A quantitative tri-fluorescent yeast two-hybrid system: from flow cytometry to *in-cellula* affinities

Supplementary Material

David Cluet, Ikram Amri, Blandine Vergier, Jérémie Léault, Astrid Audibert, Clémence Grosjean, Dylan Calabresi, and Martin Spichty[#]

Laboratoire de Biologie et de Modélisation de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, Université Lyon 1, Université de Lyon; 46 allée d'Italie; 69364 Lyon cedex 07; France.

#) corresponding author. Phone: +33 472 72 8645; Email: martin.spichty@ens-lyon.fr.

Running title: **A quantitative tri-fluorescent yeast two-hybrid system**

Computational methods for the alchemical free energy calculations

We used a standard dual topology approach where the two residues 35 and 39 of Barstar were simultaneously transformed from A to D and from D to A, respectively, as a function of the usual coupling parameter λ. The reader is referred to the literature for more details (1).

Software

Model systems were set up with the program CHARMM (2), version c39b1. Initial input files for CHARMM were generated with the CHARMM-GUI server (3) and then modified to implement the actual structural model with dual topology (see below). Molecular dynamics simulations for free energy calculations were carried out with the program NAMD, version 2.11 (4).

Model system

The crystal structure of the complex Barnase:Barstar (5) served as starting point of the structural model. For the alchemical transformation in Barstar alone (see left, vertical leg in Suppl. Fig. S3A) only the coordinates of chain D, residues 1-89 (PDB entry 1BRS), were selected. For the transformation of the complex (right vertical leg) the coordinates of Barnase (chain C, residues 1-110) were specified in addition to those of Barstar. Missing residue coordinates were added with the help of the CHARMM-GUI server. Both systems were solvated in a cubic box (side length of 80 Å) of TIP3 water molecules. With the program CHARMM, residues 35 and 39 were then replaced by residues with dual topology for aspartic acid (D) and alanine (A).

Settings for molecular dynamics

The systems were simulated with the CHARMM36 force field and periodic boundary conditions. A cutoff of 12 Å was used for short range non-bonded interactions whereas long-range electrostatic interactions were treated by Particle-Mesh Ewald (PME) with a grid spacing of 1 Å. The equation of motion was integrated with a time step of 1 fs. Short range non-bonded forces were updated every 2 fs, PME forces every 4 fs. The system was kept at constant pressure (1 atm) and temperature (298.15 K) with a piston oscillation period of 100 fs and a damping time scale of 100 fs. With these MD settings the average box size of both systems (Barstar alone and complexed with Barnase H102A) was almost identical (77.77 ± 0.05 vs. 77.64 ± 0.05 Å) in the following free energy calculations. For systems with identical (cubic) box sizes and no change in net charge (during the alchemical transformation), finite-size effects should largely cancel when comparing the two legs of the alchemical transformations (6–8).

Alchemical transformations

The electrostatic interactions of the outgoing residues (here A35, D39) are linearly removed from λ=0 to λ=0.5 and those of the incoming residues (D35, A39) linearly added from λ=0.5 to λ=1 (NAMD parameter alchElecLambdaStart=0.5). The van-der-Waals interactions are scaled down from λ=0 to λ=1 for the outgoing residues, and scaled up from λ=0 to λ=1 for the incoming residues (alchVdWLambdaEnd=1). To avoid endpoint problems, we used NAMD's soft-core potentials with a shifting coefficient of 4 Å. Non-bonded interactions within outgoing and incoming atoms were also scaled with λ to account for interactions between residues 35 and 39 (alchDecouple=off). The actual transformation was done with a windowing method where the dimer and tetramer systems were sampled by molecular-dynamics at 21 equally-separated λ-values

(from 0 to 1 with an increment parameter of 0.05). For each window we performed two separate MD simulations. Using NAMD's FEP option, we recorded on-the-fly every 1000 fs the work required to switch instantaneously the Hamiltonian from the actual window to one of the neighboring windows. For the endpoints at λ=0 and 1 we performed only one MD simulation (because there is only a single neighbor).

The free energy difference between two neighboring windows was calculated from the collected work data for forward and reverse switches using Bennett's acceptance ratio (BAR) method (9). The total free energy change for the transformation of λ from 0 to 1, (= $\Delta G_{\text{alchemical}}$, see also Suppl. Fig. S8A) was obtained by summing up all differences between neighboring windows. After an equilibration phase of 1 ns (Barstar alone) and 5 ns (complex), we monitored for blocks of 250 ps (Barstar) and 1 ns (complex) the value of $\Delta G_{\text{alchemical}}$ (Suppl. Fig. S8B). The value of $\Delta G_{\text{alchemical}}$ fluctuates around a mean value of -0.17 ± 0.08 kcal/mol in the case of the Barstar alone and 1.71 ± 0.31 kcal/mol in the case of the Barstar-Barnase complex. The errors correspond to twice the standard error of the mean. Finally, the difference between the two vertical legs of the thermodynamic cycle of Suppl. Fig. S3A can be calculated:

$$\Delta G^{\text{Barstar}}_{\text{alchemica l}} - \Delta G^{\text{complex}}_{\text{alchemical}} = -1.88 \pm 0.32 \text{ kcal/mol} \qquad \text{Eq. S1}$$

The usual error propagation rule was used for this subtraction.

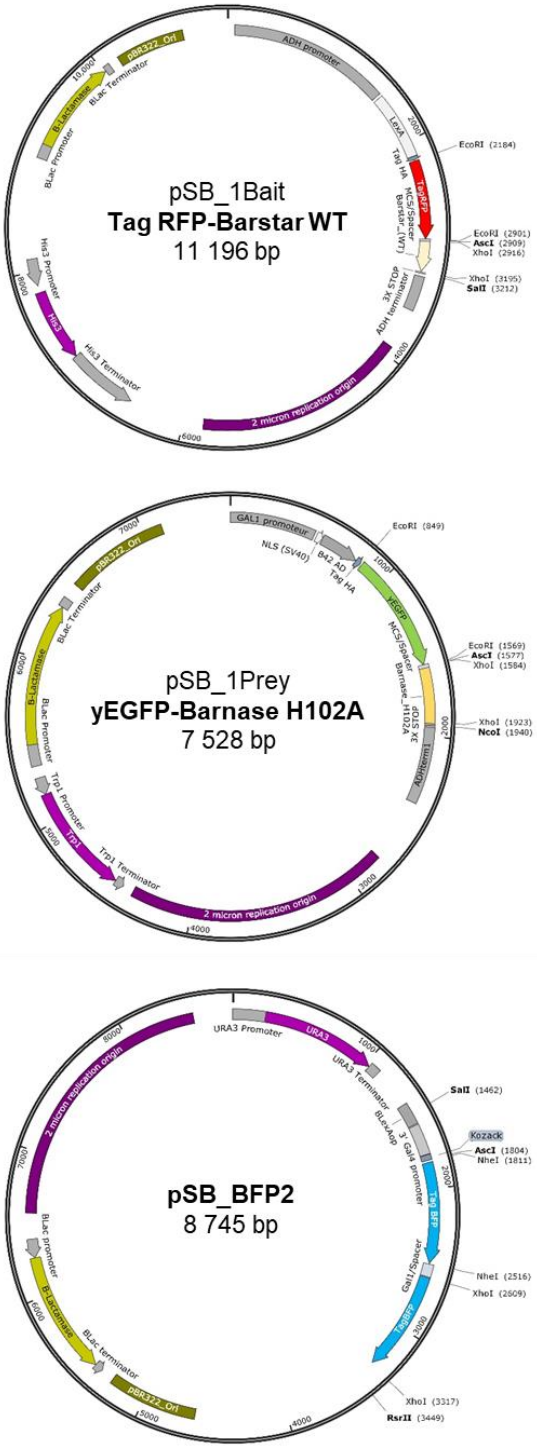
Suppl. Table S1: Primers list.

Coding Sequence	Target plasmid	Orientation	Primer name	Sequence
New expression K7 with Barstar WT	pEG-202	Forward	primSB_0001	GTTGGGGTTATTCGCAACGGCGACT
		Reverse	primSB_0002	GAAATTCGCCCGGAATTAGCTTGGCT
Tag RFP from pTag_RFP-Actin	pSB_1Bait_Barstar	Forward	primSB_0003	CCTTATGATGTGCCAGATTATGCCTCTCCCGAATTCATGG TGTCTAAGGGCGAAGAGCTGATT
		Reverse	primSB_0004	GACTGCTTTTTTCATCTCGAGGGCGCGCCCGAATTCATTA AGTTTGTGCCCCAGTTTGCTAGGG
CDS with bait extensions (Barstar mutants and Ras G12V C186A, CDK2)	pSB_1Bait_RFP	Forward	primSB_0018	GGGCACAAACTTAATGAATTCGGGCG
		Reverse	primSB_0019	AGCTTGGCTGCAGGTCGACTCACT
Barnase H102A	pSB_1Bait_RFP	Forward	primSB_0169	GGGCACAAACTTAATGAATTCGGGCGCGCCCTCGAG
		Reverse	primSB_0170	AGCTTGGCTGCAGGTCGACTCACTCACTTACTCGAGTTAT CTGATTTTTGTAAAGGTCTGATAAGCGTCC
CRaf RBD WT and A85K	pSB_1Bait_RFP	Forward	primSB_0169	GGGCACAAACTTAATGAATTCGGGCGCGCCCTCGAG
		Reverse	primSB_0173	AGCTTGGCTGCAGGTCGACTCACTCACTTACTCGAGCAGG AAATCTACTTGAAGTTCTTCTCCAATC
SRC SH3	pSB_1Bait_RFP	Forward	primSB_0169	GGGCACAAACTTAATGAATTCGGGCGCGCCCTCGAG
		Reverse	primSB_0171	AGCTTGGCTGCAGGTCGACTCACTCACTTACTCGAGGGGC GCCACGTAGTTGCTGG
BLIP1	pSB_1Bait_RFP	Forward	primSB_0169	GGGCACAAACTTAATGAATTCGGGCGCGCCCTCGAG
		Reverse	primSB_0174	AGCTTGGCTGCAGGTCGACTCACTCACTTACTCGAGGACT AAATCCCATTGCCTCTTACCCTG
ARMC1 from cDNA library	pSB_1Bait_RFP	Forward	primSB_0230	GGGCACAAACTTAATGAATTCGGGCGCGCCCTCGAGATGA ATTCTTCCACTTCCACCATGAGTGAAG
		Reverse	primSB_0231	AGCTTGGCTGCAGGTCGACTCACTCACTTACTCGAGTCAC CAATAAAATGATCTGGATAAAAAGTTTGCAGC
Emerin from cDNA library	pSB_1Bait_RFP	Forward	primSB_0232	GGGCACAAACTTAATGAATTCGGGCGCGCCCTCGAGATGG ACAACTACGCAGATCTTTCGGATAC
		Reverse	primSB_0233	AGCTTGGCTGCAGGTCGACTCACTCACTTACTCGAGCTAG AAGGGGTTGCCTTCTTCAGC
CksHs1	pSB_1Bait_RFP	Forward	primSB_0240	GGGCACAAACTTAATGAATTCGGGCGCGCCCTCGAGATGT CGCACAAACAAATTTACTATTCGGACAAATAC
		Reverse	primSB_0241	AGCTTGGCTGCAGGTCGACTCACTCACTTACTCGAGTTTC TTTGGTTTCTTGGGTAGTGGGCG
New expression K7 with Barnase H102A	pJG4-5	Forward	primSB_0010	CCTTATGATGTGCCAGATTATGCCTCTC
		Reverse	primSB_0011	CCAAACCTCTGGCGAAGAAGTCCAAA
yEGFP from pGY-LexA-GFP_KanMX	pSB_1Prey_Barnase H102A	Forward	primSB_0012	CCTTATGATGTGCCAGATTATGCCTCTCCCGAATTCATGA GCAAGGGCGAGGAGCTGTTC
		Reverse	primSB_0013	GTTGATAACCTGTGCCTCGAGGGCGCGCCCGAATTCCTTG TACAGCTCGTCCATGCCGAG
CDS with prey extensions (CRaf RBD WT and A85K, CksHs1)	pSB_1Prey_yEGFP	Forward	primSB_0020	GACGAGCTGTACAAGGAATTCGGG
		Reverse	primSB_0021	AAGTCCAAAGCTTCCATGGTCACTCACTTA
Barstar WT and mutants	pSB_1Prey_yEGFP	Forward	primSB_0165	GACGAGCTGTACAAGGAATTCGGGCGCGCCCTCGAG
		Reverse	primSB_0164	AAGTCCAAAGCTTCCATGGTCACTCACTTACTCGAGTTAA GAAAGTATGATGGTGATGTCGCAGCC
Ras G12V C186A	pSB_1Prey_yEGFP	Forward	primSB_0165	GACGAGCTGTACAAGGAATTCGGGCGCGCCCTCGAG
		Reverse	primSB_0166	AAGTCCAAAGCTTCCATGGTCACTCACTTACTCGAGTCAG GAGAGCACTGCCTTGCAGC
Nef LAI	pSB_1Prey_yEGFP	Forward	primSB_0165	GACGAGCTGTACAAGGAATTCGGGCGCGCCCTCGAG
		Reverse	primSB_0167	AAGTCCAAAGCTTCCATGGTCACTCACTTACTCGAGGCAG TTCTTGAAGTACTCCGGATGC
TEM	pSB_1Prey_yEGFP	Forward	primSB_0165	GACGAGCTGTACAAGGAATTCGGGCGCGCCCTCGAG
		Reverse	primSB_0168	AAGTCCAAAGCTTCCATGGTCACTCACTTACTCGAGCCAA TGCTTAATCAGTGAGGCACCTATC
ARMC1 from cDNA library	pSB_1Prey_yEGFP	Forward	primSB_0234	GACGAGCTGTACAAGGAATTCGGGCGCGCCCTCGAGATGA ATTCTTCCACTTCCACCATGAGTGAAG
		Reverse	primSB_0235	AAGTCCAAAGCTTCCATGGTCACTCACTTACTCGAGTCAC CAATAAAATGATCTGGATAAAAAGTTTGCAGC
Emerin from cDNA library	pSB_1Prey_yEGFP	Forward	primSB_0236	GACGAGCTGTACAAGGAATTCGGGCGCGCCCTCGAGATGG ACAACTACGCAGATCTTTCGGATAC
		Reverse	primSB_0237	AAGTCCAAAGCTTCCATGGTCACTCACTTACTCGAGCTAG AAGGGGTTGCCTTCTTCAGC
CDK2	pSB_1Prey_yEGFP	Forward	primSB_0238	GACGAGCTGTACAAGGAATTCGGGCGCGCCCTCGAGATGG AGAACTTCCAAAAGGTGGAAAAGATCG
		Reverse	primSB_0239	AAGTCCAAAGCTTCCATGGTCACTCACTTACTCGAGGAGT CGAAGATGGGGTACTGGCTT
Gal1 promoter delta Gal4 with 8 operator LexA and the Kozack sequence with a new downstream MCS	pSH18-34	Forward	primSB_0076	GGACGCAAAGAAGTTTAATAATCATATTACATGGC
		Reverse	primSB_0077	GAAAAAACTATAATGACTAAATCTCATTCAGAAGAAGTGG GGCGCGCCGCTAGC
Gal1 Nterm sequence	pSH18-34	Forward	primSB_0078	CATTCAGAAGAAGTGGGGCGCGCCGCTAGCATTGTACCTG AGTTCAATTCTAGCGCAAAGG
		Reverse	primSB_0079	CCAAGCTTGGCCAAGCCCGGACTCGAG
Tag-BFP from pTag_BFP-Actin	pSH18-34	Forward	primSB_0084	ATTCCAAGCTTGGCCAAGCCCGGACTCGAGATGAGCGAGC TGATTAAGGAGAACATGC
		Reverse	primSB_0085	CCTAGCAAACTGGGGCACAAGCTTAATTAACTCGAGTAAT AACCGGGCAGGCCATGTCTG
Gal1 terminator sequence	pSH18-34	Forward	primSB_0080	TAACTCGAGTAATAACCGGGCAGGCCATGTCTG
		Reverse	primSB_0081	TAATAAAACGCCCGTTCCCGGACG
Tag-BFP from pTag_BFP-Actin	pSB_3RO	Forward	primSB_0120	AAATCTCATTCAGAAGAAGTGGGGCGCGCCGCTAGCATGA GCGAGCTGATTAAGGAGAACATGC
		Reverse	primSB_0121	CTTTGCGCTAGAATTGAACTCAGGTACAATGCTAGCATTA AGCTTGTGCCCCAGTTTGCTAGG

Suppl. Table S2: Macsquant VYB settings for qY2H fluorescence acquisition.

Channel	Setting
FSC	229V
SSC	265V
V1	264V
B1	327V
Y1	506V

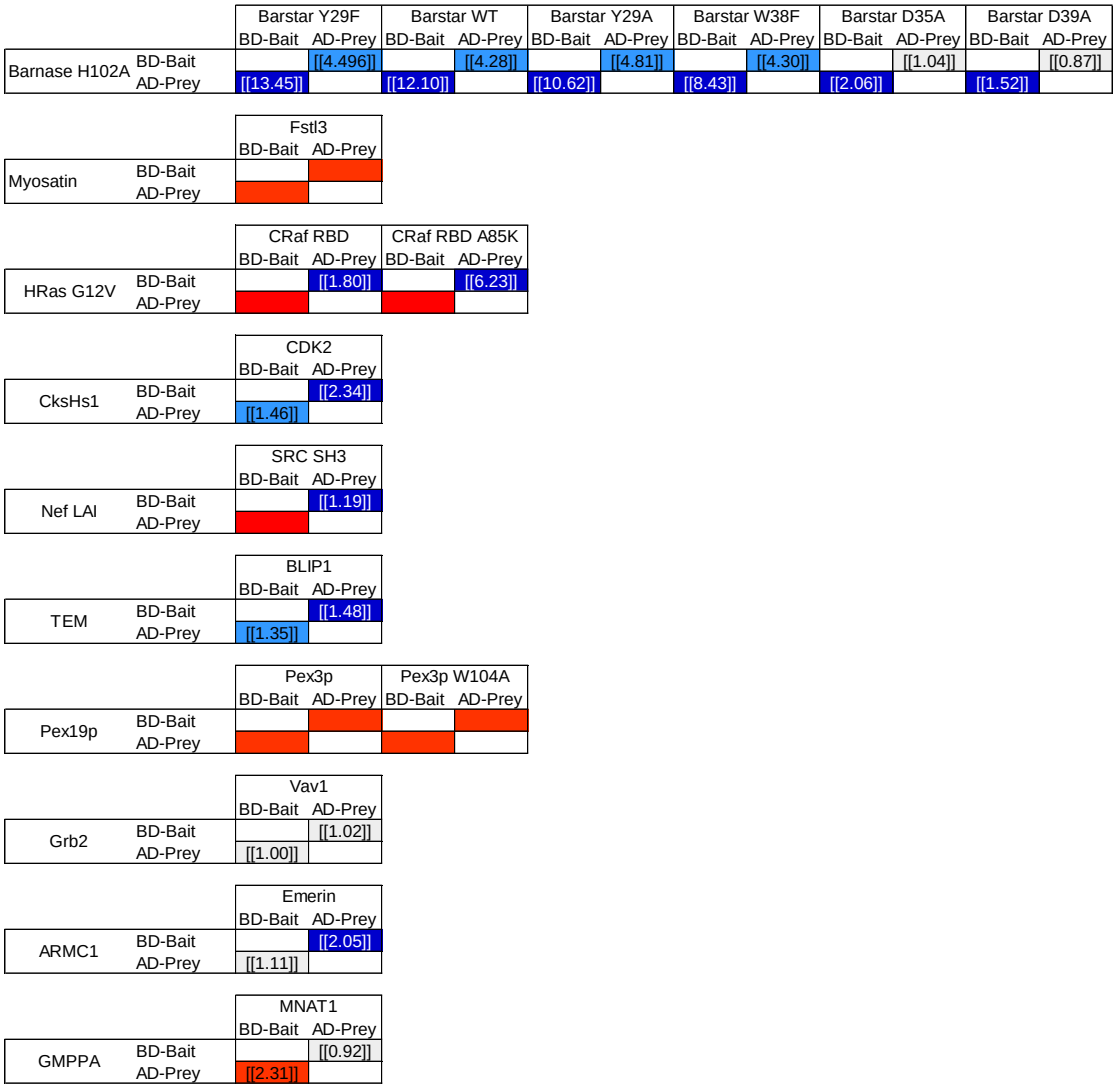
Parameter	Setting
Mixing mode	Strong
Flow speed	Medium



85

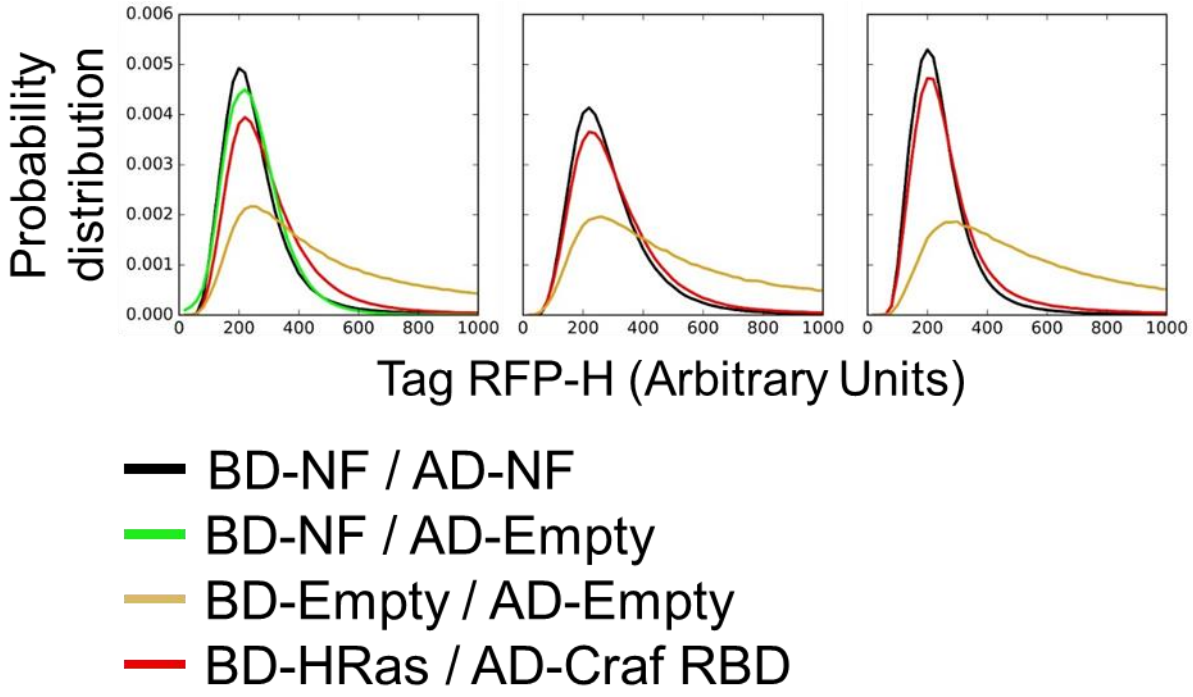
86 **Suppl. Fig. S1.** qY2H plasmid maps.

87



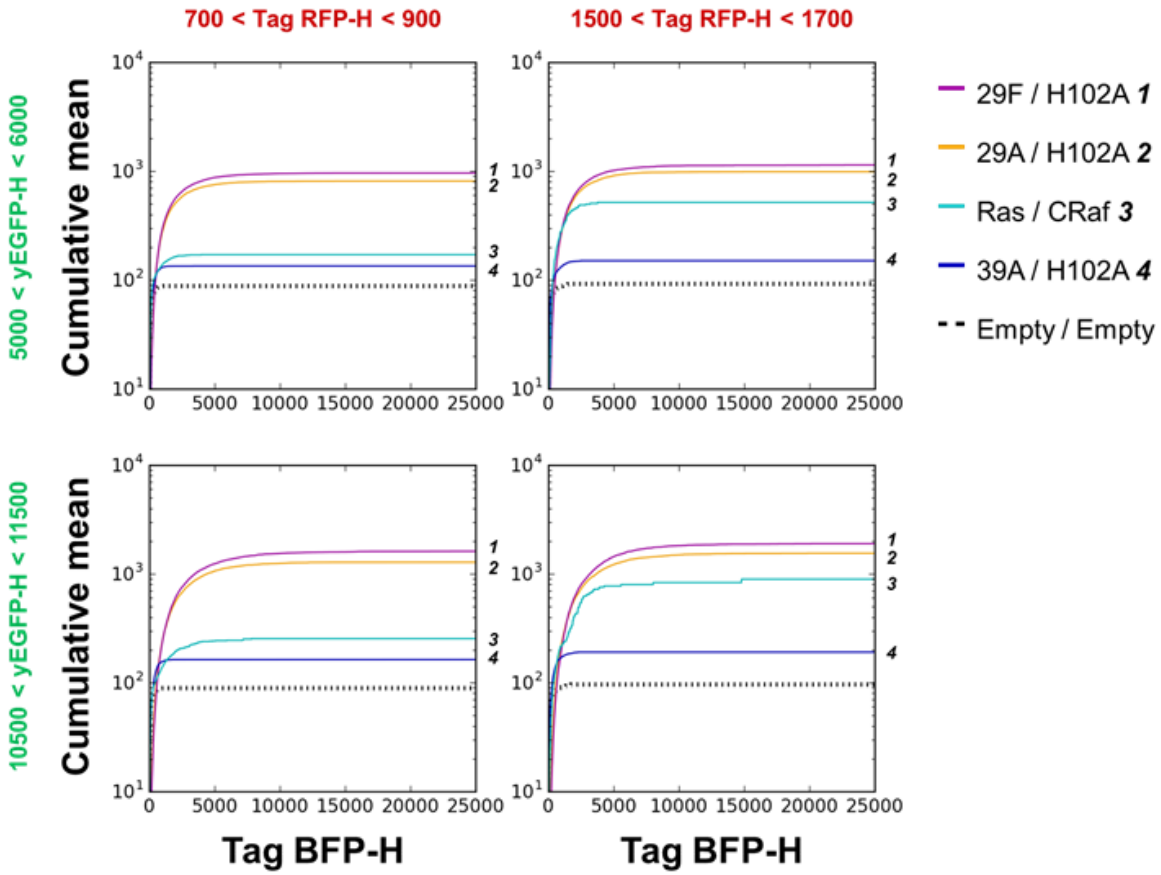
Suppl. Fig. S2.

Determination of the optimal orientation for PPI couples of Table 1. Since certain proteins showed a significant auto-activation level when used as BD-Bait fusion (see Table 2), some of the orientations could not be tested (red-colored fields). For the remaining orientations, we determined $\langle \text{Tag BFP-H} \rangle_{\text{sample}}$ / $\langle \text{Tag BFP-H} \rangle_{\text{CTRL}}$, *i.e.*, the mean reporter level of a given sample relative to BD-Empty / AD-Empty control sample; the relative reporter level was calculated either for the entire population of cells or for a double-gated subpopulation (indicated by [[…]], see "Experimental Procedures", subsection "Data analyses"). When an orientation gave a relative reporter level larger than two times the relative standard deviation of the BD-Empty / AD-Empty control (0.19), it is represented in blue. If both orientations fulfill this criterion; the orientation with the larger reporter level is indicated in dark blue.

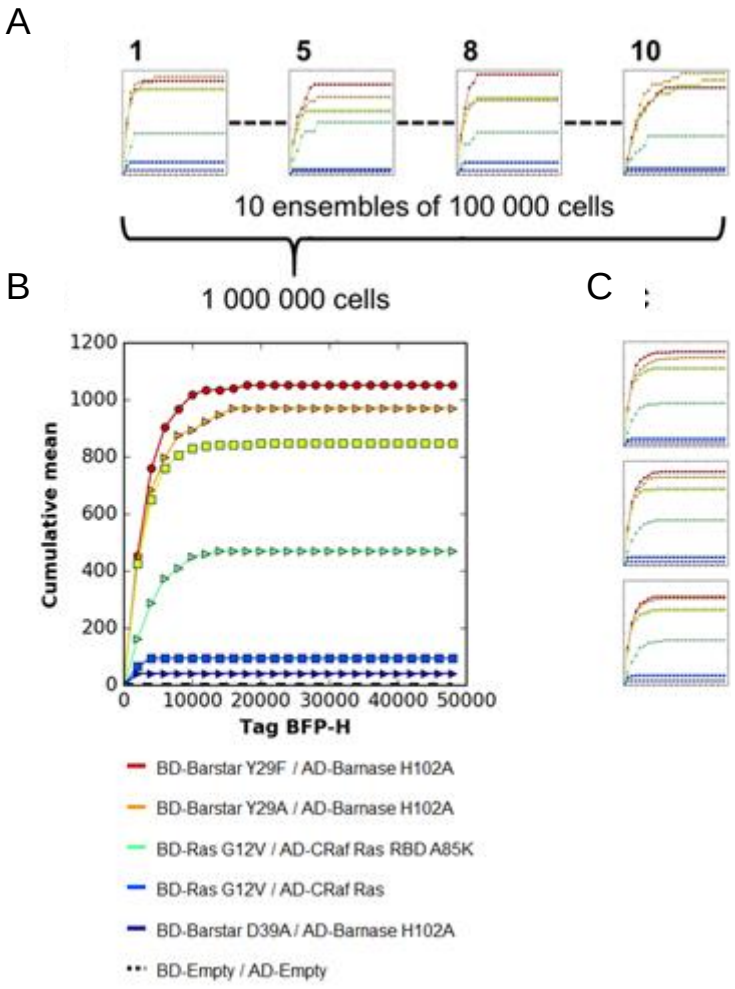


Suppl. Fig. S3.

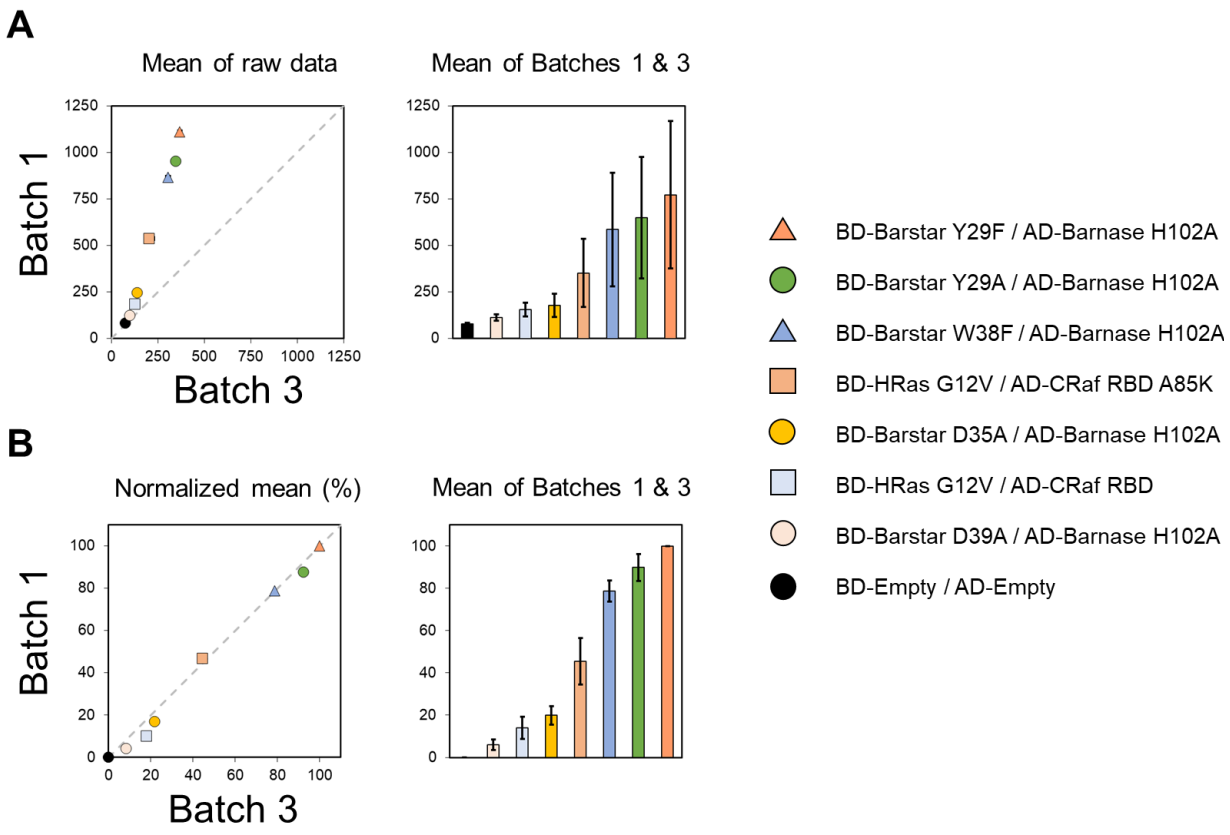
Detection of BD-HRas expression by flow-cytometry. BD-HRas is not detectable by western blot (see Fig. 3). But a weak expression level can be observed by flow-cytometry. The linear Tag-RFP-H signals were represented as probability distributions for the BD-HRas / AD-CRaf couple (Red). The non-fluorescent control, BD-NF / AD-NF, is presented in Black. In order to detect compensation issues between the yEGFP and Tag-RFP channels we added the BD-NF / AD-Empty couple that displays the strongest yEGFP-H signals. In our acquisition conditions the two negative controls have comparable distributions, when BD-HRas generates a weak but reproducible increase in the distribution (three independent experiments are presented). For comparison the BD-Empty / AD-Empty couple (orange), with an optimal Bait expression level is displayed.



Suppl. Fig. S4.
Impact of the gating region on the qY2H affinity ladder. The same double gating approach as in Fig. 4 was used with two different intervals for each channel: 700-900 and 1500-1700 for Tag RFP-H, and 5000-6000 and 10500-11500 for yEGFP-H, respectively. The four different combinations do not affect the ordering of the couples. Only their relative positions vary from one gates combination to another. The analysis was performed with samples of ten millions cells.

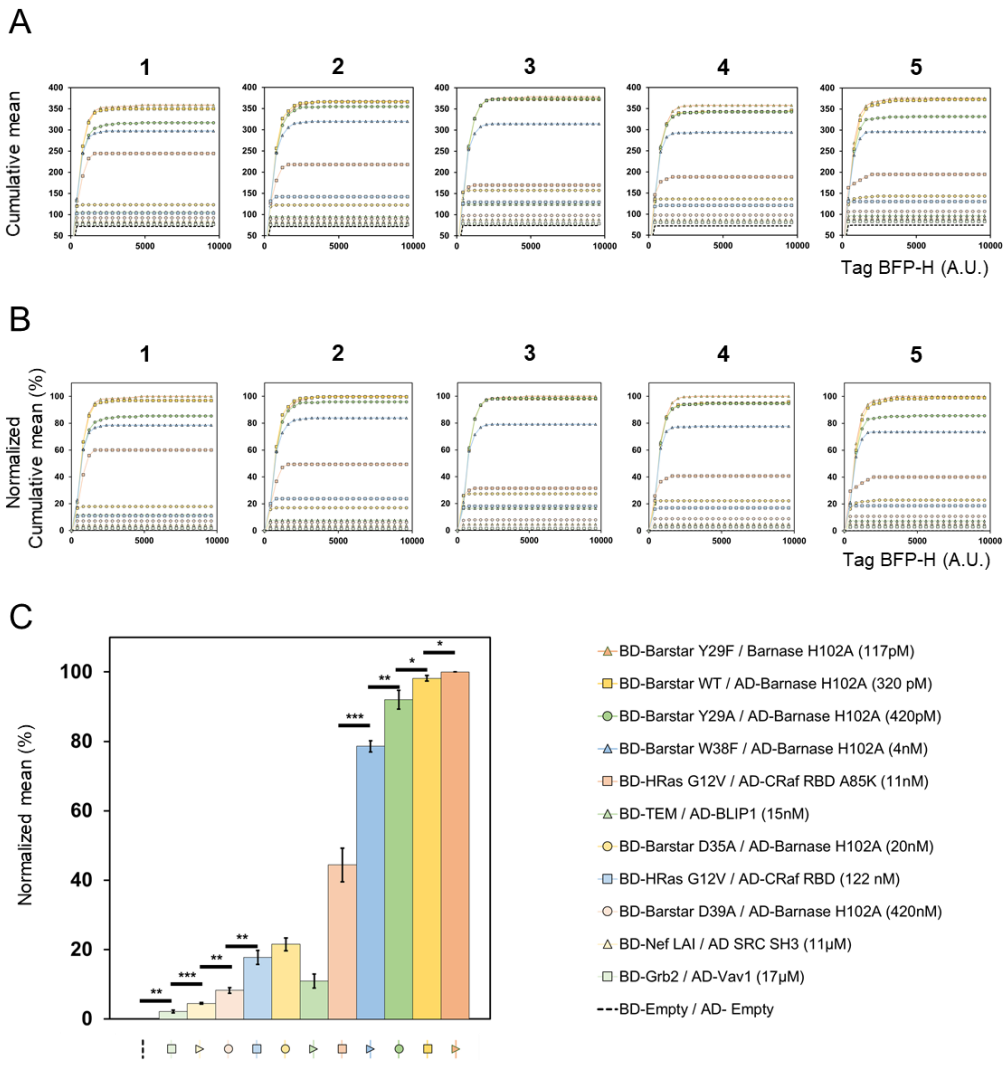


Suppl. Fig. S5.
Impact of the number of cells on the qY2H affinity ladder. *A*. Ten successive sub-ensembles of 100 000 cells from one single experiment were used to perform a qY2H affinity ladder analysis. Background signal (BD-Empty / AD-Empty) was removed for all samples. Some sub-ensembles (*e.g.*, **5**) lead to a correct order of the PPIs according to their affinities, but several sub-ensembles gave wrong results (**1**, **8**, **10**). *B*. When the ten sub-ensembles were combined to a single ensemble of one million cells, a correct affinity ladder was obtained. *C*. The affinity ladders obtained from three subsequent ensembles of one million cells are presented. They all show the same correct order as in Fig. 4.



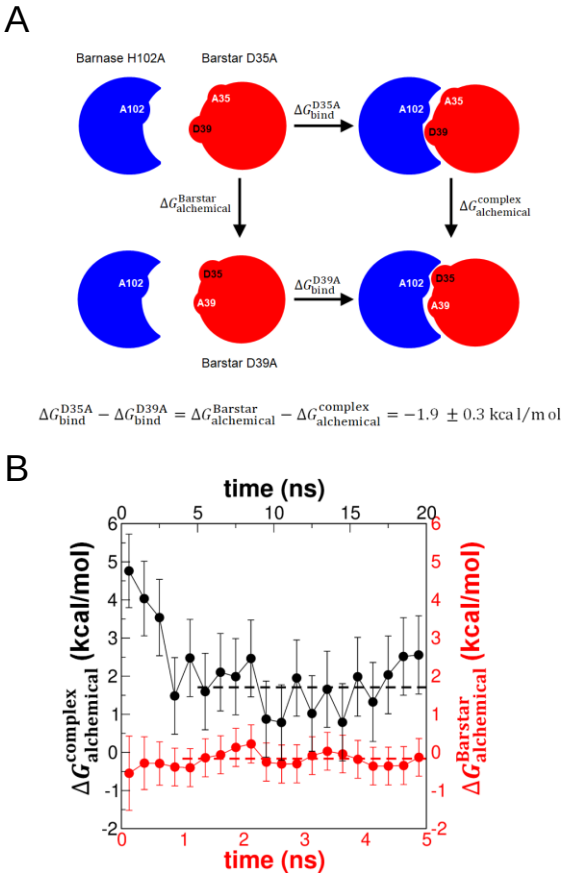
Suppl. Fig. S6

Reporter level for different batches of chemo-competent Yeasts. Several repetitions of the qY2H experiments were performed with two different batches of competent yeasts (1 and 3 of Supp. Data 1). The reporter level <TagBFP-H> of seven different couples of Table 1 was determined for each repetition by analyzing one million cells. For each couple, the average of <TagBFP-H> was calculated for the five and six repetitions originating from batch 1 and 3, respectively. *A.* The correlation of the averages between batches 1 and 3 is shown for the raw data on the left side. The curve corresponding to a perfect correlation is presented as a dashed line. The quasi-linear correlation for the seven different couples implies that a similar scaling factor applies to all repetitions originating from the same batch of yeasts. When combining Batches 1 & 3, the mean values and sample standard deviations on the right side are obtained *B.* Once the reporter levels are normalized (as explained in "Experimental procedures", subsection "Statistical analyses"), the scaling factor is largely attenuated (left side) and the (absolute and relative) sample standard deviations are reduced significantly for most couples with respect to the raw data (right side).

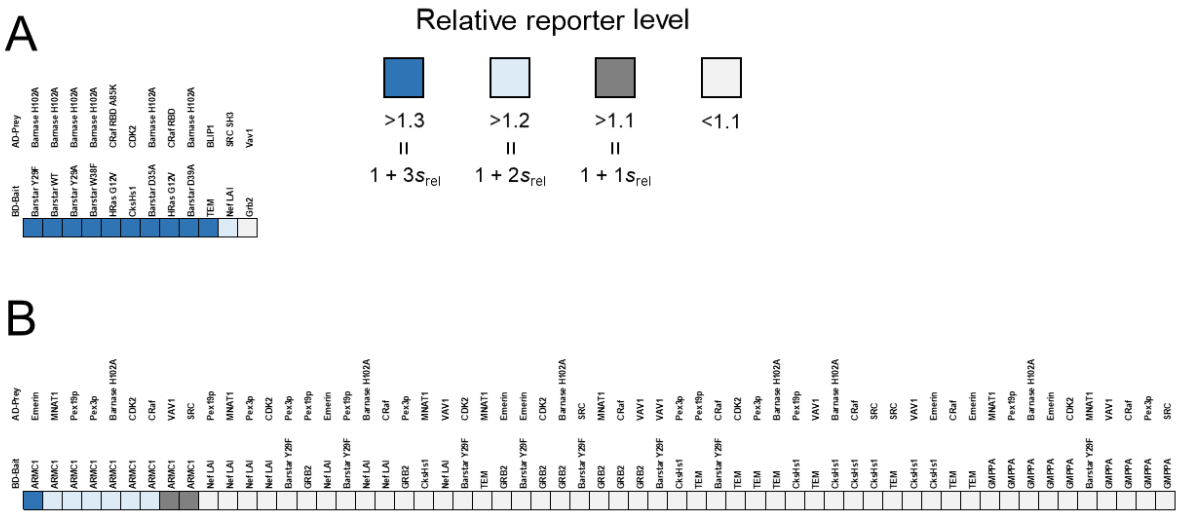


Suppl. Fig. S7

Reproducibility of the affinity ladder and its normalization. Five repetitions of the qY2H experiment were performed with the same batch of competent yeasts. For all samples, one million of cells were analyzed; the same gates as in Fig. 5 (main text) were applied. *A.* The plotting of the cumulative means (with raw data) for each repetition yielded five similar affinity ladders where the couples are ordered according to their affinity (Table 1). An exception is the couple BD-TEM / AD-BLIP1 that is ranked too low in all repetitions. Also, the couple BD-HRas G12V / AD-CRaf is ranked too high in experiment 2. Note that in experiment 4, no yeast cells were detected for BD-TEM / AD-BLIP1 in the double gated region. *B.* Affinity ladders are shown for normalized cumulative means. First, the cumulative mean of the BD-Empty / AD-Empty control sample (*i.e.*, background of the qY2H system) is removed from the cumulative mean of all samples. Second, the background-corrected cumulative means are normalized with the corresponding value of the couple BD-Barstar Y29F / AD-Barnase H102A (K_d= 117 pM). *C.* The normalized means were used to calculate the average value of the five repetitions. Error bars correspond to the standard error of the mean. The statistical significance of the difference between direct neighbors of the ladder is indicated (* : p-value < 0.05, ** : p-value < 0.01, *** : p-value < 0.001).

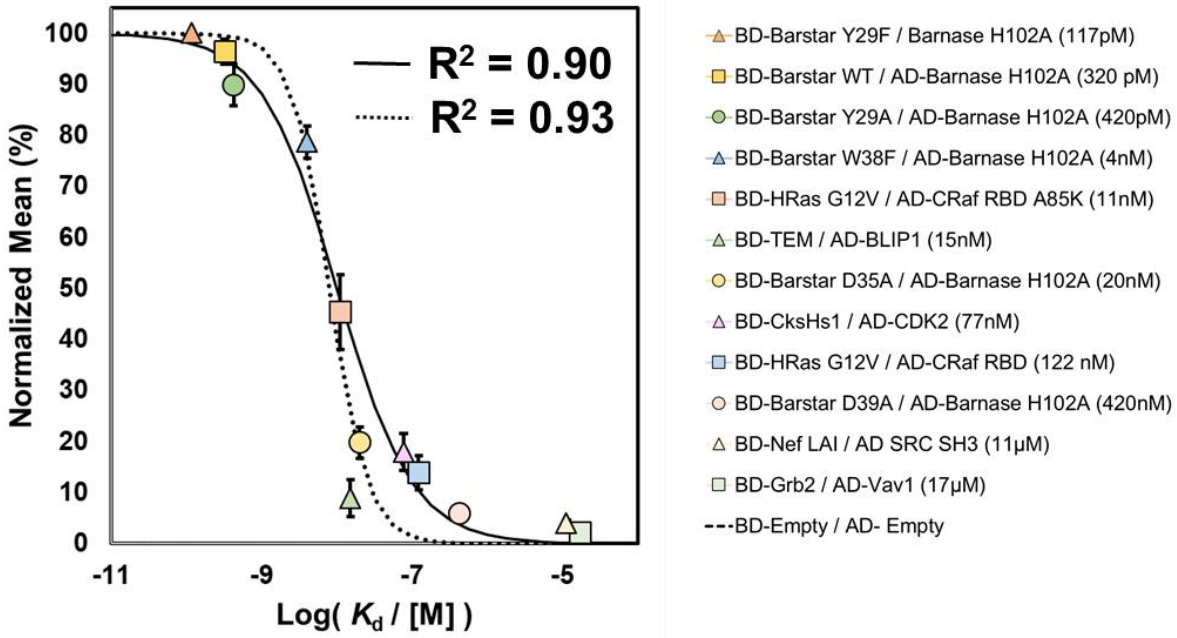


Suppl. Fig. S8.
Alchemical free energy calculations. *A.* A thermodynamic cycle is applied to calculate the difference in binding free energy for the interaction between Barnase H102A and the mutants Barstar D35A (horizontal leg, top) and D39A (horizontal leg, bottom). Because the free energy is a state function, this difference can also be obtained from the free energy difference of the corresponding alchemical transformation in Barstar alone (vertical leg, left) and the complex (vertical leg, right leg). *B.* In the block analysis the change in free energy for the alchemical transformations is plotted for consecutive blocks of 250 ps (Bartsar alone, red dots) and 1 ns (complex, black dots) of sampling. The error bars correspond to the analytical error of the maximum likelihood estimate (10). The mean value of $\Delta G_{alchemical}$ for each alchemical transformation is indicated as horizontal dashed line.



Suppl. Fig. S9.

Determination of the sensitivity and specificity of the qY2H. We determined <Tag BFP-H>sample / <Tag BFP-H>CTRL, *i.e.*, the mean reporter level of a given sample relative to BD-Empty / AD-Empty control sample; the relative reporter level was calculated for a double-gated subpopulation (see "Experimental Procedures", subsection "Data analyses"). When a couple gave a relative reporter above the margin of error of 0.19 (= 2 s_{rel} , *i.e.*, two times the relative sample standard deviation of the CTRL sample BD-Empty / AD-Empty), it is represented in blue. *A*. The subfigure presents the results obtained for twelve interactions of known affinity (Table 1). Since Pex3p and Pex19p proteins showed both a significant auto-activation level when used as BD-Bait fusion (see Table 2), the two couples Pex3p WT / Pex19p and Pex3p W104A could not be tested. The couples are ranked according to their relative reporter level. Except for the BD-Grb2 / AD-Vav1 couple, all interactions generate a reporter level that can be distinguished from the background of the system. *B*. The same analysis was performed with the couples of the Specificity Test Set. Out of the 57 couples, seven are above the margin of error, one even above 3 s_{rel}. These seven couples have the same BD-Bait fusion (BD-ARMC1) in common. The 52 other couples (91% of this set) display a relative reporter level within 1 s_{rel} of the CTRL sample, and cannot be distinguished from the system's background.



Suppl. Fig. S10.
Correlation between *in-vitro* measured K_d-values and the normalized mean of the qY2H reporter level. Correlation between *in-vitro* measured K_d-values and the normalized mean of the qY2H reporter level. The relationship follows a classical dose-response curve that can be well fitted with a sigmoid-like generalized logistic function (Richard's curve):

$$\text{Normalized Mean (NM)} = \frac{100}{1 + e^{\alpha\,(\log(K_d/[\text{M}]) + \beta)}} \qquad \text{Eq. S2}$$

with α and β two parameters to be fitted, and log the common logarithm with basis 10. When all data points (except BD-Barstar D35A / AD-Barnase H102A) are used for the least-square fitting we obtain α = 3.9 and β = 8.1 (dashed curve). Eliminating the outlier BD-TEML / AD-BLIP1, yields α = 2.0 and β = 8.0 (solid curve). Adjusted R^2 values (see section "Statistical analyses") were calculated for both parameter sets (with n=12 and k=2). Given the normalized reporter level, the dissociation constant can be determined with the inverse function:

$$\log(K_d/[\text{M}]) = \frac{\ln\left[\frac{100}{\text{NM}} - 1\right]}{\propto} - \beta \qquad \text{Eq. S3}$$

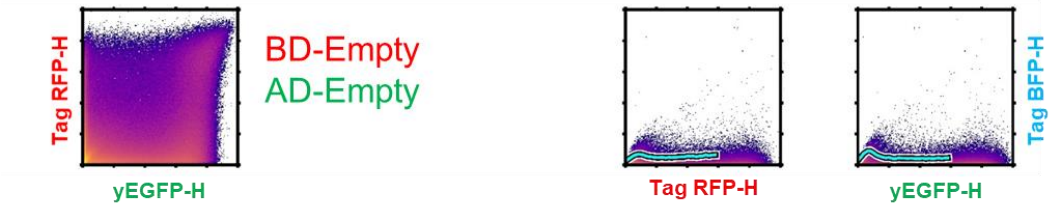
With Eq. S3 the dissociation constant of the BD-Barstar D35A / AD-Barnase H102A can be estimated to be within 18 nM (α = 3.9, β = 8.1) and 49 nM (α = 2.0, β = 8.0), in excellent agreement with independent free-energy calculations (20 nM).

Suppl. Data1.

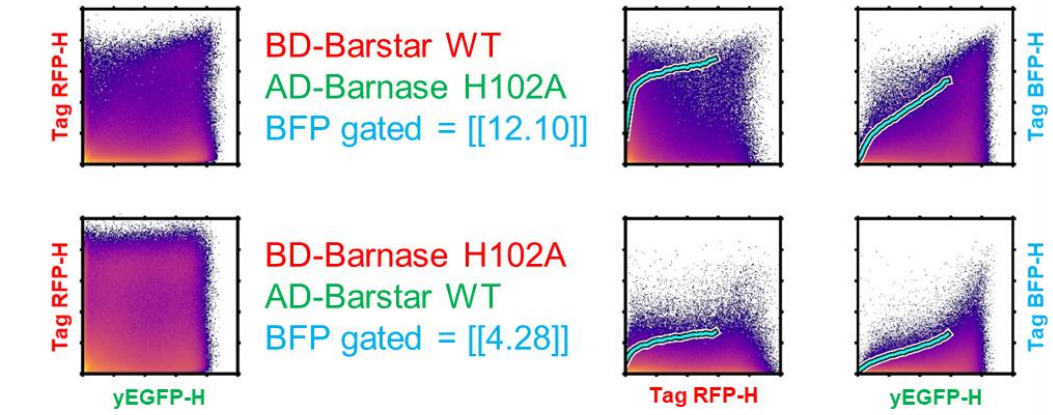
Determination of the optimal orientation of each couple. For the determination of the optimal orientation for the couples of Table 1, several batches of competent EYG42A (pBFP+) competent yeasts were used. For each batch, the corresponding BD-Empty / AD-Empty control is presented. When possible (*i.e.* no prior detection of a significant auto-activation level) both orientation were studied. One million cells were acquired for each analysis. The three channels Tag RFP-H, yEGFP-H and Tag BFP-H are displayed on the hlog-scale (see "Experimental procedures", subsection "Data analyses" for the exact settings). Patterns of correlation between Tag BFP-H and Tag RFP-H and between Tag-BFP-H and yEGFP-H are indicated by cyan curves. These curves were obtained by discretizing the cells into 30 bins based on their TagRFP-H or yEGFP-H values. For each bin the mean value of the TagBFP-H values was determined for the top 5%. In addition, the relative reporter level $<\text{Tag BFP-H}>_{\text{sample}}$ / $<\text{Tag BFP-H}>_{\text{CTRL}}$ for standardized expression levels of BD-Bait and AD-Prey is indicated in double square brackets. Standardization is achieved by applying two gates: 700 < linear Tag RFP-H <900 and 5000 < linear yEGFP-H 6000.

YEAST BATCH 1

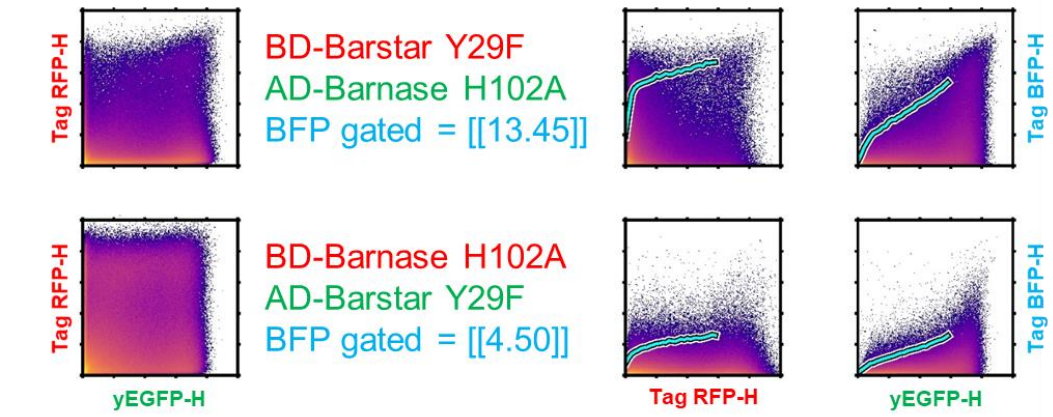
1.00 Empty / Empty



1.01 Barstar WT / Barnase H102A

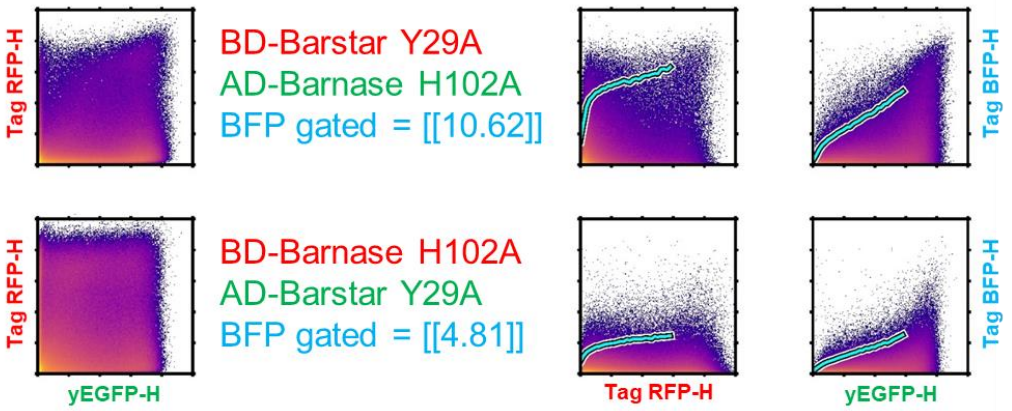


1.02 Barstar Y29F / Barnase H102A

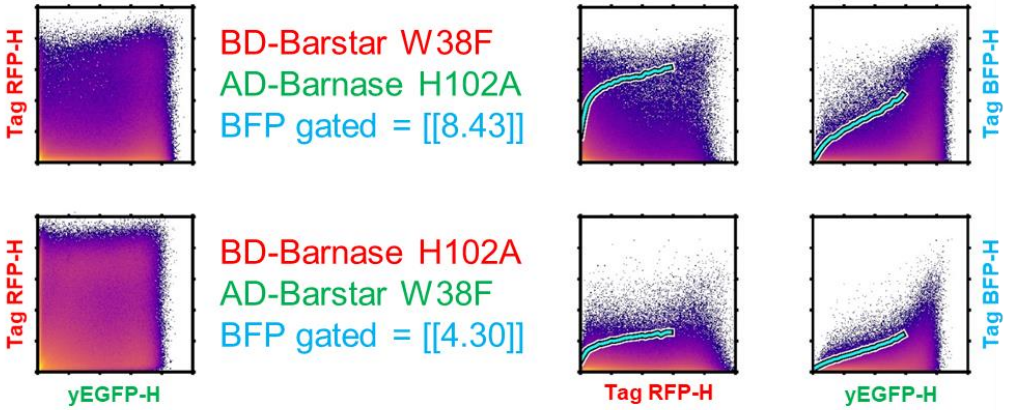


228

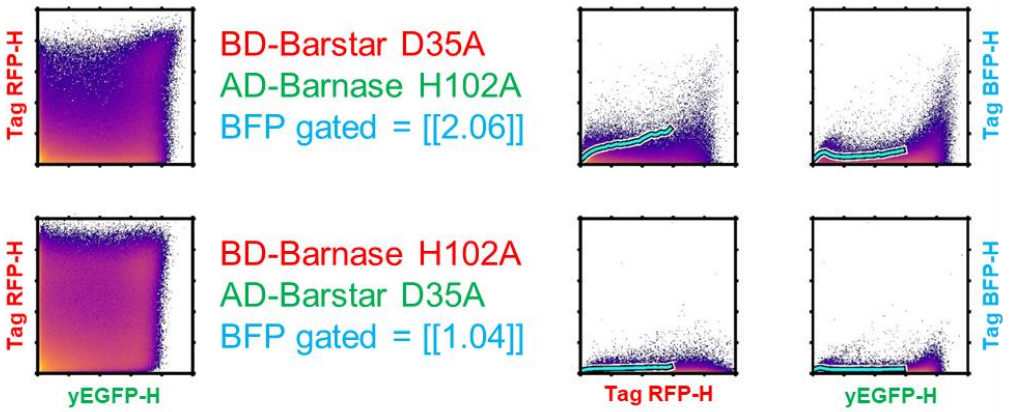
1.03 Barstar Y29A / Barnase H102A



1.04 Barstar W38F / Barnase H102A



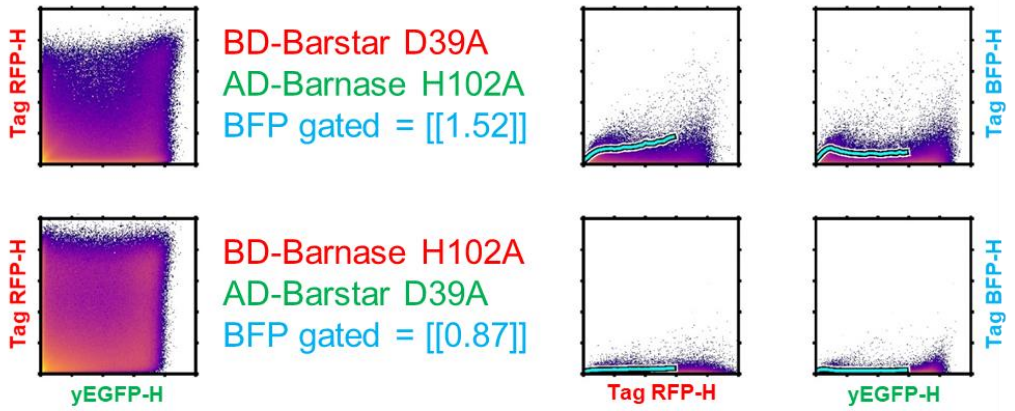
1.05 Barstar D35A / Barnase H102A



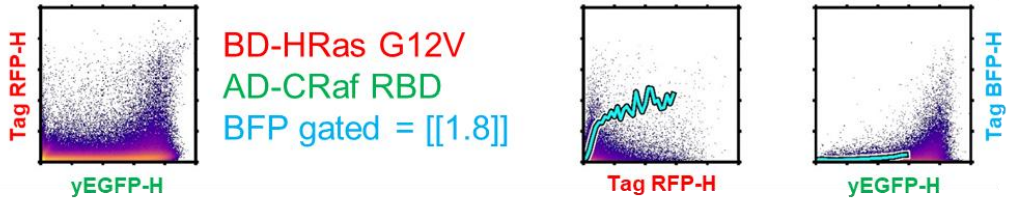
229

230

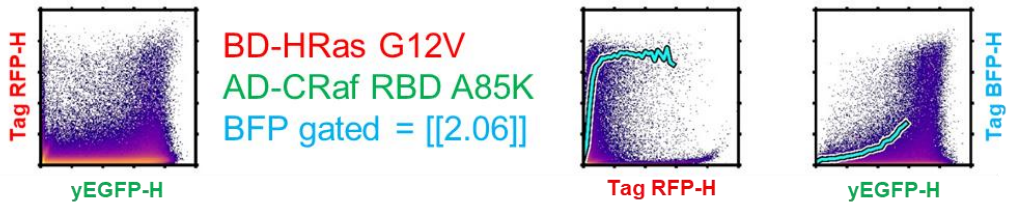
1.06 Barstar D39A / Barnase H102A



1.07 HRas G12V C186A / CRaf RBD



1.08 HRas G12V C186A / CRaf RBD A85K



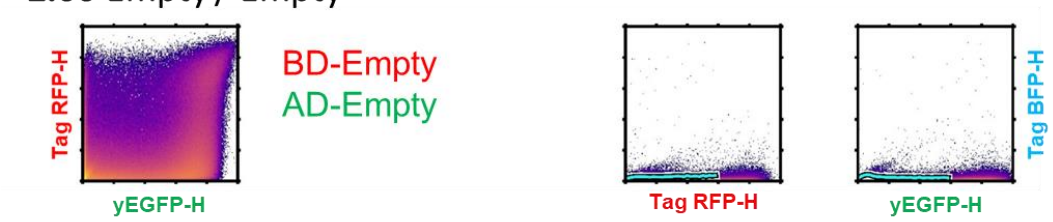
231

232

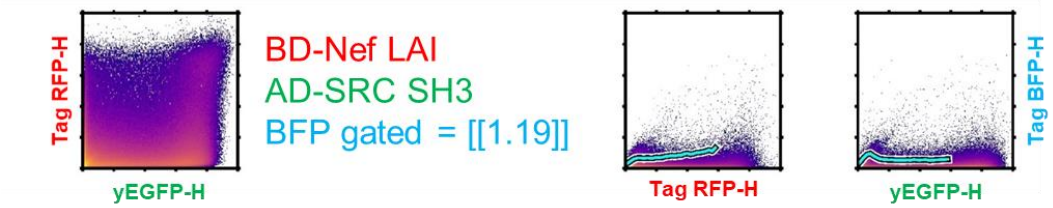
233

YEAST BATCH 2

2.00 Empty / Empty



2.01 Nef LAI / SRC SH3

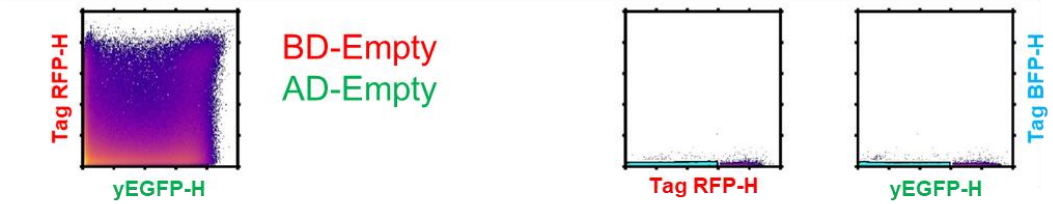


234

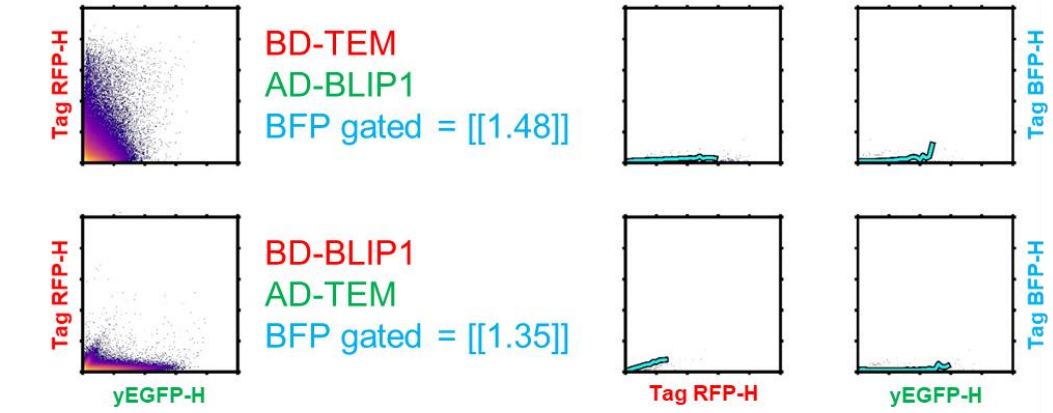
235

YEAST BATCH 3

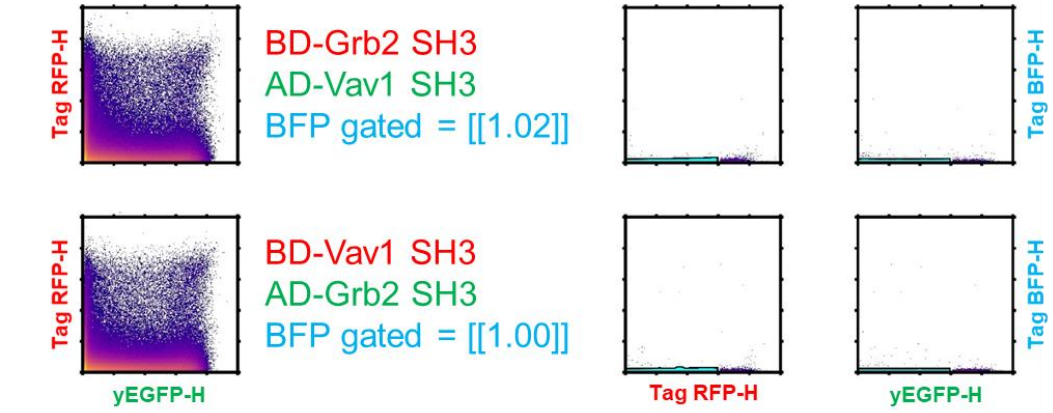
3.00 Empty / Empty



3.01 TEM / BLIP1

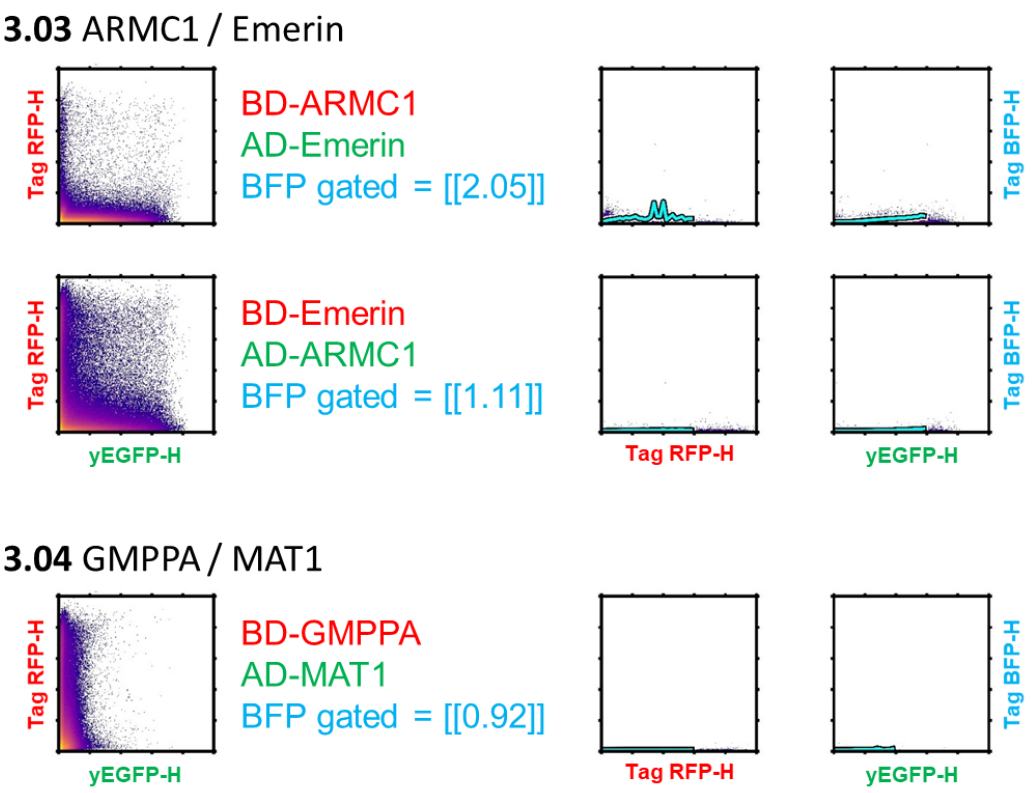


3.02 Grb2 SH3 / Vav1 SH3



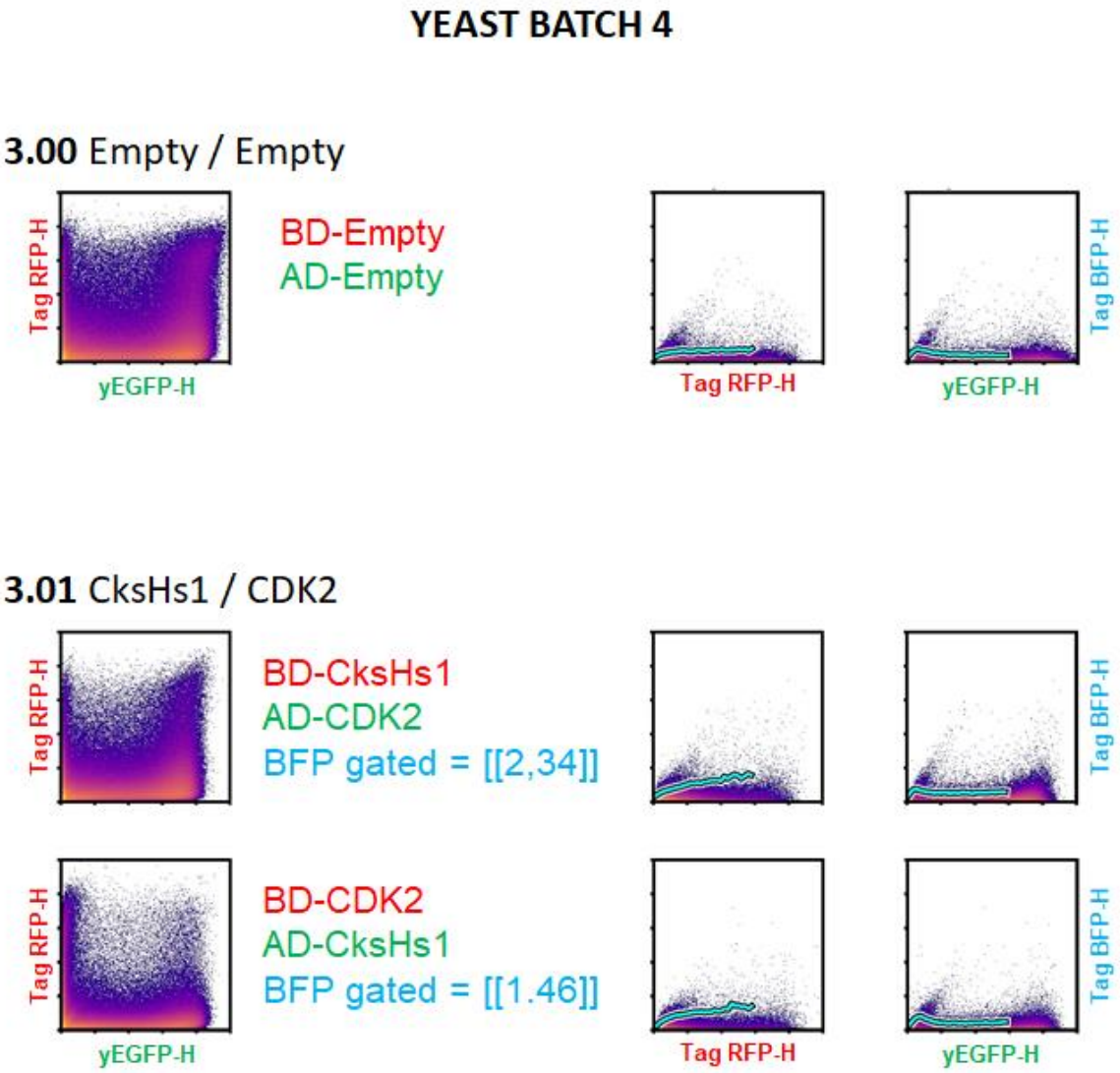
236

237


3.03 ARMC1 / Emerin
Tag RFP-H
BD-ARMC1
AD-Emerin
BFP gated = [[2.05]]
Tag BFP-H
Tag RFP-H
BD-Emerin
AD-ARMC1
BFP gated = [[1.11]]
Tag BFP-H
yEGFP-H
Tag RFP-H
yEGFP-H
3.04 GMPPA / MAT1
Tag RFP-H
BD-GMPPA
AD-MAT1
BFP gated = [[0.92]]
Tag BFP-H
yEGFP-H
Tag RFP-H
yEGFP-H

238

239


YEAST BATCH 4
3.00 Empty / Empty
Tag RFP-H
yEGFP-H
BD-Empty
AD-Empty
Tag RFP-H
Tag BFP-H
yEGFP-H
3.01 CksHs1 / CDK2
Tag RFP-H
BD-CksHs1
AD-CDK2
BFP gated = [[2,34]]
Tag BFP-H
Tag RFP-H
yEGFP-H
BD-CDK2
AD-CksHs1
BFP gated = [[1.46]]
Tag RFP-H
Tag BFP-H
yEGFP-H

240

References

1. Chipot C, Pohorille A eds. (2007) *Free energy calculations: theory and applications in chemistry and biology* (Springer, Berlin ; New York).

2. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. *J Comput Chem* 30(10):1545–1614.

3. Lee J, et al. (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* 12(1):405–413.

4. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802.

5. Buckle AM, Schreiber G, Fersht AR (1994) Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-.ANG. resolution. *Biochemistry (Mosc)* 33(30):8878–8889.

6. Hummer G, Pratt LR, García AE (1996) Free Energy of Ionic Hydration. *J Phys Chem* 100(4):1206–1215.

7. Figueirido F, Del Buono GS, Levy RM (1997) On Finite-Size Corrections to the Free Energy of Ionic Hydration. *J Phys Chem B* 101(29):5622–5623.

8. Rocklin GJ, Mobley DL, Dill KA, Hünenberger PH (2013) Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum

methods: An accurate correction scheme for electrostatic finite-size effects. *J Chem Phys* 139(18):184103.

9. Bennett CH (1976) Efficient estimation of free energy differences from Monte Carlo data. *J Comput Phys* 22(2):245–268.

10. Shirts MR, Bair E, Hooker G, Pande VS (2003) Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys Rev Lett* 91(14). doi:10.1103/PhysRevLett.91.140601.

Analysis of QY2H Data

https://github.com/dcluet/qY2H-Affinity-Ladder

Introduction

This program permits the automated generation of affinity ladders from `quantitative Yeast Two Hybrid` experiments. The program requires flow cytometry data `.fcs` files (linear scale). It generates a `.csv` table file that contains for each sample the mean reporter level. The actual affinity ladder graph is reported into a `.pdf` file. For more information, the reader is referred to our article:

A quantitative tri-fluorescent yeast two-hybrid system: from flow cytometry to in-cellula affinities

https://www.biorxiv.org/content/10.1101/553636v1

Authors

	
CLUET David	david.cluet@ens-lyon.fr
SPICHTY Martin	martin.spichty@ens-lyon.fr

License

This software is a computer program whose purpose is to automatically analyze QY2H data (`.fcs` files) and generate *in cellulo* affinity ladder.

In this respect, the user’s attention is drawn to the risks associated with loading, using, modifying and/or developing or reproducing the software by the user in light of its specific status of free software, that may mean that it is complicated to manipulate, and that also therefore means that it is reserved for developers and experienced professionals having in-depth computer knowledge. Users are therefore encouraged to load and test the software’s suitability as regards their requirements in conditions enabling the security of their systems and/or data to be ensured and, more generally, to use and operate it in the same conditions as regards security.

The fact that you are presently reading this means that– delete the_remote_branch you have had knowledge of the CeCILL license and that you accept its terms.

Requirements

This program is optimized for `Python 2.7` with the following libraries:

- `datetime`: To generate unique Analysis ID and file name.
- `FlowCytometryTools` **v 0.4.6**: To open `.fcs` files and manipulate flowcytometry data. http://eyurtsev.github.io/FlowCytometryTools/
- `glob`: To identify the `.fcs` files in the Input folder.
- `matplotlib` **v 1.5.1**: To generate the curves.
- `numpy` **v 1.13.3**: To generate and manipulate arrays.
- `os`: To handle paths of the raw data and generated files.
- `Pillow / PIL` **v 3.1.2**: To display images within the GUI of the program.
- `sys`: To permit manual abortion of the program.
- `Tkinter` **v 8.6**: To generate the GUI of the program.

Files

- README.md
- LICENSE.txt
- [] **src**
  - `Analysis_QY2H.py`
  - [] **utils**
    - `__init__.py`
    - `channels.config`
    - `Colors.py`
    - `Configuration.py`
    - `Configure_Channels.py`
    - `Ending_Window.py`
    - `Functions.py`
    - `Logo.jpg`
    - `Object_Echantillon.py`
    - `Opening_window.py`
    - `Variables.py`
- [] **doc**
  - Analysis_Configuration.jpg
  - Analysis_Progress.jpg
  - Logo_cnrs.jpg
  - Logo_ens.jpg
  - Logo_LBMC.jpg
  - Logo.jpg
  - Main_Menu.jpg
  - Results.jpg
  - Select_Channels.jpg
  - Select_File.jpg
  - Select_Input.jpg
  - Select_Output.jpg

User Guide

1 Recommendation for acquisition

Our program requires `linear` values for all fluorescence channels. Thus, be vigilant that your acquisition program is saving data as linear (even if your acquisition display is `log` or `hyper log`).

Yeast cells are usually smaller than the focused laser beam (spot) of flow cytometers. The maximum signal (= Height, H) for a given cell is obtained when the cell is fully covered by the laser spot. Thus, `H` reflects the total cellular content of the fluorophore. Therefore, we recommend to use the signal `Height (H)` of each channel.

Moreover, some flow-cytometers can apply internal corrections on specific channels. For example, the MacsquantVYB (that we used for our experiment) is correcting the `Area A` of each channels:

Area is the sum of a defined number of adjacent samples at the trigger time point divided by a scaling factor. This factor is chosen in a way that for “normal” events H=A to obtain a diagonal. The scaling factor is pressure dependent.

Thus we strongly recommend to use as much as possible **non-manipulated** values.

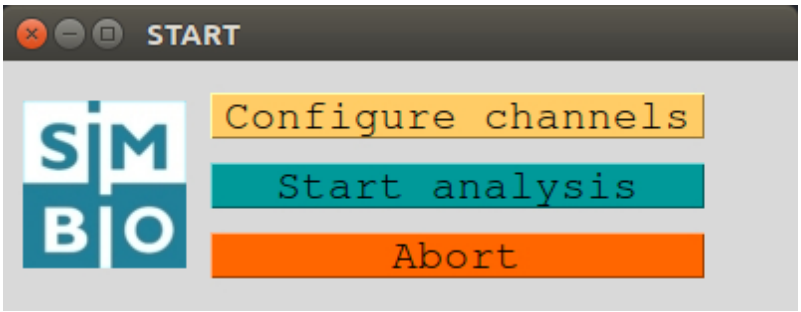
2 Main Menu

To start the program you need to execute the `Analysis_QY2H.py` python script:

```
$ python Analysis_QY2H.py
```

The main menu will propose you different functions:

1. **Configure channels** To select the channels to be used for the analysis.
2. **Start analysis** To generate a `quantitative Yeast Two Hybrid` affinity ladder from a set of `.fcs` files.
3. **Abort** To exit the program.



3 Configure the names of the channels

Before performing your first analysis, it is recommended to configure your channels. If you keep always the same acquisition settings, this step is required only once.

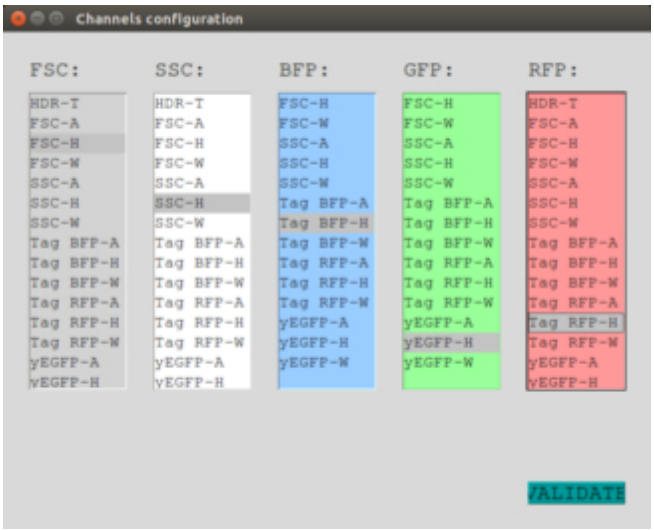
When clicking on `Configure channels`, the program prompts you to choose a `.fcs` file.

The program will identify all channels recorded in your

file. You can then attribute the correct names in the various columns. For the subsequent analysis, the column `BFP` corresponds to the `Reporter` you want to quantify. The `RFP` and the `GFP` columns correspond to the `BD-Bait` and `AD-Prey` fusion proteins respectively.

The first two columns are not used yet, but might be included in a future development of this program to sub-select a population of cells with uniform `FSC` and/or `SSC`.

The select channel names will be saved in the `channels.config` file (in the `utils` folder) when clicking on `VALIDATE`.



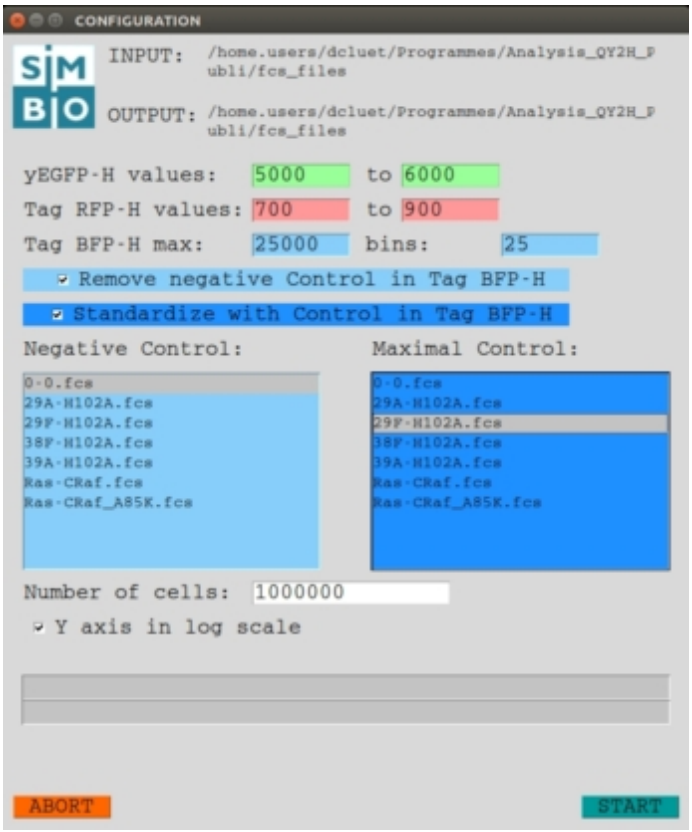
4 Perform an analysis

When clicking on `Start analysis`, the program displays the analysis configuration window. You need first to select the folder where all your files (from the same experiment) are stored.

The program will automatically find all `.fcs` files present in this folder and display them in the analysis settings interface.

You need then to specific in which folder you want the output files to be generated. By default, the program is set on the input folder.

Once the path of the `ÌNPUT` and `OUTPUT` folders are set, you have access to the analysis settings. The program will generate the `Affinity ladder` by taking a sub-ensemble of cells using gates in the `AD-Prey GFP` and `BD-Bait RFP` channels. By default the minimal and maximal values are set to those of the Fig. 4 (B and C) of our publication.

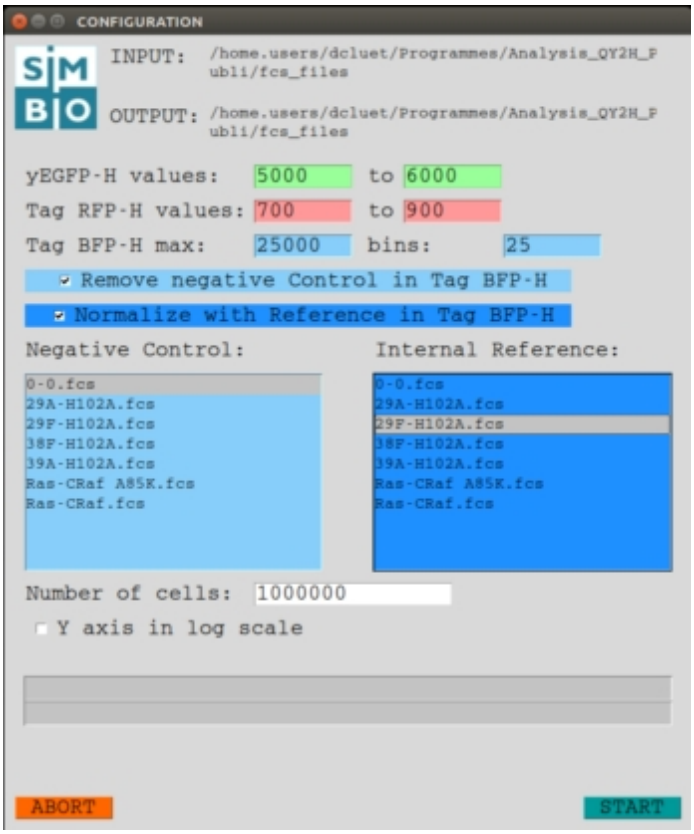


The maximum in the `Reporter (BFP)` channel, corresponds to the upper-limit (x axis) of the generated `Cumulative mean` for each sample. If the curves in the `.pdf` output file are not reaching a plateau, increase this value.

The value `BFP bins` corresponds to the number of points you want to be displayed on the final graph.

You can remove the background of the system by selecting `Remove negative Control`. Unchecking this option is useful to monitor the contribution of the background in your experiment. This information is helpful especially for the weakest interactors.

As the sensitivity of the system may vary from one batch of yeast to an other, you can ‘normalize’ (to 100) the BFP signal using an `Internal Reference`. This will allow you to better compare various experiments. We recommend to use the strongest interaction as Internal Reference.

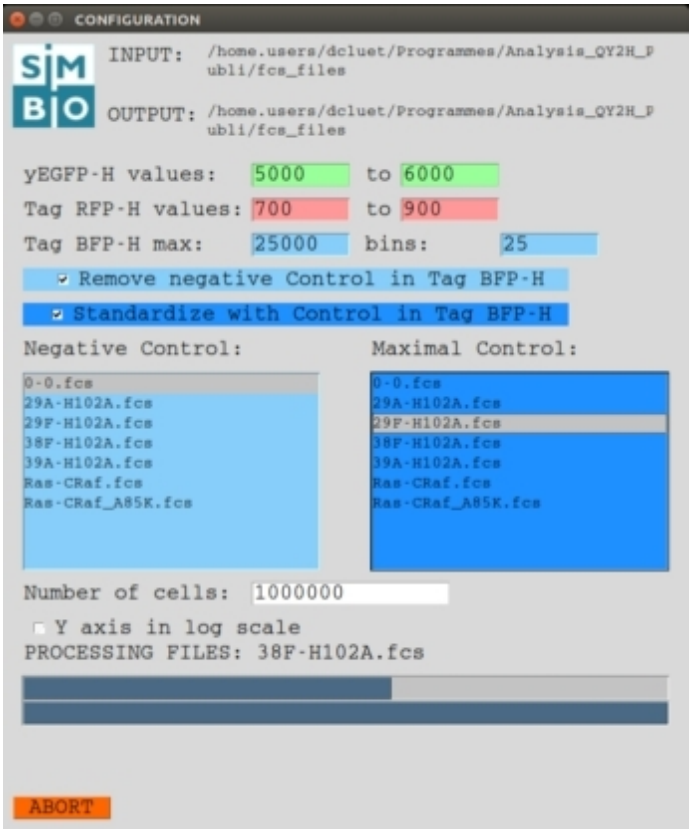


You need to specify which sample file corresponds to your `negative` control and `Internal reference`, even if no background subtraction or normalization are applied. Typically, the negative control corresponds to a qY2H experiment performed with fluorescent empty BD-Bait and AD-Prey fusion proteins. In our work, this control is called 0-0.fcs.

The value `Number of cells` corresponds to the maximum number of cells to be loaded from your file before doing the dual gating in the `AD-Prey GFP` and `BD-Bait RFP` channels. **We highly recommend you to analyse at least 1 000 000 events to obtain a reliable affinity ladder.**

You have the possibility to display the `Cumulative mean` in log or linear scale.

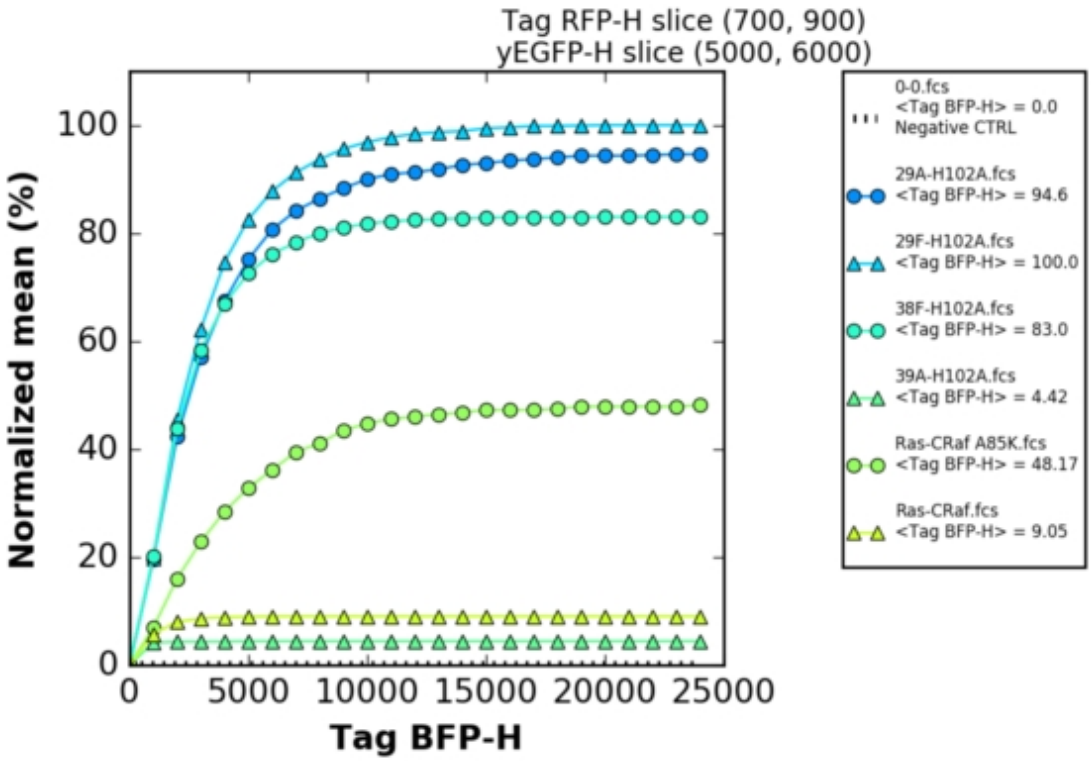
When clicking START, the program proceeds to the analysis (only if a negative control has been specified).



During the analysis the two Progress Bars inform you which file (first bar) is currently processed, and which analysis step (second bar) is performed.

Finally, the program displays the result of the analysis, with the main settings in the title. Here we present the result with the following activated options:

- **Remove negative Control**
- **Normalize with the Internal Standard**
- **Y axis in linear scale**



Click on ABORT to exit the program.

5 Output files

The program generates three files:

- A `RESULT.png` image of the graph presented at the end of the analysis
- A `.csv` table containing the `mean BFP` value for each sample file (after subtraction of the negative control and normalization, if selected)
- A `.pdf` report file, that encloses the qY2H affinity ladder graph.

The `.csv` and `.pdf` files have a common unique prefix based on the date and time of analysis. Moreover the data processing (*i.e.* background substraction and/or normalisation) is explicitly indicated.

6 Example files

The flow-cytometry files of qY2H experiments can be downloaded from http://flowrepository.org under accession numbers:

- **FR-FCM-ZYUL** (10 millions cells)
- **FR-FCM-Z25G** (1 million cells)