

Synergy Maps: Exploring Compound Combinations Using Network Visualization

Richard Lewis¹, Rajarshi Guha², Tamás Korcsmaros^{3,4} and Andreas Bender^{1†}

¹Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

²National Center for Advancing Translational Sciences, 9800 Medical Center Drive, Rockville, MD 20850, United States

³TGAC, The Genome Analysis Centre, Norwich Research Park, Norwich, UK

⁴Gut Health and Food Safety Programme, Institute of Food Research, Norwich Research Park, Norwich, United Kingdom

rl403@cam.ac.uk, tamas.korcsmaros@tgac.ac.uk, guhar@mail.nih.gov,
ab454@cam.ac.uk

Corresponding Authors: † to whom correspondence may be addressed

Abstract

High throughput screening of compound combinations has recently received a great deal of attention. Large compound libraries and the feasibility of all-pairs screening can easily generate large, information-rich datasets. To aid in the analysis of these datasets, a new visualization technique for pairwise combination screening data, termed “Synergy Maps”, is presented. In a Synergy Map, network information about the synergistic interactions of compounds is integrated with information about their properties to produce a single visualization. The result is a succinct summary of a dataset; the relationships between compound and combination properties may be investigated simultaneously, and thus may afford insight into the types and behaviors of synergy observed in the screen. An interactive web app implementation, available at richlewis42.github.io/synergy-maps, has been developed, which may find use in navigating and filtering larger scale combination datasets. This tool is applied to a recent all-pairs dataset of anti-malarials, tested against *Plasmodium falciparum*, and a preliminary analysis is given as an example, illustrating findings previously described in literature, as well as suggesting new hypotheses for future investigation.

Keywords

Compound Combinations; Mixtures; Synergy; Visualization; Network; Dimensionality Reduction;

Background

Compound combinations have recently received much interest, as they afford a number of advantages as therapeutics compared to single agent treatments across a wide range of disease areas[1–4]. The phenomenon of super-additivity of the therapeutic effect of a combination, known as *synergy*, has the potential for improved pharmaceutical treatment options in terms of increased efficacy[5] and therapeutically relevant selectivity[6], whilst reducing the risk of toxicity[7] and side-effects[8]. Two recent reviews are available on the topic[9, 10]. However, how to determine which compound combinations exhibit a desired form of synergy in a particular case is by no means clear, and the effect of multiple bioactive compounds in parallel is overall rather poorly understood.

Synergy in a combination is due to not purely additive interaction between the biological functions of the component compounds. Progress has been made in attempts to model synergy, usually by attempting to discover these interactions. For example, models incorporating flux balance analysis (FBA) have been used to correctly predict synergistic interactions in *S. cerevisiae*[11]. Enrichment analysis of molecular and pharmacological properties predicted several combinations to be synergistic, 69% of which were subsequently verified in the literature[12]. Various network approaches (such as the Stochastic Block Model[13] and the Prism algorithm [13, 14]) have been used to infer novel interactions from large incomplete drug interaction databases such as DrugBank[15, 16]. Biological network topologies of drug targets that lead to synergy have been identified through network modelling[17], and mechanisms of action of many known non-additive drug combinations have been deduced[18]. However, these models usually require heavily annotated data (such as with ATC codes, protein targets or side effect data) - a complete understanding of the origins and repercussions of synergy has not yet in general been achieved, and thus significant further work is needed, both experimental and *in silico*.

To this end, an experimental strategy for measuring synergy has been assaying all pairwise combinations for a relatively small compound library. A recently published example of this type of dataset is the DREAM Drug Sensitivity Challenge (subchallenge 2)[19], in which all combinations of 14 compounds were tested on the LY3 lymphoma cell line. The degree of synergy for each combination was indicated by the difference in growth inhibition observed by experiment from that predicted under the Bliss Independence model[20]. Other all-pairs combinatorial datasets include a 90 compound set (consisting of drugs and probes) assayed against the HCT116 colon cancer cell line[11], a set of 11 anticancer drugs tested also against HCT116[21], a set 31 antifungal compounds assayed against *Saccharomyces cerevisiae*[22, 23], and an assay of 22 antibiotics against *Escherichia coli*[14]. Each of these datasets measure dose response surfaces[5], and derive synergy metrics from those surfaces (see original papers for examples). Whilst this is currently a reasonable selection in terms of dataset size, compound variety and assay type, there is potential for many more experiments - an exciting prospect is an upcoming National Cancer Institute

Combination Screen of approximately 100 anti cancer drugs tested pairwise against the 59 NCI-60 cell lines[24].

Visualizing large numbers of combinations

The influx of this kind of combination data provides a new opportunity to analysts. Conventionally, a first step in a data focused study is an exploratory data analysis, principally focusing on informative visualization of any data collected with the goal of identifying major trends[25]. This can be challenging, due to structure of combination data, and the geometric scaling of possible combinations with respect to compound library size[26]. Two major approaches have been utilized to visualize combination data in the literature: *heatmaps* and *networks*. Heatmaps (see *Figure 1*) are featured extensively in the literature[11, 13, 14, 19, 21, 23, 27]. Compounds of the dataset are represented as rows and columns, with their corresponding combinations positioned at the intersecting elements. A color map[11, 13, 14, 19] or gradient may be used to indicate direction and/or degree of non-additivity for each combination. The compounds may be ordered according to a particular physicochemical property, grouped by targeted protein[11] or pathway[27], hierarchically clustered according to synergy profile[23] or just alphabetically[19].

Heatmaps are useful as an uncluttered static presentation of data. It is possible to identify disproportionately synergistic compounds and also compounds that behave similarly if clustering such as in Cokol et al.[23] and *Figure 1* is applied. Additionally, relevant dose-response matrices may be superimposed[11, 21, 23] to reveal different shapes of response surfaces, which may encode information of underlying biological network topology[11, 28, 29]. A drawback is that little information about the actual compounds are encoded - they may be ordered according to a physicochemical property, but this limits further possible insight into the dataset. Furthermore, for a large dataset (for example over a hundred compounds), such as those produced using high-throughput techniques[24], the heatmap quickly becomes cluttered and individual compounds become difficult to identify.

Network representation (see *Figure 2*) for all pairs combination data is also popular[3, 13, 14, 22, 23, 30] - nodes correspond to compounds, and edges to combinations, connecting their components. Edges may be coloured according to sign, and weighted according to degree of synergy. A graph layout algorithm, such as circular[31] or force-directed[32] is usually employed to position nodes. This type of representation has a tendency to become overcrowded, and threshold values may be required to limit the number of edges. Despite this, networks have the potential to scale better with dataset size than heatmaps as compounds are positioned in two dimensions rather than along a single one. A notable shortcoming (shared with heatmaps) is that the nature of the compounds in the dataset is not simultaneously well represented: it is only possible to show a few properties, through node color, size or superimposing numbers. An example of

this may be found in a recent publication[22] where the cLogP of compounds were superimposed over the relevant node, and ordered in a circle to illustrate the increased potential of lipophilic compounds to participate in synergy. Whilst this may offer insight for the specific publication, it seems unlikely that a single property will satisfactorily explain synergistic behavior for all datasets.

Hence, an improvement in chemical property representation for the visualization of compound combination screens is still very much desirable, which is the objective of the current work.

Chemical Property Visualization

Compounds have traditionally been represented under a descriptor space using a dimensionality reduction algorithm as a scatter plot; a common example is Principle Component Analysis (PCA)[33] applied to physicochemical descriptors. A state-of-the-art equivalent might be the use of Student's t-distributed Stochastic Neighbour Embedding (t-SNE)[34] on proprietary descriptors[35]. In this way, compounds may be easily compared according to their properties or features; adjacent compounds tend to share properties and behaviour in the descriptor space in question.

In this communication, we introduce a novel type of visualization for combination datasets, named “Synergy Maps”. Synergy Maps combine network and descriptor space representation to yield an information dense presentation of a combination dataset. Specifically, the approach positions the nodes of a drug-drug interaction graph in two-dimensional space using the techniques referred to in the previous section; in this way, synergistic interactions can be straightforwardly related to trends in compound properties, and thus hypotheses for the origins of the synergy might be more quickly proposed. We also introduce an interactive implementation, which enables the generation of synergy maps for novel combination datasets, and allows for exploration of synergy under different spaces, metrics and datasets. Source code is provided as a GitHub repository.

As an example, we produce synergy maps for a combination dataset of 56 antimalarials tested against *P. falciparum*, and detail a quick analysis of the resultant maps.

Implementation

The application was constructed according to the client-server paradigm: data processing, including the descriptor calculation, and subsequent dimensionality reduction, is performed in *Python* (the server process), then transferred via *JavaScript Object Notation* (JSON) to the client visualization, implemented in *JavaScript* (see *Figure 3* for details). The program can be run on any computer with *Python 2.7* and an HTML5 capable browser (tested on the latest *Internet Explorer*, *Safari* and *Google Chrome*).

Data Processing Step

An input dataset should consist of compound data in the form of a Structure-Data File (SDF), and data associated with their combinations (including calculated synergy metrics) in the form of a Comma Separated Values (CSV) file (examples provided with the repository). A script is then written (or a default one used), specifying the descriptors, dimensionality reduction techniques and synergy metrics to employ in generating the processed file (example scripts provided with the repository).

A previously collected all pairs combination dataset of 56 compounds tested against *P. falciparum*[36] was selected as an example dataset to concretely illustrate the technique. Each combination was tested in a 6x6 dose-response matrix, varying the concentration of each compound on each axis. The change in growth inhibition was measured at each dose combination, yielding a response surface. From this, 9 different synergy metrics[37] were evaluated for all 1540 combinations. These were then preprocessed into the appropriate input format.

Compounds were initially standardized using Chemaxon Standardizer[38], to ensure a consistent representation of compounds. Descriptors for each compound were calculated for physicochemical, structural and biological spaces, each of which may be of relevance to synergy (*Table 1*). Firstly, all available physicochemical descriptors were calculated using *PaDEL*[39]. Secondly, Morgan fingerprints of radius 2, and folded to 2048 bits were generated as structural descriptors using *RDKit*[40]. Finally, 1080 Naive Bayes binary models, trained using ChEMBL[41] bioactivities, were used to predict likely (human) protein targets for each structure (notably, the organism of interest is not human for the example, but these descriptors act as reasonable generic biological descriptors[42]).

The dimensionality of each space was then reduced to two dimensions using three different, yet complementary techniques (*Table 2*). Principal Component Analysis (PCA) and MultiDimensional Scaling (MDS) were run using default parameters in *scikit-learn*[43], and student's t-distributed Stochastic Neighbor Embedding (t-SNE), was employed using a perplexity of 40. This yielded nine sets of coordinates per compound.

Due to the relatively small chemical space spanned by the 56 compounds, an additional 175 diverse compounds from MIPE[37] were temporarily added to the dataset, to diversify the space covered, and so allow for a better and more consistent dimensionality reduction step. This may not be necessary for a larger and more diverse compound set, but in practice made the resultant plots more reproducible and transferable (this was especially the case for t-SNE, which has a non-convex objective function, and thus converges to different solutions each time it is run. It also allowed for a higher perplexity (roughly the expected density of neighbors) to be set, which prevents artificially large gaps opening in the dataset).

The combinations were filtered for quality: firstly through the Quality Control

score (removing those with a score of above 4) of the data producer[36], then by removing extreme values (top and bottom 2.5% of values sorted by Gamma) on a case by case basis, by checking whether their surfaces appeared unlikely to be genuine (for an example, see *Figure 4*). The synergy metrics provided were then standardized, such that an increase in synergy was represented by an increase in magnitude, and a negative sign used for antagonism for those metrics for which it was defined (*Table 3*). The processed data was then outputted as a JSON formatted file.

The output

The visualization stage uses the *Data Driven Documents* (D3) JavaScript library to produce a *Scalable Vector Graphics* (SVG) image of the network, positioning nodes according to the coordinates precalculated in the previous step. By default, blue and red edges represent synergistic and antagonistic combinations respectively, and edge width represents the extent of the interaction. Node area is used to represent the activity of the compound individually. Synergy cut-off values may be set using a slider to declutter the visualization of the many essentially additive combinations.

The resultant networks generated for the example are shown in *Figure 5*, and an annotated version of t-SNE applied to the Bayes Affinity fingerprints with pGamma (negative log of the Gamma metric from Cokol et al.[23]) synergy values is shown in *Figure 6*. This representation may allow for the most interesting observations to be made: compounds that are predicted to modulate similar protein targets, and thus potentially share similar modes of action, are clustered together; if similar interactions are observed consistently between clusters, the underlying modes of action of each cluster might be hypothesized to interact as the cause of the synergy.

The static networks provide an insight into the relationship between compound properties and synergy, but the use of *JavaScript* enabled interactivity affords more involved exploration of the data. The author's implementation may be accessed at RichLewis42.github.io/Synergy-Maps, and source code at github.com/RichLewis42/Synergy-Maps. A screen shot of the software is shown in *Figure 7*. Controls allow for the synergy metric, descriptor space or dimensionality reduction technique to be changed dynamically. This feature may be used to gain a feel of the relatedness of the different synergy metrics selected, or the different spaces and reduction algorithms. A filter controlling the minimum synergy and antagonism required for display is provided to avoid overcrowding of the visualization. Tooltips provide additional information for compounds and combinations, originally supplied as extra fields in the original files; for the dataset used in this paper, an example is "hypothesized mode of action" for compounds.

Results and Discussion

Whilst the purpose of this paper is simply to introduce a novel visualization technique rather than analyze the resulting networks, it is possible to illustrate a few observations that may be made; these could be investigated further in subsequent assays. Firstly, we can see that compounds annotated as Histone DeAcetylase (HDAC) inhibitors, which are clustered in the north-east of the *Figure 6*, appear to be the most likely compounds in the dataset to be synergistic, and specifically with the compounds in the center (these are annotated with diverse modes of action, but often were kinase or phosphatase inhibitors). This property has been reported in the literature, where the HDAC inhibitor trichostatin A was found to interact synergistically with geldanamycin, an Hsp90 inhibitor[44], in *P. falciparum*. Interestingly, NVP-AUY922, an Hsp90 inhibitor included in the dataset, clustered to the centre; this is likely where geldanamycin would also be placed due to their similar annotated modes of action. This result would be in agreement with the observed trend and suggest that the method might yield some predictive power for unknown combinations. In contrast to this, PI3K inhibitors are shown to exhibit in general disproportionately more antagonism with the other compounds in the dataset. Whilst these observations are by no means reliable by themselves, they may form a basis for further study, and provide an example in how this type of visualization may prove a useful first step in the analysis of pairwise combination data.

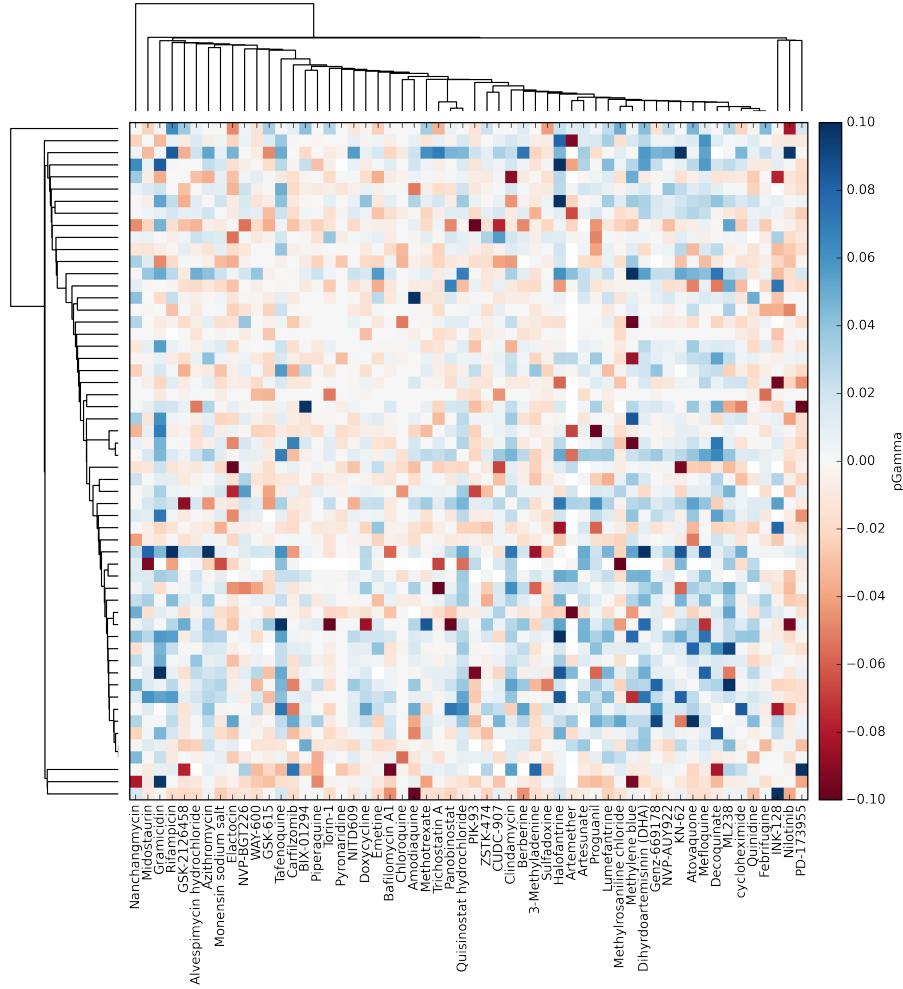
Comparisons

In the author's opinion, these observations are much less clear in the heatmap or network visualization of the data, illustrating the strength of synergy maps. However there are some problems that arise, principally in 'over fitting' an interpretation - trends may appear at random, and as such 'control' visualizations should be consulted, to provide a reality check. These can be done by scrambling compound or combination data, or using random feature representations to generate compound coordinates, as shown in *Figure 8*. Observed trends should certainly be treated with healthy skepticism, although it is likely that with the growth of high quality datasets, these chance correlations will lessen and more may be gained from the approach.

Conclusion

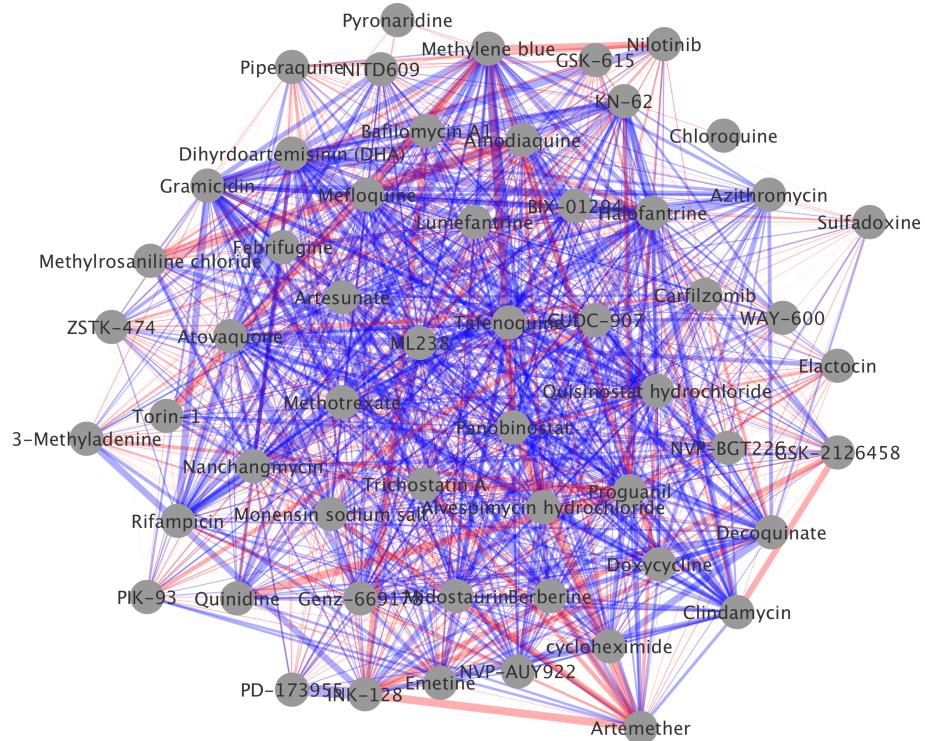
Synergy Maps, a novel method for visualization of a combination data set was presented, integrating combination-based information in a network, with compound-based information using a dimensionality reduced scatter-plot. An accompanying interactive visualization tool was also introduced, which enables fast and simple exploration and presentation of combination data. An all-pairs combination dataset assayed against *P. falciparum* was analyzed as an example, identifying several properties already reported in the literature.

Figure 1



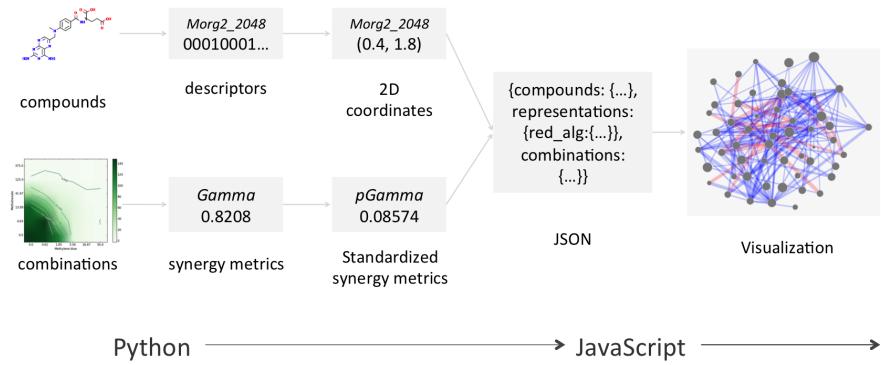
Heatmap representation of the NCATS Malaria Dataset *The heatmap, created using the Python visualization package matplotlib[45], is constructed as a matrix; rows and columns map to individual columns of the dataset, and the intersecting elements to their combination. The heatmap used the *pGamma* metric described in Table 3. Compounds were clustered according to their predicted targets, using predictions from an inhouse target prediction tool, such that compounds with a similar bioactivity profile, such as Artesunate and Artemether, cluster together.*

Figure 2



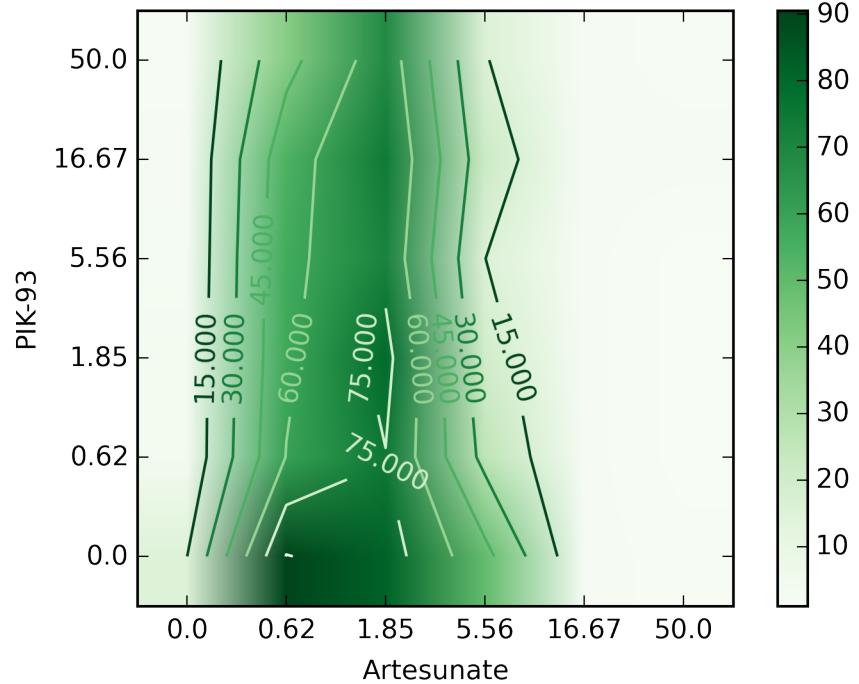
Network Representation of the NCATS Malaria Dataset This network visualization was created using Cytoscape[46]. Nodes represent compounds, whilst edges represent combinations, with thickness indicating degree of non-additivity, and red and blue indicating antagonism and synergy respectively. The layout was generated using Cytoscape’s “organic” layout routine.

Figure 3



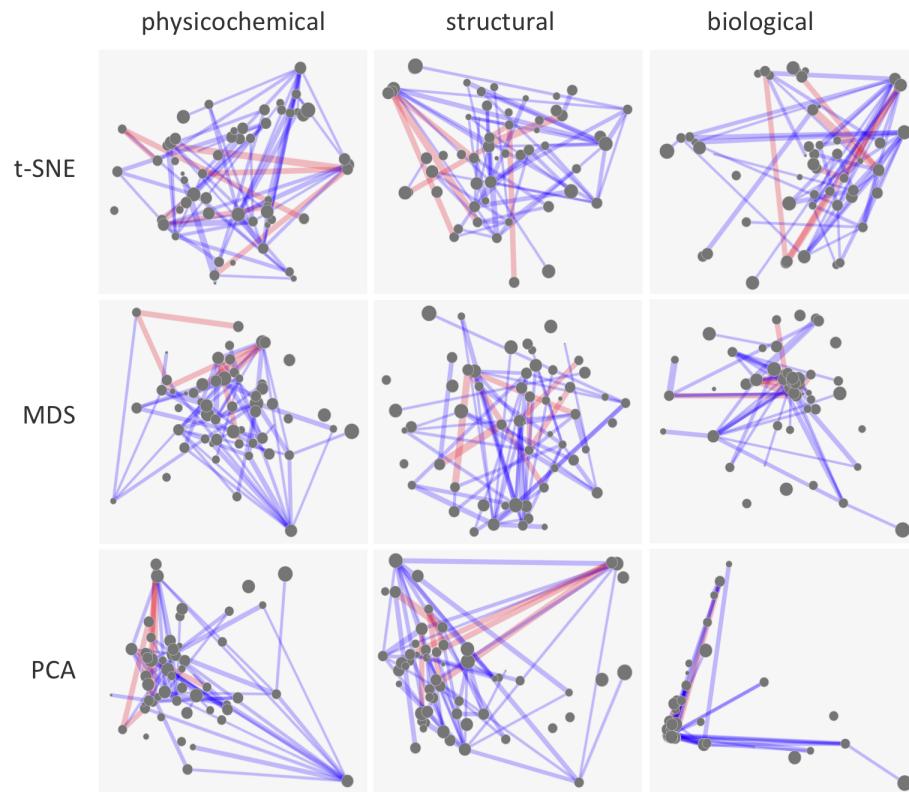
Synergy Maps Work Flow *The work flow employed by the Synergy Maps application. The raw compound and combination data is transformed in steps to yield processed data in JSON, which is then used by the JavaScript App to create the visualization. Specifically, the chosen descriptors (table 1) are generated from the supplied chemical graphs, and then reduced to two dimensions by the selected dimensionality reduction techniques (table 2). The combination data is assigned synergy values. The processed data is packaged into a JSON file.*

Figure 4



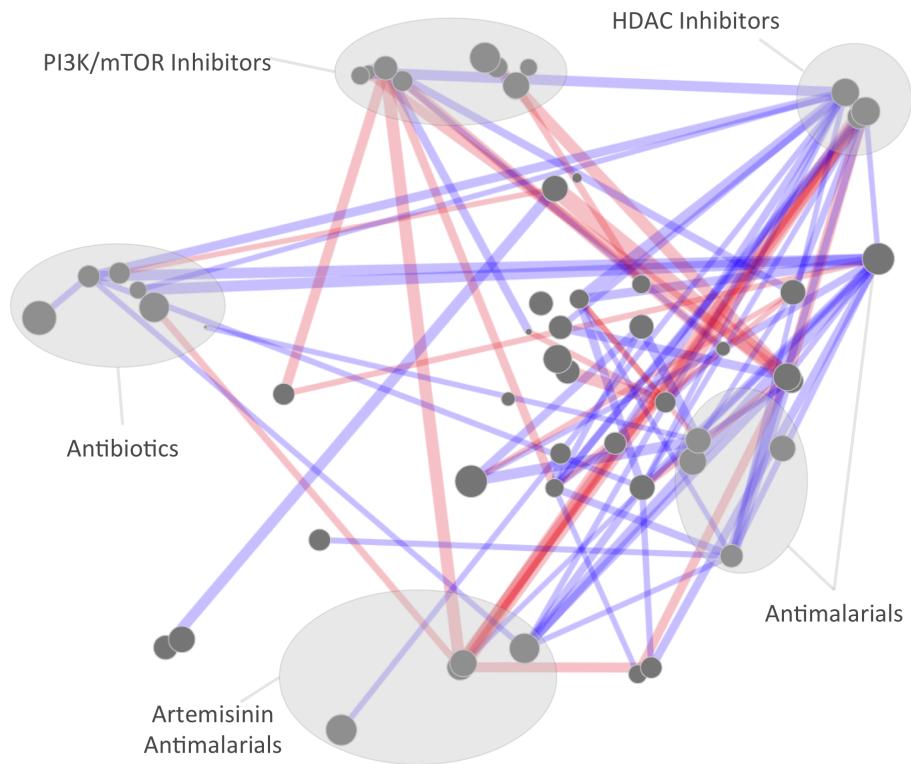
Improbable Combination Surface. The surface yields a suspiciously strongly antagonistic (-0.7) value of $p\Gamma$. The surface implies that the growth of *P. falciparum* is rescued by a low concentration of Artemeter, a known antimalarial. In fact, it seems much more likely that the zero concentration row has simply been contaminated, causing an incorrect value of $p\Gamma$.

Figure 5



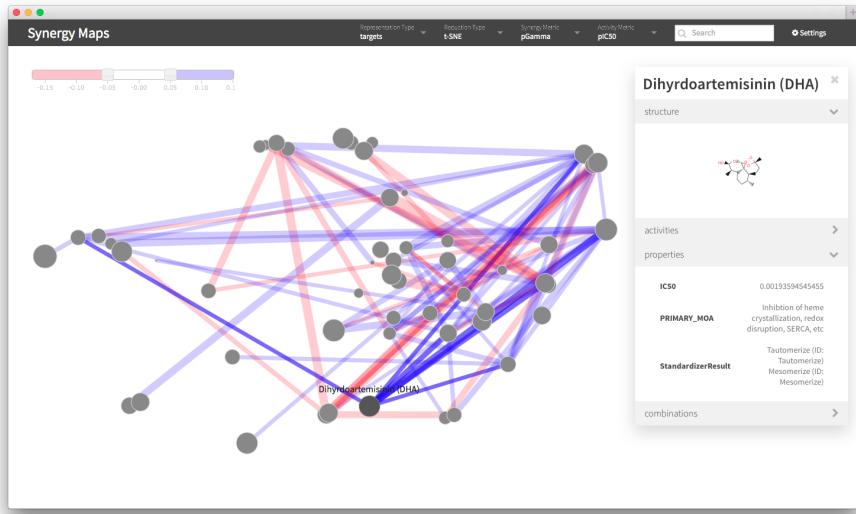
Synergy Maps Sample static networks. Nodes represent compounds, with radius indicating relative $pIC50$. Edges represent combinations, with thickness indicating degree of non-additivity, and red and blue indicating antagonism and synergy respectively. It appears that whilst PCA is a passable dimensionality reduction algorithm for physicochemical and structural space (despite concentrating points in the centre), it does not differentiate the compounds well in biological space. MDS does a little better, yet ultimately still concentrates points towards the centre, preventing compounds from being easily being differentiated. In the author's opinion, t-SNE performs well in all spaces; clear clusters can be seen, identifying groups of compounds similar in that space, yet points are still spread across space helpfully so as not to clutter the visualization.

Figure 6



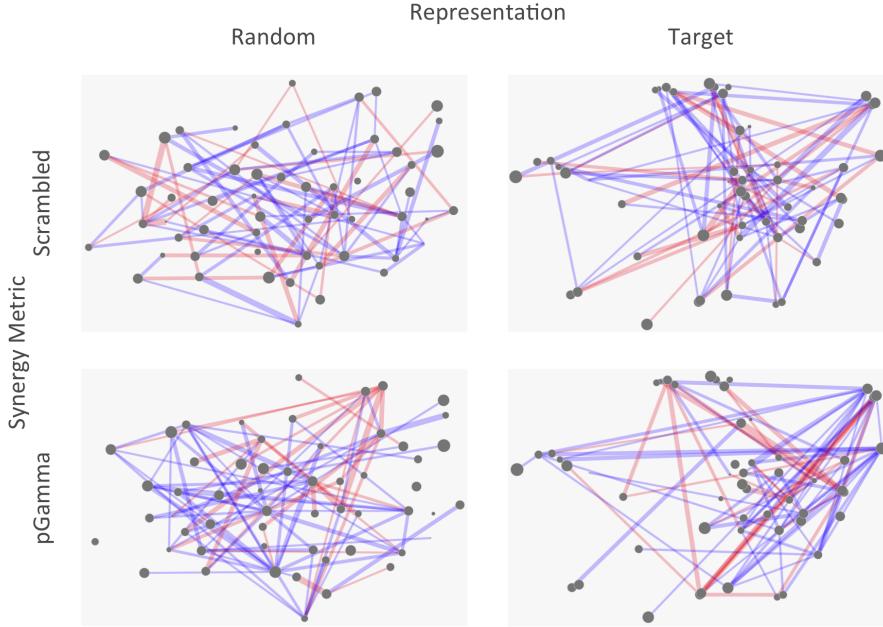
Synergy Map, represented under t-SNE reduced Biological Space Biological space, with various clusters annotated according to hypothesized mode of action or drug function. It appears that the HDAC inhibitor cluster (Including Quisinostat, Trichostatin A and Panobinostat) tends to be disproportionately synergistic compared to other clusters, whilst the PI3K/mTOR inhibitors exhibit disproportional antagonism.

Figure 7



Screen Shot of the Interactive Web Visualization *A screen shot of the interactive web visualization. The representation, reduction type, synergy metric and activity metric may be set by drop down menus in the top bar. Compounds may be searched for using the search box. A slider, shown in the top left, may be used to select threshold levels above and below which combinations should be shown. Individual compounds and combinations may be selected to bring up a tooltip, as shown for Dihydroartemisinin in this example. The tooltip will display any extra property information supplied, such as the primary mode of action in this example. Additional metadata specific to whether a compound or a combination has been selected is also given, such as the activities for a compound, and the synergies for a combination.*

Figure 8



Validation through Randomization and Scrambling *In order to assess the meaningfulness of potential hypotheses drawn from a synergy map, it is advantageous to compare with random data, such as those in the figure, to protect against spurious correlations being interpreted for more than they are. The random compound positions (leftmost maps) are generated using random feature vectors, which were then reduced using t-SNE in an identical fashion to the Bayes Affinity vectors (rightmost maps). The random positioning of compounds appears not to produce the clusters observed when real data is used. Randomly shuffling the combinations values (topmost maps) reveals more realistic maps - there are several clusters which appear to share many synergies, for which hypotheses may have been proposed, illustrating the danger of overinterpretation of synergy maps.*

Table 1

Descriptors calculated for the compounds which were used for later visualization in Synergy Maps.

Represented Space	Descriptor type	Implementation
Physicochemical	771 physicochemical descriptors	PaDEL[39]
Structural	2048 bit folded Morgan fingerprints[47] of radius 2	RDKit[40]
Biological	1080 bayes affinity fingerprint[42]	in house Naive Bayesian models

Represented Space	Descriptor type	Implementation

Three diverse spaces were selected for representation of compounds, to give maximum insight into the properties of synergy in these different spaces. Physicochemical space are likely to differentiate compounds according to their ADMET properties; Structural space will differentiate compounds according to the structure of their chemical graph; finally Biological space attempts to differentiate compounds according to their relative affinity for protein targets. Hence, with this selection of descriptors our software is able to highlight structure in multiple facets of a dataset.

Table 2

Dimensionality Reduction Techniques implemented in Synergy Maps.

Technique	Implementation
Principal Components Analysis (PCA)[33]	scikit-learn[43]
Multidimensional Scaling (MDS)	scikit-learn[43]
Student's t-distributed Stochastic Neighbour Embedding (t-SNE)	according to original publication [34]

Three differing dimensionality reduction techniques were employed; These methods provide a means to interpret the approximate structure of data in extremely high dimensional space (such as physicochemical space) on a two dimensional page. PCA locates a lower dimensional hyperplane of highest variance in a hyperspace, and projects the data onto the hyperplane. MDS attempts to preserve distances in high dimensional space with those lower dimensional space. Student's t-distributed Stochastic Neighbour Embedding also employs distance based scaling, yet imposes statistical distributions on these; it has been asserted[34] that it outperforms other methods for locating structure in high dimensional data, whilst avoiding overcrowding the centre of the low dimensional space with data points.

Table 3

Synergy metrics Calculated for the NCATS Malaria dataset.

Synergy Metric	Derivation
pGamma	negative based-10 logarithm of Gamma, from Cokol et al.[23]
Median Excess	The median of the sum of the differences between the combination responses and the single

Synergy Metric	Derivation
Num Excess	The number of combinations in a block that show a better combination responses than both
-ExcessHSA	negative of ExcessHSA, from Lehár et al.[11]
-ExcessCRX	negative of ExcessCRX, from Lehár et al. [11]
LS3x3	The minimum value of the sum of the deviations from the HSA model are evaluated on all
pBeta	negative based-10 logarithm of Beta, from Cokol et al.[23]

The synergy metrics for the dataset. These were calculated for a previous study, however a couple of alterations were required to get a consistent behaviour, specifically positive and negative values relating to synergy and antagonism respectively: pBeta and pGamma were derived from Beta and Gamma[23] by taking the negative based-10 logarithm. The original metrics specify synergy below 1, and antagonism above, so this transformation handily yields the desired mapping of antagonism to negative values. -ExcessHSA and -ExcessCRX are derived from ExcessHSA and ExcessCRX[11] by taking the negative in order to attribute positive values to synergistic interactions.

Competing Interests

The authors declare no competing financial interests.

Authors' Contribution

RG produced the data, and carried out the preliminary analysis. RL conceived of and designed the software, carried out the data analysis and drafted the manuscript. AB, TK and RG helped draft the paper. All authors read and approved the final manuscript.

Authors' information

RL is a second year PhD student under the supervision of AB, at the Centre for Molecular Informatics of the Department of Chemistry, University of Cambridge.

RG is {{...}}

TK is a Computational Biology Fellow at TGAC, The Genome Analysis Centre and a Research Leader at the Institute of Food Research, Norwich.

AB is a Lecturer in Molecular Informatics at the Centre for Molecular Informatics of the Department of Chemistry, University of Cambridge.

Acknowledgements

Azedine Zoufir, Yasaman Kalandar Motamedi, Dan Mason and Krishna Bulusu are thanked for their advice and work in the area, and the rest of the Bender Group for helpful feedback on the layout and design of the software. [Thank the malaria dataset ‘commissioners’?] EPSRC and ERC are thanked for funding.

References

1. Yuan S, Wang F, Wang J, Huang P, Chen G, Zhang H, Feng L, Wang L, Colman H, Keating MJ, Li X, Xu R-H: **Effective Elimination of Cancer Stem Cells By a Novel Drug Combination Strategy.** *STEM CELLS* 2012, **31**:23–34.
2. Hill J a, Nislow C, Ammar R, Torti D, Cowen LE: **Genetic and Genomic Architecture of the Evolution of Resistance to Antifungal Drug Combinations.** *PLoS Genetics* 2013, **9**:e1003390.
3. Tan X, Hu L, Luquette LJ, Gao G, Liu Y, Qu H, Xi R, Lu ZJ, Park PJ, Elledge SJ: **Systematic identification of synergistic drug pairs targeting HIV.** *Nature Biotechnology* 2012, **30**:1125–1130.
4. Katouli A a, Komarova NL: **Optimizing Combination Therapies with Existing and Future CML Drugs.** *PLoS ONE* 2010, **5**:e12300.
5. Berenbaum MC: **What is synergy?** *Pharmacological Reviews* 1989, **41**:93–141.
6. Krueger AS, Avery W, Heilbut AM, Johansen LM, Price ER, Rickles RJ, Short GF, Staunton JE, Jin X, Lee MS, Zimmermann GR, Borisy A a, Lehár J: **Synergistic drug combinations tend to improve therapeutically relevant selectivity.** *Nature Biotechnology* 2009, **27**:659–666.
7. Greco WR, Bravo G, Parsons JC: **The search for synergy: A critical review from a response surface perspective.** *Pharmacological Reviews* 1995, **47**:331–385.
8. Wang Y-Y, Xu K-J, Song J, Zhao X-M: **Exploring drug combinations in genetic interaction network.** *BMC Bioinformatics* 2012, **13**(Suppl 7):S7.
9. Ryall RA, Tan AC: **Systems biology approaches for advancing the discovery of effective drug combinations.** 2015, **7**:7.
10. Bulusu KC, Guha R, Mason DJ, Lewis RPI, Muratov EN, Motamedi YK, Cokol M, Bender A: **Modelling of Compound Combination Effects and Applications to Efficacy and Toxicity: State-of-the-Art, Challenges and Perspectives.** *Drug Discovery Today*.
11. Lehár J, Zimmermann GR, Krueger AS, Molnar RA, Ledell JT, Heilbut AM, Short III GF, Giusti LC, Nolan GP, Magid OA, Lee MS, Borisy AA, Stockwell

- BR, Keith CT: **Chemical combination effects predict connectivity in biological systems.** *Molecular Systems Biology* 2007, **3**.
12. Zhao X-M, Iskar M, Zeller G, Kuhn M, Van Noort V, Bork P: **Prediction of Drug Combinations by Integrating Molecular and Pharmacological Data.** *PLoS Computational Biology* 2011, **7**:e1002323.
13. Guimerà R, Sales-Pardo M: **A network inference method for large-scale unsupervised identification of novel drug-drug interactions.** *PLoS Computational Biology* 2013, **9**:e1003374.
14. Yeh P, Tschumi AI, Kishony R: **Functional classification of drugs by properties of their pairwise interactions.** *Nature Genetics* 2006, **38**:489–494.
15. Guo AC, Knox C, Wishart DS, Pon A, Law V, Banco K, Jewison T, Mak C, Neveu V, Liu P, Djoumbou Y, Ly S, Frolkis A, Eisner R, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS: **DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs.** *Nucleic Acids Research* 2010, **39**(Database):D1035–D1041.
16. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M: **DrugBank: a knowledgebase for drugs, drug actions and drug targets.** *Nucleic Acids Research* 2007, **36**(Database):D901–D906.
17. Yin N, Ma W, Pei J, Ouyang Q, Tang C, Lai L: **Synergistic and antagonistic drug combinations depend on network topology.** *PLoS ONE* 2014, **9**:e93960.
18. Jia J, Zhu F, Ma X, Cao Z, Cao ZW, Li Y, Li YX, Chen YZ: **Mechanisms of drug combinations: interaction and network perspectives.** *Nature Reviews Drug Discovery* 2009, **8**:111–128.
19. Bansal M, Yang J, Karan C, Menden MP, Costello JC, Tang H, Xiao G, Li Y, Allen J, Zhong R, Chen B, Kim M, Wang T, Heiser LM, Realubit R, Mattioli M, Alvarez MJ, Shen Y, NCI DREAM Community, Gallahan D, Singer D, Saez-Rodriguez J, Xie Y, Stolovitzky G, Califano A: **A community computational challenge to predict the activity of pairs of compounds.** *Nature Biotechnology* 2014.
20. Bliss CI: **The Toxicity of Poisons Applied Jointly.** *Annals of applied biology* 1939, **26**:585–615.
21. Severyn B, Liehr RA, Wolicki A, Nguyen KH, Hudak EM, Ferrer M, Caldwell JS, Hermes JD, Li J, Tudor M: **Parsimonious discovery of synergistic drug combinations.** *ACS Chemical Biology* 2011, **6**:1391–1398.
22. Yilancioglu K, Weinstein ZB, Meydan C, Akhmetov A, Toprak I, Durmaz A, Iossifov I, Kazan H, Roth FP, Cokol M: **Target-independent prediction of drug synergies using only drug lipophilicity.** *Journal of Chemical Information and Modeling* 2014:140715122131000.

23. Cokol M, Chua HN, Tasan M, Mutlu B, Weinstein ZB, Suzuki Y, Nergiz ME, Costanzo M, Baryshnikova A, Giaever G, Nislow C, Myers CL, Andrews BJ, Boone C, Roth FP: **Systematic exploration of synergistic drug pairs.** *Molecular Systems Biology* 2011, **7**:544.
24. Holbeck S, Collins JM, Doroshow JH: **27 NCI-60 Combination Screening Matrix of Approved Anticancer Drugs.** In *European journal of cancer*; 2012:11.
25. Tukey JW: *Exploratory Data Analysis*. Addison-Wesley; 1977.
26. Tornero-Velez R, Egeghy PP, Cohen Hubal EA: **Biogeographical Analysis of Chemical Co-Occurrence Data to Identify Priorities for Mixtures Research.** *Risk Analysis* 2011, **32**:224–236.
27. Zimmermann GR, Keith CT, Lehár J: **Multi-target therapeutics: when the whole is greater than the sum of the parts.** *Drug Discovery Today* 2007, **12**:34–42.
28. Yeh P, Kishony R: **Networks from drug-drug surfaces.** *Molecular Systems Biology* 2007, **3**.
29. Stockwell BR, Giaever G, Nislow C, Lehár J: **Combination chemical genetics.** *Nature Chemical Biology* 2008, **4**:674–681.
30. Yeh PJ, Hegreness MJ, Aiden AP, Kishony R: **Drug interactions and the evolution of antibiotic resistance.** *Nature reviews Microbiology* 2009, **7**:460–466.
31. Baur M, Brandes U: **Crossing Reduction in Circular Layouts.** In *Graph-theoretic concepts in computer science*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005:332–343.
32. Fruchterman TMJ, Reingold EM: **Graph drawing by force-directed placement.** *Software: Practice and Experience* 1991, **21**:1129–1164.
33. Pearson K: **On lines and planes of closest fit to systems of points in space.** *Philosophical Magazine Series 6* 1901, **2**:559–572.
34. Maaten L van der, Hinton G: **Visualizing Data using t-SNE.** *jmlrc-sailmitedu* 2008, **9**.
35. *Visualization Prospect - Merck Molecular Activity Challenge.* 2012.
36. Guha R: **Malaria Paper.**
37. Griner LAM, Guha R, Shinn P, Young RM, Keller JM, Liu D, Goldlust IS, Yasgar A, McKnight C, Boxer MB, Duveau DY, Jiang J-K, Michael S, Mierzwa T, Huang W, Walsh MJ, Mott BT, Patel P, Leister W, Maloney DJ, Leclair CA, Rai G, Jadhav A, Peysier BD, Austin CP, Martin SE, Simeonov A, Ferrer M, Staudt LM, Thomas CJ: **High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-celllike diffuse**

large B-cell lymphoma cells. *Proceedings of the National Academy of Sciences* 2014;201311846.

38. ChemAxon: **Standardizer**. 2013.
39. Yap CW: **PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints.** *Journal of computational chemistry* 2011, **32**:1466–1474.
40. *RDKit: Open-source cheminformatics.* <http://www.rdkit.org>; 2014.
41. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP: **ChEMBL: a large-scale bioactivity database for drug discovery.** *Nucleic Acids Research* 2011, **40**:D1100–D1107.
42. Bender A, Jenkins JL, Glick M, Zhan D, Nettles JH, Davies JW: “**Bayes affinity fingerprints” Improve retrieval rates in virtual screening and define orthogonal bioactivity space: When are multitarget drugs a feasible concept?** *Journal of Chemical Information and Modeling* 2006, **46**:2445–2456.
43. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E: **Scikit-learn: Machine Learning in Python.** *Journal of Machine Learning Research* 2011, **12**:2825–2830.
44. Pallavi R, Roy N, Nageshan RK, Talukdar P, Pavithra SR, Reddy R, Venketesh S, Kumar R, Gupta AK, Singh RK, Yadav SC, Tatu U: **Heat shock protein 90 as a drug target against protozoan infections: biochemical characterization of HSP90 from Plasmodium falciparum and Trypanosoma evansi and evaluation of its inhibitor as a candidate drug.** *The Journal of biological chemistry* 2010, **285**:37964–37975.
45. Hunter JD: **Matplotlib: A 2D graphics environment.** *Computing in Science and Engineering* 2007, **9**:99–104.
46. Christmas R, Avila-Campillo I, Bolouri H, Schwikowski B, Anderson M, Kelley R, Landys N, Workman C, Ideker T, Cerami E, Sheridan R, Bader GD, Sander C: **Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.** *AACR Education Book* 2005, **2005**:12.
47. Morgan HL: **The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service.** *Journal of Chemical Documentation* 1965, **5**:107–113.