Bioinformatics Advance Access published May 15, 2013

BIOINFORMATICS APPLICATIONS NOTE 2013, pages 1–3
doi:10.1093/bioinformatics/btt195

Systems biology

Advance Access publication April 26, 2013

OCSANA: optimal combinations of interventions from network analysis

Paola Vera-Licona[1,2,3,*], Eric Bonnet[1,2,3], Emmanuel Barillot[1,2,3] and Andrei Zinovyev[1,2,3]

[1]Institut Curie, Paris F-75248, France, [2]INSERM, U900, Paris F-75248, France and [3]Mines Paris Tech, Fontainebleau F-77300, France

Associate Editor: Martin Bishop

ABSTRACT

Targeted therapies interfering with specifically one protein activity are promising strategies in the treatment of diseases like cancer. However, accumulated empirical experience has shown that targeting multiple proteins in signaling networks involved in the disease is often necessary. Thus, one important problem in biomedical research is the design and prioritization of optimal combinations of interventions to repress a pathological behavior, while minimizing side-effects. OCSANA (optimal combinations of interventions from network analysis) is a new software designed to identify and prioritize optimal and minimal combinations of interventions to disrupt the paths between source nodes and target nodes. When specified by the user, OCSANA seeks to additionally minimize the side effects that a combination of interventions can cause on specified off-target nodes. With the crucial ability to cope with very large networks, OCSANA includes an exact solution and a novel selective enumeration approach for the combinatorial interventions' problem.

Availability: The latest version of OCSANA, implemented as a plugin for Cytoscape and distributed under LGPL license, is available together with source code at http://bioinfo.curie.fr/projects/ocsana.

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Contact: paola.vera-licona@curie.fr

Received on September 7, 2012; revised on April 18, 2013; accepted on April 19, 2013

1 INTRODUCTION

Cellular functions and activities are governed by complex signaling and regulatory networks. Diseases arise from abnormal behavior in these networks. Thus, the design of targeted therapies from a systems biology approach aims to identify which molecules to intervene in these networks, to repress a pathological behavior while minimizing side effects. Accumulated empirical experience has shown that combination or multi-component interventions are necessary to cope with the redundancy and multi-functionality that characterize biological networks (Fitzgerald *et al.*, 2006). Redundancy requires for several pathways to be targeted, as alternate routes can compensate the disrupted pathways' function. Multi-functionality implies that intervening molecules that play a central role in the cell may cause side effects, requiring alternative points of intervention (Samaga *et al.*, 2010). Some methods have been proposed to address some aspects of this problem (Hädicke and Klamt, 2011; Haus *et al.*, 2008; Klamt *et al.*, 2006). However, limited scalability of the methods and the lack of a prioritization criterion are hindering factors for their applicability to large biological networks. We introduce OCSANA, a software for the identification and prioritization of optimal minimal combinations of interventions (CIs). We define a CI as a set of nodes such that each elementary path (a path from source to target node) contains at least one node from this set. This CI set indicates the nodes to be intervened to disrupt all the identified elementary paths. The interventions can be knock outs (deletion of genes/proteins) and knock ins (overexpressions of genes/proteins). A CI is minimal if no proper subset of the CI is a CI itself, and its optimality is defined in terms of a heuristic scoring (see Section 2). To ensure the method's scalability, OCSANA includes an EXACT SOLUTION via an adaptation of Berge's algorithm (Berge, 1989) and a novel SELECTIVE ENUMERATION approach based on a weighted-greedy algorithm. The EXACT SOLUTION computes all minimal CIs of all sizes, and, similar to (Hädicke and Klamt, 2011), it is adapted to compute all CIs up to a specified size. The SELECTIVE ENUMERATION computes optimal minimal CIs up to a specified size, and it can be parametrized to identify such CIs sets by FULL ENUMERATION.

2 METHODS

OCSANA is implemented as a plugin to the open source network analysis and visualization software, Cytoscape (Shannon *et al.*, 2003). It uses the Java library BiNoM (Zinovyev *et al.*, 2008), to facilitate the import and analysis of networks. Based on the network's structure, OCSANA incorporates a scoring with three purposes: to evaluate the optimality of nodes to become part of a CI, to efficiently compute optimal solutions in the SELECTIVE ENUMERATION and to prioritize the identified minimal CIs. The scoring of a node is based on (i) the lengths of the paths from the node of interest to the targets, (ii) the type of effect on target nodes (e.g. activation/inhibition effect), (iii) side effects with respect to off-target nodes, (iv) the number of elementary paths in which the node participates and (v) the number of targets that such node can reach simultaneously. In the Supplementary File Supp1_AlgDescription, we provide a description of the scoring and the algorithms underlying the software.

*To whom correspondence should be addressed.

Algorithm 1. Algorithm outline to compute CIs

Input/mandatory: A network (directed signed graph), a set of source nodes, target nodes and a set of parameters.
Input/optional: A set of complementary nodes assigned as off-target nodes (i.e. side effects).
Output: Prioritized list of optimal CIs.
1. Pre-processing step: Compute the collection of elementary paths, that is, paths from source nodes to target nodes according to the selected parameters for the path analysis.
2. Score the nodes present in the elementary paths and sort them in a descending order.
3. Compute the so-called minimal hitting sets (MHSs) for the elementary paths according to the selected algorithm approach and sort them according to OCSANA's score. This sorted list of MHSs is the sought list of prioritized optimal CIs.

2.1 Software features

Listed below are some of the characteristics of OCSANA's software:

(1) Acquisition of the different tools from Cytoscape and BiNoM: (**a**) A graphical user interface to launch and edit networks, (**b**) Upload of networks in BioPAX, SBGN and SBML formats and (**c**) Different path analyses available for the pre-processing step.

(2) The user can select between exact solution and selective enumeration approaches.

(3) Optional analysis on side effects with respect to off-target nodes.

(4) Selection of maximum size of CIs to be identified.

(5) Visualization of results in Cytoscape.

3 PERFORMANCE TESTING AND APPLICATION

Testing performance. We used three biological networks of increasing size. The interaction graph for Epidermal Growth Factor Receptor (EGFR) signaling introduced by (Samaga *et al.*, 2009), an Epidermal Growth Factor Family (ErbB) family signaling network involved in breast cancer and a human epidermal growth factor receptor 2 positive breast cancer network (HER2+ BCN). All the details on the computations and results are included in the Supplementary File Supp2_PerformanceGraphs. We tested OCSANA's time performance to compute all CIs up to size 5 (and 6 as well for the HER2+ BCN) by its Exact Solution (Berges's algorithm) and compared it with its Selective Enumeration considering either Full Enumeration (exhaustive search) or by reducing the search space to the most optimal candidate CIs (optimal solutions). All the details on the computations and results are included in the Supplementary File Supp2_PerformanceGraphs. In Figure 1, we present a summary of the results across the three different path analyses available in OCSANA:

EGFR network. Across the different algorithms and path analyses, the time performance is identical.

ERbB family network. For shortest and optimal and suboptimal path analyses, Berge's algorithm is slower than Full

		Path Analysis Selected			
		Shortest Paths	Optimal & Suboptimal Shortest Paths	All Non Self-intersecting Paths	Graphs on the % of CIs Up to Size 5 identified by Selective Enumeration/ Optimal Solutions
(A1)	EGFR Network (144 Nodes, 266 Edges)				
(A2)	[No. Elementary Paths, No. Elementary Nodes]	[125, 49]	[234, 55]	[11050, 63]	
(A3)	Time to find all CIs up to Size 5 by Exact Sol. (Berge's alg.)	< 1s.	< 1s.	< 1s.	N/A (The search space is too small to require selective enumeration)
(A4)	Time to find all CIs up to Size 5 by Full Enumeration/Exhaustive Search	< 1s.	< 1s.	< 1s.	
(A5)	Time to find all of CIs up to Size 5 by Selective Enumeration/Optimal Sols.	< 1s.	< 1s.	< 1s.	
(B1)	ErbB Family Breast Cancer Network (336 Nodes, 492 Edges)				
(B2)	[No. Elementary Paths, No. Elementary Nodes]	[158, 77]	[473, 103]	[23668, 160] (All paths up to length 20)	
(B3)	Time to find all CIs up to Size 5 by Exact Sol. (Berge's alg.)	78 s.	66 s.	965 s.	
(B4)	Time to find all CIs up to Size 5 by Full Enumeration/Exhaustive Search	6 s.	25 s.	3401 s.	
(B5)	[Time to find all of CIs up to Size 5 by Selective Enumeration /Optimal Sols., % of search space needed]	[<1 s., 9.47%]	[<1 s., 2.29%]	[380 s., 11.19%]	
(C1)	HER2+ Breast Cancer Network (2753 Nodes, 3812 Edges)				
(C2)	[No. Elementary Paths, No. Elementary Nodes]	[534, 121]	[2538, 170]	[69805, 317] (All paths up to length 20)	
(C3)	Time to find all CIs up to Size 5 by Exact Sol. (Berge's alg.)	741 s.	133 s.	696 s.	
(C4)	Time to find all CIs up to Size 5 by Full Enumeration/Exhaustive Search	59 s.	753 s.	N/A (> 43000 s.)	
(C5)	[Time to find all of CIs up to Size 5 by Selective Enumeration /Optimal Sols., % of search space needed]	[<1 s., 1.07%]	[1 s., 0.57%]	[72 s., 0.76%]	
(C6)	Time to find all CIs up to Size 5 by Exact Sol. (Berge's alg.)	36610 s.	9965 s.	N/A (> 43000 s.)	
(C7)	Time to find all CIs up to Size 5 by Full Enumeration/Exhaustive Search	1274 s.	21786 s.	N/A (> 43000 s.)	
(C8)	[Time to find all of CIs up to Size 6 by Selective Enumeration /Optimal Sols., % of search space needed]	[484 s., 36.40%]	[1633 s., 7.36%]	N/A (> 43000 s.) 95 s. to compute 40% of CIs	

Fig. 1. Comparison of OCSANA's performance on three different biological networks under different choices of path analyses. Rows A1, B1 and C1 show the characteristics of each network. Rows A2, B2 and C2 show the number of elementary paths and nodes according to the selected source nodes and target nodes in each network. On rows A3, B3 and C3, we present OCSANA's time performance using Berges algorithm to compute all CIs up to size 5. Rows A4, B4 and C4 show the time performance to compute with the Full Enumeration all CIs up to size 5; the full enumeration is done considering the Selective Enumeration parametrized to perform an exhaustive search. On rows A5, B5 and C5, we show the time performance to compute all CIs up to size 5 by Selective Enumeration, reducing the search space to consider only optimal candidate CIs. Finally, on rows C5, C6 and C7, it is shown OCSANA's time performance to compute all CIs up to size 6 by Berge's algorithm, Full Enumeration and Selective Enumeration. All computation were made on a desktop computer with a quad-core Intel® Xeon® CPU X5472 3.00 Ghz processor and 12 GB of random access memory

ENUMERATION. For the all non–self-intersecting path analysis, Berge's approach outperforms the FULL ENUMERATION. Across the different path analyses, the SELECTIVE ENUMERATION outperforms Berge's algorithm. Finally, we see that for the three path analyses, at most 12% of the search space was needed in the SELECTIVE ENUMERATION, thus providing evidence that OCSANAs scoring aids to correctly identify all CIs within its optimal solutions.

HER2+ BCN. Under shortest path analysis, the EXACT SOLUTION performs poorer in comparison with FULL ENUMERATION and SELECTIVE ENUMERATION to compute all CIs up to size 5 and 6. For example, to compute all CIs up to size 6, the EXACT SOLUTION takes >10 h, whereas SELECTIVE ENUMERATION takes ~8 min to find the same set of CIs obtained by the EXACT SOLUTION. To compute CIs up to size 5 under optimal and suboptimal and all non–self-intersecting paths analyses, the EXACT SOLUTION outperforms the FULL ENUMERATION, whereas the SELECTIVE ENUMERATION outperforms the EXACT SOLUTION. We observe that none of the algorithms was able to compute all CIs up to size 6 in <12 h. In this case, a method allowing finding a significant number of CIs in a reasonable time becomes useful. Thereupon, we show that the SELECTIVE ENUMERATION finds 40% of all the CIs up to size 6 in just 95 s (the total number of CIs was obtained by letting the EXACT SOLUTION run for more than a day).

Application example. Using the cell-fate decision network by (Calzone *et al*., 2010), in the Supp3_ApplicationExample file, we present an application of OCSANA to identify therapeutic CIs. We show that our findings are complementary to the basic network theory approaches, and we validated them with recent literature.

4 FUTURE WORK

Future work will include modifications in the scoring method to incorporate biological data available (e.g. gene expression levels). We also contemplate to including a more realistic model of signal propagation and incorporating it in the heuristic scoring.

Funding: This work was supported by the ASSET EU FP7 collaborative project under grant no. 259348 from call FP7-HEALTH-2010-two-stage and by the grant 'Project Incitatif et Collaboratif Computational Systems Biology Approach for Cancer' from Institut Curie.

Conflict of Interest: none declared.

REFERENCES

Berge,C. (1989) *Hypergraphs. Combinatorics of Finite Sets*. Vol. 45. North-Holland Mathematical Library, Amsterdam, North-Holland.

Calzone,L. *et al*. (2010) Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Comput. Biol*., **6**, e1000702.

Fitzgerald,J.B. *et al*. (2006) Systems biology and combination therapy in the quest for clinical efficacy. *Nat. Chem. Biol*., **2**, 458–466.

Hädicke,O. and Klamt,S. (2011) Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metab. Eng*., **13**, 204–213.

Haus,U.U. *et al*. (2008) Computing knock-out strategies in metabolic networks. *J. Comput Biol*, **15**, 259–268.

Klamt,S. *et al*. (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, **7**, 56.

Samaga,R. *et al*. (2009) The Logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS Comput. Biol*., **5**, e1000438.

Samaga,R. *et al*. (2010) Computing combinatorial intervention strategies and failure modes in signaling networks. *J. Comput. Biol*., **17**, 39–53.

Shannon,P. *et al*. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*., **13**, 2498–2504.

Zinovyev,A. *et al*. (2008) BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks. *Bioinformatics*, **24**, 876–877.