

**Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig**

**Decoupled molecules with binding
polynomials of bidegree $(n,2)$**

by

Yue Ren, Johannes W. R. Martini, and Jacinta Torres

Preprint no.: 75

2017



DECOUPLED MOLECULES WITH BINDING POLYNOMIALS OF BIDEGREE $(n, 2)$

YUE REN, JOHANNES W. R. MARTINI, AND JACINTA TORRES

ABSTRACT. We present a result on the number of decoupled molecules for systems binding two different types of ligands. In the case of n and 2 binding sites respectively, we show that, generically, there are $2(n!)^2$ decoupled molecules with the same binding polynomial. For molecules with more binding sites for the second ligand, we provide computational results.

1. INTRODUCTION

In biology, a ligand is a substance that binds to a target molecule to serve a given purpose. A classical [6, 20] and intensively studied [2, 21] example is oxygen, which binds reversibly to hemoglobin to be transported through the bloodstream. Reversible mutual binding of different molecules is also a key feature in biological signal transduction [10, 11, 17, 19] and gene regulation [18].

A common model for describing equilibrium and steady states of a ligand L binding to the sites of a target molecule M comes from the grand canonical ensemble of statistical mechanics [4, 22, 34, 37]. The grand partition function, in our context also known as the *binding polynomial*, arises as the denominator of the rational function describing the average number of occupied binding sites as a function of ligand activity. In the case of a target molecule with only one binding site, this rational function is given by

$$\Psi(\Lambda) = \frac{a\Lambda}{a\Lambda + 1},$$

where the variable Λ denotes the activity of the ligand in the environment, and a is a transformation of the binding energy depending on the temperature, which is usually assumed to be constant. This equation is also known as the (sigmoid) Henderson-Hasselbalch titration curve. Titration refers to the laboratory method used to obtain this curve. For systems of molecules with n binding sites it generalizes to the Adair equation [1, 36]:

$$\Psi(\Lambda) = \frac{na_n\Lambda^n + (n-1)a_{n-1}\Lambda^{n-1} + \dots + a_1}{a_n\Lambda^n + a_{n-1}\Lambda^{n-1} + \dots + a_1 + 1}.$$

In this model, the roots of the binding polynomial play an important role for the characterization of the binding behavior of the ligand to the target molecule [7, 8, 9, 12]. The rational functions of systems with n binding sites can be represented as

sums of n Henderson-Hasselbalch curves [31, 32, 33], which means that any given system of interacting binding sites can be represented by a hypothetical molecule consisting of stochastic independent binding sites [28] possessing the same titration curve. The roots of the binding polynomial determine the binding energies of the independent pseudo-sites in this so-called *decoupled sites representation*.

For two different types of ligands, the binding polynomial has two variables representing the activities of both ligands in the environment. Seeking an analogous decoupled sites representation leads to a version in which the binding sites for the same type of ligand do not interact, but interaction terms between the sites of different ligands remain [29, 30]. Contrary to the case of one type of ligand, where the decoupled sites representation is unique up to permutation of the roots, there are several different decoupled molecules. It has been shown previously that in the case of n and 1 binding sites for the two ligands, respectively, there are $n!$ decoupled molecules. The situation becomes more complicated for general systems of n_1 and n_2 binding sites. The main goal of this paper is to prove the following theorem.

Theorem 1.1

The decoupled molecules with $(n, 2)$ sites with a fixed binding polynomial of bidegree $(n, 2)$ are the solutions to a system of $3n + 2$ unknowns: the $n + 2$ binding energies and the $2n$ interaction energies. Generically, the number of complex solutions to this system equals $4(n!)^3$. These come in $2(n!)^2$ classes under relabeling of the sites.

The article is structured as follows: In Section 2, we recall the definition of the binding polynomial and formulate the central question addressed in this work. In Section 3, we recall some results and techniques of numerical algebraic geometry, which are necessary to prove the main theorem in Section 4. We conclude the article with some experimental results in Section 5 and some open questions in Section 6.

ACKNOWLEDGEMENTS

The authors would like to thank Bernd Sturmfels for suggesting this collaboration and for his comments on a previous version of this paper, Corey Harris for his useful remarks and Jon Hauenstein for his advice regarding the BERTINI computations.

2. BACKGROUND AND FRAMEWORK

In this section, we briefly recap the algebraic framework as well as past results, and, in doing so, fix various notations. Most importantly, we introduce some shorthand notation for molecules with $(n, 2)$ sites for Sections 3 and 4.

2.1. Single type of ligand. The binding behaviour of systems with one type of ligand is governed by the energies required to bind to each site of the target molecule and the way different binding sites interact with each other. Following the notation of [31], we identify target molecules with these parameters.

Definition 2.1

A molecule M with n sites for one type of ligand is a point

$$M = (g_1, \dots, g_n, w_{1,2}, w_{1,3}, \dots, w_{n-1,n}) \in (\mathbb{C}^*)^n \times (\mathbb{C}^*)^{\binom{n}{2}}.$$

The g_i are called the *binding energies* and the $w_{i,j}$ are called the *interaction energies*; they measure, respectively, the energy at each site i and the interaction energy between sites i and j (see Figure 1). We call M *decoupled* if $w_{i,j} = 1$ for all $1 \leq i < j \leq n$.

We will consider the natural S_n action that corresponds to relabelling the sites:

$$\sigma \cdot (g_1, \dots, g_n, w_{1,2}, \dots, w_{n-1,n}) := (g_{\sigma(1)}, \dots, g_{\sigma(n)}, w_{\sigma(1),\sigma(2)}, \dots, w_{\sigma(n-1),\sigma(n)})$$

for $\sigma \in S_n$.

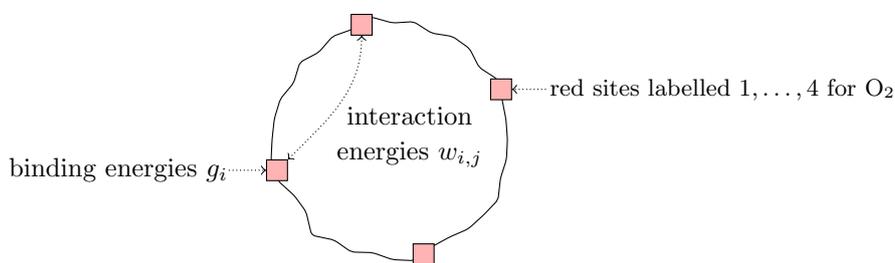


FIGURE 1. Hemoglobin with its 4 sites for oxygen.

Definition 2.2

Given a labelled target molecule M with n sites, we refer to $K := \{0, 1\}^n$ as the set of all microstates. To each microstate $k = (k_1, \dots, k_n) \in K$ we associated a microstate constant

$$g(k) := \prod_{i=1}^n \left(g_i^{k_i} \prod_{j=i+1}^n w_{i,j}^{k_i k_j} \right).$$

The *binding polynomial* is then defined as

$$P_M(\Lambda) = \sum_{k \in K} g(k) \Lambda^{|k|} \in \mathbb{C}[\Lambda].$$

It is a polynomial of degree n with constant term 1, and the map $M \mapsto P_M$ is constant on the S_n orbits, i.e. $P_M(\Lambda) = P_{\sigma(M)}(\Lambda)$ for every $M \in (\mathbb{C}^*)^{\frac{n(n+1)}{2}}$ and all $\sigma \in S_n$.

The following theorem is also known as the *decoupled sites representation*. It implies that any molecule with real binding and interaction energies can be uniquely represented by a molecule with neutral interaction energy, provided that complex binding energies are allowed. Its proof consists of a reformulation of Vieta's formulas.

Theorem 2.3 ([31, Proposition 2])

For any molecule N there exists a decoupled labelled target molecule M , unique up to relabelling of the sites, such that $P_M = P_N$.

2.2. Multiple types of ligands. In case of $d > 1$ types of ligands, we consider each binding site to be only able to take up to one type of ligand [30]. This is sensible, as we can model a single binding site capable of binding to two types of ligands as two binding sites with interaction energies set so that the two sites can never be saturated at the same time.

For our purposes, let us assume that $d = 2$. We write n_1 and n_2 for the number of sites capable of binding to the first and second ligand, respectively.

Definition 2.4

A molecule M with (n_1, n_2) sites is a point

$$M = (g_{T_1}, \dots, g_{T_{n_1}}, g_{S_1}, \dots, g_{S_{n_2}}, (w_P)_{P \subset \{T_i, S_j\}, |P|=2}) \in (\mathbb{C}^*)^{n_1+n_2} \times (\mathbb{C}^*)^{\binom{n_1+n_2}{2}},$$

where $T_1, \dots, T_{n_1}, S_1, \dots, S_{n_2}$ represent the binding sites for ligand type T and S respectively (see Figure 2) and

- $g_{T_1}, \dots, g_{T_{n_1}}$ and $g_{S_1}, \dots, g_{S_{n_2}}$ are the binding energies,
- w_P for $P \subset \{T_1, \dots, T_{n_1}, S_1, \dots, S_{n_2}\}$ with $|P| = 2$ are the interaction energies.

We call M *decoupled*, if $w_P = 1$ for $P \subset \{T_1, \dots, T_{n_1}\}$ and $P \subset \{S_1, \dots, S_{n_2}\}$.

Similar to the case $d = 1$, there is a natural $S_{n_1} \times S_{n_2}$ action that corresponds to relabelling the sites.

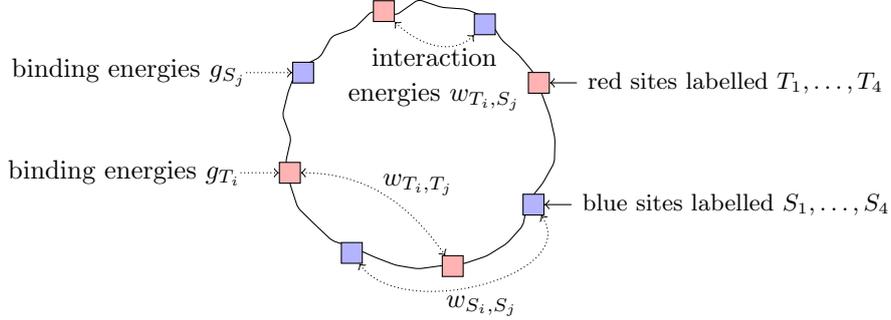


FIGURE 2. A molecule with $(4,4)$ sites.

Definition 2.5

Similarly to the case $d = 1$, we can define the binding polynomial P_M of a molecule M . Explicitly, for decoupled molecules M , P_M is a bivariate polynomial in the two (ligand) variables Λ_1 and Λ_2 ,

$$P_M(\Lambda_1, \Lambda_2) = \sum_{i=1, \dots, n_1} \sum_{j=1, \dots, n_2} a_{i,j} \Lambda_1^i \Lambda_2^j,$$

where the coefficients $a_{i,j}$ are given by

$$a_{i,j} = \sum_{\substack{I \subset \{1, \dots, n_1\} \\ J \subset \{1, \dots, n_2\} \\ |I|=i, |J|=j}} \prod_{T_i \in I} g_{T_i} \prod_{S_j \in J} g_{S_j} \prod_{\substack{S_j \in I \\ T_i \in J}} w_{\{T_i, S_j\}}. \quad (1)$$

It is a bivariate polynomial of bidegree (n_1, n_2) with constant term 1. Moreover, the map $M \mapsto P_M$ is constant on the $S_{n_1} \times S_{n_2}$ -orbits, i.e. $P_M(\Lambda) = P_{\sigma(M)}(\Lambda)$ for every $M \in (\mathbb{C}^*)^{n_1+n_2} \times (\mathbb{C}^*)^{\binom{n_1+n_2}{2}}$ and all $\sigma \in S_{n_1} \times S_{n_2}$.

In this case, the decoupled sites representation takes the following form.

Theorem 2.6 ([30, Corollary 2])

For any molecule N with $(n, 1)$ sites there exists, up to relabelling of the sites, and counted with multiplicity, $n!$ decoupled molecules M of the same type such that $P_N = P_M$.

2.3. Decoupled molecules with $(n, 2)$ sites. The main focus of this article are decoupled molecules with $(n, 2)$ sites, for which we will simplify the notation as follows: instead of T_1, \dots, T_n , we label the n binding sites of the first type with $1, \dots, n$, and, instead of S_1, S_2 , we label the two binding sites of the second type with A, B (see Figure 3), so that

- $g_1, \dots, g_n, g_A, g_B$ represent the binding energies,
- $w_{1,A}, \dots, w_{n,A}, w_{1,B}, \dots, w_{n,B}$ represent the non-trivial interaction energies.

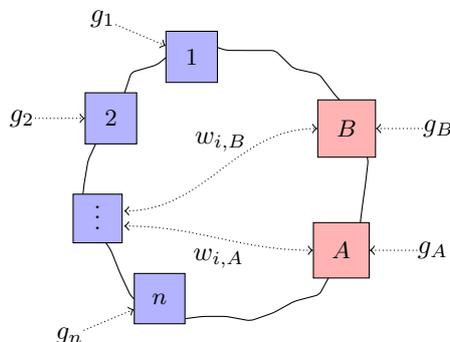


FIGURE 3. A decoupled molecule with $(n, 2)$ sites.

The formulas for the coefficients of the binding polynomial then simplify to the polynomials in System (2). For an explicit instance of the equations and their solutions, see Section 5.1. We denote the pair set of decoupled molecules with $(n, 2)$ sites and their binding polynomials of bidegree $(n, 2)$ (ignoring its constant term 1) by

$$\mathcal{M} = \{(\underline{g}, \underline{w}; \underline{a}) \in (\mathbb{C}^*)^{n+2} \times (\mathbb{C}^*)^{n \cdot 2} \times \mathbb{C}^{(n+1) \cdot (2+1)-1} \mid (\underline{g}, \underline{w}; \underline{a}) \text{ satisfies System (2)}\}.$$

$$\begin{aligned}
a_{1,0} &= g_1 + \dots + g_n, \\
&\vdots \\
a_{n,0} &= g_1 \cdot \dots \cdot g_n, \\
a_{0,1} &= g_A + g_B, \\
a_{1,1} &= g_A(g_1 w_{1,A} + \dots + g_n w_{n,A}) + g_B(g_1 w_{1,B} + \dots + g_n w_{n,B}), \\
a_{2,1} &= g_A(g_1 g_2 w_{1,A} w_{2,A} + \dots + g_{n-1} g_n w_{n-1,A} w_{n,A}) \\
&\quad + g_B(g_1 g_2 w_{1,B} w_{2,B} + \dots + g_{n-1} g_n w_{n-1,B} w_{n,B}), \\
&\vdots \\
a_{n,1} &= g_A g_1 \dots g_n w_{1,A} \dots w_{n,A} + g_B g_1 \dots g_n w_{1,B} \dots w_{n,B}, \\
a_{0,2} &= g_A g_B, \\
a_{1,2} &= g_A g_B (g_1 w_{1,A} w_{1,B} + \dots + g_n w_{n,A} w_{n,B}), \\
a_{2,2} &= g_A g_B (g_1 w_{1,A} w_{1,B} g_2 w_{2,A} w_{2,B} + \dots + g_{n-1} w_{n-1,A} w_{n-1,B} g_n w_{n,A} w_{n,B}), \\
&\vdots \\
a_{n,2} &= g_A g_B g_1 w_{1,A} w_{1,B} \dots g_n w_{n,A} w_{n,B}.
\end{aligned} \tag{2}$$

FIGURE 4. Coefficients of the binding polynomial of bidegree $(n,2)$.

3. NUMERICAL ALGEBRAIC GEOMETRY

In this section we recall some basic notions of numerical algebraic geometry and its main workhorse: homotopy continuation. For that, we regard \mathcal{M} as the kernel of the polynomial map

$$f : \underbrace{(\mathbb{C}^*)^{n+2} \times (\mathbb{C}^*)^{n \cdot 2}}_{=:X} \times \underbrace{\mathbb{C}^{(n+1) \cdot (2+1) - 1}}_{=:Y} \rightarrow \mathbb{C}^{(n+1) \cdot (2+1) - 1}$$

with

$$f(\underline{g}, \underline{w}; \underline{a}) = \begin{pmatrix} g_1 + \dots + g_n - a_{1,0} \\ \vdots \\ \vdots \end{pmatrix},$$

where $\underline{g} := (g_1, \dots, g_n, g_A, g_B)$ and $\underline{w} := (w_{1,A}, \dots, w_{n,A}, w_{1,B}, \dots, w_{n,B})$ are referred to as *unknowns*, and $\underline{a} = (a_{1,0}, \dots, a_{n,2})$ are regarded as *parameters*. We fix a projection

$$\pi_Y : X \times Y \longrightarrow Y.$$

One fundamental and important concept is that solutions vary continuously in the parameters, which is summarized in the following theorem.

Theorem 3.1 ([35, Theorem A.14.1])

If there is an isolated solution $(\underline{g}^, \underline{w}^*; \underline{a}^*) \in X \times Y$ of $f(\underline{g}, \underline{w}; \underline{a}^*) = 0$, then there are arbitrarily small euclidean open sets $\mathcal{U} \subset X$ that contain $(\underline{g}^*, \underline{w}^*)$ and such that*

(1) *$(\underline{g}^*, \underline{w}^*; \underline{a}^*)$ is the only solution of $f(\underline{g}, \underline{w}; \underline{a}^*) = 0$ in $\mathcal{U} \cap (X \times \{\underline{a}^*\})$;*

- (2) $f(\underline{g}, \underline{w}; \underline{a}') = 0$ has only isolated solutions for $\underline{a}' \in \pi_Y(\mathcal{U})$ and $(\underline{g}, \underline{w}) \in \mathcal{U} \cap (X \times \{\underline{a}^*\})$;
- (3) the multiplicity of $(\underline{g}^*, \underline{w}^*; \underline{a}^*)$ as a solution of $f(\underline{g}, \underline{w}; \underline{a}^*) = 0$ equals the sum of the multiplicities of the isolated solutions of $f(\underline{g}, \underline{w}; \underline{a}') = 0$ for $\underline{a}' \in \pi_Y(\mathcal{U})$ and $(\underline{g}, \underline{w}) \in \mathcal{U} \cap (X \times \{\underline{a}^*\})$

Example 3.2 (Vieta's Formula)

Consider the first n components of our polynomial map, which are given by (abbreviating $a_i := a_{i,0}$):

$$f(g_1, \dots, g_n; a_1, \dots, a_n) := \begin{pmatrix} g_1 + \dots + g_n - a_1 \\ g_1 g_2 + g_1 g_3 + \dots + g_{n-1} g_n - a_2 \\ \vdots \\ g_1 \cdot \dots \cdot g_n - a_n \end{pmatrix}.$$

Given any parameter $\underline{a}' \in \mathbb{C}^n$, Vieta's formula states that any solution $\underline{g}' \in \mathbb{C}^n$ to $f(\underline{g}; \underline{a}') = 0$ consists of the roots of the univariate polynomial $t^n + a'_1 t^{n-1} + \dots + a'_n$. Hence there exist a Zariski-open set $\mathcal{U} := \mathbb{C}^n \setminus \text{Disc}_x(x^n + a_1 x^{n-1} + \dots + a_n)$ such that for any $\underline{a}' \in \mathcal{U}$ there are $n!$ distinct simple solutions to $f(\underline{g}; \underline{a}') = 0$. We say that there are *generically* $n!$ solutions and refer to $\underline{a}' \in \mathcal{U}$ as a *generic* choice of parameters.

Should $x^n + a'_1 x^{n-1} + \dots + a'_n = (x - 1)^n$, then the only solution is $\underline{g}' = (1, \dots, 1)$. Theorem 3.1 implies that this solution is of multiplicity $n!$. This will be important in the proof of Lemma 4.2.

The arguably most essential tool in numerical algebraic geometry is path tracking. That is given

- a starting solution $(g', w'; a') \in X \times Y$
- a target parameter $a^* \in Y$
- a continuous path $\phi : [0, 1] \rightarrow Y$ with $\phi(1) = a'$ and $\phi(0) = a^*$

there exist, under certain circumstances [35, Theorem 7.1.6], a *solution path*

$$z : (0, 1] \rightarrow X \text{ with } z(1) = (g', w') \text{ and } f(z(t), \phi(t)) = 0.$$

However, the solution path might diverge, which is why these problems are commonly studied in a projective framework.

Example 3.3 (solutions at infinity)

The simplest example of diverging solution path is the function

$$f : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}, \quad f(x; a) = ax^2 - x,$$

with two starting solutions $(0; 1)$, $(1; 1)$, the target parameter 0 and the continuous, straight-line path $\phi : [0, 1] \rightarrow \mathbb{C}, t \mapsto 1 - t$, see Figure 5.

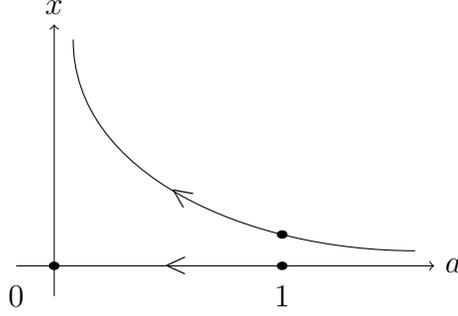


FIGURE 5. A converging and a diverging solution path.

The two solution paths are

$$\begin{aligned} z_1 &: [0, 1) \longrightarrow \mathbb{C}, & t &\longmapsto 0, \\ z_2 &: [0, 1) \longrightarrow \mathbb{C}, & t &\longmapsto \frac{1}{1-t}, \end{aligned}$$

of which the first obviously converges, while the second diverges. Note that diverging paths can only appear if parameters occur in the coefficients of non-constant monomials, which is not the case in System (2), see proof of Theorem 4.4.

4. GENERIC DECOUPLED MOLECULES WITH $(n, 2)$ SITES

In this section, we show that a binding polynomial represents generically $4 \cdot (n!)^3$ decoupled molecules with $(n, 2)$ sites. Due to the complexity of the system of polynomial equations, the proof is split in two parts. First, we study a special class of decoupled molecules and their binding polynomials. In a second step, we study their implication to the generic case.

4.1. Normalized molecules. In this subsection, we restrict ourselves to a special class of decoupled molecules and their binding polynomials. This simplifies our system of equations and allows us to show that their binding polynomials generically represent $2n!$ molecules, each of multiplicity $2(n!)^2$.

Definition 4.1

Recall that \mathcal{M} consists pairs of molecules and their binding polynomials (see Section 2.3). We define the set of all *normalized* molecules to be

$$\mathcal{M}_{\text{norm}} := \left\{ (\underline{g}, \underline{w}; \underline{a}) \in \mathcal{M} \left| \begin{array}{l} g_i = 1 \text{ for } i = 1, \dots, n \text{ and } g_A = g_B = 1 \\ w_{i,A} w_{i,B} = 1 \text{ for } i = 1, \dots, n \end{array} \right. \right\}.$$

Lemma 4.2

The projection

$$\mathcal{M} \xrightarrow{\pi_{\text{norm}}} (\mathbb{C}^*)^n \times \mathbb{C}^n, \quad (\underline{g}, \underline{w}; \underline{a}) \longmapsto (w_{1,A}, \dots, w_{n,A}; a_{1,1}, \dots, a_{n,1})$$

maps $\mathcal{M}_{\text{norm}}$ bijectively onto the affine variety V cut out by

$$\begin{aligned} a_{1,1} &= (w_{1,A} + \dots + w_{n,A}) + \left(\frac{1}{w_{1,A}} + \dots + \frac{1}{w_{n,A}} \right), \\ &\vdots \\ a_{n,1} &= w_{1,A} \cdots w_{n,A} + \frac{1}{w_{1,A} \cdots w_{n,A}}. \end{aligned} \tag{3}$$

Moreover, any point on V of multiplicity 1 is the image of a point on $\mathcal{M}_{\text{norm}}$ of multiplicity $2(n!)^2$.

Proof. The bijection follows directly from the conditions on $\mathcal{M}_{\text{norm}}$ and the equations of System (2): If $(\underline{g}, \underline{w}; \underline{a})$ is normalized, then by definition $\underline{g} = (1, \dots, 1)$ and $w_{1,B} = w_{1,A}^{-1}$. Additionally, the following parameters are uniquely determined by the following equations of System (2):

$$\begin{aligned} a_{0,1} &= g_A + g_B, & a_{0,2} &= g_A g_B, \\ a_{1,0} &= g_1 + \dots + g_n, & a_{1,2} &= g_A g_B (g_1 w_{1,A} w_{1,B} + \dots + g_n w_{n,A} w_{n,B}), \\ &\vdots & &\vdots \\ a_{n,0} &= g_1 \cdots g_n, & a_{n,2} &= g_A g_B g_1 w_{1,A} w_{1,B} \cdots g_n w_{n,A} w_{n,B}. \end{aligned}$$

The multiplicity follows from the fact that:

- any solution $(g_A, g_B; a_{0,j})$ to the two equations in the first row is of multiplicity 2 (see Example 3.2),
- the solution $(g_1, \dots, g_n; a_{i,0}) = (1, \dots, 1; \binom{n}{i})$ to the latter equations in the first column is of multiplicity $n!$,
- given $g_A = g_B = g_i = 1$, the solution $(w_{i,A}, w_{i,B}; a_{i,2}) = (1, \dots, 1; \binom{n}{i})$ to the latter equations in the second column is of multiplicity $n!$.

and from the fact that the multiplicity of the entire system equals the product of the multiplicities of the three smaller systems in our case [15, Proposition 1.29]. \square

Proposition 4.3

A normalized binding polynomial represents generically $2n!$ decoupled molecules, each of multiplicity $2(n!)^2$.

Proof. By Lemma 4.2, it suffices to show that System (3) has $2n!$ simple solutions for generic $\underline{a} = (a_{1,1}, \dots, a_{n,1}) \in \mathbb{C}^n$, or rather for $(a_{1,1}, \dots, a_{n,1}) \in \mathcal{U}$ for some Euclidean open subset $\mathcal{U} \subseteq \mathbb{C}^n$. For the sake of simplicity, we abbreviate $a_i := a_{i,1}$ and $w_i := w_{i,A}$ for $i = 1, \dots, n$. Next, we introduce n new variables μ_1, \dots, μ_n and consider the following equivalent system of $2n$ equations in the $2n$ variables

$\mu_1, \dots, \mu_n, w_1, \dots, w_n$:

$$\mu_1 = (w_1 + \dots + w_n) \quad (1)$$

$$\vdots \quad \vdots$$

$$\mu_n = w_1 \cdots w_n \quad (n)$$

$$a_1 - \mu_1 = \left(\frac{1}{w_1} + \dots + \frac{1}{w_n} \right) \quad (-1)$$

$$\vdots \quad \vdots$$

$$a_n - \mu_n = \frac{1}{w_1 \cdots w_n} \quad (-n)$$

Let \mathcal{N} be the variety cut out by the system above and let π_μ and π_a denote the three projections onto μ_i and a_i respectively.

$$\begin{array}{ccc} \mathcal{N} \subseteq (\mathbb{C}^*)^{|w_1, \dots, w_n|} \times \mathbb{C}^{|a_1, \dots, a_n|} \times \mathbb{C}^{|{\mu_1, \dots, \mu_n}|} & & \\ & \begin{array}{ccc} \downarrow \pi_a & & \searrow \pi_\mu \\ \mathbb{C}^{|a_1, \dots, a_n|} & \xleftarrow{\varphi} & \mathbb{C}^{|{\mu_1, \dots, \mu_n}|} \end{array} & \end{array}$$

We will construct a dominant (i.e. its image is Zariski dense), 2:1 rational map

$$\varphi : \mathbb{C}^n \dashrightarrow \mathbb{C}^n, \quad (w_1, \dots, w_n) \mapsto (a_1, \dots, a_n),$$

that maps $\underline{\mu}$ to the unique \underline{a} for which some \underline{w} exists such that $(\underline{w}; \underline{a}, \underline{\mu}) \in \mathcal{N}$. In short, the diagram above commutes. The image of the complement $\mathbb{C}^n \setminus \text{Disc}_x(x^n + \mu_1 x^{n-1} + \dots + \mu_n)$ will then contain an open set $\mathcal{U} \subseteq \mathbb{C}^n$, and for any $\underline{a} \in \mathcal{U}$ the system will have $2n!$ solutions: 2 solutions in $\underline{\mu}$, both outside the discriminant, and consequently also $n!$ solutions in \underline{w} for each $\underline{\mu}$.

To construct φ , observe that combining equations $(-n)$ and (n) and obtain:

$$a_n - \mu_n = \frac{1}{w_1 \cdots w_n} = \frac{1}{\mu_n},$$

which is equivalent to

$$\mu_n^2 - a_n \cdot \mu_n + 1 = 0 \quad \text{or} \quad a_n = \mu_n + \frac{1}{\mu_n}.$$

Moreover, multiplying equation $(-n)$ with $x_1 \cdots x_n$ yields

$$(a_{n-1} - \mu_{n-1}) \cdot \underbrace{x_1 \cdots x_n}_{\text{Eq.}(n)_{\mu_n}} = \underbrace{x_1 + \dots + x_n}_{\text{Eq.}(1)_{\mu_1}},$$

or, more generally, by multiplying Equation $(-i)$ with $x_1 \cdots x_n$:

$$(a_i - \mu_i) \cdot \mu_n = \mu_{n-i} \quad \text{or} \quad a_i = \mu_i - \frac{\mu_{n-i}}{\mu_n} \quad \text{for } i = 1, \dots, n-1.$$

Set

$$\begin{aligned} \varphi : \quad \mathbb{C}^n &\dashrightarrow \mathbb{C}^n, \\ (\mu_1, \dots, \mu_n) &\longmapsto \left(\mu_1 - \frac{\mu_{n-1}}{\mu_n}, \dots, \mu_{n-1} - \frac{\mu_1}{\mu_n}, \mu_n - \frac{1}{\mu_n} \right) \end{aligned}$$

By construction, φ is 2 : 1 and commutes with the projections π_μ, π_a . Moreover, it is dominant as its Jacobian,

$$J(\phi) = \begin{pmatrix} 1 & & & & -1 \\ & 1 & & -1 & \\ & & \ddots & & \\ & & & -1 & 1 \\ \frac{\mu_{n-1}}{\mu_n^2} & \frac{\mu_{n-2}}{\mu_n^2} & \dots & \frac{\mu_1}{\mu_n^2} & 1 + \frac{1}{\mu_n^2} \end{pmatrix}$$

is invertible at $(1, \dots, 1)$. □

4.2. A generic decoupled sites representation. In this subsection, we will infer from Proposition 4.3 the number of molecules a binding polynomial generically represents.

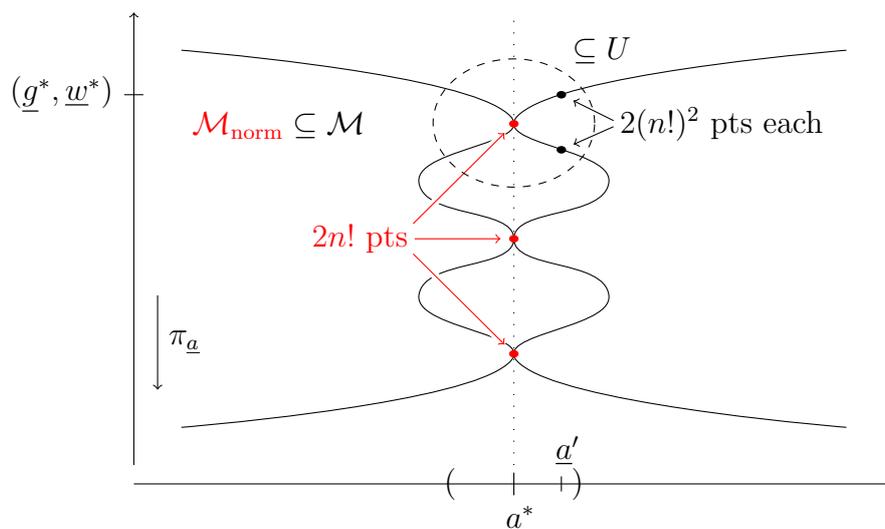


FIGURE 6. Perturbation of normalized binding polynomials.

Theorem 4.4

A binding polynomial of bidegree $(n, 2)$ represents generically $4(n!)^3$ decoupled molecules with $(n, 2)$ sites. These come in $2(n!)^2$ classes modulo the $S_n \times S_2$ action that corresponds to relabelling of the sites.

Proof. It suffices to show the claim in a euclidean open set. For that consider a generic normalized binding polynomial $\underline{a}^* \in \mathbb{C}^{(n+1)(2+1)-1}$. Proposition 4.3 states that there are $2n!$ solutions $(\underline{g}^*, \underline{w}^*; \underline{a}^*)$ to $f(\underline{g}, \underline{w}; \underline{a}^*) = 0$ of multiplicity $2(n!)^2$.

Applying Theorem 3.1 to each of the solutions, we obtain an open subset $U \subseteq X \times Y$, where $X := (\mathbb{C}^*)^{n+2} \times (\mathbb{C}^*)^{n-2}$, $Y := \mathbb{C}^{(n+1)(2+1)-1}$, such that

- for any $\underline{a}' \in \pi_{\underline{a}}(U)$, $U \cap X \times \{\underline{a}'\}$ has only isolated solutions of $f(\underline{g}, \underline{w}; \underline{a}') = 0$,
- the sum of the multiplicities of those isolated solutions is $4(n!)^3$.

It remains to show that U contains all isolated solutions of $f(\underline{g}, \underline{w}; \underline{a}') = 0$ for $\underline{a}' \in \pi_{\underline{a}}(U)$ and that the solutions are simple. Both follow from the fact that our parameters are exactly the constant terms, i.e. we can regard \mathcal{M} as the graph of the polynomial map

$$h : X \longrightarrow Y, \quad (\underline{g}, \underline{w}) \longmapsto \begin{pmatrix} g_1 + \dots + g_n \\ \vdots \\ \vdots \end{pmatrix},$$

so that $f(\underline{g}, \underline{w}; \underline{a}) = 0$ is equivalent to $h(\underline{g}, \underline{w}) = \underline{a}$. Fix $\underline{a}' \in \pi_{\underline{a}}(U)$ and a path

$$\phi : [0, 1] \rightarrow Y, \quad \phi(0) = \underline{a}^* \text{ and } \phi(1) = \underline{a}'.$$

To see that U contains all solutions to $f(\underline{g}, \underline{w}; \underline{a}') = 0$, observe that any solution $(g', w'; \underline{a}')$ has a solution path

$$z : (0, 1] \longrightarrow X \text{ with } z(1) = (g', w')$$

such that $f(z(t); \phi(t)) = 0$ for all $t \in (0, 1]$. As $\lim_{t \rightarrow 0} h(z(t)) = \lim_{t \rightarrow 0} \phi(t) = \underline{a}^*$ and h is polynomial, $\lim_{t \rightarrow 0} z(t)$ has to converge. Therefore $(\lim_{t \rightarrow 0} z(t), \underline{a}^*)$ is one of our $2n!$ solutions, which implies $(g', w'; \underline{a}') \in U$.

To see that solutions are generically simple note that a solution $(g', w'; \underline{a}')$ to $f(\underline{g}, \underline{w}; \underline{a}') = 0$ is singular if and only if the point (g', w') is a critical point of h . Hence, any solution in the following open set will be simple

$$\mathcal{U} := U \cap \pi_{\underline{a}}^{-1}(Y \setminus S),$$

where $S := \{h(g', w') \in \mathbb{C}^{\{a_{i,j}\}} \mid (g', w') \text{ critical point}\}$ consists of the images of all critical points of h . \square

5. FURTHER EXPERIMENTAL RESULTS

In this section, we provide some experimental results for $(n, 2)$ and beyond. For simplicity, we will use randomly chosen $\underline{a} \in \mathbb{C}^{(n+1)(2+1)-1}$. Moreover, we will also fix a choice of $g_{S_1}, \dots, g_{S_{n_1}}, g_{T_1}, \dots, g_{T_{n_2}}$ to factor out the natural $S_{n_1} \times S_{n_2}$ action on the roots of System (1), see Section 5.1.

All computations are done using one of the following three programs:

BERTINI[3]: A solver for polynomial equations using numerical algebraic geometry. It has built-in features for parallel path-tracking, which proved to be particularly useful for big examples.

GFAN[24] : A software package for computing Gröbner fans and tropical varieties. It features a new algorithm for computing mixed volumes using tropical homotopy methods [25].

SINGULAR[13]: A computer algebra system for polynomial computations, with special emphasis on commutative and non-commutative algebra, algebraic geometry, and singularity theory.

Code, tutorials and other auxiliary files for the computations will soon be made available under `software.mis.mpg.de`.

5.1. Explicit solutions for $(3, 2)$. Consider the following equations of System (2) for $n = 3$:

$$\begin{aligned}
 a_{1,1} &= g_A(g_1 w_{1,A} + g_2 w_{2,A} + g_3 w_{3,A}) + g_B(g_1 w_{1,B} + g_2 w_{2,B} + g_3 w_{3,B}), \\
 a_{2,1} &= g_A(g_1 g_2 w_{1,A} w_{2,A} + g_1 g_3 w_{1,A} w_{3,A} + g_2 g_3 w_{2,A} w_{3,A}) \\
 &\quad + g_B(g_1 g_2 w_{1,B} w_{2,B} + g_1 g_3 w_{1,B} w_{3,B} + g_2 g_3 w_{2,B} w_{3,B}), \\
 a_{3,1} &= g_A g_1 g_2 g_3 w_{1,A} w_{2,A} w_{3,A} + g_B g_1 g_2 g_3 w_{1,B} w_{2,B} w_{3,B}, \\
 a_{1,2} &= g_A g_B (g_1 w_{1,A} w_{1,B} + g_2 w_{2,A} w_{2,B} + g_3 w_{3,A} w_{3,B}), \\
 a_{2,2} &= g_A g_B (g_1 g_2 w_{1,A} w_{1,B} w_{2,A} w_{2,B} + g_1 g_3 w_{1,A} w_{1,B} w_{3,A} w_{3,B} + g_2 g_3 w_{2,A} w_{2,B} w_{3,A} w_{3,B}), \\
 a_{3,2} &= g_A g_B g_1 g_2 g_3 w_{1,A} w_{1,B} w_{2,A} w_{2,B} w_{3,A} w_{3,B}.
 \end{aligned}$$

Choosing

$$\begin{aligned}
 g_1 &= 2, & g_2 &= 3, & g_3 &= 5, & g_A &= 11, & g_B &= 13, \\
 a_{1,1} &= 71, & a_{2,1} &= 73, & a_{3,1} &= 79, & a_{1,2} &= 101, & a_{2,2} &= 103, & a_{3,2} &= 107,
 \end{aligned}$$

the system then simplifies to

$$\begin{aligned}
 71 &= 11(2w_{1,A} + 3w_{2,A} + 5w_{3,A}) + 13(2w_{1,B} + 3w_{2,B} + 5w_{3,B}), \\
 73 &= 11(6w_{1,A} w_{2,A} + 10w_{1,A} w_{3,A} + 15w_{2,A} w_{3,A}) \\
 &\quad + 13(6w_{1,B} w_{2,B} + 10w_{1,B} w_{3,B} + 15w_{2,B} w_{3,B}), \\
 79 &= 330w_{1,A} w_{2,A} w_{3,A} + 390w_{1,B} w_{2,B} w_{3,B}, \\
 101 &= 143(2w_{1,A} w_{1,B} + 3w_{2,A} w_{2,B} + 5w_{3,A} w_{3,B}), \\
 103 &= 143(6w_{1,A} w_{1,B} w_{2,A} w_{2,B} + 10w_{1,A} w_{1,B} w_{3,A} w_{3,B} + 15w_{2,A} w_{2,B} w_{3,A} w_{3,B}), \\
 107 &= 4290w_{1,A} w_{1,B} w_{2,A} w_{2,B} w_{3,A} w_{3,B}.
 \end{aligned}$$

Using BERTINI, we see that it has 72 roots, all of which are non-real and simple, 12 of which have strictly positive real component. Figure 7 shows a pair of complex conjugate solutions with lexicographically largest real part. The ordering on the variables is $w_{1,A}$, $w_{1,B}$, $w_{2,A}$, $w_{2,B}$, $w_{3,A}$, $w_{3,B}$, so that the first lines corresponds to the real and imaginary part of $w_{1,A}$.

```

0.733175658157242746563365475886e0 -0.525124949875722284087132912860e0
0.261871644858051814095937009460e0 -0.442937573433368024261374882297e0
0.175843060588207991285330446970e0 0.331651692189479173140340537837e0
0.691545419943605655254684211883e0 0.448271194339440245246537764856e0
0.154572711574605557323732099093e0 -0.483019502079697923238189339927e0
0.104413557544869576753196043395e0 0.326278707683474805854874986793e0

0.733175658157242746563365475898e0 0.525124949875722284087132912872e0
0.261871644858051814095937009460e0 0.442937573433368024261374882309e0
0.175843060588207991285330446974e0 -0.331651692189479173140340537843e0
0.691545419943605655254684211896e0 -0.448271194339440245246537764868e0
0.154572711574605557323732099096e0 0.483019502079697923238189339940e0
0.104413557544869576753196043398e0 -0.326278707683474805854874986799e0

```

FIGURE 7. Two complex conjugate solutions for (3, 2).

5.2. Mixed volumes. The Newton polytope of a polynomial is the convex hull of all exponent vectors of all monomials with non-zero coefficient. Given a polynomial system f_1, \dots, f_N in N variables, the *mixed volume* of their Newton polytopes is a number that equals the number of roots provided the non-zero coefficients are generic. This is known as the Bernstein-Khovanskii-Kushnirenko Theorem [5].

Figure 8 shows a table with the mixed volume for various (n_1, n_2) computed using GFAN. We see that the number for $(n_1, 1)$ and $(n_1, 2)$ corresponds with the theoretical results. Sadly, there is no apparent pattern for (n_1, n_2) with $n_2 > 2$.

Note that there exist criteria on so-called Newton-degeneracy which guarantee that the mixed volume equals the number of roots [23]. However, due to the high-dimensionality of the Newton polytopes, these were infeasible to verify for the cases of interest such as (4, 3).

	1	2	3	4	5	6
1	1	2	6	24	120	720
2		8	72	1 152	28 800	1 036 800
3			1 944	162 432	24 624 000	1 349 713 408
4				52 862 976	-	-
5					-	-
6						-

FIGURE 8. mixed volumes of the Newton polytopes of (1)

5.3. Counting solutions with multiplicity. Given a zero-dimensional polynomial ideal $I \subseteq \mathbb{C}[x]$, the dimension of $\mathbb{C}[x]/I$ as a \mathbb{C} -vector space equals the number of solutions counted with multiplicity. It can be easily read off any Gröbner basis, but computing the Gröbner basis itself is a highly challenging task [16, Section

1.8.5]. In Figure 8, red numbers mark all cases for which Gröbner bases were computeable in SINGULAR. The respective vector space dimensions (computed using the SINGULAR command `vdim`) all coincided with the mixed volume.

5.4. **Explicit solutions for $(5, 2)$ and $(4, 3)$.** For the cases $(5, 2)$ and $(4, 3)$, highlighted blue in Figure 8, we also tried to compute explicit roots using BERTINI. However, numerical instabilities arose in both cases during the computation, so that the roots computed are not complete.

For $(5, 2)$ we obtained 28737 roots, 63 short or 99.8% of the proven 28800 roots. For $(4, 3)$ we obtained 156966 roots, 5466 short or 97% of the conjectured 162432 roots.

6. OPEN QUESTIONS

We close with three open questions.

Question 6.1

What is the number of solutions for (n_1, n_2) ?

For binding polynomials of bidegree $(n, 1)$ and $(n, 2)$, the number of decoupled molecules is given by relatively simple expressions. Assuming that the mixed volumes of the Newton polytopes equals the number of solutions, Table 8 indicates a more complicated pattern in the number of decoupled molecules for $(n, 3)$. The smallest interesting example is the case $(4, 3)$ for which we conjecture that the number equals 162432.

Question 6.2

How many solutions with real, positive values for g_i and $w_{i,j}$ exist?

For univariate binding polynomials, the existence of complex roots suggests that the system does strongly interact and cannot be represented by a real decoupled system. In particular it is an indicator for “cooperativity” [27]. It is neither clear how this concept can be translated to decoupled molecules for two types of ligands nor which characteristic different decoupled molecules share. To develop an understanding, it would be helpful to determine the number of real, positive solutions for small examples.

Question 6.3

Find an algorithm to compute the minimal interaction energy that a molecule with prescribed binding polynomial has.

For univariate binding polynomials, a quantitative measure for “cooperativity” is mapping the polynomial to the minimal interaction energy which is required to generate it [26]. In more detail, the *norm* of a molecule is the product of all the *absolutes* of its interaction energies,

$$|M| = \prod |w_{i,j}|, \text{ where } |w| := \max(w, w^{-1}),$$

while the *norm* of a polynomial Φ is the minimal norm of all molecules that give rise to this polynomial. How can we calculate $|\Phi|$? It would be interesting to investigate whether the machinery that has been developed for Euclidean distance degree is applicable in our setting [14].

REFERENCES

- [1] G. S. Adair, A. V. Bock, and H. Field Jr. The hemoglobin system vi. the oxygen dissociation curve of hemoglobin. *Journal of Biological Chemistry*, 63(2):529–545, 1925.
- [2] J. Barcroft. The combinations of haemoglobin with oxygen and with carbon monoxide. ii. *Biochemical Journal*, 7(5):481, 1913.
- [3] D. J. Bates, J. D. Hauenstein, A. J. Sommese, and C. W. Wampler. Bertini: Software for numerical algebraic geometry. Available at bertini.nd.edu.
- [4] A. Ben-Naim. *Cooperativity and regulation in biochemical processes*. Springer Science & Business Media, 2001.
- [5] D. N. Bernstein. The number of roots of a system of equations. *Funkcional. Anal. i Priložen.*, 9(3):1–4, 1975.
- [6] C. Bohr, K. Hasselbalch, and A. Krogh. Ueber einen in biologischer Beziehung wichtigen Einfluss, den die Kohlensäurespannung des Blutes auf dessen Sauerstoffbindung übt. *Skandinavisches Archiv Für Physiologie*, 16(2):402–412, 1904.
- [7] W. E. Briggs. A new measure of cooperativity in protein-ligand binding. *Biophysical chemistry*, 18(1):67–71, 1983.
- [8] W. E. Briggs. Cooperativity and extrema of the hill slope for symmetric protein-ligand binding polynomials. *Journal of theoretical biology*, 108(1):77–83, 1984.
- [9] W. E. Briggs. The relationship between zeros and factors of binding polynomials and cooperativity in protein-ligand binding. *Journal of theoretical biology*, 114(4):605–614, 1985.
- [10] J.-P. Changeux and S. J. Edelstein. Allosteric mechanisms of signal transduction. *Science*, 308(5727):1424–1428, 2005.
- [11] Y. R. Cho, A. J. Maguire, A. C. Try, M. S. Westwell, P. Groves, and D. H. Williams. Cooperativity and anti-cooperativity between ligand binding and the dimerization of ristocetin a: asymmetry of a homodimer complex and implications for signal transduction. *Chemistry & biology*, 3(3):207–215, 1996.
- [12] P. R. Connelly, C. H. Robert, W. E. Briggs, and S. J. Gill. Analysis of zeros of binding polynomials for tetrameric hemoglobins. *Biophysical chemistry*, 24(3):295–309, 1986.
- [13] W. Decker, G.-M. Greuel, G. Pfister, and H. Schönemann. SINGULAR 4-1-0 — A computer algebra system for polynomial computations. Available at www.singular.uni-kl.de, 2016.
- [14] Jan Draisma, Emil Horobet, Giorgio Ottaviani, Bernd Sturmfels, and Rekha R. Thomas. The Euclidean distance degree of an algebraic variety. *Found. Comput. Math.*, 16(1):99–149, 2016.
- [15] D. Eisenbud and J. Harris. *3264 and all that—a second course in algebraic geometry*. Cambridge University Press, Cambridge, 2016.
- [16] Gert-Martin Greuel and Gerhard Pfister. *A SINGULAR introduction to commutative algebra*. Springer, Berlin, extended edition, 2008. With contributions by Olaf Bachmann, Christoph Lossen and Hans Schönemann, With 1 CD-ROM (Windows, Macintosh and UNIX).
- [17] P. S. Gutierrez, D. Monteoliva, and L. Diambra. Role of cooperative binding on noise expression. *Phys. Rev. E*, 80:011914, Jul 2009.
- [18] P. S. Gutierrez, D. Monteoliva, and L. Diambra. Cooperative binding of transcription factors promotes bimodal gene expression response. *PLOS ONE*, 7(9):1–13, 09 2012.

- [19] S. H. Ha and J. E. Ferrell. Thresholds and ultrasensitivity from negative cooperativity. *Science*, 352(6288):990–993, 2016.
- [20] K. A. Hasselbalch. Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebundenen Kohlensäure desselben, und die Sauerstoffbindung des Blutes als Funktion der Wasserstoffzahl. *Biochemische Zeitschrift.*, page 78: 112–144, 1917.
- [21] A. V. Hill. The combinations of haemoglobin with oxygen and with carbon monoxide. i. *Biochemical Journal*, 7(5):471, 1913.
- [22] T. L. Hill. *Cooperativity theory in biochemistry: steady-state and equilibrium systems*. Springer Science & Business Media, 1985.
- [23] Birkett Huber and Bernd Sturmfels. A polyhedral method for solving sparse polynomial systems. *Math. Comp.*, 64(212):1541–1555, 1995.
- [24] Anders N. Jensen. Gfan, a software system for Gröbner fans and tropical varieties. Available at home.imf.au.dk/jensen/software/gfan/gfan.html.
- [25] Anders Nedergaard Jensen. Tropical homotopy continuation, 2016.
- [26] J. W. R. Martini. A measure to quantify the degree of cooperativity in overall titration curves. *Journal of Theoretical Biology*, 2017.
- [27] J. W. R. Martini, M. Schlather, and S. Schütz. A model for carrier-mediated biological signal transduction based on equilibrium ligand binding theory. *Bulletin of Mathematical Biology*, 78(5):1039–1057, 2016.
- [28] J. W. R. Martini, M. Schlather, and G. M. Ullmann. The meaning of the decoupled sites representation in terms of statistical mechanics and stochastics. *MATCH Commun. Math. Comput. Chem*, 70(3):829–850, 2013.
- [29] J. W. R. Martini, M. Schlather, and G. M. Ullmann. On the interaction of different types of ligands binding to the same molecule part ii: systems with n to 2 and n to 3 binding sites. *Journal of Mathematical Chemistry*, pages 1–19, 2013.
- [30] J. W. R. Martini, M. Schlather, and G. M. Ullmann. On the interaction of two different types of ligands binding to the same molecule part i: basics and the transfer of the decoupled sites representation to systems with n and one binding sites. *Journal of Mathematical Chemistry*, 51(2):672–695, 2013.
- [31] J. W. R. Martini and G. M. Ullmann. A mathematical view on the decoupled sites representation. *Journal of Mathematical Biology*, 66(3):477–503, Feb 2013.
- [32] A. Onufriev, D. A. Case, and G. M. Ullmann. A novel view of ph titration in biomolecules. *Biochemistry*, 40(12):3413–3419, 2001.
- [33] A. Onufriev and G. M. Ullmann. Decomposing complex cooperative ligand binding into simple components: connections between microscopic and macroscopic models. *The Journal of Physical Chemistry B*, 108(30):11157–11169, 2004.
- [34] J. A. Schellman. Macromolecular binding. *Biopolymers*, 14(5):999–1018, 1975.
- [35] A. J. Sommese and C. W. Wampler, II. *The numerical solution of systems of polynomials*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.
- [36] M. I. Stefan and N. Le Novère. Cooperative binding. *PLOS Computational Biology*, 9(6):1–6, 06 2013.
- [37] J. Wyman and S. J. Gill. *Binding and linkage: functional chemistry of biological macromolecules*. University Science Books, 1990.

MAX-PLANCK-INSTITUT FÜR MATHEMATIK IN DEN NATURWISSENSCHAFTEN, INSELSTRASSE 22,
04103 LEIPZIG, GERMANY.

E-mail address: `yueren@mis.mpg.de`

URL: <http://personal-homepages.mis.mpg.de/yueren>

KWS SAAT SE¹, GRIMSEHLSTRASSE 31, 37574 EINBECK, GERMANY.

E-mail address: `johannes.martini@kws.com`

MAX-PLANCK-INSTITUT FÜR MATHEMATIK IN DEN NATURWISSENSCHAFTEN, INSELSTRASSE 22,
04103 LEIPZIG, GERMANY.

E-mail address: `jtorres@mis.mpg.de`

URL: <http://personal-homepages.mis.mpg.de/jtorres>

¹this research project is not associated with KWS SAAT SE