Supporting Information Appendix

Contents

Table S1: Drug List and Abbreviations

Drug (Abbreviation, Source)	**Class**	**Mode of Action**
Chloramphenicol (Cm, chloramphenicol, MP biomedicals)	Protein synthesis inhibitor (50 S target)	Inhibits peptidyl transferase activity
Doxycycline (Doxy, doxycycline hyclate, Sigma Aldrich)	Protein synthesis inhibitor (30 S target)	Inhibits binding of aminoacyl t-RNA
Erythromycin (Ery, erythromycin, Sigma Aldrich)	Protein synthesis inhibitor (macrolide)	Inhibits translocation of peptidyl t-RNA
Lincomycin (Linc, lincomycin hydrochloride, MP biomedicals)	Protein synthesis inhibitor (50 S target)	Inhibits peptidyl transferase activity
Ciprofloxacin (Cip, ciprofloxacin, Sigma Aldrich)	DNA synthesis inhibitor (fluoroquinolone)	Inhibits activity of DNA gyrase
Ofloxacin (Ofl, ofloxacin, Sigma Aldrich)	DNA synthesis inhibitor (fluoroquinolone)	Inhibits activity of DNA gyrase
Trimethoprim (Tmp, trimethoprim, Sigma Aldrich)	Folic acid synthesis inhibitor	Inhibits dihydrofolate reductase
Salicylate (Sal, sodium salicylate, Sigma Aldrich)	Pain reliever	Inducer of mar system
Tetracycline (Tet, tetracycline hydrochloride, Acros)	Protein synthesis inhibitor (30 S target)	Inhibits binding of aminoacyl t-RNA
Kanamycin (Kan, kanamycin monosulfate, Fisher)	Aminoglycoside	Stalls 30S initiation complex; induces misreading of mRNA
Tobramycin (Tobr, tobramycin, TCI)	Aminoglycoside	Binds to 30S and 50S ribosomes,
Penicillin G (PenG, penicillin G potassium, Fisher)	Cell wall synthesis inhibitor	Inhibits peptidoglycan cross-linking
Ampicillin (Amp, ampicillin sodium salt , Fisher)	Cell wall synthesis inhibitor	Inhibits peptidoglycan cross-linking
Sodium Benzoate (NaBenz, sodium benzoate, Sigma Aldrich)	Preservative	Inducer of mar system
Bacitracin (Bac, bacitracin, USB)	Polypeptide antibiotic	Interferes with cell wall construction
Sulfamonomethoxine (Sulf, sulfamonomethoxine, TCI)	Folic acid synthesis inhibitor	Inhibits PABA dihydrofolate synthase
Spectinomycin (Spect, spectinomycin dihydrochloride, Sigma Aldrich)	Protein synthesis inhibitor (aminocyclitol)	Stalls 30S initiation complex
Spiramycin (Spir, spiramycin, TCI)	Protein synthesis inhibitor (macrolide)	Inhibits translocation of peptidyl t-RNA
Naladixic Acid (Nal Acid, naladixic acid, Fisher)	DNA synthesis inhibitor (quinolone)	Inhibits activity of DNA gyrase

Table S2:

Drug Combination	R^2 (Pairwise)	R^2 (Independent)	ΔAIC	Weight for Pairwise
Sal-Ery-Cm	0.93	0.82	-109.0	> 0.999
Cm-Ery-Tmp	0.87	0.87	-5.9	0.95
Cm-Ofl-Sal	0.74	**	-909.4	> 0.999
Cm-Ofl-Tmp	0.90	0.87	-14.4	> 0.999
Dox-Ery-Linc	0.86	0.88	51.5	< 0.001
Dox-Ery-Linc-Sal	0.72	**	-161.6	> 0.999
Linc-Cm-Ofl-Tmp	0.85	0.70	-83.4	> 0.999
All Data	**0.90**	**0.33**	**-2233.2**	**> 0.999**

Table S2. Comparison of Pairwise Approximation with Independent Model
Coefficient of determination, R^2, is defined as $R^2 = 1\text{-}SS_{err} / SS_{tot}$, where SS_{err} is the residual sum of squares between model and data, and SS_{tot} is the total sum of squares (proportional to the variance of the experimental measurements). ΔAIC is the difference in AIC values between the pairwise and the independent model. The last column provides the Akaike weight in favor of the pairwise model. Drug abbreviations are given in Table S1.

** $R^2<0$, which indicates very poor fit (the mean of the data provides a better fit than the model)

Table S3: Drug Combinations Used in Combinatorial Experiments:

Total of 120 unique drug dosage combinations comprised of 93 unique 3-drug combinations. Note that some of the 3-drug combinations (e.g. ampicillin, spectinomycin, and spiramycin) are repeated at different dosages.

**Drug concentrations are given in ug/mL with the following exceptions: [Cip] is ng/mL, [Sal] is mM, and [NaBenz] is mM.

Abbreviations: Amp is ampicillin, Cip is ciprofloxacin, Tobr is tobramycin, Nal Acid is naladixic acid, Spect is spectinomycin, Spir is spiramycin, NaBenz is sodium benzoate, Tmp is trimethoprim, Sulf is sulfamonomethexate, Dox is doxyclycine, Cm is chloramphenicol, PenG is penicillin G, and Bac is bacitracin.

Label	Drug 1	[Drug 1] (ug/mL)**	Drug 2	[Drug 2] (ug/mL)**	Drug 3	[Drug 3] (ug/mL)**
	Combinatorial Experiment 1 (7 drugs): Amp, Cip, Tobr, Tmp, Nal Acid, Spect, Spir					
1	Amp	1.5	Cip	3	Tobr	1.6
2	Amp	1.5	Cip	3	Tmp	0.15
3	Amp	1.5	Cip	3	Nal Acid	1.75
4	Amp	1.5	Cip	3	Spect	10
5	Amp	1.5	Cip	3	Spir	200
6	Amp	1.5	Tobr	1.6	Tmp	0.15
7	Amp	1.5	Tobr	1.6	Nal Acid	1.75
8	Amp	1.5	Tobr	1.6	Spect	10
9	Amp	1.5	Tobr	1.6	Spir	200
10	Amp	1.5	Tmp	0.15	Nal Acid	1.75
11	Amp	1.5	Tmp	0.15	Spect	10
12	Amp	1.5	Tmp	0.15	Spir	200
13	Amp	1.5	Nal Acid	1.75	Spect	10
14	Amp	1.5	Nal Acid	1.75	Spir	200
15	Amp	1.5	Spect	10	Spir	200
16	Cip	3	Tobr	1.6	Tmp	0.15
17	Cip	3	Tobr	1.6	Nal Acid	1.75
18	Cip	3	Tobr	1.6	Spect	10
19	Cip	3	Tobr	1.6	Spir	200
20	Cip	3	Tmp	0.15	Nal Acid	1.75
21	Cip	3	Tmp	0.15	Spect	10
22	Cip	3	Tmp	0.15	Spir	200
23	Cip	3	Nal Acid	1.75	Spect	10
24	Cip	3	Nal Acid	1.75	Spir	200
25	Cip	3	Spec	10	Spir	200
26	Tobr	1.6	Tmp	0.15	Nal Acid	1.75
27	Tobr	1.6	Tmp	0.15	Spect	10

Label	Drug 1	[Drug 1] (ug/mL)**	Drug 2	[Drug 2] (ug/mL)**	Drug 3	[Drug 3] (ug/mL)**
28	Tobr	1.6	Tmp	0.15	Spir	200
29	Tobr	1.6	Nal Acid	1.75	Spect	10
30	Tobr	1.6	Nal Acid	1.75	Spir	200
31	Tobr	1.6	Spect	10	Spir	200
32	Tmp	0.15	Nal Acid	1.75	Spect	10
33	Tmp	0.15	Nal Acid	1.75	Spir	200
34	Tmp	0.15	Spect	10	Spir	200
35	Nal Acid	1.75	Spect	10	Spir	200
Combinatorial Experiment 2 (6 drugs): Amp, NaBenz, Tmp, Nal Acid, Spect, Spir						
36	Amp	0.9	NaBenz	2	Tmp	0.05
37	Amp	0.9	NaBenz	2	Nal Acid	1.25
38	Amp	0.9	NaBenz	2	Spect	9
39	Amp	0.9	NaBenz	2	Spir	200
40	Amp	0.9	Tmp	0.05	Nal Acid	1.25
41	Amp	0.9	Tmp	0.05	Spect	9
42	Amp	0.9	Tmp	0.05	Spir	200
43	Amp	0.9	Nal Acid	1.25	Spect	9
44	Amp	0.9	Nal Acid	1.25	Spir	200
45	Amp	0.9	Spect	9	Spir	200
46	NaBenz	2	Tmp	0.05	Nal Acid	1.25
47	NaBenz	2	Tmp	0.05	Spect	9
48	NaBenz	2	Tmp	0.05	Spir	200
49	NaBenz	2	Nal Acid	1.25	Spect	9
50	NaBenz	2	Nal Acid	1.25	Spir	200
51	NaBenz	2	Spect	9	Spir	200
52	Tmp	0.05	Nal Acid	1.25	Spect	9
53	Tmp	0.05	Nal Acid	1.25	Spir	200
54	Tmp	0.05	Spect	9	Spir	200
55	Nal Acid	1.25	Spect	9	Spir	200
Combinatorial Experiment 3 (5 drugs): Amp, Tobr, Nal Acid, Spect, Spir						
56	Amp	0.8	Tobr	1.6	Nal Acid	1
57	Amp	0.8	Tobr	1.6	Spect	8
58	Amp	0.8	Tobr	1.6	Spir	200
59	Amp	0.8	Nal Acid	1	Spect	8
60	Amp	0.8	Nal Acid	1	Spir	200
61	Amp	0.8	Spect	8	Spir	200
62	Tobr	1.6	Nal Acid	1	Spect	8
63	Tobr	1.6	Nal Acid	1	Spir	200

Label	Drug 1	[Drug 1] (ug/mL)**	Drug 2	[Drug 2] (ug/mL)**	Drug 3	[Drug 3] (ug/mL)**
64	Tobr	1.6	Spect	8	Spir	200
65	Nal Acid	1	Spect	8	Spir	200
Combinatorial Experiment 4 (7 drugs): Amp, Cip, Sulf, Tmp, NaBenz, Dox, Cm						
66	Amp	1.25	Cip	3.5	Sulf	1.5
67	Amp	1.25	Cip	3.5	Tmp	0.13
68	Amp	1.25	Cip	3.5	NaBenz	1.5
69	Amp	1.25	Cip	3.5	Dox	0.15
70	Amp	1.25	Cip	3.5	Cm	0.5
71	Amp	1.25	Sulf	1.5	Tmp	0.13
72	Amp	1.25	Sulf	1.5	NaBenz	1.5
73	Amp	1.25	Sulf	1.5	Dox	0.15
74	Amp	1.25	Sulf	1.5	Cm	0.5
75	Amp	1.25	Tmp	0.13	NaBenz	1.5
76	Amp	1.25	Tmp	0.13	Dox	0.15
77	Amp	1.25	Tmp	0.13	Cm	0.5
78	Amp	1.25	NaBenz	1.5	Dox	0.15
79	Amp	1.25	NaBenz	1.5	Cm	0.5
80	Amp	1.25	Dox	0.15	Cm	0.5
81	Cip	3.5	Sulf	1.5	Tmp	0.13
82	Cip	3.5	Sulf	1.5	NaBenz	1.5
83	Cip	3.5	Sulf	1.5	Dox	0.15
84	Cip	3.5	Sulf	1.5	Cm	0.5
85	Cip	3.5	Tmp	0.13	NaBenz	1.5
86	Cip	3.5	Tmp	0.13	Dox	0.15
87	Cip	3.5	Tmp	0.13	Cm	0.5
88	Cip	3.5	NaBenz	1.5	Dox	0.15
89	Cip	3.5	NaBenz	1.5	Cm	0.5
90	Cip	3.5	Dox	0.15	Cm	0.5
91	Sulf	1.5	Tmp	0.13	NaBenz	1.5
92	Sulf	1.5	Tmp	0.13	Dox	0.15
93	Sulf	1.5	Tmp	0.13	Cm	0.5
94	Sulf	1.5	NaBenz	1.5	Dox	0.15
95	Sulf	1.5	NaBenz	1.5	Cm	0.5
96	Sulf	1.5	Dox	0.15	Cm	0.5
97	Tmp	0.13	NaBenz	1.5	Dox	0.15
98	Tmp	0.13	NaBenz	1.5	Cm	0.5
99	Tmp	0.13	Dox	0.15	Cm	0.5
100	NaBenz	1.5	Dox	0.15	Cm	0.5

Label	Drug 1	[Drug 1] (ug/mL)**	Drug 2	[Drug 2] (ug/mL)**	Drug 3	[Drug 3] (ug/mL)**
Combinatorial Experiment 5 (6 drugs): PenG, Bac, Sulf, Dox, Cip, Cm						
101	PenG	20	Bac	200	Sulf	2.5
102	PenG	20	Bac	200	Dox	0.25
103	PenG	20	Bac	200	Cip	5
104	PenG	20	Bac	200	Cm	0.4
105	PenG	20	Sulf	2.5	Dox	0.25
106	PenG	20	Sulf	2.5	Cip	5
107	PenG	20	Sulf	2.5	Cm	0.4
108	PenG	20	Dox	0.25	Cip	5
109	PenG	20	Dox	0.25	Cm	0.4
110	PenG	20	Cip	5	Cm	0.4
111	Bac	200	Sulf	2.5	Dox	0.25
112	Bac	200	Sulf	2.5	Cip	5
113	Bac	200	Sulf	2.5	Cm	0.4
114	Bac	200	Dox	0.25	Cip	5
115	Bac	200	Dox	0.25	Cm	0.4
116	Bac	200	Cip	5	Cm	0.4
117	Sulf	2.5	Dox	0.25	Cip	5
118	Sulf	2.5	Dox	0.25	Cm	0.4
119	Sulf	0.25	Cip	5	Cm	0.4
120	Dox	0.25	Cip	5	Cm	0.4

Table S4:

Drug Combination	f_c
E. coli	
Sal-Ery-Cm	0.95
Cm-Ery-Tmp	0.89
Cm-Ofl-Sal	0.98
Cm-Ofl-Tmp	0.95
Dox-Ery-Linc	0.94
S. aureus	
Tet-Kan-Ery	0.93
Total (All Drugs)	**0.97**

Table S4. Validation of Pairwise Approximation. f_c represents the fraction of total three-drug correlations that are captured by the pairwise model. For each drug dosage containing nonzero amounts of all three drugs, we calculated the maximum entropy distributions P_N (N=1,2,3), which are consistent with all measurements involving N or fewer drugs. We then calculated the multi-information $I_3 = S_1 - S_3$, where S_i is the entropy of the distribution P_i. The fraction of total correlations captured by the pairwise model is then

$$f_c = \frac{\sum I_2}{\sum I_3}$$

where sums run over all data points for a given 3-drug combination. An f_c of 1 would indicate that the pairwise model captured all higher-order correlations or, equivalently, that interactions involving exactly N drugs (for N>2) do not contribute to the multi-drug effects. Cm, chloramphenicol; Dox, doxycycline; Ery, erythromycin; Kan, kanamycin; Linc, lincomycin; Ofl, ofloxacin; Sal, salicylate . Tet, tetracycline; Tmp, trimethoprim.

2 Supporting Text

2.1 Example Growth Curve

An example growth curve is shown in Figure S1.

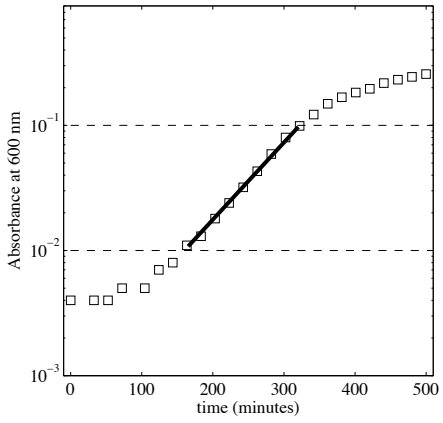

Figure S1: Growth Rate Measurement. Time series of A_{600} vs. time. Solid line, fit to exponential function. Dashed lines, region of exponential growth. Growth rate is given by the slope of the line.

2.2 Statistical Framework for Drug Combinations

The ultimate goal of our analysis is to establish a predictive relationship between the effects of small drug combinations (1- or 2-drug combinations) and the effects of larger multi-drug combinations. Because mechanistic models for large intracellular networks are often not tractable, we introduce a statistical framework which, by construction, associates drug interactions to correlations between stochastic variables. The model offers one way of establishing testable predictions by first mapping experimental measurements to moments of a joint probability distribution. The problem is then reduced to estimating the unknown distribution, which can be achieved using statistical techniques, such as entropy maximization, or (in principle) by incorporating other assumptions about the underlying physical system.

Specifically, we assume that interactions between N drugs can be modeled as correlations between N continuous stochastic variables, $X_i, (i = 1...N)$, such

that the observed growth of cells ($g_{1,2..N}$) in the presence of N drugs is given by

$$g_{1,2..N} = \langle X_1 X_2 ... X_N \rangle \tag{S1}$$

where brackets represent an expectation value over an ensemble described by the unknown probability density $P(x_1, x_2, ...x_N)$. If the variables X_i are uncorrelated, the growth reduces to a product

$$g_{1,2..N} = \langle X_1 \rangle \langle X_2 \rangle ... \langle X_N \rangle \equiv g_1 g_2 ... g_N, \tag{S2}$$

which is equivalent to Bliss independence, a common phenomenological model used in pharmacology to describe non-interacting drugs [2].

We would like to ask whether pairwise interactions between drugs can be used to predict the effects of larger combinations of drugs. Within the above framework, predicting effects of drug combinations reduces to estimating moments of the unknown distribution $P(x_1, x_2, ...x_N)$ using data on interactions between pairs of drugs. Therefore, to test our hypothesis, we must estimate higher-order moments of $P(x_1, x_2, ...x_N)$ (the effects of a multi-drug combination) using only the lower order moments (the effects of two-drug combinations). The question, then, is how does one estimate, without mechanistic assumptions or a physical model, the unknown probability distribution $P(x_1, x_2, ...x_N)$ given only information about some collection of moments of that distribution,

$$\langle f_j \rangle \equiv \int_a^b \int_a^b ... \int_a^b P(x_1, x_2, ...x_N) f_j(x_1, x_2, ...x_N) dx_1 dx_2 ... dx_N = \alpha_j. \tag{S3}$$

Entropy maximization offers one method of solving this problem by choosing a distribution consistent with known moments but that does not incorporate additional statistical structure [18, 19, 33].

In what follows, we restrict ourselves for illustrative purposes to the three-drug case, though the results are easily generalizable to any larger drug combination. To estimate $P(x_1, x_2, x_3)$, we maximize the entropy, $S(P)$, subject to the known moment constrains. The entropy, $S(P)$, is defined (up to an additive constant) as

$$S(P) = -\int_a^b \int_a^b \int_a^b P(x_1, x_2, x_3) \log \left(\frac{P(x_1, x_2, x_3)}{q(x_1, x_2, x_3)} \right) dx_1 dx_2 dx_3, \tag{S4}$$

where $q(x_1, x_2, x_3)$ is a continuous prior distribution that accounts for an a priori knowledge gleaned from, for example, physical considerations or experience. The maximization amounts to minimizing the Kullback-Leibler divergence [35] between the distributions P and q, subject to constraints on the moments. We choose the interval $[a, b]$ to be finite and take $q(x_1, x_2, x_3)$ to be a constant, which is equivalent to assuming a uniform prior distribution. We stress that our results do not depend on a specific choice of $[a, b]$, as long as some minimal conditions are met (see below).

To proceed with the estimation of $P(x_1, x_2, x_3)$, we first measured the growth response to each drug i alone (g_i) and to all pairs of drugs, (g_{ij}). To predict the

effects of a given three-drug combination, for example, we measured g_1, g_2, g_3, g_{12}, g_{13}, and g_{23}. The corresponding constraints on the distribution are simply

$$\begin{aligned}
\langle f_1 \rangle &\equiv \int_a^b \int_a^b \int_a^b P(x_1, x_2, x_3) x_1 dx_1 dx_2 dx_3 = g_1, \\
\langle f_2 \rangle &\equiv \int_a^b \int_a^b \int_a^b P(x_1, x_2, x_3) x_2 dx_1 dx_2 dx_3 = g_2, \\
\langle f_3 \rangle &\equiv \int_a^b \int_a^b \int_a^b P(x_1, x_2, x_3) x_3 dx_1 dx_2 dx_3 = g_3, \\
\langle f_4 \rangle &\equiv \int_a^b \int_a^b \int_a^b P(x_1, x_2, x_3) x_1 x_2 dx_1 dx_2 dx_3 = g_{12}, \\
\langle f_5 \rangle &\equiv \int_a^b \int_a^b \int_a^b P(x_1, x_2, x_3) x_1 x_3 dx_1 dx_2 dx_3 = g_{13}, \\
\langle f_6 \rangle &\equiv \int_a^b \int_a^b \int_a^b P(x_1, x_2, x_3) x_2 x_3 dx_1 dx_2 dx_3 = g_{23}.
\end{aligned} \tag{S5}$$

We can use Lagrange multipliers $(\lambda_0, h_1, h_2, h_3, J_{12}, J_{13}, J_{23})$ to maximize the entropy $S(P)$ subject to these constraints, which leads to

$$P(x_1, x_2, x_3) = \frac{1}{Z} \exp\left(h_1 x_1 + h_2 x_2 + h_3 x_3 + J_{12} x_1 x_2 + J_{13} x_1 x_3 + J_{23} x_2 x_3\right), \tag{S6}$$

where Z is a constant (related to λ_0) that normalizes the distribution. It can be shown that, in general, the entropy of a distribution calculated in this way corresponds to the global maximum, if it exists [19],[33].

We have labeled the Lagrange multipliers as h_i and J_{ij} in accordance with notation commonly used for the well-known Ising model, which takes a similar form [36]. In the context of our drug interaction model, h_i encodes the single-drug growth response and J_{ij} encodes information about deviations from Bliss independence for a given drug pair, with $J_{ij} > 0$ indicating antagonism and $J_{ij} < 0$ indicating synergy. We call the parameter h_i the resilience coefficient and J_{ij} the drug-drug coupling coefficient between the drugs i and j; they characterize the response to single drugs and to pairs of drugs, respectively (Fig. S2). Intuitively, the value of the resilience coefficient reflects the cell growth in response to a given concentration of one drug (Fig. S2). The resilience coefficient decreases with increasing drug concentration. The drug-drug coupling coefficient, J, reflects, for each drug dosage, the nature of interactions taking place between two given drugs (Fig. S2). For example, when J is zero, there exists no drug-drug coupling and the two drugs act independently. When J is positive, the drug pair is antagonistic and for negative values, the pair is synergistic.

2.2.1 Growth Rate Predictions and Uncertainties

In practice, we calculate the parameters h_i and J_{ij} from experimental data using a standard numerical technique that involves minimizing a dual space

Lagrangian [37]. The minimization occurs on a convex surface and can be accomplished with any unconstrained optimization algorithm. For each dosage of a given three- or four-drug combination, we performed the optimization 50 times (for 3 drugs) or 25 times (for four drugs) starting from random initial conditions drawn from a uniform distribution on the interval $[-0.5, 0.5]$. Nonphysical predictions ($g < 0$, $g > 1$) occasionally arise from strongly synergistic or strongly antagonistic combinations, and these are set to 0 (no growth) or 1 (maximum growth), respectively. While the minimization should not be prone to errors due to local minima, we find that fits of similar quality can be achieved using a range of parameter values; hence, there is some uncertainty in the location of the true minimum. Taking random initial conditions allows us to estimate this uncertainty and offer more reliable predictions. All predictions represent the mean of these trials. Error bars of the growth predictions in Figure 2 are $\pm 2\sigma$, with σ the standard deviation of the distribution of trials. Standard errors of the mean, which are between 5 and 8 times smaller, could be used instead to give a true estimate of the error associated with each prediction, but they leave the reader without a sense of σ. Uncertainties in the prediction of drug interactions, $I_{1..N} \equiv g_{12..N} - g_1 g_2 ... g_N$, (Figure 3) must incorporate standard errors from single drug measurements (g_i). Therefore, the error bars represent $\pm$ 1 standard error of the mean. For distributions of 25 or 50 trials, the standard error associated with the prediction of the first term ($g_{12..N}$) is much smaller than that of the second term ($g_1 g_2 ... g_N$). Uncertainties of the drug interaction predictions are therefore dominated by standard errors in the estimates of single drug growth rates g_i appearing in the second term.

2.2.2 Choosing the State Space

The calculation of the maximum entropy distribution requires a specific choice of state space, $[a, b]$, for each continuous stochastic variable X_i. First, we note that if the boundaries are chosen such that $[a, b] = [-\infty, \infty]$ - that is, the variables take values on the real line - a (normalizable) distribution of the form Equation S6 does not exist, because there are no constraints on the variances, $\langle X_i^2 \rangle$. In practice, this difficulty can be circumvented by choosing $[a, b]$ to be finite, which puts implicit limits on the variance of each variable. While this amounts to an additional assumption, we find empirically that the predictions of higher moments from lower moments do not depend on the choice $[a, b]$ as long as i) the distribution of the form Equation S6 is normalizable and ii) a solution to Equations S5, S6 can be found for some choice of Lagrange multipliers. The specific values of the Lagrange multipliers will of course depend on the choice of state space, but the relationship between higher moments and lower moments conforms to that given by Issesrlis' theorem in all cases where a suitable solution to Equation 5 is found. We return to this point below.

Figures S3, S4 illustrate the fit of models with different choices of (a, b) to all two-drug and single drug data. We note that these are not predictions, but simply fits to examine whether a solution to Equations S5, S6 can be found. Figure S3 illustrates that choices with $(a, b) = (0, b)$ for $b > 0$ do not provide

an accurate description of many of the measured drug interactions; that is, a valid solution cannot be found. On the other hand, the fit improves significantly when $a < 0$ and $b > 0$ (Figure S4). For sufficiently large $|b - a|$, the fit again becomes poor, likely because of the failure of numerical integration over the increasingly large state space. Hence, for all three-drug calculations, we choose $(a, b) = (-3, 4)$ (Figure S4, lower left panel), which provides an excellent fit ($R^2 > 0.99$) to the pairwise data, indicating that a solution to Equations S5, S6 is achievable. This choice is not unique, and other choices (e.g. $(a, b) = (-9, 10)$) are possible but must utilize more computational resources to calculate integrals at the same level of accuracy. For similar reasons, we choose a smaller range $(a, b) = (-1, 2)$ for four-drug predictions to allow for faster computation of the numerical integrals. The final predictions do not depend on these choices of state space, but instead only on the measured growth rates for drug pairs and single drugs. The exact same results are also obtained if we choose the variables to be discrete "spin-like" variables, as long as the value of the spin is sufficiently large (e.g. spin $= \pm 4$). In the latter case, the integrals become sums that are easily calculated.

2.2.3 Example Maximum Entropy Distributions

We illustrate example (marginal) maximum entropy distributions calculated for the drug combination salicylate, erythromycin, and chloramphenicol in Figures S5, S6. Figure S5 shows the pairwise, $P_2(x_1, x_2) \equiv \int_a^b P(x_1, x_2, x_3)dx_3$, and single variable, $P_1(x_1) \equiv \int_a^b P(x_1, x_2, x_3)dx_3dx_2$, marginal distributions for the three-drug combination at a given dose of each drug. In this figure, the concentration of chloramphenicol is 0, so these distributions describe the effects of salicylate and erythromycin alone (right panels) and in combination (left panel). Similarly, Figure S6 shows the pairwise and single variable marginal distributions for erythromycin and chloramphenicol in the absence of salicylate. Deviations from the uniform distribution ensure that the experimental measurements of pairwise drug interactions (2-body correlations) and single-drug effects (single variable means) are appropriately described by expectation values of P.

2.2.4 Isserlis' Theorem Describes Observed Moment Relationships

Empirically, we find that the moment relationships derived from our experiments are consistent with the well-known Isserlis' formula [38],

$$\langle X_i X_j X_k \rangle = \langle X_i \rangle \langle X_j X_k \rangle + \langle X_j \rangle \langle X_i X_k \rangle + \langle X_k \rangle \langle X_i X_j \rangle - 2\langle X_i \rangle \langle X_j \rangle \langle X_k \rangle, \quad \text{(S7)}$$

or in terms of the growth measurements,

$$g_{ijk} = g_i g_{jk} + g_j g_{ik} + g_k g_{ij} - 2 g_i g_j g_k. \quad \text{(S8)}$$

Similar expressions hold for higher order moments. For example,

$$g_{ijkl} = g_{il} g_{jk} + g_{ik} g_{jl} + g_{ij} g_{kl} - 2 g_i g_j g_k g_l. \quad \text{(S9)}$$

Isserlis' equations were originally proven for jointly distributed Gaussian variables, but they have also been extended to certain classes of non-Gaussian variables [39]. These relationships can be derived from the maximum entropy results using first order perturbation theory when the drug-drug coupling is small compared to the single drug effects; they are exact if the distribution $P(x)$ is Gaussian. The result (Figure S7) is perhaps not surprising, given that the choice of finite $[a, b]$ implicitly constrains the variance of the distributions.

Consider, for example, that the same relationship can also be achieved in the following way. Assume that the variables are constrained such that $\langle X_i^2 \rangle = \sigma_i^2$ for some choice of constants $\sigma_i^2 > 0$. Under these conditions, the maximum entropy distribution for variables defined on the real line is a Gaussian [33]. Therefore, Isserlis' theorem will describe the moment relationships, and the result will not depend on the specific choices of σ_i^2, as long as they are sufficiently large that a distribution satisfying all moment constraints exists.

The success of Equation S8 and, more generally, Isserlis' theorem in predicting the effects of large drug combinations is, in itself, a striking result. It suggests that one could arrive at the same predictions by assuming, at the outset, that the variables X_i come from a multi-variate Gaussian distribution. Such a relationship could arise, for example, from the Central Limit Theorem if one could argue that the underlying stochasticity of intracellular networks contributing to the multi-drug response arises from a sum of independent, or nearly independent, stochastic variables. This remains an open question for future work. Nevertheless, in practice, the simplicity of the algebraic expressions given by Isserlis renders the method useful even to those without extensive computational resources or experience.

2.2.5 Drug With Itself

In pharmacology, Bliss independence is well-known to be a poor model for the effects of a two drugs with highly similar mechanisms. In particular, it is often noted that Bliss independence cannot accurately describe an experiment where a drug is divided into two volumes which are then combined (i.e. the "interactions" of a drug with itself). Our results extend Bliss independence to account for interactions between drug pairs, which raises the question of whether the model can more accurately describe the "interaction" of a drug with itself. Applying equation 8 to a such a scenario, we have

$$\begin{aligned} g(c_1 + c_2 + c_3) =& g(c_1)g(c_2 + c_3) + g(c_2)g(c_1 + c_3) + \\ & g(c_3)g(c_1 + c_2) - 2g(c_1)g(c_2)g(c_3), \end{aligned} \tag{S10}$$

where $g(x)$ is the growth in the presence of a drug at a concentration x. One solution to this equation is given by an exponential function, which is a reasonable model for the dose-response curve of many drugs over limited concentration ranges. However, dose-response curves are typically modeled with a Hill function, $g(x) = (1 + (x/K)^n)^{-1}$, which is consistent with our single-drug data but is not a solution of equation S10. To explore the usefulness of equation S10 for

describing typical Hill-like dose-response relationships, we consider Hill functions with Hill coefficients of $n = 1$, $n = 2$, and $n = 5$ (and $K = 1$ without loss of generality). We then compare the predictions of equation S10 and the predictions of Bliss independence (given by $g(c_1 + c_2 + c_3) = g(c_1)g(c_2)g(c_3)$) with the true Hill function (Figure S8). The pairwise model significantly improves upon Bliss independence, especially when Hill coefficients are near 1, but it can not perfectly capture steep features of the dose-response curve for larger n and high drug dosages. These results suggest that the model may lose accuracy at high dosages when drug combinations involve drugs with identical mechanisms of action and steep dose-response curves. In practice, we find that dose response curves rarely have $n > 2$, and furthermore, the method works well even when drugs have similar–but not identical–modes of action (See Dox-Ery-Linc combo in main text, Figure 2). Therefore, this theoretical limitation is unlikely to be relevant in most practical situations.

2.3 Failure and Success of Bliss Independent Model

While our pairwise model performs significantly better, on the whole, than the Bliss independent model, we found that some combinations of three drugs may nevertheless be appropriately modeled with Bliss independence. Figure S9 compares predictions from Bliss independence (left) with those from the pairwise model (right) for two 3-drug combinations. In the top drug combination (Cm-Ofl-Sal), the pairwise approximation significantly outperforms the independent model. On the other hand, in the lower panels (Dox-Ery-Linc), the results from both models are highly correlated ($r \approx 0.95$) and both provide reasonable fits to the data. The latter result is particularly interesting given the strong interactions that take place between doxycycline-lincomycin (strong suppression) and doxycyline-erythromycin (strong synergy) when used in pairs (see Figure 2).

2.4 Akaike Information Criteria and Model Selection

To statistically compare the pairwise model with the independent model, we use standard model-selection techniques [40] (see Table S1 for results). Specifically, we assume that the experimental errors are independent and Gaussian distributed with unknown variance σ^2. We confirm approximate normality of residuals in Figure S10. We then calculate for each model the Akaike Information Criteria, which is given by

$$\mathrm{AIC} = -2\log(\mathcal{L}(\hat{c}|y)) + 2n \tag{S11}$$

where $\log(\mathcal{L}(\hat{c}|y))$ is the log likelihood function, y is the data, c is maximum likelihood estimate of the free parameters of the model (in this case, σ^2), and n is the number of free parameters ($n = 1$ for both models, corresponding to the unknown error variance). The AIC is an estimate of the expectation value of the relative Kullback-Leibler (KL) divergence between the fitted model and the "true mechanism" generating the observed data. The model with the

lowest AIC value among a set of models is considered the best model in that it minimizes the KL divergence between the model and statistical mechanism underlying the data. For independent Gaussian errors, AIC reduces (up to an additive constant) to

$$\mathrm{AIC} = -N \log(\hat{\sigma}^2) + 2n, \tag{S12}$$

where N is the number of observations and $\hat{\sigma}^2$ is the maximum likelihood estimate of the variance. In practice, we use a small sample estimator of AIC that includes a bias correction term

$$\mathrm{AIC} = -2 \log(\mathcal{L}(\hat{c}|y)) + 2n + \frac{2n(n+1)}{N-n-1}. \tag{S13}$$

The differences in AIC values between the pairwise model and the Bliss independent model can be converted to an Akaiki weight in favor of the pairwise model,

$$w = \frac{\exp(-\delta/2)}{\exp(-\delta/2)+1} \tag{S14}$$

where $\delta \equiv \mathrm{AIC}_{\mathrm{pair}} - \mathrm{AIC}_{\mathrm{ind}}$. Because $\exp(-\delta/2)$ is proportional to the likelihood of the pairwise model given the data, the weight w can be interpreted as a measure of the evidence in favor of the pairwise model as the best of the two models.

2.5 Predictions of 3-Drug and 4-Drug Effects

Figures S11 - S15 show predictions for three-drug (Figures S11 - S14) and four-drug (Figure S15) combinations calculated using the maximum entropy distributions (or, equivalently, using Equation S7). Each figure includes heat maps comparing experimental growth to theoretical predictions (left hand side) as well as a direct comparison of predictions vs. experiments.

2.6 Combinatorial Experiments Testing 3-Drug Predictions

In addition to exploring the entire space of 3-drug concentrations for the drug combinations listed above, we have also performed combinatorial experiments to test the predictions of our model on a broad range of 3-drug combinations, each at a single dosage. Each combinatorial experiment involves N drugs, each at a single concentration, $D_1, D_2, ...D_N$. In each experiment, we test all $\binom{N}{3}$ possible 3-drug combinations and compare the experimental results to predictions from our pairwise model. We choose N to be 5, 6, or 7 and performed 5 combinatorial experiments yielding a total of 93 unique 3-drug combinations and 120 unique dosage combinations.

Table S3 lists all drug combinations, and the corresponding comparisons between predictions and experiment are shown in Figures S16, S17 (inset, which includes error bars). The pairwise model performs remarkably well ($R^2 = 0.95$)

and significantly outperforms the naive independence model ($R^2 = 0.29$), which demonstrates the need to account for pairwise interactions.

To estimate the frequency of pure 3-body interactions, we also include a histogram (Figure S17, main figure) of the statistical deviations from the pairwise predictions. These deviations, which cannot be statistically explained by the pairwise approximation, occur when the 95 percent confidence interval of the difference $\delta = g_{exp} - g_{pred}$, where g_{exp} is the relative growth from experiment and g_{pred} is the predicted relative growth, does not contain 0. The difference between the boundary of this confidence interval and 0 is defined to be the deviation, ΔI_3 (units are relative growth rate); this deviation may arise from pure 3-drug interactions. In 74 of the 120 drug combinations, the deviation is zero ($\Delta I_3 = 0$). In the remaining 46 combinations, the deviations (unexplained drug interactions) are very small (mean$= 0.034 \pm 0.005$), with the maximum of $\Delta I_{3,max} = 0.12$.

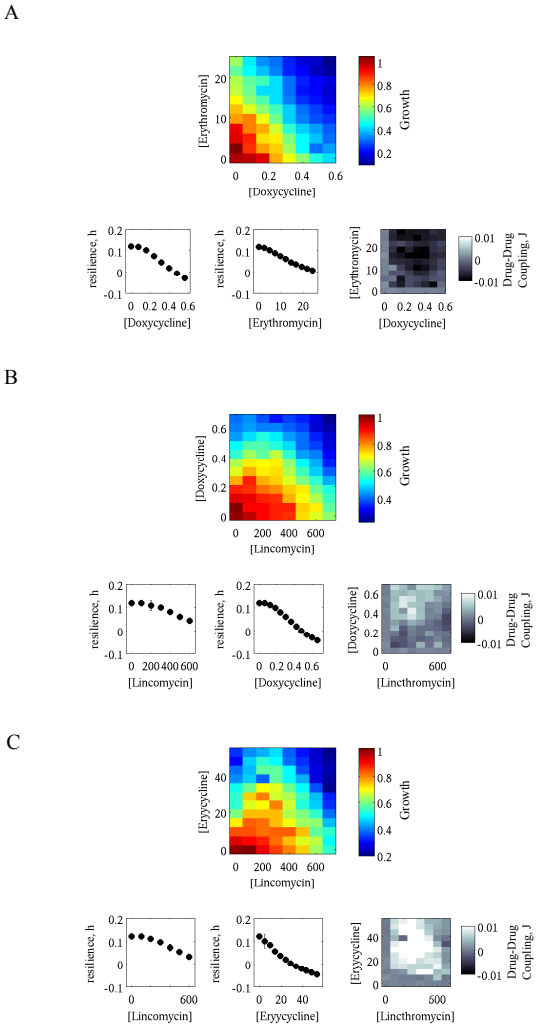

Figure S2: Experimentally determined growth rates, resilience coefficients (h), and coupling coefficients (J), of maximum entropy distribution for pairwise drug interactions. Growth rate data and maximum entropy coefficients for drug pairs (A) Doxycycline-Erythromycin (synergistic), (B) Doxycycline-Lincomycin (weakly antagonistic), and (C) Erythromycin-Lincomycin (strongly antagonistic). In each panel, top plots show heat maps of cell growth in the presence of two drugs. Cell growth is normalized by growth in the absence of drugs. Warmer colors indicate high growth rates, whereas cooler colors indicate slower growth rates. Bottom left, resilience coefficients, h, as a function of each drug in the combination. Decreasing the resilience coefficient, h, corresponds to a decrease in growth rate. Error bars: standard error of replicates (smaller than data points). Bottom right, drug-drug coupling coefficients, J, as a function of drug concentration for each drug pair. $J > 0$ corresponds to antagonism, $J < 0$ to synergy, and $J = 0$ to additivity.

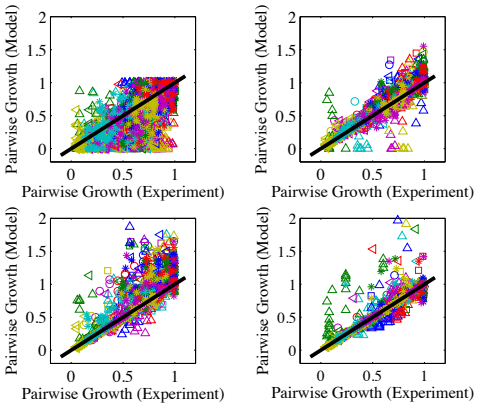

Figure S3: Fitting two-drug Data Using State Spaces with $a = 0$, $b > 0$. Upper left, $b = 1$, upper right, $b = 3$, lower left, $b = 5$, lower right, $b = 5$. Different symbols represent growth of cells in response to drug pairs drawn from different three-drug combinations (Sal-Ery-Cm, squares; Cm-Ery-Tmp, circles; Cm-Ofl-Sal, upright triangles; Cm-Ofl-Tmp, leftward triangles; Dox-Ery-Linc, stars). Black lines, line of slope 1 indicating perfect fit. Note that many data points in the lower right panel fall outside of the range of the plots.

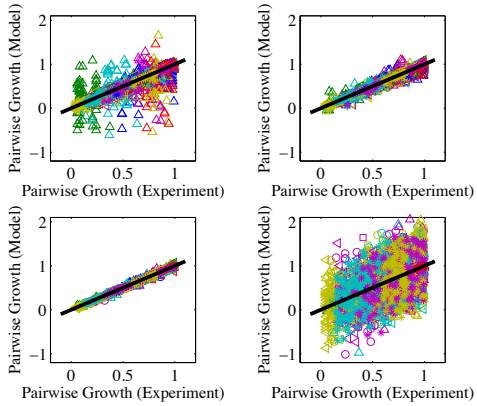

Figure S4: Fitting two-drug Data Using State Spaces with $a < 0$, $b > 0$. Upper left, $(a, b) = (-0.5, 1.5)$, upper right, $(a, b) = (-2, 3)$, lower left, $(a, b) = (-3, 4)$, lower right, $(a, b) = (-19, 20)$. Different symbols represent growth of cells in response to drug pairs drawn from different three-drug combinations (Sal-Ery-Cm, squares; Cm-Ery-Tmp, circles; Cm-Ofl-Sal, upright triangles; Cm-Ofl-Tmp, leftward triangles; Dox-Ery-Linc, stars). Black lines, line of slope 1 indicating perfect fit.

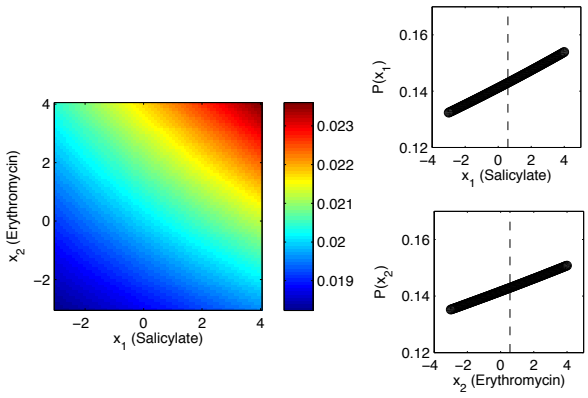

Figure S5: Example Maximum Entropy Distrubitions: Pairwise, $P_2(x_1, x_2) \equiv \int_a^b P(x_1, x_2, x_3)dx_3$ (left panel), and single variable, $P_1(x_1) \equiv \int_a^b P(x_1, x_2, x_3)dx_3dx_2$ (right panels), marginal distributions for the three-drug combination salicylate (2 mM), erythromycin (25μg/mL), and chloramphenicol (0μg/mL). Vertical dashed lines indicate averages $\langle x_i \rangle$, which correspond to single drug growth rates g_i. Drugs are arbitrarily labeled as 1 (salicylate), 2 (erythromycin), and 3 (chloramphenicol).

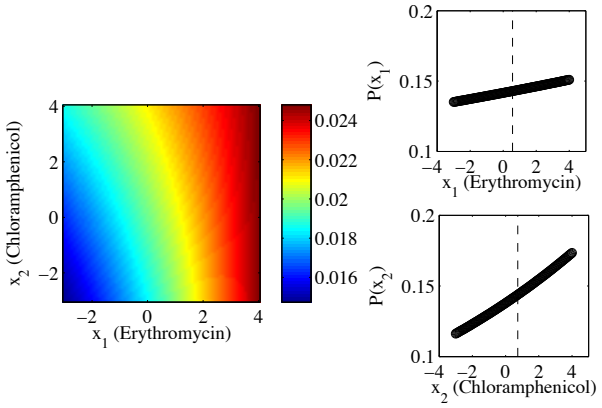

Figure S6: Example Maximum Entropy Distrubitions: Pairwise, $P_2(x_1, x_2) \equiv \int_a^b P(x_1, x_2, x_3)dx_3$ (left panel), and single variable, $P_1(x_1) \equiv \int_a^b P(x_1, x_2, x_3)dx_3dx_2$ (right panels), marginal distributions for the three-drug combination salicylate (0 mM), erythromycin (25μg/mL), and chloramphenicol (1μg/mL). Vertical dashed lines indicate averages $\langle x_i \rangle$, which correspond to single drug growth rates g_i. Drugs are arbitrarily labeled as 1 (erythromycin), 2 (chloramphenicol), and 3 (salicylate).

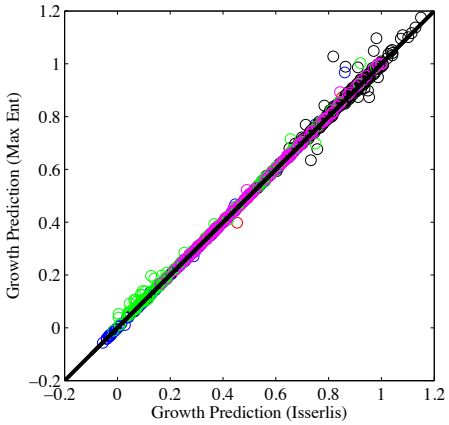

Figure S7: Comparison of Moment Relationships given by Maximum Entropy and Isserlis' Theorem: Predictions of growth in the presence of three-drug (3rd order moments) based on maximum entropy (x axis) and Isserlis' theorem (y-axis). Different colors represent different three-drug combinations.

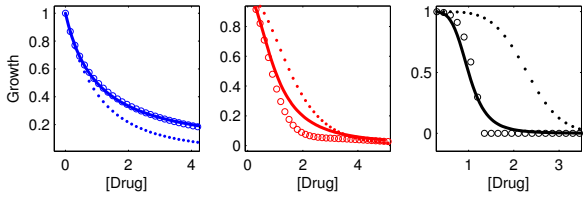

Figure S8: Drug With Itself. Predictions from Isserlis' equation (open circles) and Bliss independence (solid circles) for Hill function dose response curves (solid line) with Hill coefficients $n = 1$ (left), $n = 2$ (center), and $n = 5$ (right).

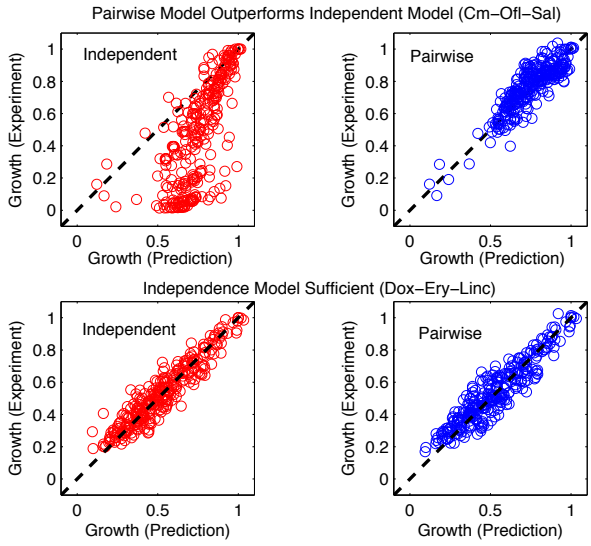

Figure S9: Comparison between independent model (left) and pairwise model (right) for two three-drug combinations: chloramphenicol-ofloxacin-salicylate (top) and doxycycline-erythromycin-lincomycin (bottom).

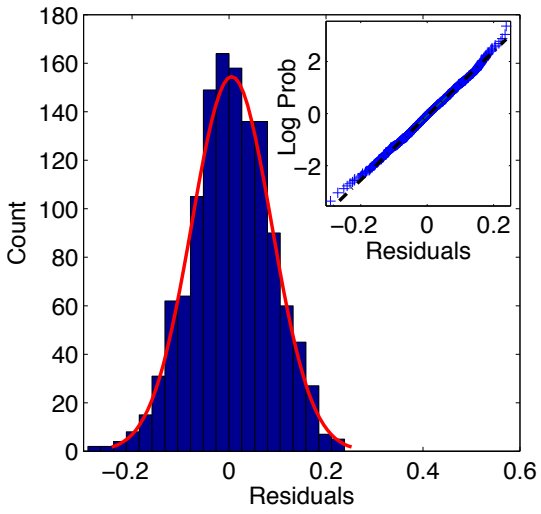

Figure S10: Verifying Normality of Residuals. Main figure, Histogram of Residuals from Pairwise Model (all drug combinations); red line, fit to normal distribution. Inset: Normal Probability Plot (straight line indicates normality).

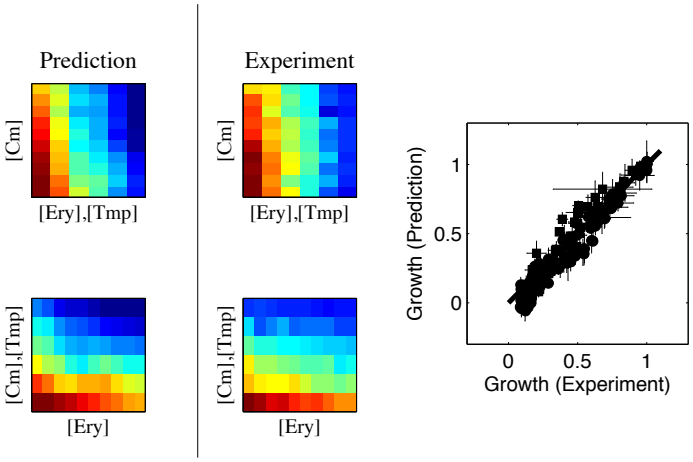

Figure S11: Comparison of Predictions with Experiments for the three-drug combination Cm-Ery-Tmp.

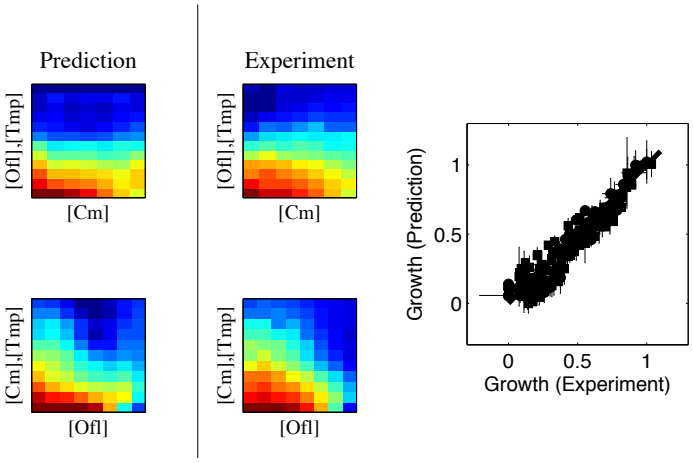

Figure S12: Comparison of Predictions with Experiments for the three-drug combination Cm-Ofl-Tmp

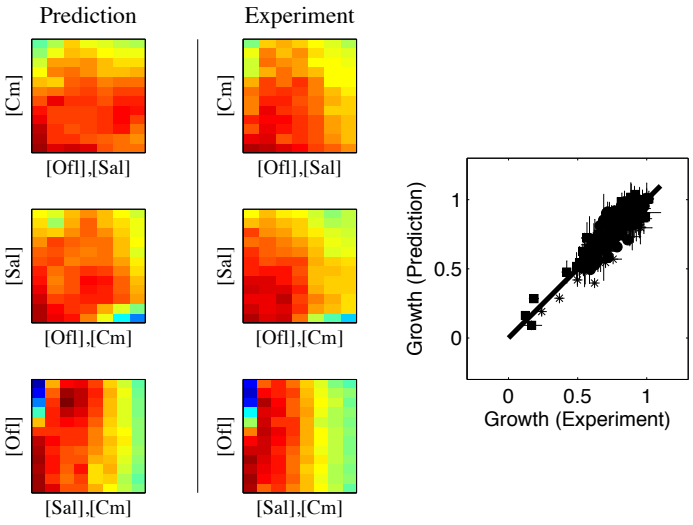

Figure S13: Comparison of Predictions with Experiments for the three-drug combination Cm-Ofl-Sal

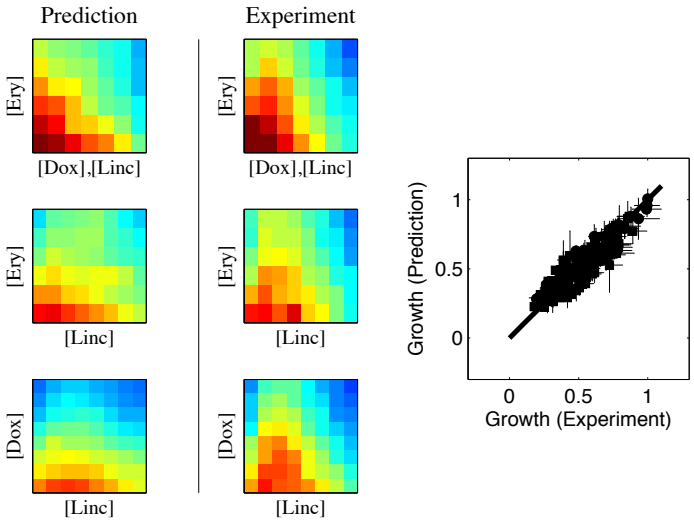

Figure S14: Comparison of Predictions with Experiments for the three-drug combination Dox-Ery-Linc

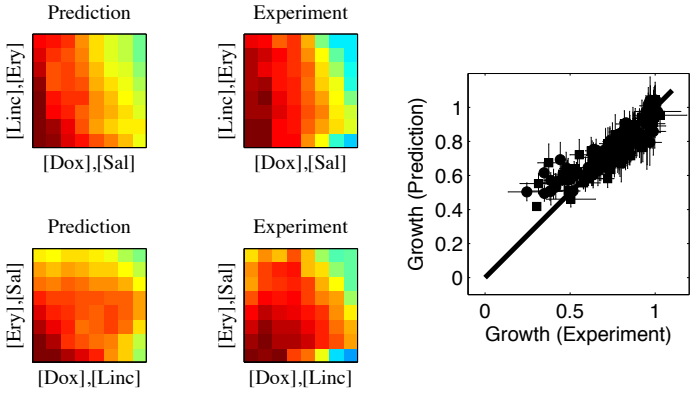

Figure S15: Comparison of Predictions with Experiments for the four-drug combination Dox-Ery-Linc-Sal

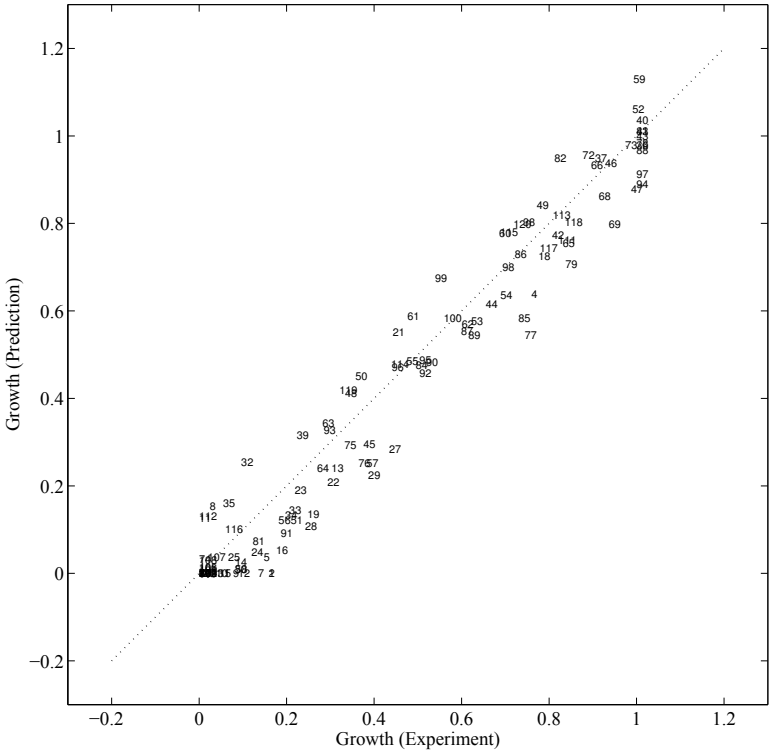

Figure S16: Comparison of Predictions with Experiments for the 3-drug Combinatorial Experiments. Each number corresponds to a 3-drug combination from the table at the end of the SI material.

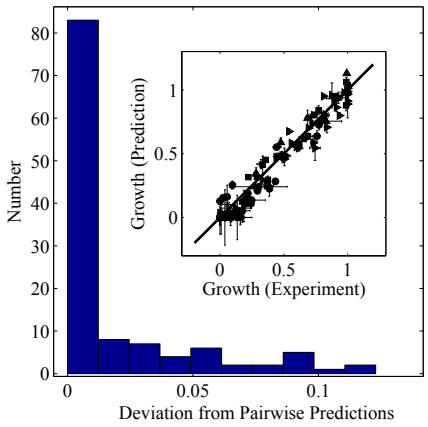

Figure S17: Histogram of Deviations from Pairwise Predictions. Deviations from the pairwise predictions occur when the 95 percent confidence interval of the difference $\delta = g_{exp} - g_{pred}$, where g_{exp} is the relative growth from experiment and g_{pred} is the predicted relative growth, does not contain 0. The difference between the boundary of this confidence interval and 0 is defined to be the deviation from pairwise predictions (units are relative growth rate). Inset: Comparison of Predictions with Experiments for the 3-drug Combinatorial Experiments. Error bars are ± standard error.

References

[33] Kapur, J N (2009), *Maximum Entropy Models in Science and Engineering*, New Age International Publishers, New Delhi.

[34] The standard Shannon measure of entropy (see, for example, [19] is a special case of this Bayesian entropy when the prior distribution is uniform.

[35] S. Kullback (1959), *Information theory and statistics*, John Wiley and Sons, NY.

[36] The traditional Ising model defines interactions between discrete spins taking values of 1 or -1. Also, it is typically assumed that the $J_{ij} = J$, independent of spin. Our model, on the other hand, resembles a continous-spin version of the disordered Ising model where both the fields h_i and the couplings J_{ij} differ for each spin.

[37] Mead, L and N Papanicolaou (1984), *Maximum Entropy in the Problem of Moments*, J. Math. Phys. 25, 2404.

[38] Isserlis, L (1916), *On Certain Probable Errors and Correlation Coefficients of Multiple Frequency Distributions with Skew Regression*, Biometrika 11: 185190.

[39] Michalowicz, J, Nichols, J, Bucholtz, J and Olson, C (2009), *A general Isserlis theorem for mixed-Gaussian random variables*, J. Stat. Phys., Volume 136 (1), 89-102.

[40] Burnham, K and Anderson, D (1998). *Model Selection and Multimodel Inference*, 2nd Edition, Springer-Verlag, New York.