# Correspondence

## Potential artefacts in protein-interaction networks

Patrick Aloy, Robert B. Russell*

First published online 20 September 2002

The construction of cell networks is an important goal for biology. Among the many networks now studied are those derived from protein-interaction data [1–5]. Recently a network built from large-scale yeast two-hybrid data [2] was found to have an inherent stability, and argued to be in accordance with compartmentalisation of cell processes [6]. Links between highly connected proteins, or hubs, and those with few connections were favoured meaning that hubs were well separated. The resulting decrease in cross-talk between hubs would increase network stability by restricting the effect of deleterious perturbations. However, we argue that the perceived stability is probably an artefact of the two-hybrid data.

In a two-hybrid experiment [7], one protein is attached to a transcription factor DNA binding domain ('bait'), and another to the activation domain ('prey'). A signal results only if an interaction occurs in the nucleus and orients the two parts of the transcription factor such that they both bind DNA and activate the transcription machinery. Since fusing proteins might disrupt structure or function, it is not surprising that interactions seen in one orientation (i.e. A as bait, B as prey) are not always seen in the other (i.e. B as bait, A as prey). In the same dataset, we observed that proteins with fewer than 30 partners show, on average, an equal preference to interact as bait or prey, although individual proteins sometimes interact only as bait or prey (Fig. 1). Similar preferences are also seen for both the 'core' data from the same yeast-two-hybrid study [2] and those from another large-scale study [1].

The situation is very different for the 23 hubs with $\geq 30$ partners: all but one strongly prefer to be bait rather than prey (Fig. 1). Given the equal bait/prey preference observed for the other proteins in the set, the probability of this occurring by chance is about $10^{-5}$. The variation in hub function (see below) makes a simple biological explanation for this bias unlikely. We argue that it is most probably caused by a systematic error in the data. Indeed removing the hubs abolishes the previously observed bait/prey asymmetry [6]. Considering all interactions, the baits interact with three partners on average compared to 1.8 for preys [6], whereas without the hubs both values are 1.7. The one hub not showing a bait preference (AGP17) is not well understood, but may truly be promiscuous regarding interacting partners, since, in contrast to most of the others, interactions with this protein are frequently seen in both directions.

The hubs considered by Maslov and Sneppen [6] consist of eight baits that interact with $\geq 90$ partners and seven preys with $\geq 30$ proteins, making a total of 1220 interactions. The hub–baits, although accounting for 1042 interactions, only are linked twice, whereas the hub–preys, with 178 interactions, are linked six times, roughly as expected from a random model. This observation also supports the idea that the baits making many interactions might be an artefact of the two-hybrid system, since only they behave differently from expected.

Further evidence for the artifactual nature of these hubs comes from yeast knockout experiments [8]: only six out of 23 are lethal when removed. This is not what one would expect given a recent analysis of a more carefully curated dataset that found a strong correlation between lethality and the number of interacting partners [9].

It is also informative to consider exceptions to the rule. For example, VMA6, a subunit of vacuolar ATPase interacts as bait with 88 proteins, but as prey with only one: VMA22, a vacuolar ATPase assembly protein. The similarity in the names alone suggests the interaction is genuine, unlike those found with the same protein as bait. These include dubious looking interactions with a metabolic enzyme (TYR1), mitochondrial citrate synthetase (CIT1), and a nuclear splicing factor (NOP13). These exceptions also demonstrate that these supposedly promiscuous proteins are *able* to interact as prey, but do so only as often as one might expect for an average protein.

What is the cause of these bait-biased hubs? Heat-shock, ribosomal and certain other proteins are apparently sticky [3], being naturally promiscuous for interaction partners. However, these are not seen among the hubs, and moreover there would be no logical reason to be promiscuous only as baits. We suspect that the bias is instead related to GAL4 function, with transcription in the absence of the C-terminal activation domain possibly accomplished by the presence of an activation domain-like acidic blob [10]. Indeed, baits are known to turn on reporter genes in the absence of any prey (e.g. [11]).

When we consider cleaner yeast-two-hybrid datasets (e.g. [1]; 'core' data of [2]), the systematic suppression of links between highly connected proteins is much less evident (data not shown) and we expect that any effect still seen is due to artefacts that are still present, but just to a lesser degree. Thus, the observed decrease in cross-talk between hubs is also likely to be an artefact of the two-hybrid data used in the study.

Despite being prone to errors, interactions identified by high-throughput methods applied to whole genomes have provided new insights into protein function. Networks derived from the data reveal unexpected features that are changing perceptions of protein synergies in the cell [12]. However, to be most useful networks should be built only with reliable data and be scrutinised constantly for features at odds with biological reality. It is dangerous to draw general conclusions about the behaviour of cell networks when using data that are error-prone and incomplete, and when the nature of systematic errors is poorly understood.

### Note added in proof

The different scales used in the figures in the response of Maslov and Sheppen (0.6-1.2 or 0.3-1.3) from those in their original paper (0-1.4) magnifies the reported effect.


PII: S0014-5793(02)03427-0

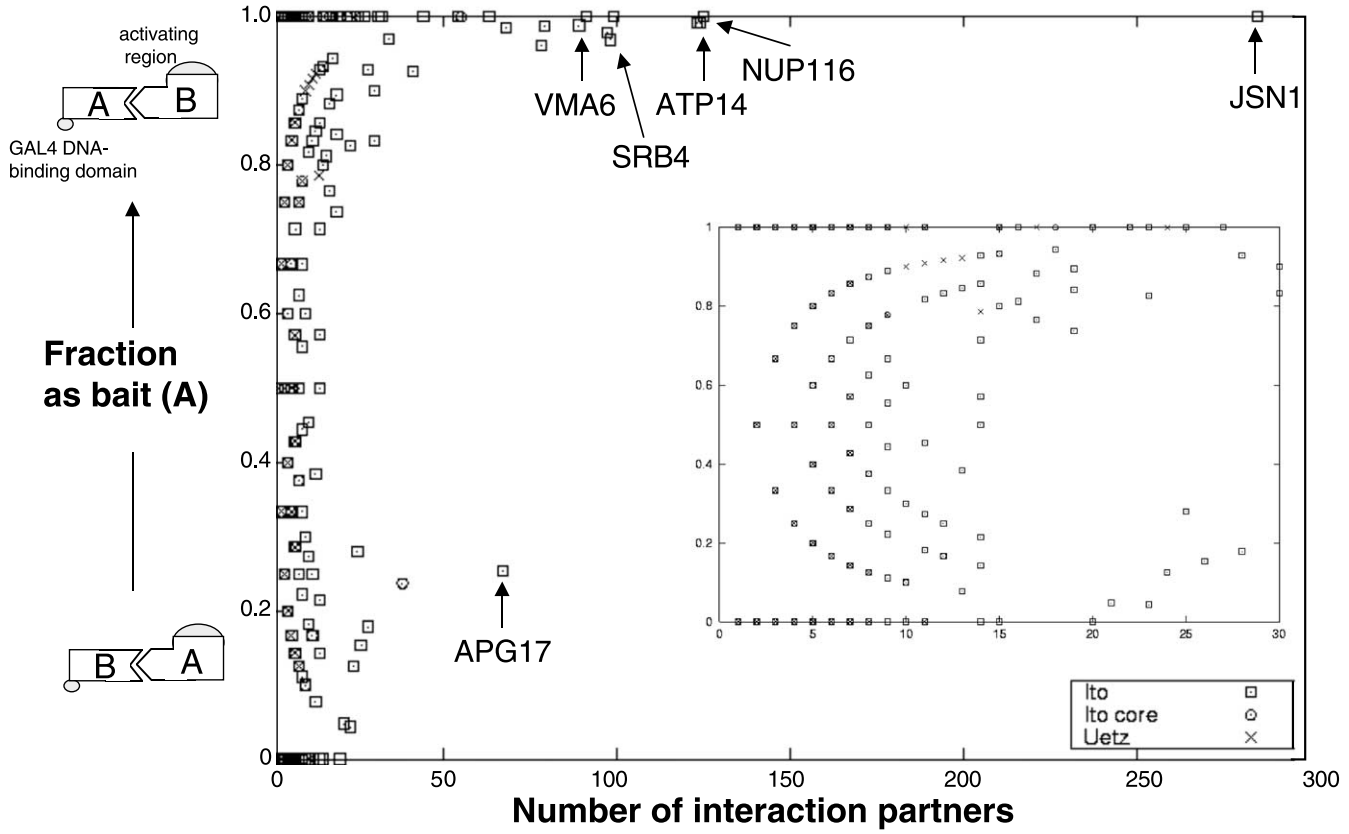


Fig. 1. Plot of the fraction of interactions occurring as bait vs. the number of interaction partners for yeast two-hybrid data. 'Ito' includes 4549 interactions and 'Ito core' a subset of 841 of these thought to be the most reliable [2]; 'Uetz' includes a total of 958 interactions [1]. Proteins discussed in the text are labelled with their yeast protein database codes. No points are obscured by the inset, which shows the subset of proteins making fewer than 30 interactions.

## References

[1] Uetz, P. et al. (2000) Nature 403, 623.
[2] Ito, T. et al. (2001) Proc. Natl. Acad. Sci. USA 98, 4569.
[3] Gavin, A.C. et al. (2002) Nature 415, 141.
[4] Ho, Y. et al. (2002) Nature 415, 180.
[5] Aloy, P. and Russell, R.B. (2002) Proc. Natl. Acad. Sci. USA 99, 5896.
[6] Maslov, S. and Sneppen, K. (2002) Science 296, 910.
[7] Fields, S. and Song, O. (1989) Nature 340, 245.
[8] Costanzo, M.C. et al. (2000) Nucleic Acids Res. 28, 73.
[9] Jeong, H., Mason, S.P., Barabasi, A.-L. and Oltvai, Z.N. (2001) Nature 411, 41.
[10] Ma, J. and Ptashne, M. (1987) Cell 51, 113.
[11] Housni, H.E., Vandenbroere, I., Perez-Morga, D., Christophe, D. and Pirson, I. (1998) Anal. Biochem. 262, 94.
[12] Wolf, Y.I., Karev, G. and Koonin, E.V. (2002) Bioessays 24, 105.

*Corresponding author. Fax: (49)-6221-387517.
*E-mail address:* russell@embl-heidelberg.de (R.B. Russell).

*EMBL, Meyerhofstrasse 1, D69117 Heidelberg, Germany*

PII: S0014-5793(02)03427-0