

**ARTICLE**

# Identifiability of equilibrium constants for receptors with two to five binding sites

Klaus Benndorf<sup>1</sup> and Eckhard Schulz<sup>2</sup>

**Ligand-gated ion channels (LGICs)** are regularly oligomers containing between two and five binding sites for ligands. Neither in homomeric nor heteromeric LGICs the activation process evoked by the ligand binding is fully understood. Here, we show on theoretical grounds that for LGICs with two to five binding sites, the cooperativity upon channel activation can be determined in considerable detail. The main requirements for our strategy are a defined number of binding sites in a channel, which can be achieved by concatenation, a systematic mutation of all binding sites and a global fit of all concentration-activation relationships (CARs) with corresponding intimately coupled Markovian state models. We take advantage of translating these state models to cubes with dimensions 2, 3, 4, and 5. We show that the maximum possible number of CARs for these LGICs specify all 7, 13, 23, and 41 independent model parameters, respectively, which directly provide all equilibrium constants within the respective schemes. Moreover, a fit that uses stochastically varied scaled unitary start vectors enables the determination of all parameters, without any bias imposed by specific start vectors. A comparison of the outcome of the analyses for the models with 2 to 5 binding sites showed that the identifiability of the parameters is best for a case with 5 binding sites and 41 parameters. Our strategy can be used to analyze experimental data of other LGICs and may be applicable to voltage-gated ion channels and metabotropic receptors.

## Introduction

The function of receptor proteins embedded in cell membranes is controlled by the binding of ligands to highly specific binding sites. Following ligand binding, the receptors exert conformational changes, highly specific as well, that evoke secondary responses in either the membrane or the cell interior. The most simple-structured ligands are ions, e.g.,  $\text{Ca}^{2+}$  ions, but most ligands are organic molecules, including highly diverse neurotransmitters, or peptide- and proteohormones.

In ionotropic receptors, the binding of ligands controls the open probability of a pore for ions. To realize a pore structure, many of these receptors are oligomers composed of between two and five subunits. These subunits can be either identical or different, resulting in homomeric or heteromeric receptors, respectively. The oligomeric nature of the receptors is what causes the number of binding sites to be predominantly larger than one. The consequence is that the activation of the receptors is controlled by the binding of several ligands and the resulting conformational changes by each binding step influence each other, resulting in the phenomenon of cooperative receptor activation. Moreover, the cooperative conformational changes of the subunits feed back to the binding steps themselves in a

reciprocal fashion (Colquhoun, 1998; Kusch et al., 2010). This complexity of processes severely limits quantitative studies on receptor activation.

For >50 yr, the allosteric Monod-Wyman-Changeux (MWC) model (Monod et al., 1965) has been used to interpret the activation of symmetric proteins composed of identical subunits. The applicability of this model is based on the simple elegance of the assumption that the subunits perform a joint “allosteric” conformational change of all subunits and that the equilibrium of this allosteric step is shifted in proportion to the number of ligands bound. To keep the models manageable, often-defined stoichiometric factors have been used. This leads to an extremely low number of required parameters for describing the activation process, one equilibrium association constant, one equilibrium constant for the allosteric conformational change, and one allosteric factor. The MWC model has been widely applied, including various ionotropic membrane receptors (e.g.,  $\text{GABA}_A$  receptors [Steinbach and Akk, 2019], nicotinic acetyl-choline receptors [nAChRs; Lee and Sine, 2005], and also cyclic nucleotide-gated [CNG] channels [Goulding et al., 1994]) that are gated from the intracellular side. Moreover, the MWC model

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

Correspondence to Klaus Benndorf: [klaus.benndorf@med.uni-jena.de](mailto:klaus.benndorf@med.uni-jena.de).

© 2023 Benndorf and Schulz. 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/>).

was extensively used for quantifying the oxygen binding in the tetrameric hemoglobin molecule (Eaton et al., 1999), though it is composed of different types of subunits, two  $\alpha$ -chains and two  $\beta$ -chains in the case of adult hemoglobin (Fermi et al., 1984). However, the simplifying assumptions of using only a few parameters and fixed stoichiometries often severely limited detailed conclusions on the interaction of the binding sites and subunits in the framework of the MWC model, either for current measurements (Colquhoun, 1998) or binding measurements (Middendorf and Aldrich, 2017a, 2017b), even when applying state-of-the-art Bayesian strategies (Hines et al., 2014). For heteromeric receptors, the identifiability becomes even worse because any analysis would require more than one equilibrium association constant. Some remedy to overcome these problems in ligand-gated ion channels (LGICs) was to study subunit cooperativity by sophisticated single-channel analyses of burst activity, e.g., in homopentameric (Beato et al., 2007) or heteropentameric glycine receptors (GlyRs; Burzomato et al., 2004). An elegant alternative single-channel approach was to coexpress subunits with impeded binding sites and a second mutation in the pore to identify the subunit composition. This approach has been applied to homotetrameric CNG (Liu et al., 1998) and BK channels (Niu and Magleby, 2002), activated by intracellular cGMP and  $\text{Ca}^{2+}$ , respectively.

Recently, we quantified activation gating in heterotetrameric CNG channels composed of two CNGA2 subunits, one CNGA4 and one CNGB1b subunit, by systematically impeding the four binding sites and subjecting the resulting 16 concentration-activation relationships (CARs) to a global fit analysis of 16 complex respective kinetic schemes (Schirmeyer et al., 2021). This approach allowed us to describe receptor activation by 32 equilibrium association constants for the 4 binding sites and 4 closed-open isomerization constants. Consistency of the subunit composition was reached by concatenating the subunits. The high degree of complexity could be managed by transposing the models to 4-D hypercubes. Due to multiple microscopic reversibilities and by equating the constants of the two A2 subunits, this fit required only 17 free parameters, which is, however, still a formidable number. Nevertheless, the precision of our fit was surprisingly high.

We therefore studied on theoretical grounds how this unexpectedly good identifiability of parameters could be reached (Benndorf et al., 2022). We showed that with this strategy indeed multiple functional parameters of the individual subunits can be determined because the constraints in this approach are sufficiently high. Furthermore, we employed a stochastic fit approach using scaled unitary (SU) start vectors to avoid any bias arising from a specific start vector. We showed that 23 parameters could be determined reliably, including 15 virtual equilibrium-association constants, 4 equilibrium constants for the closed-open isomerizations, and 4 disabling factors for the mutations. The 15 virtual equilibrium-association constants allowed us to determine the 32 equilibrium-association constants specific to the subunits of a tetrameric channel at each degree of ligand binding.

It was therefore intriguing to generalize our strategy to receptors with lower and higher numbers of binding sites.

Biologically, the incidence of receptors with four subunits and four binding sites seems to dominate over the other constellations. For ionotropic receptors, typical tetrameric examples with four binding sites are AMPA receptors (Sobolevsky et al., 2009) or CNG channels (Li et al., 2017). Representative trimeric receptors with three binding sites are purinergic P2X receptors (Kawate et al., 2009) or acid-sensing ion channels (ASICs; Jasti et al., 2007). Pentameric ionotropic receptors belong mainly to the class of cys-loop receptors, including the families of nAChRs, GlyRs, and serotonin receptors (5HTRs). Many of these receptors are formed by two  $\alpha$ -subunits and three other homolog subunits. For example, the classical nAChR of both the electric eel and the electric fish (Changeux et al., 1970; Miledi and Potter, 1971) is composed of two  $\alpha$ -subunits, one  $\beta$ -subunit, one  $\gamma$ -subunit, and one  $\delta$ -subunit (Weill et al., 1974; daCosta and Baenziger, 2013), forming only two binding sites between the adjacent subunits  $\alpha$  and  $\gamma$  as well as  $\alpha$  and  $\delta$ . On the other hand, there are also homopentameric relatives with five binding sites, e.g., the  $\alpha 7$  nAChR (McKay et al., 2007) or the  $\alpha 1$  GlyR (Zhu and Gouaux, 2021). The key determinant for our analysis is not the number of subunits but the number of functional binding sites.

Herein, we elaborate for channels of known composition with two to five binding sites how promising it is to quantify receptor activation in great detail by our strategy, to impede the affinity of defined binding sites, and to globally fit all possible CAR combinations with either wild type or impeded binding sites by complex systems of coupled kinetic schemes. In other words, we empirically study the identifiability of multiple equilibrium constants for synthetic parameter sets by globally refitting multiple simulated CARs. We demonstrate that complex models with more binding sites and possible CARs can provide more accurate and precise results than simple models with less binding sites and possible CARs.

## Materials and methods

### Terminology and specification of models

To introduce the terminology used herein, we first consider the heteromeric allosteric (HA) model for the tetrameric channel as described previously (Schirmeyer et al., 2021) and which is also shown in its “systematic representation” in Fig. S1 A with respect to the number of binding sites. The model contains 32 different equilibrium association constants for the closed states,  $K_x$ , describing the binding of the ligand to the four binding sites for each occupation pattern of the other binding sites, and the 4 closed-open isomerization constants,  $E_x$ , for the transitions of the single to the quadruple liganded channel, thereby assuming that  $E_x$  is equal at an equal degree of liganding. Notably, due to microscopic reversibility, all equilibrium association constants for the open channel are implicitly given by the respective  $K_x$  of the closed channel and the two related  $E_x$ . At zero ligand,  $P_0$  was set to the negligible value  $10^{-5}$ . A tetrameric HA model is termed in the following as the “HA<sup>4</sup>” model.

For the previous global fit analysis, mutational factors,  $f_{d,x}$ , were introduced for each mutated binding site, describing the reduction of the affinity. With all combinations of wild type and mutated binding sites, this led to maximally 16 CARs and

corresponding 16 specific HA<sup>4</sup> models. These 16 HA<sup>4</sup> models build together the complex heterotetrameric allosteric CHA<sup>4</sup> model which was finally subjected to a global fit. This terminology is used here throughout. The superscript indicates the number of binding sites.

Notably, the multifaceted network of binding steps in [Fig. S1 A](#) could be elegantly analyzed by transforming it into a 4-D hypercube ([Fig. S1 B](#)) and introducing 15 independent virtual equilibrium association constants,  $Z_x$  ([Fig. S1 C](#)), determining the 32  $K_x$  by simple ratios ([Schirmeyer et al., 2021](#)). This representation of a model will be termed herein as “cube representation” and will also be used for channels with two to five dimensions.

Because the number of binding sites can deviate from the number of subunits and to generalize our approach, we categorize the LGICs herein by the actual number of binding sites and not by the number of actual subunits. Hence, the superscript at the end indicates the number of actual binding sites. For example, a tetrameric CNG channel with four binding sites would be described by a CHA<sup>4</sup> model and a pentameric nAChR with two binding sites by a CHA<sup>2</sup> model.

### Computation of the models

For the LGICs with two to five binding sites, four types of respective HA models are considered herein. To keep track, we use the nomenclature “HA<sup>x</sup>” model, where  $x$  denotes the number of ligands. To include the effects of the mutations, each mutation generates its own HA<sup>x</sup> model for each CAR. The number of HA<sup>x</sup> models included in a complex CHA<sup>x</sup> model is given by  $n_{CAR,max}$  ([Eq. 21](#)) and is found to be 4, 8, 16, and 32 for  $x = 2, 3, 4$ , and 5, respectively.

For all computations, we assumed that microscopic reversibility ([Colquhoun and Hawkes, 1995](#)) is obeyed, resulting in the ligand binding to a defined number of independent equilibrium constants on which the others depend ([Colquhoun et al., 2004](#)).

In our previous analysis on heterotetrameric channels, we eased the considerations by interpreting the 16 closed states of our HA<sup>4</sup> models as corners of a 4-D hypercube in which the transitions, specified by the 32  $K_x$ , form the 32 edges ([Fig. S1 B; Benndorf et al., 2022](#)).

Herein, we essentially follow this logic: given state 0, the 4-D hypercube defines 15 virtual equilibrium association constants (violet lines  $Z_1$ – $Z_{15}$  in [Fig. S1 C](#)), which allow the computation of the 32  $K_x$  by respective ratios of two  $Z_x$  (see below). Microscopic reversibility is directly obeyed. For disabling the binding sites, we assume that the affinity of a binding site is reduced by a factor  $fd_x < 1$  ( $x = 1\dots4$ ). As in our previous approach, the  $fd_x$  factors were considered to affect solely the affinity at a binding site for the ligand. For the equilibrium constants of the closed-open isomerizations,  $E_x$  ( $x = 1\dots4$ ; [Fig. S1 A](#), bottom), it is assumed that they are equal at an equal degree of liganding and independent of where the binding site is occupied.  $E_0$  is set to the negligible value of  $10^{-5}$ . The total number of free parameters for the global fit of the 16 concatamers is then  $15 + 4 + 4 = 23$ .

To explain the rationale of the analysis, it is useful to skip to our earlier nomenclature. The closed state 0 is then termed C0000, where the four zeros tell that no binding site is occupied.

The other 15 closed states in the 4-D hypercube are then given by Cijkl, where i, j, k, and l can be either 0 (empty) or 1 (occupied).

If denoting for the unliganded closed state C0000 the occupation probability as  $p_c0000$ , the occupation probability  $p_cijkl$  of each other closed state Cijkl can be described by the virtual equilibrium constants  $Zijkl$

$$p_cijkl = p_c0000 \cdot Zijkl \cdot fd_1^{\gamma_1} \cdot fd_2^{\gamma_2} \cdot fd_3^{\gamma_3} \cdot fd_4^{\gamma_4} \cdot L^\alpha. \quad (1)$$

The ligand concentration  $L$  has the power of the sum of the indices  $\alpha = i + j + k + l$ , corresponding to the number of bound ligands. The factors of disabling by the mutational factor  $fd_u$  ( $u = 1\dots4$ ) of the four binding sites have the exponents  $\gamma_1$  to  $\gamma_4$ , which are also either 0 or 1. An exponent is 1 if the binding site  $u$  is mutated and has bound a ligand. Otherwise, the exponent is 0.

The occupation probability of an open state Oijkl is obtained by multiplying  $p_cijkl$  with one of the associated opening equilibrium constants  $E_\alpha$  in the form

$$p_{ijkl} = p_c0000 \cdot Zijkl \cdot fd_1^{\gamma_1} \cdot fd_2^{\gamma_2} \cdot fd_3^{\gamma_3} \cdot fd_4^{\gamma_4} \cdot L^\alpha \cdot E_\alpha. \quad (2)$$

To ease the notation, five subtotals are defined:

$$S_\alpha = \sum_{i+j+k+l=\alpha} (Zijkl \cdot fd_1^{\gamma_1} \cdot fd_2^{\gamma_2} \cdot fd_3^{\gamma_3} \cdot fd_4^{\gamma_4}), \quad (3)$$

where the summation extends over all combinations of i, j, k, and l with the same sum  $\alpha = 0\dots4$ .

Then the open probability  $P_{oc}(L)$  of the whole channel can be calculated by the ratio of two polynomials:

$$P_{oc}(L) = \left( \sum_{\alpha=0}^4 E_\alpha S_\alpha L^\alpha \right) \cdot \left[ \sum_{\alpha=0}^4 (1 + E_\alpha) S_\alpha L^\alpha \right]^{-1}. \quad (4)$$

[Eqs. 1, 2, and 3](#) can be easily adapted to the other HA<sup>x</sup> models. A vector  $\mathbf{r} = (i, j)$ ,  $\mathbf{r} = (i, j, k)$ ,  $\mathbf{r} = (i, j, k, l)$ , or  $\mathbf{r} = (i, j, k, l, m)$  is used to define the state for the HA<sup>2</sup>, HA<sup>3</sup>, HA<sup>4</sup>, or HA<sup>5</sup> model, respectively. The sum of the components of  $\mathbf{r}$  gives the sum norm  $||\mathbf{r}||$ . Including the occupation probability of the closed unliganded state  $p_c(\mathbf{0})$ , it follows

$$p_c(\mathbf{r}) = p_c(\mathbf{0}) \cdot Z(\mathbf{r}) \cdot \left( \prod_{i=1}^x fd_i^{\gamma_i} \right) \cdot L^\alpha, \quad (5)$$

$$p_o(\mathbf{r}) = p_c(\mathbf{0}) \cdot Z(\mathbf{r}) \cdot \left( \prod_{i=1}^x fd_i^{\gamma_i} \right) \cdot L^\alpha \cdot E_\alpha, \quad (6)$$

where  $\alpha$  is the sum of the components of the vector  $\mathbf{r}$ . For the subtotals  $S_\alpha$ , the summation extends over all combinations, where the components of  $\mathbf{r}$  give the same sum  $\alpha$ :

$$S_\alpha = \sum_{||\mathbf{r}||=\alpha} \left[ Z(\mathbf{r}) \cdot \left( \prod_{i=1}^x fd_i^{\gamma_i} \right) \right], \quad (7)$$

with  $\alpha = 0\dots x$  and  $S_0 = 1$  for all HA<sup>x</sup> models.

The open probability  $P_{oc}(L)$  of a HA<sup>x</sup> model ( $x = 2\dots5$ ) can then be written in the canonical form

$$P_{oc}(L) = \left( \sum_{\alpha=0}^x E_\alpha S_\alpha L^\alpha \right) \cdot \left[ \sum_{\alpha=0}^x (1 + E_\alpha) S_\alpha L^\alpha \right]^{-1}. \quad (8)$$

This equation was used for both simulating the data points and computation of the CARs by the respective model in the fits.

Since the formula requires a value for  $E_0$  though no detectable  $P_o$  was assumed with no ligand bound, we set  $E_0$  generally to the negligible value of  $10^{-5}$ .

### Simulation of data points

$P_o$  of noise-free CARs for the LGIC<sup>\*</sup>s were simulated by Eq. 8. To mimic typical experimental conditions, for each CAR seven concentrations between  $P_o = 0.01$  and  $P_o = 0.99$  were determined with an equal distance on a logarithmic scale. Then for each of these seven concentrations, five data points were generated by adding a stochastic noise,  $n_{st}$ , to the noise-free  $P_o$  value according to

$$n_{st} = 0.25 \cdot R \cdot [P_o \cdot (1 - P_o)], \quad (9)$$

where  $R$  is an equally distributed random number between -1 and 1. This noise level was also used as standard for previous simulations (Benndorf et al., 2022). It was used herein for all simulations. The result was that each CAR was determined in total by 35 stochastically scattering data points. This resulted for the maximum of possible combinations of wild type and mutated binding sites in  $n_{CARmax}$  of 4, 8, 16, and 32 CARs for the CHA<sup>2</sup>, CHA<sup>3</sup>, CHA<sup>4</sup>, and CHA<sup>5</sup> model, respectively.

### Global fit

The nlinfit routine of the Matlab software was used to globally fit the CHA models, operating with an iterative least square estimation. The maximum iteration number was set to 100. The routine provides the fit parameters,  $x_i$  ( $E_x$ ,  $Z_x$ ,  $fd_x$ ), the covariance matrix,  $Cov_{EZfd}$ , with the elements  $\sigma_{ij}$  of which the main diagonal gives, with  $i = j$ , the values of the variances and standard deviations  $\sigma_i^2$  and  $\sigma_i$  of the fit parameters, respectively. The  $Z_x$  allow to compute all respective equilibrium association constants,  $K_x$ , by simple ratios (Tables S1, S2, S3, and S4).

Relative standard deviations,  $\sigma_{i,rel}$ , of the parameters were determined according to

$$\sigma_{i,rel} = \sigma_i / x_i. \quad (10)$$

These relative standard deviations are useful for comparing the uncertainty of different parameters. Here, two different measures are used to quantify the uncertainties of the parameters in the fit, imprecision and inaccuracy (Benndorf et al., 2022).

### Imprecision

The term precision of a parameter  $x_i$  specifies its quality, without knowing that the parameter has been determined correctly. Consequently, the term imprecision,  $ipr_i$ , denotes a statistical variability of a parameter with respect to its determined mean. Hence,  $ipr_i$  is given by  $\sigma_{i,rel}$ .

The mean imprecision of a parameter  $x_i$  from  $n$  consecutive fits was calculated by

$$ipr_{i,n} = \frac{1}{n} \sum_{j=1}^n ipr_{i,j}. \quad (11)$$

From  $ipr_{i,n}$ , the mean imprecision of all  $n_p$  parameters was used as a first overall measure for the goodness of a fit according to

$$mipr_{n_p,n} = \frac{1}{n_p} \sum_{i=1}^{n_p} ipr_{i,n}. \quad (12)$$

### Inaccuracy

The term accuracy of a parameter specifies the quality of a fitted parameter  $x_i$  with respect to the true value,  $x_{i,tr}$ . This value is accessible only in simulated data as analyzed herein. The term inaccuracy,  $iac_i$ , indicates the relative deviation of a parameter  $x_i$  from the true value  $x_{i,tr}$ . One way to calculate the relative inaccuracy of a parameter  $x_i$  is given by

$$iac_i = \frac{1}{x_{i,tr}} \sqrt{(x_i - x_{i,tr})^2}. \quad (13)$$

From  $iac_i$  we calculated the mean of  $n$  consecutive fits,  $iac_{i,n}$ , according to

$$iac_{i,n} = \frac{1}{n} \sum_{j=1}^n iac_{i,j}. \quad (14)$$

From  $iac_{i,n}$  we calculated the mean inaccuracy, a second overall measure for all  $n_p$  parameters according to

$$miac_{n_p,n} = \frac{1}{n_p} \sum_{i=1}^{n_p} iac_{i,n}. \quad (15)$$

### Correlation matrices

To obtain the correlation matrix for the parameters, we first built the respective covariance matrix  $Cov_{EZfd}$  provided by the fit. It is then easy to calculate the correlation matrices  $Corr_{EZfd}$  containing the elements

$$\rho_{ij} = Cov_{EZfd}(i,j) / (\sigma_i \cdot \sigma_j). \quad (16)$$

### Stochastic procedure to identify best fits for a given set of data points

To refine the search for the best minimum among fits providing exclusively parameters with positive sign (successful<sub>p</sub> fits), we used  $MSE_x$  values provided by the nlinfit routine of MATLAB for each fit  $x$ . We identified the minimum  $MSE_x$  value,  $Min(MSE_x)$ , and calculated the difference of each  $MSE_x$  to  $Min(MSE_x)$  by

$$\Delta MSE_x = MSE_x - Min(MSE_x) \quad (17)$$

and sorted the  $\Delta MSE_x$  values with respect to their size starting at  $Min(MSE_x)$ . Then the  $\Delta MSE_x$  values were plotted for  $x \geq 1$  on a logarithmic scale along the sorted sequence. A threshold at  $\Delta MSE_x = 10^{-10}$  was set, and the fits identified in this way, termed successful<sub>p $\Delta$ MSE</sub> fits, were used for further analysis.

### Computation times

Computing 1,000 of the most challenging fits with stochastically varied scaled unitary start vectors lasted about 10 min for the least complex CHA<sup>2</sup> model and about 300 min for the most complex CHA<sup>5</sup> model (Fujitsu Lifebook U series; vPRO i7).

### Online supplemental material

Fig. S1 shows the structure of the HA<sup>4</sup> model in systematic representation and as 4-D hypercube as published earlier. Fig. S2

illustrates by bar graphs how the fit parameters were set for the HA<sup>2</sup> to HA<sup>5</sup> model. **Fig. S3** shows the inaccuracy and imprecision of the parameters for 1,000 global fits of the CHA<sup>2</sup> to CHA<sup>5</sup> model using a start vector with the true values. **Fig. S4** provides color-coded correlation matrices for the CHA<sup>x</sup> models. **Fig. S5** illustrates the effects of mutational factors on the identifiability of  $E_x$  for the CHA<sup>x</sup> models. **Fig. S6** shows the effects of weakening the mutational factors on the identifiability of the  $Z_x$  for the CHA<sup>x</sup> models. **Fig. S7** shows plots of the inaccuracy and imprecision of the parameters for equal SU start vectors with stochastically varied elements. **Fig. S8** illustrates successful<sub>p<sub>ΔMSE</sub></sub> fits with different stochastically varied SU start vectors and stochastically varied elements. **Fig. S9** provides histograms of successful<sub>p<sub>ΔMSE</sub></sub> fits with different SU vectors and stochastically varied elements. **Fig. S10** shows the fit parameters determined with different SU vectors and stochastically varied elements and compares them with the true parameters. **Fig. S11** illustrates the inaccuracy and imprecision of the parameters for successful<sub>p<sub>ΔMSE</sub></sub> fits with different SU vectors and stochastically varied elements. Tables S1, S2, S3, and S4 provide the generation of  $K_x$  from  $Z_x$  in the HA<sup>2</sup> to HA<sup>5</sup> model, respectively. In the supplemental text, the algorithms for specifying the parameters for the CHA<sup>x</sup> models are described.

## Results

### Channels to be analyzed

We chose to consider LGICs with two to five binding sites because these binding-site numbers exist in natural LGICs. The respective HA<sup>2</sup> to HA<sup>5</sup> models are shown in their systematic representation in **Fig. 1, A–D**. We first addressed the question as to what extent more or less binding sites affect the number of CAR combinations and the number of free parameters in respective fits. **Fig. 2** illustrates by the models in cube representation that their structure simply follows a respective change of the dimension  $n_D$ . HA<sup>2</sup> models have 4 states arranged in a 2-D square, HA<sup>3</sup> models have 8 states arranged in a 3-D cube, HA<sup>4</sup> models have 16 states arranged in a 4-D hypercube, and HA<sup>5</sup> models have 32 states arranged in a 5-D hypercube. The corresponding numbers of  $K_x$ ,  $n_{Kx}$ , are given by the number of edges in these structures according to

$$n_{Kx} = n_D \cdot 2^{(n_D-1)} \quad (18)$$

to be 4, 12, 32, and 80.

In the following, the analysis logic with the virtual equilibrium constants is used. Their number,  $n_{Zx}$ , is given by

$$n_{Zx} = 2^{n_D} - 1. \quad (19)$$

The respective values are 3, 7, 15, and 31 (**Fig. 3**). The  $n_{Zx}$  values also tell directly the number of independent parameters in a fit for ligand binding. The  $Z_x$  define all respective  $K_x$  (Tables S1, S2, S3, and S4).

The numbers of  $E_x$  and  $fd_x$  are given by  $n_D$  directly. The total number of parameters in a global fit is obtained by

$$n_p = n_{Zx} + 2n_D = 2^{n_D} - 1 + 2n_D, \quad (20)$$

resulting in 7, 13, 23, and 41 free parameters for the  $n_D = 2, 3, 4$ , and 5, respectively. Hence, the number of binding sites increases

the number of free parameters in a nonlinear fashion, a fact that compromises the chances for a successful fit.

On the other hand, the maximum number of CARs,  $n_{CARmax}$ , increases with increased numbers of binding sites and their mutations, generating additional constraints.  $n_{CARmax}$  is given by

$$n_{CARmax} = 2^{n_D}, \quad (21)$$

resulting in 4, 8, 16, and 32 CARs for  $n_D$  of 2, 3, 4, and 5, respectively. This fact increases the chances for a successful fit and competes with the above effect.

A plot of the ratio  $n_p/n_{CARmax}$  as a function of the dimension  $n_D$  shows a decrease in the higher dimensions (**Fig. 4**). Hence, if all possible CARs are included, this suggests that the chances for a reasonable fit are poorest in the 2-D case and largest in the 5-D case, i.e., a fit with seven parameters would be less promising than that with as much as 41 parameters, respectively, although it is a priori not trivial to conclude this because of the different structures of the models. We therefore investigated the outcome of global fits with the CHA<sup>2</sup> through CHA<sup>5</sup> model systematically using the maximum numbers of CARs given by **Eq. 21**.

### Specification of theoretical fit parameters for CHA<sup>x</sup> models

Comparison of CHA models with different structures and number of ligands requires the use of related theoretical parameters. These theoretical parameters were generated in some relation to the previous parameters for the heterotetrameric CNG channels (**Schirmeyer et al., 2021; Benndorf et al., 2022**). Their exact derivation is described in the supplemental text at the end of the PDF, and their values are illustrated in **Fig. S2**.

For  $E_x$ , we set the value of the single-liganded channel to 10<sup>-1</sup> and that of the fully liganded channel to 10<sup>2</sup>, i.e., we assumed that maximal liganding causes a 1,000-fold increase of  $E_x$  with respect to  $E_l$ . The intermediate  $E_x$  values were set equidistantly on a logarithmic scale between  $E_l$  and the maximum  $E_x$  according to **Eq. 23**.

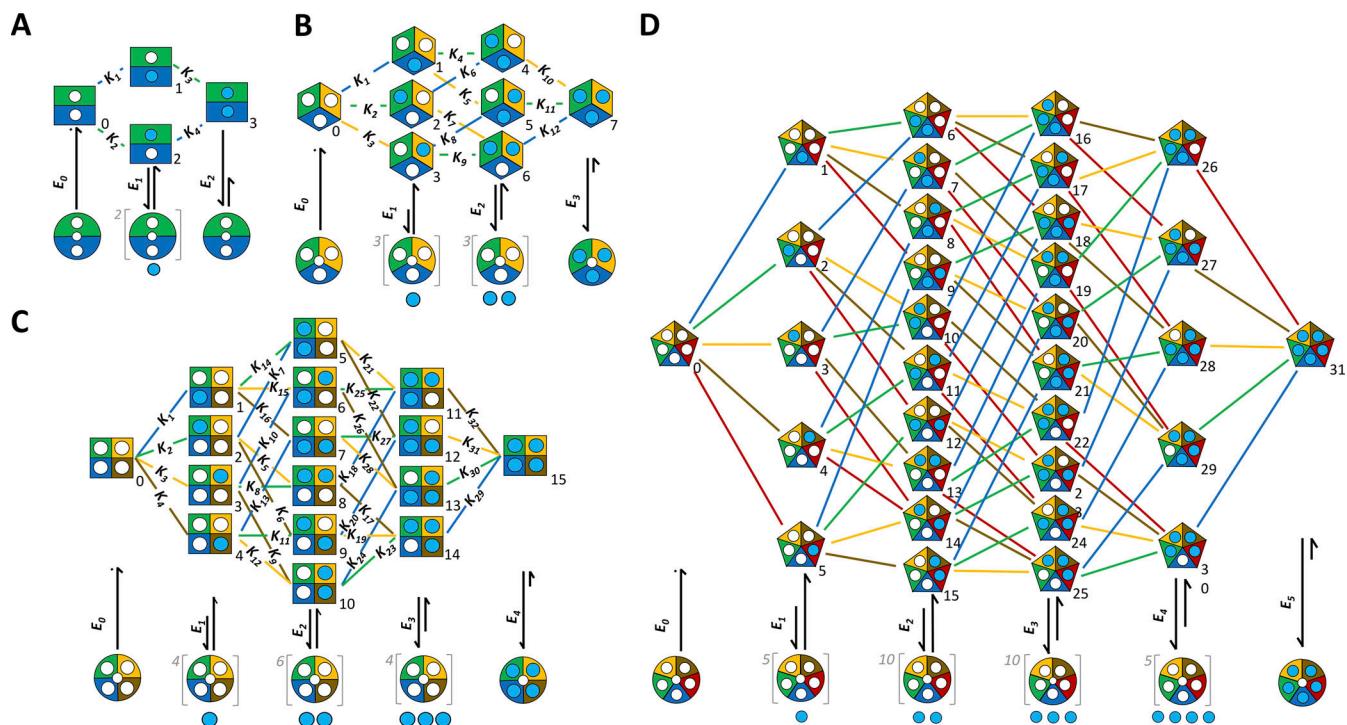
For  $fd_x$ , the situation is grossly analogous.  $fd_l$  was set to 10<sup>-4</sup> and that of the maximally impeded binding site was set to 10<sup>-2</sup>, resulting in a maximum difference between the binding sites of 100. Again, we set the intermediate  $fd_x$  values equidistantly on a logarithmic scale between  $fd_l$  and the maximum  $fd_x$  according to **Eq. 24**.

For  $Z_x$ , the situation is more complex because the number of  $Z_x$ ,  $n_{Zx}$ , rises exponentially according to **Eq. 19** and the options at an equal degree of liganding follow binomial statistics. The generation of the parameters are described by Eqs. 25, 26, 27, 28, 29, 30, 31, 32, 33, and 34.

While the values for the  $E_x$  and  $fd_x$  differ substantially for different  $x$ , quite similar to values found for heteromeric CNG channels (**Schirmeyer et al., 2021; Benndorf et al., 2022**), we chose a more challenging condition for the  $Z_x$  values to differ at an equal degree of liganding by only one order of magnitude. Moreover, they also superimpose significantly for different degrees of liganding.

### Simulating the CARs

With these equilibrium constants, we computed families of data points with the CHA<sup>2</sup>, CHA<sup>3</sup>, CHA<sup>4</sup>, and CHA<sup>5</sup> model by



**Figure 1. HA<sup>x</sup> models in systematic representation.** Blue circles, ligand; white circles, empty binding site. The set of closed–open isomerizations on the bottom is valid for all HA<sup>x</sup> models at an equal degree of liganding. For open channels with incomplete occupation of the binding sites, the ligands are drawn below the channel cartoon to illustrate that there is more than one option for occupancy. The  $K_x$  indicate the equilibrium association constants corresponding to Tables S1, S2, S3, and S4. **(A)** HA<sup>2</sup> model. **(B)** HA<sup>3</sup> model. **(C)** HA<sup>4</sup> model. **(D)** HA<sup>5</sup> model. In the HA<sup>5</sup> model, the  $K_x$  were omitted for clarity but are provided by Table S4.

including the respective maximum numbers of CARs and using Eq. 8 (Fig. 5). The noise was set to 0.25 (see Materials and methods), a value resembling the scatter in reasonable experimental measurements.

According to the set parameters for respective wild type and maximally mutated channels, these “extreme” CARs cover a similar concentration range and the CARs with intermediate degrees of mutations fill the range between the two extreme CARs, thereby developing characteristic shapes.

We first compared the fitted parameters,  $x_i$ , with the true parameters,  $x_{i,tr}$ . To increase the robustness of the result, we repeated the global fit 1,000 times, thereby generating the data points for each fit de novo according to Eq. 9. Despite taking the true values as the start vector, a portion of the fits failed, as identified by at least one negative parameter, a physically senseless result. In the following, all remaining non-failing fits with only positive parameters are termed successful<sub>p</sub> fits (Fig. 6 A), of which it is a priori not clear if they have run to the correct minimum or to another minimum. The diagram in Fig. 6 A suggests that models with higher numbers of binding sites have more constraints than fits with lower number of binding sites.

From these successful<sub>p</sub> fits, the means for each parameter and their standard deviation were computed and plotted together with the respective true parameter in bar graphs (Fig. 7). The mean relative standard deviations  $\sigma_{i,rel}$  for each parameter of 1,000 fits ( $ipr_{i,1000}$ ; Eq. 11) are plotted below each bar graph as

a fraction of 1 to illustrate the relative errors. The following conclusions can be drawn: (1) the global fits found with all models for each mean parameter is a minimum near  $x_{i,tr}$ ; (2) the errors obtained from these fits were generally fairly small, but those of the last  $E_x$  as well as  $Z_1$  and the last  $Z_x$  exceeded the others; (3) particularly well determined are the mutational factors of the binding sites,  $fd_i$ ; and (4) the  $x_i$  are apparently best determined with the CHA<sup>5</sup> model and worst with CHA<sup>2</sup> model.

To further characterize the global fits, we determined for the successful<sub>p</sub> fits the uncertainties of the parameters by two measures, the inaccuracy, the deviation of the fitted parameters from the respective true values, and the imprecision, the deviation of the fitted parameter from the determined mean (see Materials and methods). We first consider the inaccuracy that is only accessible in simulated data as used herein. The inaccuracy for the CHA<sup>5</sup> model with  $n_p = 41$  parameters is approximately similarly good as that for the CHA<sup>4</sup> model and CHA<sup>3</sup> model with  $n_p = 23$  and  $n_p = 13$  parameters, respectively, and all of these fits seem to be superior compared with the fit with the CHA<sup>2</sup> model with  $n_p = 7$  parameters (Fig. S3 A). We also compared the imprecision (Fig. S3 B), a measure that would be accessible also for real experimental data. We considered if the somewhat surprising higher goodness of the fits for the channels with the higher parameter numbers can also be obtained from this measure of uncertainty. The diagrams for the four channels confirm the above result, i.e., the channels with more binding

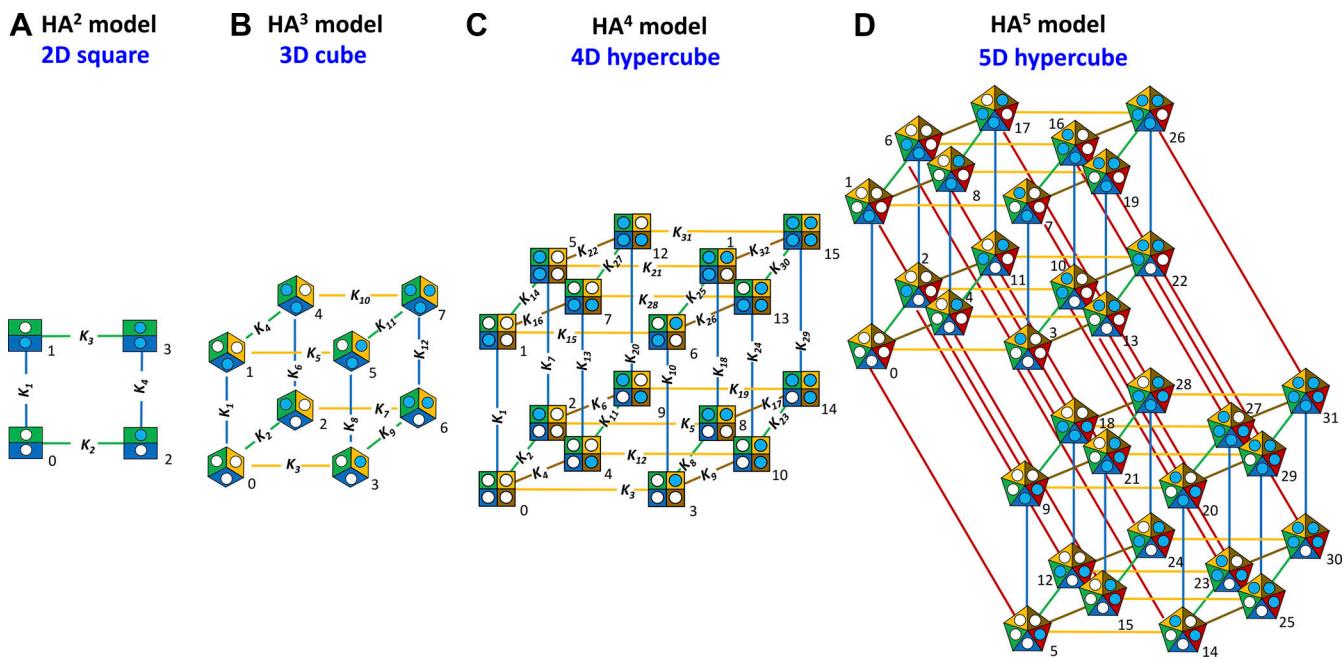


Figure 2. **Closed states of HA<sup>x</sup> models in cube representation.** The symbols correspond to those in Fig. 1. Each model in systematic representation in Fig. 1 was translated into the cube representation. The model structure follows a respective change of the dimension  $n_D$ . (A) HA<sup>2</sup> model containing four states arranged in a 2-D square. (B) HA<sup>3</sup> model containing eight states arranged in a 3-D cube. (C) HA<sup>4</sup> model containing 16 states arranged in a 4-D hypercube. (D) HA<sup>5</sup> model containing 32 states arranged in a 5-D hypercube. The  $K_x$  are indicated apart from the HA<sup>5</sup> model where they were omitted for clarity.

sites provide an apparently lower imprecision than the channels with less binding sites.

We next addressed the question of whether the fact that some of the few parameters are apparently less well-determined than the majority of the others can be attributed to special correlations among the parameters. To this end, we computed the matrix of the correlation coefficients from the covariance matrix provided by the fits of all CHA models (see Materials and methods) and plotted them in false colors (Benndorf et al., 2022) to identify conspicuous specific correlations among parameters (Fig. S4). The result was that the parameters with the highest imprecision, the last  $E_x$  and  $Z_x$ , show a pronounced negative correlation. But there are also a few other relevant negative correlations between the second to last  $E_x$  and  $Z_x$  as well as some positive correlations for both the CHA<sup>4</sup> and the CHA<sup>5</sup> model between high  $Z_x$ . Nevertheless, the most relevant result is that non-correlated parameters (colored yellow) dominate for both the CHA<sup>4</sup> and the CHA<sup>5</sup> model.

Finally, we computed for the successful<sub>p</sub> fits respective overall-measures for all parameters of a fit, of which each is already the mean of  $n$  fits as specified above. For the inaccuracy and the imprecision, we calculated the overall means,  $m_{iac_{np,n}}$  and  $m_{ipr_{np,n}}$  (Eqs. 15 and 12, respectively; Fig. 6, B and C). In the case of the correlation coefficients, we abolished the influence of different signs by taking the absolute values and taking out the ones of the main diagonal, yielding the mean absolute correlation coefficients<sub>n</sub> (Fig. 6 D).

These diagrams demonstrate that the chances to determine the parameters in the respective models are best in the models with the higher number of binding sites. To come back to our

initial question, the increasing number of constraints by the increased number of CARs in channels with more binding sites overbalances the effect introduced by the increased number of parameters. The most important result, however, is that it seems to be indeed realistic to analyze a complex system with 41 parameters, as given by the CHA<sup>5</sup> model, if only all CAR combinations with systematically impeded binding sites are available and if selecting the successful<sub>p</sub> fits for the analysis.

#### Relation between mutational effects and parameter identifiability

We next quantified how strong the mutational factors should be to obtain reasonable constraints in the global fit by using as a criterion the number of successful<sub>p</sub> fits and the imprecision. We considered the parameters  $E_x$  and  $Z_x$  (Fig. S2) and reduced systematically the mutational effects in 10 steps (weakening factor  $w = 10 \dots 0$ ) according to

$${}^w f d_y = 10^{\left\{ \frac{-w(2n_D - 1 - y)}{5(n_D - 1)} \right\}} \quad (22)$$

where  $w = 10$  specifies  ${}^{10} f d_y = f d_y$ , i.e., the maximum mutational effect, and  $w = 0$  specifies  ${}^0 f d_y = 1$ , i.e., no mutational effect or wild type channels.  $y$  runs from 1 to  $n_D$ . We then performed 250 fits at each degree of  $f d_y$  reduction (Fig. 8 A).

We first counted the number of successful<sub>p</sub> fits (Fig. 8 B). As expected, for all CHA<sup>x</sup> models, a weaker mutational effect causes a lower number of successful<sub>p</sub> fits. Moreover, at stronger mutational effects ( $w \rightarrow 10$ ), the CHA<sup>4</sup> and CHA<sup>5</sup> models produce better success than the CHA<sup>2</sup> model. The CHA<sup>3</sup> model is intermediate.

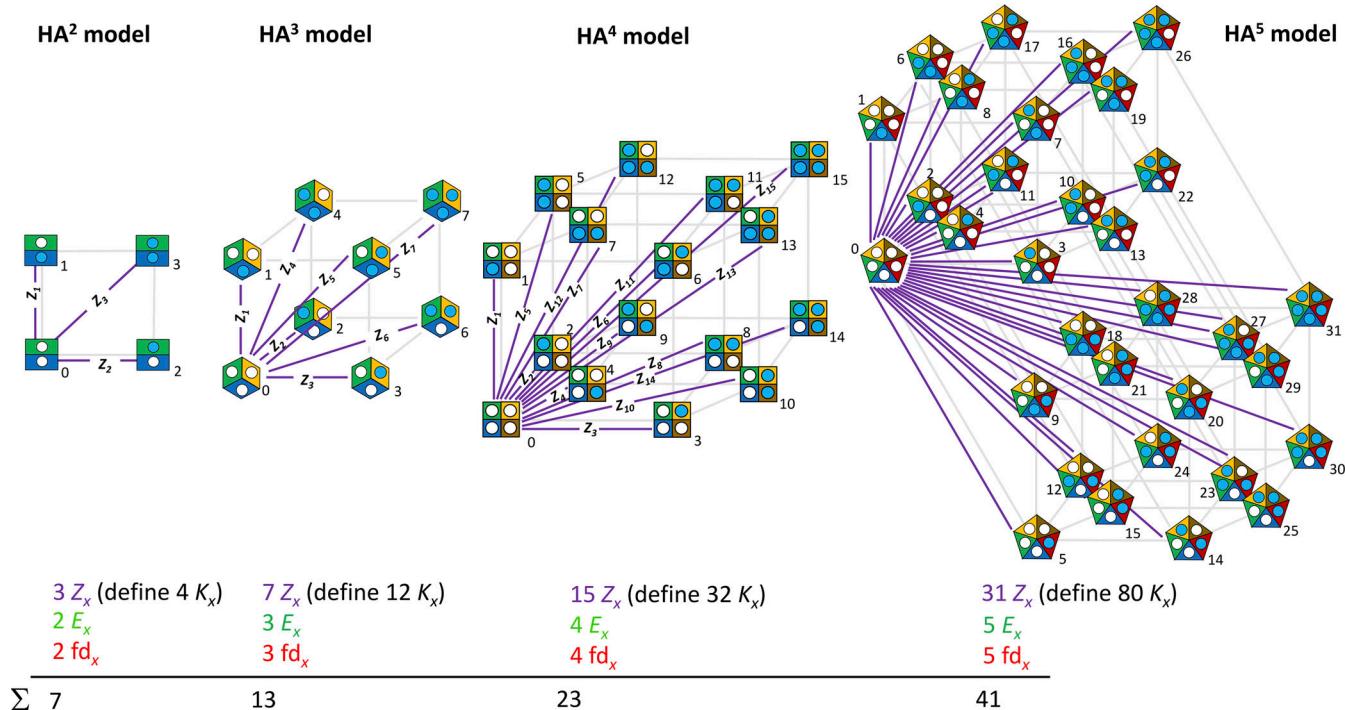


Figure 3. **Virtual equilibrium association constants for  $\text{HA}^x$  models in cube representation.** The virtual equilibrium association constants,  $Z_x$ , are indicated by the violet lines running from state 0 to the other states. The number of  $Z_x$  in a model defines directly the number of independent parameters for ligand binding. **(A)**  $\text{HA}^2$  model. **(B)**  $\text{HA}^3$  model. **(C)**  $\text{HA}^4$  model. **(D)**  $\text{HA}^5$  model. The  $Z_x$  are indicated apart from the  $\text{HA}^5$  model where they were omitted for clarity. The dimensions of the  $Z_x$  are  $M^{-1}$ ,  $M^{-2}$ ,  $M^{-3}$ ,  $M^{-4}$ , or  $M^{-5}$  for the occupation of 1, 2, 3, 4, and 5 binding sites, respectively.

When calculating the corresponding means of the imprecision (Eq. 12) over all parameters, it is also lower for the  $\text{CHA}^4$  and  $\text{CHA}^5$  model than the  $\text{CHA}^3$  model and even much lower than for the  $\text{CHA}^2$  model (Fig. 8 C). Moreover, the results show that with the selected set of parameters, for  $w$  a value of 7 or larger is a good idea, resulting for  ${}^7\text{fd}_i$  and  ${}^7\text{fd}_{\text{max}}$  to a 631- and 25-times reduced affinity, respectively.

We next considered how the different degrees of mutational effects affect the  $E_x$  and  $Z_x$  individually. Regarding  $E_x$ , it is evident that there is a pronounced reduction of the imprecision from the  $\text{CHA}^2$  to the  $\text{CHA}^4$  or  $\text{CHA}^5$  model, telling that the lower imprecision (higher precision) of fits with the more complex models, and respectively more parameters, are indeed better determined (Fig. S5). It is also notable that  $E_i$  and  $E_{\text{max}}$  are some more uncertain than the respective intermediate  $E_x$ . For  $Z_x$ , the results are basically similar. Again, there is a pronounced reduction of the imprecision from the  $\text{CHA}^2$  to the  $\text{CHA}^4$  or  $\text{CHA}^5$  model. Also,  $Z_i$  and  $Z_{\text{max}}$  are more uncertain than the respective intermediate  $Z_x$  (Fig. S6). Together, these results show that the reduced overall imprecision (Fig. 6 C) of fits with the more complex models, and respectively more parameters, is caused by a lower imprecision of both  $E_x$  and  $Z_x$ .

#### Searching the global minimum without the bias of a specific start vector

Our approach so far allowed us to substantiate that the parameter identifiability for the  $\text{CHA}^5$  model was better than

that for the  $\text{CHA}^2$  model and at least as good as that for the  $\text{CHA}^3$  and  $\text{CHA}^5$  model. Besides the fact that this situation does not match the situation of an experimenter, who, of course, does not know any true values, our approach could have been

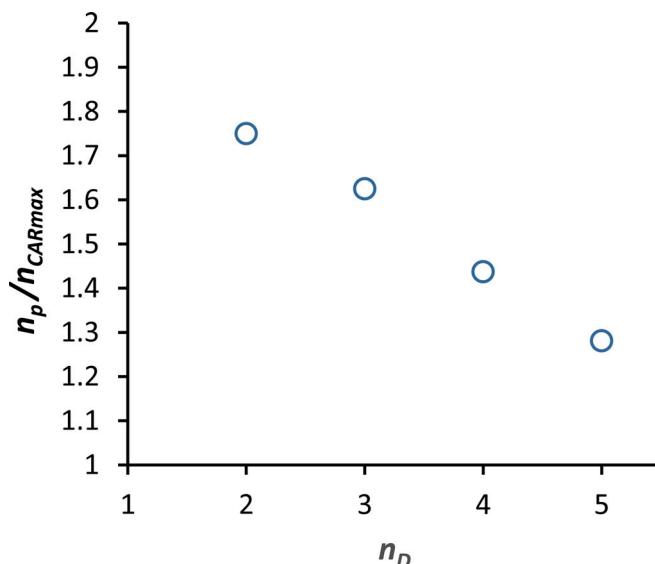
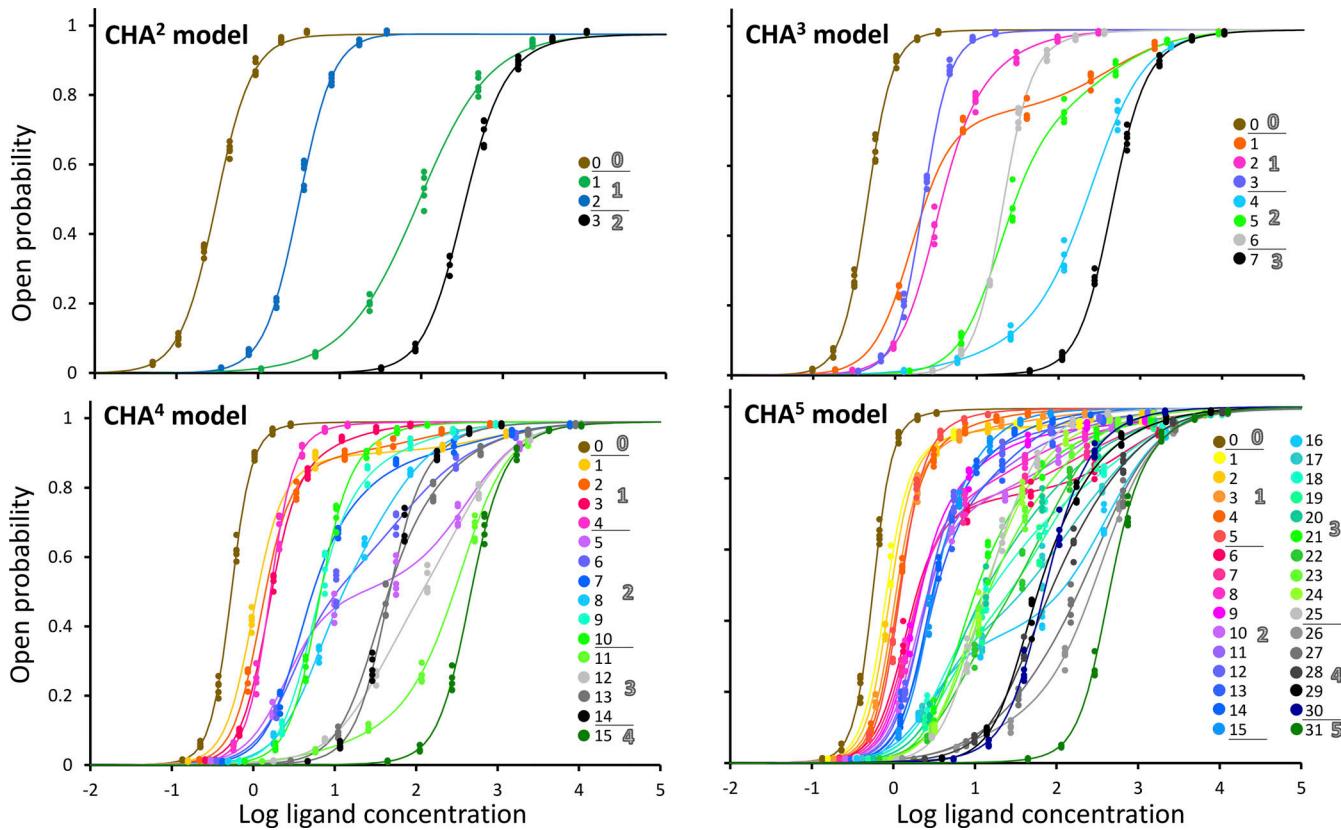


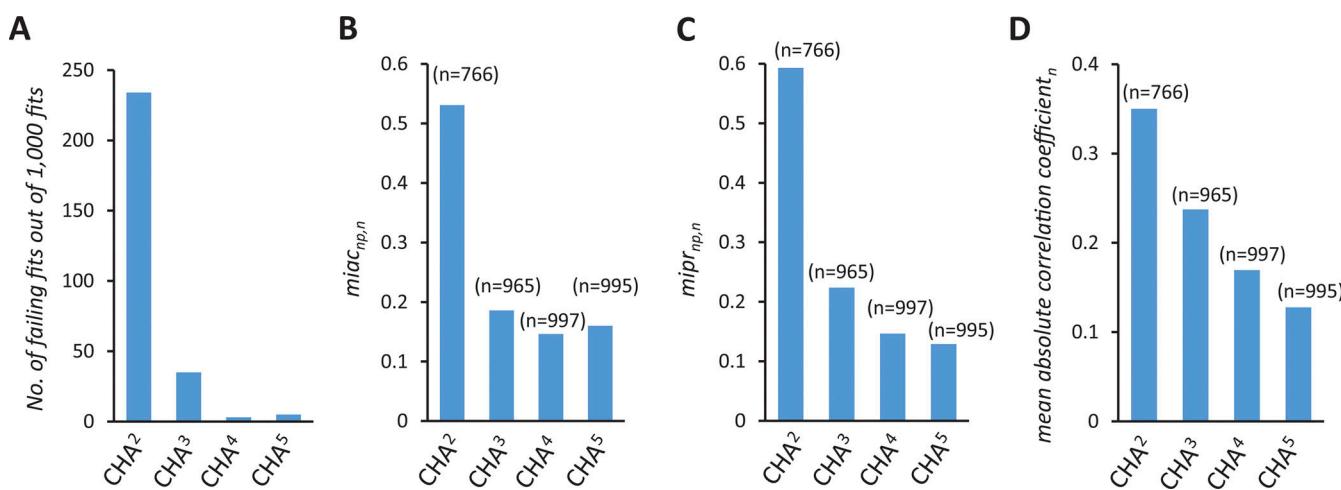
Figure 4. **Relation between number of parameters and available CARs in  $\text{CHA}^x$  models.** The ratio of the number of parameters,  $n_p$ , to the maximum number of available CARs,  $n_{\text{CARmax}}$ , is plotted versus the number of dimensions,  $n_D$ , i.e., the number of binding sites.



**Figure 5. Simulation of CARs with  $\text{CHA}^x$  models.** Using the equilibrium constants illustrated in Fig. S2, we computed the full set of CARs (Eq. 21) with the  $\text{CHA}^2$  to the  $\text{CHA}^5$  model. Noise 0.25 (see Materials and methods). For the specified parameters, the wild type and maximally mutated channels span a similar concentration range. The CARs with intermediate degrees of mutations fill the range between the two extreme CARs and develop characteristic shapes and, thus, constraints for the fit.

strongly biased by using the true values as a start vector, certainly promoting convergence to the minimum with the true parameters.

To match the requirements of an experimenter, the logic of the approach is now turned around: a set of data points for the respective maximum number of CARs is viewed as given and an



**Figure 6. Success rate and goodness of fits with the  $\text{CHA}^x$  models.** 1,000 fits with de novo generated data points each were performed with each model. The start vector consisted of the true values. The value  $n$  indicates the number of successful<sub>p</sub> fits out of 1,000 fits. **(A)**  $n$  of successful<sub>p</sub> fits out of 1,000 fits. **(B)** Mean inaccuracy,  $\text{miac}_{np,n}$ , computed over all  $n_p$  parameters according to Eq. 15. **(C)** Mean imprecision,  $\text{mipr}_{np,n}$ , computed over all  $n_p$  parameters according to Eq. 12. **(D)** Mean absolute correlation coefficients (see Materials and methods). The influence of the different signs was abolished by taking their absolute values. Also, the "1" of the main diagonal was taken out.  $n$  specifies the number of included successful fits.

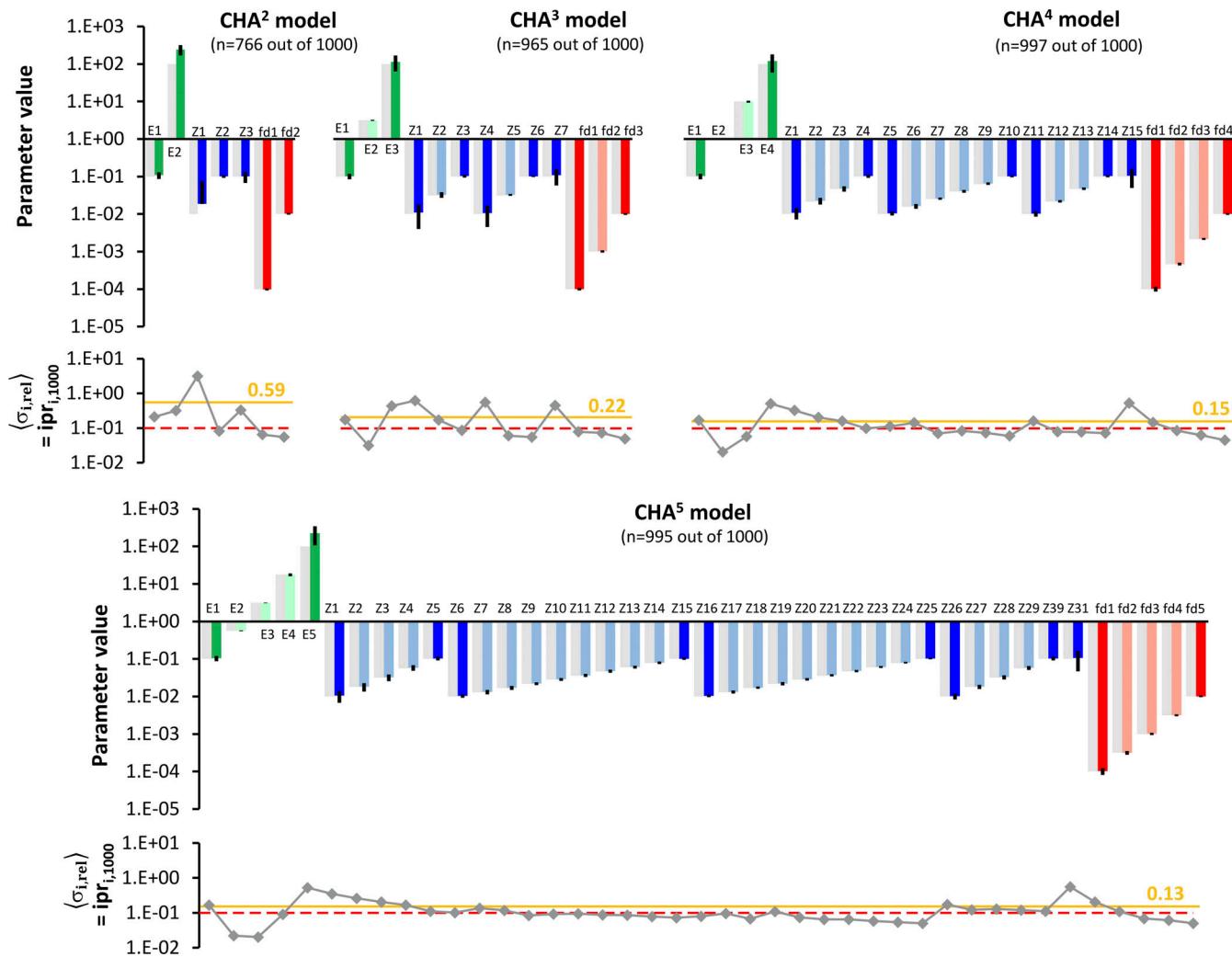


Figure 7. **Comparison of fit parameters with true parameters as a start vector.** In the bar graphs, the fitted parameters,  $x_i$ , are compared with the true parameters,  $x_{i,tr}$ . For each model, 1,000 fits were performed. As a starting vector, the true parameters were used (gray bars). The data points were de novo generated for each fit. Despite taking the true values as the start vector, part of the fits failed. The number of successful fits is given by  $n$ . Means and standard deviations of these fits are shown. The mean relative standard deviations  $\sigma_{i,rel}$  of 1,000 fits,  $ipr_{i,1000}$  (Eq. 11), below the bar graphs indicate the relative errors of each parameter. The ochre lines and numbers right specify the overall  $ipr_{i,1000}$  value for each model and the dashed red line the level of 0.1, i.e., the 10% error.

approach is developed to analyze the respective data of the CHA<sup>2</sup> to CHA<sup>5</sup> model.

To this end, we followed our previous strategy of using SU vectors, varying over a wide range, and vary each of its elements stochastically (Benndorf et al., 2022; Schirmeyer et al., 2022). If sufficiently many successful fits with parameters providing the same minimum can be obtained and the nonsuccessful fits can be reliably eliminated, any bias introduced by a specific start vector can be ruled out.

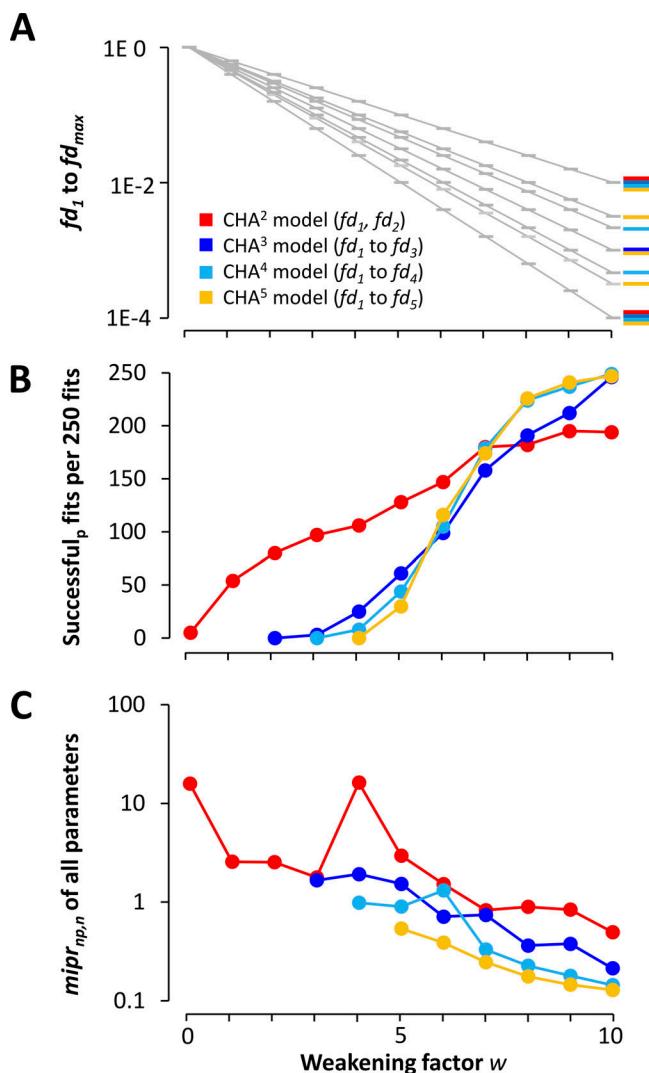
We varied the SU vector from  $10^{-6}$  to  $10^3$ , i.e., over nine orders of magnitude, resulting in 19 SU vectors (Fig. 9 A). As a stochastic factor, we used  $A = 100$ , i.e., each element of the SU vector was stochastically varied in the range of  $10^{-2}$  to  $10^2$  of its value. In total, this covers the huge range of 13 orders of magnitude for the stochastically varied SU start vector. The noise factor was again 0.25 and 1,000 fits were performed for  $n_{CAR,max}$  CARs.

We generated one set of stochastically varied data points, as described, and applied the fit routine for the 19 SU vectors à 1,000

fits. We then identified the successful<sub>p</sub> fits. Compared with the above approach using the true parameters as start vector, their portion was expectedly very low because the values of the stochastically varied SU start vector must be extremely unfortunate.

Then a second selection algorithm for successful fits out of the successful<sub>p</sub> fits was applied.

Using the  $MSE_x$  values provided by the fit routine, we identified the minimum  $MSE_x$  value,  $\text{Min}(MSE_x)$  in the 1,000 fits, calculated the difference values  $\Delta MSE_x$  of each  $MSE_x$  to  $\text{Min}(MSE_x)$  (Eq. 17), and sorted these values starting with the smallest (see Materials and methods). These sorted  $\Delta MSE_x$  values were plotted for  $x > 0$  on a logarithmic scale (Fig. 9 B). The number of  $\Delta MSE_x$  values was very low ( $<<10^{-10}$ ). At a certain  $x$ ,  $\Delta MSE_x$  jumped to much higher values  $>>10^{-10}$ , indicative of another but much worse fit minimum. These much larger  $\Delta MSE_x$  values were in no case similar to each other as were the  $\Delta MSE_x$  values near  $\Delta MSE_1 = 0$ , suggesting that no other reproducible minimum was present. We then set a threshold at



**Figure 8. Effects of mutational factors on parameter identifiability.** (A) Scheme illustrating the weakening steps for all  $fd_y$  where  $y$  runs from 1 to  $n_D$ . Using  $E_x$  and  $Z_x$  illustrated in Fig. S2, the mutational factors  $fd_y$  were systematically reduced in 10 steps (weakening factor  $w = 10 \dots 0$ ), providing weakening mutational factors  $w^{fd_y}$  according to Eq. 21.  $w = 10$  specifies  $^{10}fd_y$  the maximum mutational effect, as shown in Fig. S2, whereas  $^0fd_y$  specifies wild type channels with no weakening effect. (B) Plot of the number of successful<sub>p</sub> fits as function of  $w$ . After performing 250 fits each, the number of successful<sub>p</sub> fits was counted by eliminating all fits with physically senseless negative parameters. A weaker mutational effect decreases the number of successful<sub>p</sub> fits. At sufficiently strong mutational effects, the CHA<sup>4</sup> and CHA<sup>5</sup> model are superior over the CHA<sup>2</sup> model. (C) Plot of the mean imprecision  $mipr_{np,n}$  as function of  $w$ . The mean imprecision (Eq. 12) over all parameters is lower for the CHA<sup>4</sup> and CHA<sup>5</sup> model than for the CHA<sup>3</sup> model and, particularly, for the CHA<sup>2</sup> model.

$\Delta MSE_x = 10^{-10}$  to identify all fits in close proximity of  $\Delta MSE_1 = 0$ . Furthermore, we specified that for a given SU vector at least five fits with  $\Delta MSE_x$  values below the threshold of  $10^{-10}$  had to be present among 1,000 fits to accept a true minimum. With these assumptions, we could identify all fits close to the best fit at  $\Delta MSE_1 = 0$ . These fits are denoted successful<sub>p<sub>AMSE</sub></sub> fits in the following. In the plots, the fits with  $\Delta MSE_1 = 0$  were omitted. The diagrams show that out of the 1,000 fits, only a small

portion but sufficiently many successful<sub>p<sub>AMSE</sub></sub> fits were obtained for all four models. The number of successful<sub>p<sub>AMSE</sub></sub> fits was either equal or some smaller than that of successful<sub>p</sub> fits.

Plotting histograms for the number of successful<sub>p<sub>AMSE</sub></sub> fits as a function of the SU vector shows that it was maximal near the SU vector of  $10^{-1}$  and drops to both larger and smaller SU start vectors to zero (Fig. 10). As a matter of principle, this means that it is possible to identify a range for the SU vectors that yields the majority of successful<sub>p<sub>AMSE</sub></sub> fits. Hence, our chosen SU start vectors cover a reasonable range in which a stochastic variation of the elements of the SU start vector by the factor  $A = 100$  identifies a small but reasonable portion of successful<sub>p<sub>AMSE</sub></sub> fits. Moreover, as already indicated in Fig. 9 B, the number of successful<sub>p<sub>AMSE</sub></sub> fits for all SU vectors decreases for the more complex models.

We next calculated for each model the mean and the standard deviation for all parameters as obtained from the successful<sub>p<sub>AMSE</sub></sub> fits with all SU start vectors. These means are again plotted in bar graphs for all parameters together with the true values (Fig. 11). Also the mean relative standard deviations  $\sigma_{i,rel}$  for each parameter of 1,000 fits ( $ipr_{i,1000}$ ; Eq. 11) are plotted below each bar graph as in Fig. 7. All parameters could be determined similarly well as with the analysis above although now there was no bias introduced by any specific start vectors. This result is further confirmed by plotting the inaccuracy and the imprecision (Fig. S7), which both closely match the profiles obtained from the successful<sub>p</sub> fits using the true parameters as the start vector.

In summary, our analysis with the stochastically varied SU start vectors used identical data points for the 19,000 fits with all 19 SU vectors, a strategy that could also be run with real experimental data.

As a control, we finally repeated our stochastic approach with different sets of data points for the different SU start vectors, while for the 1,000 fits at each SU start vector, again, identical stochastically varied data points were used. This approach has the advantage that 19 different data point sets were used. If the results are similar to those with identical data points for all SU start vectors, this would confirm that the successful result shown by Figs. 9, 10, and 11; Fig. S7 does not depend on a specific data set. The corresponding Figs. S8, S9, S10, and S11 show a closely similar result to that provided by Figs. 9, 10, 11, and S7, i.e., all parameters are determined, the maximum number of successful<sub>p<sub>AMSE</sub></sub> fits was near the SU start vector of  $10^{-1}$  (Figs. S8 and S9), the parameters match the true parameters (Fig. S10), and the inaccuracy as well as the imprecision (Fig. S11) match those of the corresponding successful<sub>p<sub>AMSE</sub></sub> fits shown in Fig. S7 and also those obtained from the successful<sub>p</sub> fits shown in Fig. S3. We conclude that our approach is not essentially influenced by a major bias of selecting a specific set of data points.

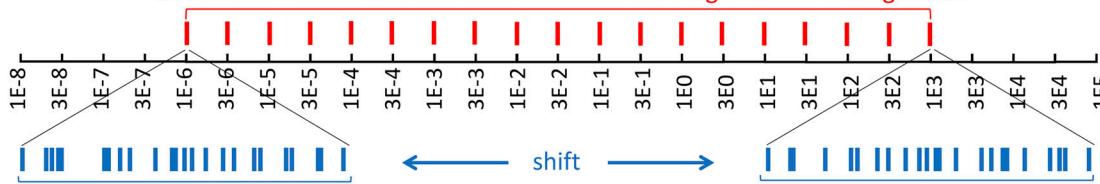
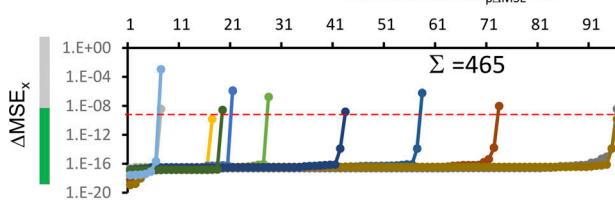
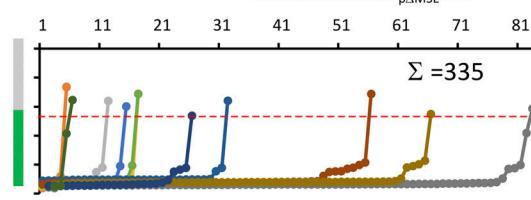
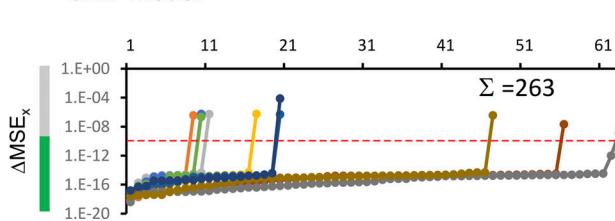
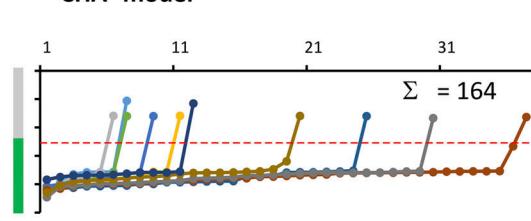
## Discussion

### Summary

In this study, we investigated the identifiability of 7–41 free parameters determining the CHA<sup>2</sup> to CHA<sup>5</sup> schemes for LGICs containing two to five binding sites, respectively. These parameters define respective larger numbers of equilibrium association constants as specified in Tables S1, S2, S3, and S4. For

**A**

19 SU start vectors with identical elements covering 9 orders of magnitude

**B****CHA<sup>2</sup> model**No. of successful  $p_{\text{AMSE}}$  fits**CHA<sup>3</sup> model**No. of successful  $p_{\text{AMSE}}$  fits**CHA<sup>4</sup> model****CHA<sup>5</sup> model**

**Figure 9. Successful $p_{\text{AMSE}}$  fits with equal SU start vector and stochastically varied elements. (A)** Scheme of the stochastic variation of the start vector. SU start vectors extended over the range from  $10^{-6}$  to  $10^3$  yielding 19 SU vectors. The SU vectors were stochastically varied by the factor  $A = 10^2$ . 1,000 fits were performed for each SU start vector. The noise factor was 0.25 and 1,000 fits were performed for each SU vector. (B) After selection of successful $p_{\text{AMSE}}$  fits, a second selection algorithm for fits was applied. Using the  $\text{MSE}_x$  values provided by the fit routine, the minimum  $\text{MSE}_x$  value,  $\text{Min}(\text{MSE}_x)$ , was identified in the 1,000 fits and the difference of each  $\text{MSE}_x$  to the  $\text{Min}(\text{MSE}_x)$  was calculated, yielding  $\Delta \text{MSE}_x$ .  $\Delta \text{MSE}_x$  values were sorted, starting with the smallest, and plotted for  $x > 0$  on a logarithmic scale. A threshold was set at  $10^{-10}$  (red dashed lines) to identify all fits in close proximity of  $\Delta \text{MSE}_1 = 0$ . All values below  $10^{-10}$  indicate successful $p_{\text{AMSE}}$  fits. For all four models, the sums of all obtained successful $p_{\text{AMSE}}$  fits out of 19,000 fits are indicated.

our comparison of the performance of the CHA<sup>2</sup> to CHA<sup>5</sup> models, we included the respective full sets of all possible CARs whose number is given by Eq. 21. The computations showed that the parameters of the most complex CHA<sup>5</sup> model with 41 parameters are markedly better determined than those of the much simpler CHA<sup>2</sup> model with only 7 parameters. The CHA<sup>3</sup> and CHA<sup>4</sup> models were in-between. The likely explanation for this apparently counterintuitive result is that the fit constraints rise steeper with the higher complexity than the degrees of freedom introduced by the growing number of parameters.

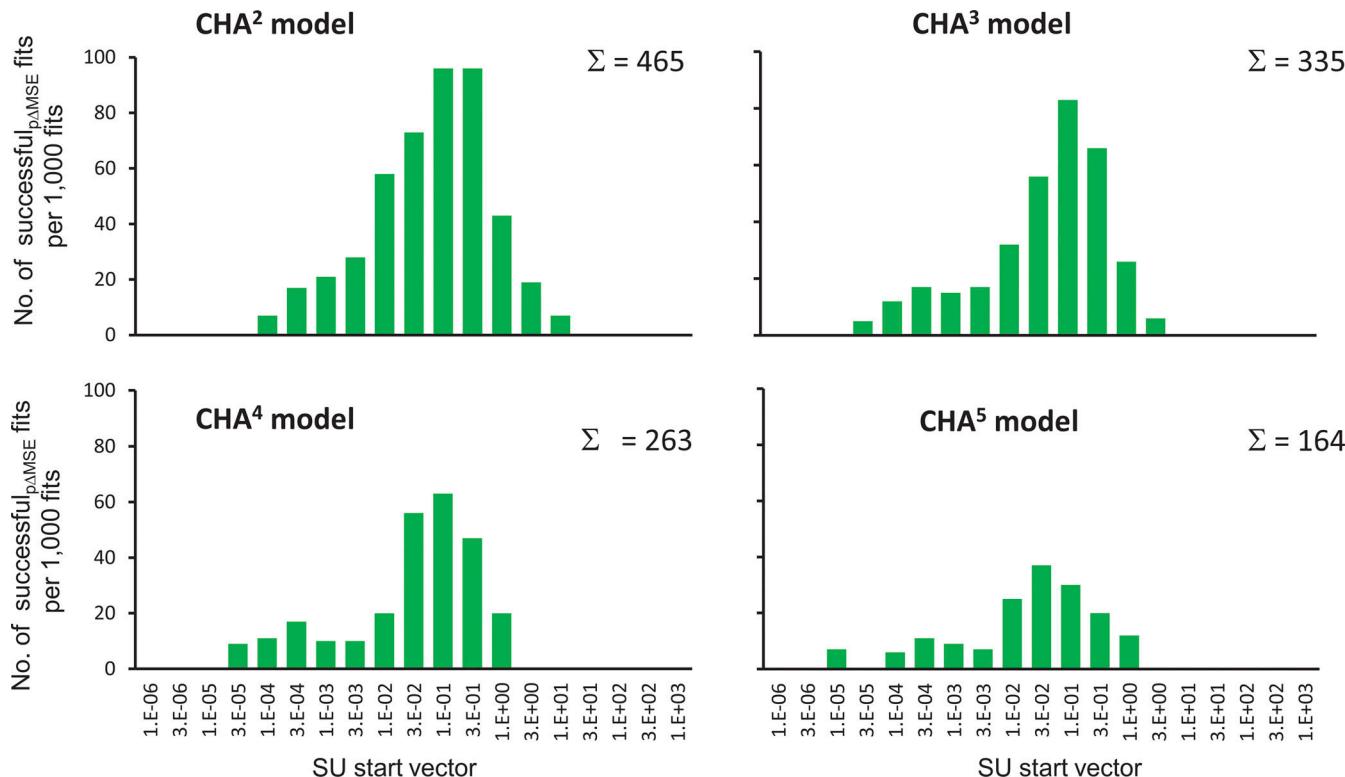
We like to emphasize again that all computations performed herein were based on the synthetic parameter sets generated by Eq. 23, 24, 25, 27, 29, and 32 in the supplemental text at the end of the PDF. Though the parameter set for LGIC<sup>4</sup>s was related to some extent to the experimental data in heterotetrameric CNG channels (Schirmeyer et al., 2021), it is synthetic, as are the parameter sets for LGIC<sup>2</sup>s, LGIC<sup>4</sup>s, and LGIC<sup>5</sup>s.

#### Assumptions for the strategy

Our strategy is based on three assumptions: (1) a known subunit composition of the channel, (2) a marked reduction of the binding affinity at the binding sites by appropriate mutations, and (3) a global fit of CARs obtained from recording macroscopic currents with coupled models.

To apply our strategy to experimental data of concatameric channels, three assumptions have to be fulfilled: (1) the architecture and function of the concatameric channels must be similar to channels built by non-concatenated subunits; (2) the mutations at the binding sites must markedly reduce but not entirely abolish the binding affinity; and (3) the CARs used for the global fits with coupled models must be of high quality.

Assumption 1 requires that the expected architecture and an approximately unchanged function of the concatenated channels can be demonstrated. Possible side effects of linking the subunits can be easily identified by comparing the CAR of the wt concatamer with the CAR of channels built by individual wt subunits. Regarding the architecture, an incorrect assembly of the mutated subunits in the concatamers has to be excluded which might appear if the subunits of two concatamers build only one functional channel while part of the subunits are “looped out” (Steinbach and Akk, 2011). This phenomenon has been originally reported for voltage-gated K<sup>+</sup> channels (McCormack et al., 1992) and later also for pentameric ligand-gated nicotinic receptors (Zhou et al., 2003; Groot-Kormelink et al., 2004). A relatively simple functional test would be to take a concatamer containing only wild type subunits and a concatamer containing only mutated subunits, shifting the CAR to much higher concentrations while the slope is unaffected. If



**Figure 10. Histograms of successful<sub>pΔMSE</sub> fits with equal SU start vector and stochastically varied elements.** For each of the 19 SU vectors (abscissa), at least five successful<sub>pΔMSE</sub> fits were required to include them in the histogram. The fits with  $\Delta MSE_1 = 0$  were omitted. Reasonable numbers of successful<sub>pΔMSE</sub> fits were obtained for all CHA<sup>X</sup> models. The number of successful<sub>pΔMSE</sub> fits was maximal near the SU vector of  $10^{-1}$  and drops to both larger and smaller SU vectors to zero. The number of successful<sub>pΔMSE</sub> fits for all SU vectors decreases for the more complex models.

coexpressing these concatamers in a 1:1 ratio, channels formed by the two concatamers only would lead to two clearly separated CAR components with unchanged steepness whereas a significant combination of the subunits of the two concatamers within a channel, while looping out others, should lead to a smeared CAR over the whole concentration range. We demonstrated the preservation of the two components for wt and mutated concatamers of homotetrameric CNGA2 channels (Wongsamitkul et al., 2016). Moreover, the consistency of our previous results on concatenated heterotetrameric CNGA4:A2: B1b:A2 channels (Schirmeyer et al., 2021) and homotetrameric HCN2 channels (Sunkara et al., 2018) suggested both a correct architecture and only minor side effects by concatenation. We therefore assume that is generally possible to follow such a strategy. It should also be noted that enforcement of subunits into a desired stoichiometry and position is only one way to define the composition of channels. In case that all subunits differ in their structure and are consistently incorporated in a channel, concatenation is not even required.

Assumption 2 of our strategy includes that the mutations strongly impede the binding sites but preserve functionality, i.e., all mutated binding sites keep their ability to bind a ligand and contribute to activation. We realized this in our simulations by following the previous strategy for tetrameric channels using impeding parameters  $f_{d_x}$  for each of the subunits, i.e., the  $f_{d_x}$  values were assumed to be unique for each binding site, independent of the activation of the other subunits. Hence, the  $f_{d_x}$

factors are helper parameters that are not relevant for interpreting the constants of the wild type channel.

For assumption 3, the global fit with coupled models, we assumed for each binding step individual equilibrium association constants,  $K_x$ , but lumped the equilibrium closed-open isomerization constants,  $E_x$ , at an equal degree or liganding (Schirmeyer et al., 2021; Benndorf et al., 2022). This significantly reduced the number of parameters. Future simulation studies will show to what extent it is possible to differentiate between the  $E_x$ . One possible analysis strategy we already tested is by including data at a second voltage and assuming that each subunit exerts an invariant subunit promotion energy on channel opening (Schirmeyer et al., 2022). Simple additivity of the subunit promotion energies led to individual  $E_x$  values for each constellation of liganding. In all fits, we obeyed the principle of microscopic reversibility, which is widely accepted to be physically meaningful though for some ion channels a violation of this principle has been suggested (Schneggenburger and Ascher, 1997; Xu and Meissner, 1998). It should also be noted that we only simulated and analyzed equilibrium constants here which do not tell anything about the dynamic response of the receptors.

#### Applicability of the approach to LGICs

We demonstrated that our fit approach can be used for LGICs with two to five binding sites (LGIC<sup>2</sup> to LGIC<sup>5</sup>) and that an accurate read-out of channel activation is possible.

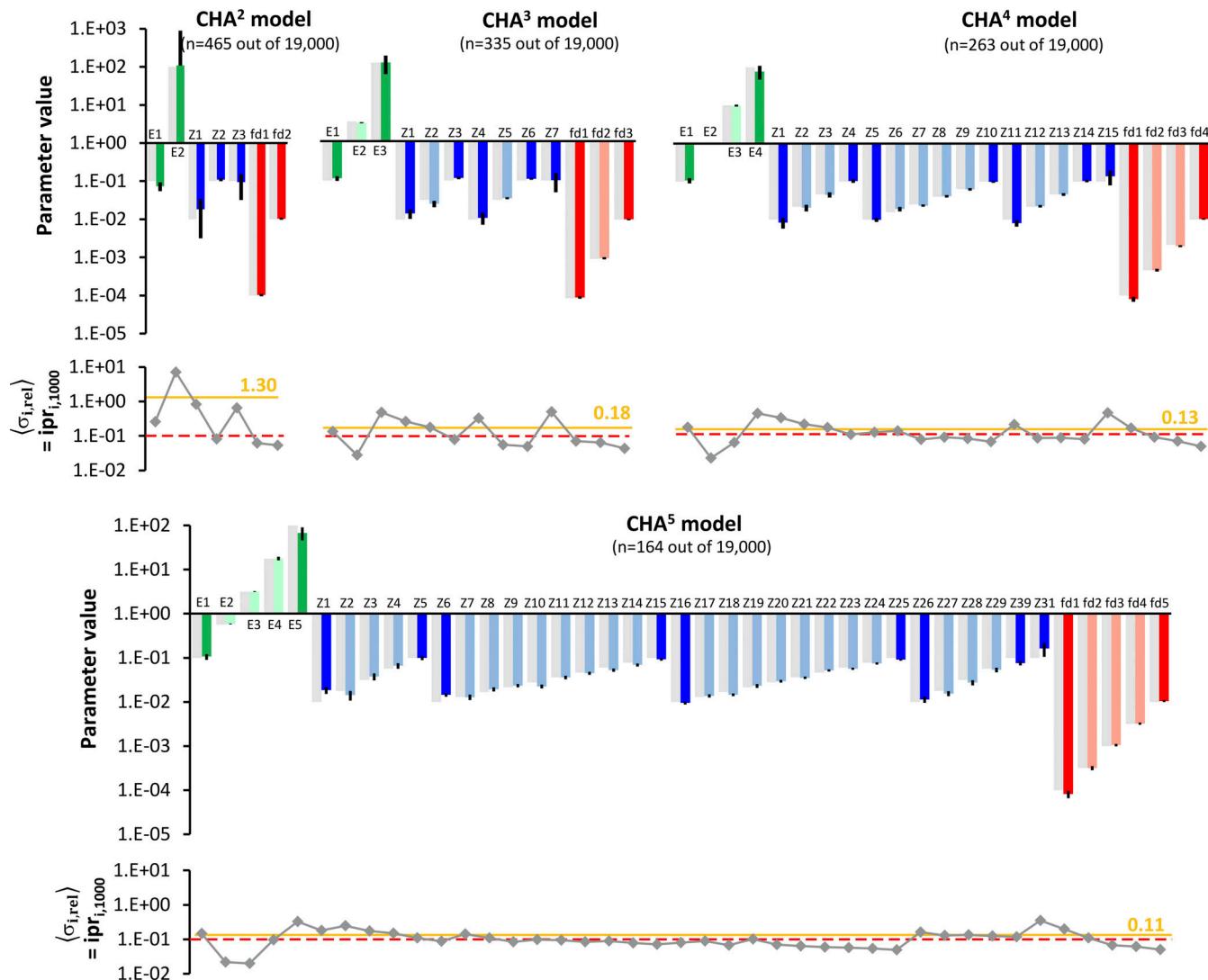


Figure 11. Comparison of fit parameters with equal SU start vector and stochastically varied elements. The same formalism as in Fig. 7. In the bar graphs, the fitted parameters,  $x_i$ , are compared with the true parameters,  $x_{i,tr}$  (gray). For each SU vector, 1,000 fits were performed. The results for the successful  $_{\Delta \text{MSE}}$  fits are shown. All parameters are determined reasonably though no bias was introduced by specific start vectors. The mean relative standard deviations  $\sigma_{i,\text{rel}}$  of 1,000 fits,  $\text{ipr}_{i,1000}$  (Eq. 11), below the bar graph indicates the relative errors of each parameter. The ochre lines and numbers right specify the overall  $\text{ipr}_{i,1000}$  value for each model and the dashed red line the level of 0.1, i.e., the 10% error.

Among LGIC<sup>3</sup>s, possible candidate channels for our approach are trimeric purinergic P2X receptors (Kawate et al., 2009; Hattori and Gouaux, 2012) and related ASICs (Jasti et al., 2007) that form between two adjacent subunits three binding sites. For P2X receptors, there are several reports that concatamers are functional (Nickle et al., 2003; Nagaya et al., 2005; Stelmashenko et al., 2012; Gusic et al., 2021) which would allow us to define the position of the mutated binding sites. For ASICs, we are not aware of respective concatamers.

Typical LGIC<sup>4</sup>s with four binding sites are AMPA receptors (Sobolevsky et al., 2009). So far there are no reports about functional concatamers in these receptors, possibly because of the problem that the N-terminus is outside. However, for related NMDA receptors, it has been demonstrated that functional concatamers can be constructed (Schorge and Colquhoun, 2003), which might open the door to applying our strategy. In contrast,

tetrameric CNG channels (Li et al., 2017) and hyperpolarization-activated cyclic nucleotide-gated channels (Lee and MacKinnon, 2017) form willingly functional concatamers that can be systematically analyzed (Ulens and Siegelbaum, 2003; Wongsamitkul et al., 2016; Sunkara et al., 2018; Schirmeyer et al., 2021).

Pentameric ionotropic receptors belong mainly to the class of cys-loop receptors. They include the families of nAChRs, GlyRs, and 5HTRs. Many of these receptors are formed by two  $\alpha$ -subunits and three other homolog subunits. For example, the classical nAChR of both the electric eel and the electric fish (Changeux et al., 1970; Miledi and Potter, 1971) is composed of two  $\alpha$ -subunits, one  $\beta$ -subunit, one  $\gamma$ -subunit, and one  $\delta$ -subunit (Weill et al., 1974; daCosta and Baenziger, 2013), forming only two binding sites between the adjacent subunits  $\alpha$  and  $\gamma$  as well as  $\alpha$  and  $\delta$ . Therefore, they would be classified herein as LGIC<sup>2</sup>s. In these channels, the specification of the binding sites need not

necessarily be forced by concatenating any subunits because the subunits form on their own channels with defined composition with defined binding sites. Pentameric receptors with five binding sites also exist, e.g., the  $\alpha 7$  nAChR (McKay et al., 2007) or the  $\alpha 1$  GlyR (Zhu and Gouaux, 2021). In our classification, they would be LGIC<sup>5</sup>s to be described by the CHA<sup>5</sup> model. It seems to be promising to investigate these or other pentameric receptors with the strategy described herein because the subunits can be linked to form functional concatameric channels (Mazzaferro et al., 2017; Prevost et al., 2021). In summary, there are many relevant candidate channels for our strategy to be applied.

### Applicability of the approach to other receptors

Our approach should also be applicable to other membrane receptors. This includes voltage-gated ion channels (VGICs), for which functional concatamers have been reported e.g., from Shaker subunits (Tytgat and Hess, 1992; Al-Sabi et al., 2010) and Kv2.1 and Kv6.4 subunits (Möller 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. Though measurement of currents in VGICs is easier than that in LGICs, quantitative analysis as described herein would be more complex because each voltage-dependent equilibrium constant requires two parameters, an equilibrium constant at zero voltage and an apparent gating charge. Moreover, our strategy should be easily usable to analyze natural heterotetramers, such as Nav channels (Jiang et al., 2020) and Cav channels (Wu et al., 2015). Furthermore, our analysis could be applied to oligomeric metabotropic receptors as well if a sufficiently valid read-out of receptor activation is available. Advancements in FRET techniques with dimeric glutamate receptors seem to be a promising strategy (Hlavackova et al., 2005; Grushevskyi et al., 2019; Kukaj et al., 2023).

Another intriguing idea is to analyze double mutations at two relevant positions in a channel or receptor, e.g., one mutation at the ligand binding site and a second in the pore gate. If both mutations exert different single effects, a systematic combination of these mutations should allow us to identify the interaction of the subunit with very high precision. However, the number of required combinations might become immediately very high. For example, introducing in a heterotetrameric channel a mutation in the binding site and another mutation in the pore gate would result in maximally  $16^2 = 256$  combinations. Nevertheless, it is very likely that such a strategy would allow us to specify very subtle interactions of subunits or binding sites in these oligomeric proteins. Notably, this idea is basically an extension of the idea of the double-mutant cycle analysis (Horovitz, 1996).

### The stochastic approach using SU start vectors

A further relevant result is that we successfully used a stochastic approach to identify sufficiently many successful<sub>pAMSE</sub> fits out of a vast number of fits by using SU vectors, scanning a range of nine orders of magnitude, and stochastically varying its elements over additional four orders of magnitude. Though the portion of successful<sub>pAMSE</sub> fits was only small compared with the

total number of fits, its results regarding accuracy and precision, quantified by inaccuracy and imprecision, respectively, were entirely reliable. Moreover, the scanning of the SU vector over the useful range of nine orders of magnitude allows an investigator to reliably identify the true parameters without prior knowledge of the parameters. For other applications, the range for the SU vectors could be easily extended. Apparently, there is a maximum incidence of the successful<sub>pAMSE</sub> fits in the range of the true parameters, which was in our case  $\sim 10^{-1}$  (c.f., Fig. S2). Our stochastic fit approach seems to be useful for any simple or complex fit problem because the investigator does not have to generate specific starting parameters, in particular if there are no reasonable guesses and the number of parameters is large.

### Conclusions

We conclude that for LGICs of known composition with two to five binding sites, our strategy to impede, but not abolish, the affinity of defined binding sites by mutations and to globally fit all respective CAR combinations of wild type and impeded binding sites by systems of respective kinetic schemes is a powerful approach to analyze subunit cooperativity of LGICs in great detail. To exclude any bias in the fit arising from specific start vectors, we show that scaled unitary start vectors stochastically varied over an extremely wide range provide reliable fit results. Remarkably, and to some extent counterintuitively, our results show that the identifiability of the parameters in complex models with more binding sites is superior to that in simple models with less binding sites.

### Data availability

The data are available from the corresponding author upon reasonable request.

### Acknowledgments

Joseph A. Mindell served as editor.

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

Author contributions: K. Benndorf designed the project, created the simulation and fit software, did the computations, designed the figures, and wrote the manuscript. E. Schulz developed the strategy for using hypercubes in the analysis.

Disclosures: The authors declare no competing interests exist.

Submitted: 26 May 2023

Revised: 22 August 2023

Revised: 27 September 2023

Accepted: 7 October 2023

### References

- Al-Sabi, A., O. Shamotienko, S.N. Dhochartaigh, N. Muniyappa, M. Le Berre, H. Shaban, J. Wang, J.T. Sack, and J.O. Dolly. 2010. Arrangement of Kv1 alpha subunits dictates sensitivity to tetraethylammonium. *J. Gen. Physiol.* 136:273–282. <https://doi.org/10.1085/jgp.200910398>
- 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>

Benndorf, K., T. Eick, C. Sattler, R. Schmauder, and E. Schulz. 2022. A strategy for determining the equilibrium constants for heteromeric ion channels in a complex model. *J. Gen. Physiol.* 154:e202113041. <https://doi.org/10.1085/jgp.202113041>

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>

Changeux, J.P., M. Kasai, and C.Y. Lee. 1970. Use of a snake venom toxin to characterize the cholinergic receptor protein. *Proc. Natl. Acad. Sci. USA.* 67:1241–1247. <https://doi.org/10.1073/pnas.67.3.1241>

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., K.A. Dowsland, M. Beato, and A.J.R. Plested. 2004. How to impose microscopic reversibility in complex reaction mechanisms. *Biophys. J.* 86:3510–3518. <https://doi.org/10.1529/biophysj.103.038679>

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. Springer US, Boston, MA. 397–482. [https://doi.org/10.1007/978-1-4419-1229-9\\_18](https://doi.org/10.1007/978-1-4419-1229-9_18)

daCosta, C.J., 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>

Fermi, G., M.F. Perutz, B. Shaanan, and R. Fourme. 1984. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.* 175: 159–174. [https://doi.org/10.1016/0022-2836\(84\)90472-8](https://doi.org/10.1016/0022-2836(84)90472-8)

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>

Groot-Kormelink, P.J., S.D. Broadbent, J.P. Boorman, and L.G. Sivilotti. 2004. Incomplete incorporation of tandem subunits in recombinant neuronal nicotinic receptors. *J. Gen. Physiol.* 123:697–708. <https://doi.org/10.1085/jgp.200409042>

Grushevskyi, E.O., T. Kukaj, R. Schmauder, A. Bock, U. Zabel, T. Schwabe, K. Benndorf, and M.J. Lohse. 2019. Stepwise activation of a class C GPCR begins with millisecond dimer rearrangement. *Proc. Natl. Acad. Sci. USA.* 116:10150–10155. <https://doi.org/10.1073/pnas.1900261116>

Gusic, M., K. Benndorf, and C. Sattler. 2021. Dissecting activation steps in P2X7 receptors. *Biochem. Biophys. Res. Commun.* 569:112–117. <https://doi.org/10.1016/j.bbrc.2021.06.101>

Hattori, M., and E. Gouaux. 2012. Molecular mechanism of ATP binding and ion channel activation in P2X receptors. *Nature.* 485:207–212. <https://doi.org/10.1038/nature1010>

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>

Hlavackova, V., C. Goudet, J. Kniazeff, A. Zikova, D. Maurel, C. Vol, J. Trojanova, L. Prézeau, J.P. Pin, and J. Blahos. 2005. Evidence for a single heptahelical domain being turned on upon activation of a dimeric GPCR. *EMBO J.* 24:499–509. <https://doi.org/10.1038/sj.emboj.7600557>

Horovitz, A. 1996. Double-mutant cycles: A powerful tool for analyzing protein structure and function. *Fold. Des.* 1:R121–R126. [https://doi.org/10.1016/S1359-0278\(96\)00056-9](https://doi.org/10.1016/S1359-0278(96)00056-9)

Jasti, J., H. Furukawa, E.B. Gonzales, and E. Gouaux. 2007. Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH. *Nature.* 449: 316–323. <https://doi.org/10.1038/nature06163>

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>

Kukaj, T., C. Sattler, T. Zimmer, R. Schmauder, and K. Benndorf. 2023. Kinetic fingerprinting of metabotropic glutamate receptors. *Commun. Biol.* 6:104. <https://doi.org/10.1038/s42003-023-04468-z>

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, C.H., and R. MacKinnon. 2017. Structures of the human HCN1 hyperpolarization-activated channel. *Cell.* 168:111–120.e11. <https://doi.org/10.1016/j.cell.2016.12.023>

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/10.1038/nature20819>

Liu, D.T., G.R. Tibbs, P. Paoletti, and S.A. Siegelbaum. 1998. Constraining ligand-binding site stoichiometry suggests that a cyclic nucleotide-gated channel is composed of two functional dimers. *Neuron.* 21: 235–248. [https://doi.org/10.1016/S0896-6273\(00\)80530-9](https://doi.org/10.1016/S0896-6273(00)80530-9)

Mazzaferro, S., I. Bermudez, and S.M. Sine. 2017.  $\alpha 4\beta 2$  nicotinic acetylcholine receptors: Relationships between subunit stoichiometry and function at the single channel level. *J. Biol. Chem.* 292:2729–2740. <https://doi.org/10.1074/jbc.M116.764183>

McCormack, K., L. Lin, L.E. Iverson, M.A. Tanouye, and F.J. Sigworth. 1992. Tandem linkage of Shaker K<sup>+</sup> channel subunits does not ensure the stoichiometry of expressed channels. *Biophys. J.* 63:1406–1411. [https://doi.org/10.1016/S0006-3495\(92\)81703-4](https://doi.org/10.1016/S0006-3495(92)81703-4)

McKay, B.E., A.N. Placzek, and J.A. Dani. 2007. Regulation of synaptic transmission and plasticity by neuronal nicotinic acetylcholine receptors. *Biochem. Pharmacol.* 74:1120–1133. <https://doi.org/10.1016/j.bcp.2007.07.001>

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>

Miledi, R., and L.T. Potter. 1971. Acetylcholine receptors in muscle fibres. *Nature.* 233:599–603. <https://doi.org/10.1038/233599a0>

Möller, 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)

Nagaya, N., R.K. Tittle, N. Saar, S.S. Della, and R.I. Hume. 2005. An inter-subunit zinc binding site in rat P2X2 receptors. *J. Biol. Chem.* 280: 25982–25993. <https://doi.org/10.1074/jbc.M504545200>

Nicke, A., J. Rettinger, and G. Schmalzing. 2003. Monomeric and dimeric byproducts are the principal functional elements of higher order P2X1 concatamers. *Mol. Pharmacol.* 63:243–252. <https://doi.org/10.1124/mol.63.1.243>

Niu, X., and K.L. Magleby. 2002. Stepwise contribution of each subunit to the cooperative activation of BK channels by Ca<sup>2+</sup>. *Proc. Natl. Acad. Sci. USA.* 99:11441–11446. <https://doi.org/10.1073/pnas.172254699>

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>

Prevost, M.S., H. Bouchenaki, N. Barilone, M. Gielen, and P.J. Corringer. 2021. Concatamers to re-investigate the role of  $\alpha 5$  in  $\alpha 4\beta 2$  nicotinic receptors. *Cell. Mol. Life Sci.* 78:1051–1064. <https://doi.org/10.1007/s00018-020-03558-z>

Schirmeyer, J., T. Eick, E. Schulz, S. Hummert, C. Sattler, R. Schmauder, and K. Benndorf. 2022. Subunit promotion energies for channel opening in heterotetrameric olfactory CNG channels. *PLOS Comput. Biol.* 18: e1010376. <https://doi.org/10.1371/journal.pcbi.1010376>

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)

Schorge, S., and D. Colquhoun. 2003. Studies of NMDA receptor function and stoichiometry with truncated and tandem subunits. *J. Neurosci.* 23: 1151–1158. <https://doi.org/10.1523/JNEUROSCI.23-04-01151.2003>

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. 2011. Use of concatemers of ligand-gated ion channel subunits to study mechanisms of steroid potentiation. *Anesthesiology*. 115: 1328-1337. <https://doi.org/10.1097/ALN.0b013e318233046a>

Steinbach, J.H., and G. Akk. 2019. Applying the monod-wyman-changeux allosteric activation model to pseudo-steady-state responses from GABA<sub>A</sub> receptors. *Mol. Pharmacol.* 95:106-119. <https://doi.org/10.1124/mol.112.113787>

Stelmashenko, O., U. Lalo, Y. Yang, L. Bragg, R.A. North, and V. Compan. 2012. Activation of trimeric P2X2 receptors by fewer than three ATP molecules. *Mol. Pharmacol.* 82:760-766. <https://doi.org/10.1124/mol.112.080903>

Sunkara, M.R., T. Schwabe, G. Ehrlich, J. Kusch, and K. Benndorf. 2018. All four subunits of HCN2 channels contribute to the activation gating in an additive but intricate manner. *J. Gen. Physiol.* 150:1261-1271. <https://doi.org/10.1085/jgp.20171935>

Tytgat, J., and P. Hess. 1992. Evidence for cooperative interactions in potassium channel gating. *Nature*. 359:420-423. <https://doi.org/10.1038/359420a0>

Ulens, C., and S.A. Siegelbaum. 2003. Regulation of hyperpolarization-activated HCN channels by cAMP through a gating switch in binding domain symmetry. *Neuron*. 40:959-970. [https://doi.org/10.1016/S0896-6273\(03\)00753-0](https://doi.org/10.1016/S0896-6273(03)00753-0)

Weill, C.L., M.G. McNamee, and A. Karlin. 1974. Affinity-labeling of purified acetylcholine receptor from *Torpedo californica*. *Biochem. Biophys. Res. Commun.* 61:997-1003. [https://doi.org/10.1016/0006-291X\(74\)90254-X](https://doi.org/10.1016/0006-291X(74)90254-X)

Wongsamitkul, N., V. Nache, T. Eick, S. Hummert, E. Schulz, R. Schmauder, J. Schirmeyer, T. Zimmer, and K. Benndorf. 2016. Quantifying the cooperative subunit action in a multimeric membrane receptor. *Sci. Rep.* 6: 20974. <https://doi.org/10.1038/srep20974>

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<sup>2+</sup> release channel by sarcoplasmic reticulum luminal Ca<sup>2+</sup>. *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)

Zhou, Y., M.E. Nelson, A. Kuryatov, C. Choi, J. Cooper, and J. Lindstrom. 2003. Human alpha4beta2 acetylcholine receptors formed from linked subunits. *J. Neurosci.* 23:9004-9015. <https://doi.org/10.1523/JNEUROSCI.23-27-09004.2003>

Zhu, H., and E. Gouaux. 2021. Architecture and assembly mechanism of native glycine receptors. *Nature*. 599:513-517. <https://doi.org/10.1038/s41586-021-04022-z>

## Supplemental material

### Supplemental methods

#### Specification of the parameters for the CHA<sup>x</sup> models

For  $E_x$ , we set the value of the single-liganded channel to  $10^{-1}$  and that of the fully liganded channel to  $10^2$ , i.e., we assumed that maximal liganding causes a 1,000-fold increase of  $E_x$  with respect to  $E_l$ . Because the number of  $E_x$  values increases with  $n_D$ , we set the intermediate  $E_x$  values equidistant between  $E_l$  and the maximum  $E_x$  on a logarithmic scale according to

$$E_x = 10^{\left(-1 + \frac{3(x-1)}{(n_D-1)}\right)}, x = (1 \dots n_D). \quad (23)$$

The values are shown in Fig. S2 (green).

For  $fd_x$ , the algorithm is analogous.  $fd_l$  was set to  $10^{-4}$  and that of the maximally impeded binding site was set to  $10^{-2}$ , i.e., we assumed a maximum difference between the binding sites of 100. Because the number of  $fd_x$  values increases with  $n_D$ , we set the intermediate  $fd_x$  values equidistant between  $fd_l$  and the maximum  $fd_x$  on a logarithmic scale according to

$$fd_x = 10^{\left(-4 + \frac{2(x-1)}{(n_D-1)}\right)}, x = (1 \dots n_D). \quad (24)$$

The values are shown in Fig. S2 (red).

For  $Z_x$ , the situation is more complex because the number  $Z_x$ ,  $n_{Zx}$ , rises exponentially according to Eq. 19 and the number of options at an equal degree of liganding are given by binomial coefficients. The lower and the higher border of  $Z_x$  was set to  $10^{-2}$  and  $10^{-1}$ , respectively, according to.

CHA<sup>2</sup> model:

$$Z_x = 10^{-2+(x-1)}, x = 1, 2, \quad (25)$$

$$Z_3 = Z_2. \quad (26)$$

CHA<sup>3</sup> model:

$$Z_x = Z_{3+x} = 10^{-2+(x-1)/2}, x = 1 \dots 3, \quad (27)$$

$$Z_7 = Z_6. \quad (28)$$

CHA<sup>4</sup> model:

$$Z_x = Z_{10+x} = 10^{-2+(x-1)/3}, x = 1 \dots 4, \quad (29)$$

$$Z_{x+4} = 10^{-2+(x-1)/5}, x = 1 \dots 6, \quad (30)$$

$$Z_{15} = Z_{14}. \quad (31)$$

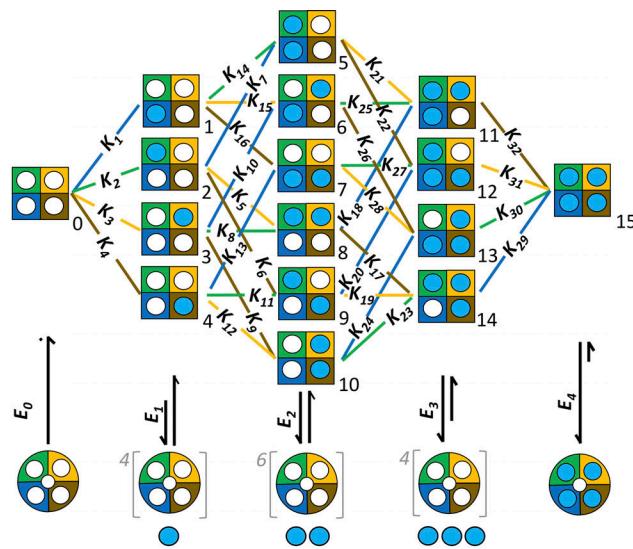
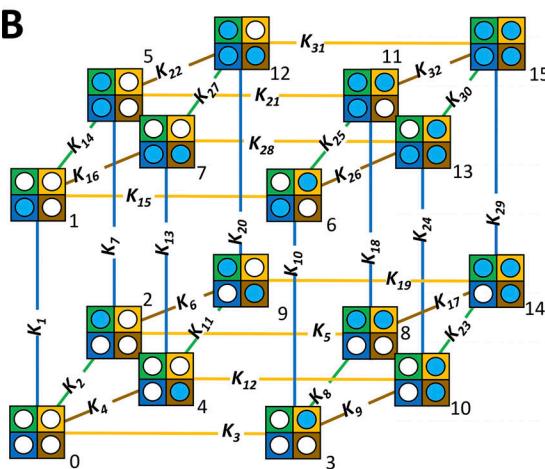
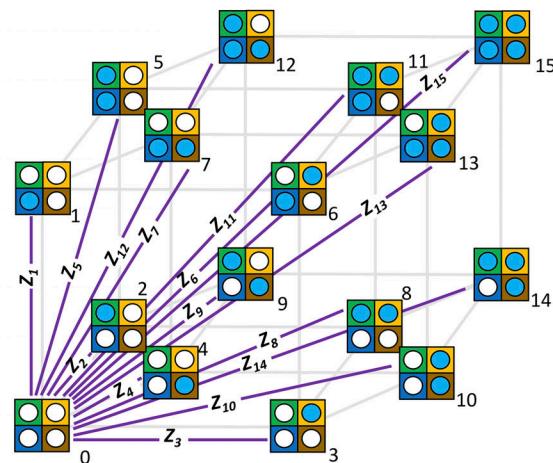
CHA<sup>5</sup> model:

$$Z_x = Z_{25+x} = 10^{-2+(x-1)/4}, x = 1 \dots 5, \quad (32)$$

$$Z_{5+x} = Z_{15+x} = 10^{-2+(x-1)/9}, x = 1 \dots 10, \quad (33)$$

$$Z_{31} = Z_{30}. \quad (34)$$

All values are illustrated in Fig. S2 (blue).

**A**

**B**

**C**


**Figure S1.  $\text{HA}^4$  model. (A)**  $\text{HA}^4$  model in systematic representation. Equilibrium association constants for ligand binding,  $K_x$  (x = 1...32). For globally fitting, 16 such intimately coupled  $\text{HA}^4$  models with multiple combinations of mutated binding sites were combined with the comprehensive  $\text{CHA}^4$  model (Schirmeyer et al., 2021). Blue circles, ligand; white circles, empty binding site. The set of closed–open isomerizations on the bottom is valid for all  $\text{HA}^4$  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). **(B)**  $\text{HA}^4$  model of A in cube representation. The closed–open isomerizations ( $E_x$ ) were omitted for clarity. **(C)** 4-D model of A in cube representation with virtual equilibrium association constants,  $Z_x$ , specifying the virtual transitions from state 0.

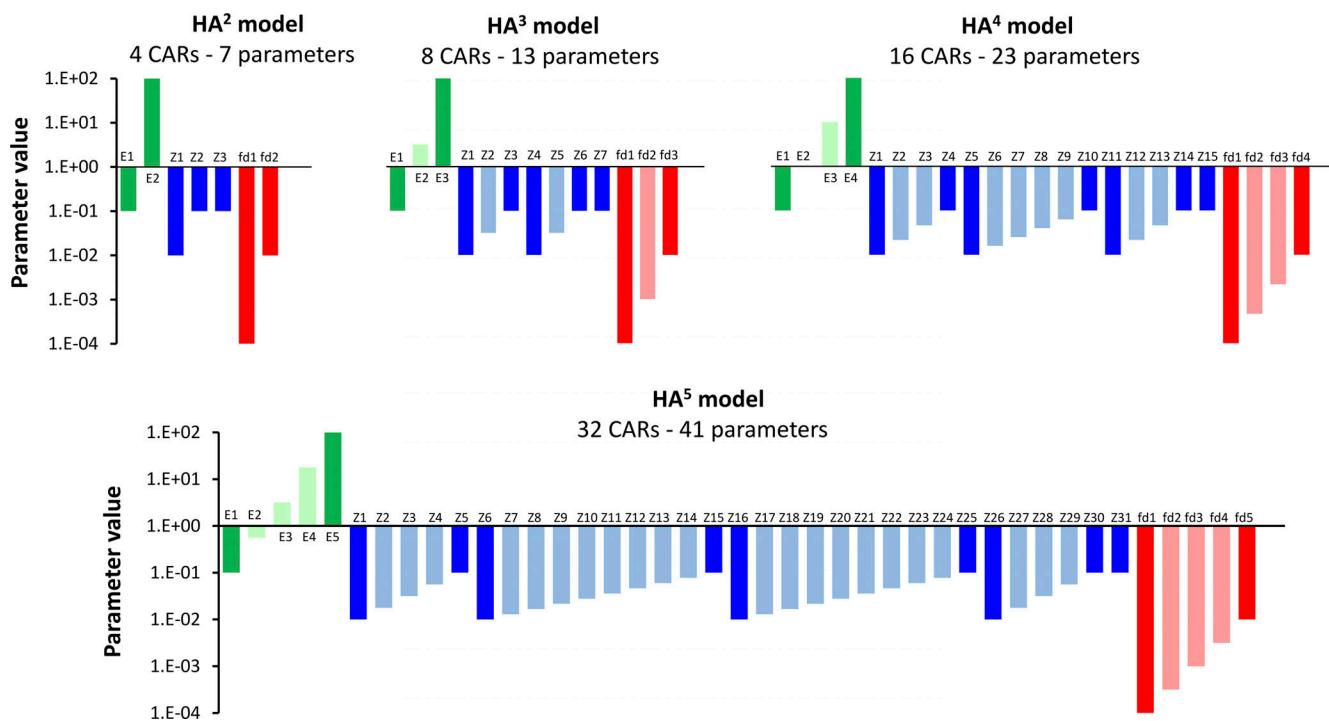


Figure S2. **Setting the theoretical fit parameters.** Theoretical parameters were created on the basis of related real data on heterotetrameric CNG channels (Schirmeyer et al., 2021; Benndorf et al., 2022; for details see Eqs. 23, 24, 25, 27, 29, and 32 in the supplemental text at the end of the PDF). The set parameters are termed throughout as true parameters,  $x_{i,tr}$ . The saturated colors indicate parameters at respective upper and lower borders and the shaded colors parameters with interpolated values.

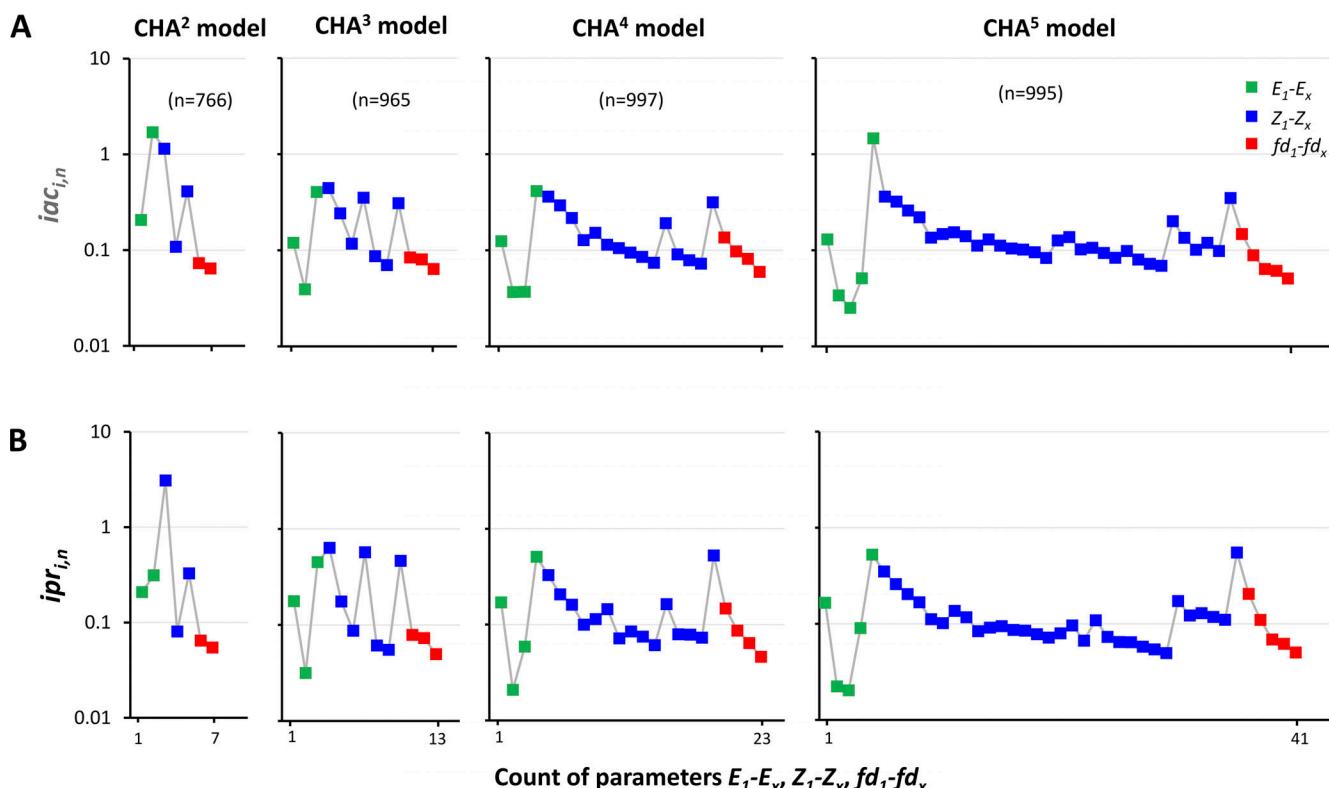
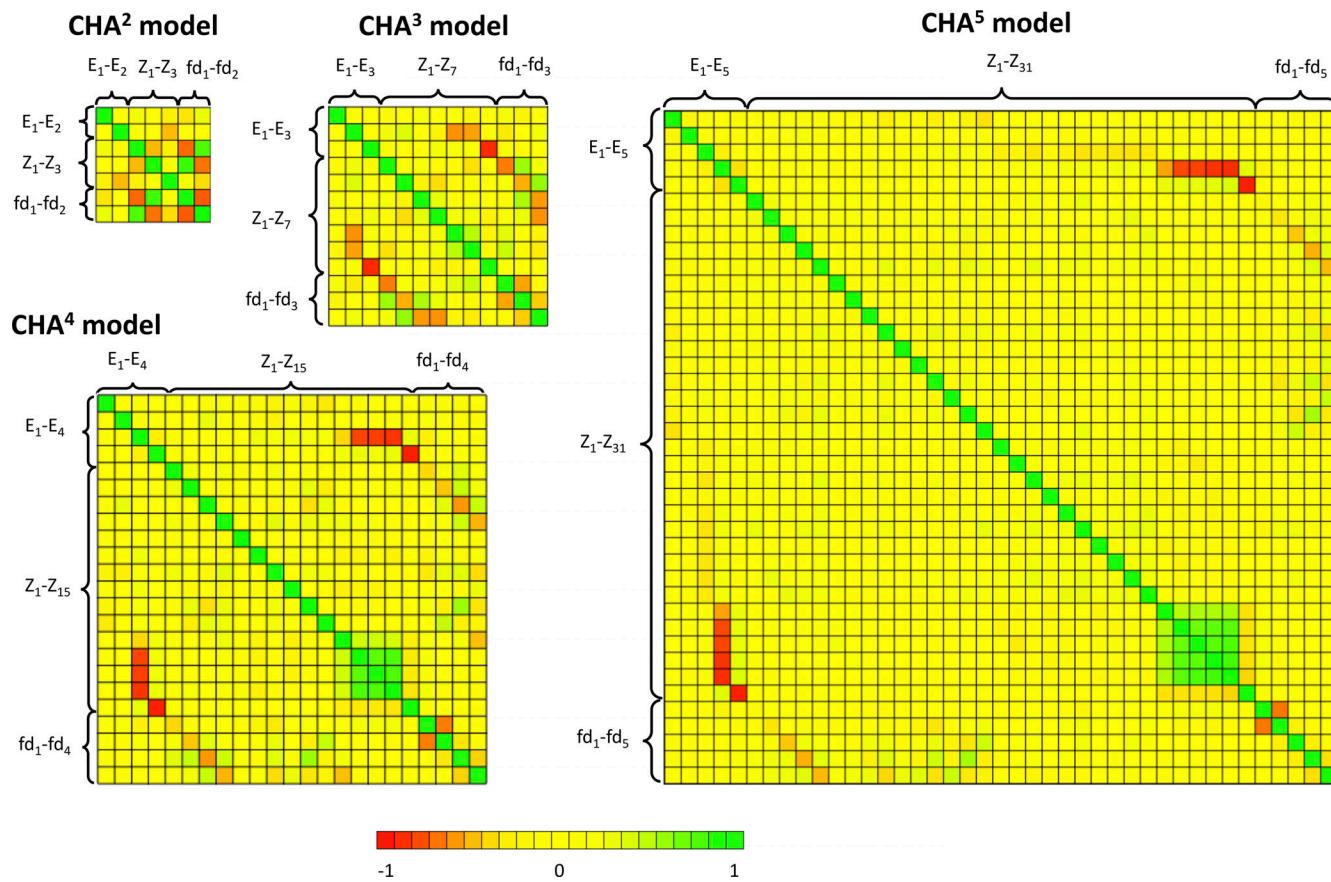


Figure S3. **Inaccuracy and imprecision of the parameters in global fits of  $\text{CHA}^x$  models.** (A) Inaccuracy. The mean inaccuracy for each parameter  $x_i$ ,  $iac_{i,n}$  was computed by Eq. 14. The number of included successful fits,  $n$ , is indicated in brackets. (B) Imprecision. The mean imprecision for each parameter  $x_i$ ,  $ipr_{i,n}$  was computed by Eq. 11. The diagrams confirm that the channels with more binding sites provide a lower imprecision than the channels with less binding sites.



**Figure S4. Color-coded correlation matrices for CHA<sup>x</sup> models.** The correlation coefficients were calculated from the covariance matrices (see Materials and methods) and plotted by false colors (see legend at the bottom; Benndorf et al., 2022). Few parameters are strongly correlated, either positive or negative. Notably, non-correlated parameters (colored yellow) dominate most for the CNA<sup>5</sup> and least for the CNA<sup>2</sup> model. For further explanation, see Simulating the CARs in the Results.

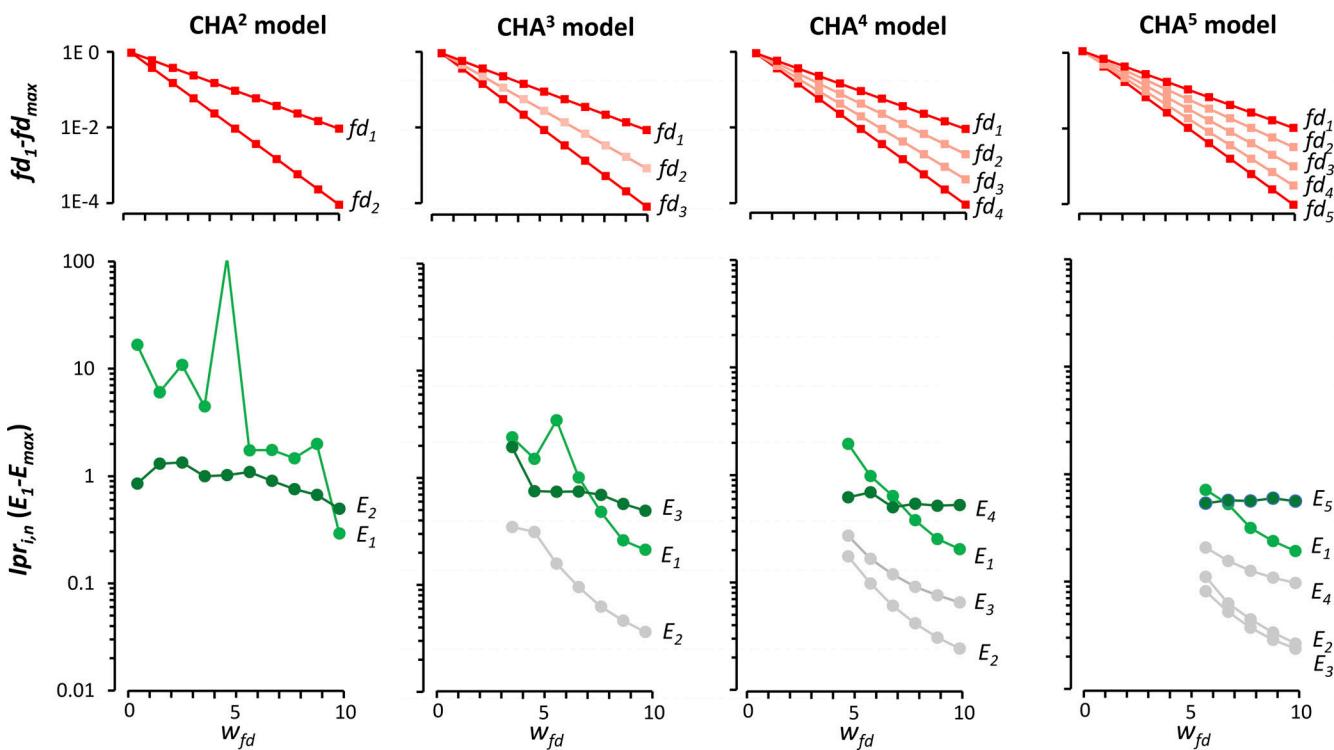


Figure S5. **Effects of mutational factors on the identifiability of  $E_x$ .** Data are provided for the CHA<sup>2</sup> to the CHA<sup>5</sup> model. At the top, the schemes for the weakening steps of all  $f_{dx}$  are indicated. The imprecision of  $E_x$  decreases from the CHA<sup>2</sup> to the CHA<sup>4</sup> or CHA<sup>5</sup> model. Also,  $E_1$  and  $E_{max}$  (green) are more uncertain than the respective intermediate  $E_x$  (gray).

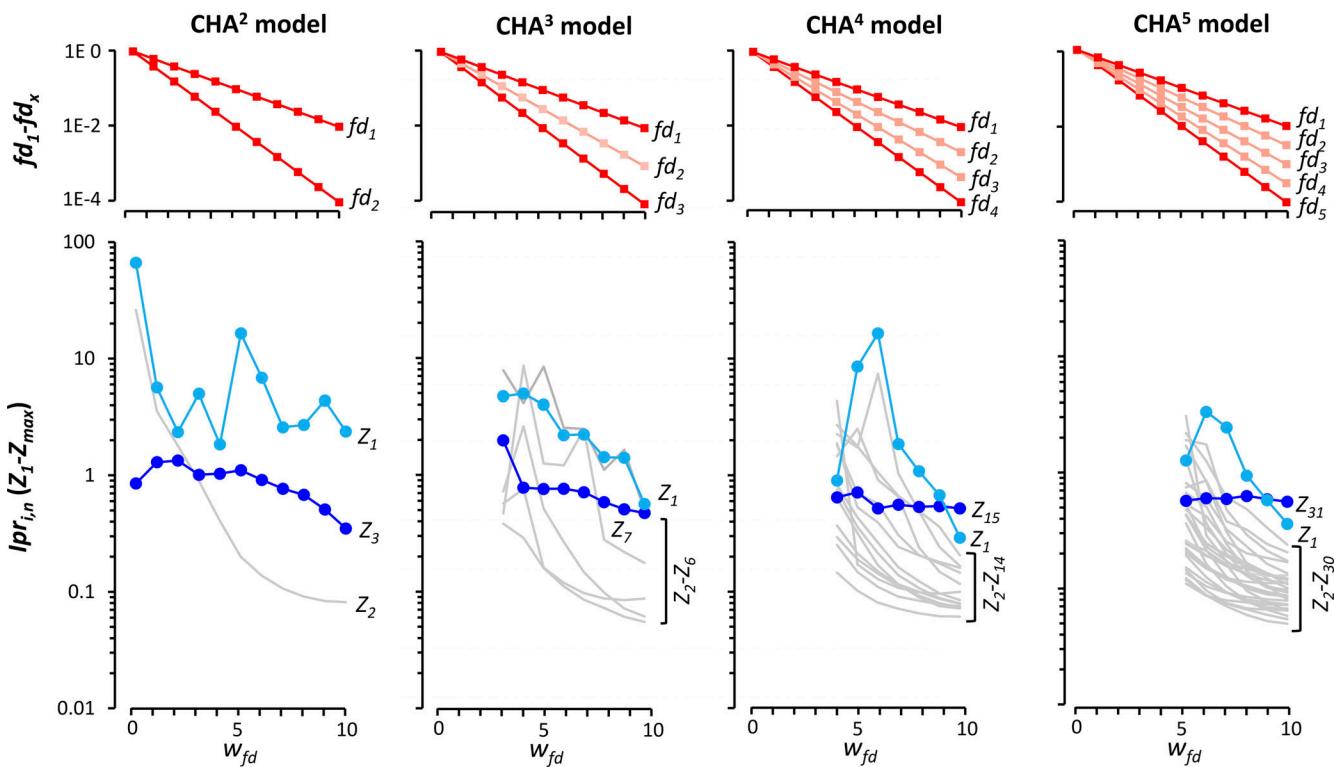


Figure S6. **Effects of mutational factors on the identifiability of  $Z_x$ .** Data are provided for the CHA<sup>2</sup> to the CHA<sup>5</sup> model. At the top the schemes for the weakening steps of all  $f_{dx}$  are indicated. The imprecision of  $Z_x$  decreases from the CHA<sup>2</sup> to the CHA<sup>4</sup> or CHA<sup>5</sup> model. As with  $E_x$ ,  $Z_1$ , and  $Z_{max}$  (green) are more uncertain than the respective intermediate  $Z_x$  (gray). For clarity, the actually discrete points for the intermediate  $Z_x$  were omitted.

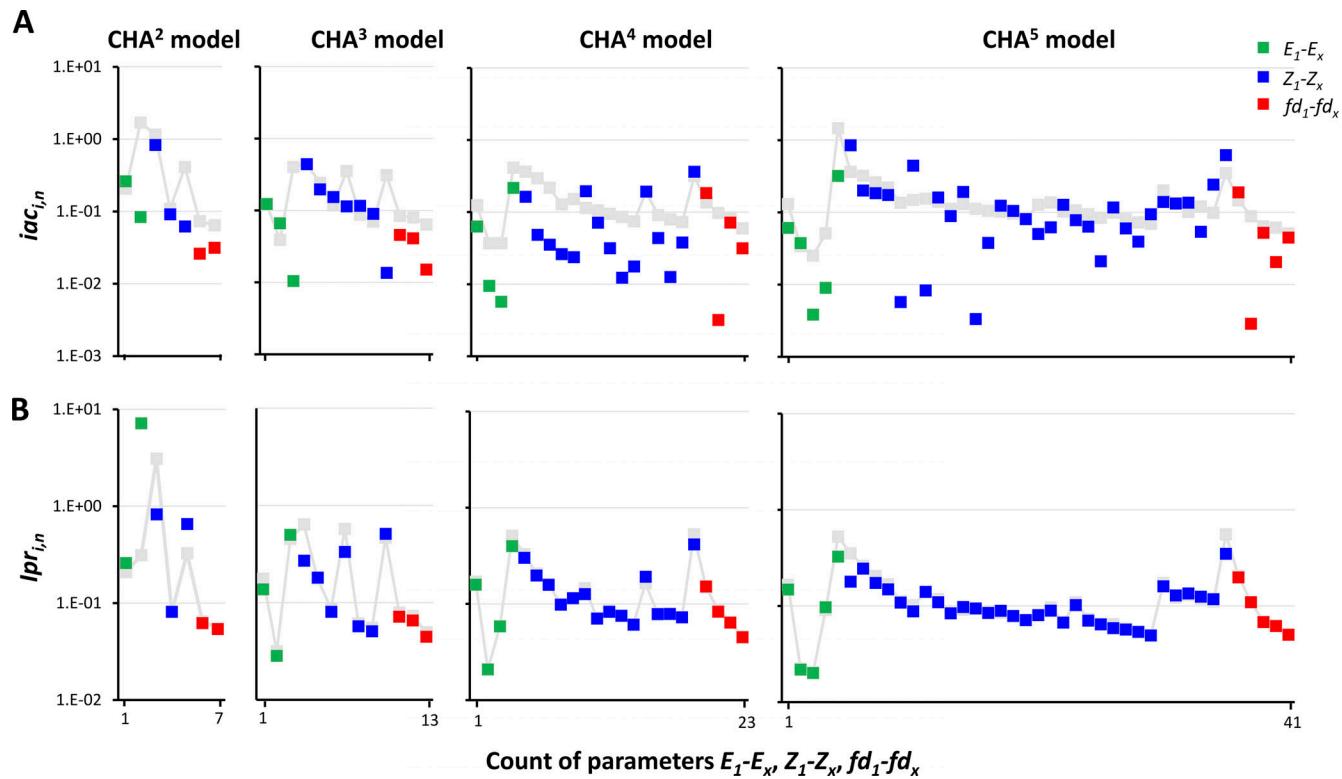


Figure S7. **Inaccuracy and imprecision with equal SU start vector and stochastically varied elements.** The values were obtained from the fits in Fig. 11. **(A)** Inaccuracy. The mean inaccuracy for each parameter  $x_i$ ,  $iac_{i,n}$ , was computed by Eq. 14. The number of included successful<sub>pΔMSE</sub> fits,  $n$ , is indicated in brackets. **(B)** Imprecision. The mean imprecision for each parameter  $x_i$ ,  $ipr_{i,n}$ , was computed by Eq. 11. The gray points indicate the values of Fig. S3 for comparison.

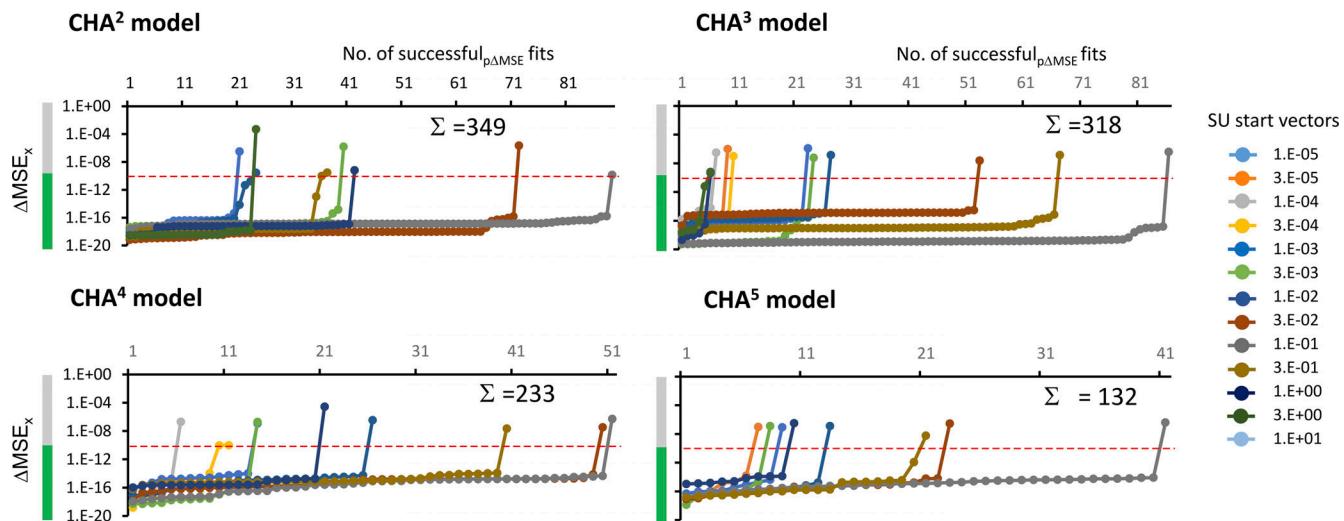


Figure S8. **Successful<sub>pΔMSE</sub> fits with different stochastically varied SU start vectors.** The variation of the scaled SU start vectors followed the algorithm described for Fig. 9 A with the difference that the data points were generated de novo for each SU vector. Successful<sub>pΔMSE</sub> fits were selected as described in Fig. 9 B by using a threshold (red dashed lines) at  $10^{-10}$  for the corresponding  $\Delta MSE_x$  values. Again, for all four models, the sums of all obtained successful<sub>pΔMSE</sub> fits out of 19,000 fits are indicated.

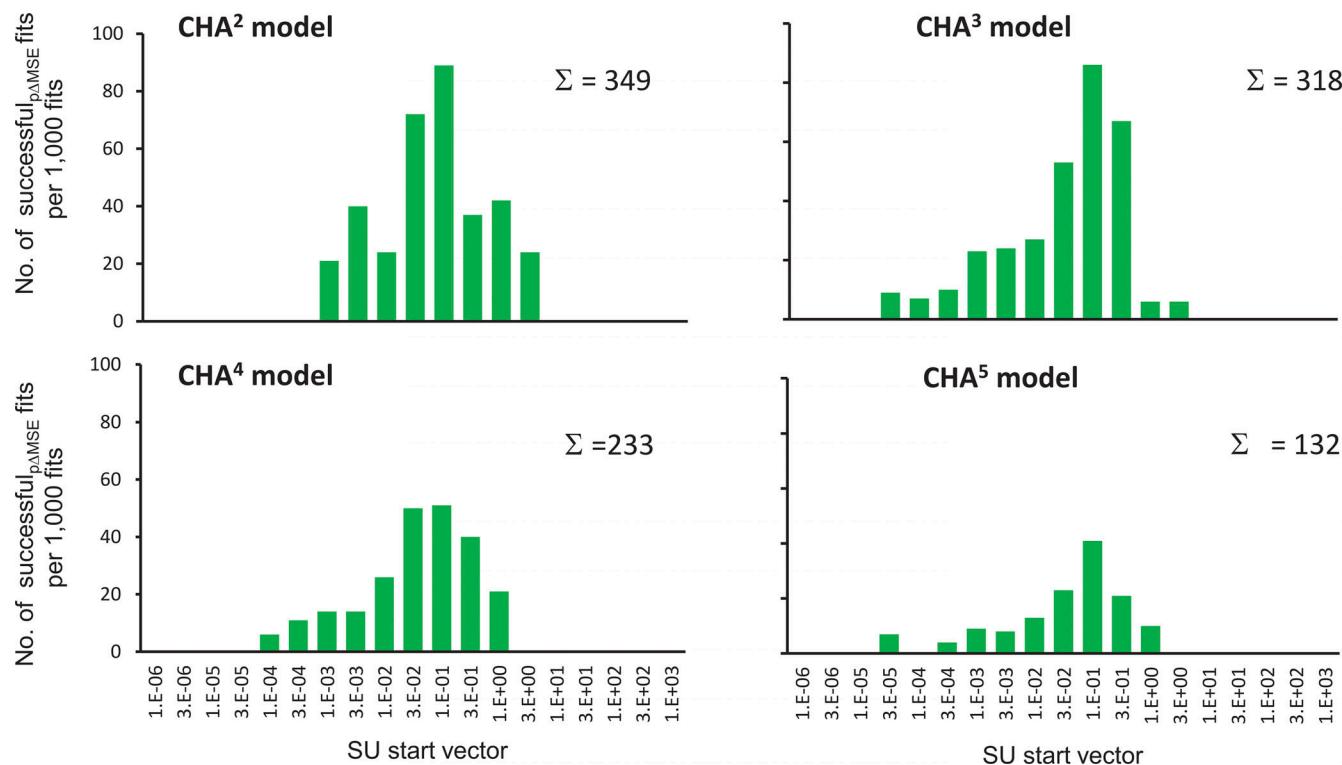


Figure S9. **Histograms of successful<sub>pΔMSE</sub> fits with different SU vectors and stochastically varied elements.** The data correspond to those in Fig. S8. Analogous to Fig. 10, for each of the 19 SU vectors (abscissa), at least five successful<sub>pΔMSE</sub> fits were required to include them in the histogram. Reasonable numbers of successful<sub>pΔMSE</sub> fits were obtained for all CHA<sup>x</sup> models. The number of successful<sub>pΔMSE</sub> fits was maximal near the SU vector of 10<sup>-1</sup>.

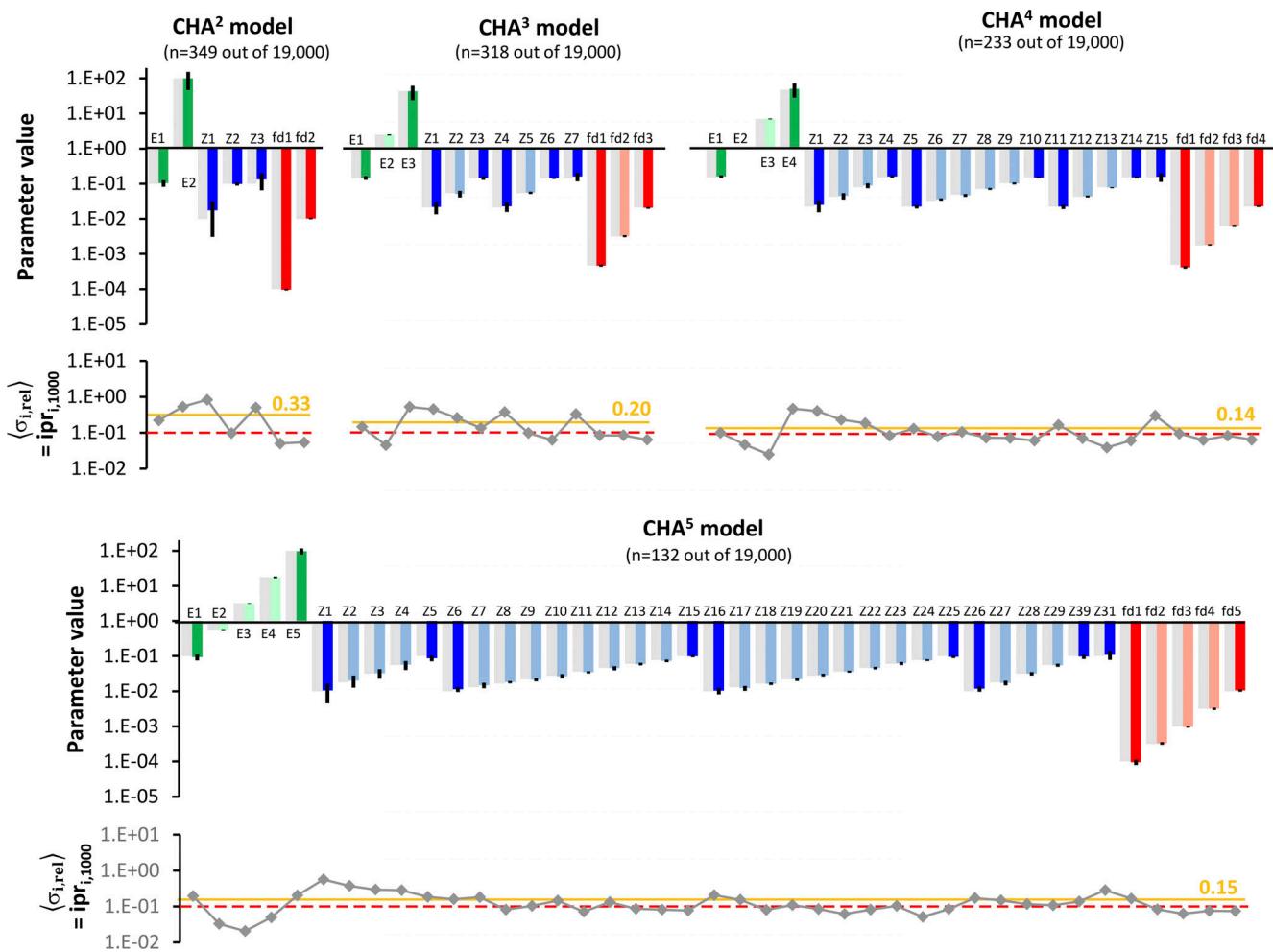


Figure S10. **Comparison of fit parameters with different SU vectors and stochastically varied elements.** Same formalism as in Figs. 7 and 11. In the bar graphs, the fitted parameters,  $x_p$ , are compared with the true parameters,  $x_{i,tr}$  (gray). Means and standard deviations of these fits are shown. The mean relative standard deviations  $\sigma_{i,rel}$  of 1,000 fits,  $ipr_{i,1000}$  (Eq. 11), below the bar graphs indicate the relative errors of each parameter. The ochre lines and numbers right specify the overall  $ipr_{i,1000}$  value for each model and the dashed red line the level of 0.1, i.e., the 10% error.

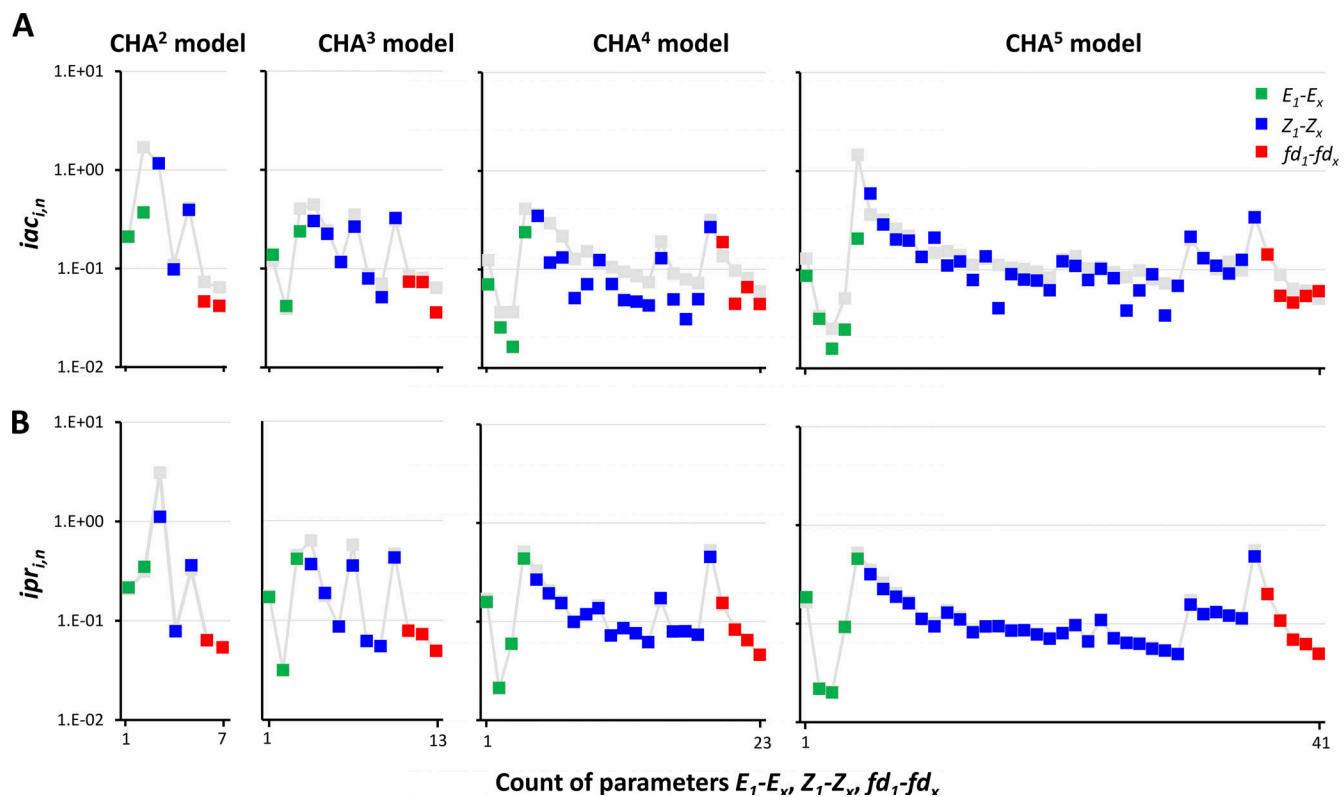


Figure S11. **Inaccuracy and imprecision of the parameters for successful  $p_{\text{AMSE}}$  fits with different SU vectors and stochastically varied elements.** Same formalism as in Figs. S3 and S7. **(A)** Inaccuracy. **(B)** Imprecision. The gray points indicate the values of Fig. S3 for comparison.

Four tables are provided online. Table S1 shows the generation of  $K_x$  from  $Z_x$  in HA<sup>2</sup> model. Table S2 shows the generation of  $K_x$  from  $Z_x$  in HA<sup>3</sup> model. Table S3 shows the generation of  $K_x$  from  $Z_x$  in HA<sup>4</sup> model. Table S4 shows the generation of  $K_x$  from  $Z_x$  in HA<sup>5</sup> model used in this study.