#### Thomas Hinze

Lehrstuhl Bioinformatik Friedrich-Schiller-Universität Jena

thomas.hinze@uni-jena.de http://users.minet.uni-jena.de/~hinze

4. Labornahe Simulation molekularer Computer



#### Weshalb labornahe Simulation?

- Modellbildung ist der Schlüssel zur formalen Beschreibung.
- Formale Beschreibung dient dem Verständnis.

#### **Simulation**

- dient der Veranschaulichung.
- macht Prozesse f
   ür den Menschen fassbar und wahrnehmbar.
- erleichtert das Verständnis und zeigt Zusammenhänge auf.
- ermöglicht Was-wäre-wenn-Szenarien.
- dient der kostengünstigen experimentellen Vorbereitung und Optimierung von Laborimplementierungen.
- hilft beim Suchen und Aufdecken von Fehlern.
- kann durch gutes Anschauungsmaterial potentielle Geldgeber überzeugen

# Begriffe

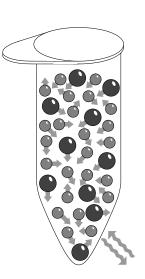
# System – Systemobjekte Systemobjekte – Transition

#### Simulationsmethoden

- statisch vs. dynamisch
- deterministisch vs. stochastisch
- endogen vs. exogen

Dynamische Simulationen im Speziellen

- ereignisgesteuert vs. zeitgesteuer
- diskret vs. kontinuierlich



Einführung

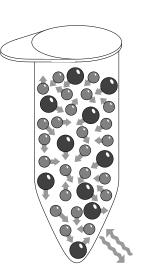
# **Begriffe**

System – Systemobjekte Systemobjekte – Transition

#### **Simulationsmethoden**

- · statisch vs. dynamisch
- deterministisch vs. stochastisch
- endogen vs. exogen

- diskret vs. kontinuierlich.



Einführung

# **Begriffe**

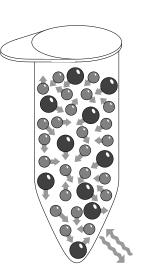
System - Systemobjekte Systemobjekte - Transition

#### Simulationsmethoden

- statisch vs. dynamisch
- deterministisch vs. stochastisch
- endogen vs. exogen

Dynamische Simulationen im Speziellen:

- ereignisgesteuert vs. zeitgesteuert
- diskret vs. kontinuierlich



# Softwaretools zur Simulation molekularer Computer Auswahl

- AMBER (Simulation von Biomolekülen und ihrer Dynamik)
- CellDesigner / COPASI (Simulation chem. Reaktionsnetzwerke)
- DNA-Haskell (Funktionales Experimentiersystem, TU Dresden)
- **EDNA** (Virtuelles Reagenzglas, Univ. Memphis)
- FunDNA (Functional Workbench, Univ. Texas)
- Hellics (DNA Computing Simulation, Univ. Oldenburg)
- MGS (Rule-Based Language for Complex Objects, Univ. Evry)
- NACST (Toolbox-Sammlung zu DNA, Seoul National Univ.)
- P-System-Simulatoren (PSim, ..., U. Sheffield, Sevilla, Milano)
- saces (Simple Artificial Chemistry Experimental System)
- Sisyphus (Sim. Syst. for Phenomena Undergoing Side Effects)
- SRSim / LAMMPS (Rule-Based Modelling in Space)
- Virtuelle Labore (Univ. Oldenburg, OFFIS)

# Simple Artificial Chemistry Experimental System

Entwickler: A. Aguillon, D. Noelpp, HTI Bern, 2005–2006

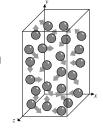
#### Idee

Dynamisches Verhalten idealer Gase kann rechnen

#### **Prinzip**

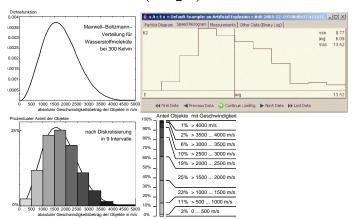
- Menge von Partikeln Moleküle mit Parametern ( $m, E_b, \#, ...$ )
- Zufällig plaziert und angestoßen entsprechend Maxwell-Boltzmann-Verteilung
- Menge von Reaktionen und globale Parameter  $A + B \longrightarrow C + D$ :  $E_{\text{activation}}$  (B oder D auch leer)
- Algorithmus: Brownsche Bewegung (random walk) mit elastischen/inelast. Stößen, Impuls- und Energieerhaltung
- Auswertung: Animation, Protokoll, Bericht, Histogramm
- Beispiele: endl. Automaten, Lösung NP-vollst. Probleme





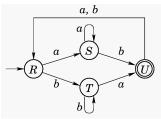
### Maxwell-Boltzmann-Verteilung – saces (II) Geschwindigkeitsverteilung der Molekülobjekte

$$\mathrm{f}_0(|\vec{v}|) = 4\pi \cdot \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} \cdot |\vec{v}|^2 \cdot \mathrm{e}^{-\frac{m|\vec{v}|^2}{2k_B T}}$$



# Simulationsbeispiel: Endlicher Automat

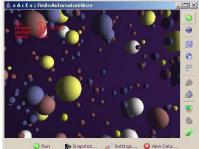
saces (III). Screenshots: A. Aguillon, D. Noelpp, HTI Bern, 2006

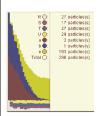


Name	Initial Count	Energy	Radius	Mass	Color
R	100	10	1	100	
5	0	10	1	100	150-35
T	0	10	1	100	
U	0	10	1	100	
a	400	1	0.3	3	
b	400	1	0.3	3	-
e	0	1	0.3	3	
New Part	ticle Class	Delete Currer	nt		



 $U + b \longrightarrow R$ 





# Sisyphus

Simulation System for Phenomena Undergoing Side Effects Entwickler: C.Frutschi<sup>1</sup>, P.Grossniklaus<sup>1</sup>, U.Hatnik<sup>2</sup>, B.Jambresic<sup>1</sup>, R.Lehmann<sup>1</sup>, <sup>1</sup>HTI Bern, <sup>2</sup>Fraunhofer-Institut IIS-EAS Gruppe ModSim, 2001–2005

Beschreibungen: http://www.hti.bfh.ch → Informatik → Diplomarbeiten

#### Idee

- Nucleotidgenaue Abbildung linearer (Doppel)Stränge
- Satz von molekularbiologischen Operationen auf DNA
- Parametrisierung von Seiteneffekteinflüssen

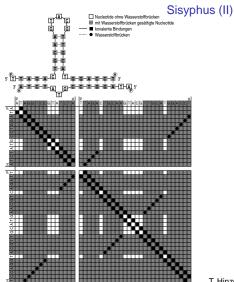
#### **Prinzip**

- Formalisierung der Operationswirkungen einschl. wichtiger Seiteneffekte
- Objektorientierte Simulation in Java
- Multimengenbasierter, restriktiver und nichtdeterministischer Ansatz
- Definition von Reagenzgläsern und ihres anfänglichen Inhaltes
- Visualisierung und Aufbereitung entstandener Reagenzglasinhalte

Konzept: T. Hinze, U. Hatnik, M. Sturm. An Object Oriented Simulation of Real Occurring Molecular Biological Processes for DNA Computing and Its Experimental Verification, In N. Jonoska, N.C. Seeman, editors, DNA Computing. Proceedings DNA7. Series Lecture Notes in Computer Science, Vol. 2340, pp. 1-13, Springer Verlag, 2002



# DNA-Stränge als Bindungsmatrizen



T. Hinze, M. Sturm. Rechnen mit DNA. Oldenbourg, 2004

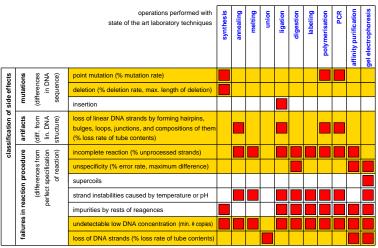
# **DNA-Operationen**

Sisyphus (III)

- Synthesis
- Annealing
- Melting
- Union
- Ligation
- Digestion
- Labeling
- Polymerisation
- Affinity Purification
- Detect (angereicherte Fluoreszenzmarker)
- Gel Electrophoresis
- ⇒ Jeweils spezifische Operations- und Seiteneffektparameter

## Seiteneffekte von DNA-Operationen

Sisyphus (IV)



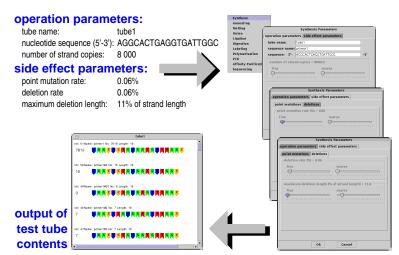
: supported in simulation tool

in brackets: statistical parameters

: significant side effect caused by the operation

# Stochastischer Simulationsansatz - Synthesis

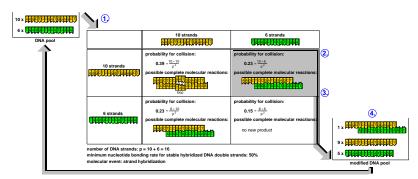
#### Sisyphus (V)



# Stochastischer Simulationsansatz - Annealing

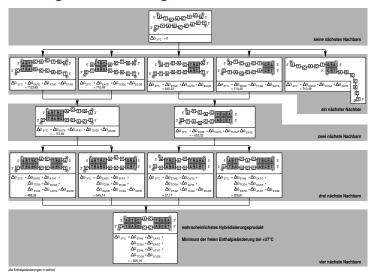
Sisyphus (VI)

#### **Iteration molekularer Events**



- 1 Create list (matrix) of molecular events and their probabilities including side effects
- 2) Select one molecular event randomly with respect to the probability distribution
- 3) Determine all possible reaction products from this molecular event and select one of them
- 4. Modify DNA pool

# Annealingprodukt mittels Nearest Neighbor Minimierung der freien Energie



# Minimierung der freien Energie: wahrscheinlichstes Hybridisierungsprodukt

Normwertetabelle (+37°C,  $K_{Na^+} = 1 \frac{\text{mol}}{\text{I}}$ )

nächster	$\Delta G$ in $\frac{\text{cal}}{\text{mol}}$
Nachbar	mor
AA/TT	-1013,27
AC/GT	-1471,09
AT/AT	-862,33
TA/TA	-499,71
TC/GA	-1284,64

nächster	$\Delta G$ in $\frac{\text{cal}}{\text{mol}}$
Nachbar	
CA/TG	-1354,90
CG/CG	-2135,74
CT/AG	-1298,15
GC/GC	-2195,79
GG/CC	-1829.39

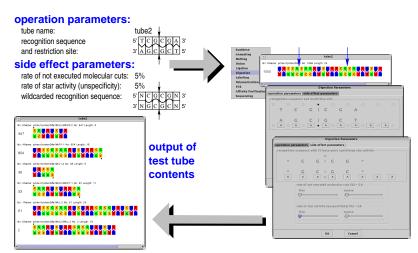
End-	$\Delta G$ in $\frac{\text{cal}}{\text{mol}}$
stücke	
EA/TE	993,11
ET/AE	993,11
EC/GE	1002,02
EG/CE	1002,02
-	•

⇒ Umrechnung bei Abweichung von den Normwerten mgl.

SantaLucia jr., H.T. Allawi, P.A. Seneviratne. Improved nearest-neighbor parameters for predicting DNA duplex stability. Biochemistry 35(11):3555–3562, 1996

# Stochastischer Simulationsansatz – Digestion

Sisyphus (IX)

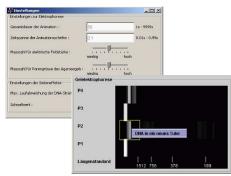


# Simulationsansatz für Gel Electrophoresis

#### Sisyphus (X)

#### **Dynamisch kontinuierlich**

- Herleitung einer mathematischen Gesetzmäßigkeit für den zu simulierenden Vorgang
- Finden einer geeigneten Abstraktionsstufe
- Diskretisierung und Animation des zeitlichen und räumlichen Ablaufs



Grafik: C. Frutschi, P. Grossniklaus. Simulation ausgewählter DNA-Computing-Operationen. HTI Bern, 2004

Sisyphus (XI)

#### Gleichgewicht zwischen elektrischer und Reibungskraft

$$F_E = F_R$$

$$F_E = q \cdot E = q \cdot \frac{U}{h}$$

$$F_R = 6 \cdot \pi \cdot \eta \cdot r \cdot v$$
 (Stokessches Gesetz)

Sisyphus (XI)

#### Gleichgewicht zwischen elektrischer und Reibungskraft

$$F_E = F_R$$
 $F_E = q \cdot E = q \cdot \frac{U}{h}$ 

Objektladung  $q$ 

Gleichspannung  $U$ 

Elektrodenabstand  $h$ 

$$F_R = 6 \cdot \pi \cdot \eta \cdot r \cdot v$$
 (Stokessches Gesetz)

Materialkonstante für Viskosität des Mediums  $\eta$ 

wirksamer Radius  $r$ 

Bewegungsgeschwindigkeit  $v$ 

⇒ Annahme: Kugel ("DNA-Knäuel") bewegt sich durch Flüssigkeit

 $F_F = F_R$ 

# Beschreibung der Gelelektrophorese

Sisyphus (XI)

#### Gleichgewicht zwischen elektrischer und Reibungskraft

$$F_E = q \cdot E = q \cdot \frac{U}{h}$$
Objektladung  $q$ 
Gleichspannung  $U$ 
Elektrodenabstand  $h$ 
 $F_R = 6 \cdot \pi \cdot \eta \cdot r \cdot v$  (Stokessches Gesetz)
Materialkonstante für Viskosität des Mediums  $\eta$ 
wirksamer Radius  $r$ 

⇒ Annahme: Kugel ("DNA-Knäuel") bewegt sich durch Flüssigkeit

Bewegungsgeschwindigkeit V

Sisyphus (XII)

#### Radius r auf Masse und Dichte zurückführen

$$V_{ ext{Kugel}} \geq V_{ ext{DNA-Molekul}}$$
  $\frac{4}{3} \cdot \pi \cdot r^3 \geq \frac{m}{\rho}$  Masse  $m$ 

#### Daraus folgt:

$$r \geq \left(\frac{3}{4 \cdot \pi} \cdot \frac{m}{\rho}\right)^{\frac{1}{3}}$$

$$r = \left(\frac{3}{4 \cdot \pi} \cdot \frac{m}{\rho}\right)^{\frac{1}{3}} \cdot G \text{ mit } G > 1$$

Sisyphus (XII)

#### Radius r auf Masse und Dichte zurückführen

$$egin{array}{lll} V_{ ext{Kugel}} & \geq & V_{ ext{DNA-Molekul}} \ rac{4}{3} \cdot \pi \cdot r^3 & \geq & rac{m}{
ho} \ & ext{Masse } m \ & ext{Dichte } 
ho \end{array}$$

$$r \geq \left(\frac{3}{4 \cdot \pi} \cdot \frac{m}{\rho}\right)^{\frac{1}{3}}$$

$$r = \left(\frac{3}{4 \cdot \pi} \cdot \frac{m}{\rho}\right)^{\frac{1}{3}} \cdot G \text{ mit } G > 1$$

Sisyphus (XII)

#### Radius r auf Masse und Dichte zurückführen

$$egin{array}{lll} V_{ ext{Kugel}} & \geq & V_{ ext{DNA-Molekul}} \ rac{4}{3} \cdot \pi \cdot r^3 & \geq & rac{m}{
ho} \ & ext{Masse } m \ & ext{Dichte } 
ho \end{array}$$

#### Daraus folgt:

$$r \geq \left(\frac{3}{4 \cdot \pi} \cdot \frac{m}{\rho}\right)^{\frac{1}{3}}$$

$$r = \left(\frac{3}{4 \cdot \pi} \cdot \frac{m}{\rho}\right)^{\frac{1}{3}} \cdot G \text{ mit } G > 1$$

Sisyphus (XIII)

#### Bewegungsgleichung der DNA-Moleküle im Gel

Zurückgelegter Weg s in Abhängigkeit von der Zeit t

$$s(t) = v \cdot t$$

$$s(t) = \frac{q \cdot E}{6 \cdot \pi \cdot \eta \cdot \left(\frac{3 \cdot m}{4 \cdot \pi \cdot \rho}\right)^{\frac{1}{3}} \cdot G} \cdot t$$

$$= \frac{q}{6 \cdot \pi \cdot \left(\frac{3}{4 \cdot \pi \rho}\right)^{\frac{1}{3}} \cdot G} \cdot \underbrace{\frac{E}{\eta}}_{\text{idealisierend als Lauf-konstante aufgefasst}} \cdot \underbrace{\frac{1}{m^{\frac{1}{3}}}}_{\text{DNA-unabhängige Konstante}} \cdot t$$

$$= C \cdot \underbrace{\frac{E}{\eta} \cdot \frac{1}{m^{\frac{1}{3}}}}_{\text{Im}} \cdot t$$

Sisyphus (XIII)

#### Bewegungsgleichung der DNA-Moleküle im Gel

Zurückgelegter Weg s in Abhängigkeit von der Zeit t

$$s(t) = v \cdot t$$

$$s(t) = \frac{q \cdot E}{6 \cdot \pi \cdot \eta \cdot \left(\frac{3 \cdot m}{4 \cdot \pi \cdot \rho}\right)^{\frac{1}{3}} \cdot G} \cdot t$$

$$= \frac{q}{6 \cdot \pi \cdot \left(\frac{3}{4 \cdot \pi \rho}\right)^{\frac{1}{3}} \cdot G} \cdot \frac{E}{\eta} \cdot \frac{1}{m^{\frac{1}{3}}} \cdot t$$

$$idealisierend als Laufkonstante aufgefasst Konstante
$$= C \cdot \frac{E}{\eta} \cdot \frac{1}{m^{\frac{1}{3}}} \cdot t$$$$

⇒ Laufabweichungen (Seiteneffekteinflüsse) durch randomisiertes Variieren von DNA-Molekülmassen *m* nachbilden

Sisyphus (XIII)

#### Bewegungsgleichung der DNA-Moleküle im Gel

Zurückgelegter Weg s in Abhängigkeit von der Zeit t

$$s(t) = v \cdot t$$

$$s(t) = \frac{q \cdot E}{6 \cdot \pi \cdot \eta \cdot \left(\frac{3 \cdot m}{4 \cdot \pi \cdot \rho}\right)^{\frac{1}{3}} \cdot G} \cdot t$$

$$= \frac{q}{6 \cdot \pi \cdot \left(\frac{3}{4 \cdot \pi \rho}\right)^{\frac{1}{3}} \cdot G} \cdot \underbrace{\frac{E}{\eta}}_{\text{idealisierend als Laufkonstante aufgefasst}} \cdot \underbrace{\frac{1}{m^{\frac{1}{3}}}}_{\text{DNA-unabhängige Konstante}} \cdot \underbrace{\frac{1}{m^{\frac{1}{3}}}}_{\text{DNA-abhängige Konstante}}$$

⇒ Laufabweichungen (Seiteneffekteinflüsse) durch randomisiertes Variieren von DNA-Molekülmassen *m* nachbilden

Sisyphus (XIII)

#### Bewegungsgleichung der DNA-Moleküle im Gel

Zurückgelegter Weg s in Abhängigkeit von der Zeit t

$$s(t) = v \cdot t$$

$$s(t) = \frac{q \cdot E}{6 \cdot \pi \cdot \eta \cdot \left(\frac{3 \cdot m}{4 \cdot \pi \cdot \rho}\right)^{\frac{1}{3}} \cdot G} \cdot t$$

$$= \frac{q}{6 \cdot \pi \cdot \left(\frac{3}{4 \cdot \pi \rho}\right)^{\frac{1}{3}} \cdot G} \cdot \underbrace{\frac{E}{\eta}}_{\text{idealisierend als Lauf-konstante aufgefasst}} \underbrace{\frac{1}{m^{\frac{1}{3}}}}_{\text{DNA-unabhängige}} \underbrace{\frac{1}{m^{\frac{1}{3}}}}_{\text{DNA-abhängige}} \cdot t$$

$$= C \cdot \frac{E}{\eta} \cdot \frac{1}{m^{\frac{1}{3}}} \cdot t$$

⇒ Laufabweichungen (Seiteneffekteinflüsse) durch randomisiertes Variieren von DNA-Molekülmassen *m* nachbilden

Sisyphus (XIV)

#### Bestimmung der Molekülmasse aus der Bindungsmatrix

Nucleotid bzw.	chemische	Masse	Molekulargewicht
Strangendenmarkierung	Summenformel	$m_{ m X}$ in ${ m g}$	$M_{ m X}$ in $rac{ m g}{ m mol}$
A	$C_{10}H_{12}N_5O_6P$	$5,4668 \cdot 10^{-22}$	329, 2
С	C <sub>10</sub> H <sub>10</sub> N <sub>3</sub> O <sub>7</sub> P	$5,2337 \cdot 10^{-22}$	315, 2
G	$C_{10}H_{12}N_5O_7P$	$5,7324 \cdot 10^{-22}$	345, 2
T	$C_{10}H_{12}N_2O_8P$	$5,3004 \cdot 10^{-22}$	319, 2
Н	OH	$2,8243 \cdot 10^{-23}$	17,0
P	PO <sub>4</sub>	$1,5770 \cdot 10^{-22}$	95,0
В	$C_{10}H_{16}N_2O_3S$	$4,0572 \cdot 10^{-22}$	244,3

Bandenintensität von Anzahl Wasserstoffbrücken abhängig

Maßzahl:  $\#w = 2 \cdot \#AT + 3 \cdot \#CG$ 

Maßzahl in Graustufe (Helligkeitswert) umrechnen

Aufsummieren aller Helligkeitswerte über die Gellänge pro Zeitschritt

Sisyphus (XIV)

#### Bestimmung der Molekülmasse aus der Bindungsmatrix

Nucleotid bzw.	chemische	Masse	Molekulargewicht
Strangendenmarkierung	Summenformel	$m_{ m X}$ in ${ m g}$	$M_{ m X}$ in $rac{ m g}{ m mol}$
A	$C_{10}H_{12}N_5O_6P$	$5,4668 \cdot 10^{-22}$	329, 2
С	C <sub>10</sub> H <sub>10</sub> N <sub>3</sub> O <sub>7</sub> P	$5,2337 \cdot 10^{-22}$	315, 2
G	$C_{10}H_{12}N_5O_7P$	$5,7324 \cdot 10^{-22}$	345, 2
T	$C_{10}H_{12}N_2O_8P$	$5,3004 \cdot 10^{-22}$	319, 2
Н	OH	$2,8243 \cdot 10^{-23}$	17,0
P	PO <sub>4</sub>	$1,5770 \cdot 10^{-22}$	95,0
В	$C_{10}H_{16}N_2O_3S$	$4,0572 \cdot 10^{-22}$	244,3

#### Bandenintensität von Anzahl Wasserstoffbrücken abhängig

Maßzahl:  $\#w = 2 \cdot \#AT + 3 \cdot \#CG$ 

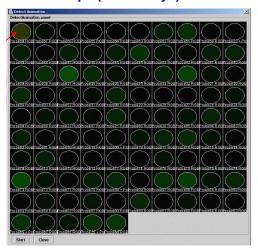
Maßzahl in Graustufe (Helligkeitswert) umrechnen

Aufsummieren aller Helligkeitswerte über die Gellänge pro Zeitschritt

#### Simulationsansatz - Detect

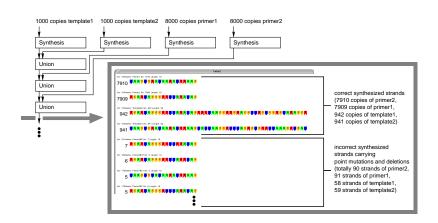
Sisyphus (XV)

#### Simulation von DNA-Chips (Microarrays)



R. Lehmann, B. Jambresic. Simulation der Arbeitsweise eines DNA-Chips. HTI Bern, 2005

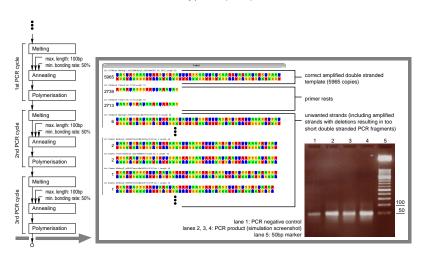
# Beispiel für Operationsfolgen – PCR Sisyphus (XVI)



⇒ Sichtbarmachen von Seiteneffekteinflüssen

# Beispiel für Operationsfolgen – PCR

Sisyphus (XVII)

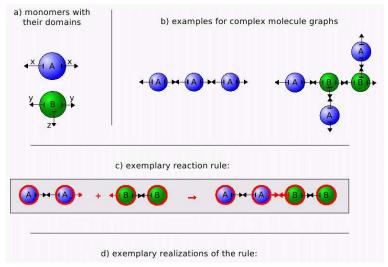


# Eigenschaften von Sisyphus

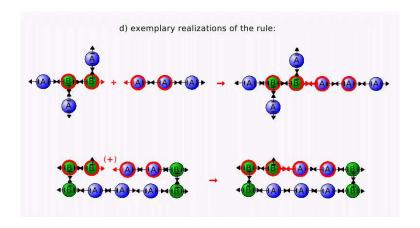
Sisyphus (XVIII)

	Olayphua (XVIII)											
Eiger	DNA-Operationen nschaften der beschriebenen Simulationen	Synthesis	Union	Split	Melting	Digestion	Labeling	Polymerisation	Affinity Purification	Annealing	Ligation	Gel Electrophoresis
	Simulationsergebnis primärstrukturabhängig											
eţe	Simulationsergebnis sekundärstrukturabhängig											
ausgewertete Parameter	Simulationergebnis von Masse der Objekte abhängig											
ausge	Simulationsergebnis von Ort oder Lage der Objekte abhängig											
	Simulationsergebnis von Bewegungsgeschwindigkeit der Objekte abhängig											
	statisch											
l .	dynamisch diskret											
imula	dynamisch kontinuierlich											
genutzte Simula- tionsmethoden	deterministisch (unter Vernachlässigung der Seiteneffekteinflüsse)											
	stochastisch (unter Vernachlässigung der Seiteneffekteinflüsse)											
	exogen (bzgl. Temperaturspezifität)											
berücksichtigte Seiteneffekte	Punktmutationen											
	Deletions											
	unerwünschte nichtlineare DNA (Artefakte)											
	Strangverluste											
berüc	unvollständiger Prozessverlauf											
"	unspezifische Wirkungen im Prozessverlauf											

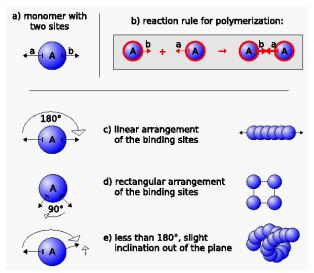
# SRSim: Simulator für mol. Dynamik im Raum (I)

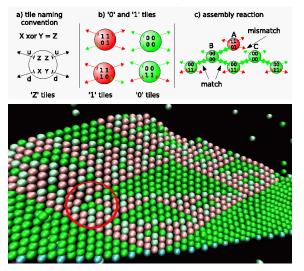


## SRSim: Simulator für mol. Dynamik im Raum (II)



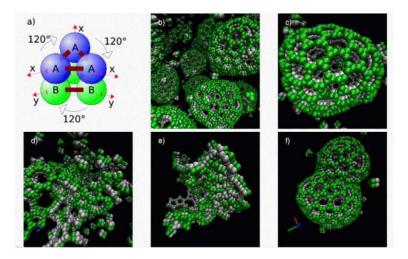
# SRSim: Simulator für mol. Dynamik im Raum (III)





G. Grünert, B. Ibrahim, T. Lenser, T. Hinze, P. Dittrich. Rule-Based Modeling in Space, BMC Bioinformatics, 2010

# SRSim: Simulator für mol. Dynamik im Raum (V)



# Zusammenfassung

- Simulation dient der kostengünstigen Vorbereitung und Optimierung von Laborexperimenten zur Implementierung molekularer Computer.
- Jede Simulation ist nur so gut wie das zugrunde liegende mathematische Modell und der verwendete Datenpool.
- Simulationen k\u00f6nnen immense Hardwareressourcen. beanspruchen.
- Verantwortung in der Auswahl/Kombination verwendeter Simulationsmethoden
- Hauptbestandteile: Modellbeschreibung, Parametererfassung, Rechenkern, Aufbereitung der Ergebnisse
- Entwicklung von Simulationstools einhergehend mit fortlaufender Validation und Verifikation