

A mathematical view on the decoupled sites representation

Johannes W. R. Martini · G. Matthias Ullmann

Received: 20 June 2011 / Revised: 14 November 2011 / Published online: 25 February 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract The decoupled sites representation (DSR) is a theoretical instrument which allows to regard complex pH titration curves of biomolecules with several interacting proton binding sites as composition of isolated, non-interacting sites, each with a standard Henderson–Hasselbalch titration curve. In this work, we present the mathematical framework in which the DSR is embedded and give mathematical proofs for several statements in the periphery of the DSR. These proofs also identify exceptions. To apply the DSR to any molecule, it is necessary to extend the set of binding energies from \mathbb{R} to a stripe within \mathbb{C} . An important observation in this context is that even positive interaction energies (repulsion) between the binding sites will not guarantee real binding energies in the decoupled system, at least if the molecule has more than four proton binding sites. Moreover, we show that for a given overall titration curve it is not only possible to find a corresponding system with an interaction energy of zero but with any arbitrary fix interaction energy. This result also effects practical work as it shows that for any given titration curve, there is an infinite number of corresponding hypothetical molecules. Furthermore, this implies that—using a common definition of cooperative binding on the level of interaction energies—a meaningful measure of cooperativity between the binding sites cannot be defined solely on the basis of the overall titration. Consequently, all measures of cooperativity based on

GMU was supported by the Deutsche Forschungsgemeinschaft through SFB 840 research project B2.

J. W. R. Martini (✉)
Institut für Mathematische Stochastik, Georg-August Universität Göttingen,
Göttingen, Germany
e-mail: jmartin2@uni-goettingen.de

G. M. Ullmann
Bioinformatik/Strukturbiologie, Universität Bayreuth, Bayreuth, Germany
e-mail: Matthias.Ullmann@uni-bayreuth.de

the overall binding curve do not measure the type of cooperativity commonly defined on the basis of interaction energies. Understanding the DSR mathematically provides the basis of transferring the DSR to biomolecules with different types of interacting ligands, such as protons and electrons, which play an important role within electron transport chains like in photosynthesis.

Keywords Decoupled sites representation · Protonation · Binding polynomial · Interaction energy · Binding energy · Ligand binding

Mathematics Subject Classification (2010) 08B99 · 92E99 · 26C05 · 26C15

1 Introduction

The investigation of pH-dependent average binding of protons to binding sites of a molecule in equilibrium is a classical field of chemistry. Proton binding equilibria are a key for understanding biological molecular mechanisms since proton binding leads also to a change of the charge distribution. Thus, it can affect the catalytic center of an enzyme, the affinity to the substrate (or to another type of ligand), and the tertiary structure of macromolecules such as proteins and nucleic acids (Garcia-Moreno 1995). Moreover, electron or proton transport chains in oxidation processes and photosynthesis can be described by binding properties of the carrier proteins (Becker et al. 2007; Medvedev and Stuchebrukhov 2006; Till et al. 2008). Certainly, the mathematical description of titration curves of individual sites of macromolecules is much more complicated than that of small molecules with only one binding site. The more complicated shape of the titration curve of macromolecules is a result of interaction between the different sites (Bombarda and Ullmann 2010; Onufriev and Ullmann 2004; Tanford and Kirkwood 1957). This interaction can lead to titration curves exhibiting strong deviations from the classical Henderson–Hasselbalch (HH) curves (Ackers et al. 1983; Bashford and Karplus 1991; Bombarda and Ullmann 2010; Onufriev and Ullmann 2004) given by

$$\langle x \rangle = \frac{10^{pK_a - pH}}{1 + 10^{pK_a - pH}}. \quad (1)$$

Thus, Eq. (1) has to be generalized adequately. Since the numerator and the denominator are polynomial functions of the ligand activity (of degree one), a generalization leads to the concept of binding polynomials (bp) of degree n describing the overall titration properties of a molecule with n proton binding sites (Cantor and Schimmel 1980; Schellman 1975; Wyman and Gill 1990). The decoupled sites representation (DSR) was developed to connect macroscopic and microscopic titration behavior and to find an easy expression for the complex shape of microscopic titration curves (Onufriev et al. 2001; Onufriev and Ullmann 2004). The main result is that for any given macromolecule with n proton binding sites and interaction between these sites, there exists a molecule with n binding sites with the same binding polynomial but without intrinsic interaction. Note that the “existence” refers to the mathematical point of view, in terms of a tuple of n binding energies and not in terms of a certain

chemical structure. Even though Onufriev et al. (2001) described how to calculate the corresponding decoupled system, the mathematical structure of the problem was not investigated deeply. This lack hampered further progress such as transferring the DSR to macromolecules with different types of ligands (e.g. protons and electrons). We close this gap by investigating the mapping of binding energies and interaction energies of a molecule to its binding polynomial. This mapping shows that all molecules sharing the same binding polynomial are elements of an algebraic variety. Due to the special structure of the equations defining the sub-variety consisting of the molecule without interaction, no methods of algebraic geometry are required to prove the existence¹ and uniqueness² of the decoupled system. We show that the DSR can be generalized. For a given binding polynomial, it is not only possible to find a system without interaction with this binding polynomial, but it is possible to find systems with any interaction energy t with the same binding polynomial as long as the interaction energy is the same between all sites. We call this extension of the theory the generalized DSR, which states that the DSR is not a special result of the “lack of interaction”, but a consequence of identical interaction energies. Moreover, this implies that—using the common definition of cooperative binding on the level of interaction energies—a meaningful measure for cooperativity cannot only be based on the data provided by the overall titration curve.

2 Basics of proton binding

2.1 Molecules with several proton binding sites

Let us regard a certain type of (bio)molecule M with n proton binding sites. The proton binding properties of the molecule can be characterized by n binding energies G_1^M, \dots, G_n^M and $\frac{n(n-1)}{2}$ pairwise interaction energies $W_{1,2}^M, \dots, W_{1,n}^M, \dots, W_{n-1,n}^M$, where $W_{i,j}^M$ is the interaction energy of the i th and j th proton binding site. As titration properties are determined by these energies, every other system of proton binding sites with identical binding and interaction energies is regarded as equal to M , as the average overall titration as well as the titration curve of every individual site are identical. Thus, every molecule can be identified with at least one element

$$M \in \mathbb{R}^{\frac{n(n+1)}{2}}. \quad (2)$$

(We will show later that it is necessary to extend the domain of energies from \mathbb{R} to a stripe within \mathbb{C} .) Moreover, there is no natural order of the binding sites within a molecule. Thus, one has to notice that one and the same molecule can be identified with several tuples which is illustrated by Example 1.

Example 1 Let $n = 3$. Then the tuple

$$(G_1, G_2, G_3, W_{1,2}, W_{1,3}, W_{2,3}) = (1, 2, 3, 1, 2, 3)$$

¹ To prove the existence means to show that the corresponding variety is not the empty set.

² The decoupled system is unique with respect to an equivalence relation.

and

$$(G_1, G_2, G_3, W_{1,2}, W_{1,3}, W_{2,3}) = (2, 3, 1, 3, 1, 2)$$

belong to the same molecule.

This is a result of the disorderliness of the binding sites: One can number the binding sites in any order. However, it is important that the interaction energies are permuted accordingly. This property motivates the definition of an equivalence relation.

Definition 1 Let

$$a = (G_1^a, G_2^a, \dots, G_n^a, \dots, W_{n-1,n}^a)$$

and

$$b = (G_1^b, G_2^b, \dots, G_n^b, \dots, W_{n-1,n}^b) \in \mathbb{R}^{\frac{n(n+1)}{2}}.$$

Then a is equivalent to b (Notation: $a \sim b$) if and only if there exists a permutation σ of $(1, \dots, n)$ such that

$$a = (G_{\sigma(1)}^b, G_{\sigma(2)}^b, \dots, G_{\sigma(n)}^b, W_{\sigma(1),\sigma(2)}^b, W_{\sigma(1),\sigma(3)}^b, \dots, W_{\sigma(n-1),\sigma(n)}^b). \tag{3}$$

This means that every molecule can be identified with exactly one element

$$M \in \mathbb{R}^m / \sim \tag{4}$$

with $m = \frac{n(n+1)}{2}$.

Having defined the set of molecules or more general systems, we investigate how an element M is mapped to its bp, which is an element of $\mathbb{R}[\Lambda]$ (the polynomial ring in one variable Λ and coefficients in \mathbb{R}).

2.2 The binding polynomial

The bp of a molecule M is defined by

$$\Phi(M) := \sum_{k \in K} \exp(-\beta G^k) \Lambda^{l(k)} \tag{5}$$

where k is a microstate and K is the set of all microstates³ (Schellman 1975; Wyman and Gill 1990). A microstate $k = (x_1^k, \dots, x_n^k)$ is an n -Tupel with $x_i^k \in \{0, 1\}$ describing the protonation state of the molecule by

$$x_i^k = 1 \iff \text{the } i \text{ th binding site is protonated in microstate } k.$$

³ The cardinality of K is given by $\#K = 2^n$, if each site can exist in two states: protonated and deprotonated.

Moreover,

$$l(k) := \sum_{i=1}^n x_i^k, \tag{6}$$

G^k is the standard Gibbs free energy of microstate k , $\beta \in \mathbb{R}$ is constant (depending on the temperature which is assumed constant) and Λ is the variable of the polynomial (Schellman 1975; Wyman and Gill 1990). For the sake of simplicity, in this work, we set $\beta := 1$, abandon all units and define

$$G^k := \sum_{i=1}^n x_i^k G_i + \sum_{i=1}^n \sum_{j>i}^n x_i^k x_j^k W_{i,j}, \tag{7}$$

where G_i denotes free binding energy of site i if all other sites are unoccupied (compare Onufriev et al. 2001; Schellman 1975; Wyman and Gill 1990).

With these simplifications Eq. (5) rewrites

$$\Phi(M) := \sum_{k \in K} \left[\prod_{i=1}^n \left(\exp(-x_i^k G_i^M) \prod_{i<j} \exp(-x_i^k x_j^k W_{i,j}^M) \right) \Lambda^{l(k)} \right]. \tag{8}$$

We see here that the map

$$\begin{aligned} \Phi : \mathbb{R}^m / \sim &\longrightarrow \mathbb{R}[\Lambda] \\ M &\mapsto \Phi(M) \end{aligned}$$

can be factored into $\Phi = \Phi_2 \circ \Phi_1$ with

$$\begin{aligned} \Phi_1 : \mathbb{R}^m / \sim &\longrightarrow \mathbb{R}^{+m} / \sim \\ (G_1, \dots, G_n, \dots, W_{n-1,n}) &\mapsto (\exp(-G_1), \dots, \exp(-G_n), \dots, \exp(-W_{n-1,n})) \end{aligned} \tag{9}$$

and

$$\begin{aligned} \Phi_2 : \mathbb{R}^{+m} / \sim &\longrightarrow \mathbb{R}[\Lambda] \\ (g_1, g_2, \dots, g_n, \dots, w_{n-1,n}) &\mapsto \sum_{k \in K} \left[\prod_{i=1}^n \left(g_i^{x_i^k} \prod_{i<j} w_{i,j}^{x_i^k x_j^k} \right) \Lambda^{l(k)} \right]. \end{aligned} \tag{10}$$

As notation we use

$$(g_1, \dots, g_n, \dots, w_{n-1,n}) := \Phi_1(G_1, \dots, G_n, \dots, W_{n-1,n})$$

to indicate that we are considering the image of Φ_1 . Additionally, we call

$$\exp(-G_i) \text{ and } \exp(-W_{i,j})$$

binding and interaction constant to underline that these values are not energies. Moreover, $\mathbb{R}[\Lambda]$ denotes the polynomial ring in the variable Λ and coefficients in \mathbb{R} . Note that Φ_1 is well defined as

$$a \sim b \iff \Phi_1(a) \sim \Phi_1(b).$$

Φ_1 being a bijection, we can work with its image values—that is every molecule M is described by $(g_1^M, \dots, g_n^M, \dots, w_{n-1,n}^M)$ —and concentrate on the investigation of Φ_2 : The map is well defined as for two representatives of the same equivalence class

$$a \sim b \implies \Phi_2(a) = \Phi_2(b).$$

Moreover, we can rewrite Φ_2 without using microstates $k \in K$ but with focus on the coefficients of $\Phi(M)$:

$$\begin{aligned} & \Phi_2(g_1, \dots, w_{n-1,n}) \\ &= \prod_{i \in N} \left(g_i \prod_{i < j} w_{i,j} \right) \Lambda^n + \sum_{p \in N} \left[\prod_{i \in N, i \neq p} \left(g_i \prod_{i < j, j \neq p} w_{i,j} \right) \right] \Lambda^{n-1} \\ &+ \sum_{\substack{(p_1, p_2) \in N^2 \\ p_1 \neq p_2}} \left[\prod_{\substack{p_1 \neq i \neq p_2 \\ i < j \\ p_1 \neq j \neq p_2}} \left(g_i \prod_{i < j} w_{i,j} \right) \right] \Lambda^{n-2} + \dots + \sum_{i \in N} g_i \Lambda + 1 \quad (11) \end{aligned}$$

with $N := \{1, \dots, n\}$. Thus, Φ_2 maps a system of several interacting protonation sites to its bp which determines the overall titration curve (Corollary 1).

2.3 Titration curves of a certain site and of overall proton binding

Let $M = (g_1, \dots, g_n, w_{1,n}, \dots, w_{n-1,n})$ be a molecule. Moreover, let g^k denote the free energy of a microstate $k \in K$

$$g^k := \prod_{i \in N} \left(g_i^{x_i^k} \prod_{i < j} w_{i,j}^{x_i^k x_j^k} \right).$$

Then the average protonation of site i is given by

$$\langle x_i \rangle = \frac{Z_i(M)}{\Phi(M)} \tag{12}$$

with

$$\begin{aligned} Z_i(M) &= \sum_{\{k \in K | x_i^k = 1\}} g^k \Lambda^{l(k)} \\ &= g_i \prod_{\substack{m \in N \\ m \neq i}} \left(g_m w_{m,i} \prod_{\substack{j \neq i \\ m < j}} w_{m,j} \right) \Lambda^n \\ &\quad + g_i \sum_{\substack{p \in N \\ p \neq i}} \left[\prod_{\substack{m \in N \\ p \neq m \neq i}} \left(g_m w_{m,i} \prod_{\substack{i \neq j \neq p \\ m < j}} w_{m,j} \right) \right] \Lambda^{n-1} + \dots + g_i \Lambda, \end{aligned} \tag{13}$$

and $\Phi(M)$ denoting the bp of the molecule. Equation (13) means that only those microstates in which site i is protonated define the polynomial $Z_i(M)$ (Cantor and Schimmel 1980; Onufriev et al. 2001; Schellman 1975; Wyman and Gill 1990). Consequently, the overall titration curve has the shape

$$\langle x \rangle = \frac{\sum_{i=1}^n Z_i(M)}{\Phi(M)}. \tag{14}$$

Proposition 1 *Let M be a molecule with*

$$\Phi(M) = a_n \Lambda^n + a_{n-1} \Lambda^{n-1} + a_{n-2} \Lambda^{n-2} + \dots + a_1 \Lambda + 1.$$

Then its overall titration curve is given by

$$\frac{na_n \Lambda^n + (n - 1)a_{n-1} \Lambda^{n-1} + (n - 2)a_{n-2} \Lambda^{n-2} + \dots + a_1 \Lambda}{\Phi(M)}.$$

Proof The p -th coefficient a_p of $P(\Lambda)$ is given by

$$a_p = \sum_{\{k \in K | l(k) = p\}} g^k.$$

We compare it to the p -th coefficient a_p^i of Z_i which is

$$a_p^i = \sum_{\substack{\{k \in K\} \\ l(k) = p, x_i^k = 1}} g^k.$$

Thus, every g^k with $l(k) = p$ is summand of p coefficients a_p^i . Consequently,

$$\sum_{i=1}^n a_p^i = p \sum_{\{k \in K | l(k)=p\}} g^k = p a_p.$$

□

Corollary 1 *The overall titration curve is determined by the bp.*

Proof This is a direct consequence of Proposition 1. □

3 The decoupled sites representation

3.1 The extension of the domain of energies

Onufriev et al. (2001) presented the main result of the DSR, saying that for every molecule $M = (g_1^M, \dots, w_{n-1,n}^M)$ there exists exactly one system $L = (g_1^L, \dots, g_n^L, 1, \dots, 1)$ such that

$$\Phi_2(M) = \Phi_2(L).$$

However, it is possible that this might require the use of complex energies with a non-zero imaginary part which is illustrated by Example 2.

Example 2 Let M be a molecule with two interacting protonation sites and

$$(g_1^M, g_2^M, w_{1,2}^M) = (1, 1, 2).$$

Another molecule $L = (g_1^L, g_2^L, 1)$ with the same bp solves, according to Eq. (11), the system

$$\begin{aligned} g_1^L g_2^L &= g_1^M g_2^M w_{1,2}^M = 2 \\ g_1^L + g_2^L &= g_1^M + g_2^M = 2. \end{aligned}$$

Solving these equations gives the unique solution $L = (g_1^L, g_2^L, 1) = (1 + i, 1 - i, 1)$ which is equal to $(1 - i, 1 + i, 1)$ as we are dealing with equivalence classes.

This statement can easily be generalized to all molecules with two (or more) binding sites:

Lemma 1 *A molecule described by $(g_1, g_2, w_{1,2}) \in \mathbb{R}^{+3}$ requires the use of complex binding constants with a non-zero imaginary part to be presented as decoupled system if and only if*

$$(g_1 + g_2)^2 < 4g_1 g_2 w_{1,2}.$$

Proof

$$\Phi_2(g_1, g_2, w_{1,2}) = g_1 g_2 w_{1,2} \Lambda^2 + (g_1 + g_2) \Lambda + 1$$

A system $(d, e, 1)$ without interaction has to solve the equations:

$$\begin{aligned} de &= g_1 g_2 w_{1,2} \\ d + e &= g_1 + g_2. \end{aligned}$$

Thus, d has to solve

$$d^2 - (g_1 + g_2)d + g_1 g_2 w_{1,2} = 0$$

which shows that $d \notin \mathbb{R}$ if and only if

$$(g_1 + g_2)^2 < 4g_1 g_2 w_{1,2}. \quad (15)$$

□

This example shows that if the DSR shall be valid for all molecules, it is necessary to allow $g_i, w_{i,j} \in \mathbb{C} \setminus \{0\} =: \mathbb{C}^*$. This set would be appropriate as it guarantees the existence of a decoupled system with the same bp, and additional bps with coefficients in $\mathbb{C} \setminus \mathbb{R}$ do not have to be considered. However, complex numbers with imaginary part $iy \neq 0$ pose a problem for physical interpretation as it might not be regarded as binding “energy”. A discussion of this phenomenon can be found in Sect. 6.

The following corollary illustrates that complex binding energies with non-zero imaginary part are an indicator for negative interaction energies in the original molecule in the case of two binding sites.

Corollary 2 *Let $(g_1, g_2, w_{1,2}) \in \mathbb{R}^{+3}$ be a molecule with two ligand binding sites and interaction constant $w_{1,2} \leq 1$. Then the binding constants of the corresponding decoupled system are real.*

Proof

$$w_{1,2} \leq 1 \implies g_1^2 + (2 - 4w_{1,2})g_1 g_2 + g_2^2 \geq (g_1 - g_2)^2 \geq 0.$$

This implies

$$(g_1 + g_2)^2 \geq 4g_1 g_2 w_{1,2}$$

which proves the statement due to Lemma 1. □

Remark 1 Numerical analysis in the work of Onufriev et al. (2001) led to the conjecture that repulsion ($w_{i,j} < 1 \forall i, j$) is sufficient for the binding energies of the corresponding decoupled system to be real. For the case of two binding sites this statement was shown in Corollary 2. However, this assumption is not always true:

Example 3 Let us regard the molecule

$$\begin{aligned} \Phi_1(M) &= (g_1, g_2, g_3, g_4, g_5, w_{1,2}, w_{1,3}, w_{1,4}, w_{1,5}, w_{2,3}, w_{2,4}, w_{2,5}, w_{3,4}, w_{3,5}, w_{4,5}) \\ &= \left(2, 2, 2, 2, 2, \frac{1023}{1024}, \frac{1023}{1024}, \frac{1}{1024}, \frac{1023}{1024}, \frac{1023}{1024}, \frac{1}{1024}, \frac{1}{1024}, \frac{1}{1024}, \frac{1}{1024}, \frac{1}{1024}, \frac{1}{1024} \right) \end{aligned}$$

with positive interaction energies (repulsion). However, its bp has two complex roots with non-zero imaginary part which can be shown easily by an investigation of its extremes. This example demonstrates that even if repulsion is assumed decoupling can require the use of complex binding constants (see Corollary 4).

Having illustrated the importance of an extended domain of energies, we have to adapt the definitions. In Eq. (10) the domain and image space of the map Φ_2 has to be changed to

$$\Phi_2 : \mathbb{C}^{*m} / \sim \longrightarrow \mathbb{C}[\Lambda] \tag{16}$$

where the definition of the equivalence relation “ \sim ” in Eq. (3) is not affected. However, one has to think about the appropriate domain of Φ_1 . As $\exp : \mathbb{C} \longrightarrow \mathbb{C}^*$ is not bijective it is necessary to change expression (9) to

$$\Phi_1 : D^m / \sim \longrightarrow \mathbb{C}^{*m} / \sim \tag{17}$$

with $D := \{x + iy | x \in \mathbb{R}, y \in [-\pi, \pi[] \subset \mathbb{C}$. Consequently, Eq. (4) rewrites

$$M \in D^m / \sim. \tag{18}$$

To simplify notation we will use

$$\mathbb{D} := D^m / \sim \text{ and } \mathbb{H} := \mathbb{C}^{*m} / \sim.$$

3.2 The decoupled sites representation

With this framework we can express the DSR as proposition.

Proposition 2 (The decoupled sites representation)

Let $M = (g_1^M, \dots, g_n^M, \dots, w_{n-1,n}^M) \in \mathbb{H}$ be a molecule. Then there exists a unique system $L = (g_1^L, \dots, g_n^L, 1, \dots, 1) \in \mathbb{H}$ such that

$$\Phi_2(M) = \Phi_2(L).$$

The proof of Proposition 2 is based on the properties of Φ_2 . The argumentation is the following: We define $O_n := \{(g_1, g_2, \dots, g_n, 1, \dots, 1) \in \mathbb{H}\}$, the set of all molecules without interaction, and show that Φ_2 is injective on O_n . Moreover, we show that

$$\Phi_2(O_n) = \Phi_2(\mathbb{H}) =: \text{Im}(\Phi_2),$$

i.e., the image set of the restriction is still the same. This means that for every $M \in \mathbb{H}$ there exists at least one element $L \in O_n$ with

$$\Phi_2(M) = \Phi_2(L).$$

Moreover, this means that

$$\Phi_{2,O_n} : O_n \longrightarrow \text{Im}(\Phi_2)$$

is also surjective to $\text{Im}(\Phi_2)$, thus bijective and consequently there exists an inverse map

$$\Phi_{2,O_n}^{-1} : \text{Im}(\Phi_2) \longrightarrow O_n.$$

This gives

$$L = \Phi_{2,O_n}^{-1} \circ \Phi_2(M). \tag{19}$$

Bijectivity of Φ_{2,O_n} guarantees existence of the inverse map and thus uniqueness of L . In the following, we will explain every statement required for the proof of Proposition 2 in detail. However before, we will analyze the properties of the unrestricted map Φ_2 . Let

$$P_1(\Lambda) := a_n \Lambda^n + a_{n-1} \Lambda^{n-1} + \dots + 1 \in \text{Im}(\Phi_2)$$

be a polynomial. Φ_2 is not injective as every preimage p of $P_1(\Lambda)$ has to fulfill a system of n equations, given by a_1, \dots, a_n , with $\frac{n(n+1)}{2} = m$ variables represented by the binding and interaction constants (see Eq. 11). This system of equations describes an affine algebraic variety of dimension larger than 0. Thus, a solution solving all equations simultaneously is not unique. However, if the domain is reduced by fixing some variables, a restriction of Φ_2 allows a one-to-one identification between $\text{Im}(\Phi_2)$ and the restricted domain. If all $w_{i,j}$ are fixed, for example $w_{i,j} = 1 \forall i, j \in N$, which means reducing the domain to O_n , the system of equations which has to be solved is reduced to:

$$a_n = \prod_{i \in N} g_i, \quad a_{n-1} = \sum_{k \in N} \prod_{\substack{i \in N \\ i \neq k}} g_i, \dots, \quad a_1 = \sum_{k \in N} g_k. \tag{20}$$

This new system (20) consists of n non-linear equations with n variables $\{g_i\}_{i \in N}$, and shows a certain type of symmetry since the role of g_i does not depend on i .

Remark 2 A solution $p \in \mathbb{C}^{*n}$ to system (20) can be regarded as an element of \mathbb{H} by extending it to

$$(g_1, \dots, g_n, 1, \dots, 1) \in \frac{\mathbb{C}^{*n} \times (1, \dots, 1)}{\sim} = O_n.$$

This means solving system (20) equals finding a preimage of $P_1(\Lambda)$ w.r.t. Φ_{2, O_n} .

A lot of important statements about properties of Φ_{2, O_n} including the results required for the proof of Proposition 2 are a consequence of the following essential consideration: The Fundamental Theorem of algebra allows to define the map

$$\Phi_3 : \text{Im}(\Phi_2) \longrightarrow O_n$$

from a polynomial to its roots. It is well defined due to the equivalence relation of O_n . As the set $\text{Im}(\Phi_2)$ is “normed“ by $a_0 = 1$, Φ_3 is bijective with inverse map

$$\begin{aligned} \Phi_3^{-1} : O_n &\longrightarrow \text{Im}(\Phi_2) \\ \Phi_3^{-1}(g_1, \dots, g_n, 1, \dots, 1) &= \prod_{i=1}^n (\Lambda - g_i) \cdot \prod_{i=1}^n \left(-\frac{1}{g_i}\right). \end{aligned} \tag{21}$$

The factor $\prod_{i=1}^n \left(-\frac{1}{g_i}\right)$ is required to ensure $a_0 = +1$. Moreover, we define the bijective function

$$\begin{aligned} \Phi_4 : O_n &\longrightarrow O_n \\ \Phi_4(g_1, \dots, g_n, 1, \dots, 1) &= \left(-\frac{1}{g_1}, \dots, -\frac{1}{g_n}, 1, \dots, 1\right). \end{aligned} \tag{22}$$

Lemma 2 $\Phi_{2, O_n} = \Phi_3^{-1} \circ \Phi_4$.

Proof Calculating $\Phi_3^{-1} \circ \Phi_4(g_1, \dots, g_n, 1, \dots, 1)$ gives Eq. (11) with $w_{i,j} = 1 \ \forall i, j$. □

Actually, this lemma is sufficient to prove Proposition 2, due to the following corollary. However, we will highlight the important facts once more.

Corollary 3 Φ_{2, O_n} is bijective.

Proof As $\Phi_{2, O_n} = \Phi_3^{-1} \circ \Phi_4$, it is a composition of two bijective maps and consequently bijective. □

Corollary 4 Let $P(\Lambda) := a_n \Lambda^n + a_{n-1} \Lambda^{n-1} + \dots + 1 \in \text{Im}(\Phi_2)$ be a polynomial and let $S = (g_1, \dots, g_n, 1, \dots, 1)$ be in O_n . Then

$$\begin{aligned} \Phi_2(S) &= P(\Lambda) \\ &\iff \end{aligned} \tag{23}$$

$\left(-\frac{1}{g_1}, \dots, -\frac{1}{g_n}\right)$ is a permutation of the roots (with multiplicity) of $P(\Lambda)$.

Proof $\Phi_{2,O_n} = \Phi_3^{-1} \circ \Phi_4 \iff \Phi_3 \circ \Phi_{2,O_n} = \Phi_4$ □

Corollary 5 $\Phi_{2,O_n}(O_n) = \text{Im}(\Phi_2)$.

Proof Every polynomial $P(\Lambda) \in \text{Im}(\Phi_2)$ has n roots $\Lambda_1, \dots, \Lambda_n$ (with multiplicity). According to Corollary 4 $(-\frac{1}{\Lambda_1}, \dots, -\frac{1}{\Lambda_n}, 1 \dots 1)$ is a preimage in O_n . □

Proof (of Proposition 2) The proof of Proposition 2 follows from the previous results, as already described. □

Remark 3 The use of the Fundamental Theorem of algebra for the definition of Φ_3 was permitted by the extension of the set of constants. Note that the statements of Lemma 2 and Corollary 3 are sufficient to prove Proposition 2. However, Corollary 4 is a useful tool we would like to highlight. Moreover, note that this presented proof includes the argumentation for the proof of Vieta’s formulas, which could also be used directly to deduce Corollary 4 and also to prove the DSR.

An important point is that the existence of a unique non-interacting system for every bp is not a result of the lack of interaction, but of the reduced number of variables in the system of equations (20). This means for every other fixed value of $w_{i,j} = t \forall i, j$ there is a system with identical bp. This statement is a generalization of the DSR.

Proposition 3 (Generalized DSR)

Let $M = (g_1^M, \dots, g_n^M, \dots, w_{n-1,n}^M) \in \mathbb{H}$ be a molecule. Then $\forall t \in \mathbb{C}^*$ there exists a unique system $L = (g_1^L, \dots, g_n^L, t, \dots, t) \in \mathbb{H}$ such that

$$\Phi_2(M) = \Phi_2(L).$$

Proof We define $O_n(t) := \{(g_1, \dots, g_n, \dots, w_{n-1,n}) \in \mathbb{H} | w_{i,j} = t \forall i, j\}$ the set of all molecules with interaction constants $w_{i,j} = t$. Furthermore, we define the bijective map

$$\begin{aligned} \Phi_{5,t} : \text{Im}(\Phi_2) &\longrightarrow \text{Im}(\Phi_2) \\ \Phi_{5,t}(a_n \Lambda^n + a_{n-1} \Lambda^{n-1} + \dots + a_2 \Lambda^2 + a_1 \Lambda + 1) & \\ = t^{\frac{n(n-1)}{2}} a_n \Lambda^n + t^{\frac{(n-1)(n-2)}{2}} a_{n-1} \Lambda^{n-1} + \dots + t a_2 \Lambda^2 + a_1 \Lambda + 1. & \end{aligned} \tag{24}$$

Thus, we receive for the restriction $\Phi_{2,O_n(t)}$ of Φ_2 on $O_n(t)$:

$$\Phi_{2,O_n(t)} = \Phi_{5,t} \circ \Phi_3^{-1} \circ \Phi_4. \tag{25}$$

The rest of the argumentation is analogue to the proof of Proposition 2. □

Remark 4 The generalized DSR described in Proposition 3 is a generalization of the DSR to any value t of all interaction constants $w_{i,j}$. However, all interaction constants are fixed to the same value. A further generalization to any value of $(w_{1,2}, w_{1,3}, \dots) = (t_1, t_2, \dots)$ with $t_i \neq t_j$ is also possible but the proof of Proposition 3 can not be transferred directly, as the equivalence relation on the set leads to impermutability

of the roots of the polynomial. This means the order of the roots is important, and it is not possible to factor out a power of t which is used to define $\Phi_{5,t}$. As fixing $(w_{1,2}, w_{1,3}, \dots) = (t_1, t_2, \dots)$ leads to a system of n equations and n variables one might assume that the existence of a solution is obvious, however this is not clear as the system does not consist of linear equations. Instead methods of algebraic geometry have to be used. We will present a proof for the case $n = 3$ later.

4 Special considerations of the case $n = 2$

At first we will give an example for Proposition 3 for $n = 2$.

Example 4 Let M be a molecule with two interacting binding sites described by $\Phi_1(M) = (a, b, c)$. Then there exists a system $(d, e, 4)$ with the same bp

$$\Phi_2(a, b, c) = abc\Lambda^2 + (a + b)\Lambda + 1.$$

A system $(d, e, 4)$ with the same bp has to solve the equations:

$$\begin{aligned} 4de &= abc \\ d + e &= a + b \end{aligned}$$

Thus, d has to solve

$$d^2 - (a + b)d + \frac{abc}{4} = 0.$$

For example, the system $(a, b, c) = (1, 1, 2)$ shares its bp

$$P(\Lambda) = 2\Lambda^2 + 2\Lambda + 1$$

with the system $(1 + \frac{1}{\sqrt{2}}, 1 - \frac{1}{\sqrt{2}}, 4)$.

Another possibility to receive this result is using Eq. (25) to see:

$$\Phi_2(a, b, c) = \Phi_2(d, e, t) = \Phi_{5,t} \circ \Phi_3^{-1} \circ \Phi_4(d, e, t)$$

which gives for $t = 4$ and $(a, b, c) = (1, 1, 2)$

$$\Phi_4^{-1} \circ \Phi_3 \circ \Phi_{5,4}^{-1} \circ \Phi_2(1, 1, 2) = (d, e, 4).$$

Moreover, it is not only possible to fix $w_{1,2}$ but a certain constant g_i to find a unique solution.

Proposition 4 *Let $\Phi_1(M) = (a, b, 1)$ be a system without intrinsic interaction. Then for every $d \in \mathbb{C}^* \setminus \{a + b\}$ there exists a system $L = (d, a + b - d, \frac{ab}{d(a+b-d)})$ which has the same bp. Moreover, all systems sharing the same bp are of this shape.*

Proof Calculating the bp of L proves the first result. The argumentation to prove the second statement is the following: Let L be a molecule possessing the same bp, then L has to solve the following equations:

$$g_1^L g_2^L w_{1,2}^L = ab$$

$$g_1^L + g_2^L = a + b.$$

Defining $d := g_1^L$ proves the second result. □

Remark 5 Note that in Proposition 4 it is not possible to find a system

$$L = (d, a + b - d, \frac{ab}{d(a + b - d)})$$

if

$$d = a + b,$$

as this leads to a division by zero. Moreover, there is no $t \in \mathbb{C}$ such that

$$e^t = 0 = a + b - d.$$

4.1 Average protonation curves of individual sites

In the following, we will dedicate our attention to the average protonation $\langle x_i \rangle$ of every site x_i in the case of two binding sites. Let $\Phi_1(M) = (a, b, c)$ be a molecule with two proton binding site. According to Sect. 2, the average protonation curves of site A and B are given by

$$\langle x_A \rangle = \frac{abc\Lambda^2 + a\Lambda}{abc\Lambda^2 + (a + b)\Lambda + 1}$$

$$\langle x_B \rangle = \frac{abc\Lambda^2 + b\Lambda}{abc\Lambda^2 + (a + b)\Lambda + 1}.$$

Onufriev et al. (2001) derived that it is always possible to represent every titration curve of a certain site as convex combination of the HH titration curves of the sites of the corresponding decoupled system. We will give a proof for this being true for nearly every case. However, this proof identifies exceptions, too.

Proposition 5 *Let $\Phi_1(M) = (a, b, 1)$ be a system without intrinsic interaction with $a \neq b$ and let $\langle x_i \rangle$ denote the average protonation of site $i \in \{A, B\}$. Moreover, let $\Phi_1(L) = (d, a + b - d, \frac{ab}{d(a + b - d)})$ be another system with the same bp ($d \neq a + b$) and let $j \in \{D, E\}$ denote the binding sites of L . Then there exists a unique $t = (\frac{d-b}{a-b}) \in \mathbb{C}$ such that*

$$\langle x_D \rangle = t\langle x_A \rangle + (1 - t)\langle x_B \rangle \tag{26}$$

and

$$\langle x_E \rangle = (1 - t)\langle x_A \rangle + t\langle x_B \rangle. \tag{27}$$

Proof The average protonation of the sites of system M can be written as

$$\begin{aligned} \langle x_A \rangle &= \frac{ab\Lambda^2 + a\Lambda}{ab\Lambda^2 + (a + b)\Lambda + 1} \\ \langle x_B \rangle &= \frac{ab\Lambda^2 + b\Lambda}{ab\Lambda^2 + (a + b)\Lambda + 1}. \end{aligned}$$

As the systems have the same bp the corresponding titration curves of the second system can be written as:

$$\begin{aligned} \langle x_D \rangle &= \frac{ab\Lambda^2 + d\Lambda}{ab\Lambda^2 + (a + b)\Lambda + 1} \\ \langle x_E \rangle &= \frac{ab\Lambda^2 + (a + b - d)\Lambda}{ab\Lambda^2 + (a + b)\Lambda + 1} \end{aligned}$$

Consequently, for $t = \frac{d-b}{a-b}$ the equalities (26) and (27) hold. Conversely, let t solve (26) and (27) then it is of form $t = \frac{d-b}{a-b}$ and consequently unique. \square

In this situation t is neither necessarily in $[0, 1] \subset \mathbb{R}$, nor in \mathbb{R} at all. This fact has to be emphasized as it shows that t can not be interpreted as a probability of protonation of the original sites at a certain pH. The importance of the restriction $a \neq b$ is highlighted by the next example.

Example 5 Let $(a, b, 1) = (2, 2, 1)$ be the decoupled system to $(d, e, w_{d,e}) = (1, 3, \frac{4}{3})$. Then it is impossible to present $\langle x_D \rangle$ as a convex combination of $\langle x_A \rangle$ and $\langle x_B \rangle$, as $\langle x_A \rangle = \langle x_B \rangle$ implies

$$t\langle x_A \rangle + (1 - t)\langle x_B \rangle = t\langle x_A \rangle + (1 - t)\langle x_A \rangle = \langle x_A \rangle \quad \forall t \in \mathbb{C}$$

and

$$\langle x_A \rangle \neq \langle x_D \rangle.$$

This can also be seen the following way: We look for $t_1, t_2 \in \mathbb{C}$ such that

$$t_1\langle x_A \rangle + t_2\langle x_B \rangle = \langle x_D \rangle.$$

In particular, this means the numerator of $t_1\langle x_A \rangle + t_2\langle x_B \rangle$ equals the numerator of $\langle x_D \rangle$ and consequently t_1, t_2 have to solve the equations

$$\begin{aligned} t_1 ab + t_2 ab &= dew_{d,e} = ab \\ t_1 a + t_2 b &= d. \end{aligned}$$

This can be rewritten

$$\underbrace{\begin{pmatrix} 1 & 1 \\ a & b \end{pmatrix}}_{:=A} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} 1 \\ d \end{pmatrix}$$

The first line means $t_2 = 1 - t_1$. In the case $a = b$ we receive $\det(A) = 0$ which means there is no solution if $d \neq a$ and an infinite number of solutions if $a = b = d$, as every $(t_1, t_2) = (t, 1 - t)$ for any $t \in \mathbb{C}$ solves the system.

5 Special considerations of the case n=3

The next proposition generalizes Proposition 3 even further for the case of three binding sites.

Proposition 6 *Let $\Phi_1(M) = (a, b, c, 1, 1, 1)$ be a molecule. Then for every $(t_1, t_2, t_3) \in \mathbb{C}^{*3}$ there exists a system $\Phi_1(L) = (g_1, g_2, g_3, t_1, t_2, t_3)$ such that*

$$\Phi_2(a, b, c, 1, 1, 1) = \Phi_2(g_1, g_2, g_3, t_1, t_2, t_3). \tag{28}$$

Proof Equation (28) means (g_1, g_2, g_3) is a solution to the system

$$\begin{aligned} f_1 : & \quad abc = t_1 t_2 t_3 g_1 g_2 g_3 \\ f_2 : & \quad ab + ac + bc = t_1 g_1 g_2 + t_2 g_1 g_3 + t_3 g_2 g_3 \\ f_3 : & \quad a + b + c = g_1 + g_2 + g_3 \end{aligned}$$

We will show that a solution exists, independently of the choice of (a, b, c) and (t_1, t_2, t_3) . To this end, we regard equations f_1, f_2, f_3 as polynomials in

$$\mathbb{C}[g_1, g_2, g_3, a, b, c, t_1, t_2, t_3].$$

We used the computational algebra system Magma to calculate the Gröbner basis (w.r.t. the lexicographic order: $g_1 > g_2 > g_3 > a > b > c > t_1 > t_2 > t_3$) of the corresponding ideal $\langle f_1, f_2, f_3 \rangle$. This Gröbner basis consists of 11 polynomials. The second elimination ideal is generated by the last polynomial which is of degree six in g_3 with constant term $a_0 \neq 0$ (as $a, b, c, t_1, t_2, t_3 \neq 0$). This means that for any choice of $a, b, c, t_i \in \mathbb{C}^*$ we will find six solutions of g_3 (with multiplicity). The Extension Theorem (Cox et al. 2008, p. 165) tells us that those partial solutions can be extended to solutions to the first elimination ideal if the leading coefficients of the generators (regarded as polynomials in g_2) of the first elimination ideal do not all vanish at the partial solution. Looking at the second polynomial, we see that the leading coefficient is t_1 . This means, in the present situation of $t_1 \neq 0$, all solutions can be extended. Regarding the first polynomial of the Gröbner basis shows that the leading coefficient is $1 \neq 0$. This means those solutions can be extended further to full solutions of the whole system. Obviously, this leads to six simultaneous solutions of equations f_1, f_2, f_3 (with multiplicity). For additional information about the Gröbner

basis, the Extension and the Elimination Theorem see Cox et al. (2005, 2008). The corresponding Magma code can be found in the appendix. \square

We will give an example.

Example 6 Let $\Phi_1(M) = (2, -\frac{1}{2}, 4, 1, 1, 1)$ be a system without interaction. We look for molecules $\Phi_1(L) = (d, e, f, 2, 4, 3)$ with the same bp and use the computer algebra system Maxima to calculate the solutions of the system

$$\begin{aligned} -\frac{1}{6} &= def \\ 5 &= 2de + 4df + 3ef \\ \frac{11}{2} &= d + e + f. \end{aligned}$$

The six solutions are $(d, e, f) \in$

$$\begin{aligned} &\{(5.313833028641072, -0.10698947820485, 0.29315667609982), \\ &(4.928390901432182, 0.62565997888068, -0.054051184028884), \\ &(0.60206626291414, 4.953815261044177, -0.055881066899869), \\ &(0.31463862460357, -0.10021762785637, 5.285578747628083), \\ &(-0.083397327599447, 5.199003322259136, 0.38439389576735), \\ &(-0.075531559612602, 0.4287288758266, 5.146802325581396)\}. \end{aligned}$$

All molecules described by the different solutions have the same bp and the same interaction energies but different binding energies. Note that the precision is necessary to see that the bps are equal.

A nice by-product of Corollary 4 is the following potential algorithm for the calculation of a molecule (d, e, f, t_1, t_2, t_3) with a given bp in a different way.

Proposition 7 *Let $\Phi_1(M) = (a, b, c, 1, 1, 1)$ be a molecule without interaction. Moreover, let $\Phi_1(L) = (d, e, f, t_1, t_2, t_3)$ be a molecule possessing the same bp. Then (d, e, f) is a fixed point of the following potential algorithm:*

$$\begin{aligned} d_1 &:= 1, & e_1 &:= 1, & f_1 &:= 1 \\ P_1(\Lambda) &:= \frac{abc}{t_1 t_2 t_3} \Lambda^3 + \left(\frac{ab + ac + bc}{t_1} + d_1 f_1 \left(1 - \frac{t_2}{t_1}\right) + e_1 f_1 \left(1 - \frac{t_3}{t_1}\right) \right) \\ \Lambda^2 + (a + b + c)\Lambda + 1(\Lambda_{i-1,1}, \Lambda_{i-1,2}, \Lambda_{i-1,3}) &:= \text{Roots of } P_{i-1} \\ (d_i, e_i, f_i) &:= \left(-\frac{1}{\Lambda_{i-1,1}}, -\frac{1}{\Lambda_{i-1,2}}, -\frac{1}{\Lambda_{i-1,3}} \right). \end{aligned}$$

Moreover, every fixed point (h, i, j) of this potential algorithm satisfies

$$\Phi_2(a, b, c, 1, 1, 1) = \Phi_2(h, i, j, t_1, t_2, t_3).$$

Proof (d, e, f) being a fixed point means that $(-\frac{1}{d}, -\frac{1}{e}, -\frac{1}{f})$ are the roots of

$$P(\Lambda) := \frac{abc}{t_1 t_2 t_3} \Lambda^3 + \left(\frac{ab + ac + bc}{t_1} + df \left(1 - \frac{t_2}{t_1}\right) + ef \left(1 - \frac{t_3}{t_1}\right) \right) \Lambda^2 + (a + b + c)\Lambda + 1.$$

According to Corollary 4, $(-\frac{1}{d}, -\frac{1}{e}, -\frac{1}{f})$ are the roots of the polynomial

$$Q(\Lambda) := def \Lambda + (de + df + ef)\Lambda + (d + e + f)\Lambda + 1.$$

To show that (d, e, f) is a fixed point, we have to show that $P(\Lambda) = Q(\Lambda)$ which means

$$\frac{abc}{t_1 t_2 t_3} = def \tag{29}$$

$$\left(\frac{ab + ac + bc}{t_1} + df \left(1 - \frac{t_2}{t_1}\right) + ef \left(1 - \frac{t_3}{t_1}\right) \right) = (de + df + ef) \tag{30}$$

$$(a + b + c) = (d + e + f). \tag{31}$$

Correctness of Eqs. (29)–(31) is a result of

$$\Phi_2(a, b, c, 1, 1, 1) = \Phi_2(d, e, f, t_1, t_2, t_3).$$

Let (h, i, j) be a fixed point. Then $(-\frac{1}{h}, -\frac{1}{i}, -\frac{1}{j})$ are the roots of the corresponding polynomial. This means (h, i, j) fulfills Eqs. (29)–(31) and consequently

$$\Phi_2(a, b, c, 1, 1, 1) = \Phi_2(h, i, j, t_1, t_2, t_3).$$

□

Remark 6 Two open questions concerning this potential algorithm are whether it always converges to a fixed point (attraction of a solution) and which of the six possible solutions will be found. The attraction of the solution is not obvious as a small perturbation of the coefficients may have a huge effect on the roots of the polynomial.

Example 7 We have implemented this potential algorithm using the computer algebra system Maxima to calculate a solution of Example 6. Let $\Phi_1(M) = (2, -\frac{1}{2}, 4, 1, 1, 1)$ be a system without interaction. We look for a system $\Phi_1(L) = (d, e, f, 2, 4, 3)$ with the same bp. The algorithm described in Proposition 7 gives for 1000 iteration steps:

$$(d, e, f) = (4.92839118225807, 0.62565999882172, -0.054051181079786).$$

This represents the second solution of Example 6.

These results draw the following picture: For a given bp $P(\Lambda)$ and interaction constants t_1, t_2, t_3 there are in general six different corresponding molecules possessing

the bp $P(\Lambda)$. If two interaction constants are equal e.g. $t_1 = t_2$, the role of g_2 and g_3 in the system of equations given by the coefficients and the map (Eq. 11) is identical. This means if $(g_1, g_2, g_3, w_1, w_2, w_3) = (d, e, f, t_1, t_1, t_3)$ is a solution, then (d, f, e, t_1, t_1, t_3) is a solution, too. Due to the equivalence relation (Eq. 3) both solutions are equivalent which means that there are not more than three different solutions. In the case of $t_1 = t_2 = t_3$ all solutions coincide, resulting in the uniqueness of the solution which is the statement of the (generalized) DSR.

5.1 Average protonation curves of individual sites

The next step is to investigate whether Proposition 5 can be extended to the situation of three binding sites. We show that this is possible in a certain way. Let there be two molecules $L = (a, b, c, 1, 1, 1)$ and $M = (d, e, f, w_1, w_2, w_3) \in \mathbb{H}$, sharing the same bp:

$$\Phi_2(L) = abc\Lambda^3 + (ab + ac + bc)\Lambda^2 + (a + b + c)\Lambda + 1.$$

Let A, B, C, D, E, F denote the corresponding binding sites. The average protonation of sites A, B, C are described by

$$\langle x_A \rangle = \frac{abc\Lambda^3 + (ab + ac)\Lambda^2 + a\Lambda + 1}{\Phi_2(L)} \tag{32}$$

$$\langle x_B \rangle = \frac{abc\Lambda^3 + (ab + bc)\Lambda^2 + b\Lambda + 1}{\Phi_2(L)} \tag{33}$$

$$\langle x_C \rangle = \frac{abc\Lambda^3 + (ac + bc)\Lambda^2 + c\Lambda + 1}{\Phi_2(L)} \tag{34}$$

We look for $t_{d,1}, t_{d,2}, t_{d,3}$ such that

$$\langle x_D \rangle = t_{d,1}\langle x_A \rangle + t_{d,2}\langle x_B \rangle + t_{d,3}\langle x_C \rangle \tag{35}$$

with

$$\langle x_D \rangle = \frac{abc\Lambda^3 + (dew_1 + dfw_2)\Lambda^2 + d\Lambda + 1}{\Phi_2(L)}. \tag{36}$$

Equality of the numerators means $(t_{d,1}, t_{d,2}, t_{d,3})$ has to solve the following system of linear equations:

$$\underbrace{\begin{pmatrix} 1 & 1 & 1 \\ a & b & c \\ ab + ac & ab + bc & ac + bc \end{pmatrix}}_{:=A} \begin{pmatrix} t_{d,1} \\ t_{d,2} \\ t_{d,3} \end{pmatrix} = \begin{pmatrix} 1 \\ d \\ dew_1 + dfw_2 \end{pmatrix}. \tag{37}$$

If $\det(A) \neq 0$, the solution $(t_{d,1}, t_{d,2}, t_{d,3})$ is unique. The same is true for $(t_{e,1}, t_{e,2}, t_{e,3})$ and $(t_{f,1}, t_{f,2}, t_{f,3})$. Calculating $(t_{i,1}, t_{i,2}, t_{i,3})$ for the situation of Example 6 shows that the solutions are not permutations of each other. However,

$$t_{d,i} + t_{e,i} + t_{f,i} = 1 \quad \forall i \in \{1, 2, 3\}.$$

For this reason, one might suggest that Proposition 5 can be generalized to the case $n = 3$.

Proposition 8 *Let $\Phi_1(M) = (a, b, c, 1, 1, 1)$ and $\Phi_1(L) = (d, e, f, w_1, w_2, w_3)$ have the same bp. Moreover, let*

$$\det(A) := \det \begin{pmatrix} 1 & 1 & 1 \\ a & b & c \\ ab + ac & ab + bc & ac + bc \end{pmatrix} \neq 0.$$

Then there exists a unique matrix

$$T := \begin{pmatrix} t_{d,1} & t_{d,2} & t_{d,3} \\ t_{e,1} & t_{e,2} & t_{e,3} \\ t_{f,1} & t_{f,2} & t_{f,3} \end{pmatrix}$$

such that

$$T \begin{pmatrix} \langle x_A \rangle \\ \langle x_B \rangle \\ \langle x_C \rangle \end{pmatrix} = \begin{pmatrix} \langle x_D \rangle \\ \langle x_E \rangle \\ \langle x_F \rangle \end{pmatrix}$$

and

$$t_{j,1} + t_{j,2} + t_{j,3} = 1 = t_{d,i} + t_{e,i} + t_{f,i} \quad \forall i \in \{1, 2, 3\}, \forall j \in \{d, e, f\}.$$

Proof As $\det(A)$ is assumed non-zero a unique solution to equation (37) exists. We can rewrite equation (37)

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & bc - ac & bc - ab \\ 0 & 0 & c - a - \frac{bc-ab}{c} \end{pmatrix} \begin{pmatrix} t_{d,1} \\ t_{d,3} \\ t_{d,2} \end{pmatrix} = \begin{pmatrix} 1 \\ dew_1 + dfw_2 - ab - ac \\ d - a - \frac{dew_1 + dfw_2 - ab - ac}{c} \end{pmatrix}.$$

This means

$$t_{d,2} = \frac{dc - dew_1 - dfw_2 + ab}{c^2 - ac - bc + ab}.$$

The same calculation for e and f gives

$$t_{e,2} = \frac{ec - dew_1 - efw_3 + ab}{c^2 - ac - bc + ab}$$

and

$$t_{f,2} = \frac{fc - dfw_2 - efw_3 + ab}{c^2 - ac - bc + ab}.$$

Thus,

$$t_{d,2} + t_{e,2} + t_{f,2} = 1.$$

Calculating $t_{j,3}$ and $t_{j,1}$ proves Proposition 8. \square

6 Biological consequences and interpretations of the described phenomena

6.1 Complex energies and cooperativity

If the DSR shall be applicable for any possible molecule, it is necessary to extend the set of binding energies from the real numbers \mathbb{R} to a larger subset of \mathbb{C} . In this work, we found two new results in this context. First, we showed that if the decoupled system shall be unique, the set of energies has to be restricted to D (Eq. 17). Secondly, we showed that repulsion in the original system is not sufficient to guarantee real binding energies in the corresponding decoupled system (Example 3). Consequently, the question arises how complex energies can be interpreted physically. A direct physical meaning of a non-zero imaginary part of a binding energy is not obvious to us. However, Corollary 2 shows that a non-zero imaginary part of the binding energies of the decoupled system L implies that the interaction energy in the original system $\Phi_1(M) = (a, b, w)$ is negative ($w > 1$). We use the term “positive-cooperative binding” for this phenomenon, according to the following definition.

Definition 2 Let $\Phi_1(M) = (g_1, \dots, g_n, w_{1,2}, \dots, w_{n-1,n})$ be a molecule. The ligand is said to bind

- positive-cooperatively to sites i and j if and only if $w_{i,j} > 1$.
- non-cooperatively to sites i and j if and only if $w_{i,j} = 1$.
- negative-cooperatively to sites i and j if and only if $w_{i,j} < 1$.

Thus, positive-cooperative binding to site i and j means that binding of the ligand to site i increases the affinity of the ligand to site j and vice versa. This definition is commonly used (Berg et al. 2007; Ben-Naim 2001; Ullmann and Ullmann 2011) and is adequate to characterize biochemical regulatory mechanisms by pointing out whether a ligand enhances or represses the binding to other sites.

In the case of two binding sites we have the following implications: According to Corollary 2, for a molecule $\Phi_1(M) = (a, b, w) \in \mathbb{R}^{+3}$ with two ligand binding sites and non-positive-cooperative binding ($w \leq 1$) the roots of the bp, and consequently the binding constants of the corresponding decoupled system will always be real. As the bp has only real and positive coefficients, its real roots have to be negative which means the binding constants are positive and the binding energies are real, too. Thus, complex binding energies with non-zero imaginary part of the decoupled system

indicate positive-cooperative binding of the ligand in the original system (Onufriev and Ullmann 2004). Conversely, real roots of the bp do not imply non-positive cooperation between the ligands which is illustrated by the following example.

Example 8 Let $\Phi_1(M) = (0.1, 2.9, \frac{100}{29})$ be a molecule with two ligand binding sites. The bp is given by

$$\Phi(M) = \Lambda^2 + 3\Lambda + 1.$$

The corresponding decoupled system is $(1.5 + \sqrt{\frac{5}{4}}, 1.5 - \sqrt{\frac{5}{4}}, 1)$. Obviously, the roots of the bp are real and the ligand binds positive-cooperatively to the molecule M .

However, it is not possible to regard truly complex binding energies of a decoupled system in general as indicator for positive-cooperative binding of the ligands which was illustrated in Example 3.

6.2 Consequences of the generalized DSR on the interpretation of overall titration curves

The statement of the generalized DSR is that for any given bp and any interaction constant $t \in \mathbb{C}^*$, we can find a hypothetical molecule L with $w_{i,j}^L = t \forall i, j$ possessing the given bp. This is a generalization as it shows that for a given titration curve, there are not only two systems—one with interaction and one without—sharing the same bp, but there is an infinite number of molecules sharing the same titration curve. A major consequence of this statement for experimental work is that using Definition 2, it is not possible to distinguish from the overall titration curve between positive-cooperative binding on one hand and negative- or non-cooperative binding on the other hand. This also implies that it is impossible to define a meaningful measure of cooperativity only based on the overall titration curve. Importantly, using Definition 2, a binding curve with two-step shape does not imply non-positive-cooperative ligand binding to the molecule which is demonstrated by the example $\Phi_1(M) = (2, 10^7, 10^3)$. In the following example three systems of qualitatively different ligand binding type are presented which share the same bp.

Example 9 The following molecules are examples of positive-, negative-, or non-cooperative ligand binding, but share the same overall titration curve:

$$(2, 1, 0.5), (1.5 + \sqrt{1.25}, 1.5 - \sqrt{1.25}, 1), (0.1, 2.9, \frac{100}{29}).$$

Remark 7 Note that in literature several measures of cooperativity can be found which are only based on the overall titration curve. The most famous one is the Hill coefficient. However, it is obvious that a concept of cooperativity based only on the data provided by the overall titration curve has to differ from the presented Definition 2. For this reason, we think the use of different terms for the types of cooperativity is necessary. Moreover, we think a precise definition of the concept of cooperativity, which

is measured by the Hill coefficient, on the level of interaction energies is required to make this concept clear and to distinguish between both concepts.

6.3 Identified constraints of statements in the periphery of the DSR and their physical interpretation

In Proposition 4, we showed that for a given bp

$$\Phi_2(a, b, 1) = a_2\Lambda^2 + a_1\Lambda + 1$$

and any binding constant $d \in \mathbb{C}^* \setminus \{a+b\}$, we can find a molecule $\Phi_1(M) = (d, e, w)$ possessing this bp. The exception $d = a + b = a_1$ can be interpreted the following way: If both molecules share the same bp, the equation

$$a + b = d + e$$

is valid. This means the sum of the exponential of the negative binding energies has to be equal. Thus, if d is changed, this equation will allow to adapt e accordingly. If d comes close to $a + b$ and thus e close to zero, the distribution of the binding energies on the two binding sites of the corresponding molecule will become extremely asymmetric. A small absolute value of the binding constant e has to be compensated by a great interaction constant w . The exception $d = a + b$ corresponds to the limit case of $e = 0$ ($G_e = \infty$) and $|w| = \infty$ ($W = -\infty$). It can be interpreted physically as molecule with only one ligand binding site since an infinite amount of energy is required to bind the ligand to the second binding site. In Proposition 5, we mentioned the exception that the presentation of the titration curve of an individual site of the original molecule as linear combination of the titration curves of the individual sites of the decoupled system will not be possible if the binding energies of the corresponding decoupled system are identical. This means the presentation is not possible in the case of an absolutely symmetric decoupled system (in terms of equally distributed binding energies). Analogously to the exception in Proposition 4, this situation can be regarded as a limit case. If both binding sites have the same individual titration curve, the linear combination will be identical for any weight. In the following, we will investigate which properties a molecule has to have to possess a symmetric decoupled system in the case of two binding sites.

Proposition 9 *Let $\Phi_1(M) = (a, b, w)$ be a molecule. The corresponding decoupled system has the shape $(d, d, 1)$ if and only if*

$$d = 0.5a + 0.5b$$

which is equivalent to

$$\langle x_D \rangle = 0.5\langle x_A \rangle + 0.5\langle x_B \rangle.$$

Proof Let (a, b, w) and $(d, d, 1)$ share the same bp $\Phi(M)$. Then $abw = d^2$, $a + b = 2d$, and

$$\Phi(M)\langle x_D \rangle = d^2\Lambda^2 + d\Lambda = abw\Lambda^2 + 0.5(a + b)\Lambda = \Phi(M)(0.5\langle x_A \rangle + 0.5\langle x_B \rangle).$$

Conversely, let (a, b, w) and $(d, e, 1)$ share the same bp and let

$$\langle x_D \rangle = 0.5\langle x_A \rangle + 0.5\langle x_B \rangle.$$

Then $d = 0.5a + 0.5b$. As the molecules share the same bp this gives $d + e = a + b$ and thus $d = e$. \square

This proposition can be generalized to molecules with three binding sites which describes the exception of Proposition 8.

7 Conclusions and outlook

In this paper, we presented the mathematical framework in which the DSR is embedded and which allowed to prove its consistence with standard binding theory (Proposition 2). However, to use it as a “general law” we showed that it is necessary to extend the set of binding energies from \mathbb{R} to a larger subset of \mathbb{C} (Lemma 1). We found that, contradictorily to a former conjecture, repulsion ($w_{i,j} < 1 \forall i, j$) of the ligands is not sufficient to guarantee real roots of the binding polynomial. This was illustrated by Example 3. Moreover, we showed that the DSR is the result of the special structure of a system of polynomial equations with the binding energies of the decoupled system as variables and generalized its statement (Proposition 3). The fact that for any given tuple $(t, \dots, t) \in \mathbb{C}^{*n}$ of interaction constants and any bp there exists a molecule with these interaction constants and this bp shows that it is absolutely impossible to deduce the binding or interaction energies unambiguously from a given overall titration curve (Propositions 3 and 6). Moreover, our discussion underlines that the term cooperativity is used for different concepts in literature. Using a common definition of cooperativity, it is impossible to define a meaningful measure based only on the overall titration curve. We consider the mathematical nature of the set \mathbb{H} derived from \mathbb{C}^m and the equivalence relation “ \sim ” as well as the connection to the polynomial ring $\mathbb{C}[\Lambda]$ to be of interest for further investigation. The maps within the polynomial ring and within \mathbb{H} turned out to be a useful tool to calculate molecules sharing the same bp (Eq. 25). However, the next major step would be to use methods from algebraic geometry to transfer the DSR to bps in $\mathbb{C}[\Lambda_1, \Lambda_2]$ which describe the binding behavior of molecules with different types of ligands such as protons and electrons. This topic is of large interest as proteins of electron transport chains or photosynthesis are often located in membranes with pH-gradient. Applying the DSR to those molecules might help understanding the biophysical nature of these essential processes which are the basis of life.

Acknowledgments First of all, we thank Martin Schlather for comments on the original manuscript. Moreover, we thank Tobias Dorsch for discussing ideas.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix

We used the following short Magma code for the proof of Proposition 6:

```
R<x1,x2,x3,t1,t2,t3,a,b,c> := PolynomialRing(Rationals(), 9);
eqns := [x1*x2*x3*t1*t2*t3-a*b*c,
x1*x2*t1 + t2*x1*x3 + t3*x2*x3 - a*b-a*c- c*b,
x1 + x2 + x3 - c-a-b];
I := ideal<R | eqns>;
GB:=GroebnerBasis(I);
#GB;
GB[11];
GB[10];
GB[2];
GB[1];
```

The algorithm described in Proposition 7 was implemented the following way:

```
a:2;
b:-0.5;
c:4;
w1:2;
w2:4;
w3:3;

a3:(a*b*c/(w1*w2*w3));
a1:(a+b+c);

d:1;
e:1;
f:1;

block(for p: 1 thru 1000 do (
block(a2:(a*b+a*c+b*c)/(w1)+(1-w2/w1)*d*f+(1-w3/w1)*e*f,
g(x):= a3*x^3 + a2*x^2 + a1*x+1,
[l[1],l[2],l[3]]:allroots(g(x)),
h:ev(x,l[1]),
i:ev(x,l[2]),
j:ev(x,l[3]),
d:-1/h,
e:-1/i,
f:-1/j
)))
;
```

References

- Ackers GK, Shea MA, Smith FR (1983) Free energy coupling within macromolecules: the chemical work of ligand binding at the individual sites in co-operative systems. *J Mol Biol* 170:223–242
- Bashford D, Karplus M (1991) Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. *J Phys Chem* 95(23):9556–9561

- Becker T, Ullmann RT, Ullmann GM (2007) Simulation of the electron transfer between the tetraheme subunit and the special pair of the photosynthetic reaction center using a microstate description. *J Phys Chem B* 111(11):2957–2968
- Ben-Naim A (2001) *Cooperativity and Regulation in Biochemical Processes*. Kluwer Academic/Plenum Publishers, New York
- Berg J, Tymoczko J, Stryer L (2007) *Biochemistry*. W.H. Freeman, New York
- Bombarda E, Ullmann GM (2010) pH-Dependent pKa values in proteins—a theoretical analysis of protonation energies with practical consequences for enzymatic reactions. *J Phys Chem B* 114(5):1994–2003. PMID: 20088566
- Cantor CR, Schimmel PR (1980) *Biophysical Chemistry. Part III. The Behavior of Biological Macromolecules*, 1st edn. W. H. Freeman, New York
- Cox D, Little J, O’Shea D (2005) *Using Algebraic Geometry*, 2nd edn. Springer, Berlin
- Cox DA, Little J, O’Shea D (2008) *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 3rd edn. Springer, New York
- Garcia-Moreno BE (1995) Probing structural and physical basis of protein energetics linked to protons and salt. *Methods Enzymol* 259:512–538
- Medvedev E, Stuchebrukhov A (2006) Kinetics of proton diffusion in the regimes of fast and slow exchange between the membrane surface and the bulk solution. *J Math Biol* 52:209–234. doi:[10.1007/s00285-005-0354-2](https://doi.org/10.1007/s00285-005-0354-2)
- Onufriev A, Case DA, Ullmann GM (2001) A novel view of pH titration in biomolecules. *Biochemistry* 40(12):3413–3419
- Onufriev A, Ullmann GM (2004) Decomposing complex cooperative ligand binding into simple components: Connections between microscopic and macroscopic models. *J Phys Chem B* 108(30):11157–11169
- Schellman JA (1975) Macromolecular binding. *Biopolymers* 14:999–1018
- Tanford C, Kirkwood JG (1957) Theory of protein titration curves. I. General equations for impenetrable spheres. *J Am Chem Soc* 79(20):5333–5339
- Till MS, Essigke T, Becker T, Ullmann GM (2008) Simulating the proton transfer in gramicidin A by a sequential dynamical monte carlo method. *J Phys Chem B* 112(42):13401–13410. PMID: 18826179
- Ullmann RT, Ullmann GM (2011) Coupling of protonation, reduction and conformational change in azurin from *Pseudomonas aeruginosa* investigated with free energy measures of cooperativity. *J Phys Chem B* 115:10346–10359
- Wyman J, Gill SJ (1990) *Binding and Linkage: Functional Chemistry of Biological Macromolecules*. University Science Books, Mill Valley