

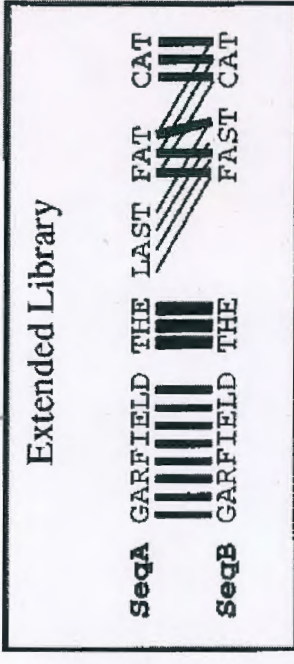
# Erweitern der Alignment-Sammlung

## Ursprüngliche Alignment-Sammlung

<b>SeqA</b>	GARFIELD THE LAST FAT CAT	<b>Prim. Weight = 88</b>	<b>SeqB</b>	GARFIELD THE VERY FAST CAT	<b>Prim. Weight = 100</b>
<b>SeqC</b>	GARFIELD THE LAST FAST CAT		<b>SeqC</b>	GARFIELD THE FAST CAT	
<b>SeqD</b>	GARFIELD THE LAST FA-T CAT	<b>Prim. Weight = 77</b>	<b>SeqD</b>	GARFIELD THE FA-T CAT	<b>Prim. Weight = 100</b>
<b>SeqA</b>	GARFIELD THE LAST FAT CAT	<b>Prim. Weight = 100</b>	<b>SeqC</b>	GARFIELD THE VERY FAST CAT	
<b>SeqB</b>	GARFIELD THE LAST FAST CAT		<b>SeqD</b>	GARFIELD THE FA-T CAT	

## Erweiterte Alignment-Sammlung

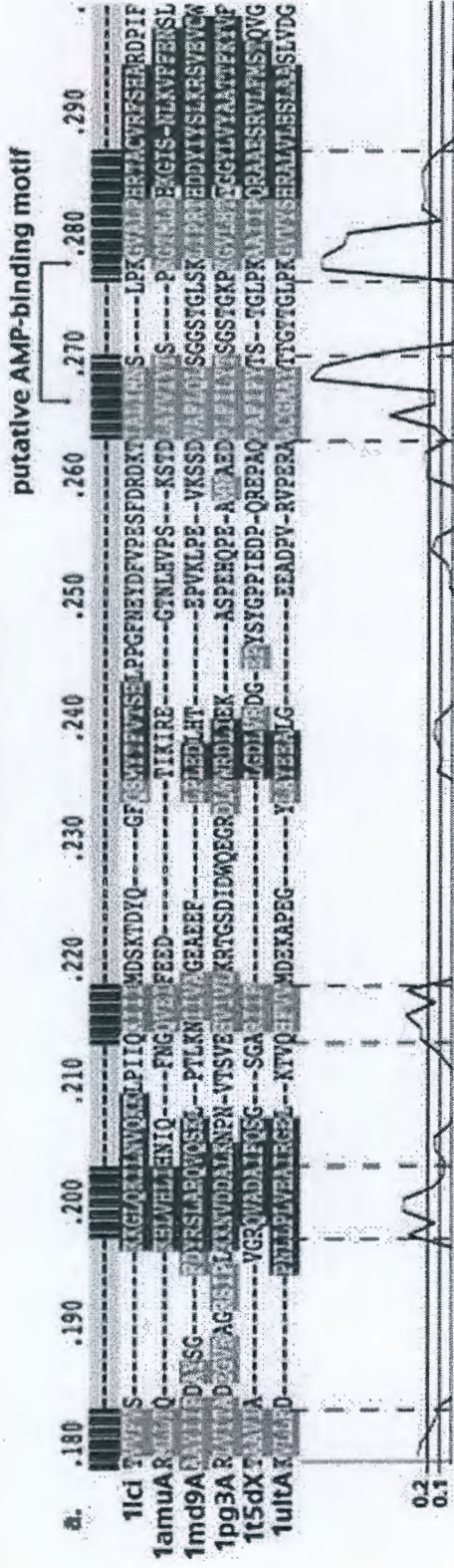
<b>SeqA</b>	GARFIELD THE LAST FAT CAT	<b>Weight = 88</b>
<b>SeqB</b>	GARFIELD THE FAST CAT	
<b>SeqA</b>	GARFIELD THE LAST FAT CAT	<b>Weight = 77</b>
<b>SeqC</b>	GARFIELD THE VERY FAST CAT	
<b>SeqB</b>	GARFIELD THE FAST CAT	
<b>Seq1</b>	GARFIELD THE LAST FAT CAT	<b>Weight = 100</b>
<b>SeqD</b>	GARFIELD THE FA-T CAT	
<b>SeqB</b>	GARFIELD THE VERY FAST CAT	



Dynamic Programming

**SeqA** GARFIELD THE LAST FA-T CAT  
**SeqB** GARFIELD THE ---- FAST CAT

# BALIiBASE



- Problem: Woher weiß ich, welche AS mit welchen AS aligniert werden müssen? Was ist die "biologische Wahrheit"?
- Antwort: Verwende 3D-Struktur-Alignments



## Evaluation auf BALiBASE

Aligner	RV11	RV12	RV20	RV30	RV40	RV50
MSAProbs	<b>74.63</b>	<b>94.86</b>	<b>94.35</b>	<b>88.20</b>	92.32	<b>90.90</b>
MUSCLE	65.75	92.32	91.50	84.23	86.31	85.28
MAFFT	69.18	93.68	93.62	87.81	<b>92.53</b>	90.14
Probalign	71.27	94.65	93.54	86.45	92.21	89.12
ProbCons	74.00	94.59	93.70	87.54	90.03	90.15
ClustalW	58.16	88.36	88.79	77.14	78.94	76.91

sum-of-pairs score (SPS): korrekt alignierte  
 Buchstabenpaare, in Prozent (von  $\binom{k}{2}N$ )

## Evaluation auf BALiBASE

Aligner	RV11	RV12	RV20	RV30	RV40	RV50
MSAProbs	<b>53.70</b>	<b>87.45</b>	<b>53.93</b>	<b>63.44</b>	61.04	<b>61.43</b>
MUSCLE	43.31	82.00	42.22	47.67	45.32	47.51
MAFFT	48.35	84.46	48.88	61.83	59.99	58.29
Probalign	48.57	86.77	46.69	59.72	<b>61.23</b>	54.36
ProbCons	52.76	86.82	50.80	60.05	53.61	59.52
ClustalW	32.53	75.58	33.86	38.17	39.82	36.50

column score (CS): *vollständig korrekt* alignierte Spalten,  
in Prozent (von  $N$ )



## Evaluation auf BALiBASE

Aligner	SPS	CS	Time (hh:mm:ss)
MSAProbs	<b>89.09</b>	<b>64.51</b>	1:12:56
MUSCLE	84.33	53.17	0:16:11
MAFFT	87.50	61.07	0:41:05
Probalign	87.78	60.68	8:05:35
ProbCons	88.31	61.89	5:29:15
ClustalW	78.65	44.75	0:18:56

Durchschnittliche Scores SPS und CS, Laufzeiten

(Liu et al., 2010)