rxncon software and visualised via Biographer (Biographer). In the same way, the *rxncon* software can be used to generate the contingency matrix, the reaction graphs, the regulatory graph, and, via BioNetGen (Blinov *et al*, 2004), the SBML file that constitute the basis for the process description. These generations are fully automated and hence the framework addresses the issue of (ii) automatic network visualisation without further assumptions and—in the case of the contingency matrix and regulatory graph—without any simplifications.

Generation of mathematical models

The contingency matrix is a template for automatic generation of mathematical models. Each elemental reaction corresponds to a basic (context-free) rule in a rule- or agent-based model (Table II), or, in other words, a set of rules that share a reaction centre (Chylek *et al*, 2011). All contextual constrains on an elemental reaction is defined in a single row in the contingency matrix, and this row defines the elemental reaction's implementation in the rule-based format. The basic rule suffices if there are no known modifiers of a particular elemental reaction (i.e., only '0' and '?' apart from the intersection with its own state(s) (which is always 'x' for a product state and '!' for a source state)). Every other contingency splits the expression in two rules; one when that elemental state is true and one when it is false. The number of rules needed only increases with the number of quantitative modifiers ('K+' and 'K−') as the qualitative modifiers sets the rate constant to zero in either the 'true' (for 'x') or false (for '!') case (see Supplementary information for details). The expansion to rules is fully defined in our data format and the *rxncon* software tool automatically generates the input file for the computational tool BioNetGen (Blinov *et al*, 2004). This file can be used for rule-based modelling, network-free simulation and creation of SBML files. The translation to and from the rule-based format is unambiguous in both directions, and we illustrate this with translation of a rule-based model of the pheromone response pathway (yeastpheromonemodel.org). This model contains lumped reactions which we translate to combinations of elemental reactions, resulting in a different equation structure but the same functionality given appropriate choice of rate constants (Supplementary Table S3). Furthermore, we cannot distinguish different identical proteins in, for example, homodimers, and can therefore not define strict *trans*-reactions within such dimers. Apart from these issues, we can reproduce the same model with only cosmetic/nomenclature differences (see Supplementary information for details). Hence, the framework addresses the issue of (iii) automatic model generation from the database of biological information.

Mapping the MAP kinase network

As a benchmark, we have used the presented framework and an extensive literature search to create a comprehensive map for the yeast MAP kinase network (Supplementary Table S1). Reactions have been defined with specific residues and domains whenever experimental support was sufficient. The degree of experimental evidence has been evaluated manually and individually for each entry, and references to primary research papers supporting each interaction have been included in the reaction and contingency lists (column 'PubMedIdentifier(s)'). We have used mechanistic data on reactions (C1) and a combination of mechanistic and genetic data on contingencies (C2) between reactions and reactants' states from primary research literature. The mapping is based solely on primary research papers and *de facto* shown data to ensure a high-quality network reconstruction. We chose to exclude almost all genetic data as indirect effects cannot be ruled out even in well-performed genetic screens. Finally, we decided not to include spatial data, as we found information especially on regulation of (re)localisation too sparse. To the best of our knowledge, we have eliminated all questionable information from the compiled data set, and convincing reactions lacking solid mechanistic evidence have been included but clearly and distinctly labelled.

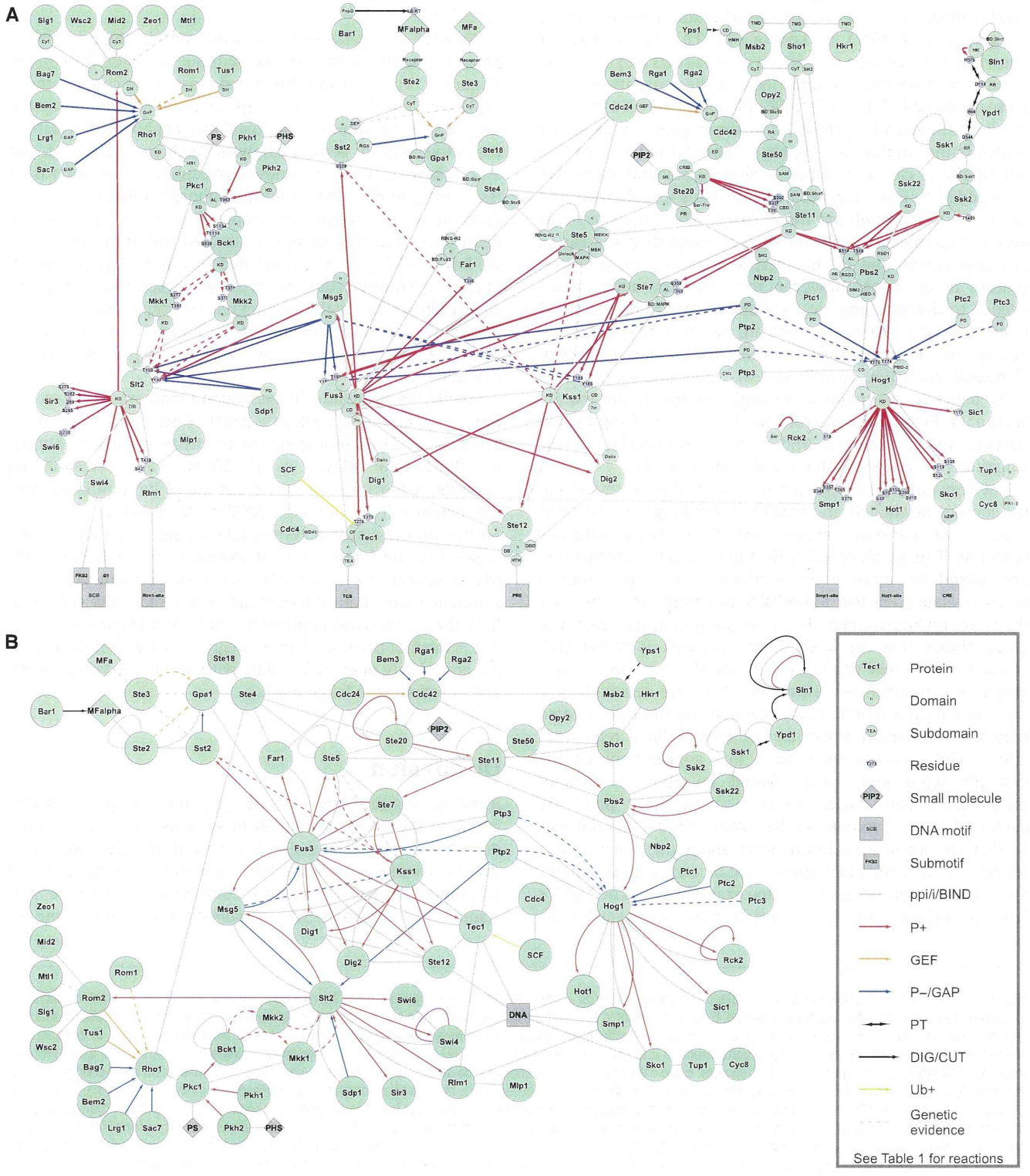
The MAP kinase network contains 84 components, 181 elementary states and 222 elementary reactions, corresponding to many hundreds of thousands of specific states. This network is large enough to be a severe challenge to the established visualisation and analysis methods. We did in fact fail to generate the complete state space and terminated the BioNetGen expansion after the first three iterations which generated 207, 1524 and 372 097 specific states, respectively. We use a range of graphical formats to visualise different aspects of this highly complex network. First, we display the network topology in the *reaction graphs* (Figure 2). These

Table II Implementation of elemental reactions in the rule-based format

Elemental reaction	BioNetGen rule implementation	
Interactions ("ppi", "i" or "bind")	A(B) + B(A) <-> A(B!1).B(A!1)	kf, kr
Intra-protein interactions (ipi)	A(A1,A2) <-> A(A1!1,A2!1)	kf
Phosphorylations (P+)	A + B(Psite~U) -> A + B(Psite~P)	kf
Autophosphorylations (AP)	A(Psite~U) -> A(Psite~P)	kf
Phosphotransfers (PT)	A(Psite~P) + B(Psite~U) <-> A(Psite~U) + B(Psite~P)	kf, kr
Dephosphorylations (P−)	A + B(Psite~P) -> A + B(Psite~U)	kf
Nucleotide exchanges (GEF)	A + B(GnP~U) -> A + B(GnP~P)	kf
Nuclease activations (GAP)	A + B(GnP~P) -> A + B(GnP~U)	kf
Ubiquitination (Ub+)	A + B(UBsite~U) -> A + B(UBsite~UB)	kf
Proteolytic cleavages (CUT)	A + B(Domain~U) -> A + B(Domain~truncated)	kf
Degradations (DEG)	A + B -> A	kf

The table displays how the different elemental reactions in Table I are translated to the rule-based format. See Supplementary information for additional details.

figures show that the number of characterised phosphorylation reactions vastly outnumbers that of characterised dephosphorylation reactions (68 to 16; Figure 2A), and that several well-established processes are only supported by genetic data (including the entire MAP kinase cascade below Pkc1; Figure 2B, dashed lines). The reaction graph also allows comparison between the established pathway architecture and the unbiased global protein – protein interaction studies and synthetic lethal networks (Figure 3A and B, respectively).



In the *contingency matrix* (Figure 4), we visualise the combined knowledge we have about the MAP kinase system (C1 and C2). The core matrix (red block of rows and blue block of columns) describe all the elemental reactions, elemental states and the (possible) contingencies of reaction on states. The black fields here show when there is no overlap between the components in the reactions and those defined in the states. Therefore, the matrix will always be sparsely populated. However, we also see that most of the remaining fields are grey; that is, effect not known ('?'). This means that our knowledge of reactions (C1; which defines rows and columns) is much stronger than our knowledge of the causality between these reactions (C2; the cells). We only have data on a minority of all possible contingencies, and these gaps are explicitly shown in the contingency matrix. It should also be noted that not all effects can be ascribed to single elemental states. We have added an outer layer of Boolean states (purple rows and columns) to account for these cases. The Boolean states describe complex mechanisms such as scaffolding and can in principle correspond to the specific states of, for example, process descriptions. However, they are only added when needed to describe empirical results. Note that only a small fraction of the states are Boolean, which reflects the low abundance of empirical data on the combinatorial effect of elemental states (i.e., specific states). Therefore, we believe it to be better to use mapping strategies which do not require such data. Finally, the matrix contains a layer of inputs and outputs (grey; columns and rows, respectively). These constitute the system's interface with the outside.

The *regulatory graph* (Figure 5) displays the information in the contingency matrix graphically, by showing how reactions produce or consume states, and how states influence reactions. This graph contains the full C1 and C2 information, and would fall apart without either. In fact, the isolated reaction–state pairs that fall outside the graph do so because they have no known incoming or outgoing contingencies. The graph shows that the MAP kinase network is rather well connected, as most reactions are indeed linked in a single graph by contingencies. However, there are relatively few input and output points; many reactions do not have known regulators and many states do not have defined regulatory effects. Only reaction–state pairs that appear between the system's input and output would be able to transmit information. This means either that all other pairs are irrelevant for the dynamics of the signal-transduction process, or that we are lacking information about their role in this process. In fact, such lose ends might be excellent candidates for targeted empirical analysis. One example would be Msb2's binding to Cdc42, which is reported to be important for the pseudohyphal differentiation pathway; raising the question of whether this binding is regulated in response to the stimuli that activate this part of the MAP kinase network. Another point that stands out is the almost complete lack of (documented) information exchange between pathways. The exception is the Sho branch of the Hog pathway, which is closely intertwined with the mating pathway, as both are activated by the shared MAP kinase kinase kinase Ste11 and parts of the cell polarity machinery.

We have also generated a network map in the established *process description* format, but without complex formations (Figure 6). This decision eliminated most of the combinatorial explosion and the need for implicit assumptions. However, there is still uncertainty in the specific phosphorylation state of the active state of certain catalysts, such as Ssk2, Ste11 and Ste7. Likewise, we do not know if phosphorylation order is an issue for proteins with multiple phosphorylation sites. In contrast to the regulatory graph (Figure 5), the process description becomes more complicated the more unknowns we have and Figure 6 is simplified (compare Supplementary Figure S2). However, the limited process description in Figure 6 clearly shows the catalyst–target relationships, and reinforces the impression that very few of the known phosphorylation reactions are balanced by known dephosphorylation reactions.

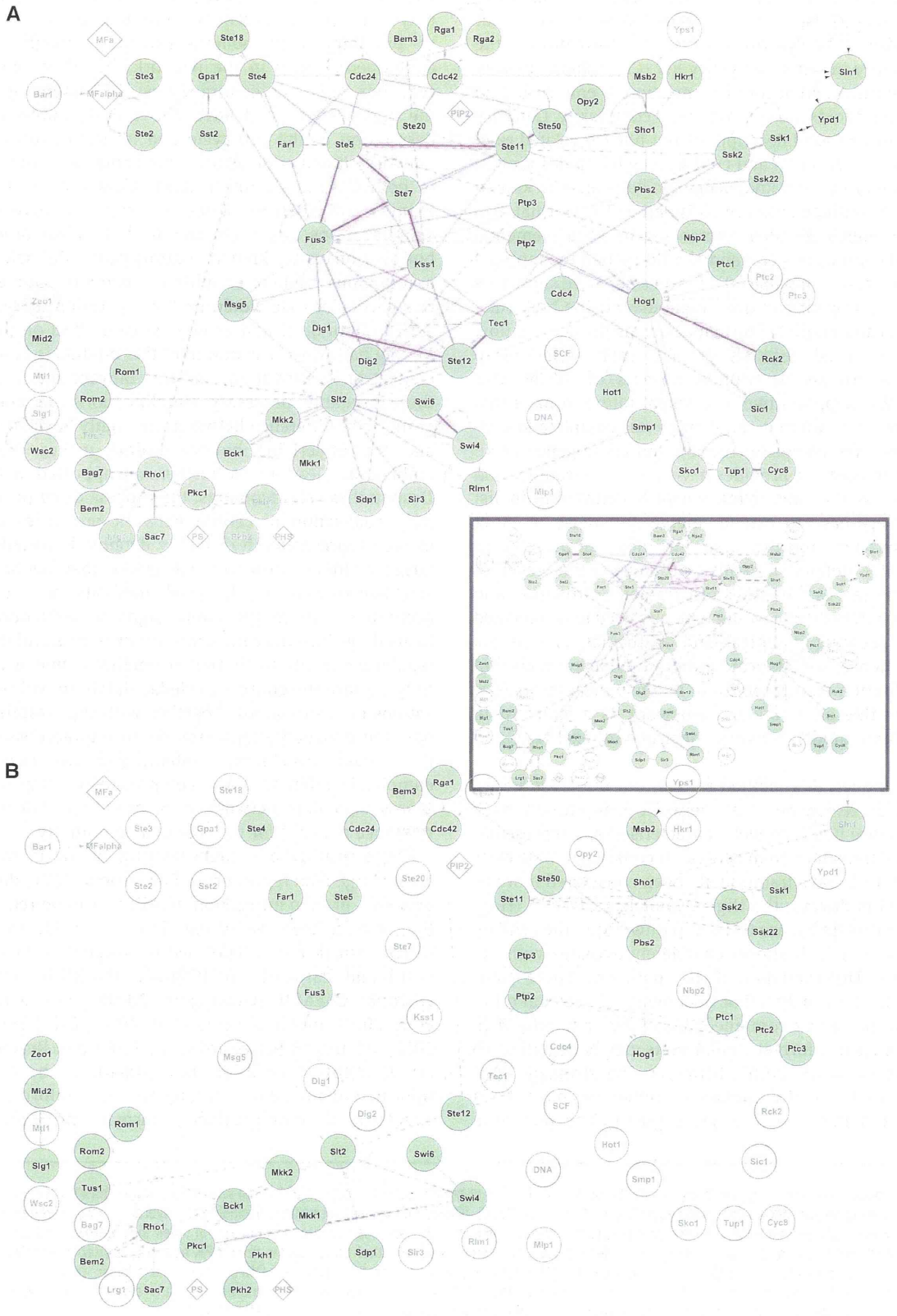
Finally, we automatically generated a mathematical description of the entire network as a proof of principle. The *rxncon* software used the contingency matrix to generate the input file for BioNetGen (Blinov *et al*, 2004). The corresponding network is too large to create but could be simulated with the network-free simulator NFSim (Sneddon *et al*, 2011). Further analysis of this system falls outside the scope of this paper, but the input file to BioNetGen and/or NFSim with trivial parameters is included as a supplement. Hence, a complete mathematical model can be automatically generated from the reaction and contingency data, and to our knowledge this is the first framework that integrates network definition at the granularity of empirical data with automatic visualisation and automatic model creation.

Discussion

It is clear that the complexity of signal-transduction networks is one of the major challenges in systems biology, impeding our ability to visualise, simulate and ultimately understand these networks. This issue has been widely recognised and substantial efforts have been committed to improve and standardise our tools for visualisation and modelling of

Figure 2 The reaction graph compactly displays the topology of the *S. cerevisiae* MAP kinase network. (**A**) The reaction graph of the MAPK network displays the components as nodes and the reactions as edges. Each component is defined by a central major node and peripheral minor nodes indicating domains, subdomains and specific residues (blue). When interacting domains and target residues are known, reactions are displayed as edges between these minor nodes. In contrast, the condensed reaction graph (**B**) displays each component as a single node, and each type of reaction between two nodes as a single edge. Nodes are either proteins (circles), small molecules (diamonds) or DNA (square). Edge colours indicate reaction type (co-substrates and co-products): Grey; protein–protein interaction (N/A), red; phosphorylation (− ATP, + ADP), orange; guanine nucleotide exchange (− GTP, + GDP), blue; dephosphorylation or GTPase activation (+ P_i), gold; ubiquitination (− ubiquitin, − ATP, + ADP, + P_i), black; phosphotransfer or proteolytic cleavage (N/A). The domain layout in (A) prioritises readability and domain organisation does not reflect linear sequence or protein structure. Arrowheads indicate directionality for unidirectional or reciprocal catalytic modifications. Reactions for which we found no direct evidence but which are supported by convincing genetic data has been included as dashed lines. Note the much higher frequency of reported phosphorylation reactions as compared with dephosphorylation reactions; in total the network includes 68 phosphorylation reactions but only 16 dephosphorylation reactions (A).

cellular networks (Hucka *et al*, 2003; Le Novere *et al*, 2009). These standardisation efforts are essential for data exchange and reusability, but many of the existing tools are unsuitable for definition, visualisation and mathematical modelling of large networks. The arguably most important problems are the combinatorial complexity, the granularity difference between



empirical and theoretical data, and the lack of exchange formats between different theoretical descriptions. Here, we have introduced a new framework for network definition at the same granularity as most empirical data. This format was already available for C1 (reaction) information, as our list of elemental reactions uses the same format as high-throughput data (PSICQUIC). We describe contextual information at the same granularity in our contingency list (C2), which not only allows an intuitive and accurate translation of empirical data but also largely avoids the combinatorial complexity. Contrary to state transition based descriptions but like the related rule-based format, the reaction and contingency based description becomes smaller the less knowledge we have as only known reactions and contingencies are considered. This format also provides for highly detailed referencing as each elemental reaction and contingency can and should be tied to empirical evidence (i.e., research paper(s)). Furthermore, we show that this format is stringent and unambiguously define both rule-based models and graphical formats, such as the activity flow diagram (condensed reaction graph), entity relationship diagram and process description formats of SBGN. Our framework also supports two new visualisation formats that we introduce here and that can display our complete knowledge database (the complete reaction and contingency lists). Finally, our framework provides a very high reusability and extendibility, as the underlying network definition—in list format—is very easy to extend, merge and reuse in other context, which is not the case for most graphically or mathematically defined systems. Of course, this level of definition still leaves the issues of parameter estimation and graphical layout, but these would typically need to be repeated even when merging graphical and mathematical network definitions. Hence, we advocate a more fundamental level of network definition than graphical or mathematical formalism. We envisage this or a similar framework as a standard to greatly facilitate model/network construction, exchange and reusability.

We have applied this method to map out the MAP kinase network of *S. cerevisiae*. This network was chosen as a benchmark since it is both well characterised and representative for signal transduction in general. It consists of three clear subgraphs, which have traditionally been considered more or less insulated pathways; the High Osmolarity Glycerol (Hog) pathway, the Protein Kinase C (PKC) pathway and the MATing (MAT) pathway, which almost completely overlaps with the PseudoHyphal Differentiation (PHD) pathway. These pathways have also been mapped or documented in several other efforts. KEGG presents a combined map of the traditional MAP kinase pathways in a format similar to its metabolic pathways (Kanehisa *et al*, 2006, 2010). However, the stringent edge definitions used for the metabolic networks have been abandoned and this is a 'biologist's graph'. The picture is similar with the maps of yeast MAP kinase pathways at Science STKE (e.g., Thorner *et al*, 2005). For example, these maps display Ste11 with four upstream regulators, but it is unclear how they regulate Ste11 and how their contributions combine (e.g., AND or OR?). Therefore, these network maps may provide an excellent introduction to the networks by providing a components list and a rough idea of the components' roles in the network, but they neither define reactions (C1) nor contingencies (C2) unambiguously. On the opposite end, we have the recently published process description of the cell cycle and its surrounding signalling network (Kaizu *et al*, 2010). This contains explicit definition of both C1 and C2 information. However, the tremendous number of specific states in such a network forces simplifications, which not only leads to a loss of knowledge, but also mixes up known contingencies (C2) with arbitrary assumptions made to simplify the network. One example in this particular case would be the separation of the upstream activation of Ste11 and its downstream effect on the Hog and Mating pathways. The output of this module is defined by the context of its activation, and this information is lost due to these arguably necessary simplifications. In addition, the granularity difference between the highly specific map states and the underlying biological data makes the mapping ambiguous, leading to further unsupported assumptions. Despite these shortcomings, the process description is useful for visualisation of certain network properties due to the explicit representation of highly detailed knowledge such as target residues. However, we stress that neither of these established and widely used methods are sufficient to accurately capture the entire signal-transduction network. Instead, we introduce the contingency matrix and the bipartite regulatory graph as alternative methods, which are able to fully capture the entire knowledge database without simplifications or assumptions. Together with the established methods, these visualisations provide an unprecedented view on the chosen benchmark system, and we trust that this completely referenced and comprehensive map of the MAP kinase signalling network in *S. cerevisiae* will be a useful reference material for the research community.

These results have direct bearing on the many efforts to create large data repositories. Pure reaction (C1) data, such as protein – protein interaction networks, can be retrieved using the standardised Molecular Interaction Query Language (MIQL; which our reaction list is designed to be compatible with) and PSICQUIC (PSICQUIC). PSICQUIC accesses, for example, ChEMBL (Overington, 2009), BioGrid (Breitkreutz *et al*, 2010), IntAct (Aranda *et al*, 2010), DIP (Xenarios *et al*, 2002), MatrixDB (Chautard *et al*, 2009) and Reactome (Croft *et al*, 2010). Several of these databases have additional information including contingency (C2) information and a standardised (non-graphical) format for definition and

Figure 3 The condensed reaction graph is an excellent tool for visualisation of high-throughput data. (**A**) Physical interactions within the MAPK network. The global protein–protein interaction network was retrieved from Biogrid (Stark *et al*, 2006), filtered for physical interactions excluding two hybrid, and visualised on the condensed reaction graph (Figure 2A). Purple edges indicate protein–protein interactions and their thickness indicates the number of times they were picked up, ranging from a single time (dashed line) to 19 times. Nodes that appear faded have no interactions with any other component in the MAPK network reported in this data set. Note that the nodes that do not correspond to single ORFs would be excluded automatically (e.g., the SCF complex, DNA, lipids). The smaller, boxed network display the corresponding two-hybrid interaction network. (**B**) Genetic interactions within the MAPK network. Synthetic lethal interactions were retrieved from Biogrid and visualised as per (A). Also quantitative data, such as mutant phenotypes and gene expression levels, can be directly visualised on the network.

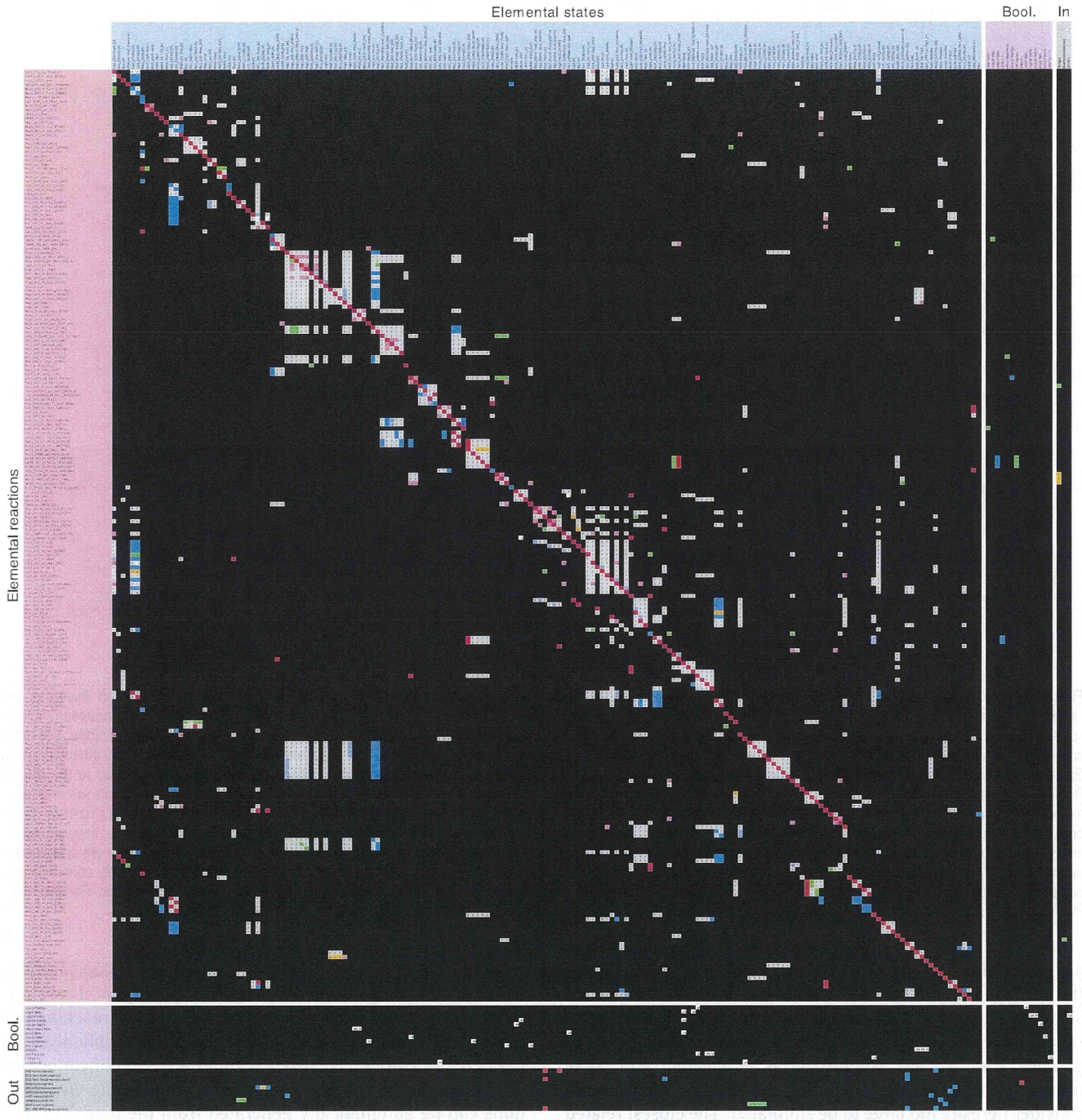


Figure 4 The contingency matrix provides a complete description of the network or network module. The core contingency matrix is spanned by the elemental reactions (rows, in red) and the elemental states (columns, in blue). The additional blocks are derived from the contingency list and contain the formation rules (rows) and effects (columns) of Boolean states (both purple) as well as the output of (rows) and input to (columns) the network (both grey). The cells in the matrix define how each reaction (row) depends on each state (column). The effects range from being absolutely required ('!'), via positive effector ('K + '), no effect ('0') and negative effector ('K–') to absolutely inhibitory ('x'), or it can be unknown or undefined ('?'). Each Boolean state is defined by a single operator ('AND' or 'OR') for the elemental states, other Booleans and/or inputs that defines it. The contingency matrix displayed here contains the complete MAPK network. Note that the contingency matrix is sparsely populated. This is both because most combinations of reactions and states lack overlap in components (black squares) and because we have very limited knowledge of the possible contingencies (grey squares). Overall, the information on what reactions can occur is much more abundant than on how they are regulated.

retrieval would further improve the usefulness of these resources and facilitate further analysis of the stored information. The framework we propose here provides such a format with the key advantage of including export to mathematical models. Since mathematical modelling is the most central and natural step to bring the knowledge in these databases into a useful form, where quantitative systems properties can most exhaustively be analysed, the introduction of such an export is

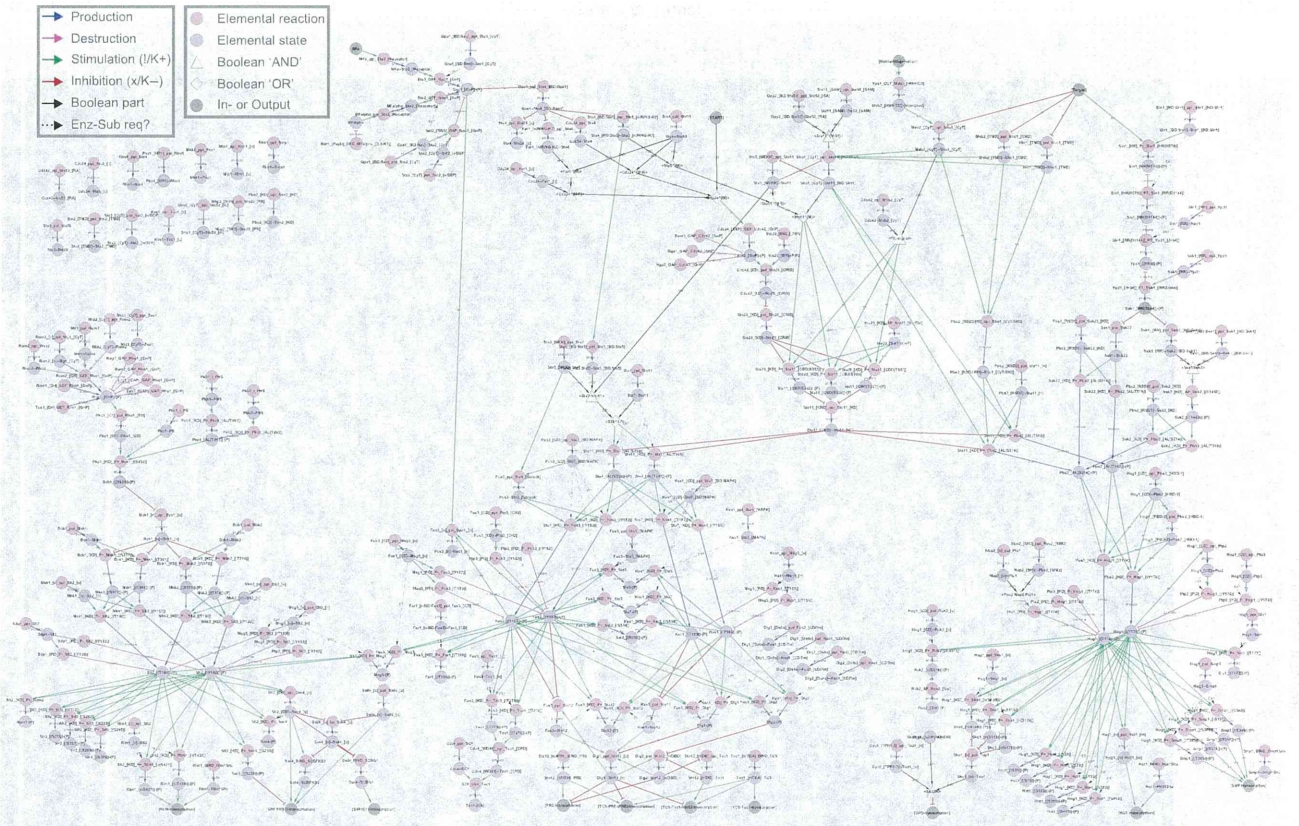


Figure 5 The regulatory graph visualise the causality between reactions and reveals the regulatory structure of the network. This bipartite graph illustrates the relationships between the reactions (red nodes) and states (blue nodes) within the network. Edges from reactions to states define how states are produced (blue) or consumed (purple), and each such edge corresponds to a single elemental reaction. Edges from states to reactions define how states regulate other reactions, and each such edge correspond to a single contingency (Green; absolute requirement ('!') or positive effector ('K + '), red; negative effector ('K–') or absolutely inhibitory ('x')). Booleans are used when the effect on a reaction cannot be attributed to single elemental states (white diamonds (OR) or triangles (AND) connected to the states/Booleans/inputs that define them with black lines). Inputs are displayed in grey and connected to the elemental reaction(s) they influence. Likewise, outputs are displayed in grey and connected to the states they are influenced by. Signals can be followed through the network from external cues (grey; top) to transcriptional response (grey; bottom) as all edges are directional. Reactions without input are not (known to be) regulated and would therefore be expected to have constant rates; likewise states without output have no (defined) impact on the system. We have also included likely but undocumented requirements for enzyme–substrate bindings before catalysis as dashed lines. The regulatory graph is the only graphical representation using the complete information in the contingency matrix, and hence the only complete and completely graphical visualisation of the network. It is also the most potent visualisation to evaluate the degree of knowledge about the network. For example, visualisation of high-throughput data would result in disconnected reaction–state pairs only, due to the lack of regulatory information (no C2 data).

an important step forward. This framework is still not as flexible as direct model definition but it provides distinct advantages. Formulating models directly using classical state transition reactions is either subjective or very cumbersome in practice due to the combinatorial explosion, and state transition based models for the networks of the size we consider here are too large to be simulated. The closest related modelling framework is rule-based modelling, in which models can be formulated without these combinatorial explosion problems, and it is also to a rule-based format that we export our models. However, the classical rule-based modelling frameworks lack all the database properties of our framework, such as the contingency matrix and its export to various novel visualisation formats. In short, one could therefore say that our framework combines the best of existing knowledge databases with new visualisation tools and rule-based modelling.

In conclusion, we present a method to document and visualise signal-transduction networks that improves on previous strategies in the following respects; (I) it allows concise mapping at the same granularity as biological data, hence pre-empting the need for implicit, unsupported assumptions, (II) it allows referencing of each elemental reaction and contingency separately and handles unknowns explicitly, (III) the network can be visualised without any simplifications or assumptions that increase the uncertainty, (IV) the visualisations can be automatically generated from the data files, (V) the network definition is a template from which a mathematical model can be automatically generated (VI) and exported to SBML and (VII) the supplied template and *rxncon* tool makes the method immediately useful for anyone with an interest in signal transduction. Hence, our framework bridge three critical levels of signal-transduction network analysis; definition, visualisation and mathematical modelling, as well as empirical data and theoretical analysis.

Materials and methods

The MAP kinase network map is based on the papers listed below. The specific reference(s) are listed for each reaction and contingency

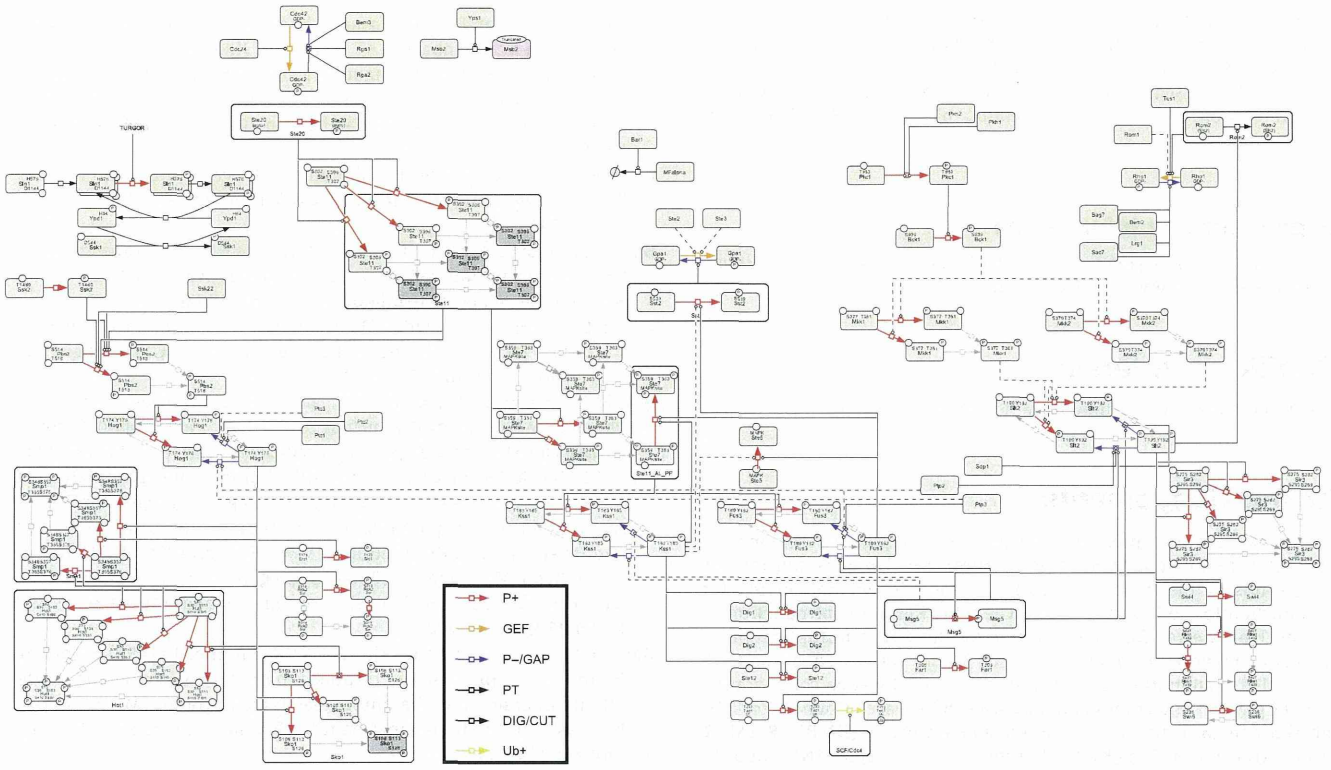


Figure 6 The limited process description displays all posttranslational modifications and their catalysts, but excludes complex formation. Each specific internal state is represented as a distinct node, although some intermediate phosphorylation states have been excluded. Phosphorylations are indicated with red arrows (ATP as co-substrate and ADP as co-product), GEF reactions as orange arrows (− GTP, + GDP), and dephosphorylation or GAP reactions as blue arrows (+ P_i). Only a fraction of the catalytic modifications have a known catalyst for both forward and reverse reactions, and the required state of the catalyst known is in even fewer cases. Therefore, even this highly simplified process description includes uncertainty in the required states of both catalysts and substrates. In this visualisation, this uncertainty has been shown by using a single catalysis arrow from a box including all potentially active state of the catalyst to the basic state of the substrate (completely unphosphorylated for kinase reactions, or completely phosphorylated for phosphatase reactions). While these simplifications are unsupported, including additional catalytic arrows would be equally arbitrary with the added drawback of making the figure more complex (see Supplementary Figure S2). Despite the need for implicit assumptions, the process description is useful as it is very explicit and intuitive to read.

individually in the reaction and contingency lists in the 'PubMedIdentifier(s)' column with their PMID number.

(Ai *et al*, 2002; Alepuz *et al*, 2003; Alepuz *et al*, 2001; Andrews and Herskowitz, 1989; Andrews and Moore, 1992; Apanovitch *et al*, 1998; Baetz and Andrews, 1999; Baetz *et al*, 2001; Ballon *et al*, 2006; Bao *et al*, 2004; Bao *et al*, 2010; Bar *et al*, 2003; Bardwell *et al*, 1996; Bardwell *et al*, 1998a; Bardwell *et al*, 1998b; Bender and Sprague, 1986; Bilsland-Marchesan *et al*, 2000; Blumer *et al*, 1988; Breitkreutz *et al*, 2001; Bruckner *et al*, 2004; Butty *et al*, 1998; Chou *et al*, 2004; Chou *et al*, 2006; Cismowski *et al*, 2001; Clark *et al*, 1993; Collister *et al*, 2002; Cook *et al*, 1996; Crosby *et al*, 2000; Cullen *et al*, 2004; Davenport *et al*, 1999; de Nadal *et al*, 2003; Dodou and Treisman, 1997; Doi *et al*, 1994; Dolan *et al*, 1989; Dowell *et al*, 1998; Drogen *et al*, 2000; Elion *et al*, 1993; Errede *et al*, 1993; Escote *et al*, 2004; Feng *et al*, 1998; Fitch *et al*, 2004; Flandez *et al*, 2004; Flotho *et al*, 2004; Friant *et al*, 2001; Garcia-Gimeno and Struhl, 2000; Garrison *et al*, 1999; Gartner *et al*, 1998; Gartner *et al*, 1992; Good *et al*, 2009; Green *et al*, 2003; Guo *et al*, 2009; Hagen *et al*, 1986; Hagen *et al*, 1991; Hahn and Thiele, 2002; Heenan *et al*, 2009; Heise *et al*, 2010; Ho *et al*, 2002; Horie *et al*, 2008; Inagaki *et al*, 1999; Inouye *et al*, 1997a; Inouye *et al*, 1997b; Irie *et al*, 1993; Jacoby *et al*, 1997; Jung *et al*, 2002; Kamada *et al*, 1995; Kamada *et al*, 1996; Ketela *et al*, 1999; Kim *et al*, 2010; Kim *et al*, 2008; Kranz *et al*, 1994; Kusari *et al*, 2004; Lamson *et al*, 2002; Lee and Levin, 1992; Leeuw *et al*, 1995; Leeuw *et al*, 1998; Li *et al*, 1998; Liu *et al*, 2005; MacKay *et al*, 1991; MacKay *et al*, 1988; Madden *et al*, 1997; Madhani and Fink, 1997; Madhani *et al*, 1997; Maeda *et al*, 1995; Maeda *et al*, 1994; Maleri *et al*, 2004; Mapes and Ota, 2004; Martin *et al*, 2000; Mattison and Ota, 2000; Mattison *et al*, 1999; Medici *et al*, 1997; Melcher and Thorner, 1996; Metodiev *et al*, 2002; Miyajima *et al*, 1987; Murakami *et al*, 2008; Nasmyth and Dirick, 1991; Nehlin *et al*, 1992; Neiman and Herskowitz, 1994; Nern and Arkowitz, 1998; Nern and Arkowitz, 1999; Nonaka *et al*, 1995; Olson *et al*, 2000; Ostrander and Gorman, 1999; Ozaki *et al*, 1996; Paravicini and Friedli, 1996; Parnell *et al*, 2005; Pascual-Ahuir *et al*, 2001; Peter *et al*, 1996; Peterson *et al*, 1994; Philip and Levin, 2001; Posas and Saito, 1997; Posas and Saito, 1998; Posas *et al*, 1998; Posas *et al*, 1996; Proft *et al*, 2005; Proft *et al*, 2001; Proft and Serrano, 1999; Proft and Struhl, 2002; Raicu *et al*, 2005; Raitt *et al*, 2000; Rajavel *et al*, 1999; Reiser *et al*, 2000; Remenyi *et al*, 2005; Rep *et al*, 2000; Rep *et al*, 1999; Roberts and Fink, 1994; Schmelzle *et al*, 2002; Schmidt *et al*, 1997; Schmidt *et al*, 2002; Schmitz *et al*, 2002; Shi *et al*, 2005; Shimada *et al*, 2004; Sidorova and Breeden, 1993; Siegmund and Nasmyth, 1996; Siekhaus and Drubin, 2003; Simon *et al*, 1995; Skowyra *et al*, 1997; Smith *et al*, 2002; Soler *et al*, 1995; Song *et al*, 1996; Taba *et al*, 1991; Takahashi and Pryciak, 2007; Tao *et al*, 2002; Tarassov *et al*, 2008; Tatebayashi *et al*, 2003; Tatebayashi *et al*, 2007; Tatebayashi *et al*, 2006; Tedford *et al*, 1997; Truckses *et al*, 2006; Truman *et al*, 2009; Vadaie *et al*, 2008; Valtz *et al*, 1995; Varanasi *et al*, 1996; Verna *et al*, 1997; Vilella *et al*, 2005; Wang and Konopka, 2009; Wang *et al*, 2005; Warmka *et al*, 2001; Wassmann and Ammerer, 1997; Watanabe *et al*, 1994; Watanabe *et al*, 1995; Watanabe *et al*, 1997; Winters and Pryciak, 2005; Wu *et al*, 2006; Wu *et al*, 1999; Wu *et al*, 1995; Wu *et al*, 2004; Wurgler-Murphy *et al*, 1997; Yablonski *et al*, 1996; Yamamoto *et al*, 2010; Yesilaltay and Jenness, 2000; Young *et al*, 2002; Yuan and Fields, 1991; Zarrinpar *et al*, 2004;

Zarrinpar *et al*, 2003; Zarzov *et al*, 1996; Zeitlinger *et al*, 2003; Zhan *et al*, 1997; Zhan and Guan, 1999; Zhao *et al*, 1995; Zheng and Guan, 1994; Zheng *et al*, 1994; Zhou *et al*, 1993).

The methods used are an integral part of the results and are outlined in that section. For additional details, please see Supplementary information.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We thank past and present colleagues for helpful discussions; in particular Akira Funahashi, Noriko Hiroi and Douglas Murray for suggestions in the conception phase, Hans-Michael Kaltenbach for introduction to the bipartite graph, Clemens Kühn for introduction to NFsim, Jens Nielsen for the suggestion to use matrix multiplication to calculate network distance and Nina Arens for proofreading. We acknowledge support from JSPS and SSF (Japan-Sweden collaborative postdoc grant to MK), Lions and the Swedish Research Council (to GC), the German Ministry for Education and Research (BMBF, SysMO2 project Translucent 2 to EK), the European Commission (UNICELLSYS, Grant 201142, AQUAGLYCEROPORIN, Grant 35995, CELLCOMPUT, Grant 043310 and SYSTEMSBIOLOGY, Grant 514169, all to SH and EK) and from the MULTIDISCIPLINARY BIO Sweden-Japan initiative (Sweden: Foundation for Strategic Research SSF and Vinnova, Japan: Japan Science and Technology Agency JST) to SH and HK. Work in the laboratory of SH was also supported by a grant from the Swedish Research Council (Grant 2007-4905).

Author contributions: HK initiated the mapping project. MK conceived the framework. GC and MK developed the framework with input from all the authors. CFT and MK mapped the MAP kinase network. FK and RP implemented the framework with guidance from GC and MK. FK created the *rxncon* software tool. MK drafted the first manuscript with help from GC. SH and EK contributed biological and theoretical background knowledge, respectively. SH, EK and HK provided the research environments and contributed to completion of the manuscript. All authors read, edited and approved the final manuscript.

References

Ai W, Bertram PG, Tsang CK, Chan TF, Zheng XF (2002) Regulation of subtelomeric silencing during stress response. *Mol Cell* **10:** 1295–1305

Alepuz PM, de Nadal E, Zapater M, Ammerer G, Posas F (2003) Osmostress-induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II. *EMBO J* **22:** 2433–2442

Alepuz PM, Jovanovic A, Reiser V, Ammerer G (2001) Stress-induced map kinase Hog1 is part of transcription activation complexes. *Mol Cell* **7:** 767–777

Andrews BJ, Herskowitz I (1989) Identification of a DNA binding factor involved in cell-cycle control of the yeast HO gene. *Cell* **57:** 21–29

Andrews BJ, Moore LA (1992) Interaction of the yeast Swi4 and Swi6 cell cycle regulatory proteins in vitro. *Proc Natl Acad Sci USA* **89:** 11852–11856

Apanovitch DM, Slep KC, Sigler PB, Dohlman HG (1998) Sst2 is a GTPase-activating protein for Gpa1: purification and characterization of a cognate RGS-Galpha protein pair in yeast. *Biochemistry* **37:** 4815–4822

Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, Feuermann M, Ghanbarian AT, Kerrien S, Khadake J, Kerssemakers J, Leroy C, Menden M, Michaut M, Montecchi-Palazzi L, Neuhauser SN, Orchard S, Perreau V, Roechert B, van Eijk K *et al* (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res* **38:** D525–D531

Baetz K, Andrews B (1999) Regulation of cell cycle transcription factor Swi4 through auto-inhibition of DNA binding. *Mol Cell Biol* **19:** 6729–6741

Baetz K, Moffat J, Haynes J, Chang M, Andrews B (2001) Transcriptional coregulation by the cell integrity mitogen-activated protein kinase Slt2 and the cell cycle regulator Swi4. *Mol Cell Biol* **21:** 6515–6528

Ballon DR, Flanary PL, Gladue DP, Konopka JB, Dohlman HG, Thorner J (2006) DEP-domain-mediated regulation of GPCR signaling responses. *Cell* **126:** 1079–1093

Bao MZ, Schwartz MA, Cantin GT, Yates JR3rd, Madhani HD (2004) Pheromone-dependent destruction of the Tec1 transcription factor is required for MAP kinase signaling specificity in yeast. *Cell* **119:** 991–1000

Bao MZ, Shock TR, Madhani HD (2010) Multisite phosphorylation of the Saccharomyces cerevisiae filamentous growth regulator Tec1 is required for its recognition by the E3 ubiquitin ligase adaptor Cdc4 and its subsequent destruction in vivo. *Eukaryot Cell* **9:** 31–36

Bar EE, Ellicott AT, Stone DE (2003) Gbetagamma recruits Rho1 to the site of polarized growth during mating in budding yeast. *J Biol Chem* **278:** 21798–21804

Bardwell L, Cook JG, Chang EC, Cairns BR, Thorner J (1996) Signaling in the yeast pheromone response pathway: specific and high-affinity interaction of the mitogen-activated protein (MAP) kinases Kss1 and Fus3 with the upstream MAP kinase kinase Ste7. *Mol Cell Biol* **16:** 3637–3650

Bardwell L, Cook JG, Voora D, Baggott DM, Martinez AR, Thorner J (1998a) Repression of yeast Ste12 transcription factor by direct binding of unphosphorylated Kss1 MAPK and its regulation by the Ste7 MEK. *Genes Dev* **12:** 2887–2898

Bardwell L, Cook JG, Zhu-Shimoni JX, Voora D, Thorner J (1998b) Differential regulation of transcription: repression by unactivated mitogen-activated protein kinase Kss1 requires the Dig1 and Dig2 proteins. *Proc Natl Acad Sci USA* **95:** 15400–15405

Bender A, Sprague GFJr. (1986) Yeast peptide pheromones, a-factor and alpha-factor, activate a common response mechanism in their target cells. *Cell* **47:** 929–937

Bilsland-Marchesan E, Arino J, Saito H, Sunnerhagen P, Posas F (2000) Rck2 kinase is a substrate for the osmotic stress-activated mitogen-activated protein kinase Hog1. *Mol Cell Biol* **20:** 3887–3895

Biographer http://code.google.com/p/biographer/

Blinov ML, Faeder JR, Goldstein B, Hlavacek WS (2004) BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics (Oxford, England)* **20:** 3289–3291

Blinov ML, Yang J, Faeder JR, Hlavacek WS (2006) Depicting signaling cascades. *Nature Biotechnology* **24:** 137–138author reply 138

Blumer KJ, Reneke JE, Thorner J (1988) The STE2 gene product is the ligand-binding component of the alpha-factor receptor of Saccharomyces cerevisiae. *J Biol Chem* **263:** 10836–10842

Borisov NM, Chistopolsky AS, Faeder JR, Kholodenko BN (2008) Domain-oriented reduction of rule-based network models. *IET Syst Biol* **2:** 342–351

Breitkreutz A, Boucher L, Tyers M (2001) MAPK specificity in the yeast pheromone response independent of transcriptional activation. *Curr Biol* **11:** 1266–1271

Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, Lin ZY, Breitkreutz BJ, Stark C, Liu G, Ahn J, Dewar-Darch D, Reguly T, Tang X, Almeida R, Qin ZS, Pawson T, Gingras AC, Nesvizhskii AI, Tyers M (2010) A global protein kinase and phosphatase interaction network in yeast. *Science (New York, NY)* **328:** 1043–1046

Bruckner S, Kohler T, Braus GH, Heise B, Bolte M, Mosch HU (2004) Differential regulation of Tec1 by Fus3 and Kss1 confers signaling specificity in yeast development. *Curr Genet* **46:** 331–342

Butty AC, Pryciak PM, Huang LS, Herskowitz I, Peter M (1998) The role of Far1p in linking the heterotrimeric G protein to polarity establishment proteins during yeast mating. *Science* **282:** 1511–1516

Chautard E, Ballut L, Thierry-Mieg N, Ricard-Blum S (2009) MatrixDB, a database focused on extracellular protein-protein and protein-carbohydrate interactions. *Bioinformatics (Oxford, England)* **25:** 690–691

Chou S, Huang L, Liu H (2004) Fus3-regulated Tec1 degradation through SCFCdc4 determines MAPK signaling specificity during mating in yeast. *Cell* **119:** 981–990

Chou S, Lane S, Liu H (2006) Regulation of mating and filamentation genes by two distinct Ste12 complexes in Saccharomyces cerevisiae. *Mol Cell Biol* **26:** 4794–4805

Chylek LA, Hu B, Blinov ML, Emonet T, Faeder JR, Goldstein B, Gutenkunst RN, Haugh JM, Lipniacki T, Posner RG, Yang J, Hlavacek WS (2011) Guidelines for visualizing and annotating rule-based models. *Mol Biol Syst* **7:** 2779–2795

Cismowski MJ, Metodiev M, Draper E, Stone DE (2001) Biochemical analysis of yeast G(alpha) mutants that enhance adaptation to pheromone. *Biochem Biophys Res Commun* **284:** 247–254

Clark KL, Dignard D, Thomas DY, Whiteway M (1993) Interactions among the subunits of the G protein involved in Saccharomyces cerevisiae mating. *Mol Cell Biol* **13:** 1–8

Collister M, Didmon MP, MacIsaac F, Stark MJ, MacDonald NQ, Keyse SM (2002) YIL113w encodes a functional dual-specificity protein phosphatase which specifically interacts with and inactivates the Slt2/Mpk1p MAP kinase in S. cerevisiae. *FEBS Lett* **527:** 186–192

Conzelmann H, Fey D, Gilles ED (2008) Exact model reduction of combinatorial reaction networks. *BMC Syst Biol* **2:** 78

Cook JG, Bardwell L, Kron SJ, Thorner J (1996) Two novel targets of the MAP kinase Kss1 are negative regulators of invasive growth in the yeast Saccharomyces cerevisiae. *Genes Dev* **10:** 2831–2848

Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, Jupe S, Kataskaya I, Mahajan S, May B, Ndegwa N, Schmidt E, Shamovsky V, Yung C, Birney E, Hermjakob H *et al* (2010) *Reactome: a database of reactions, pathways and biological processesNucleic Acids Res 39(Database issue): D691–D697*

Crosby JA, Konopka JB, Fields S (2000) Constitutive activation of the Saccharomyces cerevisiae transcriptional regulator Ste12p by mutations at the amino-terminus. *Yeast* **16:** 1365–1375

Cullen PJ, Sabbagh WJr., Graham E, Irick MM, van Olden EK, Neal C, Delrow J, Bardwell L, Sprague GFJr. (2004) A signaling mucin at the head of the Cdc42- and MAPK-dependent filamentous growth pathway in yeast. *Genes Dev* **18:** 1695–1708

Danos V (2007) Rule-based modelling of cellular signalling. *Cell Signal* 17–41

Davenport KD, Williams KE, Ullmann BD, Gustin MC (1999) Activation of the Saccharomyces cerevisiae filamentation/invasion pathway by osmotic stress in high-osmolarity glycogen pathway mutants. *Genetics* **153:** 1091–1103

de Nadal E, Casadome L, Posas F (2003) Targeting the MEF2-like transcription factor Smp1 by the stress-activated Hog1 mitogen-activated protein kinase. *Mol Cell Biol* **23:** 229–237

Dodou E, Treisman R (1997) The Saccharomyces cerevisiae MADS-box transcription factor Rlm1 is a target for the Mpk1 mitogen-activated protein kinase pathway. *Mol Cell Biol* **17:** 1848–1859

Doi K, Gartner A, Ammerer G, Errede B, Shinkawa H, Sugimoto K, Matsumoto K (1994) MSG5, a novel protein phosphatase promotes adaptation to pheromone response in S. cerevisiae. *EMBO J* **13:** 61–70

Dolan JW, Kirkman C, Fields S (1989) The yeast STE12 protein binds to the DNA sequence mediating pheromone induction. *Proc Natl Acad Sci USA* **86:** 5703–5707

Dowell SJ, Bishop AL, Dyos SL, Brown AJ, Whiteway MS (1998) Mapping of a yeast G protein betagamma signaling interaction. *Genetics* **150:** 1407–1417

Drogen F, O'Rourke SM, Stucke VM, Jaquenoud M, Neiman AM, Peter M (2000) Phosphorylation of the MEKK Ste11p by the PAK-like kinase Ste20p is required for MAP kinase signaling in vivo. *Curr Biol* **10:** 630–639

Elion EA, Satterberg B, Kranz JE (1993) FUS3 phosphorylates multiple components of the mating signal transduction cascade: evidence for STE12 and FAR1. *Mol Biol Cell* **4:** 495–510

Errede B, Gartner A, Zhou Z, Nasmyth K, Ammerer G (1993) MAP kinase-related FUS3 from S. cerevisiae is activated by STE7 in vitro. *Nature* **362:** 261–264

Escote X, Zapater M, Clotet J, Posas F (2004) Hog1 mediates cell-cycle arrest in G1 phase by the dual targeting of Sic1. *Nat Cell Biol* **6:** 997–1002

Faeder JR, Blinov ML, Goldstein B, Hlavacek WS (2005) Rule-based modeling of biochemical networks. *Complexity* **10:** 22–41

Feng Y, Song LY, Kincaid E, Mahanty SK, Elion EA (1998) Functional binding between Gbeta and the LIM domain of Ste5 is required to activate the MEKK Ste11. *Curr Biol* **8:** 267–278

Fitch PG, Gammie AE, Lee DJ, de Candal VB, Rose MD (2004) Lrg1p Is a Rho1 GTPase-activating protein required for efficient cell fusion in yeast. *Genetics* **168:** 733–746

Flandez M, Cosano IC, Nombela C, Martin H, Molina M (2004) Reciprocal regulation between Slt2 MAPK and isoforms of Msg5 dual-specificity protein phosphatase modulates the yeast cell integrity pathway. *J Biol Chem* **279:** 11027–11034

Flotho A, Simpson DM, Qi M, Elion EA (2004) Localized feedback phosphorylation of Ste5p scaffold by associated MAPK cascade *J Biol Chem* **279:** 47391–47401

Friant S, Lombardi R, Schmelzle T, Hall MN, Riezman H (2001) Sphingoid base signaling via Pkh kinases is required for endocytosis in yeast. *EMBO J* **20:** 6783–6792

Garcia-Gimeno MA, Struhl K (2000) Aca1 and Aca2, ATF/CREB activators in Saccharomyces cerevisiae, are important for carbon source utilization but not the response to stress. *Mol Cell Biol* **20:** 4340–4349

Garrison TR, Zhang Y, Pausch M, Apanovitch D, Aebersold R, Dohlman HG (1999) Feedback phosphorylation of an RGS protein by MAP kinase in yeast. *J Biol Chem* **274:** 36387–36391

Gartner A, Jovanovic A, Jeoung DI, Bourlat S, Cross FR, Ammerer G (1998) Pheromone-dependent G1 cell cycle arrest requires Far1 phosphorylation, but may not involve inhibition of Cdc28-Cln2 kinase, in vivo. *Mol Cell Biol* **18:** 3681–3691

Gartner A, Nasmyth K, Ammerer G (1992) Signal transduction in Saccharomyces cerevisiae requires tyrosine and threonine phosphorylation of FUS3 and KSS1. *Genes Dev* **6:** 1280–1292

Good M, Tang G, Singleton J, Remenyi A, Lim WA (2009) The Ste5 scaffold directs mating signaling by catalytically unlocking the Fus3 MAP kinase for activation. *Cell* **136:** 1085–1097

Green R, Lesage G, Sdicu AM, Menard P, Bussey H (2003) A synthetic analysis of the Saccharomyces cerevisiae stress sensor Mid2p, and identification of a Mid2p-interacting protein, Zeo1p, that modulates the PKC1-MPK1 cell integrity pathway. *Microbiology* **149:** 2487–2499

Guo S, Shen X, Yan G, Ma D, Bai X, Li S, Jiang Y (2009) A MAP kinase dependent feedback mechanism controls Rho1 GTPase and actin distribution in yeast. *PLoS One* **4:** e6089

Hagen DC, McCaffrey G, Sprague GFJr. (1986) Evidence the yeast STE3 gene encodes a receptor for the peptide pheromone a factor: gene sequence and implications for the structure of the presumed receptor. *Proc Natl Acad Sci USA* **83:** 1418–1422

Hagen DC, McCaffrey G, Sprague GFJr. (1991) Pheromone response elements are necessary and sufficient for basal and pheromone-induced transcription of the FUS1 gene of Saccharomyces cerevisiae. *Mol Cell Biol* **11:** 2952–2961

Hahn JS, Thiele DJ (2002) Regulation of the Saccharomyces cerevisiae Slt2 kinase pathway by the stress-inducible Sdp1 dual specificity phosphatase. *J Biol Chem* **277:** 21278–21284

Heenan EJ, Vanhooke JL, Temple BR, Betts L, Sondek JE, Dohlman HG (2009) Structure and function of Vps15 in the endosomal G protein signaling pathway. *Biochemistry* **48:** 6390–6401