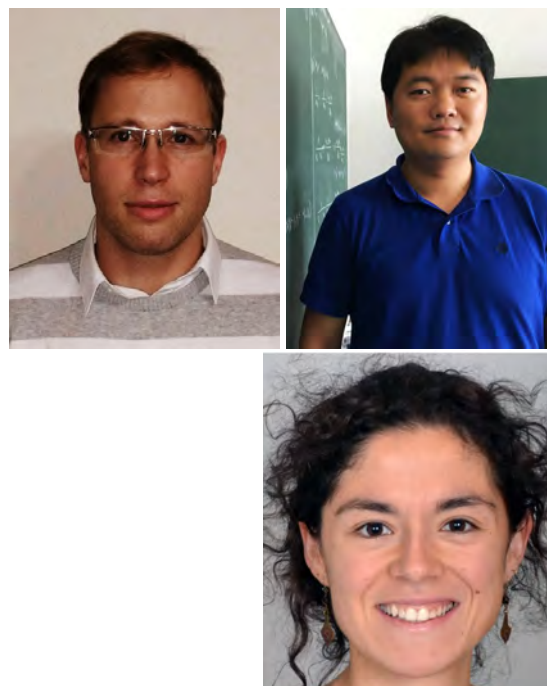


Über die Anzahl entkoppelter Moleküle mit demselben Bindepolynom

J. W. R. Martini (KWS SAAT SE^a)
 Y. Ren (Max-Planck-Institut MIS, Leipzig)
 J. Torres (Max-Planck-Institut MIS, Leipzig)

johannes.martini@kws.com
 yueren@mis.mpg.de
 jtorres@mis.mpg.de

^adieses Projekt ist nicht mit KWS SAAT SE affiliert



Vorwort

Dies ist ein kurzer Bericht über den Inhalt der Artikel [8, 9] und vor allem [13]. Für genauere Details verweisen wir auf diese Artikel. Begleitmaterial zu den Computereperimenten ist auf <https://software.mis.mpg.de> zu finden.

Einführung

Ein Ligand ist ein Stoff der reversibel an spezifische Bindestellen eines Biomoleküls binden kann. Dieses Biomolekül kann mehrere Bindestellen für unterschiedliche Liganden besitzen. Ein prominentes Beispiel hierfür ist Hämoglobin (siehe Abb. 1) welches vier Bindestellen für Sauerstoff, sowie eine weitere für 2,3-Bisphosphoglycerat besitzt. Letzteres moduliert hierbei die Affinität der vier Sauerstoffbindestellen.

Ein gängiges Modell, um Gleichgewichte und stationäre Zustände solcher Systeme zu beschreiben kommt vom sog. großkanonischen Ensemble der statistischen Mechanik [14]. Deren Partitionsfunktion, in unserem Kontext auch als *Bindungspolynom* bekannt, ist der Nenner der rationalen Funktion, welches die durchschnittliche Anzahl an belegten Bindestellen in Relation zur Ligandenaktivität beschreibt. Die Nullstellen dieses Polynoms spielen eine wichtige Rolle in der Charakterisierung des Bindungsverhaltens des Systems [4].

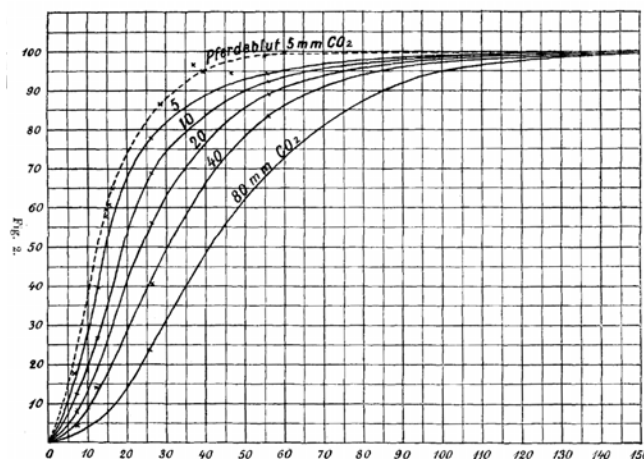


Abbildung 1: Sauerstoffkonzentration im Blut in Relation zur Sauerstoff- und Kohlenstoffdioxidkonzentration in der Luft, aus einer historischen Arbeit von Bohr, Hasselbalch und Krogh [3]

Der Einfachheit halber beschränken wir uns auf Systeme mit zwei Liganden, die wir T und S nennen. Wir sagen, ein Molekül M hat (n_1, n_2) Bindestellen, falls es n_1 Bindestellen für T und n_2 Bindestellen für S besitzt. Markieren wir diese mit T_1, \dots, T_{n_1} und S_1, \dots, S_{n_2} , so wird M beschrieben durch einen Punkt

$$M = \left(g_{T_1}, \dots, g_{T_{n_1}}, g_{S_1}, \dots, g_{S_{n_2}} \right)_{(WP)_{P \subseteq \{T_i, S_j\}, |P|=2}}$$

$$\in (\mathbb{C}^*)^{n_1+n_2} \times (\mathbb{C}^*)^{\binom{n_1+n_2}{2}}$$

wobei (siehe Abb. 2):

- g_{T_i}, g_{S_j} in Zusammenhang mit den Bindungsenergien an den Stellen T_i, S_j stehen,
- w_P in Zusammenhang mit der Wechselwirkung zwischen den beiden Stellen in P steht.

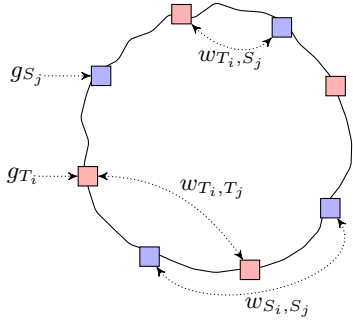


Abbildung 2: Ein Molekül mit (4,4) Bindestellen.

Für ein solches Molekül M ist das Bindungspolynom p_M ein bivariates Polynom vom Grad (n_1, n_2) ,

$$p_M(X, Y) = \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} a_{i,j} X^i Y^j,$$

dessen Koeffizienten $a_{i,j}$ wiederum selbst polynomiell von M abhängig sind (siehe Gleichungen (*) und (**)).

Normalerweise sind nur reellwertige, positive M von Interesse. Allerdings lässt sich im Fall von nur einem Liganden, jedes Molekül einzigartig durch ein komplexes entkoppeltes Molekül ohne Interaktion ($w_P = 1$ für alle P) repräsentieren [11]. “Repräsentieren” bedeutet hier, dass beide Moleküle dasselbe Bindungspolynom besitzen. Bei zwei verschiedenen Typen von Liganden bedeutet “entkoppelt”, dass keine Wechselwirkung zwischen den Bindestellen vom gleichen Typ existiert, genauer gesagt $w_P = 1$ für $P \subseteq \{T_1, \dots, T_{n_1}\}$ und $P \subseteq \{S_1, \dots, S_{n_2}\}$.

Wir beschäftigen uns mit der Frage, wie viele entkoppelte Moleküle sich in Systemen mit zwei Liganden ein Bindungspolynom teilen. Wir leiten explizite Formeln für Moleküle mit $(n, 1)$ und $(n, 2)$ Bindestellen her und untersuchen grössere Moleküle mittels Methoden der Computeralgebra.

Moleküle mit $(n, 1)$ Bindestellen

Sei M ein Molekül mit $(n, 1)$ Bindestellen: n Stellen für den ersten Liganden (als $1, \dots, n$ markiert) und eine Stelle für den zweiten Liganden (als A markiert) (siehe Abb. 3).

Dann ist das Bindungspolynom von M ein bivariates Polynom vom Grad $(n, 1)$:

$$p_M(X, Y) = \sum_{i=0}^n \sum_{j=0}^1 a_{i,j} X^i Y^j.$$

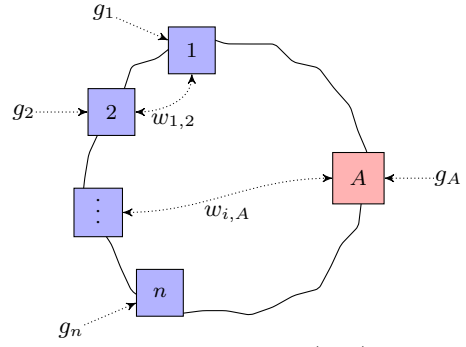


Abbildung 3: Ein Molekül M mit $(n, 1)$ Bindestellen. Ist M entkoppelt, so ist $w_{i,j} = 1 \forall i, j \in \{1, \dots, n\}$ (aber nicht notwendigerweise $w_{i,A} = 1$).

Ist nun M entkoppelt, oder suchen wir ein entkoppeltes Molekül M mit dem obigen Bindungspolynom, so müssen dessen Bindungsenergien g_1, \dots, g_n, g_A und Wechselwirkungen $w_{i,A}$ zwischen den Bindestellen unterschiedlichen Typs folgendes algebraisches System lösen:

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

Wir betrachten (*) als ein parametrisiertes System polynomieller Gleichungen, mit Parametern $a_{i,j}$, $(i, j) \neq (0, 0)$, und Variablen $g_i, g_A, w_{i,A}$.

Das Gleichungssystem (*) ist symmetrisch unter der folgenden $S_n \times S_1$ -Wirkung, was einer Ummarkierung der Bindestellen entspricht:

$$\begin{aligned} (\sigma, 1) \cdot (g_1, \dots, g_n, g_A, w_{1,A}, \dots, w_{n,A}) \\ = (g_{\sigma(1)}, \dots, g_{\sigma(n)}, g_A, w_{\sigma(1),A}, \dots, w_{\sigma(n),A}). \end{aligned}$$

Durch Umstellen und Anwendung von Vieta’s Formeln sieht man, dass (*) für generische Parameter $a_{i,j}$ immer $(n!)^2$ Lösungen besitzt. Diese kommen in $n!$ Orbits unter der $S_n \times S_1$ Wirkung, weswegen es generisch $n!$ verschiedene (unmarkierte) Moleküle gibt, die ein gegebenes Polynom als Bindungspolynom haben [8].

Ein simpler Trick, um die Gruppenwirkung für das Lösen konkreter Instanzen aus (*) herauszuteilen, ist eine Wahl der g_1, \dots, g_n, g_A fest vorzugeben. So kann man, wenn man zum Beispiel das System für zufällig gewählte Parameter lösen will, einfach $a_{i,1}$ für $i > 1$ und g_1, \dots, g_n, g_A zufällig wählen, anstatt alle $a_{i,j}$ zufällig zu wählen.

Moleküle mit $(n, 2)$ Bindestellen

Sei M ein Molekül mit $(n, 2)$ Bindestellen: $1, \dots, n$ für den ersten Ligand und A, B für den zweiten (siehe Abb. 4).

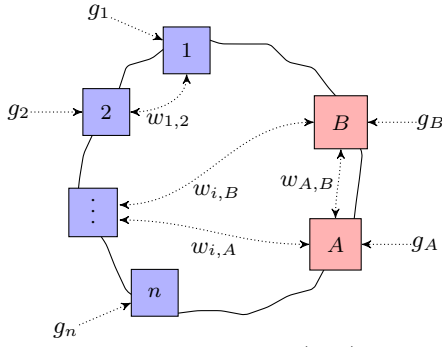


Abbildung 4: Ein Molekül M mit $(n, 2)$ Bindestellen. Ist M entkoppelt, so ist $w_{i,j} = 1 = w_{A,B}$ für alle i, j .

Dann ist das Bindungspolynom von M ein bivariates Polynom vom Grad $(n, 2)$:

$$p_M(X, Y) = \sum_{i=0}^n \sum_{j=0}^2 a_{i,j} X^i Y^j,$$

Ist M entkoppelt, so sind die Koeffizienten gegeben durch:

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

Betrachten wir in dem konkreten Fall $n = 3$ ein zufällig gewähltes Polynom wie das folgende (beachte, dass g_i, g_A, g_B fest vorgegeben sind um die natürliche $S_3 \times S_2$ Wirkung auszuteilen):

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

so liefert uns BERTINI [1] 72 Lösungen, wobei keine davon reell ist (siehe Abb. 5 für die BERTINI Eingabe).

```

CONFIG
% refine endpoints to 20 digits
SharpenDigits: 20;

% used to classify sing vs nonsing
CondNumThreshold: 1e15;

% track paths more accurately
ODEPredictor: 2;
TrackTolBeforeEG: 1e-8;
TrackTolDuringEG: 1e-8;
FinalTol: 1e-10;

% values at infinity
SecurityMaxNorm: 1e8;
MaxNorm: 1e8;
END;
INPUT
variable_group w1,w2,w3,w4,w5,w6;
function f1,f2,f3,f4,f5,f6;

f1=(11*(2*w1+3*w3+5*w5)+13*(2*w2+3*w4+5*w6)-71)
/70;
f2=(11*(6*w1*w3+10*w1*w5+15*w3*w5)+13*(6*w2*w4+10*
w2*w6+15*w4*w6)-73)/70;
f3=(330*w1*w3*w5+390*w2*w4*w6-79)/300;
f4=(143*(2*w1*w2+3*w3*w4+5*w5*w6)-101)/150;
f5=(143*(6*w1*w2*w3*w4+10*w1*w2*w5*w6+15*w3*w4*w5*
w6)-103)/120;
f6=(4290*w1*w2*w3*w4*w5*w6-107)/1000;
END;

```

Abbildung 5: BERTINI Eingabe für ein beliebig gewähltes Polynom mit $(3, 2)$ Bindestellen. Beachte die spezielle Konfiguration, um numerischen Fehlern vorzubeugen.

Konzentrieren wir uns in der Nullstellenmenge \mathcal{M} des Systems (**) auf sog. normierte Moleküle, d.h. Moleküle mit

$$g_i = g_A = g_B = 1 \text{ und } w_{i,A} \cdot w_{i,B} = 1,$$

so kann man beweisen, dass diese eine Multiplizität von $2n!$ besitzen und Verzweigungspunkte mit Verzweigungsgrad $(2n!)^2$ von folgender Projektion sind:

$$\begin{aligned} \mathcal{M} &\subseteq \mathbb{C}^{\{a_{i,j}\}} \times (\mathbb{C}^*)^{\{g_i, g_A, g_B\}} \times (\mathbb{C}^*)^{\{w_{i,A}, w_{i,B}\}} \\ &\quad \downarrow \\ &(\mathbb{C}^*)^{\{g_i, g_A, g_B\}} \times (\mathbb{C}^*)^{\{w_{i,A}, w_{i,B}\}}. \end{aligned}$$

Demnach gibt es zu einem generisches Bindungspolynom vom Bigrad $(n, 2)$ exakt $4(n!)^3$ entkoppelte Moleküle, bzw. $2(n!)^2$ Orbits unter der $S_n \times S_2$ Wirkung [13]. Für den Fall $n = 3$ macht das 864 Lösungen oder 72 Orbits unter der $S_3 \times S_2$ Wirkung.

Experimentelle Ergebnisse

Um eine obere Schranke für die Anzahl der Moleküle mit dem gleichen generischen Bindungspolynom von Grad (n, m) modulo der natürlichen $S_n \times S_m$ Wirkung zu bestimmen, betrachten wir die gemischten Volumina der jeweiligen Gleichungssysteme. Diese stimmen nach Bernstein's Theorem mit den Anzahl der Lösungen überein, falls die Koeffizienten generisch sind [2].

Abb. 6 listet diese tabellarisch für kleine (n, m) , o.B.d.A. $n \geq m$. Diese wurden mit GFAN [6] berechnet, welches einen neuen Algorithmus basierend

auf tropischen Homotopiemethoden benutzt [7]. Zuerst erkennen wir, dass für $(n, 1)$ und $(n, 2)$ das gemischte Volumen mit den bewiesenen Werten in [8, 13] übereinstimmt.

	1	2	3	4	5	6
1	1	2	6	24	120	720
2		8	72	1152	28800	1036800
3			1944	162432	24624000	1349713408
4				52862976	-	-
5					-	-
6						-

Abbildung 6: gemischte Volumina für kleine (n, m)

Als nächstes benutzen wir für die kleineren Fälle Gröbnerbasen um die Anzahl der Lösungen symbolisch zu bestimmen. Die roten Zahlen heben die mit SINGULAR [5] erfolgreich berechneten Fälle hervor. Hierdurch wird erstmals für den Fall $(3, 3)$ die Anzahl der Lösungen bestimmt und bestätigt, dass diese ebenfalls mit dem gemischten Volumen übereinstimmt.

Schließlich wurde für die beiden nächstgrößeren, blau gekennzeichneten Fälle versucht die Anzahl der Lösungen numerisch mittels BERTINI [1] zu berechnen. Dies stellte sich als extrem rechenintensiv und numerisch schwierig heraus. Der Fall $(3, 4)$ benötigte umgerechnet 6 CPU Jahre (Debian server with Intel Xeon E7-8837, 2.67GHz), und bereits der Fall $(5, 2)$ bedurfte einer speziellen Konfiguration, um numerischen Instabilitäten vorzubeugen (siehe Abb. 5). Letzteres war uns in beiden Rechnungen leider nicht komplett möglich wegen eventuell nicht alle Lösungen gefunden wurden:

Für $(5, 2)$ erhielten wir 28737 Lösungen, 63 weniger als oder 99.8% der bewiesenen 28800 Lösungen. Für $(4, 3)$ erhielten wir 156966 Lösungen, 5466 weniger als oder 97% der durch die gemischten Volumina vermuteten 162432 Lösungen.

Offene Fragen

(1) Für Bindungspolynome vom Grad $(n, 1)$ und $(n, 2)$ ist die Zahl dazugehöriger entkoppelter Moleküle durch einfache Formeln beschreibbar. Nehmen wir an, dass die gemischten Volumina des Gleichungssystems und die Anzahl der Lösungen übereinstimmen, zeichnet Abbildung 6 ein komplizierteres Muster für die Zahl entkoppelter Moleküle vom Grad $(n, 3)$. So enthält beispielsweise die vermutete Anzahl entkoppelter Moleküle für den Fall $(3, 4)$ (162432) den Primfaktor 47. Die Anzahl entkoppelter Moleküle wäre hier also nicht einfach ein Produkt von Fakultäten der Grade.

(2) Für univariate Bindungspolynome wird die Existenz von nicht-reellen Nullstellen als Indikator für Kooperativität angesehen [10, 12]. Es ist weder klar, wie dieses Konzept auf Moleküle mit zwei Typen von Liganden übertragen werden könnte, noch welche gemeinsamen Eigenschaften verschiedene entkoppelte Moleküle teilen. Um ein Verständnis hierfür zu entwickeln, wäre es beispielsweise nützlich, die Zahl reeller, positiver Lösungen für kleine Fälle zu berechnen.

Literatur

- [1] D. J. Bates, J. D. Hauenstein, A. J. Sommese, and C. W. Wampler. Bertini: Software for numerical algebraic geometry. Available at bertini.nd.edu.
- [2] D. N. Bernstein. The number of roots of a system of equations. *Funkcional. Anal. i Priložen.*, 9(3):1–4, 1975.
- [3] 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.
- [4] W. E. Briggs. The relationship between zeros and factors of binding polynomials and cooperativity in protein-ligand binding. *J Theor Biol*, 114(4):605–614, 1985.
- [5] 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.
- [6] Anders N. Jensen. Gfan, a software system for Gröbner fans and tropical varieties. Available at home.imf.au.dk/jensen/software/gfan/gfan.html.
- [7] Anders Nedergaard Jensen. Tropical homotopy continuation, 2016.
- [8] 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 site *J Math Chem*, 51(2):672–695, 2013.
- [9] 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 *J Math Chem*, 51(2):696–714, 2013.
- [10] J. W. R. Martini, L. Diambra and M. Habeck. Cooperative binding: a multiple personality. *J Math Biol*, 72(7): 1747–1774, 2016.
- [11] A. Onufriev, D. A. Case and G. M. Ullmann A novel view of pH titration in biomolecules. *Biochem*, 40(12):3413–3419, 2001.
- [12] A. Onufriev, D. A. Case and G. M. Ullmann. Decomposing complex cooperative ligand binding into simple components: connections between microscopic and macroscopic models. *J Phys Chem B*, 108 (30):11157–11169, 2004.
- [13] Y. Ren, J. W. R. Martini, and J. Torres. Decoupled molecules with binding polynomials of bidegree $(n, 2)$. arXiv:1711.06865, Submitted.
- [14] J. A. Schellman. Macromolecular binding. *Biopolymers*, 14 (5): 999–1018, 1975.