

ARTICLE

A strategy for determining the equilibrium constants for heteromeric ion channels in a complex model

Klaus Benndorf¹ , Thomas Eick¹ , Christian Sattler¹ , Ralf Schmauder¹ , and Eckhard Schulz² 

Ligand-gated ion channels are oligomers containing several binding sites for the ligands. However, the signal transmission from the ligand binding site to the pore has not yet been fully elucidated for any of these channels. In heteromeric channels, the situation is even more complex than in homomeric channels. Using published data for concatamers of heteromeric cyclic nucleotide-gated channels, we show that, on theoretical grounds, multiple functional parameters of the individual subunits can be determined with high precision. The main components of our strategy are (1) the generation of a defined subunit composition by concatenating multiple subunits, (2) the construction of 16 concatameric channels, which differ in systematically permuted binding sites, (3) the determination of respectively differing concentration-activation relationships, and (4) a complex global fit analysis with corresponding intimately coupled Markovian state models. The amount of constraints in this approach is exceedingly high. Furthermore, we propose a stochastic fit analysis with a scaled unitary start vector of identical elements to avoid any bias arising from a specific start vector. Our approach enabled us to determine 23 free parameters, including 4 equilibrium constants for the closed-open isomerizations, 4 disabling factors for the mutations of the different subunits, and 15 virtual equilibrium-association constants in the context of a 4-D hypercube. From the virtual equilibrium-association constants, we could determine 32 equilibrium-association constants of the subunits at different degrees of ligand binding. Our strategy can be generalized and is therefore adaptable to other ion channels.

Introduction

The function of receptor proteins is essential for the homeostasis of living cells and, thus, of whole organisms. Receptor proteins are functionally controlled by the binding of ligands to highly specific binding sites. This binding triggers specific conformational changes, evoking secondary responses. Ligands can be simply ions such as Ca^{2+} ions or molecules of highly different size such as diverse neurotransmitters or peptide- and proteohormones. Ionotropic receptors are receptors in which the binding of ligands controls the openness of a pore for ions. Many of these ionotropic receptors are oligomers composed of several identical or homologous subunits. This results in a respective number of binding sites that are formed either by the subunits alone (e.g., ionotropic AMPA receptors [Sobolevsky et al., 2009], cyclic nucleotide-gated channels [Li et al., 2017]) or at interfaces between adjacent subunits (e.g., purinergic P2X receptors [Kawate et al., 2009], nicotinic acetylcholine receptors [daCosta and Baenziger, 2013]). Consequently, the activation of the receptors is controlled by several binding steps. Further, taking into account that the subunits can specifically influence each other upon activation in a cooperative fashion and that the

underlying conformational changes reciprocally feed back to the binding steps themselves (Colquhoun, 1998; Kusch et al., 2010), analyzing receptor activation is a highly challenging task.

For decades, the Monod-Wyman-Changeux (MWC) model (Monod et al., 1965) has been applied to interpret cooperative ligand-induced activation in proteins. It was thereby assumed that a symmetric oligomeric protein of identical subunits performs a joint “allosteric” conformational change of all subunits and that only the equilibrium of the allosteric step is shifted in proportion to the number of ligands bound. To keep the model simple, fixed stoichiometric factors are typically used, leaving only one equilibrium constant for ligand association (K), one equilibrium constant for the allosteric conformational change (E_0), and one fixed allosteric factor (f) with changing power (Fig. S1). The MWC model has been widely applied, ranging from oxygen binding to hemoglobin (Eaton et al., 1999) to various ionotropic membrane receptors (e.g., GABA_A receptors [Steinbach and Akk, 2019], nicotinic acetylcholine receptors [Lee and Sine, 2005], cyclic nucleotide-gated channels [Goulding et al., 1994]). However, the uniqueness and accuracy

¹Institute of Physiology II, Jena University Hospital, Friedrich Schiller University Jena, Jena, Germany; ²Schmalkalden University of Applied Sciences, Faculty of Electrical Engineering, Schmalkalden, Germany.

Correspondence to Klaus Benndorf: Klaus.Benndorf@med.uni-jena.de.

© 2022 Benndorf et al. This article is distributed under the terms of an Attribution-Noncommercial-Share Alike-No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms/>). After six months it is available under a Creative Commons License (Attribution-Noncommercial-Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).

of identifiable parameters in state models is often overestimated, as discussed for both electrophysiological data (Colquhoun, 1998) and data on ligand binding (Middendorf and Aldrich, 2017a, b), even when including a Bayesian framework (Hines et al., 2014). Unfortunately, more limitations arise when analyzing heteromeric ion channels containing different binding sites because a much larger number of different equilibrium constants has to be distinguished. As a consequence, thorough functional studies are very rare. In some cases, sophisticated single-channel analyses have shed light into the intricate gating mechanisms (Burzomato et al., 2004; Beato et al., 2007).

Recently, we investigated activation gating in heterotetrameric cyclic nucleotide-gated channels, composed of two CNGA2 subunits, one CNGA4 subunit and one CNGB1b subunit. We analyzed a full set of 16 concatamers with defined functional and disabled binding sites, providing 16 different concentration-activation relationships (CARs), and subjected these CARs to a global fit analysis (Schirmeyer et al., 2021). This allowed us to quantify receptor activation by means of a complex model containing 32 equilibrium association constants, together with disabling factors for the four binding sites and four closed-open isomerization constants. Because of multiple constraints, given by microscopic reversibility and the simplifying assumption that the two A2 subunits are functionally equal, we could reduce the number of free fit parameters to 17, which is still a challenging number. Notably, the precision of our fit, given by a low parameter variance, was still surprisingly high.

This encouraged us to investigate on theoretic grounds why such a high precision could be achieved at all, how many CARs are required, which functional changes mutations should preferably evoke, and how “unnoisy” the data should be. We used the determined parameters of our 16 concatameric CNGA2:A4:B1b channels, containing all possible combinations of a disabling binding site, to simulate data. We then systematically permuted all combinations of CARs at variable noise and number of data points and refitted these simulated data to derive conclusions on the determinateness of the parameters. To minimize the risk of biasing in the fits, we use a stochastic fit approach with a scaled unitary (SU) start vector of identical elements. In the proposed framework, our results show that intimately coupled CARs, carrying well-defined mutations, are a powerful source of information, not only to quantify the effects of the mutations but also to identify the gating operation of the individual subunits in a WT heteromeric channel.

Materials and methods

Data simulation

The 16 CARs for the 16 HA models were simulated by using the data determined for the A4-A2(1)-B1b-A2(2) concatamer (Schirmeyer et al., 2021). The parameters are listed in Table S1.

Global fit

The nlinfit routine of Matlab software, determining parameters by an iterative least square estimation, was used to globally fit

between 1 and 16 CARs. The maximum iteration number was set to 100. The routine provides the 23 fit parameters, x_i ($4 \times Ex, 15 \times Z_{xxxx}, 4 \times fdx$), the covariance matrix, Cov_{EZf} , with the 23×23 elements σ_{ij} of which the main diagonal gives, with $i = j$, the 23 variances and SDs σ_i^2 and σ_i of the fit parameters, respectively.

Relative SDs, $\sigma_{i,rel}$, of the parameters Ex , Z_{xxxx} , and fdx were determined according to

$$\sigma_{i,rel} = \sigma_i / x_i. \quad (1)$$

From the 15 Z_{xxxx} values, specifying the virtual equilibrium association constants in the 4-D hypercube, the 32 K_{xxxx} values, specifying the equilibrium association constants in the scheme of Fig. 1, left, were obtained by the ratios provided by Table S2. To keep track, we transferred the 15 Z_{xxxx} to Z_1-Z_{15} (Table S7) and computed the 32 K_i by $K_i = Z_m/Z_n$ ($i = 1 \dots 32; m, n = 1 \dots 15$) as specified in Table S2. For example, the calculation of $K_5 = Z_8/Z_2$ reads in long-term $K_{0110} = Z_{0110}/Z_{0100}$, where the not-underlined 1 in K_{0110} denotes that the binding of the ligand occurs at the third subunit, whereas the underlined 1 denotes that the second subunit was already occupied previously.

To calculate the SDs σ_{Ki} of the 32 K_i , we computed the covariance matrix Cov_K of the association constants with 32×32 elements from the covariance matrix of Z_i ($i = 1 \dots 15$), Cov_Z , according to

$$Cov_K = B \times Cov_Z \times B_T, \quad (2)$$

using the rules of matrix multiplication. The matrix B is of type 32×15 , and B_T , the corresponding transposed matrix, of type 15×32 . The elements of B are the partial derivatives of the K_i with respect to the Z_j ,

$$B_{ij} = \frac{\partial K_i}{\partial Z_j}, \quad (3)$$

For the example of $K_5 = Z_8 / Z_2$ (Table S2), the two elements in row 5 of matrix B are $B_{52} = \frac{\partial K_5}{\partial Z_2} = -\frac{Z_8}{Z_2^2}$ and $B_{58} = \frac{\partial K_5}{\partial Z_8} = \frac{1}{Z_2}$. All other elements in row 5 of matrix B are 0, because K_5 depends only on Z_2 and Z_8 .

Coming back to the SDs, σ_{Ki} , of the $32 \times K_i$ values, they are then given by

$$\sigma_{Ki} = \sqrt{Cov_{Ki}}. \quad (4)$$

The relative SDs, $\sigma_{Ki,rel}$, of the constants K_i were determined with

$$\sigma_{Ki,rel} = \sigma_{Ki} / K_i. \quad (5)$$

The dimensionless relative SDs, $\sigma_{i,rel}$ or $\sigma_{Ki,rel}$, are useful for comparing the uncertainty of different parameters or equilibrium association constants, respectively. Here, two different measures are used to quantify the uncertainties, “imprecision” and “inaccuracy.”

Imprecision: The term “precision” of a parameter x_i or a derived equilibrium association constant K_i specifies its quality, without any guarantee that the parameter has been determined correctly. Hence, the term imprecision (ipr) denotes a statistical variability with respect to the determined mean. Herein, we use $ipr_i = \sigma_{i,rel}$ for a parameter and $ipr_{Ki} = \sigma_{Ki,rel}$ for an equilibrium association constant.

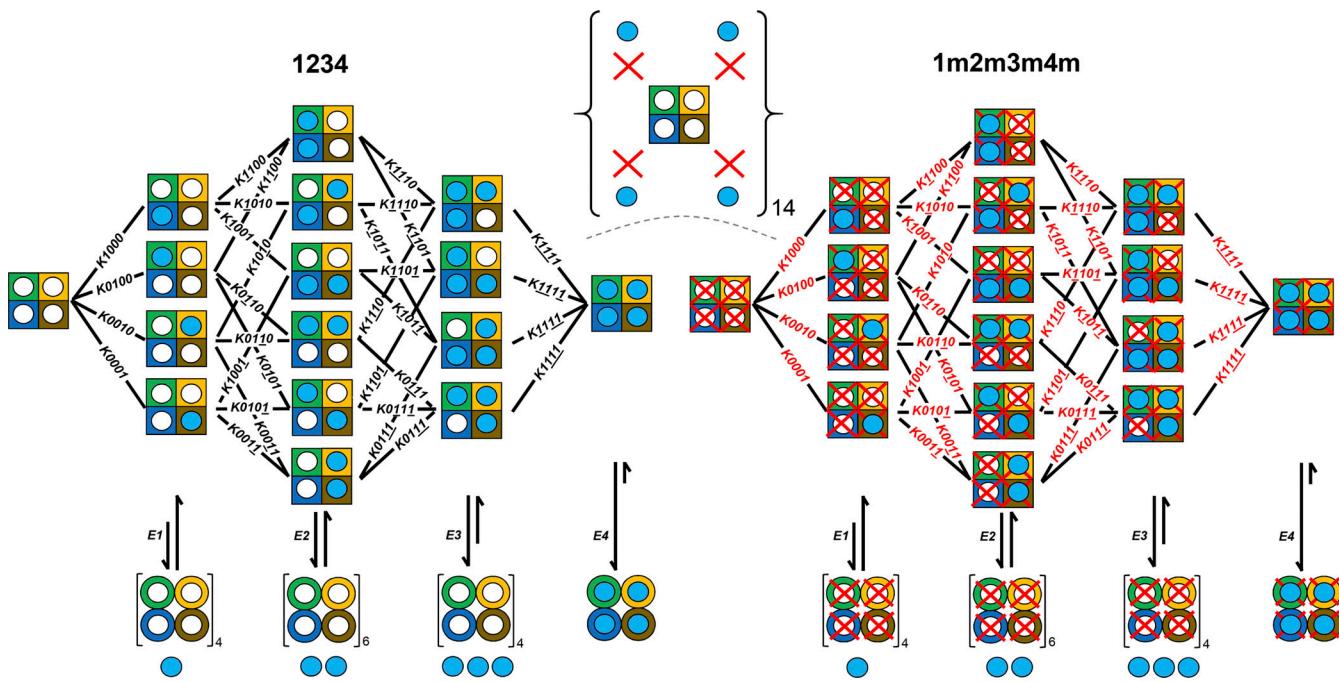


Figure 1. **Structure of the CHA₁₆ model.** 16 HA models for the concatamers 1234 (WT, left) through 1m2m3m4m (quadruple mutated, right) form, together with the 14 other models of Figs. S2, S3, and S4, the complete CHA₁₆ model. In equilibrium association constants for ligand binding, Kxxxx, x can be 0, 1, and 1 for an empty binding site, a binding site to actually occupy, and a preoccupied binding site, respectively. An HA model has 32 Kxxxxs, either black or red for liganding a WT and mutated subunit, respectively. For the maximum CHA₁₆ model, the total number of black and red Kxxxxs is 64. Blue circles, ligands; white circles, empty binding sites; red crosses, disabled binding sites. The set of closed–open isomerizations on the bottom is valid for all HA models. For the single-, double-, and triple-ligated open channel, the ligands are drawn below the channel cartoon to illustrate that there are 4, 6, and 4 options of occupancy (brackets).

Downloaded from http://jgpress.org/jgp/article-pdf/154/6/e202113041/180710/jgp_202113041.pdf by guest on 10 February 2026

The imprecision of the parameters was also used to calculate an overall measure for the goodness of fit based on all n parameters. Three types of means of the imprecision were tested:

The arithmetic mean

$$ipr_{i,am} = \frac{1}{n} \sum_{i=1}^n \sigma_{i,rel}. \quad (6a)$$

$$ipr_{Ki,am} = \frac{1}{n} \sum_{i=1}^n \sigma_{Ki,rel}. \quad (6b)$$

The squared mean (Euclidean distance)

$$ipr_{i,sm} = \left(\frac{1}{n} \sum_{i=1}^n \sigma_{i,rel}^2 \right)^{\frac{1}{2}}. \quad (7a)$$

$$ipr_{Ki,sm} = \left(\frac{1}{n} \sum_{i=1}^n \sigma_{Ki,rel}^2 \right)^{\frac{1}{2}}. \quad (7b)$$

The geometric mean (length of an edge in an n -dimensional hypercube)

$$ipr_{i,gm} = \left(\prod_{i=1}^n \sigma_{i,rel} \right)^{\frac{1}{n}}. \quad (8a)$$

$$ipr_{Ki,gm} = \left(\prod_{i=1}^n \sigma_{Ki,rel} \right)^{\frac{1}{n}}. \quad (8b)$$

Inaccuracy: The term “accuracy” of a parameter specifies the quality of a fitted parameter x_i with respect to the true value, x_{itr} . Herein we use the ratio

$$accuracy_i = x_i / x_{itr}. \quad (9)$$

The term inaccuracy indicates the relative deviation of a parameter or equilibrium association constant from the true value according to

$$\alpha_{i,rel} = \frac{1}{x_{itr}} \sqrt{(x_i - x_{itr})^2} \quad (10a)$$

$$\alpha_{Ki,rel} = \frac{1}{K_{itr}} \sqrt{(K_i - K_{itr})^2}, \quad (10b)$$

respectively. Because we analyzed only simulated data herein, x_{itr} and K_{itr} values are available. The inaccuracy was used as a second measure for the uncertainty of either a parameter or an equilibrium association constant.

To calculate also for the inaccuracy an overall measure for goodness of a fit, we again tested three types of means:

The arithmetic mean

$$iac_{i,am} = \frac{1}{n} \sum_{i=1}^n \alpha_{i,rel}. \quad (11a)$$

$$iac_{Ki,m} = \frac{1}{n} \sum_{i=1}^n \alpha_{Ki,rel}. \quad (11b)$$

The squared mean (Euclidean distance)

$$iac_{i,sm} = \left(\frac{1}{n} \sum_{i=1}^n \alpha_{i,rel}^2 \right)^{\frac{1}{2}}. \quad (12a)$$

$$iac_{Ki,sm} = \left(\frac{1}{n} \sum_{i=1}^n \alpha_{Ki,rel}^2 \right)^{\frac{1}{2}}. \quad (12b)$$

The geometric mean (length of an edge in an n -dimensional hypercube)

$$iac_{i,gm} = \left(\prod_{i=1}^n \alpha_{i,rel} \right)^{\frac{1}{n}}. \quad (13a)$$

$$iac_{Ki,gm} = \left(\prod_{i=1}^n \alpha_{Ki,rel} \right)^{\frac{1}{n}}. \quad (13b)$$

Correlation matrices

To obtain the correlation matrix for the 4 Ex and the 15 Z_i , we first built the respective 19×19 covariance matrix Cov_{EZ} provided by the fit. It is then easy to calculate the 19×19 correlation matrix $Corr_{EZ}$ containing the elements

$$\rho_{ij} = Cov_{EZ}(i,j) / (\sigma_i, \sigma_j) \quad (14a)$$

To obtain the correlation matrix for the 4 Ex and the 32 K_i , we adapted Eq. 2 to build the respective 36×36 covariance matrix Cov_{EK} . Now B (36×19) contains the partial derivatives of the 36 constants with respect to the 19 parameters and B_T (19×36) is the transposed matrix. The elements of the 36×36 correlation matrix $Corr_{EK}$ are given by

$$\rho_{ij} = Cov_{EK}(i,j) / (\sigma_{EKi}, \sigma_{EKj}). \quad (14b)$$

The theoretical determinability of the parameters

In the online supplemental materials, we consider for the case of noiseless data how many CARs and data points are required to determine all parameters of an HA model uniquely. For the example of six CARs with nine data points each, all calculation steps are described. If the data set contains only as many data points as model parameters (fewer than nine data points per CAR), we show that the system of equations to be solved becomes nonlinear. This would require iterative numerical calculations and includes the possibility that more than one solution exists. The right solution might be found by including more data points. Hence, the need for iterative approaches and additional data points arises already for the case of noise-free data if the exact rational functions for $P_o(L)$ cannot be calculated at first. In noisy data, the loss of precision can be compensated for by additional CARs and data points, as demonstrated by the fits herein.

Fitting a Gaussian function to histograms

Stochastically varying parameters were fitted with a Gaussian function according to

$$y = \frac{A}{w_{\alpha i,rel}} \exp \left[\frac{-(x_i - x_{itr})^2}{2w_{\alpha i,rel}^2} \right], \quad (15)$$

where A is an amplitude factor, $w_{\alpha i,rel}$ is the width (SD), x_i is the actual parameter, and x_{itr} is the respective true value.

Calculation of a normalized measure for the mean squared error (MSE)

To refine the search for the best minimum among the successful fits, as identified by containing positive parameters only, they

were sorted with respect to the MSE value, starting with the smallest. In this sequence of MSE values, a normalized measure, $C(MSE_x)$, was calculated for each MSE_x in an iterative manner according to

$$C(MSE_x) = MSE_x \left[\frac{1}{(x-1)} \sum_{n=1}^{x-1} MSE_n \right]^{-1}, \quad (x = 2, 3...) \quad (16)$$

$C(MSE_x)$ reaches 0 if the next MSE_x is identical to the mean of all previous MSE_1 to MSE_{x-1} . $C(MSE_1)$ is 0 and therefore not included on the log scale of Fig. 13 C. $C(MSE_x)$ increases already at very subtle changes of the MSE.

Online supplemental material

Fig. S1 shows the structure of the common MWC model. Figs. S2, S3 and S4 show HA models of the concatamers with indicated disabled binding sites. They build, together with the two models in Fig. 1, the 16 HA models of the CHA model. Fig. S5 shows the comparison of fits with 35 either equidistant or clustered data points. Fig. S6 shows histograms of correlation coefficients obtained by the global fit. Fig. S7 shows effects of data noise on the fit at stochastic variation of the true start vector. Table S1 shows constants used to simulate the 16 CARs. Table S2 provides equilibrium association constants and their genesis. Table S3 provides a comparison of true values, fitted parameters, and errors for an example fit. Table S4 shows a matrix of correlation coefficients for the fit parameters Ex , Z_i , and fdx . Table S5 shows a matrix of correlation coefficients for Ex and K_i . Table S6 summarizes effects of stochastic variation of the true start vector on the fit outcome at different conditions. Table S7 assigns the parameters Z_1-Z_{15} to the 15 Z_{xxxx} . Supplemental text derives on theoretical grounds that noiseless data require already at least six CARs for determining all parameters and that for data with realistic noise the situation is more complex.

Results

All computations were performed with the heteromeric allosteric (CHA_x) model (Fig. 1 together with Figs. S2, S3, and S4), consisting of between 1 and 16 individual HA models differing by the disabling mutations of the binding sites (Fig. 1, top; Schirmeyer et al., 2021). The suffix x in a CHA_x model specifies the combination of the included CARs differing by the disabling mutations (Table 1). Notably, the mutations disable the respective binding sites by only hindering the binding, but not knocking it entirely out; i.e., a disabled binding site can still be occupied by a ligand, but the affinity of the binding site is lowered. For a disabled subunit, the respective equilibrium association constant of the WT subunit is multiplied by a specific disabling factor $fd < 1$. The HA model consists of 16 closed states (C0000...C1111). Among the closed states, a network of 32 transitions with the equilibrium association constants K_{xxxx} is spanned. Here $x = 0$, $x = 1$, and $x = 1$ specify an empty binding site, a binding site that is actually occupied, and a preoccupied binding site, respectively. In the context of our previous work, the four digits denote the subunits A4, A2(1), B1b, and A2(2), counted from the N- to C-terminus of the concatamer. The

Table 1. Schedule of CAR combinations with all 95 combinations of disabled binding sites

No. of CARs included	Specification and #number of mutated CARs																No. of CAR combination
	#0	#1					#2					#3					#4
		CHA ₁	CHA ₂	CHA ₃	CHA ₄	CHA ₅	CHA ₆	CHA ₇	CHA ₈	CHA ₉	CHA ₁₀	CHA ₁₁	CHA ₁₂	CHA ₁₃	CHA ₁₄	CHA ₁₅	CHA ₁₆
16	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1110	0111	1101	1011	1111	95
15	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1110	0111	1101	1011		94
14	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1110	0111	1101			93
	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1110	0111	1101	1011		92
	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1110	0111	1101	1011		91
	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	0111	1101	1011	1011		90
	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1110	0111	1101			89
13	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1110	0111	1101			88
	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1110	0111	1101			87
	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1110	0111	1011			86
	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	0111	1011	1011			85
	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1101	1011	1011			84
12	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1110					83
	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	0111					82
	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	1101					81
	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011					1011	80
11	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011						79
10	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101							78
	0000	1000	0100	0010	0001	1100	1010	1001	0110		0011						77
	0000	1000	0100	0010	0001	1100	1010	1001		0101	0011						76
	0000	1000	0100	0010	0001	1100	1010		0110	0101	0011						75
	0000	1000	0100	0010	0001	1100		1001	0110	0101	0011						74
9	0000	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011						73
	0000	1000	0100	0010	0001	1100	1010	1001	0110								72
	0000	1000	0100	0010	0001	1100	1010	1001		0101							71
	0000	1000	0100	0010	0001	1100	1010		0110	0101							70
	0000	1000	0100	0010	0001	1100		1001	0110	0101							69
	0000	1000	0100	0010	0001		1010	1001	0110	0101							68
	0000	1000	0100	0010	0001	1100	1010	1001			0011						67
	0000	1000	0100	0010	0001	1100	1010		0110		0011						66
	0000	1000	0100	0010	0001		1010	1001	0110		0011						65
	0000	1000	0100	0010	0001			1010	1001		0011						64
	0000	1000	0100	0010	0001	1100	1010			0101	0011						63
	0000	1000	0100	0010	0001	1100		1001		0101	0011						62
	0000	1000	0100	0010	0001		1010	1001		0101	0011						61
	0000	1000	0100	0010	0001	1100			0110	0101	0011						60
	0000	1000	0100	0010	0001		1010		0110	0101	0011						59
	0000	1000	0100	0010	0001			1001	0110	0101	0011						58
8	0000	1000	0100	0010	0001	1100	1010	1001									57
	0000	1000	0100	0010	0001	1100	1010		0110								56
	0000	1000	0100	0010	0001	1100		1001	0110								55
	0000	1000	0100	0010	0001		1010	1001	0110								54
	0000	1000	0100	0010	0001	1100	1010			0101							53
	0000	1000	0100	0010	0001	1100		1001		0101							52
	0000	1000	0100	0010	0001		1010	1001		0101							51
	0000	1000	0100	0010	0001	1100			0110	0101							50
	0000	1000	0100	0010	0001		1010		0110	0101							49
	0000	1000	0100	0010	0001			1001	0110	0101							48
	0000	1000	0100	0010	0001	1100	1010	1001			0011						47
	0000	1000	0100	0010	0001	1100	1010		0101		0011						46
	0000	1000	0100	0010	0001		1010	1001		0101	0011						45
	0000	1000	0100	0010	0001			1010		0101	0011						44
	0000	1000	0100	0010	0001	1100			0110	0101	0011						43
	0000	1000	0100	0010	0001		1100		0110		0011						42
7	0000	1000	0100	0010	0001		1100		0110	0101	0011						41
	0000	1000	0100	0010	0001			1100		0110		0011					40
	0000	1000	0100	0010	0001		1100		0110	0101	0011						39
	0000	1000	0100	0010	0001			1100		0101	0110	0011					38
	0000	1000	0100	0010	0001				1100		0101	0111					37
	0000	1000	0100	0010	0001					0110		0011					36
	0000	1000	0100	0010	0001				1001			0011					35
	0000	1000	0100	0010	0001					0101							34
	0000	1000	0100	0010	0001	1100					0011						33
	0000	1000	0100	0010	0001					0110	0101						32
6	0000	1000	0100	0010	0001					1001							28
	0000	1000	0100	0010	0001						0101						27
	0000	1000	0100	0010	0001						0110						26
	0000	1000	0100	0010	0001					1010	1001						25
	0000	1000	0100	0010	0001						1001						24
5	0000	1000	0100	0010	0001							0110					23
	0000	1000	0100	0010	0001							0101					22
	0000	1000	0100	0010	0001							0110					21
	0000	1000	0100	0010	0001							0101					20
4	0000	1000	0100	0010	0001								0110				19
	0000	1000	0100	0010	0001								0101				18
	0000	1000	0100	0010	0001								0110				17
	0000	1000	0100	0010	0001								0101				16
3	0000	1000	0100	0010	0001									00			

The combinations span from the WT concatamer 0000 to the quadruple mutated concatamer 1111. The notation of the four digits is that the first (N-terminal) subunit is left and the last (C-terminal) subunit is right. 0 and 1 denote a WT and a mutated binding site, respectively.

model contains 15 corresponding open states (O1000...O1111) if at least one binding site is occupied. The equilibrium constants for the closed-open isomerizations (E_x) were assumed to be equal for an equal degree of liganding, resulting in the four different constants $E1...E4$. The disabling effects of the mutations are given by the factors $f1...f4$ that are specific for each subunit and independent of the occupation of the other subunits. As in the previous report (Schirmeyer et al., 2021), transitions between the open states are not considered explicitly but are actually determined due to microscopic reversibility (Colquhoun and Hawkes, 1995).

In order to generalize our considerations for any other tetrameric allosteric protein, we enabled different ligand binding for all four subunits. According to our previous approach, we used for the analysis a 4-D hypercube specifying 15 virtual equilibrium association constants $Zxxxx$ (Fig. 2, A and B; Schirmeyer et al., 2021). In total, this results in 23 free parameters for the full CHA_{16} model used herein, comprising 4 E_x , and 4 f_x in addition to the 15 $Zxxxx$. The number of 23 free parameters is larger by 6 than that used in our previous study on experimental data (Schirmeyer et al., 2021).

How many CARs are required and useful?

We simulated a data set by using the constants Z_i , E_x , and f_x determined experimentally from A4-A2-B1b-A2 concatamers of heterotetrameric CNG channels (Schirmeyer et al., 2021). The parameters are provided in Table S1. To these computed noiseless data, we added noise whose amplitude was proportional to $P_o(1 - P_o)$, generating a maximum amplitude at $P_o = 0.5$, as is typical for channel noise. For the fit, the data points were weighted respectively. Subsequently, these data were globally fitted with a Levenberg–Marquardt algorithm (see Materials and methods).

We started with the maximal fit constraints in the CHA_{16} model (1111; CAR combination 95) and considered first the effect of systematically reducing the CAR number to reach finally 1 CAR in the $CHA_1 = HA$ model (0000; CAR combination 1), thereby determining the fit quality for all 95 possible CAR combinations (Table 1). An example of a fitted CHA_{16} model is shown in Fig. 3. Consider first the values of the fit parameters Z_i and E_x (Fig. 4 A). It becomes immediately obvious that the parameters are poorly determined at combinations with low CAR numbers. Here, many fits fail as indicated by the large peaks. The situation progressively improves toward combinations with higher CAR numbers. We reproduced this 20 times and always obtained very similar results. On theoretical grounds, we state that noiseless data, which are unrealistic in experiments, require at least six CARs for determining all parameters (see Supplemental materials at the end of the PDF). In noisy data, however, the situation is more complex.

To compare the SDs σ_i of the parameters provided by the fit, relative SDs $\sigma_{i,rel}$ were determined by σ_i/x_i . These relative SDs, $\sigma_{i,rel}$, are dimensionless. They allowed us to compare the precision of the determined parameters, x_i . $\sigma_{i,rel}$ is therefore termed here imprecision, ipr_i , with $ipr_i = \sigma_{i,rel}$ (see also Materials and methods). The imprecision for the fit shown in Fig. 3 is shown in Fig. 4 B, further confirming that the goodness of the fit increased with the number of the included CARs. For the example of CAR combination 95 (CHA_{16}), the determined parameters, x_i , and the relative errors, $ipr_i = \sigma_{i,rel}$, are also listed in Table S3.

We next computed from the 15 Z_i the 32 K_i (Fig. 4 C), according to the relationships provided by Table S2, and also the respective imprecisions, ipr_{Ki} (see Materials and methods;

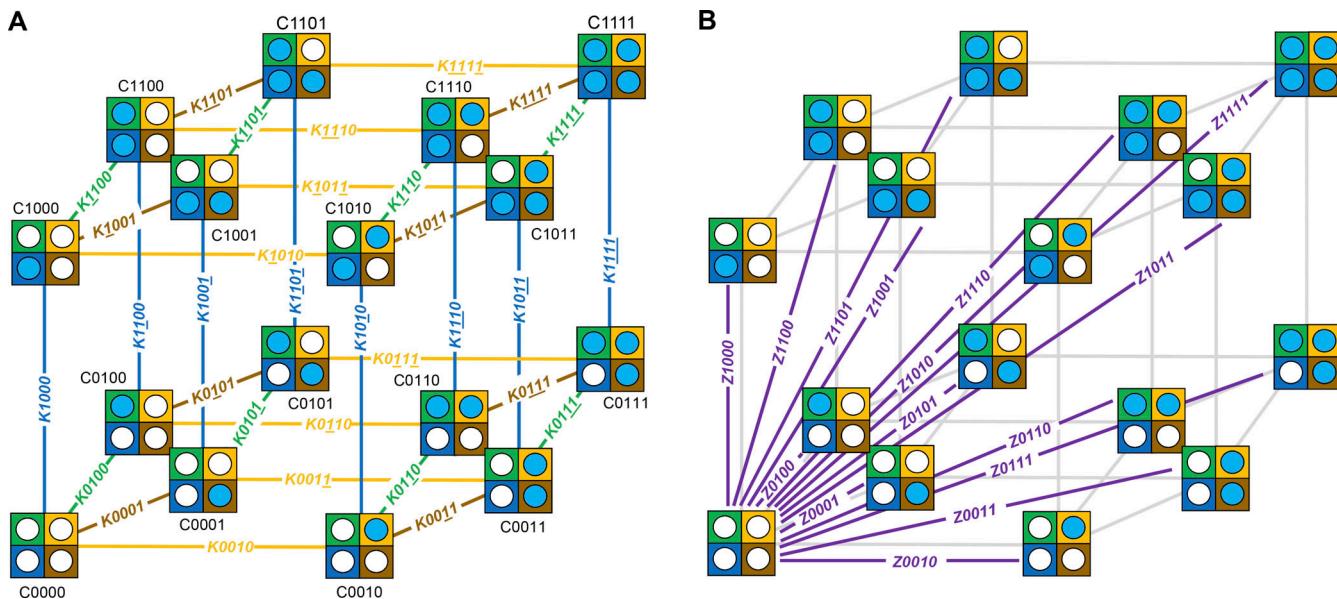


Figure 2. Analysis of the association constants by a 4-D hypercube. (A) Scheme of the 4-D hypercube. The 16 closed states (C0000...C1111) of a HA model are the corners of a 4-D hypercube in which the $Kxxxx$ denote the 32 edges ($x = 0$ for an empty subunit, $x = 1$ for a subunit to be occupied, and $x = 1$ for a preoccupied subunit). The colored lines indicate liganding of the respective subunits. **(B)** Virtual equilibrium association constants $Z0001...Z1111$ (violet lines) specifying the independent parameters.

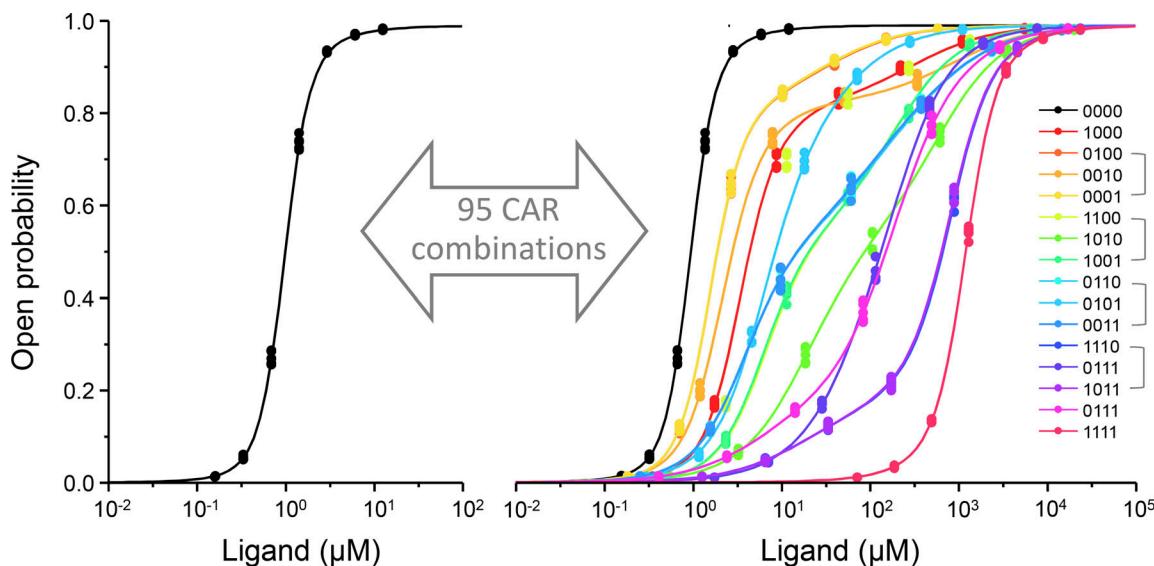


Figure 3. **Fit of 16 CARs with the CHA₁₆ model.** 7 ligand concentrations and 5 data points at each ligand concentration were chosen, summing up for the 16 CARs to 560 data points. The noise level was set to 0.25. The conditions were our standard conditions. The values for Z_i and Ex as well as K_i are provided by Fig. 4, A and C, respectively. The fdx values (imprecisions) were $fd1 = 7.022$ (0.0354), $fd2 = 7.42 \times 10^{-3}$ (0.0289), $fd3 = 1.11 \times 10^{-4}$ (0.0357), and $fd4 = 7.30 \times 10^{-3}$ (0.0262). The brackets in the legends indicate closely similar data points and curves. The big gray double arrow illustrates that the CAR combination is varied.

Fig. 4 D). Accordingly, the K_i values are determined correctly when including 12 or more CARs and, notably, the imprecisions even further decrease in the direction to CHA₁₆, varying over about two orders of magnitude between 0.003 and some more than 0.3. Nevertheless, it is remarkable that for high numbers of included CARs (≥ 12) all K_i are determined, as are also all Ex (Fig. 4, C and D) and fdx (see legend to Fig. 3).

Subsequently, the $ipr_i = \sigma_{i,rel}$ values were used to derive an easy-to-handle overall mean reporting the goodness of the fits. We tested three different means: the arithmetic mean, $ipr_{i,am}$ (Eq. 6a), the squared mean, $ipr_{i,sm}$ (Eq. 7a), and the geometric mean, $ipr_{i,gm}$ (Eq. 8a). For the fit shown in Figs. 3 and 4, A–D, the plot of these means along the 95 CAR combinations shows that all three means provide a useful measure in order to demonstrate the advantage of including 12, or better, 16 CARs in the fits (Fig. 5 A).

Our analysis of simulated data allowed us also to compute a second measure for the goodness of a fit, the inaccuracy. Based on the convention that accuracy of a parameter specifies the quality of a parameter with respect to the true value, x_i/x_{itr} (Eq. 9), the term inaccuracy indicates the relative deviation of the parameters from their true value according to Eq. 10a. Thus, in analogy to the imprecision, the inaccuracy $iac_i = \alpha_{i,rel}$ can be used as a second measure for the uncertainty of the fit. The inaccuracy is also dimensionless (see Materials and methods). To derive an appropriate easy-to-handle overall mean for the inaccuracy, we again tested the three different means of $\alpha_{i,rel}$: the arithmetic mean, $iac_{i,am}$ (Eq. 11a), the squared mean, $iac_{i,sm}$ (Eq. 12a), and the geometric mean, $iac_{i,gm}$ (Eq. 13a). For the fit shown in Figs. 3 and 4, A–D, the plot of these values along the 95 CAR combinations indicates that, in analogy to the imprecision, also these three means provide a useful measure for demonstrating the advantage of including 12, or better, 16 CARs in the fits (Fig. 5

B). In order to compare the quality of global fits herein, we decided to use in the following preferentially the squared means, $ipr_{i,sm}$ and $iac_{i,sm}$. This avoids erroneous distortions possibly arising from either negative ($ipr_{i,am}$, $iac_{i,am}$) or zero values ($ipr_{i,gm}$, $iac_{i,gm}$) of any parameters.

Effect of the number of included data points

In the above analysis, the observed decrease of the uncertainty of the fits in the direction toward higher CAR numbers could have been caused by both an increased number of data points and different effects of the mutations. To distinguish between the two effects, we aimed to analyze the uncertainties at a constant number of data points in all selected CARs, meaning that CAR combination 1 (0000; WT) contains 16 times the number of data points compared to CAR combination 95 (1111). Therefore, we tested to use the same number of $35 \times 16 = 560$ data points in all CARs. To avoid unwanted grouping effects when calculating the number of data points in clustered data, appearing when successively increasing the number of data points per CAR, we chose to perform this analysis with equidistant data in the CARs. We therefore first tested if the results with equidistant data match those obtained with clustered data as shown in Fig. 3. As expected, both imprecision and inaccuracy dropped similarly for equidistant and clustered data in the direction towards increased CAR numbers (Fig. S5).

This allowed us to compare fits with a constant number of 560 data points in all selected CARs (Fig. 6, A and B), with the condition that the successive reduction of CARs reduces the data point number in steps of 35 data points (Fig. 6 C). As expected, the relationships coincide when 16 CARs were included with all 560 data points in both relationships. When progressively decreasing the CAR number from 16 down to 1, the imprecision increased similarly. Only at intermediate CAR numbers, there seemed to be

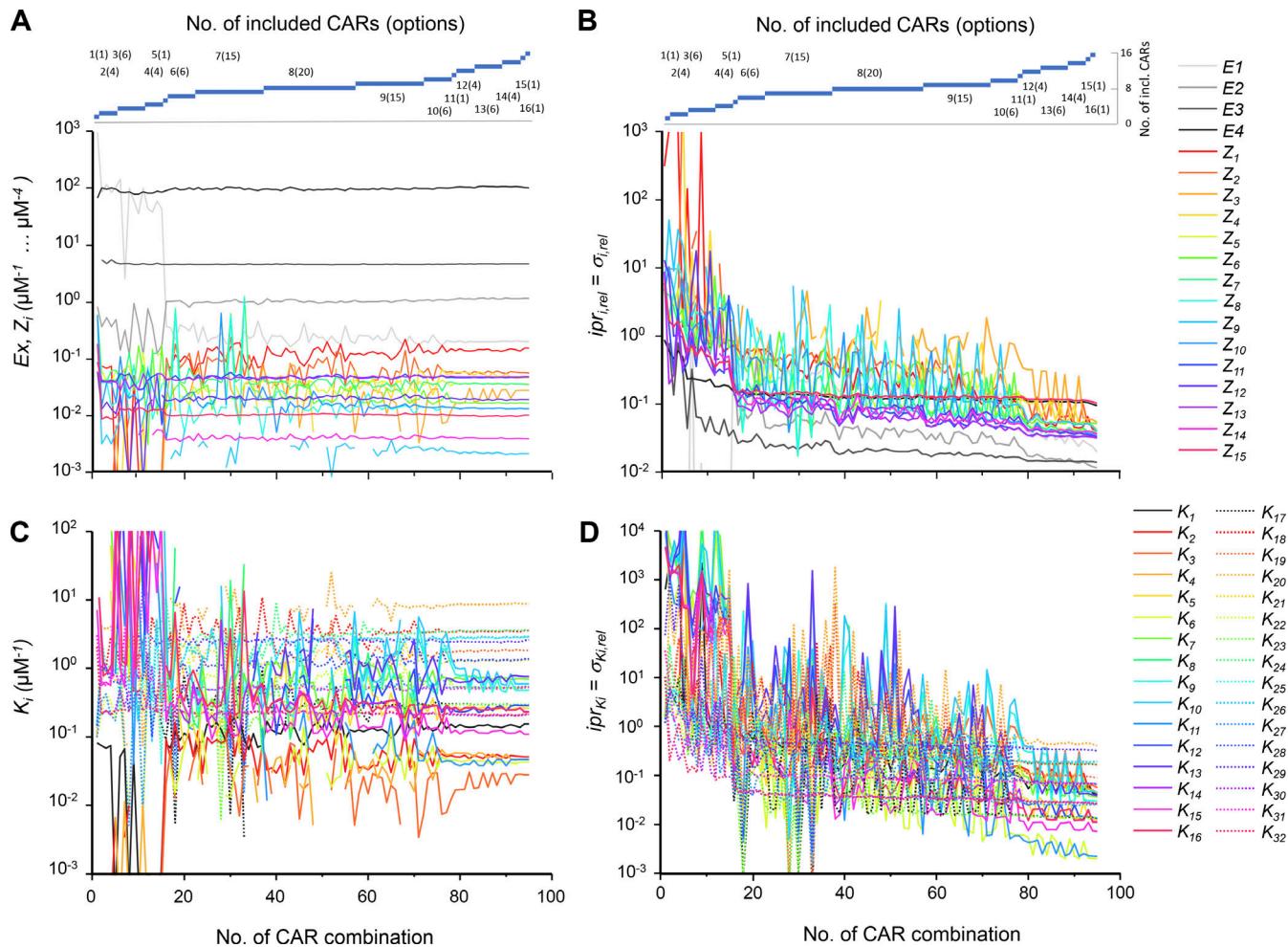


Figure 4. Effect of inclusion of different CAR numbers on Z_i , Ex , and K_i . Standard conditions. The abscissas indicate the number of the CAR combination according to Table 1. (A) Z_i and Ex as functions of the included CARs. The dimensions of Z_1 – Z_4 , Z_5 – Z_{10} , Z_{11} – Z_{14} , and Z_{15} are μM^{-1} , μM^{-2} , μM^{-3} , and μM^{-4} , respectively. The Ex are dimensionless. (B) $\sigma_{i,rel}$ of Z_i and Ex as functions of the included CARs. Including more CARs decreases the imprecision of the fit. (C) K_i as function of the included CARs. The K_i were computed from the Z_i values according to the ratios specified in Table S2. (D) $\sigma_{i,rel}$ of K_i as a function of the included CARs (see also Materials and methods).

a slightly better precision with the constant data point number of 560 (inset in Fig. 6 C). Despite this small difference, the lower uncertainty at increased numbers of included CARs remains when the total data point number was kept constant, suggesting that this effect is mainly caused by the effect of the mutations.

Effect of disabling mutations on the uncertainties

To quantify the effects of the mutations on the goodness of the global fit, we scanned a wide range of mutational effects while leaving the number of fit points constant. Practically, we started with no mutational effect, i.e., the 16 CARs were equal apart from the stochastic noise of the data points (left superimposed relationship in Fig. 7 A). We then systematically decreased each fdx 110 times by a constant factor such that it reached after 100 steps the value used for our standard conditions (top scheme in Fig. 7, B and C; cf. legend to Fig. 3).

The plots along the 95 CAR combinations show that the K_i were determined properly when the four fdx values were reduced about 60 times (Fig. 7, B and C), although moderate

improvement of the uncertainty of some K_i was still obtained when decreasing the fdx values further. Fig. 8, A and B, show for the mean imprecision and the inaccuracy three plots each (see Materials and methods) to confirm that this result does not depend on the chosen type of mean. Together, this shows that strong mutational effects provide the best results for accurately determining the equilibrium constants of the HA model and that an at least \sim 60-fold increase of the affinity by a mutation is a good idea.

Accuracy of non-failing fits at CAR numbers below 16

Next, we addressed the question to what extent our global fit analysis is successful to determine the true parameters for lesser CARs than 16. To this end, we ran the fits of the 95 CARs maximally 100 times each and counted the number of fits with only positive parameters, thereby again using the start vector with the true values (Fig. 9 A). In line with the results above, low CAR numbers generated only a low incidence of successful fits, whereas combinations with 12 or more CARs (CAR combinations

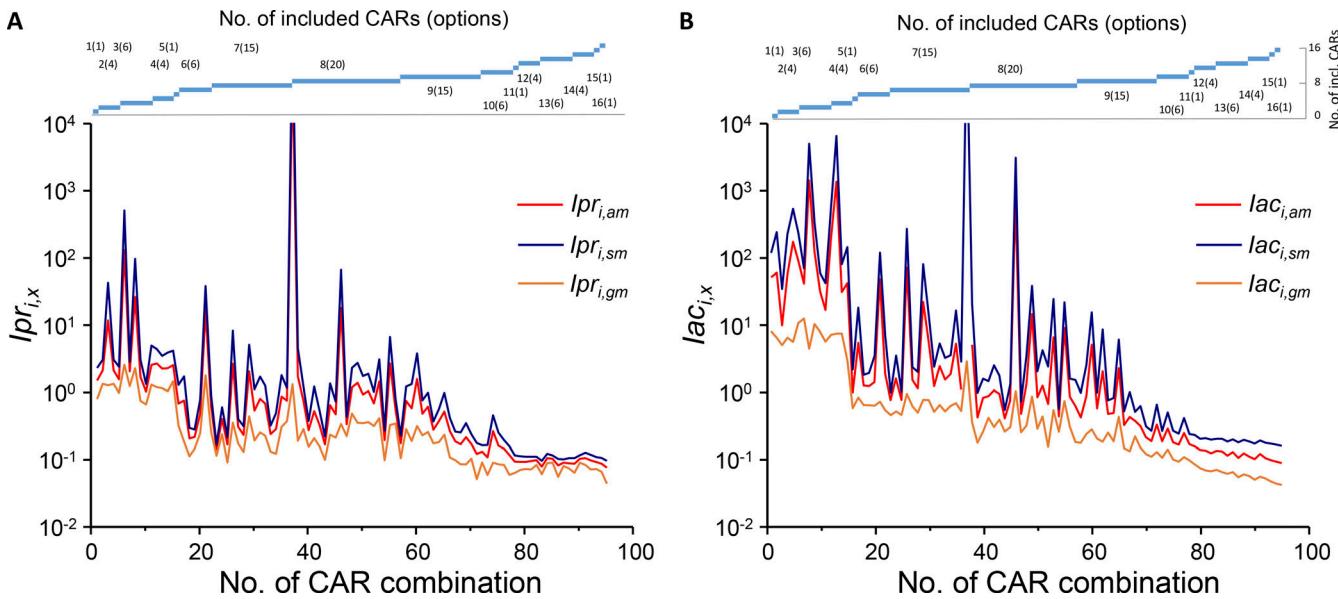


Figure 5. Comparison of three means of both imprecision and inaccuracy. (A) Plot of three mean imprecisions, $ipr_{i,x}$, as functions of the 95 CAR combinations. The values were computed by Eqs. 6a, 7a, and 8a (see Materials and methods). All three means drop similarly if more CARs are included in the fit, indicating that all three means are similarly appropriate for testing the precision of a fit. (B) Plot of the mean inaccuracies, $lac_{i,x}$, as functions of the 95 CAR combinations. The values were computed by Eqs. 11a, 12a, and 13a (see Materials and methods). The mean inaccuracies also drop similarly if more CARs are included, suggesting that these three mean inaccuracies are similarly appropriate for testing the accuracy of a fit.

80–95; cf. Table 1) generated consistently 100% successful fits. It is remarkable that successful fits can also appear exceptionally at CAR combinations with 6 or less CARs (CAR combinations 1–16; cf. Table 1) though the value of 6 is the theoretical border for a successful fit in noiseless data (see Supplemental materials).

We then calculated for the selected fits the inaccuracy for each parameter x_i (Eq. 10a). The plot of the mean inaccuracy (Eq. 12b) of these values as function of the 95 CARs (Fig. 9 B) reveals that the inaccuracy essentially decreases in the direction towards the highest CAR number of 16. Hence, the best determinability is obtained with 16 CARs and a still reasonable determinability with 12–15 CARs.

Accuracy of parameters and the equilibrium association constants

Next, we considered how well our approach determines the 23 fit parameters. With the optimal condition of including 16 CARs, we varied the data points stochastically, and fitted the CAR combination 1,000 times. As read-out we calculated the accuracy as defined by Eq. 9. The distributions of the resulting 1,000 values per parameter were then used to build 23 respective histograms (Fig. 10 A). These histograms are all bell-shaped with a mean near $x_i/x_{itr} = 1$ but differ significantly in their width. To quantify this width, the histograms were fitted by a Gaussian function (Eq. 15; fits not shown), yielding the SD as measure for the goodness of each parameter. A bar graph of these SDs (Fig. 10 B) reveals that all values apart from $E4$ and Z_{15} are <0.1 and that $E1$, $E2$, and $E3$ are smaller than the others. The latter fact makes it promising for the future to distinguish between the equilibrium constants of the closed–open isomerizations.

Because the experimentalist likes to know preferentially the equilibrium association constants K_1 – K_{32} in the kinetic scheme

of Fig. 1 left, we also calculated the respective histograms (Fig. 10 C) by the relations given in Table S2 and determined the SDs accordingly (Fig. 10 D). The obtained standard deviations of K_1 – K_{28} are between 0.08 and 0.14, telling that the accuracy of these constants is still very good. K_{29} – K_{32} are somewhat larger due to the large value of Z_{15} used for their calculation. This indicates an increased vagueness for determining the binding of the fourth ligand. So far, these results demonstrate a robust minimum for our global fit of 16 intimately coupled models which enables to determine the formidable number of 23 parameters x_i if only an appropriate start vector is available.

Correlations

The excellent goodness of our global fit raised the question as to the correlation between the parameters, x_i . We therefore calculated the correlation matrix (Table S4) from the covariance matrix (Eq. 14a). As usual, a perfect correlation has the value 1, a perfect anti-correlation -1 , and no correlation 0. Accordingly, the $n = 23$ parameters themselves produce a correlation coefficient of 1, as indicated in the main diagonal. The remaining $(n + 1)n / 2 - n = 253$ correlation coefficients between all parameters are also illustrated graphically in a color-coded manner (Fig. 11 A). Most correlation coefficients are between -0.3 and $+0.3$, suggesting low correlation (Fig. S6 A). There is only a single pronounced anti-correlation between $E4$ and Z_{15} , where the coefficient is -0.99 .

To relate these considerations better to the state model of Fig. 1, left, we computed a respective correlation matrix for the 4 Ex and 32 K_i by using Eqs. 2 and 14b (see Materials and methods). This resulted in a matrix with $n = 36$ to $(n + 1)n / 2 - n = 630$ correlation coefficients (Table S5 and Fig. 11 B). Also here, most of the correlation coefficients are between -0.3 and

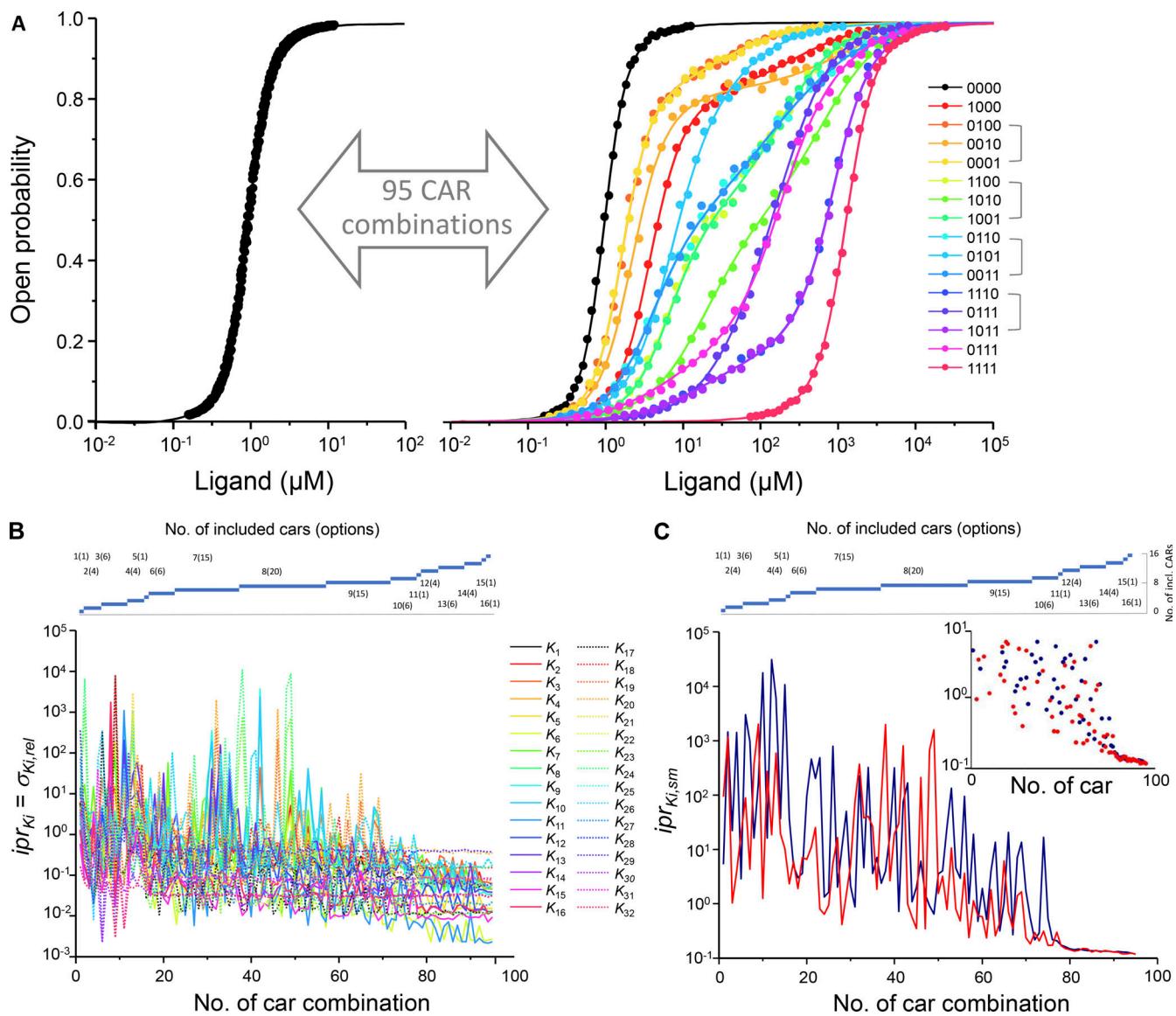


Figure 6. Effects of the data point number on the goodness of the global fit. (A) Fit of 16 CARs with the CHA₁₆ model. 35 equidistant ligand concentrations were used as data points in the CHA₁₆ model, and the number of data points per CAR was increased such that reducing the CAR number kept in total 560 data points. The noise level was set to 0.25. The brackets in the legends indicate closely similar data points and curves. The big gray double arrow illustrates that the CAR combination is varied. (B) Plot of the imprecisions $ipr_{K_i,rel}$ for the 32 K_i values as function of the 95 CAR combinations. (C) Plot of the mean imprecision, $ipr_{K_i,sm}$, for the 32 K_i values as function of the 95 CAR combinations for a constant total number of 560 data points (red) compared with 35 data points per CAR (blue). At intermediate numbers of included CARs, $ipr_{K_i,sm}$ is slightly lower for a constant total number of 560 data points (inset).

+0.3 (Fig. S6 B). However, there are also some more positive and negative values: (1) the fourth binding steps show a pronounced positive correlation between each other with coefficients close to 1 (lower right corner of the plot). (2) There is a negative correlation between the fourth binding steps and E4. To some extent, this resembles the relationship between the binding step and the subsequent conformational change in the much simpler del Castillo-Katz scheme (Colquhoun, 1998). (3) A similar tendency, although not as strong, is indicated between all third binding steps and E3. (4) There is a moderate anticorrelation among the three possible second binding steps. All anticorrelations contribute as coupled reactions to the population of an open state (e.g., K_{29} – K_{32} with E4), generating

only a small component of the overall signal. Thus, the anti-correlation indicates a balance of the deviations of one equilibrium by an opposing deviation of another. Prospectively, inclusion of more concentrations at which these states contribute more and additional read-outs (e.g., by fluorescence) could decrease the uncertainty in such anticorrelations.

Minimum of the global fit

To further confirm our observation of excellent convergence with 16 CARS and 23 parameters, we performed another test. For our standard conditions of 7 × 5 data points for each of the 16 CARs, we fitted only 22 parameters while systematically fixing the 23rd parameter. We then systematically varied the 23rd

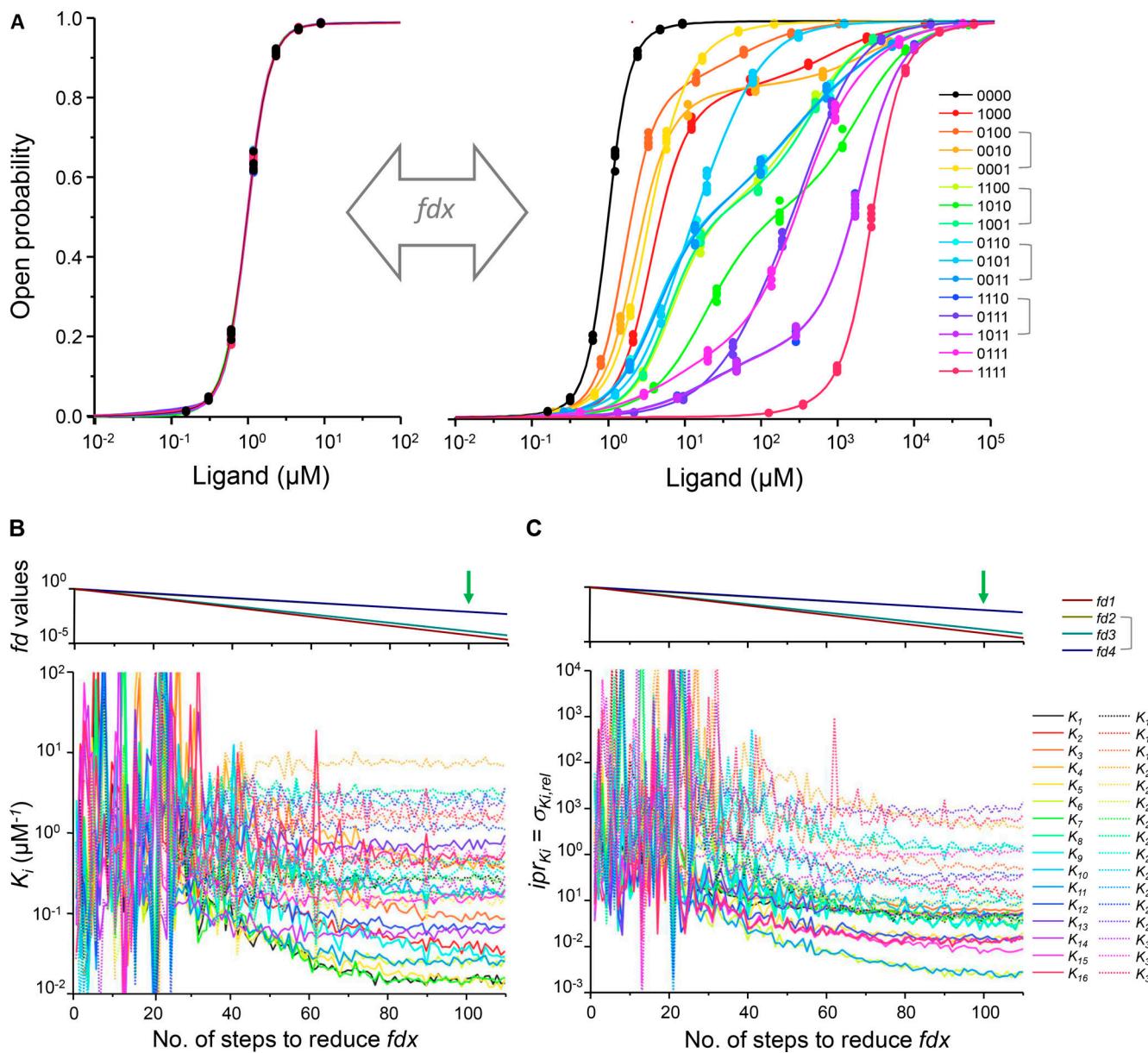


Figure 7. Mutational effect on the goodness of the global fit. (A) Fit of 16 CARs with the CHA₁₆ model. The fit was started with no mutational effect, i.e., all f_{dx} values were set to 1 (left). Then each f_{dx} was systematically decreased 110 times by a factor to reach, after 100 steps, the true f_{dx} value, and 10 steps further. The big gray double arrow illustrates that the f_{dx} values were varied. The brackets in the legends indicate closely similar data points and curves. (B) K_i as function of the factor by which f_{dx} was decreased to reach, after 100 steps, the true value (arrow). The four f_{dx} values are plotted in the top scheme. (C) Imprecisions ipr_{Ki,rel} as functions of the factor by which f_{dx} was decreased to reach, after 100 steps, the true value. The values correspond to the plot in B.

parameter by the factor B in the proximity of the true value that was determined before by the fit with 23 free parameters. In the case of a robust minimum, this should lead to a minimum MSE at $B = 1$. Matching our observation of excellent convergence, a minimum was observed for all parameters (Fig. 12, A–C). This minimum was steep for all parameters apart from the two anticorrelated parameters E4 and Z₁₅, for which it was much flatter. Nevertheless, these two parameters also showed a minimum. These results show that in the proximity of the identified minimum of the fits there is no continuum of parameter vectors as described for binding data when using a simple model with only two binding steps and an allosteric

factor but without the helpful power of mutagenesis (Middendorf and Aldrich, 2017a).

Stochastic variation of the true start vector

The weak point so far is that we used the true values as start vector, because it is possible that more minima exist in the 23-dimensional parameter space that the fit does not reach, thus biasing the fit. We therefore decided to abandon the strategy of using a start vector with the true parameters. In a first step, we started to stochastically vary the parameters of the true start vector and to count the incidence of successful fits, containing exclusively positive parameters. We evaluated the

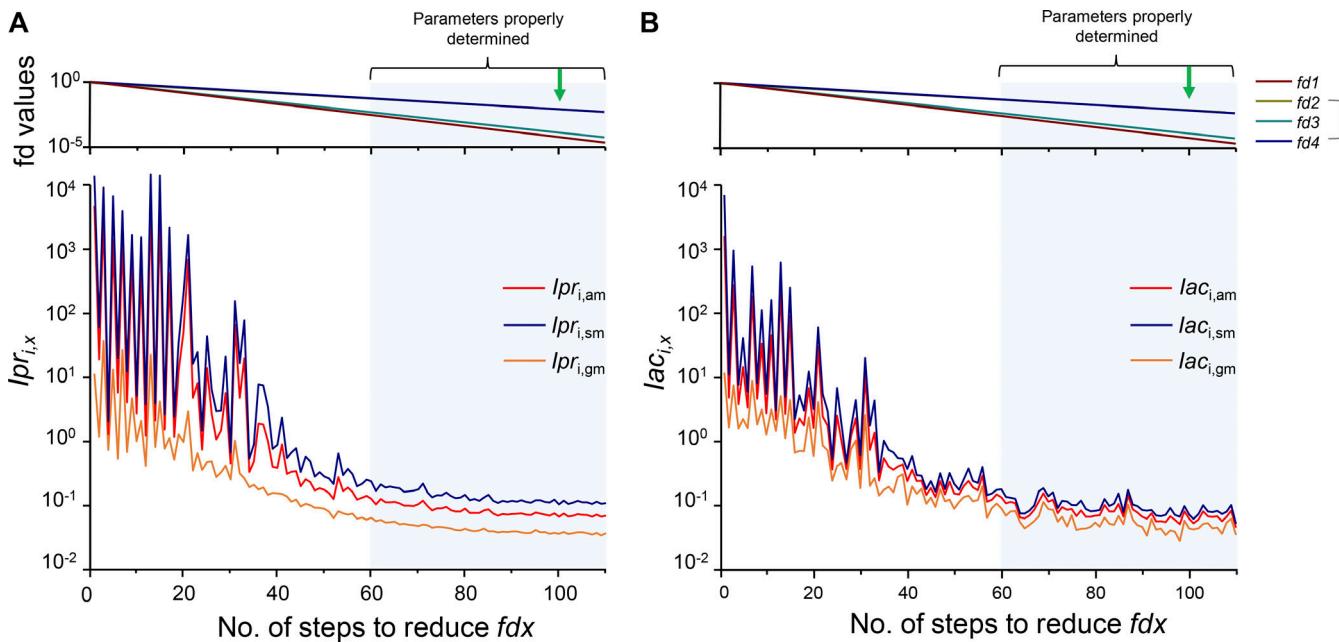


Figure 8. Mean uncertainty of the fits as function of the mutational effect. **(A)** Plot of three mean imprecisions, $lpr_{K_i,am}$, $lpr_{K_i,sm}$, and $lpr_{K_i,gm}$ (Eqs. 6b, 7b, and 8b) corresponding to Fig. 7, A and B. **(B)** Plot of the three mean inaccuracies, $lac_{K_i,am}$, $lac_{K_i,sm}$, and $lac_{K_i,gm}$ (Eqs. 11b, 12b, and 13b) corresponding to Fig. 7, A and B. The shaded areas illustrate that mutations decreasing the affinity by ≥ 60 -fold generate properly determined parameters.

mean imprecision (Eq. 6a) and the mean inaccuracy (Eq. 11a). Our general protocol was to fit the full set of 16 CARs 1,000 times.

Because our analysis now contains two stochastic influences, the noise of the data points and the variation of the start vector, we tested first the effect of a different data noise on the

incidence of successful fits at unchanged stochastic variation of the start vector. For our standard conditions of 560 data points and noise factor 0.25, all parameters of the true start vector were varied by the stochastic factor $A = 5$ in both directions, resulting in total in a 25-fold variation range for each parameter (Fig. S7 A). In this representative example, 895 fits were successful, as

Downloaded from https://jgp.oxfordjournals.org/ at 14/06/2014 13:04:11, IP: 182.170.10.195, on 10 February 2026

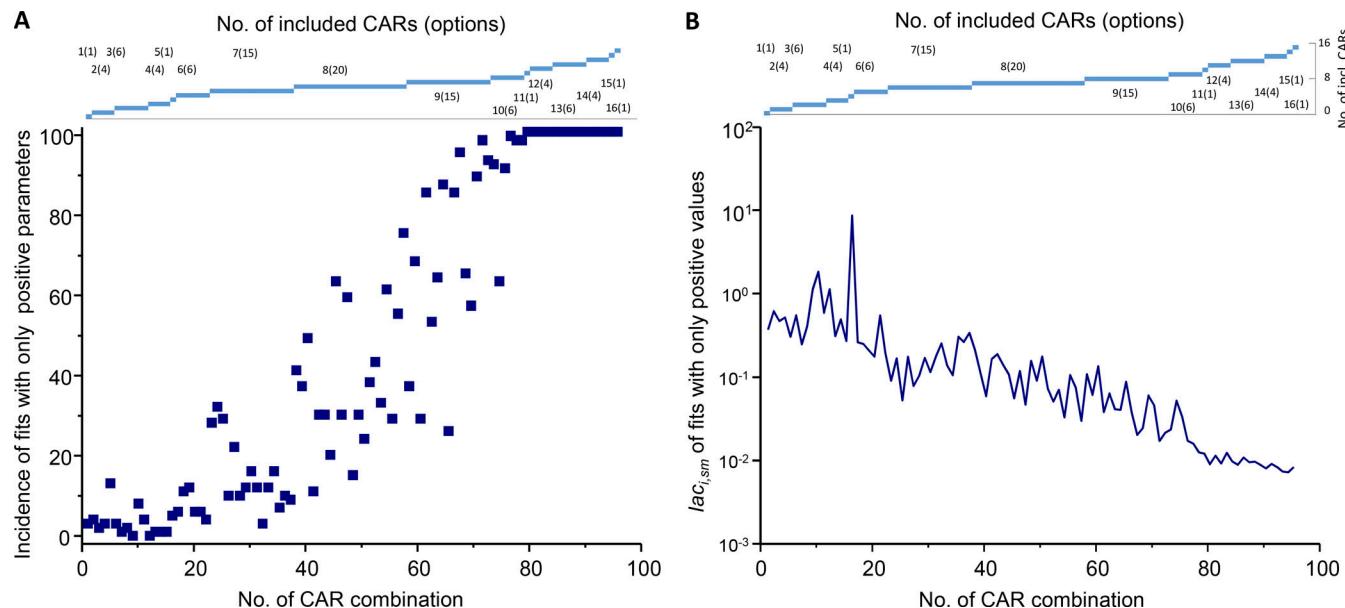


Figure 9. Inaccuracy in selected successful fits at different CAR combinations. **(A)** Incidence of fits with positive parameters only. The 95 CARs were fitted 100 times each, and the number of fits with only positive parameters, the criterion for a successful fit, was counted. The start vector contained the true values. Low CAR numbers generate only a low incidence of successful fits, whereas combinations with ≥ 12 CARs (cf. Table 1) generated successful fits in all cases. **(B)** Plot of the mean inaccuracy $lac_{i,sm}$ as a function of the 95 CARs of fits with only positive parameters. The mean was formed first for each successful fit over all parameters x_i by Eq. 12a and then over all successful fits specified in A. The inaccuracy decreases toward the highest CAR number of 16.

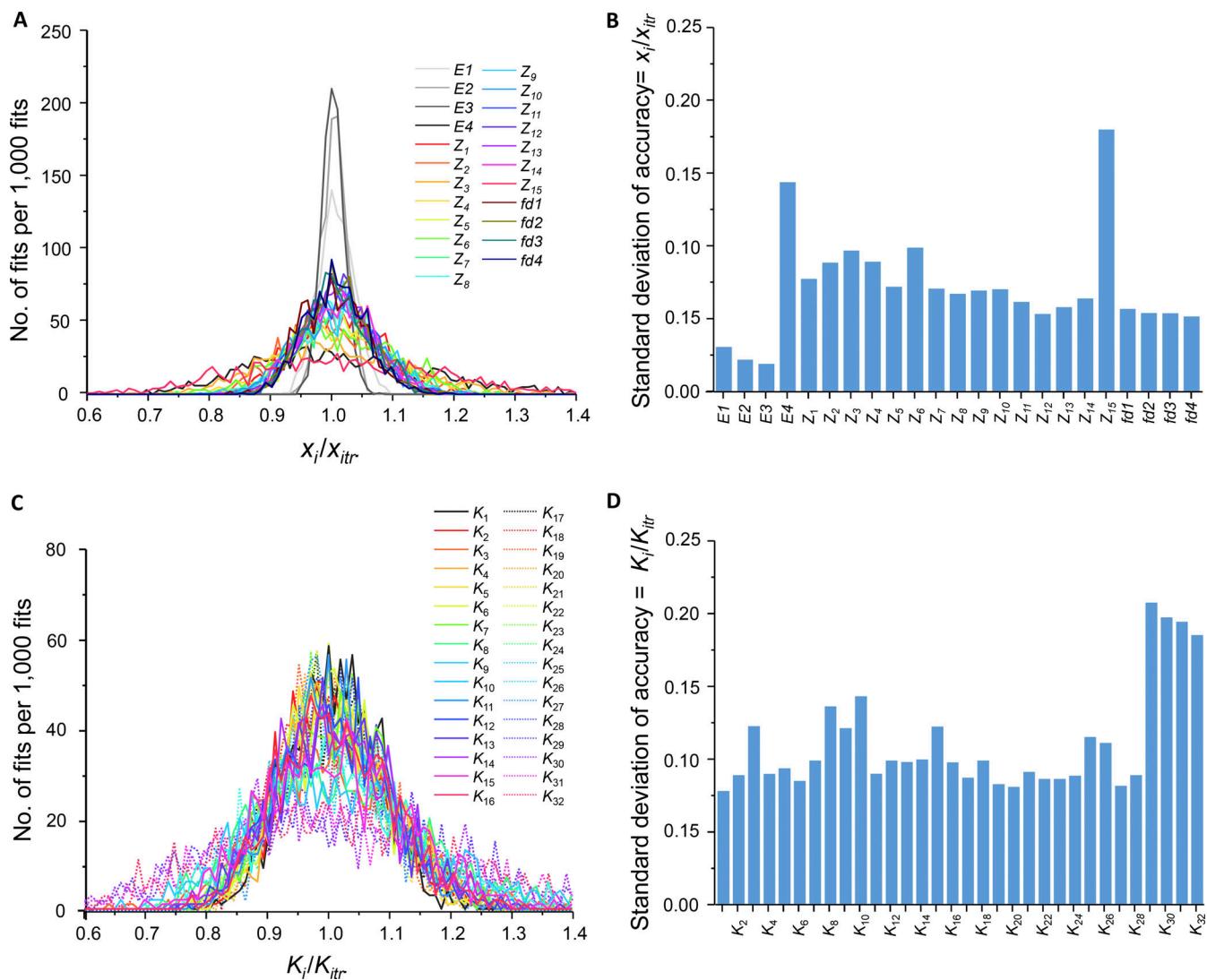


Figure 10. Accuracy of parameters and equilibrium association constants. The optimal condition of including 16 CARs (CAR combination 95) was used. This CAR combination was fitted 1,000 times while stochastically varying the data points from fit to fit. The distribution of the accuracy, x_i/x_{itr} (Eq. 9), was evaluated. **(A)** Superimposed histograms of the distribution of accuracies for all 23 determined parameters. Bin width 0.01. The histograms are bell-shaped, differing significantly in their width. **(B)** Bar graph of the SDs obtained by fitting the histograms in A with a Gaussian function (Eq. 15). The SDs of all parameters are <0.1 . **(C)** Superimposed histograms of the distribution of accuracies for the 32 equilibrium association constants K_i . Bin width 0.01. **(D)** Bar graph of the SDs obtained by fitting the histograms in C with a Gaussian function (Eq. 15). The SDs of the K_i were computed by Eqs. 2–4. The values are on average larger than the SDs of the Z_i (B) from which they were computed (see Materials and methods).

also indicated by the reasonable imprecision and inaccuracy (fit no. 1 in Table S6). Interestingly, in the failing fits, the tendency of the parameters $E4$ and Z_{15} to become negative was elevated (gray curve in Fig. S7 D), presumably mirroring the pronounced anticorrelation between these two parameters. When increasing the data noise eightfold to 2 (fit no. 2 in Table S6 and Fig. S7 B), we obtained in total 853 successful fits, which is only moderately less than obtained with the data noise 0.25. This indicates that at a poor data noise, the constraints inherent in the 16 CARs are still high, although both the imprecision and the inaccuracy have conspicuously increased. Data completely free of noise produce a rate of successful fits not dissimilar to the tested noisy data but with much lower imprecision and inaccuracy (fit no. 3 in Table S6; and Fig. S7, C and D). Together, these results suggest that the

data noise has less influence on the rate of successful fits than the amplitude of the variation of the start vector.

We then tested the effect of increasing the amplitude of the stochastic variation of the true start vector by increasing the stochastic variation factor A of all parameters from 5 to 100, resulting in an increased total range for each parameter from 2.5×10^1 -fold to 10^4 -fold, respectively (fit no. 4 in Table S6). Compared to the standard conditions, the number of successful fits has decreased, but it is most remarkable that this enormously elevated stochastic variation of the start vector left 492 fits successful, with imprecision and inaccuracy similar to those of control conditions (fit no. 1 in Table S6). Apparently, our elimination strategy of failing fits by the criterion of at least one negative parameter is useful to identify the true minimum.

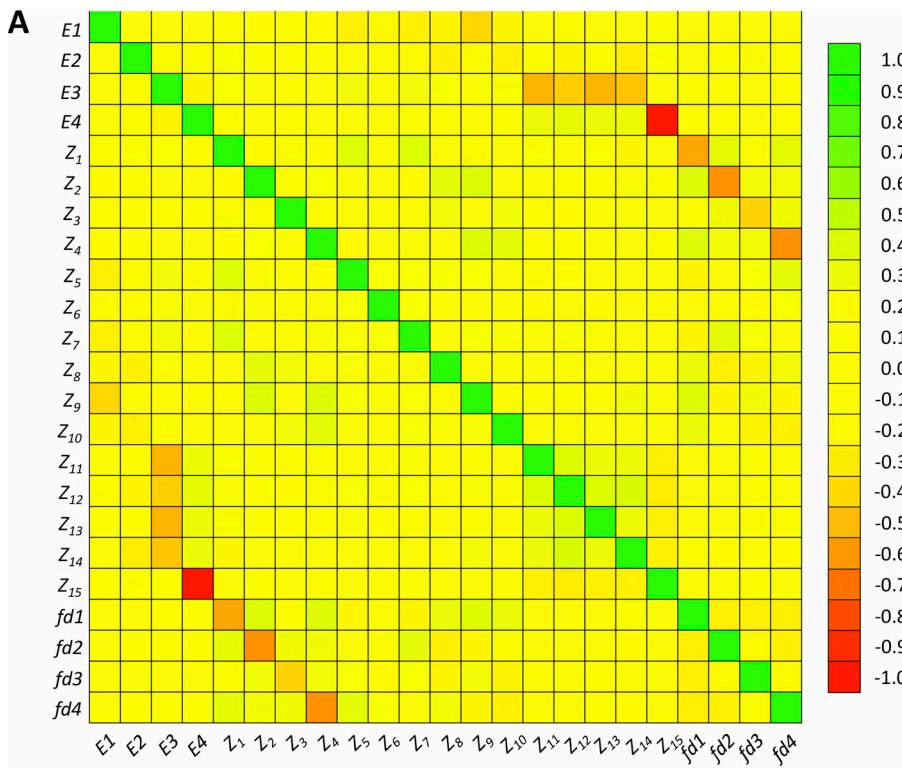
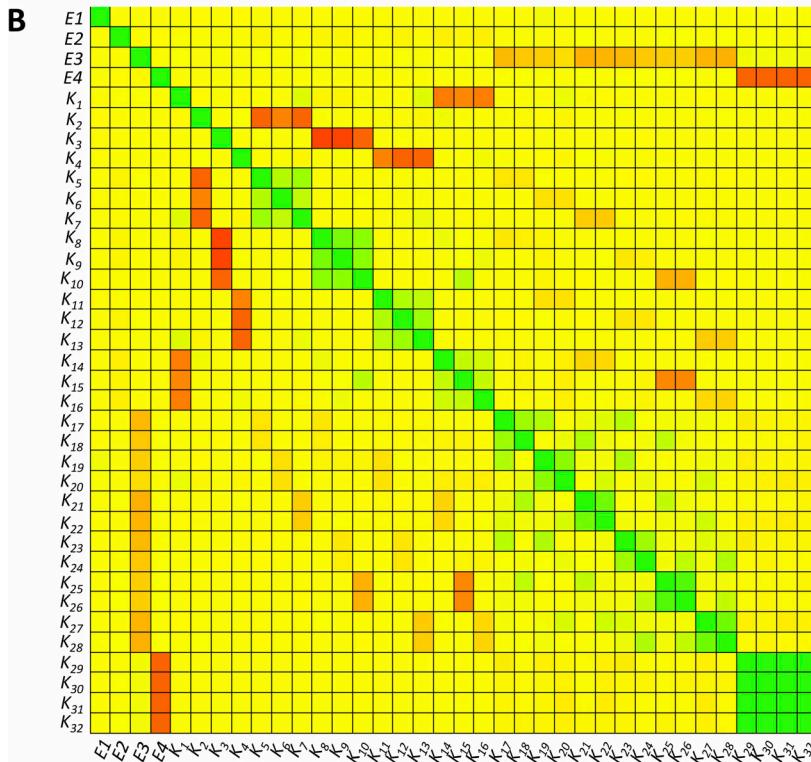


Figure 11. **Color-coded matrices of correlation coefficients. (A)** Parameters Ex , Z_i , and fdx . The exact numbers are indicated in Table S4. The data are the mean of 100 converging fits. The start vector was set to the true values. Noise factor 0.5. The main diagonal illustrates the correlations of the parameters with themselves, resulting in correlation coefficients of 1. Most other correlation coefficients are -0.3 to 0.3 , with few exceptions. **(B)** Ex and K_i . The exact numbers are indicated in Table S5. The data are the mean of the same 100 converging fits as in A. Most correlation coefficients are -0.3 to 0.3 , but there are both more positive and negative values than in A.



Stochastic variation of a SU start vector

The observed resilience of the fit success at an enormous stochastic variation of the true start vector prompted us to try SU start vectors, containing identical elements, and to vary them stochastically. If enough successful fits are obtained and the nonsuccessful fits can still be identified, a major source of bias would be removed.

We systematically scanned SU start vectors of 10^{-6} to 10^1 and used the same stochastic factor, $A = 10^2$. The success rate was maximal for an SU start vector of 10^{-1} (Fig. 13 A), which is near the logarithmic mean of 2.8×10^{-2} for all parameters used for the simulation. This shows that without prior knowledge about the parameters, (1) a sufficient number of successful fits can be

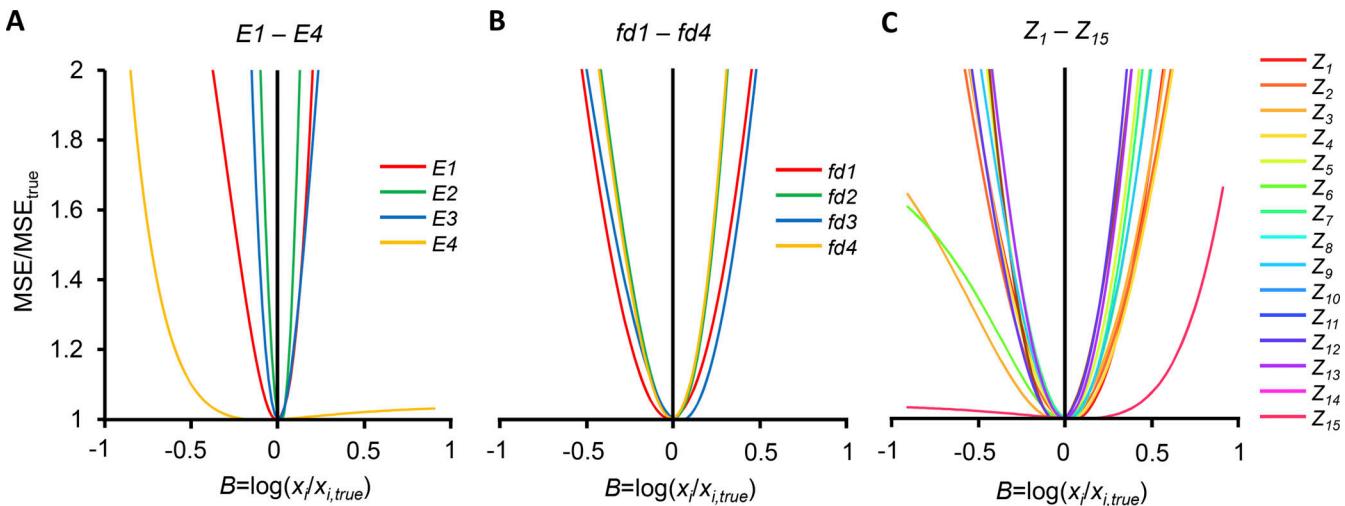


Figure 12. The minimum of the global fit. For our standard conditions ($7 \times 5 \times 16 = 560$ data points), fits with only 22 parameters were carried out while fixing the 23rd parameter and varying it by the factor B in the proximity of the true value as obtained by the 23-parameter fit. B was varied from 1/8 to 8 in logarithmically equidistant steps. The MSE was normalized to the true MSE_{true} value of the respective parameter. **(A–C)** Plots of the indicated fit parameters. There are robust minima of MSE/MSE_{true} for all parameters apart from the anticorrelated parameters $E4$ and Z_{15} , for which much flatter minima were found.

Downloaded from [http://jgpess.org/jgp/article-pdf/154/6/e202113041/1802710/jgp_202113041.pdf](http://jgpress.org/jgp/article-pdf/154/6/e202113041/1802710/jgp_202113041.pdf) by guest on 10 February 2026

obtained and (2) a useful range for a reasonable SU start vector can be identified.

We next considered the distribution of the minima among the successful fits. For a given SU start vector, all fits were performed with identical data points, providing information about the influence of the stochastic variation of the SU start vector on the minimum. To this end, the fits were sorted with respect to the MSE, starting with the smallest. For all MSE_x of this sequence, a normalized measure, $C(MSE_x)$, was iteratively calculated (Eq. 16; see Materials and methods). $C(MSE_x)$ increases at already very subtle changes.

Plotting $C(MSE_x)$ as function of the increasing MSE_x shows that with SU start vectors generating several tens of successful fits (SU start vectors 3×10^{-2} to 3×10^{-1}), the vast majority of fits produced consistently $C(MSE_x)$ values $<10^{-10}$ or even $<10^{-13}$ (Fig. 13 B). Rare outliers produced excessively larger $C(MSE_x)$ values.

To select for each SU start vector fits with nearly identical MSE, we set a threshold to 10^{-10} (red dashed line in Fig. 13 B). When taking further into account that for a given SU start vector at least five $C(MSE_x)$ values below the threshold are required, this allowed us to select fits with SU start vectors between 3×10^{-5} and 10^0 (Fig. 13 A). Both larger and smaller SU start vectors produced less successful fits and were not considered further.

The fits below the threshold of $C(MSE_x) < 10^{-10}$ are highly consistent, because this threshold means a difference of an MSE_x value in the 11th digit after the decimal sign with respect to the mean of MSE_1 through MSE_{x-1} . In contrast, for the fits above the threshold, we observed in no case a comparable consistency. This supports the conclusion that all 238 subthreshold fits obtained with the SU start vectors of 3×10^{-5} to 10^0 represent the best minimum. To show that the successful and consistent fits indeed match the true parameters, we plotted all determined parameters for the respective SU start vectors (Fig. 13 C). This shows that the successful and consistent fits below the threshold

allowed us to determine the true parameters excellently, independent of the value of the actual SU start vector.

In conclusion, our stochastic approach using SU start vectors can help to identify the true values of the 23 parameters without bias given by specific start vectors. Nevertheless, we note that these results make it very likely that the consistently identified minimum is the global minimum of the fit, but they do not finally exclude that another minimum in the huge 23-dimensional parameter space exists.

Discussion

In this study, we investigated the determinability of 23 free parameters for a model scheme of a heterotetrameric ligand-gated ion channel. We derived from these parameters in total 40 constants. Our strategy is based on knowing the subunit composition of the channel, disabling of different binding sites of the subunits by mutations, and globally fitting CARs obtained from macroscopic currents. One major assumption of our strategy is that the mutations disable the binding sites but preserve functionality, i.e., the subunits of all concatamers, including the mutated ones, can still be activated. Another assumption is that the principle of microscopic reversibility holds, being aware that for some ion channels a violation of this principle has been suggested (Schneggenburger and Ascher, 1997; Xu and Meissner, 1998).

A first relevant result is that we identified for 12 or more CARs, preferably the full set of 16 CARs (CHA₁₆ model), a network of fit constraints of unusual intensity. The success of the strategy is mainly based on introducing the four helper parameters, fdx , specific factors for disabling the affinity of the binding domains. The fdx values were assumed to be unique for each subunit, independent of the activation of the other subunits. These four helper parameters are not relevant for the

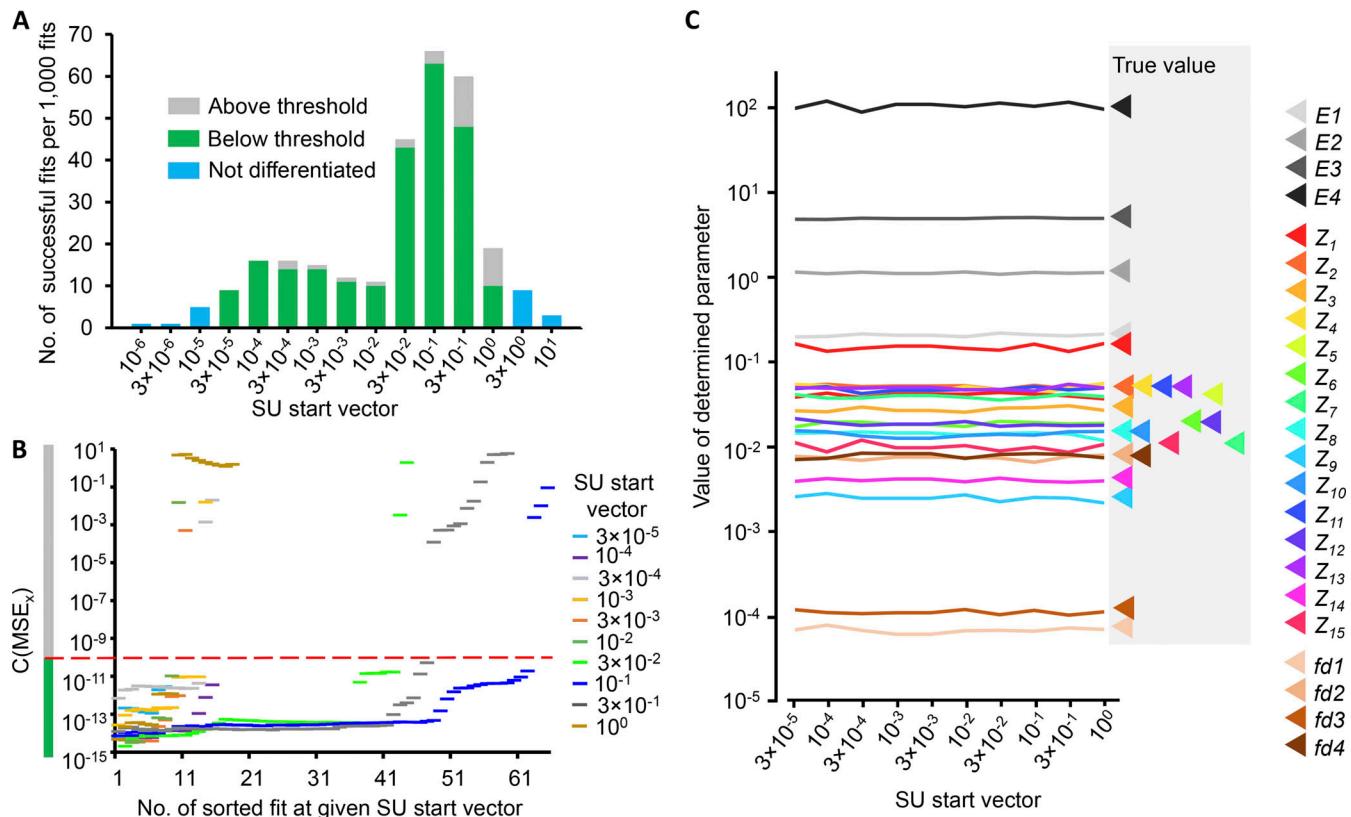


Figure 13. Successful fits with SU start vectors. SU start vectors, containing 23 identical start parameters, were varied by the stochastic factor $A = 10^2$. 1,000 fits were performed for each SU start vector. SU start vectors of 10^{-6} to 10^0 were tested. **(A)** Histogram of successful fits for different SU start vectors. Among the successful fits for SU start vectors of 3×10^{-5} to 10^0 , the differentiation with respect to a threshold refers to B. **(B)** Refined search for the fit minimum with SU start vectors of 3×10^{-5} to 10^0 . The fits were sorted with respect to the MSE value, starting with the smallest. A normalized measure for the x th element in this sequence, $C(MSE_x)$, was computed with Eq. 16 (see Materials and methods). The diagram shows the plot of $C(MSE_x)$ as a function of the increasing MSN_x number. The vast majority of fits consistently produced $C(MSE_x)$ values $<10^{-10}$ or even $<10^{-13}$. Setting a threshold to 10^{-10} (red line) allowed for the elimination of successful fits with higher MSE values. Other thresholds of $\sim 10^{-13}$ to $\sim 10^{-5}$ would also produce satisfying results. The tiny variability of $C(MSE_x)$ of 10^{-14} to 10^{-13} can be explained by the use of data points with different noise for different SU start vectors. **(C)** Plot of all determined parameters for SU start vectors varying from 3×10^{-5} to 10^0 . The dimensions of Z_1 – Z_4 , Z_5 – Z_{10} , Z_{11} – Z_{14} , and Z_{15} are μM^{-1} , μM^{-2} , μM^{-3} , and μM^{-4} , respectively. The other parameters are dimensionless. The true values are indicated. The accuracy of the determined parameters does not depend on a specific SU start vector.

interpretation of the constants of the WT channel, although being essential for the global fit.

A second relevant result is that we identified in this global fit for all 23 parameters a minimum that is remarkably robust, apart from the two strongly anticorrelated parameters $E4$ and Z_{15} . Even for them, a flat minimum was shown (Fig. 12).

A third relevant result is that we proposed a stochastic approach to consistently identify a minimum with the true 23 parameters when using a SU start vector. This SU start vector avoids any bias by more specific start vectors and thus proves the power of our systematic mutations in the concatamers together with the global fit strategy. This approach is in fact simple. First, a stochastic scan has to be performed to identify the range for useful SU start vectors by counting the successful fits with only positive parameters. Second, in a range with enough successful fits, rare fits with larger and inconsistent MSE values can be identified and discarded. The MSE of the selected fits is extremely consistent and provides the final result, which herein matched exclusively the true parameter values used in the simulations, independent of the SU start vectors

(Fig. 13 C). If a continuum of minima (Middendorf and Aldrich, 2017a) were present in our simulated data, we speculate that we would not have observed the consistent minimal MSE (Fig. 13 B).

The proposed strategy of fitting complex networks of coupled HA models with different patterns of disabling, resulting in a respective complex CHA model, is apparently an exceptionally fortunate condition to reliably determine multiple parameters. This does not automatically mean that a sufficiently large data set leads necessarily to similarly consistent results for any model, even if a model is simple compared with the models considered herein. This is because, as shown, the goodness of the fit depends on various influences including the number of CARs, the noise amplitude, the chosen start vectors, and most of all, the strength of the disabling effects of the mutations (Figs. 7 and 8). Nevertheless, our considerations allow us to say that more CARs and strong disabling effects significantly promote the determinability of the parameters. It seems to be promising to elaborate in the future a unifying theory for all these influences.

The number of 23 free parameters in our global fits seems high for investigators who are used to work with single equations

and few parameters. We like to present another argument that 23 parameters for 16 CARs is in fact not a large number: when comparing the 23 parameters with $16 \times 2 = 32$ parameters required to fit 16 CARs separately by single Hill functions, the number is already significantly smaller. When reflecting in addition the necessity of two components to describe CARs, the respective double Hill functions require five parameters. For example, when assuming six double Hill functions, the total number of parameters obtained from the data would increase to $10 \times 2 + 6 \times 5 = 50$. This exceeds the 23 parameters substantially. In return, this tells how strong the constraints are when globally fitting our system of intimately coupled models.

As noted above, the analyses presented herein are based on real data of concatameric heterotetrameric CNGA4:A2:Blb:A2 channels, where 17 parameters were previously determined. The higher number of 23 in the present theoretical study results from the assumption that each of the subunits was allowed to adopt a different equilibrium association constant, whereas in our previous study we assumed that the two A2 subunits are identical. Hence, the extended considerations herein enable a wider use of our approach for other heterotetrameric ligand-gated ion channels.

For the future, it seems to be promising to extend our analysis to dynamic data, evoked for instance by jumps of the ligand concentration, and to determine, in addition to equilibrium constants, rate constants for conformational changes. It seems to be promising to identify general basics for a successful fit with our approach independent of a specific model, covering the relations between the number of parameters, the structure of the parameter vector, the structure of the model, and the size of the stochastic factor A .

Our approach is straightforward and applicable to experimental data accordingly. Beyond heterotetrameric ligand-gated ion channels, heterotrimeric and heteropentameric ligand-gated ion channels also seem to be candidates for our approach, if the subunit composition is exactly known. Such approaches would require a 3-D cube and 5-D hypercube, respectively, leaving the basic strategy similar. Another field of applications of our strategy are classic voltage-gated channels. In analogy, our strategy seems to be suitable to identifying the correct stoichiometry in heterotetrameric channels if part of the contributing subunits do not form functional channels on their own, as for example, Kv2.1/Kv6.4 channels (Moller et al., 2020; Pisupati et al., 2020). Here, a promising idea is to affect the voltage-sensor domain of the subunits by appropriate mutations, evoking shifts of steady-state activation to more positive voltages. To some extent, this is analogous to the disabling effects used herein for ligand-gated channels. Moreover, this strategy should be usable to analyze natural heterotetramers, such as Nav channels (Jiang et al., 2020) and Cav channels (Wu et al., 2015).

Acknowledgments

Christopher J. Lingle served as editor.

The authors declare no competing financial interests.

This work was supported by the Research Unit 2518 DynIon (project P2 to K. Benndorf) of the Deutsche Forschungsgemeinschaft.

Author contributions: K. Benndorf created the simulation and fit software and did most of the computations. T. Eick, R. Schmauder, C. Sattler, and E. Schulz contributed to the analysis and some software routines. K. Benndorf, C. Sattler, and T. Eick designed and prepared the figures. K. Benndorf designed the project and wrote the manuscript.

Submitted: 21 October 2021

Revised: 11 February 2022

Accepted: 18 March 2022

References

Beato, M., V. Burzomato, and L.G. Sivilotti. 2007. The kinetics of inhibition of rat recombinant heteromeric alphabeta glycine receptors by the low-affinity antagonist SR-95531. *J. Physiol.* 580:171–179. <https://doi.org/10.1113/jphysiol.2006.126888>

Burzomato, V., M. Beato, P.J. Groot-Kormelink, D. Colquhoun, and L.G. Sivilotti. 2004. Single-channel behavior of heteromeric alphabeta glycine receptors: An attempt to detect a conformational change before the channel opens. *J. Neurosci.* 24:10924–10940. <https://doi.org/10.1523/JNEUROSCI.3424-04.2004>

Colquhoun, D. 1998. Binding, gating, affinity and efficacy: The interpretation of structure-activity relationships for agonists and of the effects of mutating receptors. *Br. J. Pharmacol.* 125:924–947. <https://doi.org/10.1038/sj.bjp.0702164>

Colquhoun, D., and A.G. Hawkes. 1995. The principles of the stochastic interpretation of ion-channel mechanisms. In *Single-Channel Recording*. B. Sakmann, and E. Neher, editors. Plenum Press, New York. 397–482.

daCosta, C.J.B., and J.E. Baenziger. 2013. Gating of pentameric ligand-gated ion channels: Structural insights and ambiguities. *Structure*. 21: 1271–1283. <https://doi.org/10.1016/j.str.2013.06.019>

Eaton, W.A., E.R. Henry, J. Hofrichter, and A. Mozzarelli. 1999. Is cooperative oxygen binding by hemoglobin really understood? *Nat. Struct. Biol.* 6: 351–358. <https://doi.org/10.1038/7586>

Goulding, E.H., G.R. Tibbs, and S.A. Siegelbaum. 1994. Molecular mechanism of cyclic-nucleotide-gated channel activation. *Nature*. 372:369–374. <https://doi.org/10.1038/372369a0>

Hines, K.E., T.R. Middendorf, and R.W. Aldrich. 2014. Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach. *J. Gen. Physiol.* 143:401–416. <https://doi.org/10.1085/jgp.20131116>

Jiang, D., H. Shi, L. Tonggu, T.M. Gamal El-Din, M.J. Lenaeus, Y. Zhao, C. Yoshioka, N. Zheng, and W.A. Catterall. 2020. Structure of the cardiac sodium channel. *Cell*. 180:122–134.e10. <https://doi.org/10.1016/j.cell.2019.11.041>

Kawate, T., J.C. Michel, W.T. Birdsong, and E. Gouaux. 2009. Crystal structure of the ATP-gated P2X(4) ion channel in the closed state. *Nature*. 460:592–598. <https://doi.org/10.1038/nature08198>

Kusch, J., C. Biskup, S. Thon, E. Schulz, V. Nache, T. Zimmer, F. Schwede, and K. Benndorf. 2010. Interdependence of receptor activation and ligand binding in HCN2 pacemaker channels. *Neuron*. 67:75–85. <https://doi.org/10.1016/j.neuron.2010.05.022>

Lee, W.Y., and S.M. Sine. 2005. Principal pathway coupling agonist binding to channel gating in nicotinic receptors. *Nature*. 438:243–247. <https://doi.org/10.1038/nature04156>

Li, M., X. Zhou, S. Wang, I. Michailidis, Y. Gong, D. Su, H. Li, X. Li, and J. Yang. 2017. Structure of a eukaryotic cyclic-nucleotide-gated channel. *Nature*. 542:60–65. <https://doi.org/10.2210/pdb5h3o/pdb>

Middendorf, T.R., and R.W. Aldrich. 2017a. Structural identifiability of equilibrium ligand-binding parameters. *J. Gen. Physiol.* 149:105–119. <https://doi.org/10.1085/jgp.201611702>

Middendorf, T.R., and R.W. Aldrich. 2017b. The structure of binding curves and practical identifiability of equilibrium ligand-binding parameters. *J. Gen. Physiol.* 149:121–147. <https://doi.org/10.1085/jgp.201611703>

Moller, L., G. Regnier, A.J. Labro, R. Blunck, and D.J. Snyders. 2020. Determining the correct stoichiometry of Kv2.1/Kv6.4 heterotetramers, functional in multiple stoichiometrical configurations. *Proc. Natl. Acad. Sci. USA*. 117:9365–9376. <https://doi.org/10.1073/pnas.1916166117>

Monod, J., J. Wyman, and J.P. Changeux. 1965. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* 12:88–118. [https://doi.org/10.1016/s0022-2836\(65\)80285-6](https://doi.org/10.1016/s0022-2836(65)80285-6)

Pisupati, A., K.J. Mickolajczyk, W.O. Hancock, and T. Jegla. 2020. What is the correct stoichiometry of Kv2.1:Kv6.4 heteromers? *Proc. Natl. Acad. Sci. USA.* 117:29288–29289. <https://doi.org/10.1073/pnas.2017827117>

Schirmeyer, J., S. Hummert, T. Eick, E. Schulz, T. Schwabe, G. Ehrlich, T. Kukaj, M. Wiegand, C. Sattler, R. Schmauder, et al. 2021. Thermodynamic profile of mutual subunit control in a heteromeric receptor. *Proc. Natl. Acad. Sci. USA.* 118. e2100469118 <https://doi.org/10.1073/pnas.2100469118>

Schneggenburger, R., and P. Ascher. 1997. Coupling of permeation and gating in an NMDA-channel pore mutant. *Neuron.* 18:167–177. [https://doi.org/10.1016/s0896-6273\(01\)80055-6](https://doi.org/10.1016/s0896-6273(01)80055-6)

Sobolevsky, A.I., M.P. Rosconi, and E. Gouaux. 2009. X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature.* 462:745–756. <https://doi.org/10.1038/nature08624>

Steinbach, J.H., and G. Akk. 2019. Applying the Monod-Wyman-Changeux allosteric activation model to pseudo-steady-state responses from GABA receptors. *Mol. Pharmacol.* 95:106–119. <https://doi.org/10.1124/mol.118.113787>

Wu, J., Z. Yan, Z. Li, C. Yan, S. Lu, M. Dong, and N. Yan. 2015. Structure of the voltage-gated calcium channel Cav1.1 complex. *Science.* 350:aad2395. <https://doi.org/10.1126/science.aad2395>

Xu, L., and G. Meissner. 1998. Regulation of cardiac muscle Ca^{2+} release channel by sarcoplasmic reticulum luminal Ca^{2+} . *Biophys. J.* 75:2302–2312. [https://doi.org/10.1016/S0006-3495\(98\)77674-X](https://doi.org/10.1016/S0006-3495(98)77674-X)

Supplemental material

The 4-D hypercube as a useful tool

Let us consider the 16 closed states $C_v = Cijkl$ ($v = 0 \dots 15$) of a tetrameric ion channel, arranged in the corners of a 4-D cube, where the binary components of the 4-D vector $v = (ijkl)$ indicate whether the subunit in question is liganded (1) or not (0). A well-manageable system to describe the equilibrium properties is based on using “virtual,” or rather “generalized,” equilibrium constants $Z_v = Zijkl$ ($v = 1 \dots 15$) between state $C0000$ and state $Cijkl$. In this set of 15 generalized equilibrium constants, $Zijkl$ has some advantages over the use of the actual equilibrium constants $Kxxxx$ between adjacent closed states. First, the $Zijkl$ form a tree with only one single node in $C0000$ and they are therefore evidently free from any microscopic reversibilities. Second, the equations for determining the $Zijkl$ from experimental data are low order or even linear compared to those with $Kxxxx$. Third, all desired $Kxxxx$ can be calculated from the $Zijkl$.

This is demonstrated for the example K_{1011}

$$K_{1011} = Z_{1011}/Z_{1010}, \quad (S1)$$

where the 1 in K_{1011} denotes that the binding of the ligand occurs at the fourth subunit, whereas the 1 denotes that the first and third subunit are already occupied.

With the help of $Zijkl$, the occupation probability P_cijkl of any closed state $Cijkl$ can now be written as a multiple of P_c0000 of the empty “ground state” $C0000$, namely

$$P_cijkl = P_c0000 \cdot Zijkl \cdot L^\alpha. \quad (S2)$$

The power $\alpha = i + j + k + l$ of the ligand concentration L counts the number of occupied subunits.

With the related equilibrium constants for the closed-open isomerizations, $Eijkl$, the occupation probability P_oijkl of the open state $Oijkl$ is given by

$$P_oijkl = P_c0000 \cdot Eijkl \cdot Zijkl \cdot L^\alpha. \quad (S3)$$

Also, openings without bound ligands are included in Eq. S3 with $Z0000 = 1$ and $E0000 \neq 0$.

The occupation probabilities of all open states add up to the open probability $P_o(L)$ of the whole channel, where the summation extends over all combinations of 0 and 1 for i, j, k, l :

$$P_o(L) = \sum_{i,j,k,l} p_{ijkl} = P_c0000 \cdot \sum_{i,j,k,l} Eijkl \cdot Zijkl \cdot L^\alpha \quad (S4)$$

The occupation probabilities of all open and closed states add up to 1

$$1 = \sum_{i,j,k,l} (p_{ijkl} + p_{ijkl}) = P_c0000 \cdot \sum_{i,j,k,l} (1 + Eijkl) \cdot Zijkl \cdot L^\alpha \quad (S5)$$

The sums in Eqs. S4 and S5 can be divided into subtotals with equal number $\alpha = i + j + k + l$ of bound ligands:

$$P_o(L) = P_c0000 \cdot \sum_{\alpha=0}^4 \left(\sum_{i+j+k+l=\alpha} Eijkl \cdot Zijkl \right) \cdot L^\alpha \quad (S6)$$

$$1 = P_c0000 \sum_{\alpha=0}^4 \left[\sum_{i+j+k+l=\alpha} (1 + Eijkl) \cdot Zijkl \right] \cdot L^\alpha \quad (S7)$$

It is useful to introduce abbreviations for the subtotals in Eqs. S6 and S7

$$a_\alpha = \sum_{i+j+k+l=\alpha} Eijkl \cdot Zijkl \quad (S8)$$

$$b_\alpha = \sum_{i+j+k+l=\alpha} (1 + Eijkl) \cdot Zijkl, \quad (S9)$$

where the summations cover all combinations with the same value of $\alpha = 0 \dots 4$. For the special case $\alpha = 0$ follows

$$b_0 - a_0 = 1. \quad (S10)$$

Now Eqs. S6 and S7 simplify to

$$P_o(L) = P_c 0000 \cdot \sum_{\alpha=0}^4 a_{\alpha} \cdot L^{\alpha} \quad (S11)$$

$$1 = P_c 0000 \cdot \sum_{\alpha=0}^4 b_{\alpha} \cdot L^{\alpha}. \quad (S12)$$

Using Eq. S12, one can eliminate $P_c 0000$ in Eq. S11 and calculate the probability of the whole channel as a rational function of the ligand concentration L with five coefficients a_{α} in the numerator and five coefficients b_{α} in the denominator

$$P_o(L) = \frac{a_0 + a_1 L + a_2 L^2 + a_3 L^3 + a_4 L^4}{b_0 + b_1 L + b_2 L^2 + b_3 L^3 + b_4 L^4}. \quad (S13)$$

How many parameters can be determined exactly from a data set?

To determine the 10 parameters of the rational function (Eq. S13), one needs 10 conditions. If exact values for the open probabilities $p_i = P_o(L_i)$ are available at nine concentrations L_i , $i = 1 \dots 9$, one gets nine equations from Eq. S13:

$$a_0 + a_1 L_i + a_2 L_i^2 + a_3 L_i^3 + a_4 L_i^4 - b_0 p_i - b_1 L_i p_i - b_2 L_i^2 p_i - b_3 L_i^3 p_i - b_4 L_i^4 p_i = 0 \quad (S14)$$

Together with $b_0 - a_0 = 1$ (Eq. S10), this is a uniquely solvable system of 10 linear equations for 10 unknown coefficients a_{α} and b_{α} . Ergo, one can always get the rational function (Eq. S13) from nine different noise-free data points.

Next, it is demonstrated how the Z_{ijkl} and E_{α} can be calculated from a_{α} and b_{α} . In our model we assumed that the equilibrium constants for the closed-open isomerizations depend only on the number α of bound ligands, $E_{ijkl} = E_{\alpha}$, with $\alpha = i + j + k + l$. Then the coefficients in Eqs. S8 and S9 and their difference are

$$a_{\alpha} = E_{\alpha} \sum_{i+j+k+l=\alpha} Z_{ijkl}, \quad (S15)$$

$$b_{\alpha} = (1 + E_{\alpha}) \sum_{i+j+k+l=\alpha} Z_{ijkl}, \quad (S16)$$

and

$$b_{\alpha} - a_{\alpha} = \sum_{i+j+k+l=\alpha} Z_{ijkl}. \quad (S17)$$

The special case $\alpha = 0$ yields with $Z0000 = 1$ the values $a_0 = E_0$ and $b_0 = 1 + E_0$.

It is then of interest to determine for a given model and noise-free data $P_o(L)$ the maximal number of parameters in the absence of any additional information.

Homotetrameric channel

First, a model for a homotetrameric channel with equivalent subunits but different equilibrium association constants K_{α} for each binding step α is considered. Here, all the generalized constants Z_{ijkl} with the same number of ones in the four indices $ijkl$ have the same value Z_{α} . From Eq. S15, one gets

$$b_{\alpha} - a_{\alpha} = \sum_{i+j+k+l=\alpha} Z_{ijkl} = \binom{4}{\alpha} Z_{\alpha}, \quad (S18)$$

where the binomial coefficient $\binom{4}{\alpha}$ provides the number of combinations with the same Z_{α} . The Z_{α} can be determined from Eq. S18, and the K_{α} can be calculated by Eq. S1 according to

$$K_{\alpha} = Z_{\alpha} / Z_{\alpha-1}, (\alpha = 1 \dots 4). \quad (S19)$$

For the parameters, E_{α} follows from Eqs. S15 and S17:

$$E_{\alpha} = a_{\alpha} / (b_{\alpha} - a_{\alpha}), (\alpha = 0 \dots 4) \quad (S20)$$

This shows that the nine parameters $E_0 \dots E_4$ and $K_1 \dots K_4$ for the model of the homotetrameric channel can be theoretically calculated from nine different data points.

Heterotetrameric channel

For the model of a heterotetrameric channel, more than nine equilibrium parameters are required. Because one CAR contains only nine conditions, additional experimental information is needed. It is therefore useful to have further constraints by including

additional CARs from channels containing subunits with a defined number of severely disabled but still functional binding sites. We assume that this mutational effect can be described by factors of disabling, fd_u ($u = 1 \dots 4$ for subunit u). According to Eqs. S15, S16, and S17, this yields for each CAR (index β)

$$a_{\alpha,\beta} = E_\alpha \sum_{i+j+k+l=\alpha} Zijkl \cdot fd_1^a \cdot fd_2^b \cdot fd_3^c \cdot fd_4^d \quad (S21)$$

$$b_{\alpha,\beta} = (1 + E_\alpha) \sum_{i+j+k+l=\alpha} Zijkl \cdot fd_1^a \cdot fd_2^b \cdot fd_3^c \cdot fd_4^d \quad (S22)$$

and

$$b_{\alpha,\beta} - a_{\alpha,\beta} = \sum_{i+j+k+l=\alpha} Zijkl \cdot fd_1^a \cdot fd_2^b \cdot fd_3^c \cdot fd_4^d \quad (S23)$$

The exponents a , b , c , and d are either 0 or 1. Such an exponent is 1 if the subunit is mutated and has bound a ligand. Otherwise, it is 0 (Schirmeyer et al., 2021). It is assumed for the following considerations that nine noise-free data points for six different CARs each are available and that the rational functions for each CAR can be calculated with Eqs. S14 and S10. It is then possible to calculate all the 24 parameters of the model ($E_0 \dots E_4$, $fd_1 \dots fd_4$, $15 \times Zijkl$) from these coefficients $a_{\alpha,\beta}$ and $b_{\alpha,\beta}$. A simple way is saying that $\beta = 0$ denotes the WT channel, $\beta = 1 \dots 4$ denote the CARs with one mutation at subunit 1...4, and $\beta = 5$ denotes a channel mutated on more than one binding place.

In the following it is shown that, under these assumptions, the unique calculation of all constants is possible. Part of the parameters can be calculated in different ways, starting with the E_α according to Eqs. S21 and S23.

$$E_\alpha = a_{\alpha,\beta} / (b_{\alpha,\beta} - a_{\alpha,\beta}), \quad (\alpha = 0 \dots 4; \text{ any } \beta) \quad (S24)$$

For the channels with only one mutation, we need only the correct factor fd_β . From Eq. S23 follows for $\beta = 1 \dots 4$ the simplification

$$b_{\alpha,\beta} - a_{\alpha,\beta} = \sum_{i+j+k+l=\alpha} Zijkl \cdot fd_\beta^a \quad (S25)$$

For the fourth coefficients with $\alpha = i + j + k + l = 4$, the sum contains only one summand

$$b_{4,\beta} - a_{4,\beta} = Z1111 \cdot fd_\beta, \quad (\beta = 1 \dots 4) \quad (S26)$$

yielding for the WT CAR

$$b_{4,0} - a_{4,0} = Z1111, \quad (\beta = 0). \quad (S27)$$

Now, it is evident that the calculation of the factors of disabling by mutations can be performed by

$$fd_\beta = (b_{4,\beta} - a_{4,\beta}) / (b_{4,0} - a_{4,0}), \quad (\beta = 1 \dots 4). \quad (S28)$$

As a consequence of Eqs. S21 or S22, for $\alpha = 4$ this can again be performed in different ways:

$$fd_\beta = a_{4,\beta} / a_{4,0} = b_{4,\beta} / b_{4,0}, \quad (\beta = 1 \dots 4) \quad (S29)$$

If the factors fd_β are known, all $Zijkl$ can be calculated by linear equations. The most simple case is the calculation of $Z1111$

$$Z1111 = b_{4,0} - a_{4,0} = (b_{4,\beta} - a_{4,\beta}) / fd_\beta \quad (S30)$$

For calculating the other $Zijkl$, it is necessary to solve systems of linear equations. In the cases $\alpha = 1$ and $\alpha = 3$, four different $Zijkl$ have to be calculated and a system of four linear equations is needed, e.g., by employing Eq. S25 with $\beta = 1 \dots 4$.

For the case $\alpha = 2$, the number of parameters $Zijkl$ is elevated to six, resulting in six linear equations. This can be realized with Eq. S17 for $\beta = 0$, with Eq. S25 for $\beta = 1 \dots 4$ and Eq. S23 for $\beta = 5$. This is the reason for the need of six different CARs in such an analysis of noiseless data.

According to Eq. S1, all constants $Kxxxx$ can be calculated from the $Zijkl$.

Finally, we emphasize again that the above considerations describe an unrealistic theoretical case of noiseless data. A critical point is also the calculation accuracy if the equation systems are not well conditioned. However, our considerations show that different calculation pathways are possible for most parameters of the model. This redundancy arises because 6 CARs measured at 9 concentrations each yield 54 conditions for the calculation of only 24 parameters.

This raises the question of whether only 24 data points (e.g., 6 points for each of 4 CARs) would be enough to calculate the 24 parameters of the model without first reconstructing the rational functions of the CARs involved. We get the 24 needed equations

from Eq. S14, but the coefficients a_k and b_k must be replaced according to Eqs. S21 and S22. This would result in 24 equations which are non-linear because of the now unavoidable products of the constants Z_{ijkl} with factors fd_β . As consequence, this system would have to be calculated numerically and, generally, could have more than one solution. If the data set contains at least a 25th data point, one could use it to test all the solutions and find the right one.

It should be stressed again that all these considerations are only valid for the theoretical case of noiseless data.

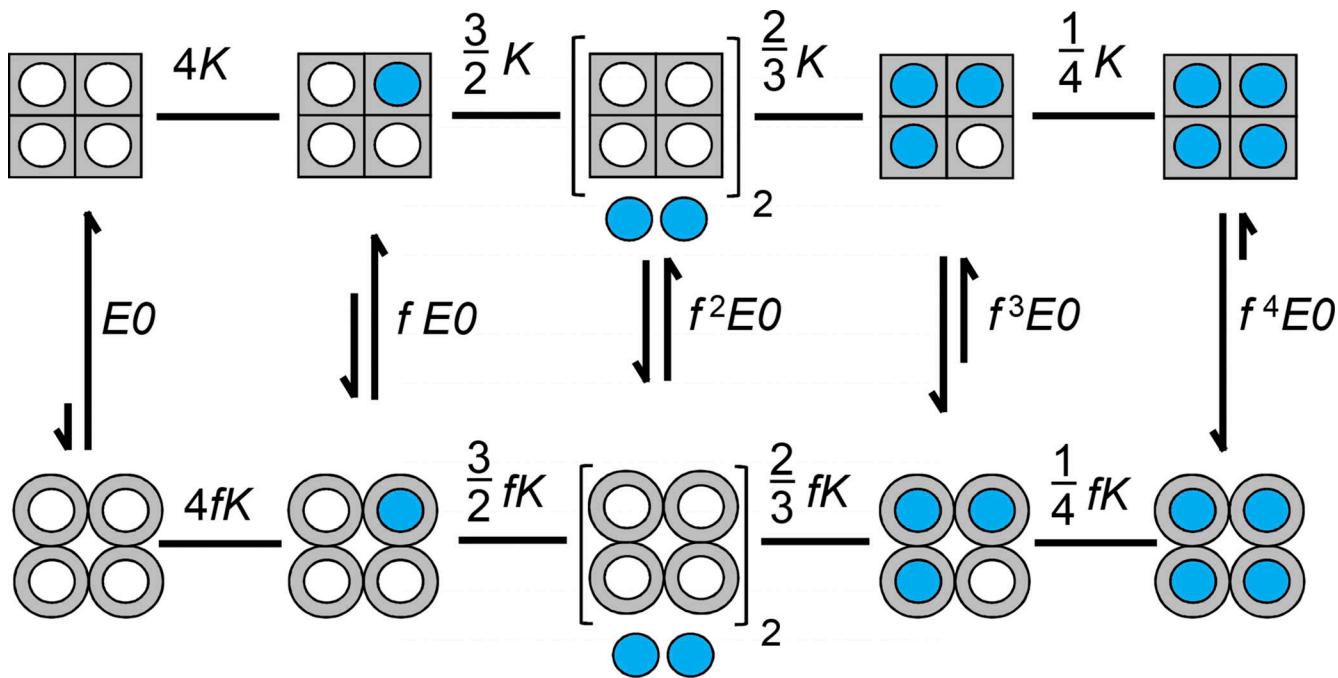
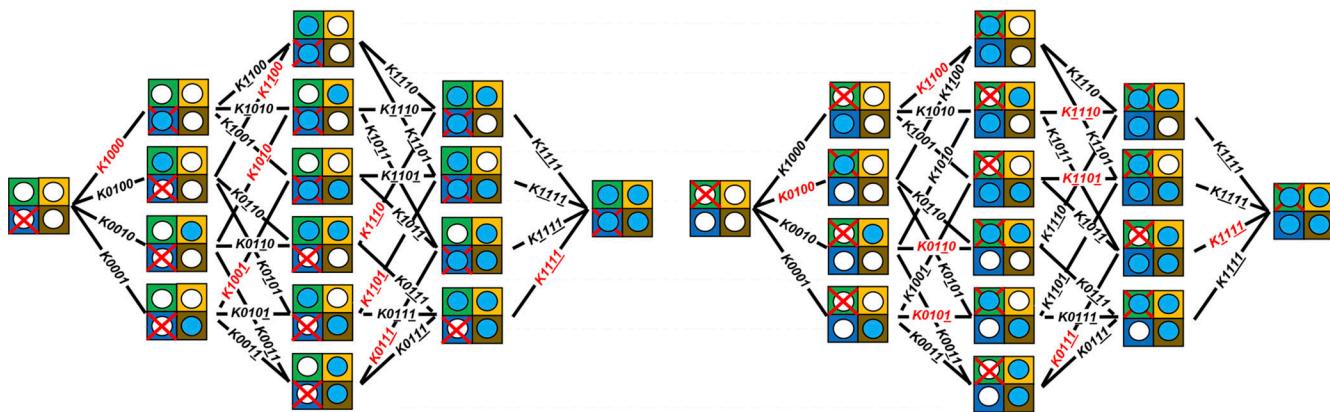


Figure S1. Structure of the MWC model. The empty binding site (white circle) of each subunit can bind one ligand (blue circle) in either the resting (gray squares) or activated (gray circles) state. The equilibrium association constant is equal for all binding steps in the resting protein (K ; mol^{-1}), and in the activated protein (fK ; mol^{-1}), resulting in the indicated stoichiometric factors. f is a constant allosteric factor. A joint allosteric step with the equilibrium isomerization constant f^nE_0 ($f > 1$, $n = 0 \dots 4$) leads to progressive activation. For the double-liganded channel, the ligands are drawn below the channel cartoon to illustrate that there are two options for binding, adjacent or diagonal.

1m234

12m34



123m4

1234m

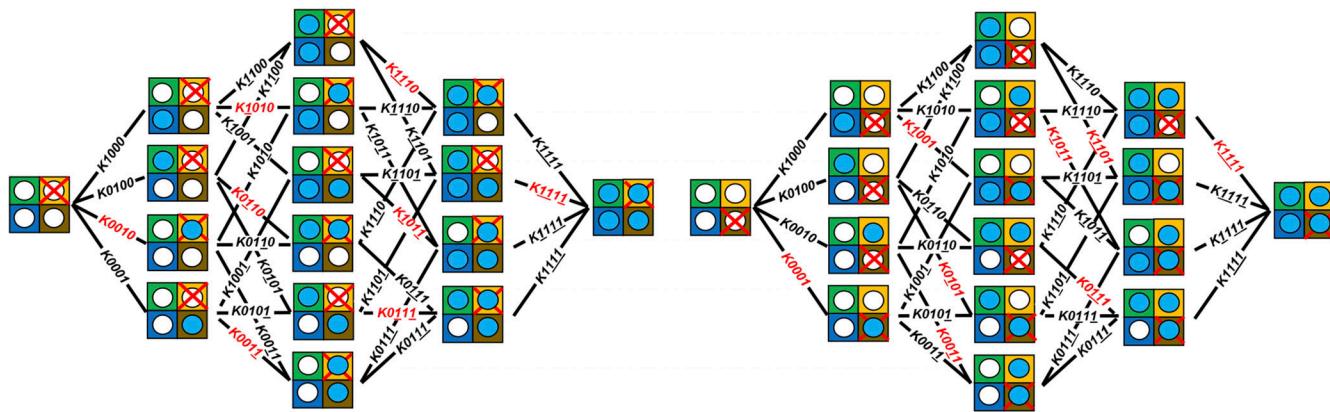


Figure S2. HA models for four concatamers containing one disabled binding site. The concatamers build, together with the two models in Fig. 1 and the models in Figs. S3 and S4, the 16 models of the CHA model. Blue circles represent a ligand, white circles represent an empty binding site, and a red cross on a white circle represents a disabled binding site. Equilibrium association constants for ligand binding, K_{xxxx} ($x = 0$ for an empty subunit, $x = 1$ for a subunit to be occupied, and $x = 1$ for a preoccupied subunit) are indicated in black for a WT and red for a disabled subunit. One HA model contains 32 K_{xxxx} , either black or red. Shown is only the network for ligand binding. Each state model has to be complemented by the scheme of the closed–open isomerizations shown at the bottom of Fig. 1.

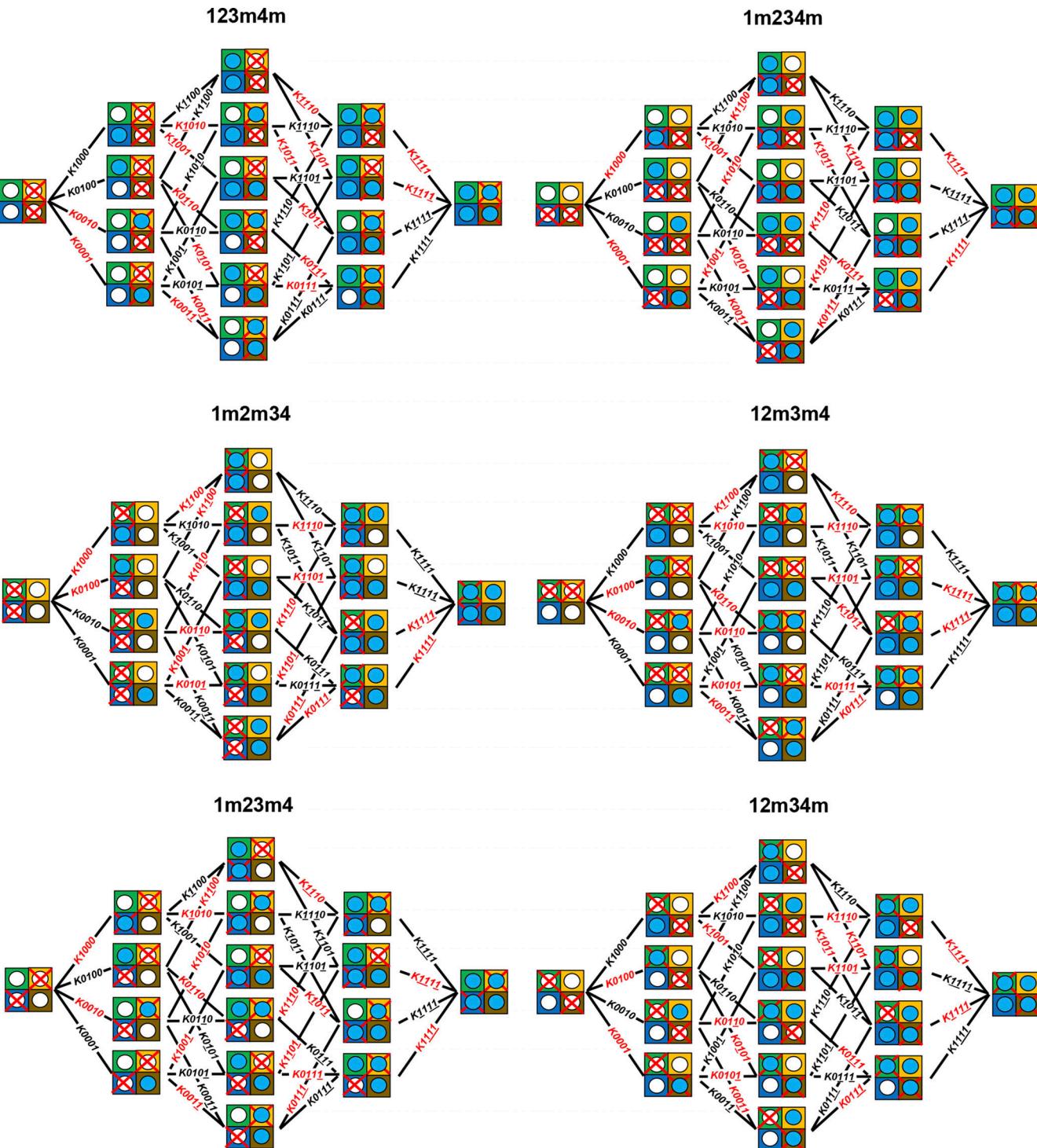
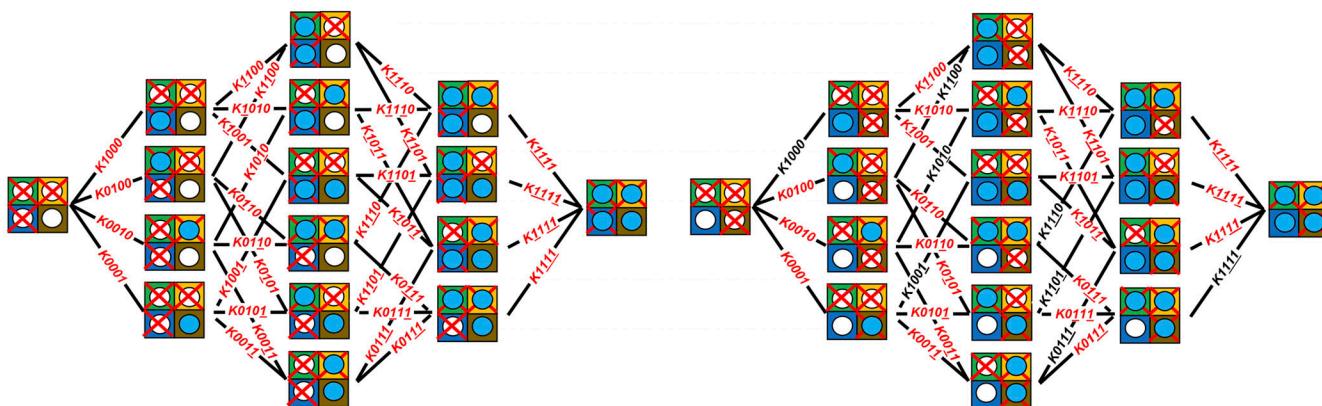
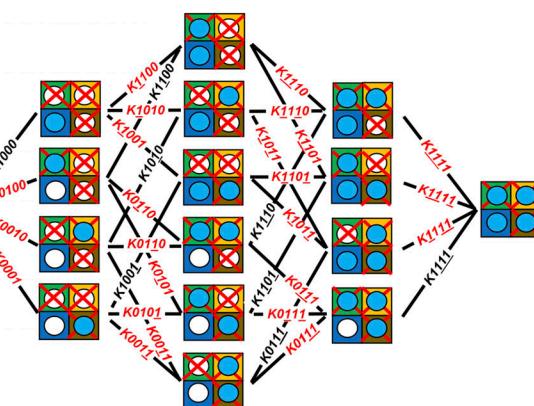


Figure S3. **HA models for six concatamers containing two disabled binding sites.** The concatamers build, together with the two models in Fig. 1 and the models in Figs. S2 and S4, the 16 models of the CHA model. For further explanation see legend to Fig. S2.

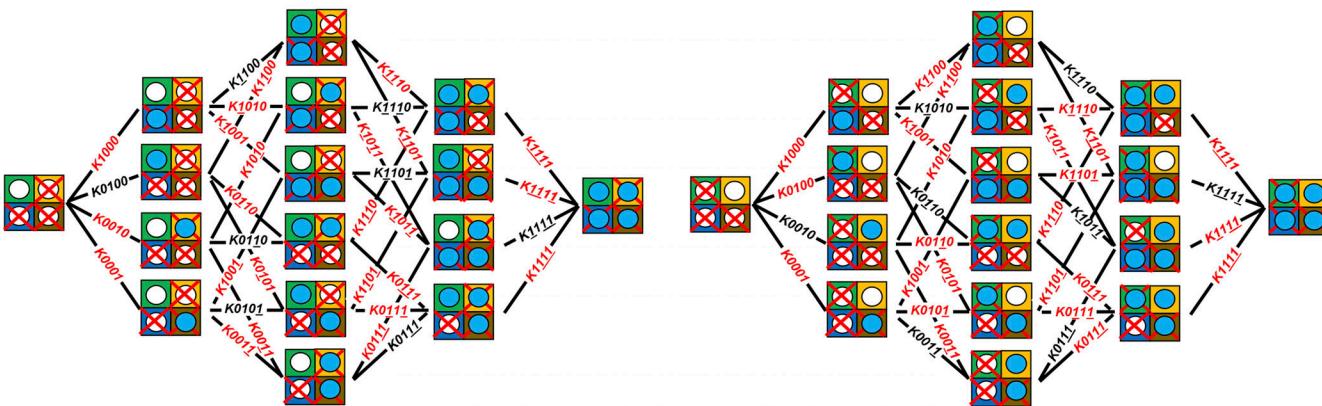
1m2m3m4



12m3m4m



1m23m4m



1m2m34m

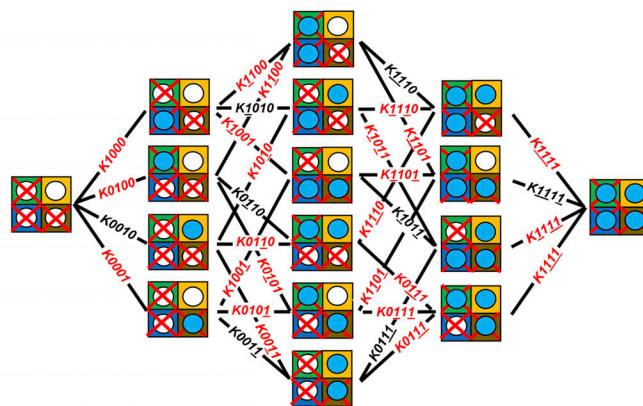


Figure S4. **HA models for four concatamers containing three disabled binding sites.** The concatamers build, together with the two models in Fig. 1 and the models in Figs. S2 and S3, the 16 models of the CHA model. For further explanation see legend to Fig. S2.

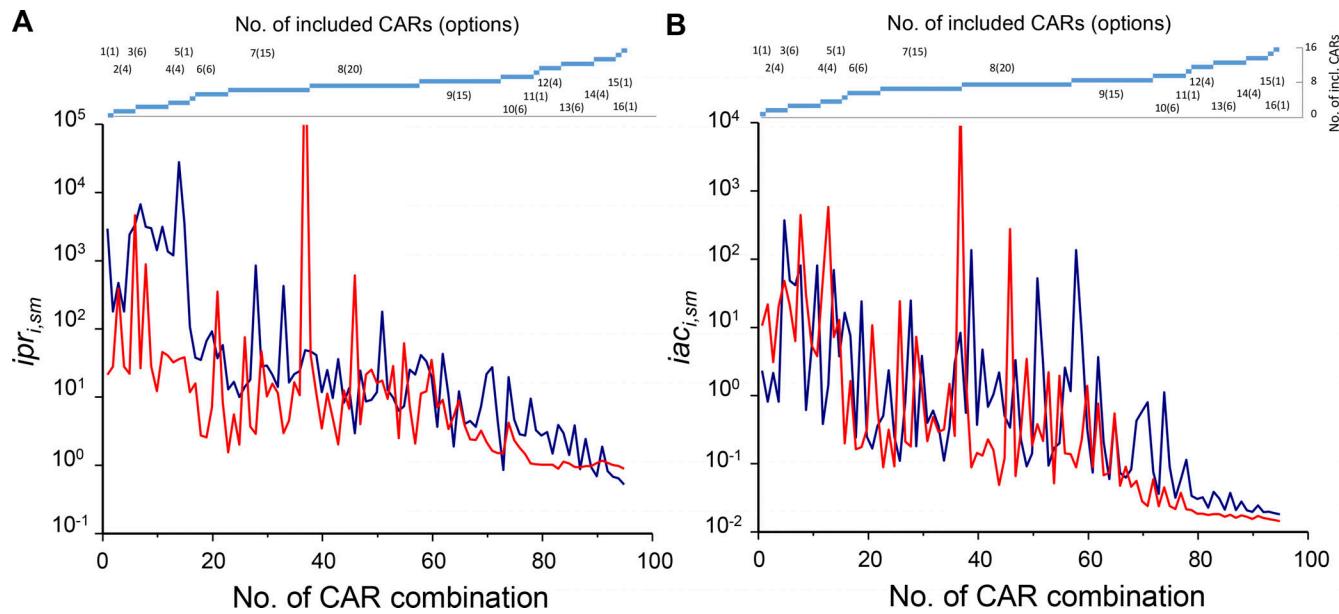


Figure S5. **Comparison of fits with 35 either equidistant or clustered data points.** Equidistant data points (blue) were generated for each CAR specifically, running from $P_o = 0.01$ to $P_o = 0.99$. **(A and B)** For clustering, the 35 data points (red) were grouped to 7 equidistant portions containing 5 data points each. All 95 CAR combinations were tested (cf. Table 1). **(A)** Imprecision. **(B)** Inaccuracy.

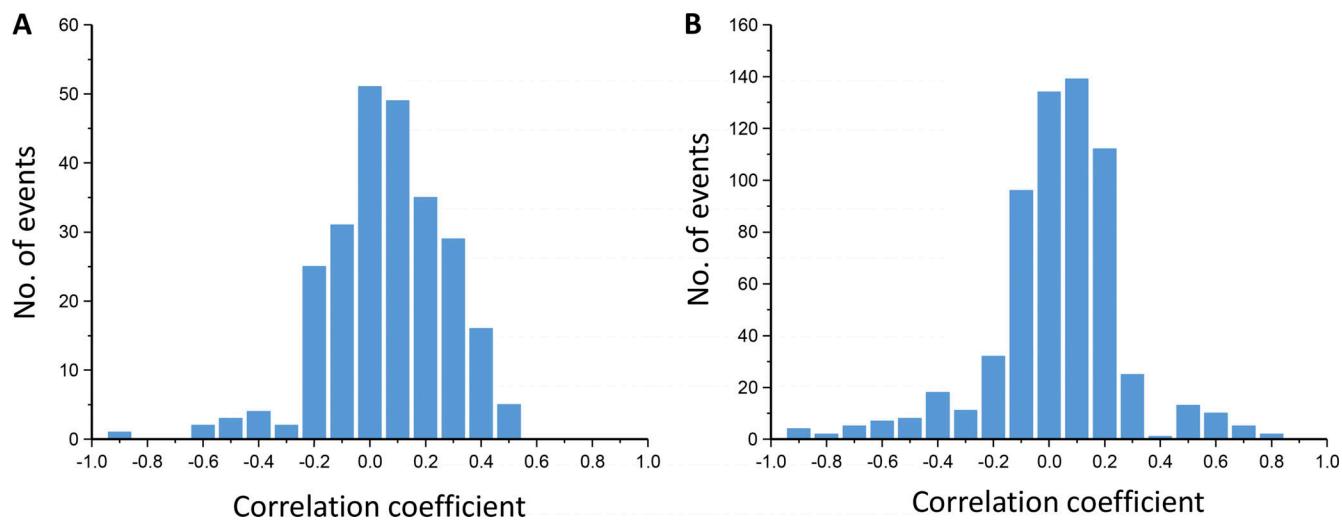


Figure S6. Correlation coefficients. **(A)** Histogram of correlation coefficients between the fit parameters Ex , Z_i , and fdx . The variances of the parameters themselves (correlation coefficient = 1) were omitted. Data are the mean of 100 converging fits. Noise factor 0.5. Most correlation coefficients are -0.3 to 0.3 , with few exceptions. An outstanding negative correlation exists between Z_{15} and $E4$. For individual values, see Table S4. **(B)** Histogram of correlation coefficients between the Ex and K_i . The K_i values were calculated from the Z_i values as described in Materials and methods. For individual values, see Table S5.

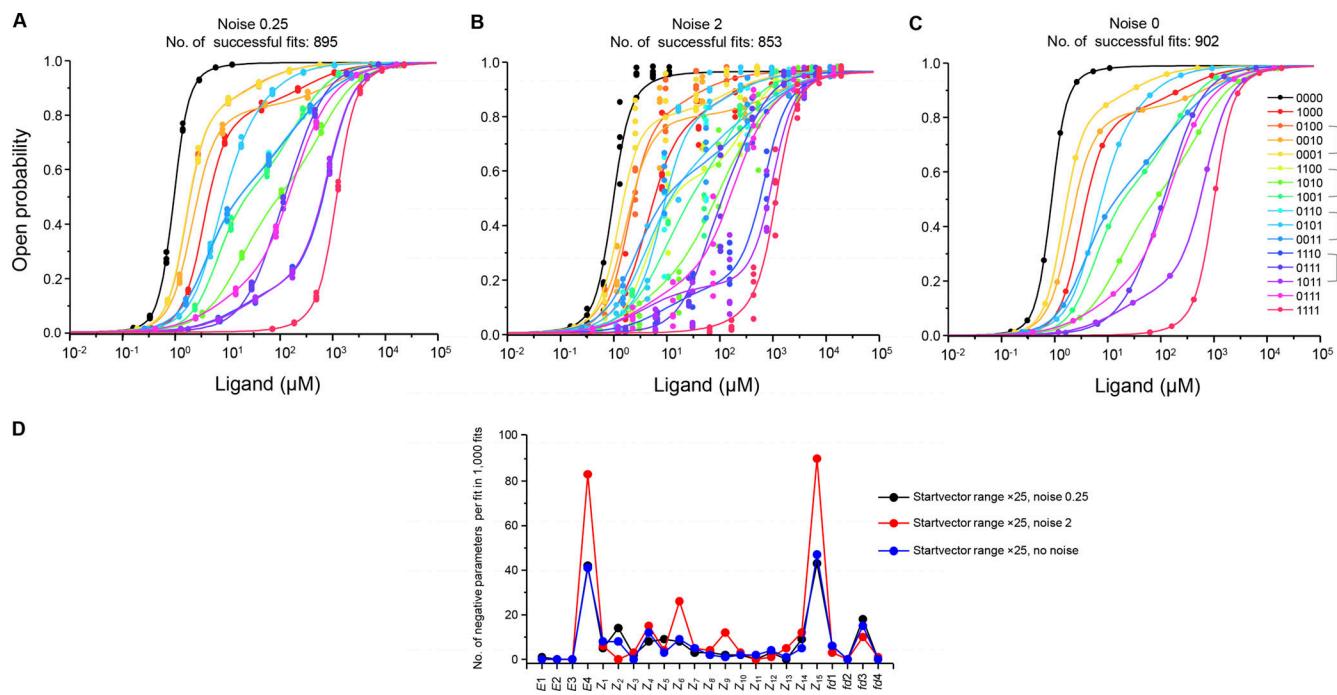


Figure S7. Effect of data noise on the fit at stochastic variation of the true start vector. Five data points at seven concentrations were used to refit the CHA₁₆ model 1,000 times. Representative examples of successful fits are shown. Each parameter was stochastically varied 5-fold in both directions, resulting in a 25-fold total variation range. **(A)** Standard conditions with a noise level of 0.25. **(B)** Noise level 2.0. **(C)** No noise. **(D)** Incidence of negative parameters for the fits in A–C.

Provided online are seven tables. Table S1 shows constants used to simulate the 16 CARs. Table S2 provides equilibrium association constants and their genesis. Table S3 provides a comparison of true values, fitted parameters, and errors for an example fit. Table S4 shows a matrix of correlation coefficients for the fit parameters Ex , Z_i , and fdx . Table S5 shows a matrix of correlation coefficients for Ex and K_i . Table S6 summarizes effects of stochastic variation of the true start vector on the fit outcome at different conditions. Table S7 assigns the parameters Z_1 – Z_{15} to the 15 Z_{xxxx} .