

3.7.2023: TREFFPUNKT WIRKSTOFFENTWICKLUNG

„Mehr Logik weniger Statistik? Logische Regeln in kleinen Datensätzen finden.“

**Zuse-Institut Berlin (ZIB)
PD Dr. Marcus Weber,
Computational Molecular Design**



Lipinski's Regel besagt, dass ein oral wirksames Medikament im Allgemeinen **nicht mehr als einen Verstoß** gegen die folgenden Kriterien aufweist:

Nicht mehr als 5 Wasserstoffbrückendonoren

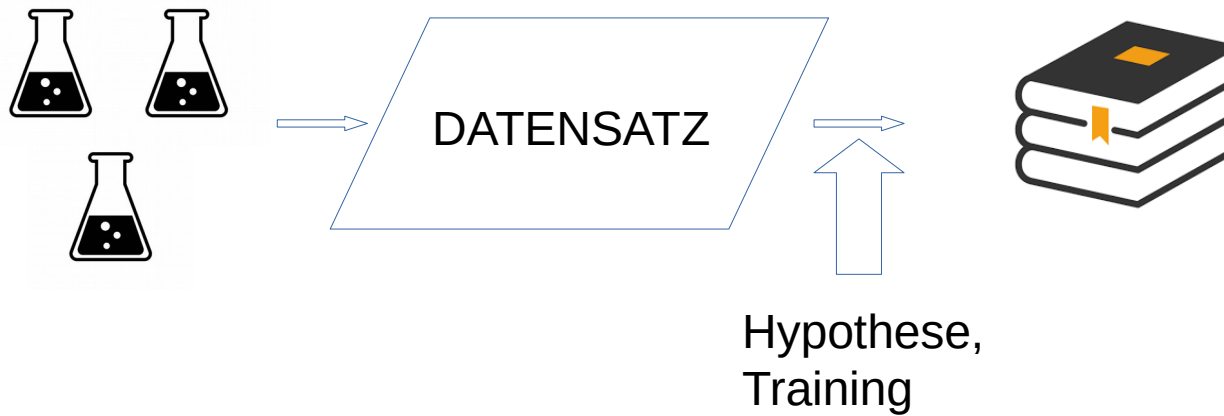
Nicht mehr als 10 Wasserstoffbrückenbindungsakzeptoren

Eine Molekülmasse von weniger als 500 Dalton

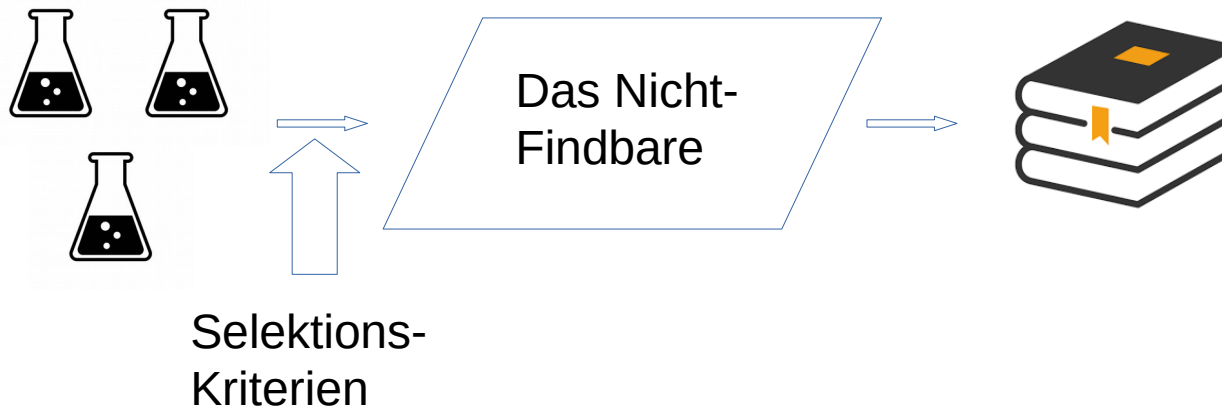
Ein berechneter Oktanol-Wasser-Verteilungskoeffizient,
der 5 nicht überschreitet

Quelle: Wikipedia

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (March 2001). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings". *Advanced Drug Delivery Reviews*. 46 (1–3): 3–26.



Muster, die ich in einem Datensatz oft wiederfinde, offenbaren eine Regel

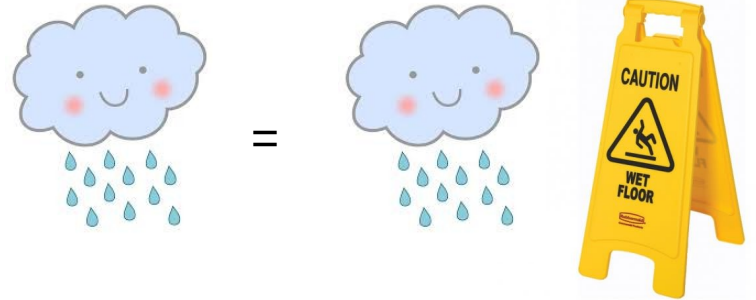


Muster, die ich nicht finden kann, offenbaren eine Regel

◀
Wenn es regnet, dann wird die Erde nass

Wenn es regnet, dann wird die Erde nass

$r = r * n$



* bedeutet „UND“

Wenn es regnet, dann wird die Erde nass

$r = r * n$

$r + r = r * n + r$

* bedeutet „UND“

+ bedeutet „ENTWEDER ODER“

Wenn es regnet, dann wird die Erde nass

$r = r * n$

$r + r = r * n + r$

$0 = r * n + r$

* bedeutet „UND“

+ bedeutet „ENTWEDER ODER“

0 bedeutet „LEERE AUSWAHL“

Wenn es regnet, dann wird die Erde nass

$$r = r * n$$

$$r + r = r * n + r$$

$$0 = r * n + r$$

$$0 = r * (n + 1)$$

* bedeutet „UND“

+ bedeutet „ENTWEDER ODER“

0 bedeutet „LEERE AUSWAHL“

1 bedeutet „ALLES AUSWÄHLEN“

Wenn es regnet, dann wird die Erde nass

$$r = r * n$$

$$r + r = r * n + r$$

$$0 = r * n + r$$

$$0 = r * (n + 1)$$



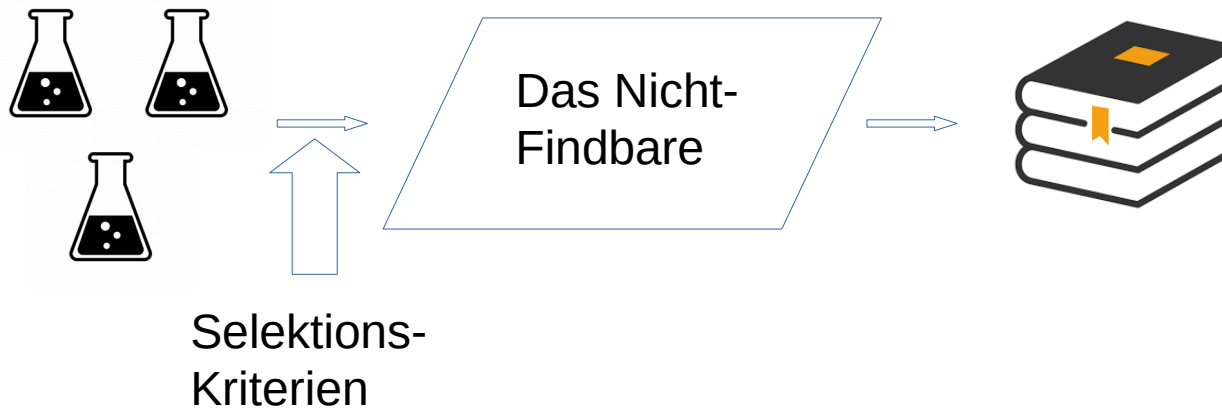
* bedeutet „UND“

+ bedeutet „ENTWEDER ODER“

0 bedeutet „LEERE AUSWAHL“

1 bedeutet „ALLES AUSWÄHLEN“

Nicht Findbar: Es regnet und die Erde wird nicht nass.



Muster, die ich nicht finden kann, offenbaren eine Regel



Lipinski's Regel besagt, dass ein oral wirksames Medikament im Allgemeinen **nur maximal eines** der folgenden Kriterien aufweist:

Mehr als 5 Wasserstoffbrückendonoren (A)

Mehr als 10 Wasserstoffbrückenbindungsakzeptoren (B)

Eine Molekülmasse von mehr als 500 Dalton (C)

Ein berechneter Oktanol-Wasser-Verteilungskoeffizient,
der 5 überschreitet (D)

$AB=0$, $AC=0$, $AD=0$, $BC=0$, $BD=0$, $CD=0$

{AB, AC, AD, BC, BD, CD, ...}

Addition von Regeln:

$$AB + AC = 0 \quad A(B+C)=0$$

Multiplikation mit Bedingungen:

$$AB * AC = 0 \quad ABC=0$$

$$AB * E = 0$$

Das Nicht-Findbare

Das „Nicht-Findbare“ ist ein Ideal eines Booleschen Polynomrings

{AB, AC, AD, BC, BD, CD, ...}

Das Nicht-
Findbare



Output bisher...

```
1 -> a^2 + a
2 -> A*a
3 -> A^2 + A
4 -> P*a
5 -> I*a + a
6 -> A*p
7 -> A*I + A*P + A*h
8 -> O*a + a*p
9 -> A*O + A*P
10 -> p^2 + p
11 -> h^2 + h
12 -> P*h + P
13 -> P^2 + P
14 -> I*p + h*p + a*h + a*p + T + P + h + p + A
15 -> I*h + A*h + I + h + A + 1
16 -> I*P + I*h + P*p + A*P + A*h + I + P + h + A + 1
17 -> I^2 + I
18 -> a^o + a*p
19 -> T*a + a*p
20 -> A^o + A*P
21 -> A*T + A*P + A*h + A
22 -> a*t + a*h + a
23 -> O*p + I*p + a*p + p
24 -> O*h + h*p + A*h + a*h + O + h + p + A + a + 1
25 -> O*p + I*p + P*p + h*p + A*h + a*h + O + h + A + a + 1
26 -> I^o + O*P + P*p + h*p + A*P + A*h + a*h + h + p + A + a + 1
27 -> o*p + I*p + P*p + h*p
28 -> h^o + A*h + a*h + a*p + o + h + A + a + 1
29 -> T*p + I*p + P*p + a*p + p
30 -> T*h + A*h + a*p + T + A
31 -> P^o + A*t + P*p + h*p + A*P + A*h + a*h + t + o + I + p + A + a + 1
32 -> P*T + I*p + h*p + A*P + a*h + a*p + T + P + h + p + A
33 -> I^o + I*p + P*p + A*P + o + p
34 -> I*T + P*T + I*p + P*p + h*p + A*P + a*h + a*p + I + P + h + p + A + 1
35 -> o^2 + o
36 -> p*t + a*p
37 -> h*t + h*p + a*p + t + h + p + 1
38 -> P*t + P^o + A*t + P*p + h*p + A*h + a*h + t + o + I + p + A + a + 1
39 -> I*t + P*t + I^o + P^o + A*t + h*p + A*P + A*h + a*h + p + A + a
40 -> O^o + I^o + h*p + A*h + a*h + o + h + p + A + a + 1
41 -> O*T + I*p + P*p + A*P + a*p + p
42 -> O*t + h*p + A*P + A*h + a*h + h + p + A + a + 1
43 -> o^2 + o
44 -> T^o + P^o + I*p + h*p + A*P + A*h + a*p + T + o + P + p + a + 1
45 -> T^2 + T
46 -> o*t + A*t + I*p + A*h + a*h + a*p + t + I + p + A + a + 1
47 -> T*t + O*t + A*t + I*p + h*p + A*h + a*p + t + T + P + h + p + a
48 -> t^2 + t
49 -> a*h*p
```

5. „a(l+1)“

Wenn a dann l

22. „a(t+h+1)“

Wenn a dann entweder t oder h

49. „ahp“

Stets ist eine der Bedingungen

a,h,p nicht erfüllt.

◀
M. Weber, K. Fackeldey: The Complexity of Comparative Text Analysis - "The Gardener is always the Murderer" says the Fourth Machine, arXiv:2012.07637 [cs.CL], 2020.

M. Weber, K. Fackeldey: The Mathematics of Comparing Objects, arXiv:2201.07032 [cs.AI], 2022.

M. Weber, O. Yürük: Coding Reliability with Aclus -- Did I correctly characterize my observations?, arXiv:2207.02855 [math.AC], 2022.

Noor-Cella Bena: Finden von Regeln und Mustern in komplexen Forschungsdaten mit Hilfe von Booleschen Polynomringen, Bachelorarbeit, 2023.