

# Computational Life Science am BreSCA

**ZeTeM Oberseminar  
03. November 2005**

**Manfred Nölte**

# Computational Life Science am BreSCA

- Die Struktur des BreSCA
- Zugang der Mathematik zur Computational Life Science
- Optimierung von Oligonukleotid-Bibliotheken für DNA-Mikroarrays

# Die Struktur des BreSCA

-

## Bremen Center for Scientific Computing and Applications Bremer Kompetenzzentrum für wissenschaftliches Rechnen

Die institutionellen Mitglieder von BreSCA sind

- Zentrum für Technomathematik ZeTeM
- Zentrum für Computational Material Science CMS
- (Zentrum für Computational Life Science CLS)
- Bremer Kompetenzzentrum für Höchstleistungsrechnen BremHLR

# Computational Life Science am BreSCA

- Die Struktur des BreSCA
- Zugang der Mathematik zur Computational Life Science
- Optimierung von Oligonukleotid-Bibliotheken für DNA-Mikroarrays

# Was ist Computational Life Science?

- Was ist Life Science?

Biologie, Medizin, Gesundheitswissenschaften,  
Biochemie, Biophysik, ...

- Biotechnologie, Molekularbiologie, Genetik, ...
- Onkologie, Virologie, Züchtung, Lebensmitteltechnologie, Ökosystemforschung, Umweltforschung, Biodiversität, Drug Design / Pharmatechnik, Labortechnik, DNA Analytik, ...
- Anwendung finden: Bioinformatik, Mikrosystemtechnik, Mikrosensorik, Nanotechnologie, ...

- **Computational Life Science** setzt Methoden der **Mathematik** und **Informatik** für wissenschaftliche Fragestellungen innerhalb der **Life Science** ein

# Zugang der Mathematik zur Computational Life Science

- Computational Life Science (Universität zu Lübeck) =  
50% Mathematik +  
25% Informatik + 25% Chemie, Biologie und Physik
- Mathematische Modellierung
- Simulation
- Bild- und Signalverarbeitung
  - Verarbeitung von Messdaten  
(z.B. Hybridisierungssignale, LCMS-Daten, Pyrogramme, ...)
- Interpretation/Auswertung von Messdaten (Datenanalyse, Klassifikation, ...)
- **Experiment-Design und -Optimierung**

# Computational Life Science am BreSCA

- Die Struktur des BreSCA
- Zugang der Mathematik zur Computational Life Science
- Optimierung von Oligonukleotid-Bibliotheken für DNA-Mikroarrays

# Experiment-Design am Beispiel der Optimierung von **Oligonukleotid-Bibliotheken** für **DNA-Mikroarrays**

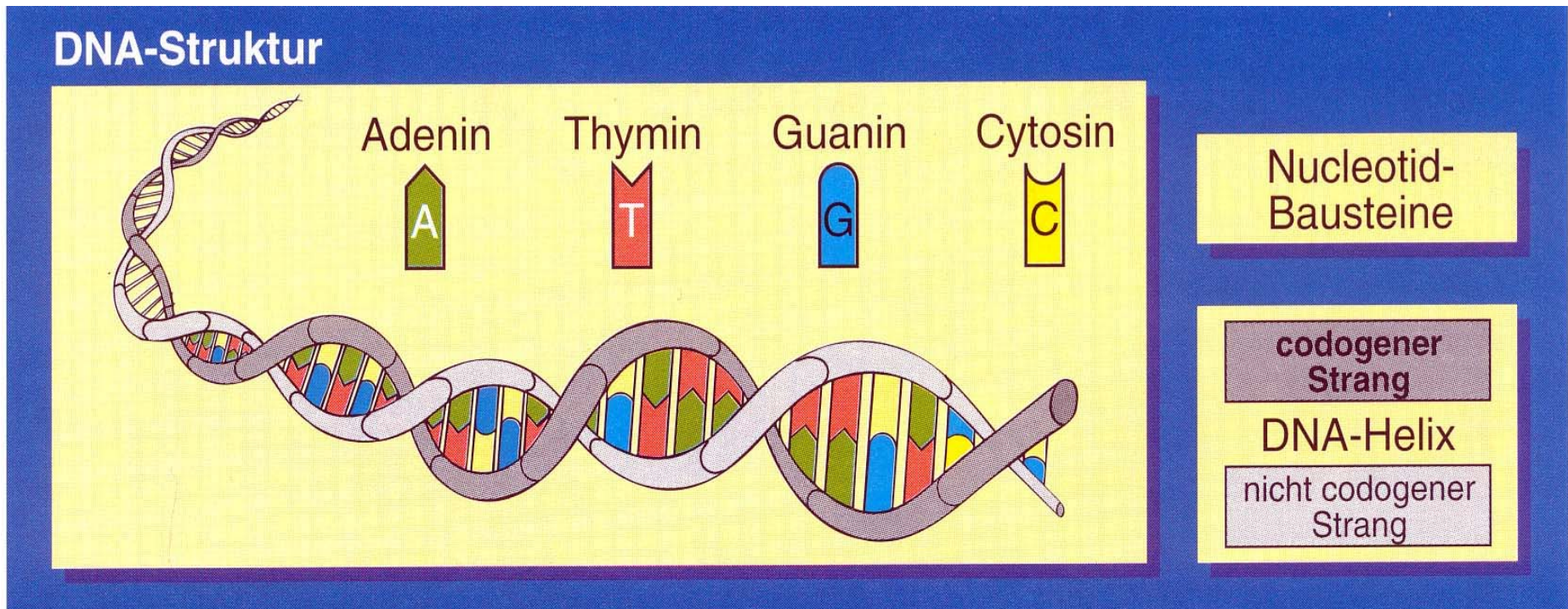
- **Oligonukleotide** sind „kurze“ DNA-Sequenzen
- **DNA-Mikroarrays** sind eine hochgradig parallelisierte Form der DNA-Analytik
  - Anwendungen sind Genexpression
  - ... sowie die Identifikation von Spezies und Krankheitsgenen
- **CAG – Centrum für angewandte Gensensorik**
  - BMBF Projekt zur Geräteentwicklung für eine Spezies Identifikation
  - Ausgründung der Bioinformatik Firma iSenseIt
  - EU Projekte „fish and chips“ und „Nanoparts“ bei Prof. Blohm
- Die Fragestellung aus Sicht der Biologie
  - Welche Fänger-Oligonukleotide/ DNA-Sonden müssen bei einer vorgegebenen Aufgabenstellung verwendet werden, um eine optimale Analytik mit DNA-Mikroarrays zu entwickeln?



# Optimierung von Oligonukleotid-Bibliotheken

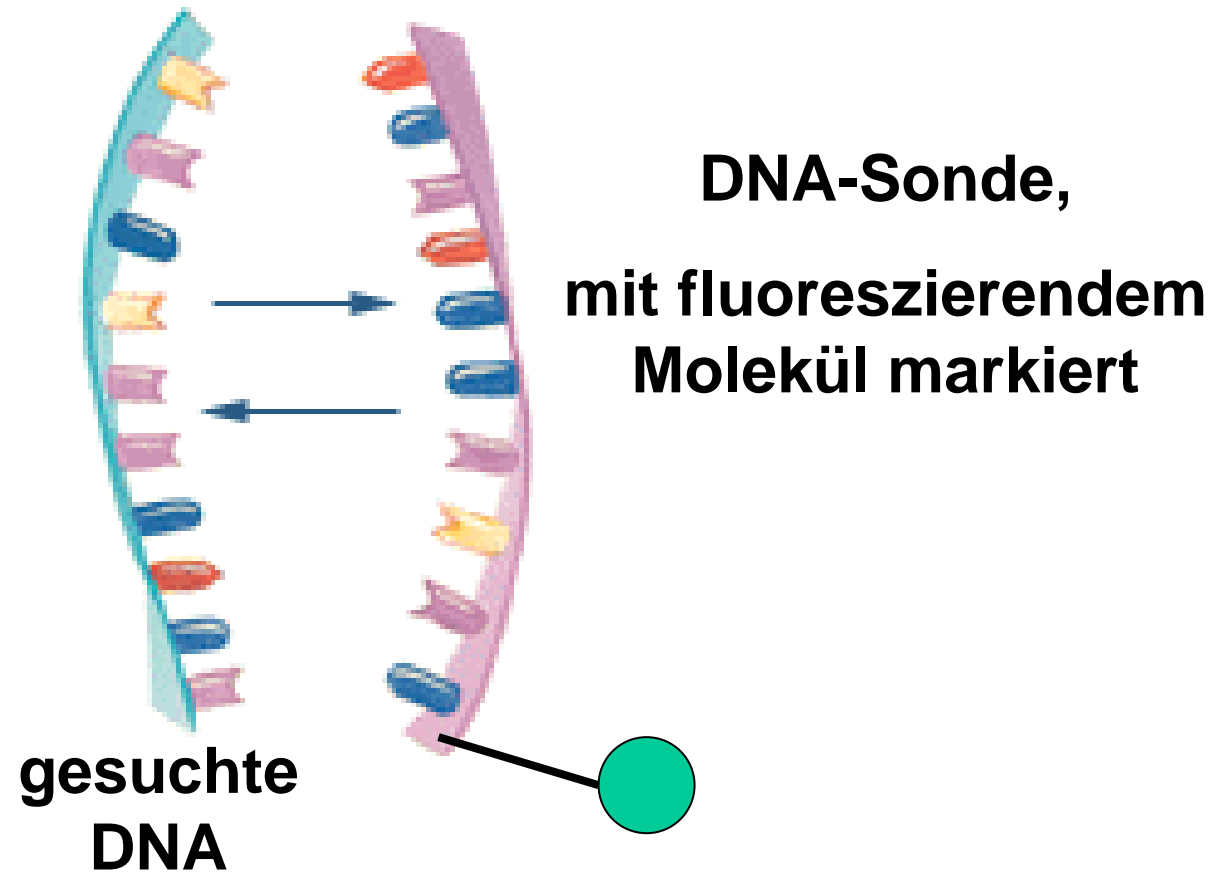
- Einführung
- Objekte und deren Eigenschaften
- Bewertungsfunktionen
- Formalisierung - Problemanalyse und Aufgabenspezifikation
- Optimierungs-Algorithmen
- Anwendungen

# DNA Struktur – Information steckt in der Basensequenz

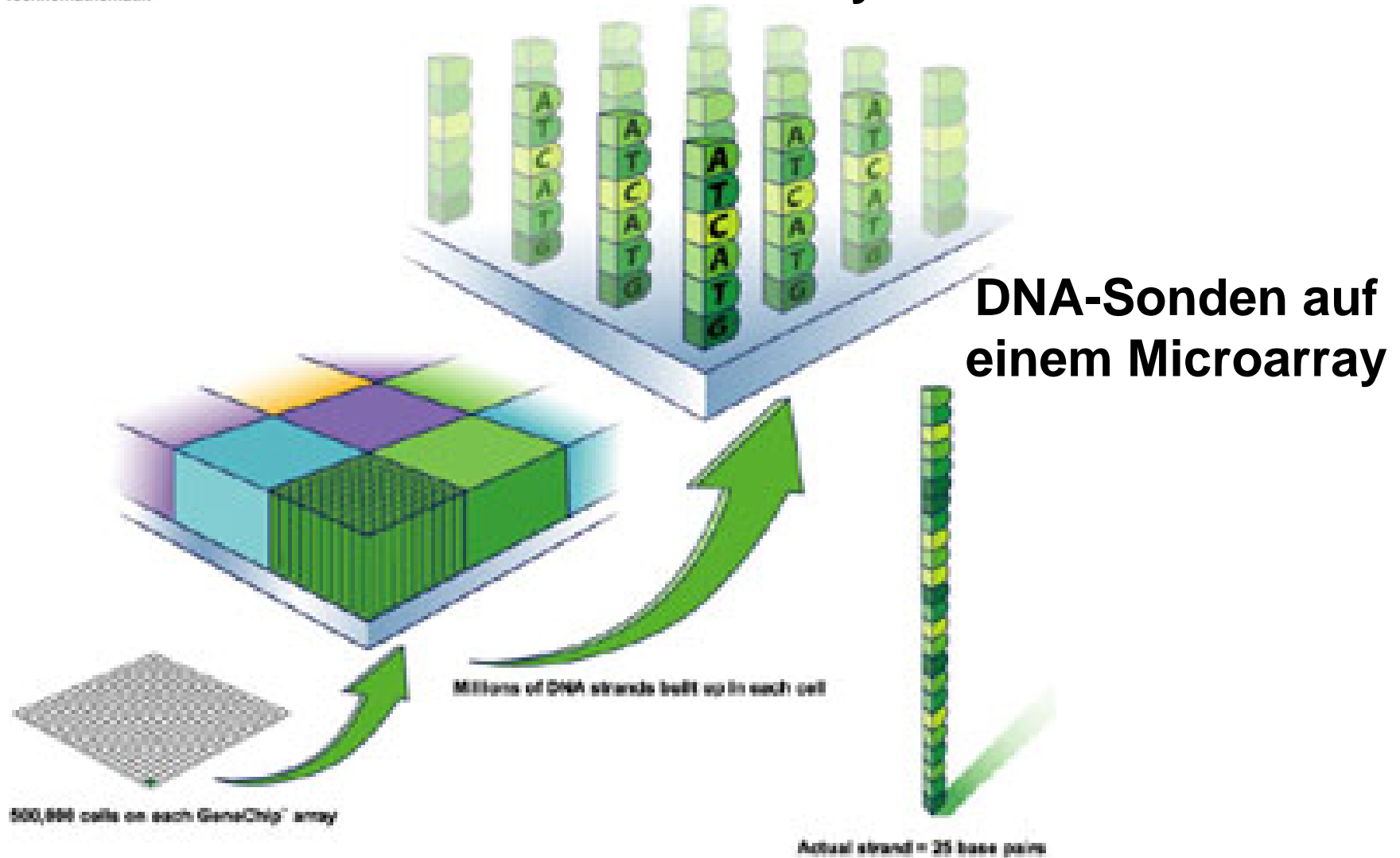


Aus: Folienserie des Fonds der  
chemischen Industrie

## Prinzip einer DNA Sonde



# DNA-Mikroarrays

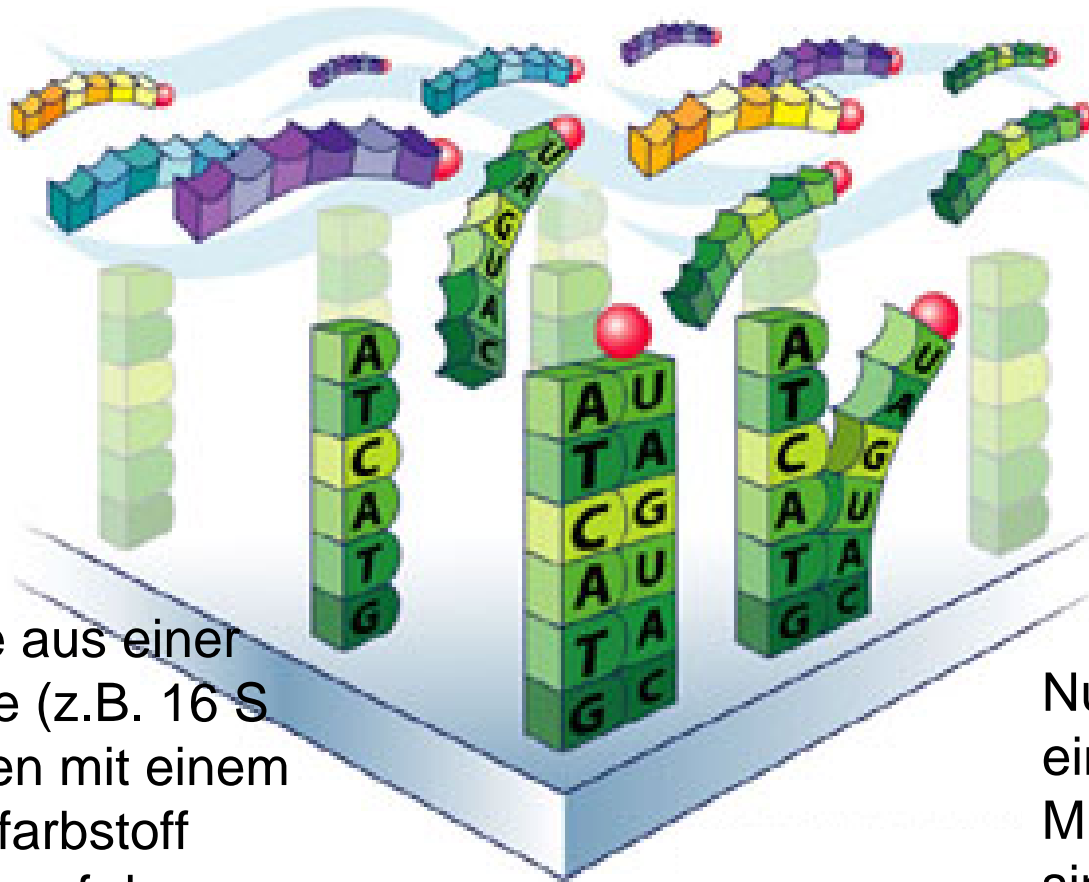


# DNA-Mikroarrays



Pro Microarray können mehrere Tausend Sensoren nach einem vorgegebenen Muster auf einem Träger aufgebracht sein.

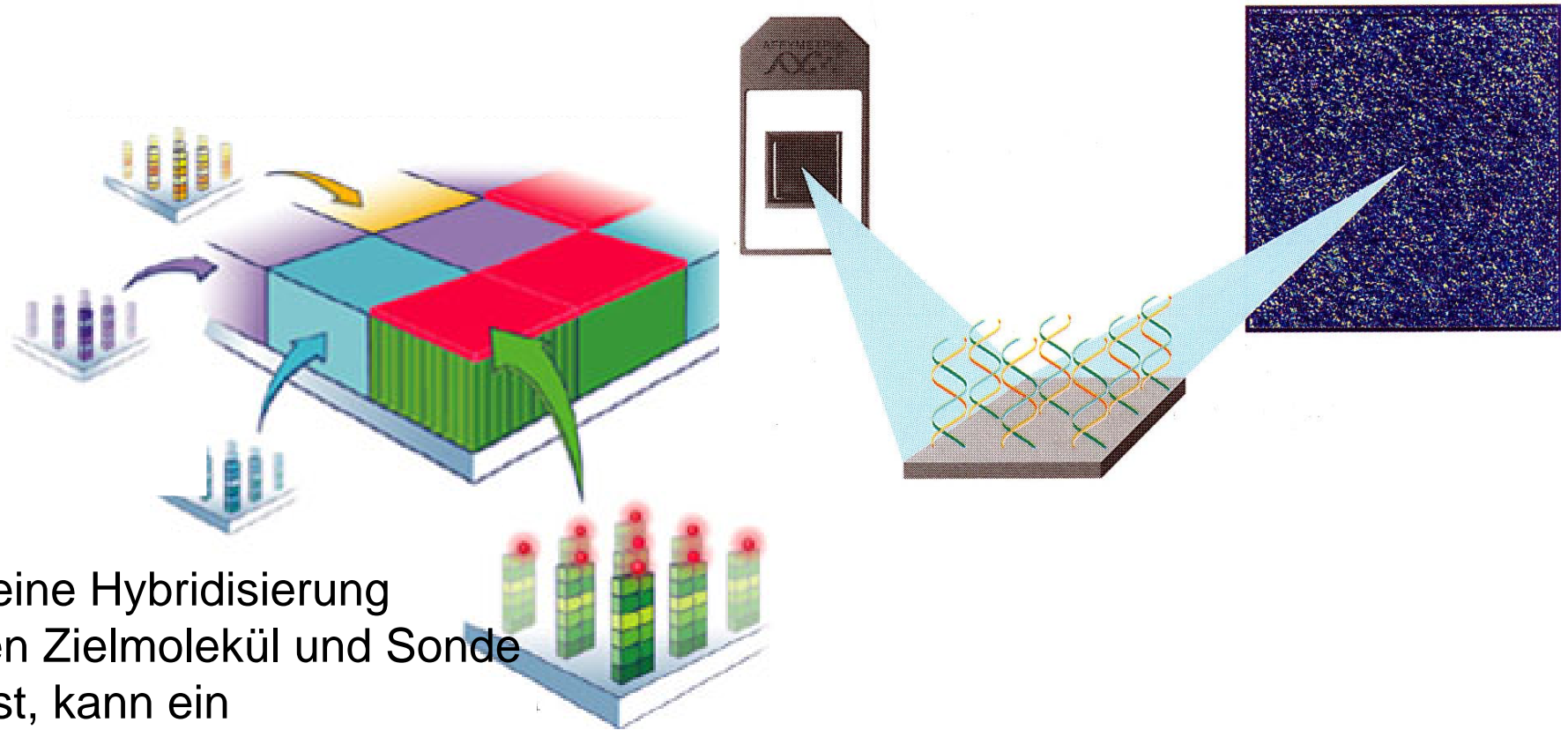
# DNA-Mikroarrays



Zielmoleküle aus einer Umweltprobe (z.B. 16 S rRNA) werden mit einem Fluoreszenzfarbstoff markiert und auf den Microarray aufgetragen.

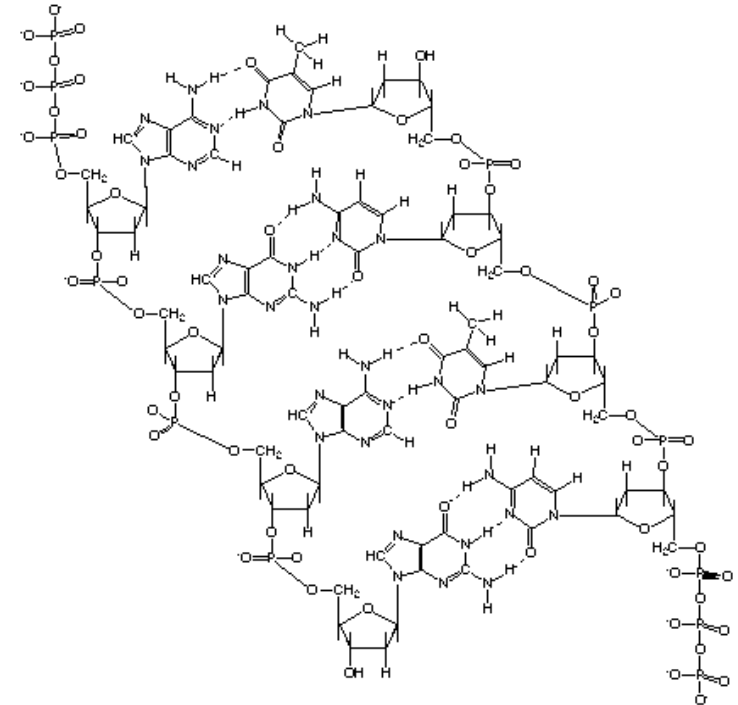
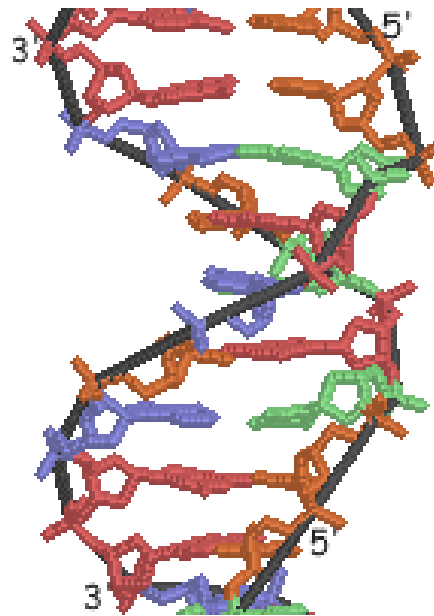
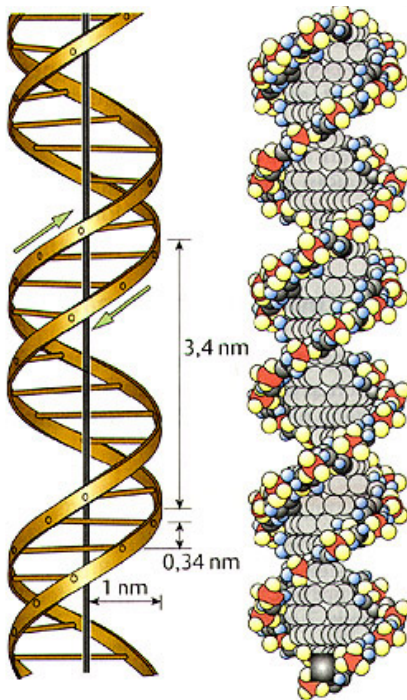
Nur Zielmoleküle, die zu einer Sonde auf dem Microarray komplementär sind, werden an den Chip binden.

# DNA-Mikroarrays



Sofern eine Hybridisierung zwischen Zielmolekül und Sonde erfolgt ist, kann ein Fluoreszenzsignal detektiert werden.

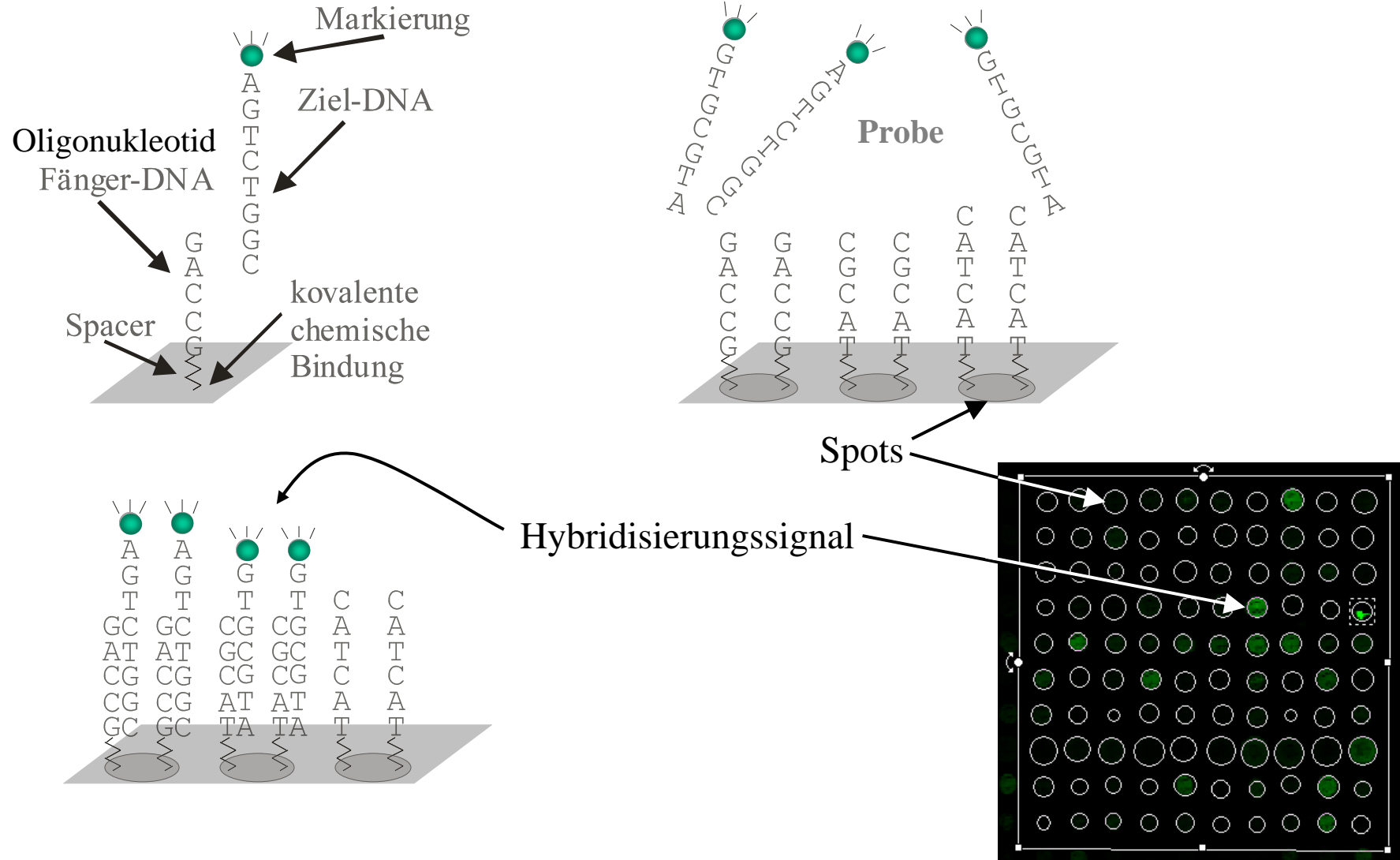
# DNA



5'ATCCGAAGCT 3'  
| | | | | | | | | |  
3'TAGGCTTCGA 5'



# Funktionsprinzip der DNA-Mikroarrays

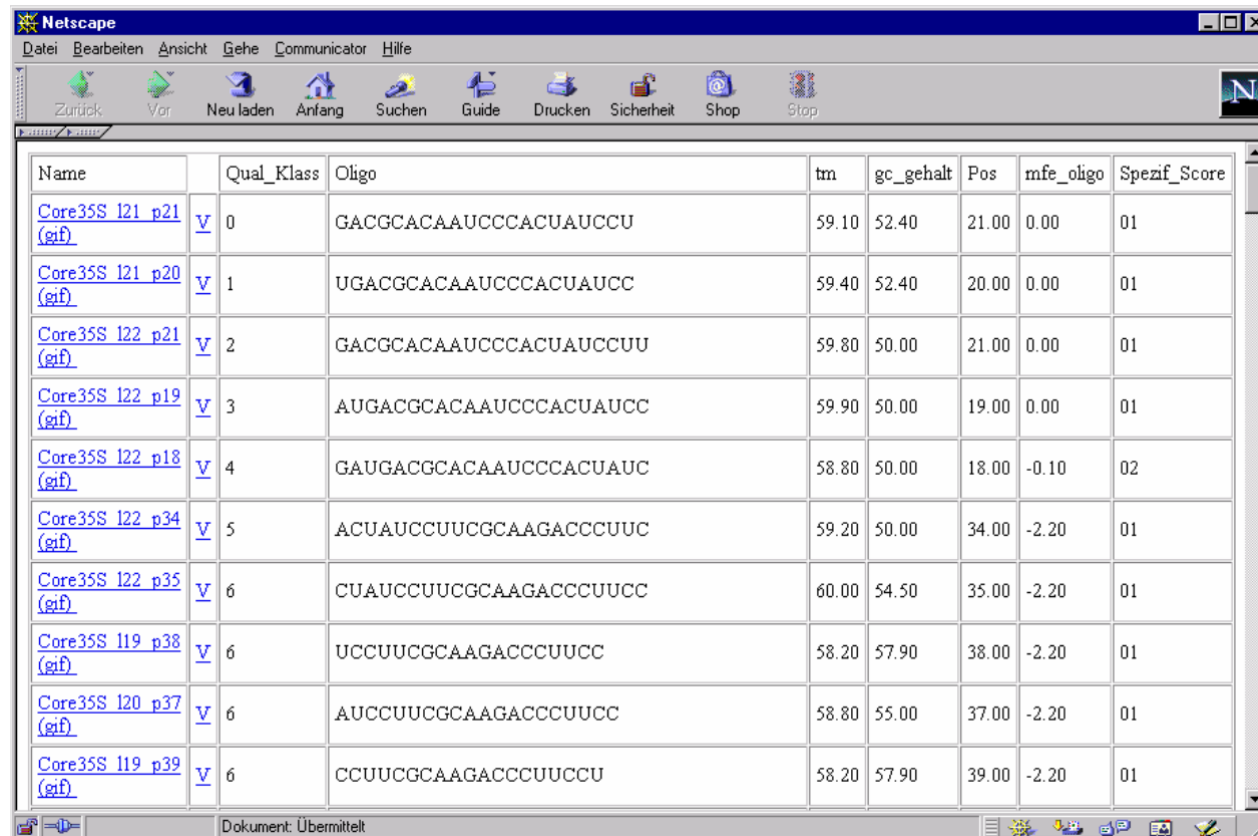


# Optimierung von Oligonukleotid-Bibliotheken für DNA-Mikroarrays

- **Problem:** Bei einer manuellen Erstellung von Oligonukleotid-Bibliotheken
  - sind große Mengen von Sequenz-Daten zu verarbeiten
  - kommt es zu einem langwierigen und fehleranfälligen Entwicklungsprozess
  - sind zahlreiche Kriterien zu berücksichtigen. Die zwei wesentlichen Aspekte:
    - **maximierte Sensitivität und Spezifität der Oligonukleotid-Bibliothek**
    - **optimale Hybridisierungs-Effizienzen der Fänger-Oligonukleotide**
- **Fragestellung und Zielsetzung:**
  - **aus Sicht der Biologie:** Welche Sequenzen/ Fänger-Oligonukleotide müssen bei einer vorgegebenen Aufgabenstellung - z.B. Organismen-Identifikation - verwendet werden, um eine optimale Analytik zu entwickeln?
  - **aus Sicht der Bioinformatik:** Entwicklung von Algorithmen zum automatisierten Design von optimierten Oligonukleotid-Bibliotheken

# Oligonukleotid-Bibliothek

## Oligonukleotid- Bibliothek:



Name	Qual_Klass	Oligo	tm	gc_gehalt	Pos	mfe_oligo	Spezif_Score
<a href="#">Core35S 121 p21 (gif)</a>	V 0	GACGCACAAUCCACUAUCCU	59.10	52.40	21.00	0.00	01
<a href="#">Core35S 121 p20 (gif)</a>	V 1	UGACGCACAAUCCACUAUCC	59.40	52.40	20.00	0.00	01
<a href="#">Core35S 122 p21 (gif)</a>	V 2	GACGCACAAUCCACUAUCCUU	59.80	50.00	21.00	0.00	01
<a href="#">Core35S 122 p19 (gif)</a>	V 3	AUGACGCACAAUCCACUAUCC	59.90	50.00	19.00	0.00	01
<a href="#">Core35S 122 p18 (gif)</a>	V 4	GAUGACGCACAAUCCACUAUC	58.80	50.00	18.00	-0.10	02
<a href="#">Core35S 122 p34 (gif)</a>	V 5	ACUAUCCUUCGCAAGACCCUUC	59.20	50.00	34.00	-2.20	01
<a href="#">Core35S 122 p35 (gif)</a>	V 6	CUAUCCUUCGCAAGACCCUUCC	60.00	54.50	35.00	-2.20	01
<a href="#">Core35S 119 p38 (gif)</a>	V 6	UCCUUCGCAAGACCCUUCC	58.20	57.90	38.00	-2.20	01
<a href="#">Core35S 120 p37 (gif)</a>	V 6	AUCCUUCGCAAGACCCUUCC	58.80	55.00	37.00	-2.20	01
<a href="#">Core35S 119 p39 (gif)</a>	V 6	CCUUCGCAAGACCCUUCU	58.20	57.90	39.00	-2.20	01

**Ziel-Sequenzen:**

Varianten der 5'UTR des Hepatitis C-Virus (HCV)

**Nichtziel-Sequenzen:**

humane DNA

**Hybridisierungsprotokoll:** 0,9 M NaCl, 50% Formamid, Hybridisierung bei RT, geeignet für Oligonukleotid-Längen 25-32bp, u.s.w.

# Optimierung von Oligonukleotid-Bibliotheken

- Einführung
- Objekte und deren Eigenschaften
- Bewertungsfunktionen
- Formalisierung - Problemanalyse und Aufgabenspezifikation
- Optimierungs-Algorithmen
- Anwendungen

## Eigenschaften der „Objekte“

- **Ziel- und Nichtziel-Sequenzen ( $t \in M$ )**
  - hohe oder niedrige Konserviertheit
  - Zugehörigkeit zu einer Sequenzklasse (z.B. Genotyp 1a bei Hepatitis C-Viren)
  - Sekundärstrukturen
- **Fänger-Oligonukleotide ( $x \in K$ )**
  - „Treffer“  $r_p$ ,  $f_p$  und „Nicht-Treffer“  $r_n$ ,  $f_n \rightarrow$  Sensitivität und Spezifität
  - Sekundärstrukturen
  - Position auf der Sekundärstruktur der Ziel-Sequenz
  - Schmelztemperatur, GC-Gehalt, Oligonukleotid-Länge
  - Sequenzmerkmale wie „GC-Clamp“, Basenwiederholungen

Diese Eigenschaften werden durch **Bewertungsfunktionen** (Scores) für einen algorithmischen Ansatz zugänglich gemacht.

# Optimierung von Oligonukleotid-Bibliotheken

- Einführung
- Objekte und deren Eigenschaften
- Bewertungsfunktionen
- Formalisierung - Problemanalyse und Aufgabenspezifikation
- Optimierungs-Algorithmen
- Anwendungen

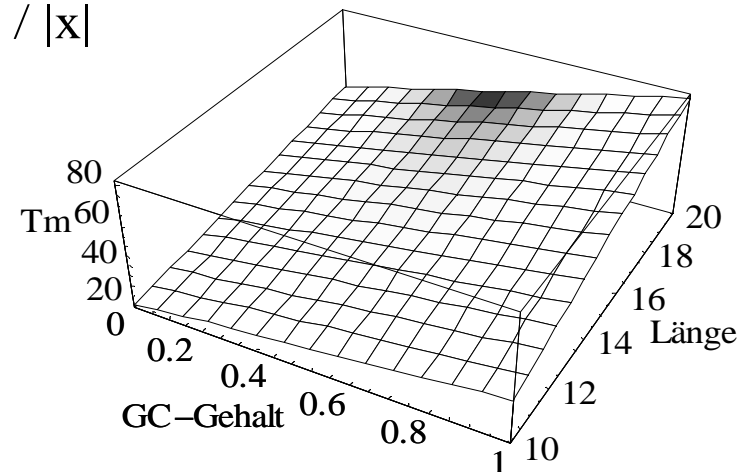
## Bewertungsfunktionen für die Eigenschaften der Fänger-Oligonukleotide x

- Match(x), sens(x), spez(x): resultieren aus den „Treffern“ des Oligonukleotids
- $\Delta G(x)$ : Sekundärstrukturen der Fänger-Oligonukleotide
- $\Delta\Delta G(x, t)$  und thdist(x, t): Position auf der Sekundärstruktur der Ziel-Sequenz

Schmelztemperatur:  $T_M = 2 \#[AT] + 4 \#[GC]$  Wallace-Regel

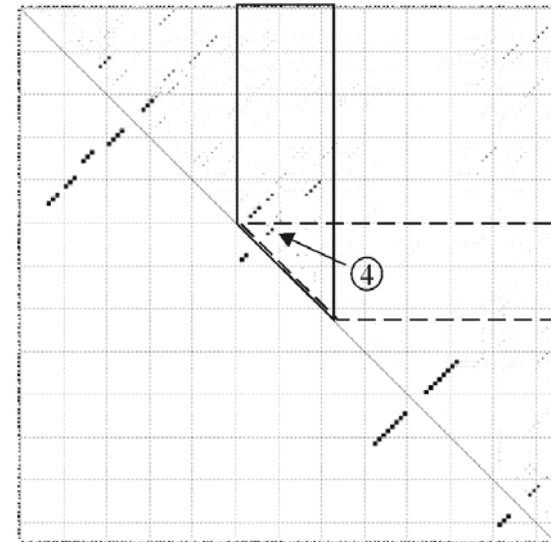
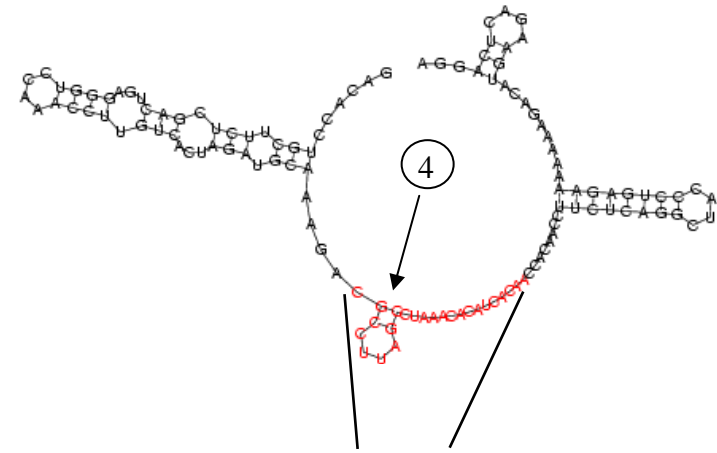
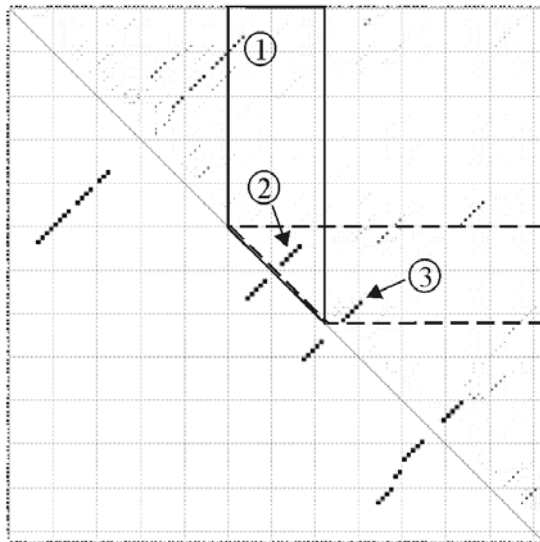
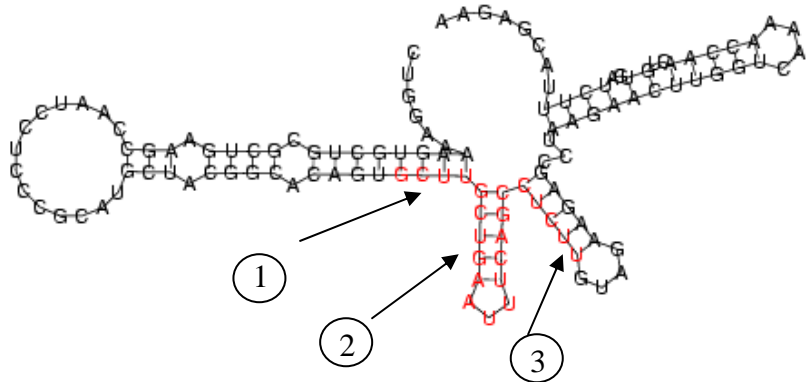
Oligonukleotid-Länge:  $|x| = \#[AT] + \#[GC] = \#[A] + \#[T] + \#[G] + \#[C]$

GC-Gehalt:  $\%GC = \#[GC] / |x|$



$$T_m = \Delta H^\circ / (\Delta S^\circ + R \ln C_T) \text{ mit } \Delta H = \sum (f_{ij} \cdot H_{ij}), \Delta S \text{ analog}$$

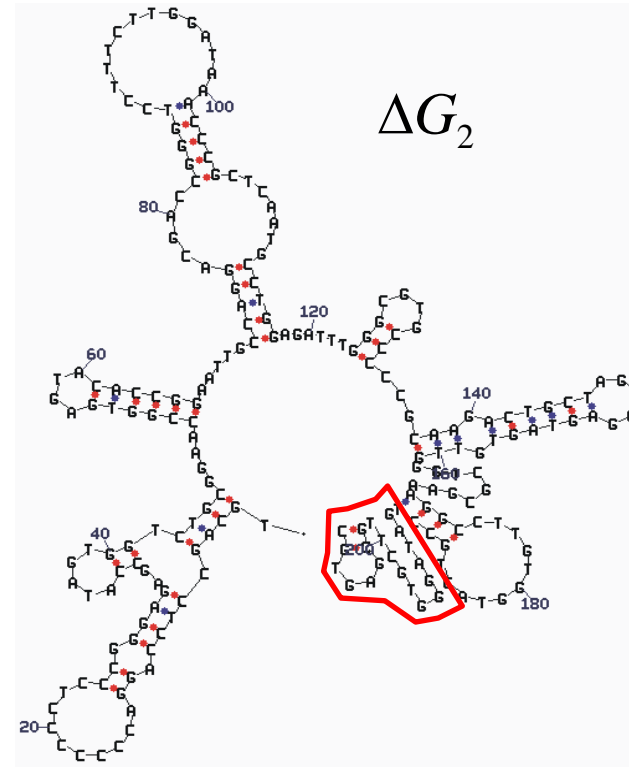
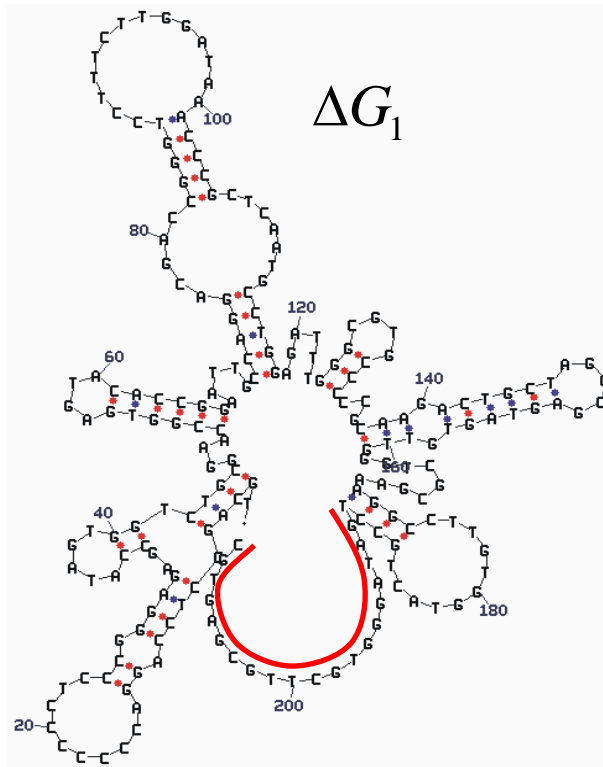
# Bewertungsfunktionen für die Eigenschaften der Ziel-Sequenzen t



...AAAGACGCUUAGCCUAAACAGAGUAGACACACCAGA...



# $\Delta\Delta G$ -Ansatz zur Sekundärstruktur-Bewertung

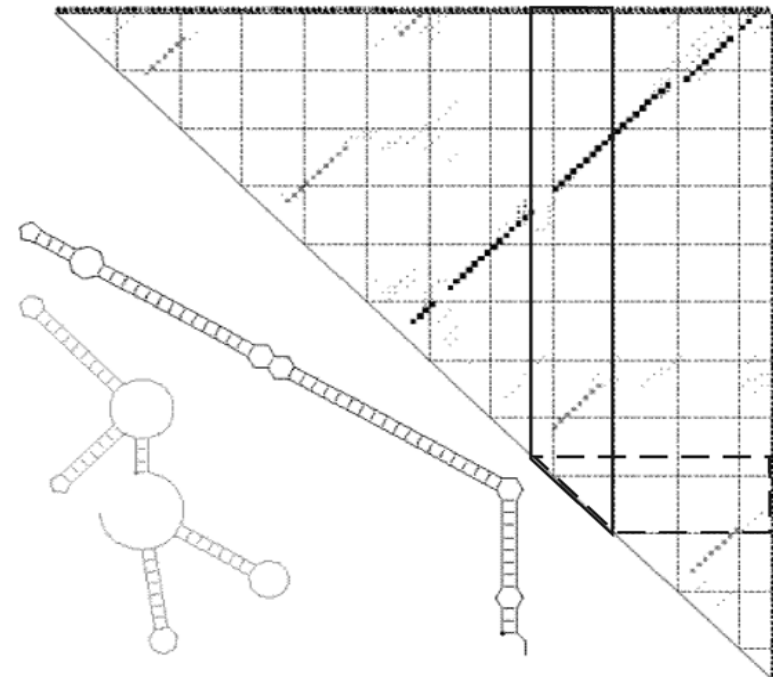
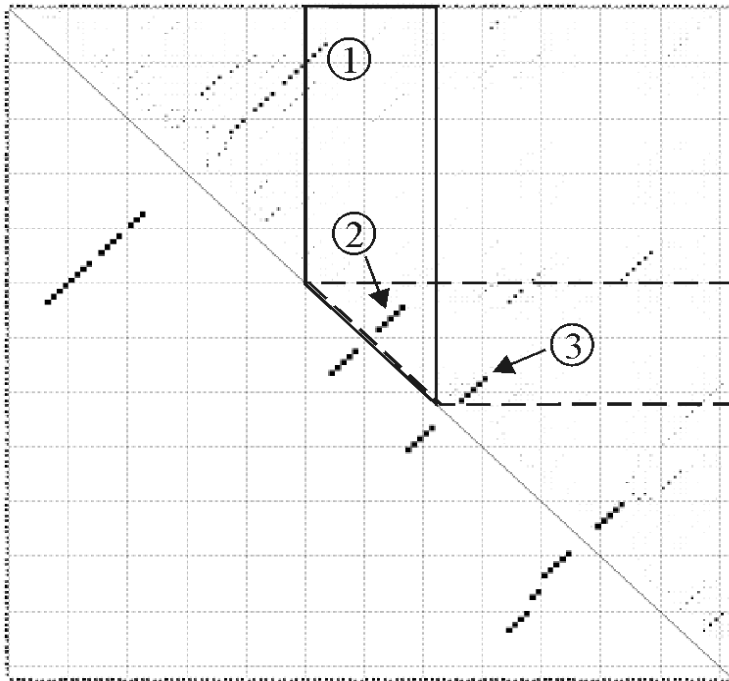


Sekundärstruktur mit und ohne Basenpaarungs-Bedingung

$$\Delta\Delta G(x, t) = \Delta G_1 - \Delta G_2 \geq 0$$

$$\text{quot}\Delta G(x, t) = \frac{\Delta G_2}{\Delta G_1} \geq 1$$

# Sekundärstruktur-Bewertung mit der Matrix der Basenpaarwahrscheinlichkeiten



$$\text{sek}(x, t) = \sum \text{[triangle shape]} + \sum \text{[trapezoid shape]}$$

$\text{sek}(x, t)$  hat Vorteile gegenüber  $\Delta\Delta G(x, t)$  bei Ziel-Sequenzen  $t$ , dessen Ensemble von Sekundärstrukturen mehrere stabilste Strukturen, mit einer geringen  $\Delta G$ -Differenz, enthält.

Dreiecksmatrix und Sekundärstruktur rechts aus

Flamm, C., Hofacker, I. L., Maurer-Stroh, S., Stadler, P. F., and Zehl, M. (2001): Design of Multi-Stable RNA Molecules, RNA, 7: 254-265.

# Optimierung von Oligonukleotid-Bibliotheken

- Einführung
- Objekte und deren Eigenschaften
- Bewertungsfunktionen
- Formalisierung - Problemanalyse und Aufgabenspezifikation
- Optimierungs-Algorithmen
- Anwendungen

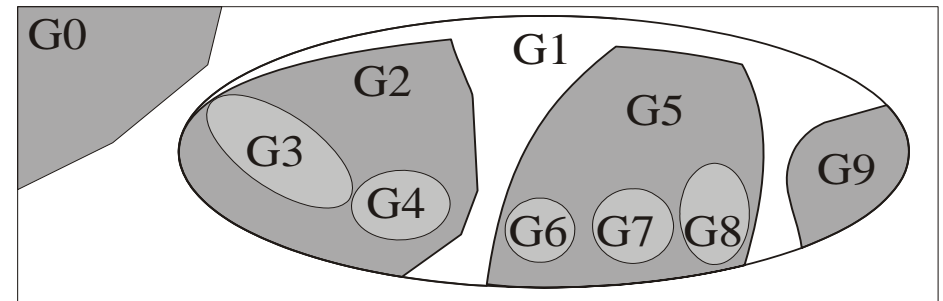
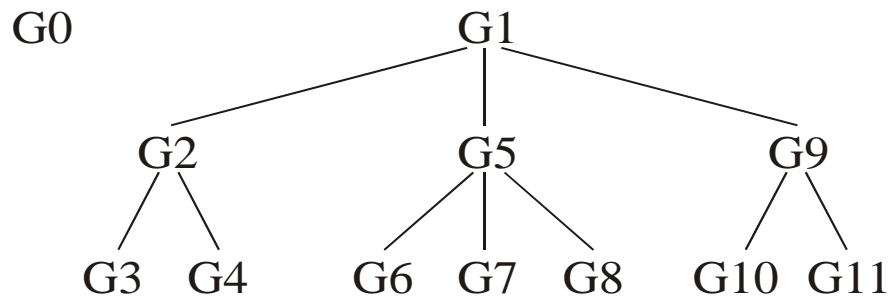
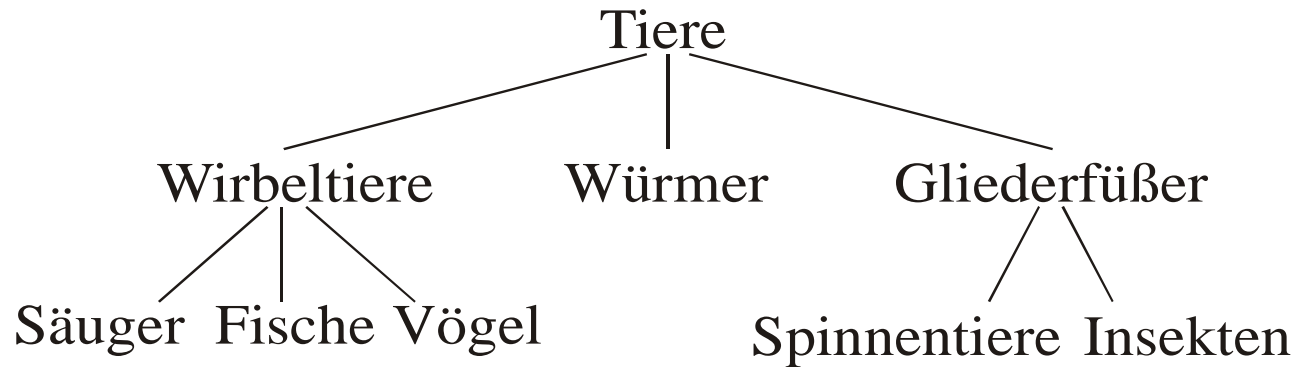
# Problemanalyse und Aufgabenspezifikation

- **Definition von Ziel- und Nichtziel-Sequenzen**  
unter Berücksichtigung hierarchisch angeordneter Sequenz-Klassen  $G_i$
- **Definition und Vorhersage von „positiven und negativen Signalen“**  
mit Hilfe der Grenzwerte  $g$ ,  $g_Z$  und  $g_N$ , dem Redundanz-Niveau  $r \in \mathbb{N}$  und dem Toleranz-Niveau  $s \in \mathbb{N}$
- Sensitivität einer ganzen Oligonukleotid-Bibliothek  $L$ :  $sens_r(L)$ 
  - möglichst viele der Ziel-Sequenzen treffen  
(Maximierung der *richtig-positiven*)
- Spezifität einer ganzen Oligonukleotid-Bibliothek  $L$ :  $spez_s(L)$ 
  - möglichst wenige der Nichtziel-Sequenzen treffen  
(Minimierung der *falsch-positiven*)

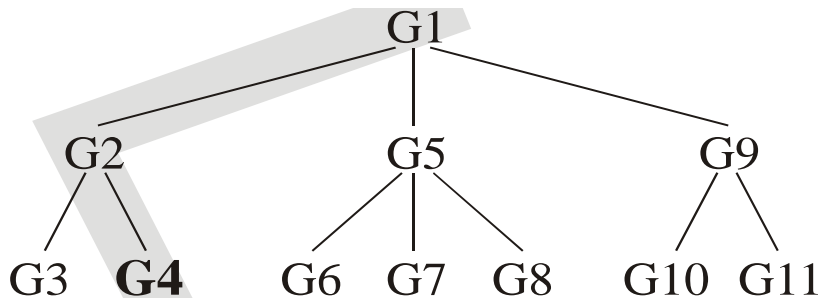
	Ziel-Klasse	Nichtziel-Klasse
Signal positiv	richtig-positive: rp	falsch-positive: fp
Signal negativ	falsch-negative: fn	richtig-negative: rn

# Definition von Ziel- und Nichtziel-Sequenzen - I

## Aus einer Phylogenie abgeleitete hierarchische Struktur von Sequenz-Klassen



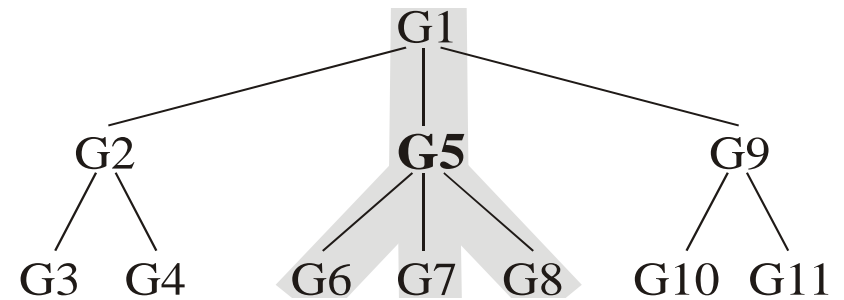
## Definition von Ziel- und Nichtziel-Sequenzen - II



Ziel-Sequenzen: Die Menge G4  
 Nichtziel-Sequenzen:  $G3 \cup G5 \cup G9$

$$M = G4$$

$$A = G3 \cup G5 \cup G9$$



Ziel-Sequenzen: Die Menge G5  
 Nichtziel-Sequenzen:  $G2 \cup G9$

$$M = G5$$

$$A = G2 \cup G9$$

## Problemanalyse und Aufgabenspezifikation

- **Definition von Ziel- und Nichtziel-Sequenzen**  
unter Berücksichtigung hierarchisch angeordneter Sequenz-Klassen  $G_i$
- **Definition und Vorhersage von „positiven und negativen Signalen“**  
mit Hilfe der Grenzwerte  $g$ ,  $g_Z$  und  $g_N$ , dem Redundanz-Niveau  $r \in \mathbb{N}$  und dem Toleranz-Niveau  $s \in \mathbb{N}$
- Sensitivität einer ganzen Oligonukleotid-Bibliothek  $L$ :  $sens_r(L)$ 
  - möglichst viele der Ziel-Sequenzen treffen  
(Maximierung der *richtig-positiven*)
- Spezifität einer ganzen Oligonukleotid-Bibliothek  $L$ :  $spez_s(L)$ 
  - möglichst wenige der Nichtziel-Sequenzen treffen  
(Minimierung der *falsch-positiven*)

	Ziel-Klasse	Nichtziel-Klasse
Signal positiv	richtig-positive: rp	falsch-positive: fp
Signal negativ	falsch-negative: fn	richtig-negative: rn

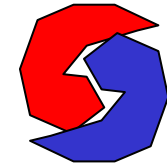


# Definition und Vorhersage von „positiven und negativen Signalen“

Bewertungsfunktionen für die **Hybridisierungseffizienz**  
und damit der Definition von „**Treffern**“:

$H(x, t)$  angelehnt an die Hamming-Distanz

$thdist(x, t)$  „thermodynamische Distanz“ zwischen  $x$  und  $t$   
(Differenz der freien Enthalpien  $\Delta G$ )



$H(x, t)$	Ziel-Klasse	Nichtziel-Klasse
0	rp0	fp0
1	rp1	fp1
2	fn2	rn2
$\geq 3$	fn3	rn3

Treffermenge des Oligonukleotids  $x$ :

$$Match(x) := \{ t \in M \mid H(x, t) \leq g \}$$

←  $g = 1$

	Ziel-Klasse	Nichtziel-Klasse
0 MM = PM	rp0	fp0
1 MM	fn1	fp1
2 MM	fn2	fp2
$\geq 3$ MM	fn3	rn3

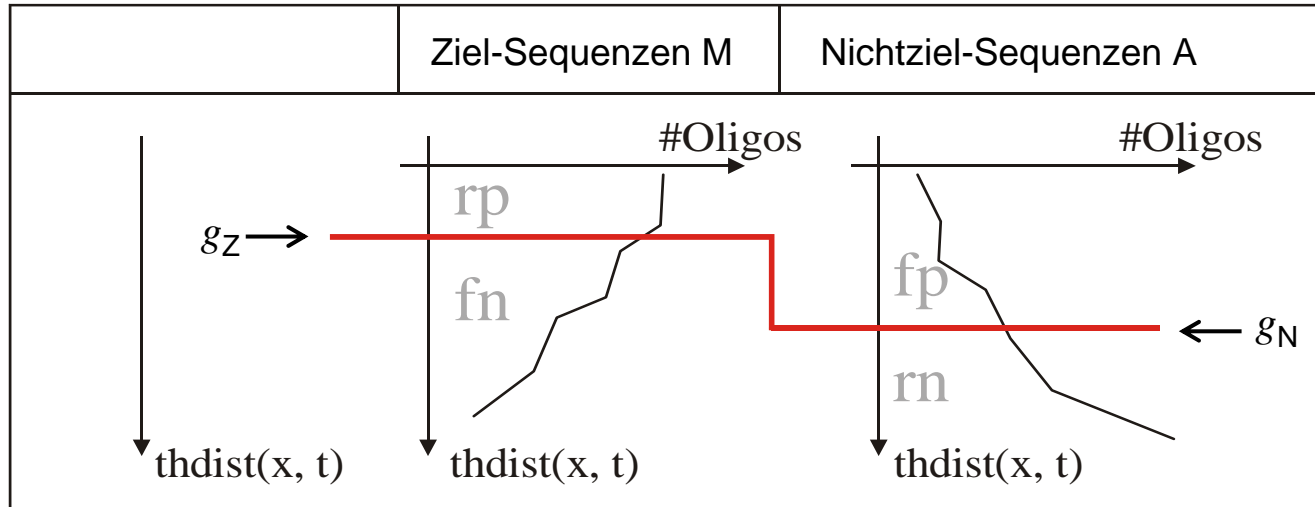
Grenzwerte  $g_Z$  und  $g_N$ :

←  $g_Z = 0$  für alle  $t \in M$ :  $H(x, t) \leq g_Z$

←  $g_N = 2$  für alle  $t \in A$ :  $H(x, t) \leq g_N$



# Definition und Vorhersage von „positiven und negativen Signalen“



für alle  $t \in M$  gibt es  
eine positives Signal,  
wenn  $\mathbf{thdist}(x, t) \leq g_Z$  ;  
für alle  $t \in A$   
wenn  $\mathbf{thdist}(x, t) \leq g_N$

Redundanz- und  
Toleranz-Niveaus

$$r \text{ und } s: \quad rp := |\{t \in M \mid m(t, L_i) \geq r\}| ; \quad fp := |\{t \in A \mid m(t, L_i) > s\}|$$

$$fn := |\{t \in M \mid m(t, L_i) < r\}| ; \quad rn := |\{t \in A \mid m(t, L_i) \leq s\}|$$

Zu den Mengen der Ziel- und Nichtziel-Sequenzen können nun zusammen mit der Treffermenge  $Match(x)$  eines Oligonukleotids  $x$  oder der Menge  $Match(L)$  einer ganzen Oligonukleotid-Bibliothek  $L$  die Sensitivität und die Spezifität berechnet werden.

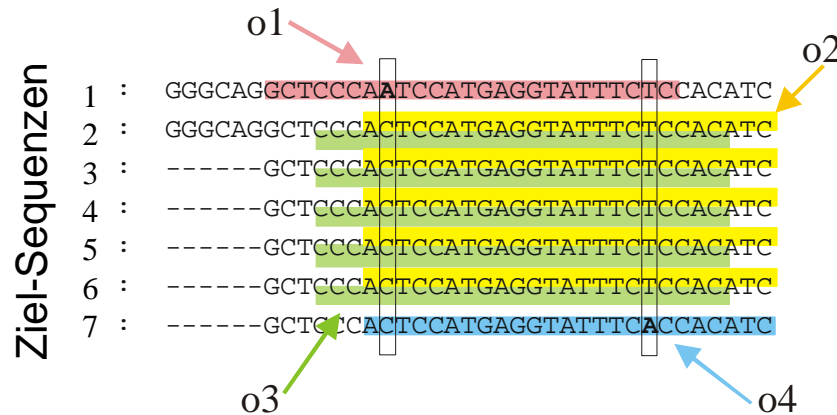
## Problemanalyse und Aufgabenspezifikation

- **Definition von Ziel- und Nichtziel-Sequenzen**  
unter Berücksichtigung hierarchisch angeordneter Sequenz-Klassen  $G_i$
- **Definition und Vorhersage von „positiven und negativen Signalen“**  
mit Hilfe der Grenzwerte  $g$ ,  $g_Z$  und  $g_N$ , dem Redundanz-Niveau  $r \in \mathbb{N}$  und dem Toleranz-Niveau  $s \in \mathbb{N}$
- **Sensitivität einer ganzen Oligonukleotid-Bibliothek  $L$ :  $sens_r(L)$** 
  - möglichst viele der Ziel-Sequenzen treffen  
(Maximierung der *richtig-positiven*)
- **Spezifität einer ganzen Oligonukleotid-Bibliothek  $L$ :  $spez_s(L)$** 
  - möglichst wenige der Nichtziel-Sequenzen treffen  
(Minimierung der *falsch-positiven*)

	Ziel-Klasse	Nichtziel-Klasse
Signal positiv	richtig-positive: rp	falsch-positive: fp
Signal negativ	falsch-negative: fn	richtig-negative: rn

# Sensitivität: Warum kombinatorische Optimierung?

Mit der Menge der Ziel-Sequenzen  $M$  und  $P = \text{Match}(K)$  wird das Tupel  $(M, P)$  als „Set Cover“-Problem bezeichnet.



**Ziel:** maximale Sensitivität (Überdeckung) mit möglichst wenigen Oligonukleotiden.

**Beispiel:**  $K = \{o1, o2, o3, o4\}$  ;  
 $M = \{\text{die sieben Zeilen}\}$   
 $\text{Match}(o_i) = \{\text{die jeweils farbig hervorgehobenen Zeilen}\}$

Sortierung der Oligonukleotide nach Sensitivität:

Oligo	Sensitivität
o2	5/7
o3	5/7
o1	1/7
o4	1/7

Oligonukleotid-Bibliothek	Strategie	Sensitivität
$L1 = \{o2, o3, o1\}$	$S1 = \text{„nehme die drei besten Oligonukleotide“}$	6/7
$L2 = \{o2, o1, o4\}$	$S2 = \text{„berücksichtige Kombinationen“}$	7/7

## Problemanalyse und Aufgabenspezifikation

- **Definition von Ziel- und Nichtziel-Sequenzen**  
unter Berücksichtigung hierarchisch angeordneter Sequenz-Klassen  $G_i$
- **Definition und Vorhersage von „positiven und negativen Signalen“**  
mit Hilfe der Grenzwerte  $g$ ,  $g_Z$  und  $g_N$ , dem Redundanz-Niveau  $r \in \mathbb{N}$  und dem Toleranz-Niveau  $s \in \mathbb{N}$
- Sensitivität einer ganzen Oligonukleotid-Bibliothek  $L$ :  $sens_r(L)$ 
  - möglichst viele der Ziel-Sequenzen treffen  
(Maximierung der *richtig-positiven*)
- Spezifität einer ganzen Oligonukleotid-Bibliothek  $L$ :  $spez_s(L)$ 
  - möglichst wenige der Nichtziel-Sequenzen treffen  
(Minimierung der *falsch-positiven*)

	Ziel-Klasse	Nichtziel-Klasse
Signal positiv	richtig-positive: rp	falsch-positive: fp
Signal negativ	falsch-negative: fn	richtig-negative: rn

## „Set Cover“-Problem mit der Nebenbedingung Spezifität

Mit der Menge der Ziel-Sequenzen  $M$ ,  $P = Match(K)$  und der Menge der Nichtziel-Sequenzen  $A$  wird das Tripel  $(M, P, A)$  als „Set Cover“-Problem mit der Nebenbedingung Spezifität bezeichnet.

**Beispiel:**  $M = \{ 5'UTR\text{-Sequenzen des Hepatitis C-Virus Genotyp 1} \}$   
 $K = \{ \text{alle Teilsequenzen der } t \in M, \text{ die vorgegebene Kriterien erfüllen} \}$   
 $Match(o_i)$  und damit  $P = Match(K)$  ergibt sich aus der Wahl von  $g_Z$  und  $g_N$   
 $A = \{ \text{menschliche Sequenzen und die HCV-Sequenzen der Genotypen 2, 3 und 4} \}$

	Ziel-Klasse	Nichtziel-Klasse
Signal positiv	richtig-positive: rp	falsch-positive: fp
Signal negativ	falsch-negative: fn	richtig-negative: rn

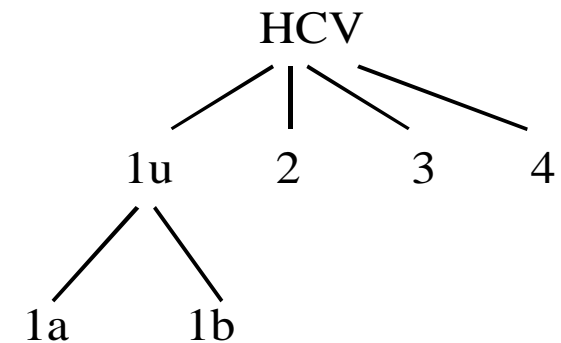
$$\text{sens}(x) = rp / (rp + fn)$$

$$\text{spez}(x) = rn / (rn + fp)$$

## Aufgabenspezifikation

Aus einer Hierarchie z.B. (HCV, (GT1 (GT1a, GT1b), GT2, GT3))  
werden mehrere (M, P, A)-Probleme abgeleitet.

**Parameter:**  $(\min T_m, \max T_m) = (60^\circ\text{C}, 64^\circ\text{C})$   
 $(\min \text{Len}, \max \text{Len}) = (26\text{bp}, 32\text{bp})$   
 $(\min \text{GC}, \max \text{GC}) = (45\%, 55\%)$   
 $(g_Z, g_N) = (0, 1)$        $(r, s) = (1, 0)$



**sonstige Parameter:** Hybridisierungs-Temperatur, Salzgehalt,  
DNA/RNA-Parametersätze der “nearest neighbor interactions”,  
“ $\Delta G$  correction terms” und der “microarray correction term”

**Aufgabe:** Treffe mit möglichst wenigen Oligonukleotiden  $x$  einer Oligonukleotid-Bibliothek  $L$  möglichst viele Ziel-Sequenzen  $t \in M$  (Sensitivität) und möglichst wenige aus der Menge der Nichtziel-Sequenzen  $A$  (Spezifität) und optimiere dabei bzgl. der Bewertungsfunktionen  $\Delta G(x)$  sowie  $\Delta\Delta G(x, t)$  oder  $\text{sek}(x, t)$

## iQserve – parameter form

- Oligonucleotide length: 50-63nt
- Melting Temperature: 70-85°C
- GC content: 40-60%
  
- Salt concentration: 100mM
- Formamide concentration: 5%
- DNA concentration: 0.1mM
  
- Position: 600E..E
- Oligo sequence type: DNA
- Target sequence type: DNA
- Reverse complement: no
- Maximal mismatches: 10 (or thermodynamical score: 2500)
- Delimitation list: default hum\_est hum\_refchr  
hum\_refrns hum\_refcum

## Optimierung von Oligonukleotid-Bibliotheken

- Einführung
- Objekte und deren Eigenschaften
- Bewertungsfunktionen
- Formalisierung - Problemanalyse und Aufgabenspezifikation
- Optimierungs-Algorithmen
- Anwendungen



# Heuristische kombinatorische Optimierungs-Algorithmen

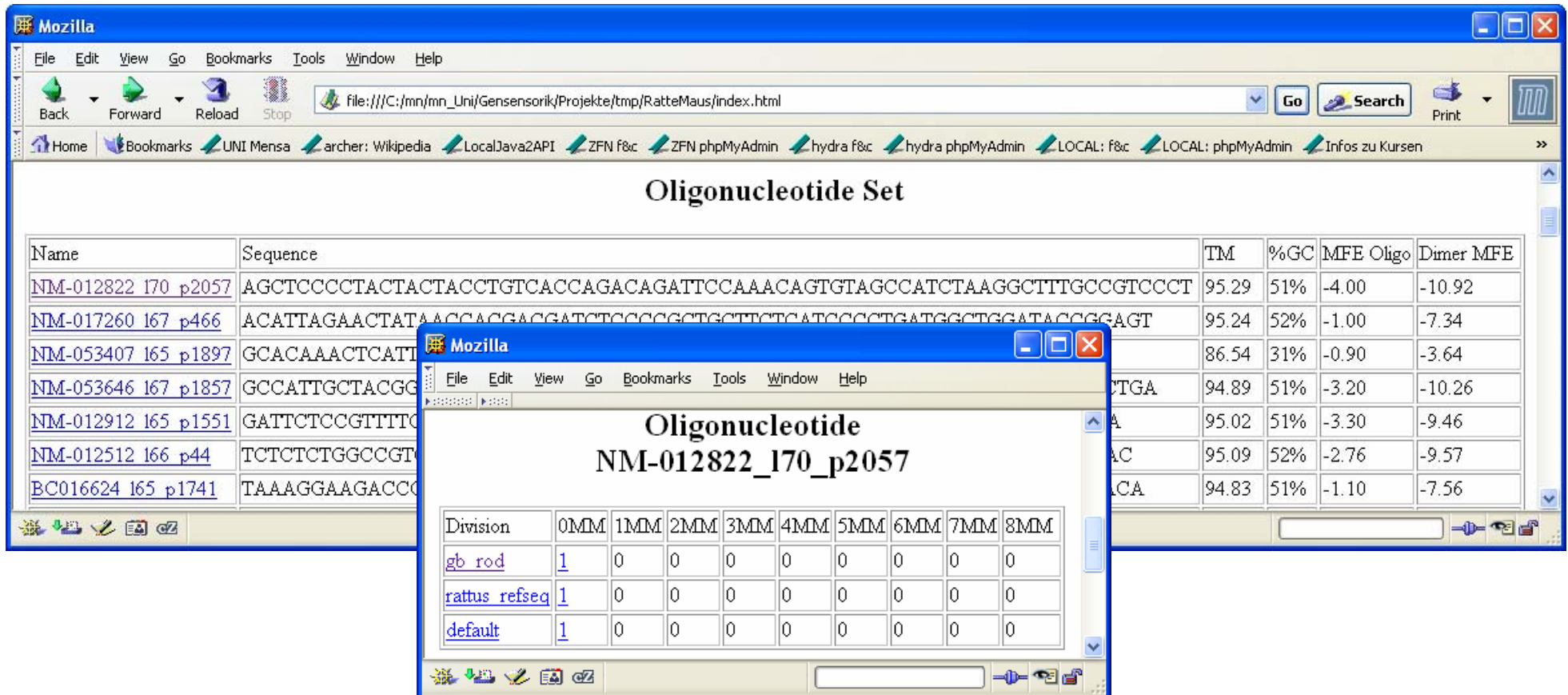
- Modifiziertes „*Greedy Set Covering*“
- Kombination von Gradientenabstieg und Competition
- Genetische Algorithmen

# Optimierung von Oligonukleotid-Bibliotheken

- Einführung
- Objekte und deren Eigenschaften
- Bewertungsfunktionen
- Formalisierung - Problemanalyse und Aufgabenspezifikation
- Optimierungs-Algorithmen
- Anwendungen

# Anwendungen – Fish-and-Chips

## 16 europäische Fischarten



**Oligonucleotide Set**

Name	Sequence	TM	%GC	MFE Oligo	Dimer MFE
<a href="#">NM-012822_170_p2057</a>	AGCTCCCCTACTACTACCTGTCACCAGACAGATTCCAAACAGTGTAGCCATCTAAGGCTTTGCCGTCCCT	95.29	51%	-4.00	-10.92
<a href="#">NM-017260_167_p466</a>	ACATTAGAACTATAACCACGAGGATCTCCCGCTGCTCTCATCCCTGATGGCTGGATACCGGAGT	95.24	52%	-1.00	-7.34
<a href="#">NM-053407_165_p1897</a>	GCACAAACTCATT	86.54	31%	-0.90	-3.64
<a href="#">NM-053646_167_p1857</a>	GCCATTGCTACGG	94.89	51%	-3.20	-10.26
<a href="#">NM-012912_165_p1551</a>	GATTCTCCGTTTT	95.02	51%	-3.30	-9.46
<a href="#">NM-012512_166_p44</a>	TCTCTCTGGCCGT	95.09	52%	-2.76	-9.57
<a href="#">BC016624_165_p1741</a>	TAAAGGAAGACC	94.83	51%	-1.10	-7.56

**Oligonucleotide**  
**NM-012822\_170\_p2057**

Division	0MM	1MM	2MM	3MM	4MM	5MM	6MM	7MM	8MM
<a href="#">gb rod</a>	1	0	0	0	0	0	0	0	0
<a href="#">rattus refseq</a>	1	0	0	0	0	0	0	0	0
<a href="#">default</a>	1	0	0	0	0	0	0	0	0

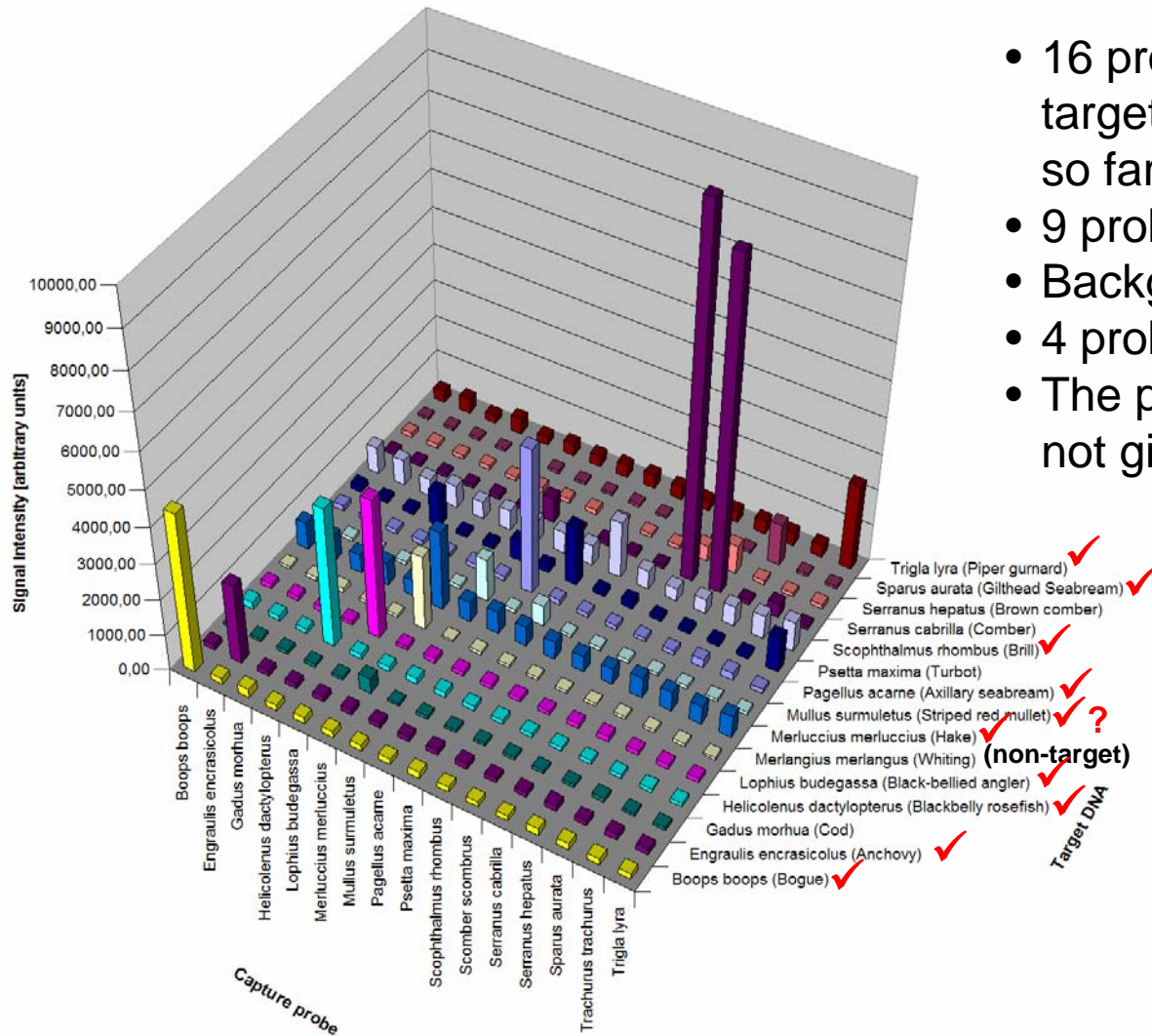
iQserve:Präsentation der HTML-Dokumentation

file:///C:/mn/mn\_Uni/Gensensorik/Projekte/fish-and-chips\_2005-07-05/default.Auslieferung/output/index.html

HTML-Dokumentation mit NCBI-Link

file:///C:/mn/mn\_Uni/Gensensorik/Projekte/tmp/RatteMaus/index.html

# Hybridisierungssignale des Chip Designs für das EU Projekt „fish and chips“



- 16 probes have been tested with 14 target and one non-target species so far
- 9 probes are species specific
- Background noise is low
- 4 probes are not species specific
- The probe for *Gadus morhua* does not give a signal

# Vorteile und Anwendungsmöglichkeiten von iQserve

- Ermöglicht das automatische Design von kombinatorisch optimierten Oligonukleotid-Bibliotheken (manuell nicht lösbar)
- Reduzierung des Entwicklungsprozesses auf wenige Tage
- Es können berücksichtigt werden:
  - große Mengen von Sequenz-Daten
  - zahlreiche Parameter und Kriterien
- Die fehleranfällige, manuelle Bearbeitung großer Datenmengen entfällt
- Ohne großen Mehraufwand können Parameter des Experiment-Designs geändert werden (z.B die Oligonukleotid-Länge)
- Elemente des Hybridisierungsprotokolls und die labortechnischen Möglichkeiten Hybridisierungssignale zu diskriminieren, können zur Parametrisierung des Systems verwendet werden