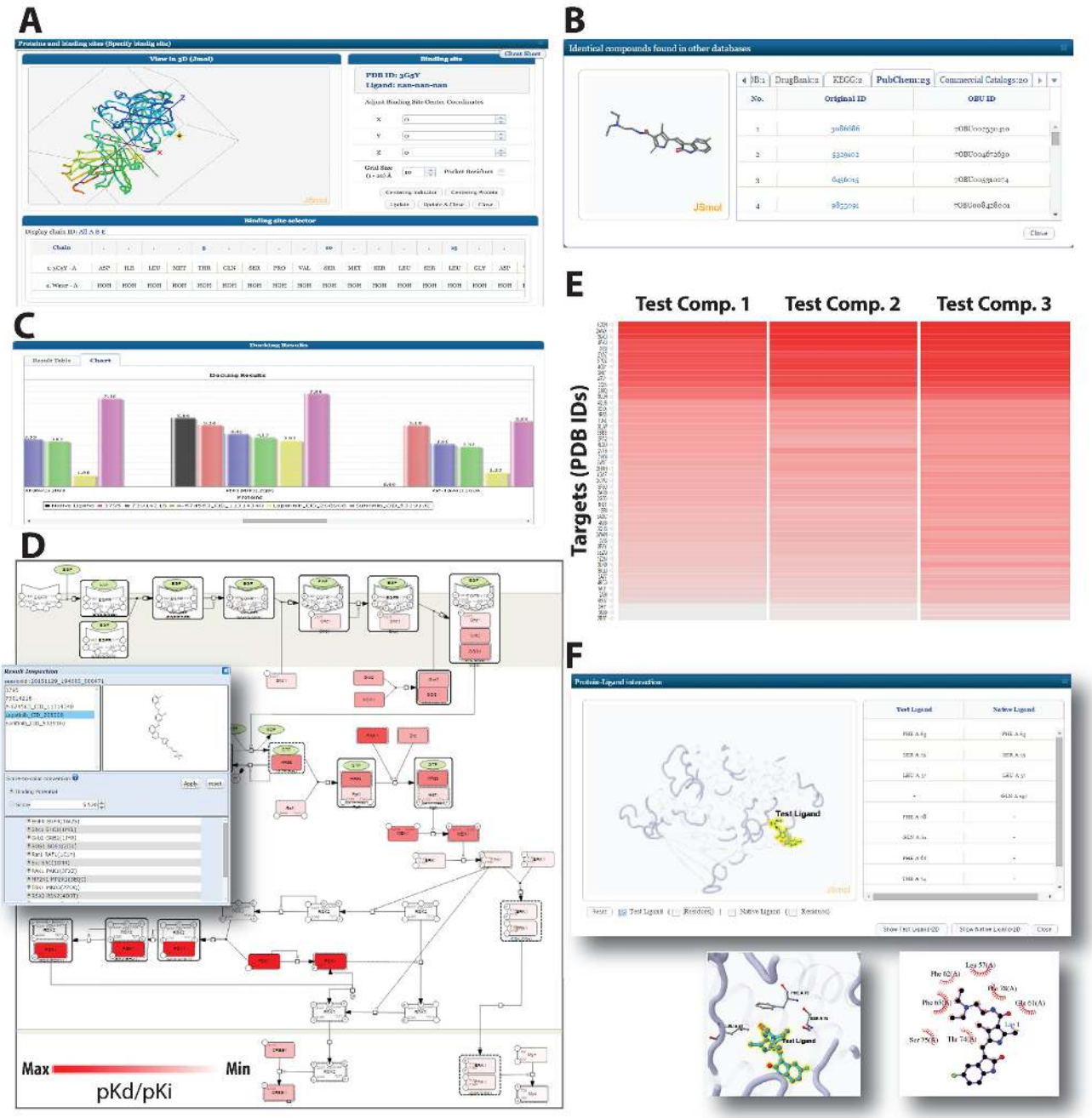

Figure 1. Screenshots of the systemsDock web interface. (**A**) Interactive functions for binding site specifications are accessed by clicking on the displayed protein structure or amino acids listed in the sequence table to define the location of the preferred binding site. Users can adjust x-y-z coordinates to refine the location. (**B**) Links are provided for the test compound in external databases, as well as to visualize the compound in 3D. (**C**) Prediction results are furnished in an interactive histogram. Docking scores for each compound are grouped by proteins. By clicking on one of the bars, molecular binding interactions can be graphically shown in 2D/3D for structure-based investigation as shown in (**F**). (**D** and **E**) Visualizing results through a pathway map provided by the user or using a heat map. Colors of proteins are displayed as white-to-red scales or as white and red according to the docking scores. Click on a colored node (i.e. protein) to display binding interactions in 2D/3D as shown in (**F**). (**F**) Visualizing protein–ligand binding interactions of the test compound or native ligand in 2D/3D. Protein residues involved in the binding interaction are automatically identified. For reference, those that interacted with a native ligand, if available, are also listed. Clicking on any of the residue entries listed allows users to center and display the specified residue for closer inspection.

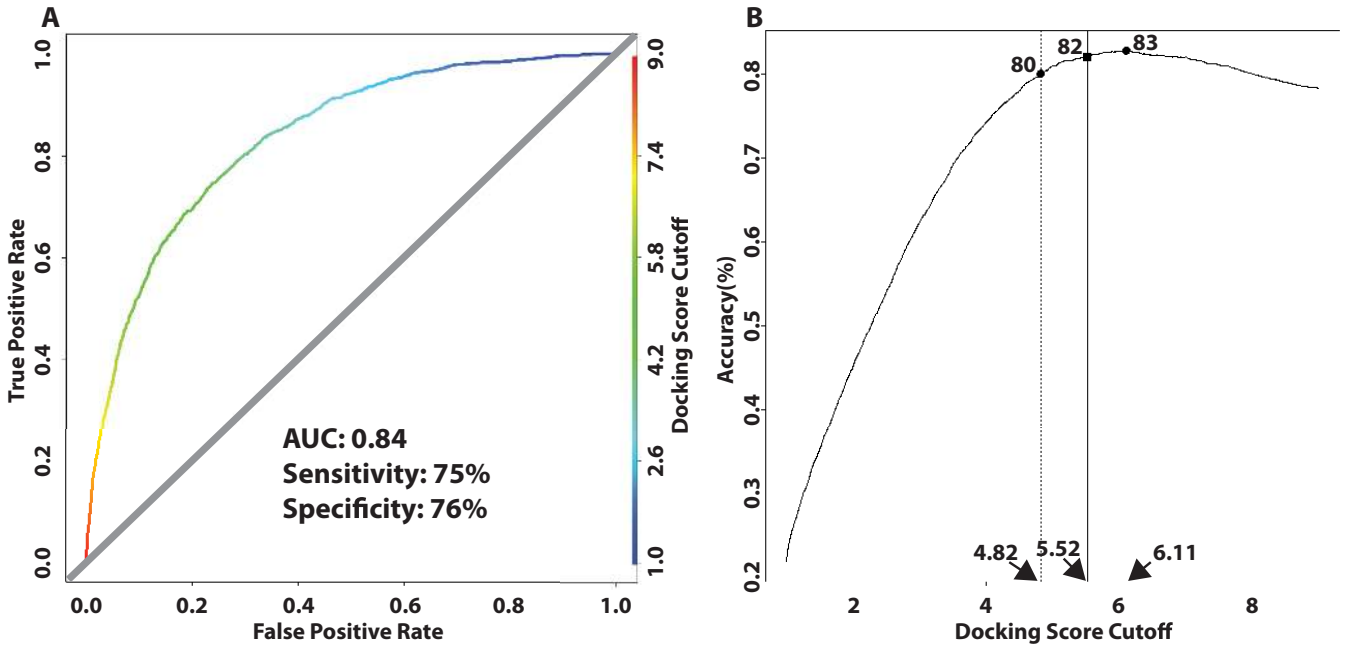

Figure 2. Receiver operating characteristic (ROC) analysis for evaluating performance of docking simulation using docK-IN on the demanding bioassay benchmark (32). The test included the results of 63 test compounds against 120 kinases. (**A**) ROC plot measured by a series of docking score cutoffs. As a reference, the gray line represents theoretical random results. The colored ROC curve is the fraction of the true positives out of the total actual positives versus the fraction of false positives out of the total actual negatives. The calculated AUC (Area Under the ROC Curve), sensitivity and specificity are 0.84, 75 and 76%, respectively. (**B**) Prediction accuracy ((True Positive + True Negative)/Total population) measured by a series of docking score cutoffs. Good accuracy level (80–83%) was observed in the range of cutoff scores from 4.82 to 6.11 (pK_d). The docking score 5.52 (pK_d) residing within the range is equal to a dissociation constant (K_d) of 3 μM, which is conventionally used to classify ligand binding activity.

is as follows: $0.9 \leq AUC \leq 1$ is considered excellent; $0.80 \leq AUC < 0.9$ is good; $0.70 \leq AUC < 0.8$ fair; $0.50 \leq AUC < 0.7$ poor; and $AUC < 0.5$ represents failure (no better than a random classifier). Taking the optimal cutoff value at the point of the ROC curve that is closest to a perfect classifier (i.e. the top-left corner of Figure 2A), we obtained the values of 75% sensitivity (i.e. compounds correctly identified as active) and 76% specificity (compounds correctly identified as inactive). According to the accuracy assessment ((True Positive + True Negative)/Total population) shown in Figure 2B, a good accuracy level (80–83%) was observed when cutoff scores were in the range of 4.82 to 6.11 (pK_d). A docking score of 5.52 (pK_d) is equal to a dissociation constant (K_d) of 3 μM, which is conventionally used to classify ligand binding activity. As demonstrated by the results of ROC analysis, docK-IN performed well on prediction, and was able to classify the activity of test compounds with good accuracy when the docking score at 5.52 was set as classifier.

Result output and inspection

Docking scores, predicted binding affinities for target proteins, are displayed in an interactive table and histogram (Figure 1C). By clicking on a table entry or histogram bar, molecular binding interactions can be graphically displayed in 2D/3D for structure-based investigation. Analyzed by LIGPLOT (35), protein residues involved in intermolecular interactions will be highlighted for rapid inspection. Screening result files from systemsDock are rich in details. Tabulated docking scores (csv file), docked poses (SDF file), processed protein structures (PDB file), and predicted binding interactions (png file) may be downloaded. The pathway map (SBML file, svg/png file), if available, may be saved. Depending upon availability of the pathway map, docking scores of the test compound against specified network proteins are optionally converted into a white-to-red color scale (indicating binding strength) or they may be classified as active (red) or inactive (white) relative to a specified cutoff. Colored results are then projected to the pathway map to directly display predicted binding affinities (Figure 1D). For larger data visualization and analysis, an interactive heat map with sorting and zooming functions is also available for users (Figure 1E), through which users can efficiently identify potential binders or known targets from plenty of tests.

systemsDock provides a set of efficient functions to manipulate the pathway map. For example, a bird's-eye-view function facilitates navigating over a pathway map. An object-list panel lists object information (e.g. species, proteins and reactions) shown on the map. Then, by simply by clicking on an entry one can center and highlight the object in the view. Users can efficiently investigate the bioactivity of a test compound against numerous proteins through the colored pathway, allowing a perspective inspection for results throughout a complex signaling network. By clicking on colored nodes (proteins), a molecule visualizer (JSmol) will be launched to illustrate the intermolecular interactions in 2D/3D (Figure 1F). Screening results are retrievable by a given web link or session ID. There is no login requirement

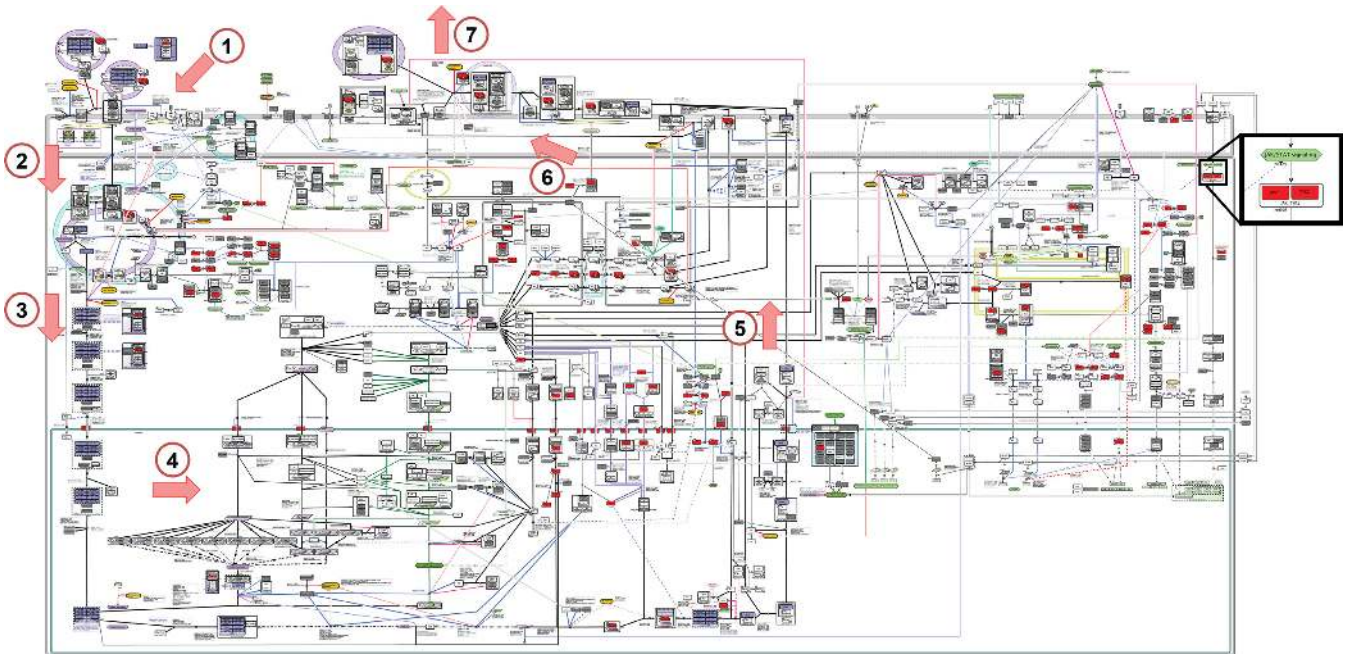

Figure 3. Prediction of a test compound (PubChem CID: 3795) against 180 FluMap proteins using systemsDock. FluMap describes the intermolecular network of influenza A virus life cycle from (1) entry, (2) endocytosis, (3) fusion, (4) transcription/translation, (5) assembly, (6) packaging to (7) budding. Docking scores are converted into white/red color according to a cutoff value at 5.52 pK_d (white: inactive; red: active; gray: not in the test or no results). Results are projected on the pathway map. The JAK1 protein, which plays a major role in the JAK/STAT signaling pathway, is enclosed in a black frame. See Supplementary Data S5 for a high resolution inspection.

to use systemsDock, but historical sessions and results may be retrieved and managed after login verification.

APPLICATION AND CASE STUDY

We applied systemsDock to investigate compounds for anti-influenza therapy. For a preliminary test, we validated two neuraminidase (NA) inhibitors, namely Oseltamivir and Peramivir, against a pool of proteins involved in the influenza infection process to verify their binding potentials. As Oseltamivir is a prodrug, we tested both Oseltamivir and Oseltamivir acid. We ranked the screening results by docking scores. Among the test proteins, the major target (i.e. NA) of the inhibitors was identified as a potential binder whilst the docking scores (Oseltamivir acid: 6.5; Oseltamivir: 6.9; Peramivir: 6.9) to this protein are seen higher than others (avg. 5.7), within the top 10 out of the 50 predicted targets (detailed in Supplementary Data S4-1). Conventional anti-influenza drugs target viral proteins (e.g. M2 or NA), but drug resistance has emerged due to the rapidity of viral mutation. Recent studies have shifted the focus to targeting host factors instead of viral proteins, and among such studies, JAK1 was recognized as a potential antiviral drug target through virus-host interactome analysis (36). Here we studied JAK1 inhibitors and applied systemsDock using compound structures with a deeply curated influenza infection pathway called FluMap (13).

The pathway map of FluMap describes the molecular pathway of the influenza A virus replication cycle (i.e. entry, endocytosis, transcription/translation, assembly and budding). It is composed of both viral and host factors (i.e. proteins, mRNAs etc.) together with over 450 reactions, enabling computational network, pharmacology-based analyses. FluMap is detailed on its website (http://www.influenza-x.org/flumap/) and the map file in SBML format is available (13).

Screening the FluMap pathway provides an opportunity to identify effective anti-viral host targets. For example, compounds with kinase inhibition activity may interrupt the protein assembly phase of the viral life cycle. Here we display the results of screening a test compound (PubChem CID: 3795) against 180 proteins of interest (Supplementary Data S4-2), and we project the predicted binding potential on the FluMap pathway to demonstrate systemsDock's capability. The compound also shows *in vitro* efficacy in follow-up validation on siRNA screening.

A colored FluMap shows network-based screening results (Figure 3). A dissociation constant of 3 μM is a conventional value to define ligand binding, and it is equal to a docking score (pK_d) of 5.52. We therefore set this docking score as the cutoff to classify the test compound's activity. Proteins in gray indicate cases not in our test because 3D structures were not available, or because no binding mode was generated by the docking tool. The compound showed a good binding selectivity in that it interacted with only 20% of proteins tested. Although it interacted with several non-viral proteins involved in different signaling pathways, only a small number of them (including FPPS, p38, Hsp90 and ERK) were identified as critical factors for the virus life cycle identified by our previous controllability analysis (13). This implied a lower potential for off-target effects. It is interesting to note that the compound may interact with viral proteins such as NA, HA, NP, which means that anti-viral

efficacy may be deduced by multi target effects in combination with host and virus factors.

CONCLUSION AND OUTLOOK

systemsDock is the first web service enabling drug developers to carry out network pharmacology-based prediction and analysis by integrating results from structural biology with systems biology. Its user-friendly GUI interface simplifies essential operations for large-scale screening. Using the predictive docking approach, systemsDock can test a large number of target proteins with good prediction accuracy. This will reduce the number of tests for bioassay. As an example, we demonstrated the application of systemsDock in investigation of anti-influenza agents and targets using network pharmacology. Progress in pathway curation as well as open resources provide a great opportunity to rationally optimize drug polypharmacology for pharmaceutical research. Together with a curated pathway map, systemsDock helps to comprehensively characterize the underlying mechanism of a drug candidate and to interpret its cascading effects, improving the prediction of drug efficacy and safety.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We would like to thank our collaborators in the Systems Biology Institute (SBI) and IMSUT, The University of Tokyo, as well as Takeshi Abe, Samik Ghosh, Vipul Gupta, Kenichiro Arakaki, Steve Aird and Reuven Pnini for insightful comments.

FUNDING

HD-Physiology Project of the Japan Society for the Promotion of Science (JSPS); JSPS KAKENHI [26730152]; Institute of Medical Science, The University Of Tokyo (IMSUT) Joint Research Project [2014-343]; Project for Enhancement of Biomedical Industry in Okinawa, Japan; and Advanced Research & Development Programs for Medical Innovation from the Japan Agency for Medical Research and Development (AMED). Funding for open access charge: Okinawa Institute of Science and Technology Graduate University (OIST).

Conflict of interest statement. None declared.

REFERENCES

1. Zapun,A., Contreras-Martel,C. and Vernet,T. (2008) Penicillin-binding proteins and beta-lactam resistance. *FEMS Microbiol. Rev.*, **32**, 361–385.
2. Geldenhuys,W.J. and Van der Schyf,C.J. (2013) Designing drugs with multi-target activity: the next step in the treatment of neurodegenerative disorders. *Expert Opin. Drug Discov.*, **8**, 115–129.
3. Kerkela,R., Woulfe,K.C., Durand,J.B., Vagnozzi,R., Kramer,D., Chu,T.F., Beahm,C., Chen,M.H. and Force,T. (2009) Sunitinib-induced cardiotoxicity is mediated by off-target inhibition of AMP-activated protein kinase. *Clin. Transl. Sci.*, **2**, 15–25.
4. Hopkins,A.L. (2008) Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.*, **4**, 682–690.
5. Yıldırım,M.A., Goh,K.-I., Cusick,M.E., Barabási,A.-L. and Vidal,M. (2007) Drug-target network. *Nat. Biotechnol.*, **25**, 1119–1126.
6. Yamanishi,Y., Kotera,M., Moriya,Y., Sawada,R., Kanehisa,M. and Goto,S. (2014) DINIES: drug–target interaction network inference engine based on supervised analysis. *Nucleic Acids Res.*, **42**, W39–W45.
7. Gfeller,D., Grosdidier,A., Wirth,M., Daina,A., Michielin,O. and Zoete,V. (2014) SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Res.*, **42**, W32–W38.
8. Ghosh,S., Matsuoka,Y., Asai,Y., Hsin,K.-Y. and Kitano,H. (2011) Software for systems biology: from tools to integrated platforms. *Nat. Rev. Genet.*, **12**, 821–832.
9. Oda,K. and Kitano,H. (2006) A comprehensive map of the toll-like receptor signaling network. *Mol. Syst. Biol.*, **2**, 1–20.
10. Caron,E., Ghosh,S., Matsuoka,Y., Ashton-Beaucage,D., Therrien,M., Lemieux,S., Perreault,C., Roux,P.P. and Kitano,H. (2010) A comprehensive map of the mTOR signaling network. *Mol. Syst. Biol.*, **6**, 453.
11. Oda,K., Matsuoka,Y., Funahashi,A. and Kitano,H. (2005) A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.*, **1**, 1–17.
12. Mizuno,S., Iijima,R., Ogishima,S., Kikuchi,M., Matsuoka,Y., Ghosh,S., Miyamoto,T., Miyashita,A., Kuwano,R. and Tanaka,H. (2012) AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. *BMC Syst. Biol.*, **6**, 52.
13. Matsuoka,Y., Matsumae,H., Katoh,M., Eisfeld,A.J., Neumann,G., Hase,T., Ghosh,S., Shoemaker,J.E., Lopes,T.J. and Watanabe,T. (2013) A comprehensive map of the influenza A virus replication cycle. *BMC Syst. Biol.*, **7**, 97.
14. Kanehisa,M., Goto,S., Sato,Y., Furumichi,M. and Tanabe,M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.*, **40**, D109–D114.
15. Croft,D., Mundo,A.F., Haw,R., Milacic,M., Weiser,J., Wu,G., Caudy,M., Garapati,P., Gillespie,M. and Kamdar,M.R. (2014) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **42**, D472–D477.
16. Mi,H., Muruganujan,A. and Thomas,P.D. (2013) PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.*, **41**, D377–D386.
17. Matsuoka,Y., Funahashi,A., Ghosh,S. and Kitano,H. (2014) Modeling and simulation using CellDesigner. *Methods Mol. Biol.*, **1164**, 121–145.
18. Matsuoka,Y., Ghosh,S., Kikuchi,N. and Kitano,H. (2010) Payao: a community platform for SBML pathway model curation. *Bioinformatics*, **26**, 1381–1383.
19. Hsin,K.-Y., Ghosh,S. and Kitano,H. (2013) Combining machine learning systems and multiple docking simulation packages to improve docking prediction reliability for network pharmacology. *PLoS One*, **8**, e83922.
20. Plewczynski,D., Łaźniewski,M., Augustyniak,R. and Ginalski,K. (2011) Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.*, **32**, 742–755.
21. Consortium,U. (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.*, **41**, D43–D47.
22. Benson,D.A., Cavanaugh,M., Clark,K., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2013) GenBank. *Nucleic Acids Res.*, **41**, D36–D42.
23. Bento,A.P., Gaulton,A., Hersey,A., Bellis,L.J., Chambers,J., Davies,M., Krüger,F.A., Light,Y., Mak,L. and McGlinchey,S. (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.*, **42**, D1083–D1090.
24. Hornbeck,P.V., Zhang,B., Murray,B., Kornhauser,J.M., Latham,V. and Skrzypek,E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.*, **43**, D512–D520.
25. Wang,Y., Xiao,J., Suzek,T.O., Zhang,J., Wang,J. and Bryant,S.H. (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.*, **37**, W623–W633.
26. Law,V., Knox,C., Djoumbou,Y., Jewison,T., Guo,A.C., Liu,Y., Maciejewski,A., Arndt,D., Wilson,M. and Neveu,V. (2014)

DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.*, **42**, D1091–D1097.
27. Liu,T., Lin,Y., Wen,X., Jorissen,R.N. and Gilson,M.K. (2007) BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.*, **35**, D198–D201.
28. Trott,O. and Olson,A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, **31**, 455–461.
29. Teramoto,R. and Fukunishi,H. (2008) Consensus scoring with feature selection for structure-based virtual screening. *J. Chem. Inf. Model.*, **48**, 288–295.
30. Houston,D.R. and Walkinshaw,M.D. (2013) Consensus docking: improving the reliability of docking in a virtual screening context. *J. Chem. Inf. Model.*, **53**, 384–390.
31. Liu,Z., Li,Y., Han,L., Li,J., Liu,J., Zhao,Z., Nie,W., Liu,Y. and Wang,R. (2014) PDB-wide collection of binding data: current status of the PDBbind database. *Bioinformatics*, **31**, 405–412.
32. Davis,M.I., Hunt,J.P., Herrgard,S., Ciceri,P., Wodicka,L.M., Pallares,G., Hocker,M., Treiber,D.K. and Zarrinkar,P.P. (2011) Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.*, **29**, 1046–1051.
33. Hanley,J.A. and McNeil,B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.
34. Sing,T., Sander,O., Beerenwinkel,N. and Lengauer,T. (2005) ROCR: visualizing classifier performance in R. *Bioinformatics*, **21**, 3940–3941.
35. Wallace,A.C., Laskowski,R.A. and Thornton,J.M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.*, **8**, 127–134.
36. Watanabe,T., Kawakami,E., Shoemaker,J.E., Lopes,T.J., Matsuoka,Y., Tomita,Y., Kozuka-Hata,H., Gorai,T., Kuwahara,T. and Takeda,E. (2014) Influenza virus-host interactome screen as a platform for antiviral drug development. *Cell Host Microbe*, **16**, 795–805.