

Systems biology

A logical model of HIV-1 interactions with the T-cell activation signalling pathway

Oyebode J. Oyeyemi, Oluwafemi Davies, David L. Robertson*, Jean-Marc Schwartz*

Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK

*To whom correspondence should be addressed.

Associate Editor: Igor Jurisica

Received on July 1, 2014; revised on October 24, 2014; accepted on November 20, 2014

Abstract

Motivation: Human immunodeficiency virus type 1 (HIV-1) hijacks host cellular processes to replicate within its host. Through interactions with host proteins, it perturbs and interrupts signaling pathways that alter key cellular functions. Although networks of viral–host interactions have been relatively well characterized, the dynamics of the perturbation process is poorly understood. Dynamic models of infection have the potential to provide insights into the HIV-1 host interaction.

Results: We employed a logical signal flow network to model the dynamic interactions between HIV-1 proteins and key human signal transduction pathways necessary for activation of CD4⁺ T lymphocytes. We integrated viral–host interaction and host signal transduction data into a dynamic logical model comprised of 137 nodes (16 HIV-1 and 121 human proteins) and 336 interactions collected from the HIV-1 Human Interaction Database. The model reproduced expected patterns of T-cell activation, co-stimulation and co-inhibition. After simulations, we identified 26 host cell factors, including MAPK1&3, Ikbb-Ikky-Ikka and PKA, which contribute to the net activation or inhibition of viral proteins. Through in silico knockouts, the model identified a further nine host cell factors, including members of the PI3K signalling pathway that are essential to viral replication. Simulation results intersected with the findings of three siRNA gene knockout studies and identified potential drug targets. Our results demonstrate how viral infection causes the cell to lose control of its signalling system. Logical Boolean modelling therefore provides a useful approach for analysing the dynamics of host–viral interactions with potential applications for drug discovery.

Contact: jean-marc.schwartz@manchester.ac.uk or david.robertson@manchester.ac.uk

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Human immunodeficiency virus type 1 (HIV-1) is the pathogenic agent responsible for acquired immune deficiency syndrome. HIV-1 cannot replicate on its own, requiring host cellular machinery to reproduce. The virus accomplishes this by interacting with its host via thousands of highly specific molecular interactions (Fu *et al.*, 2009; Ptak *et al.*, 2008). These interactions constitute a complex network that drives cellular functions to viral-favourable conditions and cause a host response to infection, altering the normal cellular functioning of the host system. The HIV-1, host cell factor interaction network is complicated with numerous direct and indirect

interactions involving viral–viral factors, viral–host cell factors and host–host cell factor interactions in the exploited host cellular systems (Pinney *et al.*, 2009). A primary goal of HIV-1 research is to obtain a detailed and coherent understanding of infection and viral replication. Critical to this goal is an understanding of viral–host interactions, the host's immune response and anti-viral mechanisms that play key roles in combating infection. This network of interactions may provide valuable insights into the development of treatment regimes and the identification of novel drug targets (Fu *et al.*, 2009; Nikolsky *et al.*, 2005; Ptak *et al.*, 2008). Extensive HIV-1, human protein interaction data have been collated in the HIV-1

Human Interaction Database (HHID). This resource gives detailed descriptions of experiments on HIV-1 host molecular interactions curated from the primary literature. Each interaction is recorded, described and linked to the National Library of Medicine PubMed identification numbers describing it—the gene Entrez ID and the NCBI Reference Sequence protein accession numbers (Fu *et al.*, 2009; Ptak *et al.*, 2008).

Computational modelling techniques have been employed to create static HIV-1 interaction networks, but these have not usually considered spatial and temporal aspects. Static models provide useful information on basic topological properties of HIV-1-host interactions (Dickerson *et al.*, 2010). For instance, van Dijk *et al.* (2010) constructed an HIV-1-human protein interaction network using HIV-1 host protein–protein interactions contained in HHID and reported that, depending on the cellular process being perturbed, viral proteins interact both directly and indirectly with HIV dependency factors (HDFs). Therefore, studying the HDF network has the potential to improve our understanding of HIV–host interactions. However, the cell is in a dynamic state. External and internal signals influence protein interactions and hence cellular functions. As a consequence of these signals, different cellular processes are halted or activated. Processes may become active only under certain conditions or in a particular cell type. Static networks alone do not account for such complexities. In a dynamic system, the effect of specific signals on its function can be determined. Construction of dynamic models of virus–host interactions will therefore improve our understanding of the viral hijack of the host cell's signal transduction network. Signalling pathway networks can explain the dynamic flow of signals between molecules in a cell. These networks not only represent the proteins and interactions involved, but they also show the direction of signal flow. An application of computer modelling techniques to understanding cellular interaction dynamics is the logical Boolean model formalism (Klamt *et al.*, 2006). Logical Boolean models are qualitative signal flow networks that provide a means to dynamically represent complex networks lacking in kinetic details and quantitative data. They reveal functional dependencies and non-linear relationships between components of the system, represent its dynamic response following from input signals and perturbations and predict end states which the system is likely to reach either as steady states or attractor cycles (Klamt *et al.*, 2007). The logical model can also be used to determine properties such as dependencies, feedback loops, robustness and sensitivity. Logical Boolean models find applications in many biological phenomena. For example, they have been used to reproduce knockout and overexpression profiles of mutant mice (Herrmann *et al.*, 2012) and to predict that the overexpression of Src results in increased endocytosis of EGFR in the absence/low amount of the epidermal growth factor in human mammary epithelial cells (Helikar *et al.*, 2013).

Of all immune cells, CD4+ T lymphocytes are the most important cell type involved in HIV-1 infection. HIV-1 enters into these cells via the interaction of its envelope proteins with the CD4 receptor and CXCR4 or CCR5 co-receptors. Besides viral entry, HIV-1's interaction with these receptors can trigger signals that activate multiple signalling pathways stimulating cellular processes within the host cell. In this study, we have computationally analysed signalling pathways essential for T-cell activation through the T-cell receptor complex (TCR), its co-stimulators and its co-inhibitors. Since chemokines, apart from inducing chemotaxis, also amplify T-cell activation signals, we have included chemokine signalling pathways. We thus present the first integrated logical model of HIV-1 interactions with T-cell activation pathways. By plotting the interdependencies between host and viral proteins, we identify host proteins

that inhibit or activate HIV-1 proteins in these pathways. By simulating loss of functions, i.e. in silico knockouts, we determine putative points for potential drug intervention. We demonstrate that insights into HIV-1 human protein interactions can be obtained by integrating HHID and other biological data into a modelling framework. This approach will be transferable to other pathways.

2 Methods

2.1 Data collection and model building

Host cell factor (HCF) interactions were obtained from 338 papers from the primary literature. These were guided by pathway data from the Kyoto Encyclopaedia of Genes and Genomes (Kanehisa *et al.*, 2004) and WikiPathways (Kelder *et al.*, 2012). Only molecular interactions in which direct involvement could be established between any two HCFs or between viral proteins and HCFs were considered as interactions in the model. We used HHID (2012 update) as the primary source to obtain viral–host protein interaction data. Each interaction between human and viral proteins was manually curated to establish that it is a direct physical interaction. Indirect interactions were excluded. The nature of each interaction and a brief description of the dynamics and combinations employed are detailed in [Supplementary Table S1](#). In presenting the interactions between two proteins, 'X' and 'Y', we considered the interaction between them to be 'activating' if it allows Y to perform its biological function and 'inhibitory' if it prevents it from performing its biological function. We have built our model with the program CellNetAnalyzer, a comprehensive user interface for analysis of cellular networks that runs in MATLAB (Klamt *et al.*, 2007), available at <http://www2.mpi-magdeburg.mpg.de/projects/cna/cna.html>. Three siRNA gene knockout studies by Brass *et al.* (2008), Zhou *et al.* (2008) and König *et al.* (2008) were accessed to identify HDFs involved in these pathways. The pathway was visualized using the System Biology Graph Tool in Certus Technologies' Electronic Lab Notes software.

2.2 Interaction graph-based study

Interaction graphs show relationships between nodes (molecules) and arcs (interactions, between molecules, also referred to as edges). For any interaction between two nodes, a sign-directed arc indicates the direction of interaction between the two nodes. Converting pathway information into an interaction graph enables the calculation of dependencies between nodes. Dependencies reveal to which extent a node in the network requires another node to be biologically active. It is the inverse of influence, which describes the extent of involvement of a source node in the biological activation or inhibition of the target node. We use an algorithm that analyses binary interactions (i.e. activating and inhibitory) to describe the influence of a molecule on another (Klamt *et al.*, 2007; Saez-Rodriguez *et al.*, 2007) based on the combination of directed paths between them (Klamt *et al.*, 2006). A path reveals a series of interactions between molecules that link them in a single direction. Paths consist of multiple arcs that connect nodes, where each arc could be positive (indicating activation) or negative (indicating inhibition). In computing the influence, both the shortest positive and negative paths (if they exist) are taken into consideration. Also, the presence of an adjoining node (a node with a negative feedback loop along the shortest path, either positive or negative) would weaken the strength of its activation or inhibition.

For example, if we consider the nodes A and D, where multiple paths link A to D, the influence of A on D is determined by the existence

of the shortest positive and negative paths (or both) linking A to D. If all paths from A to D include positive arcs only, A is said to totally activate D. In general, the influence of A on D would be as follows:

1. 'Ambivalent' if both positive and negative paths are present between the two nodes such that the net positive or negative influence cannot be easily ascertained but depends on the kinetics of the system.
2. 'Strong activation' if there exists only positive paths between the two nodes and there is no adjoining nodes with a negative feedback loop.
3. 'Weak activation' if the shortest path between the two nodes is positive and there is at least one adjoining node with a negative feedback loop.
4. 'Strong inhibition' if there exists only negative paths between the two nodes and there is no adjoining nodes with a negative feedback loop.
5. 'Weak inhibition' if the shortest path between the two nodes is negative and there is at least one adjoining node with a negative feedback loop.
6. 'No effect' if no path exists between the two nodes.

For every type of influence of A on D, D depends on A. That is, if A strongly activates D (influence), D is strongly activated (dependency) by A. The dependency matrix captures this relationship (see [Supplementary Fig. S3](#) for further details).

2.3 Understanding signal propagation in pathways

Signal propagation explains how signals are transduced from input nodes (cell receptors, etc.) through the intermediate nodes to the output node. After examining model interdependencies, we analysed our network as a logical model. A logical model consists of nodes (molecules) connected by edges (interactions) in a Logical Interaction Hypergraph (LIH). Each node is assigned a state, which defines its condition for a biological property, for example, concentration. Likewise, each edge is assigned a rule that expresses the nature of the interaction (activation/inhibition) between reactant nodes and product. Logical models use discrete values to express the state of nodes: a value 0 represents a node that is switched OFF, this means the molecule is in a biologically inactive state; a value 1 represents a node that is switched ON, the molecule is biologically active and is able to transduce signals. To represent interaction effects between proteins, we have employed a Boolean representation called the Disjunctive Normal Form ('sum or products') that uses 'AND', 'OR' and 'NOT' functions to describe the combination of arcs required to activate or inhibit a node. For example, column 3, rows 16 and 150 in [Supplementary Table S1](#) is represented as follows:

$$\neg \text{CBL} + \text{TCR (complex)} + \text{ABL} \rightarrow \text{ZAP70}$$

$$\text{ZAP70} \rightarrow \text{CRK}$$

The first expression states that TCR (complex) and ABL are required to activate ZAP70, and that CBL must be inactivated. For the second expression, ZAP70 activates CRK. Biologically, this means if we switch ON TCR and ABL but switch OFF CBL, active TCR (complex) and active ABL would activate ZAP70. ZAP70, which is now active, would activate CRK.

2.4 Logical steady-state analysis

We investigated *in vivo* signal propagation of the activation pathway in the T cell using logical steady-state analysis ([Klamt et al., 2007](#)). Logical steady states (LSSs) are situations where the state of

each node is stationary with respect to all the incoming Boolean functions. LSS analysis presents a good platform to study signal transfer and response to external stimuli in networks ([Klamt et al., 2007](#)). The algorithm determines a state for intermediate and output nodes in the model based on interaction combinations of input node(s) whose states are already given ([Klamt et al., 2007](#)). It changes the states of each node in a dynamic manner based on the LIH. At LSS, the system no longer changes dynamically, unless a new perturbation is applied ([Klamt et al., 2006](#)). Given a set of input values, the Boolean algorithm assigns a determined state to each node in the model. The state of each node is determined by the Boolean function it is associated with. If there are no feedback loops, the LSS computed is said to be complete as signals are propagated from input straight to the output. But feedback loops often lead to multi-stationarity or iterations that could prevent the existence of LSS or create a partial LSS ([Xue and Miller-Jensen, 2012](#)). Note, a partial LSS does still provide functional information on the model and negative feedback loops could give insights into points in the network whose parameterization would change its dynamic behaviour. In a partial LSS, CellNetAnalyzer represents nodes which do not reach a stationary state as 'undetermined'. We have run the logical steady-state analysis for different input conditions and node deletions.

2.5 Test for putative drug targets

To simulate the action of drugs on proteins in the pathway, we employed an *in silico* knockout technique. We set the value of a particular node to 0 to exclude it from interactions. We then determined the new LSS for this set of conditions.

3 Results

3.1 Model integration reveals the complex interdependent nature of the signal transduction pathway

We have constructed a model comprising 137 nodes (16 HIV-1 proteins and 121 human proteins) involved in 336 interactions. The model consists of three main layers: the input, intermediate and output layers. The input layer consists of nodes from which signals enter into the pathway. The intermediate layer consists of nodes that transduce signals from the input layer, and the output layer consists of nodes through which signals leave the pathway to other cellular processes. Antigens and cell membrane receptors (7 in total) that are important in T-cell activation or inhibition are the main inputs in the model. Six molecules whose activation is independent on the signalling pathway have also been set as input. Details of molecules set as inputs are given in [Supplementary Table S3](#).

The HIV-1 T-cell pathway model represents the transduction of signals from antigens which activates the TCR complex, chemokines that stimulate the chemokine receptors, co-receptor molecules (CD4 and CD45), T-cell activation co-stimulators, inducible T-cell co-stimulator (ICOS) and the CD28 molecule and co-inhibitors, cytotoxic T-lymphocyte-associated protein 4 (CTLA4) and the programmed cell death protein 1 (PD-1). It includes interactions between HCFs in the pathway and also their relations with viral proteins. Major pathways described in the model include the TCR signalling pathway, the chemokine signalling pathway, the calcium signalling pathway, the MAPK signalling pathway, the PI3K-Akt and JAK/STAT signalling pathways that relate to processes like IL-2, IL-4, IL-5, IL-10 interferon gamma and tumour necrosis factor alpha production, cytokine production, cell survival, migration, apoptosis and regulation of actin cytoskeleton.

The T-cell activation model is highly interconnected with host protein nodes linking each other in a complex hypergraph. Forty percent of the successor (products) could be activated by at least two separate interactions. PI3K can be activated as a result of six different interactions. Of the 223 T-cell activation interactions we observed, 16 interactions were inputs and therefore had no predecessors, 158 interactions required only one predecessor, 30 had two predecessors, 16 of them had three predecessors, and one and two had four and six predecessors, respectively. All interactions had only one successor (product) except 16 output interactions. This means that 22% of product nodes in the model have more than one predecessor node which combine to initiate their biological functions. Fifty-one host molecules in our results interact with at least one viral protein. This represents a significant degree of interaction (37.2%) with the host signalling pathway.

3.2 Model validation

The model was built based on local interactions described in primary literature. We have used only direct interactions in building the model. For example, if A interacts with B and B interacts with C, the interaction A–B is direct, while the interaction A–C is indirect. Some experimental groups may have observed that C is activated as a result of A being activated; however, only if this interaction is shown to occur in the absence of B should a direct link between A and C be inferred. As the validity of our model would depend on the accuracy of the host signalling pathway interactions, we have first sought to validate the host signalling pathway interactions described in our model. Therefore, we excluded all virus–host protein interactions from analysis during model validation. Logical Boolean models allow us to determine the states of proteins during signal propagation through the use of ‘ON’ and ‘OFF’ states that describe if the protein is biologically active and able to activate its downstream effectors or not (Klamt et al., 2006). This can be achieved through logical steady-state analysis. Given a set of input conditions (Klamt et al., 2006), a logical Boolean model would fix states for downstream effectors (Klamt et al., 2006) according to the Boolean value assigned to it.

We set input conditions as described in [Supplementary Table S3](#) and tested if our model would reproduce known global properties using logical steady-state analysis. A sample of 25 global interactions obtained from 23 primary literature papers was used and we reproduced 21 (~84%) of them. For example, CD28 co-stimulation is not required for c-Jun N-terminal kinase activation in T cells, CD28 co-stimulation of TCR activation increases BCL2-like 1 levels, inhibits the BCL2-antagonist of cell death (Bader and Hogue, 2003) and amplifies MAPK14 (p38) and RAC-VAV-RHOA pathways (Cronin and Penninger, 2007). The chemokine signalling pathway reportedly plays a role in T-cell activation and when it is activated along with TCRs, it amplifies the effects of TCR stimulation; our model reproduced these findings. Some other notable reproduction of literature observations include inhibition of the AKT-dependent activation of NFkB and calcium signalling dependent activation of NFAT when co-stimulator CD28 and co-inhibitor CTLA4 are both activated along with the T cell receptor (Parry et al., 2005), and the inhibition of the RASGRP1-dependent pool of IL-2 production via MEK1/ERK. Some results could not be replicated, for example, Patsoukis et al. (2012) state that PD-1 inhibits the MEK/Erk and AKT pathway; although our results indicated partial inhibition of the pathway, they did not indicate full inhibition. Nonetheless, it is a positive outcome to have 84% of predictions confirmed. The small number of unconfirmed predictions are most

likely due to the presence of feedback loops that mitigate effects in the system and thereby prevent a clear change in protein levels. Full details of our model validation results are provided in [Supplementary Table S3](#).

3.3 HIV-1 ensures activation of pathways

In the model, we first analysed the dependency of each protein node on other nodes so as to better understand the viral–host protein relationships. To understand the effect of HIV-1 on cellular function, we compared the uninfected T-cell activation pathway with the infected HIV-1 host pathway. Thus, we removed HIV-1 interactions from the model and generated the dependency matrix ([Supplementary Fig. S1](#)) for all elements in the pathway. In the dependency matrix for the uninfected pathway, each host molecule is strongly or weakly activated or inhibited by its predecessors and strongly or weakly activates or inhibits its successors. In general, host molecules either have no effect or ambivalent effects on other elements that are not their upstream or downstream effectors ([Fig. 1](#)). The most frequent relationship observed is ‘no effect’ represented by black boxes, followed by ambivalent effects in [Supplementary Fig. S1](#).

Next, we added HIV-1 interactions and repeated the procedure to obtain the dependency matrix ([Supplementary Fig. S2](#)). A comparison of the HIV-1 host pathway dependencies with the uninfected pathway dependencies reveals how viral proteins hijack the pathway. An important observation is that viral proteins could activate host proteins without the need of the predecessors of these host proteins. For example, the NCK adaptor protein (Verbanck et al., 2013) and the p21 protein (Cdc42/Rac)-activated kinase (PAK) (Zenke et al., 2004) are two proteins responsible for actin cytoskeleton regulation in T cells. PAK is normally activated in T lymphocytes by NCK. But HIV-1 Nef binds and activates PAK (rows 141 and 283 in [Supplementary Table S1](#)). In the presence of Nef, if for any reason NCK becomes inactive within the host, PAK will still be biologically active and actin skeleton reorganization would still take place. This trend is observed throughout the model. One consequence is that the host does not have full control over its immunological pathways in the presence of viral proteins. Although our model could not ascertain the extent of activation of these host molecules by viral proteins, it shows that the virus tries to ensure the activation of the T-cell pathway.

Our results indicate that viral–host interactions create multiple activation routes that lead to the same functional output. An evident observation from comparison of dependencies of the uninfected and infected pathways is the shift in dependencies of host proteins, due to interactions with viral proteins. Most factors move from strong or weak activation or inhibition effects to generally ambivalent effects. For example, in dependency results of the uninfected pathway ([Supplementary Fig. S1](#)), MAPK1&3 (Erk) is weakly activated by 11 host molecules, weakly inhibited by three host molecules, while 50 host molecules are ambivalent and 57 host molecules in the model have no effect on it. In the dependency result of the HIV-1 host pathway, however ([Supplementary Fig. S2](#)), MAPK1&3 (Erk) is weakly activated by five host molecules, weakly inhibited by one, ambivalent for 108 and unaffected by 23. More nodes have ambivalent effects and fewer nodes have no effect on the HIV-1 chemokine pathway. This shows that viral proteins function to maintain the pathways in an activated state and suppress the normal checks and controls in each pathway by weakening the dependence of the HCFs on their upstream nodes.

In the uninfected pathway dependencies, the cAMP-dependent protein kinase (PKA) family of genes is weakly inhibited by chemokine, CCR5/CXCR4, GRK, GRKL, Gai and the GaiGBY protein

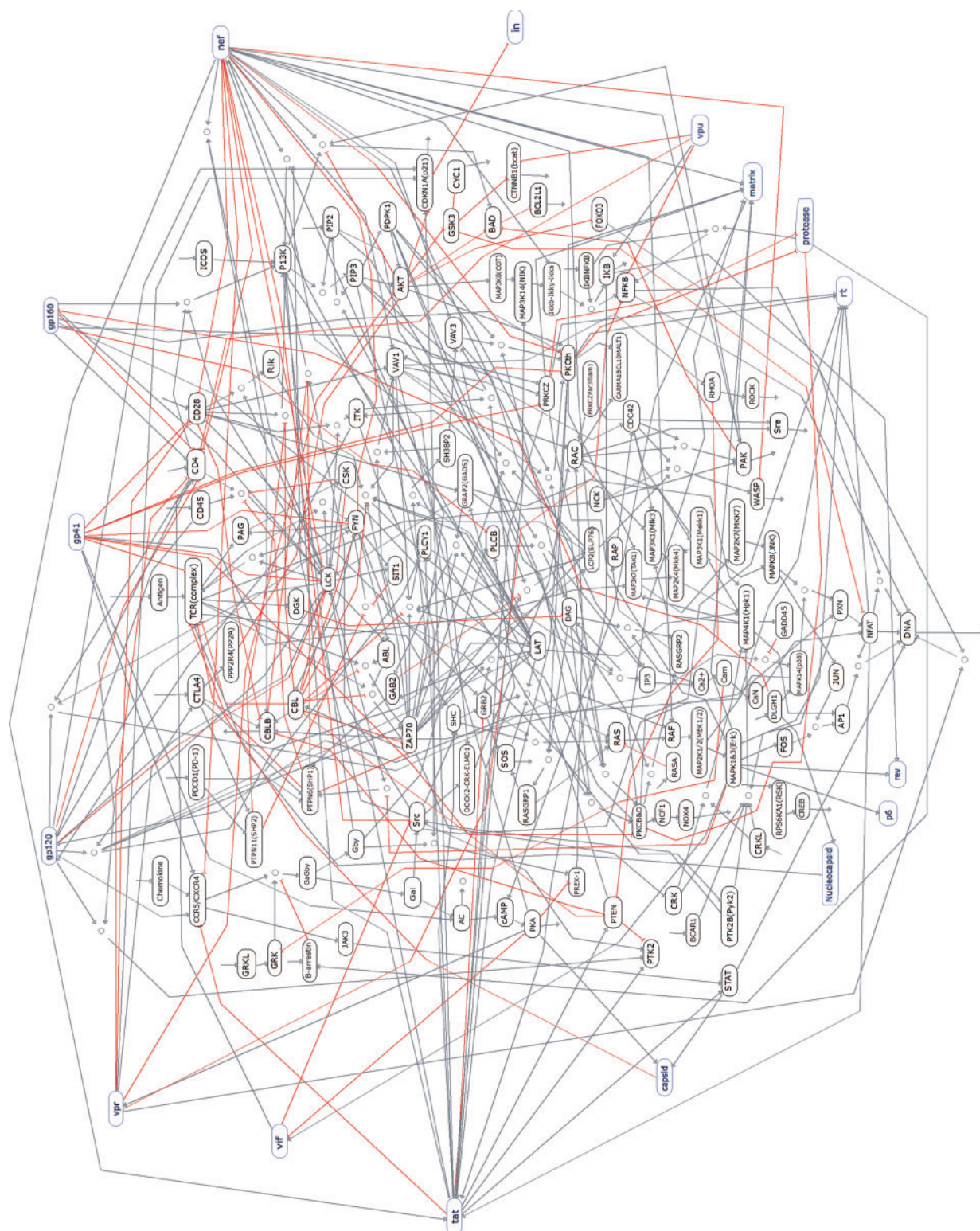


Fig. 1. Model of the T-cell activation signalling pathway in a CD4+ T cell with associated HIV-1 interactions. Molecules are shown in boxes with host proteins in grey and viral proteins in blue. Interactions between molecules are shown as edges between the molecule nodes: grey arrowed lines represent activation and red lines represent inhibition. Circles represent the AND logical operator that connects more than one reactant to the product

Table 1. Dependencies of HIV-1 proteins on HCFs in the host pathway

HIV protein	Weak activation	Total activation	Weak inhibition	Total inhibition	Ambivalent	No effect
Nef					114	23
gp160				DLGH1		136
gp120	ITK				113	23
gp41						137
Rev					114	23
Vpu						137
Vpr	PAK, NCK, AC, cAMP, PKA, Vpr				108	23
Tat	CARMA1BCL10MALT1, GSK3, IKB, IKBNFKB, ikkb-ikky-ikka, MAP3K14 (NIK), MAP3K8 (p38)		NFAT, CaN		105	23
Vif					114	23
Matrix	NCK, PAK				112	23
Capsid	STAT, JAK3				112	23
Nucleocapsid		DLGH1				136
Rt					114	23
Protease					114	23
P6	SOS, RASGRP1, RAS, RAF, MAPK1&3 (ERK), MAP2K1/2(MEK1/2)		RASA		107	23
In				CDKN1A (p21)	114	22

Note: Numbers in columns six and seven represent HCFs that have neutral (ambivalent) and no effect, respectively. Host molecules printed in blue represent dependencies coinciding with direct reactions. Molecules in red are viral proteins. A colour version of this table is available online at <http://bioinformatics.oxfordjournals.org/>.

complex (Supplementary Fig. S1). Four nodes in our model, including B-arrestin and the calcium signalling models Ca2+, Cam and IP3, weakly activate PKA. But in HIV-1 host pathway dependencies (Supplementary Fig. S2), many factors that strongly inhibited PKA in the uninfected pathway now have ambivalent effects. The viral protein matrix along with PAK, NCK, AC and cAMP weakly activate PKA. Except for AC and cAMP, the HCFs necessary for the activation or inhibition of PKA in the HIV-1 host pathway are completely different to those in the uninfected pathway. This suggests that HIV-1 induces the activation of PKA and cAMP-dependent processes which are normally inhibited in the chemokine signalling pathway.

The protein NFκB is totally inhibited by four HCFs (AC, cAMP, GADD45 and PKA), and totally activated by 26 proteins in the uninfected pathway dependencies (Supplementary Fig. S1). In the HIV-1 host pathway dependencies (Supplementary Fig. S2), however, it is neither inhibited nor activated by any protein. NFκB plays a key role in regulating the immune response to infection and anomalies in its regulation have been linked with viral infection, cancer and autoimmune diseases (Sheikh and Huang, 2003). Because NFκB is involved in DNA transcription, it has been suggested that NFκB may play a role in viral transcription from latency, this is an indicator of a loss of control of the T-cell activation system (Hiscott et al., 2001; Nimmerjahn et al., 2004; Wu et al., 1995). Another example is the protein paxillin (PXN) that is totally activated by 21 host proteins and inhibited by four host proteins in the uninfected pathway. In the HIV-1 host pathway, however, there are only five weakly activating proteins affecting PXN with about 109 ambivalent factors and 23 factors having no effect on it. Also, the regulatory host protein PTPN6 (SHP1) shows this trend. In the uninfected pathway dependency results, it is weakly inhibited by 12 proteins and weakly activated by 3 proteins. However, in the infected pathway, it is neither activated nor inhibited by any protein. The trend is constant for all proteins in the dependency results.

The dependency matrix also permits the assessment of dependencies of viral proteins via the pathway connections. Knowledge of the dependencies of viral proteins on HCFs in the pathway could offer insight on how HIV-1 depends on host proteins during replication (Klamt et al., 2006). For example, if a host protein activates a viral protein, deletion of this host protein could reduce the biological efficiency of the viral protein. Table 1 shows the dependencies of HIV-1 proteins on the pathway. HIV-1 proteins were shown to depend on a total of 26 host proteins for activation or inhibition. Out of 29 dependencies, only nine coincided with reaction expressions used to build the model. This shows that the dependency results predicted 20 new dependency states (Table 1).

3.4 Viral–host interactions activate more host proteins in the T-cell signalling network

Although resting T cells can be infected with HIV-1, viral replication takes place preferentially in active T cells (Margolick et al., 1987). Therefore in defining control conditions, we have chosen the state whereby the T cell is fully activated. A fully activated T cell requires the activation of the co-stimulator and TCR receptor. Also because the activation of chemokine receptor has been shown to contribute to T-cell activation both in our model and previously (Wong and Fish, 2003), we switched ON the chemokine input node. Because inhibition of T-cell activation is necessary to prevent autoimmunity and the ligands CD80 and CD86 are shared by CD28 and CTLA4 (Smith-Garvin et al., 2009), we have also switched ON co-inhibitors in our model. In other words, all input nodes, receptors and non-receptors required for activation of the model were switched ON, while input nodes whose activation contributes to the inactivation of the model, except for the co-inhibitors, were switched OFF. This scenario was set as our control. The full details of values assigned to each input node in our model can be found in Supplementary Table S2.

As we did for dependencies, we have considered both the uninfected T-cell activation pathway and the HIV-1 host pathway. In

Table 2. HIV-1 nodes that are switched ON/OFF by knocking out an HCF

Host cell protein	Viral protein switched ON	Viral protein switched OFF
PDPK1	Protease	
PI3K	Protease	
PIP2	Protease	
PIP3	Protease	
PKA		Vpr
PKCB&D	Protease	
cAMP		Vpr
DLGH1	gp160	Nucleocapsid
MAPK1&3 (Erk)		Rev, p6, Vif

Note: Table represents the effect of knocking out a host molecule (column 1) on switching ON (column 2) or OFF (column 3) viral proteins

calculating the LSS for the HIV-1 host pathway, we have considered the HIV-1 proteins to be in an inactive state and needing to be activated by a host molecule via the signal transduction pathway. The basis for our approach is the knowledge that viral proteins are biologically inactive outside their hosts but becomes active in them due to their interaction with host proteins. Thus, the viral nodes (proteins) in the model were not assigned values (states) but set as undetermined. Some viral proteins such as gp120, Nef and Tat interact with many HCFs in the model, whereas others interact with fewer proteins. The LSS for both pathways was computed at control conditions too (Supplementary Fig. S4). We compared LSS results of the uninfected host pathway with the HIV-1 pathway. Under control conditions in the uninfected pathway, about 16 HCFs were switched OFF in our model: VAV3, PLCy1, FOXO3, GAB2, CRK, PKA, LCP2 (SLP76), WASP, SH3BP2, cAMP, LAT, BAD, CDKNA1 (p21), GSK-3, AC and ZAP70 (full names are written in Supplementary Table S2). In the HIV-1 host pathway, four host proteins, (FOXO3, GSK3, AC and BAD), along with three HIV-1 proteins, (gp160, Integrase (in) and Protease), were switched off. This shows that more proteins in the uninfected pathway are switched off when compared with those of the HIV-1 host pathway.

3.5 Model perturbation identifies potential drug targets

The logical model provides an approach to study the effect of perturbations resulting from disruptions in the signal transduction pathway due to external stimuli or mutation (Klamt *et al.*, 2006). To understand how perturbations from mutations and drugs could potentially alter the balance of HIV-1-host interactions, we knocked out host molecules from our model. If a protein Y in a signal transduction pathway is rendered non-functional due to mutation or the effect of a drug, transfer of signals from Y would be disrupted and therefore activation or inhibition of downstream nodes could be affected. This scenario can be modelled *in silico* via 'knockout' or 'knockin' of interactions and/or nodes in the logical model. To knock out a node, we exclude the node from the model completely by assigning the state '0' to it (i.e. switched OFF), leaving all other states from the control intact. All host molecules in the model were knocked out one at a time with the other control conditions maintained and the LSS at these states were determined. We recorded viral proteins that are switched off at LSS for each protein being knocked out, and have listed these as putative drug targets (Table 2). Knocking out some host molecules switched off one or more viral proteins. For example, knocking out Erk switched off Rev, p6 and Vif. This predicts that removal of the protein

Table 3. HCFs in the uninfected T-cell activation pathway indicated as HIV-1 dependency factors in siRNA gene knockout studies

Brass <i>et al.</i>	Zhou <i>et al.</i>	Konig <i>et al.</i>
CD4	AKT	GRK
LCP2 (SLP76)	MAP2K7 (MKK7)	NFKB
NFKB	MAP3K14 (NIK)	
Ikkb-Ikky-Ikka	NFKB (Konig <i>et al.</i>)	
AKT	CD4	
PKA	CXCR4	
CXCR4	ADRBK1	
RAP		

Note: Host molecules in red indicate overlaps in the siRNA studies and dependency matrix results. A colour version of this table is available online at <http://bioinformatics.oxfordjournals.org/>.

MAPK1&3 (Erk) *in vivo* could impair the function of these three viral proteins. The removal of discs, large homolog 1 (DLGH1) protein switched off nucleocapsid but switched on the envelope glycoprotein complex gp160, suggesting that DLGH1 is an inhibitor of gp160. The gp160 node, which was off at LSS in the HIV-1 host model, was switched on by DLGH1. The PI3K pathway is heavily implicated in our results to inhibit protease as its proteins, when switched off, turn on protease, which was inhibited at control. Also, we knocked out different combinations of host molecules indicated by our dependency matrix results to have significant inhibitory or activating effects on each viral protein. For example, we knocked out both STAT and JAK3. JAK3 has been indicated in Table 1 to activate capsid. In general, the combinatory knockout of host molecules does not proffer any new conclusions. According to Drugbank data (Wishart *et al.*, 2008), five of nine HCFs identified in Table 2, namely MAPK1/3(ERK), PKA, PDPK1, PI3K and PKCB&D, are known drug targets.

3.6 siRNA comparison of results

To test the validity of our results, we compared model predictions with three high-throughput siRNA gene knockout studies from Brass *et al.* (2008), Zhou *et al.* (2008) and Konig *et al.* (2008). They identified 283, 303 and 293 HDFs respectively that are essential for viral replication in humans. siRNA gene knockout studies are a form of RNA interference experiment achieved by transfection of artificial siRNAs which are processed to suppress expression of particular genes in the cell. The cell with a specific gene inhibited is infected with HIV and viral replication monitored. Genes whose absence hampered viral replication are noted as essential for viral replication. As the associated human proteins are not lethal to the cell when absent, they represent ideal drug targets.

Although the siRNA screens were expected to identify the same targets, the experimental procedures, cell types and materials they used differed leading to limited overlap in results. For example, (Bushman *et al.*, 2009) compared the three siRNA studies used here and found only a small overlap (<7%). We have extracted host molecules from these studies that are genes in the T-cell activation pathway (Table 3). NFKB appears in all three siRNA studies, and as many as four HCFs (CD4, NFKB, AKT and CCR5/CXCR4) intersected between the Brass *et al.* (2008) and Zhou *et al.* (2008) studies. In total, 12 HDFs from these 3 studies were host molecules in our model. Three of these, PKA, ikkb-ikky-ikka and MAP3K14, were identified in our dependency matrix results (Table 1). Only PKA intersects with our logical steady-state results (Table 2). Konig

et al.'s (2008) data had no intersect with either dependency matrix or logical steady-state results. Brass *et al.*'s (2008) data had two intersects with dependency matrix results and one with logical steady-state results.

3.7 Comparing logical steady-state analysis results with dependency matrix results

There is overlap in the results from both the LSS and dependency matrix approaches. For example, the AC-cAMP-PKA pathway, which is normally inactive in the chemokine signalling pathway, activates Vpr. It is interesting to note the influence of PKA. PKA is activated via phosphorylation by cAMP in the chemokine signalling pathway and is responsible for a variety of cellular functions (Rodriguez and Kranias, 2005). In the HIV-1 host dependency matrix results, it was observed that PKA is weakly activated and its activation pattern is similar to Vpr. The HIV-1 host LSS results show that deactivation of PKA consequently deactivate sVpr, a viral protein critical for viral replication. As PKA was also identified by siRNA studies, there is evidence that supports the necessity of PKA in HIV-1 replication.

The MEK1/2-Erk pathway is also identified in both dependency and LSS results. Interestingly, MAPK1/3 (ERK) is identified as necessary for the activation of the viral proteins Vif, P6 and Rev in both tests. This protein is an integration point for many cellular signals and has functions in cell proliferation, differentiation, transcription regulation and development (Avruch *et al.*, 2001). All three viral proteins are involved in different key stages of viral replication: Vif assists in viral assembly and maturation, Rev circumvents host cell defence mechanism by countering mRNAs splicing to aid viral replication and P6 aids viral infectivity and virus core assembly. The results suggest that Erk is a very important host protein in viral replication. This result is similar to cancer cells as it was noted that uncontrolled expression of MEK1/2-Erk pathway leads to cancer (Downward, 2003). Switching off the host protein DLGH1 activated gp160 and deactivates nucleocapsid. Potentially, viral cells may control host signal transduction dynamics intermittently inhibiting or activating host proteins to favour viral replication needs at various stages of infection. In total, four HCFs (DLGH1, MAPK1&3 (Erk), cAMP and PKA) appeared in both the dependency and logical steady state results.

4 Discussion

We have used logical models as a novel approach to construct HIV-1 network dynamics, because this technique is efficient in the absence of detailed quantitative information. Pathogenic viral infections and cancer have been traced to malfunctioning signalling pathways that induce disrupted cellular responses to signals and produce a diseased state (Xue and Miller-Jensen, 2012). Using logical models, the effect of viral protein interactions on downstream cellular components can be monitored.

Our dependency results have predicted that HIV-1 interactions with proteins in these signalling pathways reduce the dependence of the host proteins on upstream HCFs in the signalling pathway for their activation or deactivation while ensuring that the pathways remain activated. We suggest that HIV-1 has evolved to disrupt the normal checks and balances that regulate each pathway, thereby favouring processes that enable its replication. The possibility of bypassing a stage in the pathway involving host cells interacting with HIV-1 has also been observed in our model, with PKA as an example. The dependency matrix provides a way to analyse the effect

of viral interactions on the host system. Our model has reproduced experimental viral-host protein interaction effects, and gives an added advantage of permitting comparison of the effect of viral proteins on all host proteins in the pathway with their consequences for other host-host protein interactions.

Our approach has also predicted global interactions not reported in the HHID (e.g. PAK and NCK activating Vpr). HIV-1 host dependency results have suggested that positive feedback loops exist between interacting pairs: Tat-GSK3 and Tat-MAP3K14 (NIK) (Nikolsky *et al.*, 2005). The feedback loop Tat-GSK3 is recorded in HHID and verifies the interaction in the database. Not only do dependency matrices predict dependencies, they also predict the influence of viral proteins on host molecules. Supplementary Table S5 shows the influence of viral proteins on host molecules. One of the host molecules in the influence table, ADRBK1, was also identified by siRNA gene knockout study by Zhou *et al.* (2008). In the combined dependency (Table 1) and influence table (Supplementary Table S5), the HIV-1 T-cell dependency made 33 global predictions; two of these 33 are found in HHID. This highlights the predictive power of logical models. Dependency matrices also allow us to exclude or include interactions in analysis based on differences due to cell type or stage of infection. With given cell type or infection stage information, dependency matrix results could be used to correct false positives and negatives that are recorded in the HHID.

Logical steady-state analysis provides a way to simulate the changes in signal transduction cascades due to mutations or external stimuli like drug actions. In our model, under controlled conditions, we discover that viral-host interactions activate proteins in the model that are normally switched OFF in signal transduction pathways. In addition to the known behaviour that viral invasion of the host cell triggers in the immune system, our results indicate that as a consequence of infection the normal checks and balances in signal transduction are bypassed, facilitating viral infection. The identification of these proteins is important as they represent fixed or immutable targets for antiviral therapeutics. Known drug targets among our LSS results are MAPK1/3 (ERK), PKA, PDPK1, PI3K and PKCB&D (Wishart *et al.*, 2008). LSS analysis therefore provides an avenue to identify potential drug targets for experimental validation and would be very useful in further *in silico* pathogenic studies. Our results suggest that viral-host interactions do contribute to the activation of more host proteins and hence aid activation of processes downstream of these proteins. It also suggests that the virus does not in any way impede these signals but instead amplifies them.

In conclusion, we have demonstrated that insights into HIV-1 human protein interactions can be obtained by integrating direct HHID interactions and other biological data into a logical modelling framework. It is yet unclear how infection dynamics will change in CD4+ T-cell signalling when other pathways are included. Therefore, HIV-1 host infection models need to be constructed on a larger scale to encompass all pathways and cell types involved in HIV-1 infection. This modelling framework could also be used to elucidate other signalling pathways to better understand signal flow information and rationally identify potential drug intervention points.

Acknowledgements

Special thanks from the authors to Chengkun Wu, Roger Ptak, William Fu, Daniel Jamieson, Caroline Paget and Mauricio Vieira Kritz for their assistance during the course of this work.

Conflict of Interest: None declared.

References

- Avruch, J. *et al.* (2001) Ras activation of the Raf kinase: tyrosine kinase recruitment of the MAP kinase cascade. *Recent Prog. Horm. Res.*, **56**, 127–155.
- Bader, G.D. and Hogue, C.W. (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, **4**, 2.
- Brass, A.L. *et al.* (2008) Identification of host proteins required for HIV infection through a functional genomic screen. *Science*, **319**, 921–926.
- Bushman, F.D. *et al.* (2009) Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog*, **5**, e1000437.
- Cronin, S.J. and Penninger, J.M. (2007) From T-cell activation signals to signaling control of anti-cancer immunity. *Immunol. Rev.*, **220**, 151–168.
- Dickerson, J.E. *et al.* (2010) The biological context of HIV-1 host interactions reveals subtle insights into a system hijack. *BMC Syst Biol*, **4**, 80.
- Downward, J. (2003) Targeting RAS signalling pathways in cancer therapy. *Nat. Rev. Cancer*, **3**, 11–22.
- Fu, W. *et al.* (2009) Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res*, **37**, D417–D422.
- Helikar, T. *et al.* (2013) A comprehensive, multi-scale dynamical model of ErbB receptor signal transduction in human mammary epithelial cells. *PLoS One*, **8**, e61757.
- Herrmann, F. *et al.* (2012) A Boolean model of the cardiac gene regulatory network determining first and second heart field identity. *PLoS One*, **7**, e46798.
- Hiscott, J. *et al.* (2001) Hostile takeovers: viral appropriation of the NF-kappaB pathway. *J. Clin. Invest.*, **107**, 143–151.
- Kanehisa, M. *et al.* (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.
- Kelder, T. *et al.* (2012) WikiPathways: building research communities on biological pathways. *Nucleic Acids Res.*, **40**, D1301–D1307.
- Klamt, S. *et al.* (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, **7**, 56.
- Klamt, S. *et al.* (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol.*, **1**, 2.
- Konig, R. *et al.* (2008) Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*, **135**, 49–60.
- Margolick, J.B. *et al.* (1987) Amplification of HTLV-III/LAV infection by antigen-induced activation of T cells and direct suppression by virus of lymphocyte blastogenic responses. *J. Immunol.*, **138**, 1719–1723.
- Nikolsky, Y. *et al.* (2005) Biological networks and analysis of experimental data in drug discovery. *Drug Discov. Today*, **10**, 653–662.
- Nimmerjahn, F. *et al.* (2004) Active NF-kappaB signalling is a prerequisite for influenza virus infection. *J. Gen. Virol.*, **85**(Pt 8), 2347–2356.
- Parry, R.V. *et al.* (2005) CTLA-4 and PD-1 receptors inhibit T-cell activation by distinct mechanisms. *Mol Cell Biol*, **25**, 9543–9553.
- Patsoukis, N. *et al.* (2012) PD-1 inhibits T cell proliferation by upregulating p27 and p15 and suppressing Cdc25A. *Cell Cycle*, **11**, 4305–4309.
- Pinney, J.W. *et al.* (2009) HIV-host interactions: a map of viral perturbation of the host system. *AIDS*, **23**, 549–554.
- Ptak, R.G. *et al.* (2008) Cataloguing the HIV type 1 human protein interaction network. *AIDS Res. Hum. Retroviruses*, **24**, 1497–1502.
- Rodriguez, P. and Kranias, E.G. (2005) Phospholamban: a key determinant of cardiac function and dysfunction. *Arch. Mal. Coeur Vaiss.*, **98**, 1239–1243.
- Saez-Rodriguez, J. *et al.* (2007) A logical model provides insights into T cell receptor signaling. *PLoS Comput. Biol.*, **3**, e163.
- Sheikh, M.S. and Huang, Y. (2003) Death receptor activation complexes: it takes two to activate TNF receptor 1. *Cell Cycle*, **2**, 550–552.
- Smith-Garvin, J.E. *et al.* (2009) T cell activation. *Annu. Rev. Immunol.*, **27**, 591–619.
- van Dijk, D. *et al.* (2010) Identifying potential survival strategies of HIV-1 through virus-host protein interaction networks. *BMC Syst. Biol.*, **4**, 96.
- Verbanck, M. *et al.* (2013) A new unsupervised gene clustering algorithm based on the integration of biological knowledge into expression data. *BMC Bioinformatics*, **14**, 42.
- Wishart, D.S. *et al.* (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, **36**, D901–D906.
- Wong, M.M. and Fish, E.N. (2003) Chemokines: attractive mediators of the immune response. *Semin Immunol*, **15**, 5–14.
- Wu, B.Y. *et al.* (1995) Regulation of human retroviral latency by the NF-kappa B/I kappa B family: inhibition of human immunodeficiency virus replication by I kappa B through a Rev-dependent mechanism. *Proc. Natl Acad. Sci. USA*, **92**, 1480–1484.
- Xue, Q. and Miller-Jensen, K. (2012) Systems biology of virus-host signaling network interactions. *BMB Rep.*, **45**, 213–220.
- Zenke, F.T. *et al.* (2004) p21-activated kinase 1 phosphorylates and regulates 14-3-3 binding to GEF-H1, a microtubule-localized Rho exchange factor. *J. Biol. Chem.*, **279**, 18392–18400.
- Zhou, H. *et al.* (2008) Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe*, **4**, 495–504.