

Visualization of pinfold simulations

Sebastian Pöttsch

Faculty of Mathematics and Computer Science
University of Leipzig

Overview

1. Introduction

1.1 Protein folding problem

1.2 HP-model

1.3 Pinfold simulation

2. Information Visualization

2.1 Shneiderman's mantra

2.2 Overview

2.3 Zoom and filter

2.4 Detail views

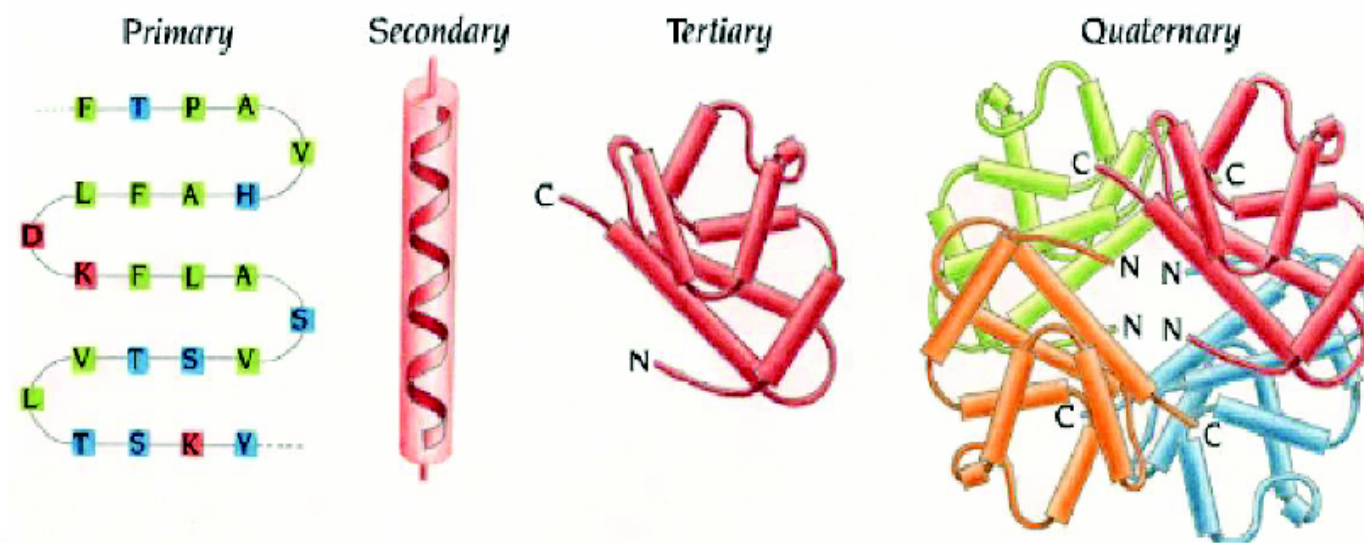
3. Pinfoldvis Demo

1. Introduction

- Proteins:
 - Complex macromolecules
 - Chains of amino acids linked by peptide bonds
 - Typical a few hundred amino acids
 - Essential functions in organisms
 - Function depends on the structure
 - Complex structures

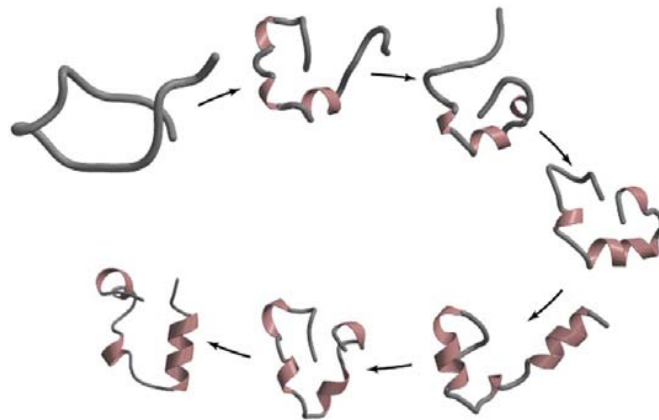
1. Introduction

- Protein structure:
 - Primary: amino acid sequence
 - Secondary: local conformations (α helices, β sheets, β turns)
 - Tertiary: real three dimensional structure
 - Quaternary: connection of polypeptide chains



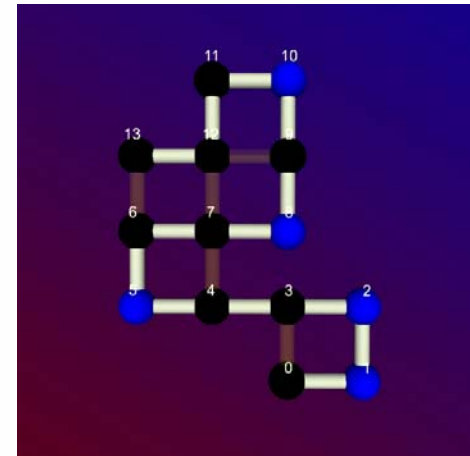
1.1 Protein folding problem

- Protein folding:
 - Problem: Given amino acid sequence → native structure
 - Not completely understood
 - Important to understand
 - Known forces (hydrogen bonds, hydrophobic interactions, Van der Waals force, disulfide bonds)



1.2 HP-Model

- By Dill and Lau, 1989
 - Amino acids → beads, unit size
 - Bonds → straight sticks, unit length
 - Positions fixed to lattice points
 - Two classes $H = \{A, C, I, L, M, F, W, Y, V\}$ and $P = \{R, N, D, E, Q, G, H, K, P, S, T\}$
 - Simple energy function
- HPNX-Model
 - H (hydrophobic), P (positive), N (negative), X (uncharged)

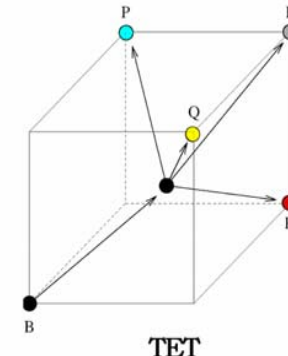
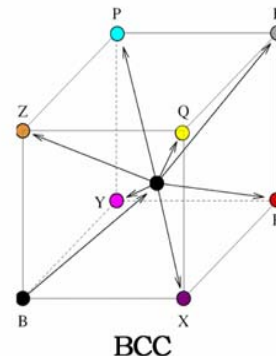
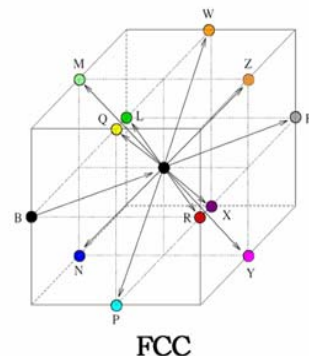
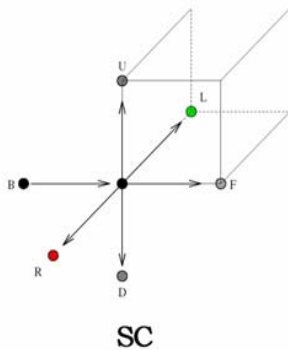
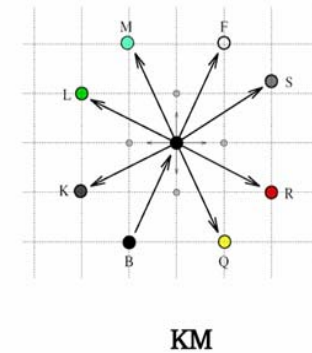
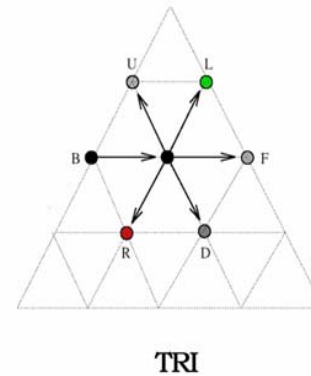
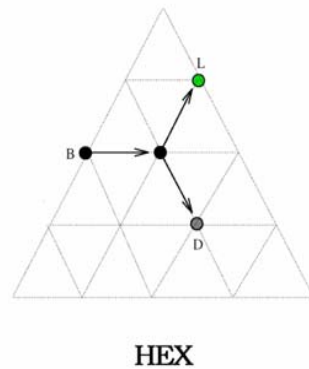
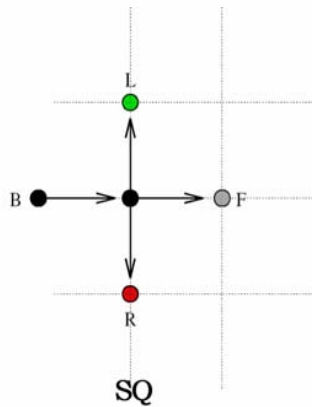
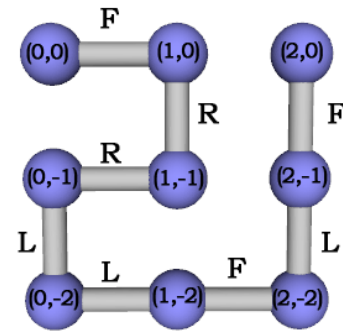


	H	P		
H	-1	0		
P	0	0		

	H	P	N	X
H	-4	0	0	0
P	0	1	-1	0
N	0	-1	1	0
X	0	0	0	0

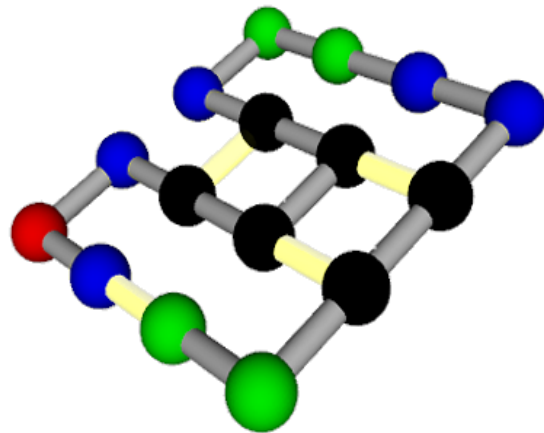
1.3 Pinfold simulation

- Based on HP/HPNX-Model
- Simulates folding on lattices
- Structure description with self-avoiding walks (SAW)



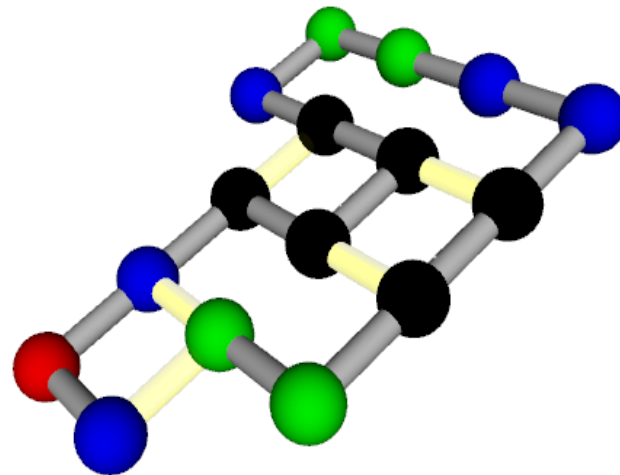
1.3 Pinfold simulation

- Example
 - Model: HPNX
 - Lattice: SQ
 - Sequence: NNHHPPNNPHHHHPXP



a)

a) Structure: FLFFLFFLLFRRFLL
Energy: -13

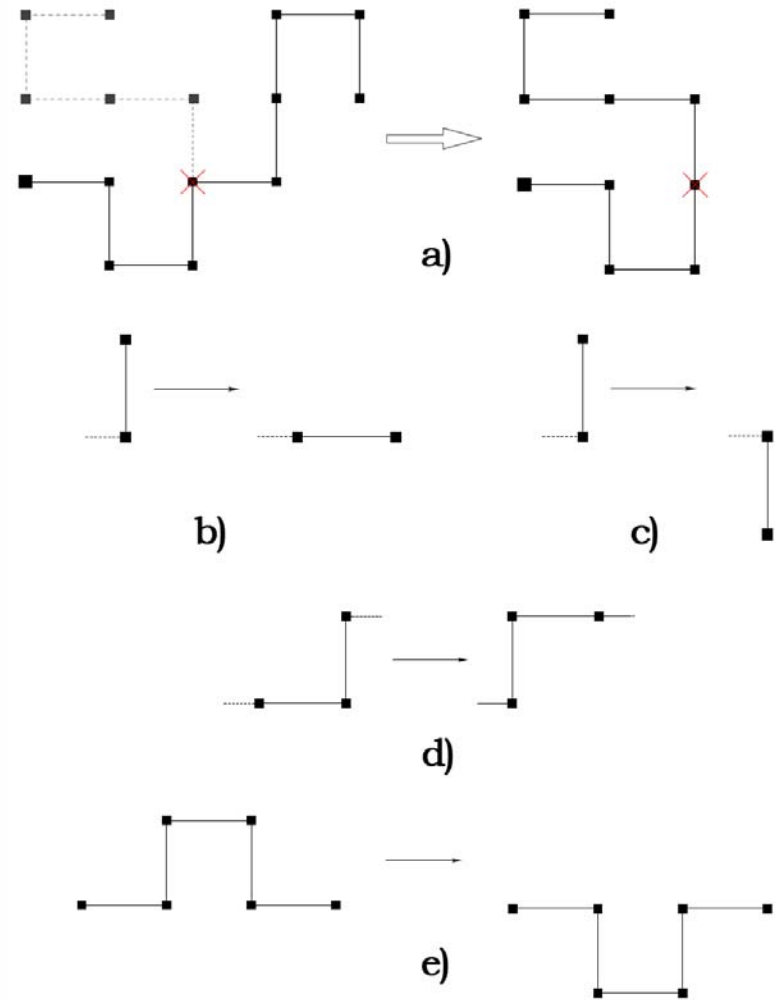


b)

b) Structure: FLFFLFFLLFRRLFL
Energy: -14

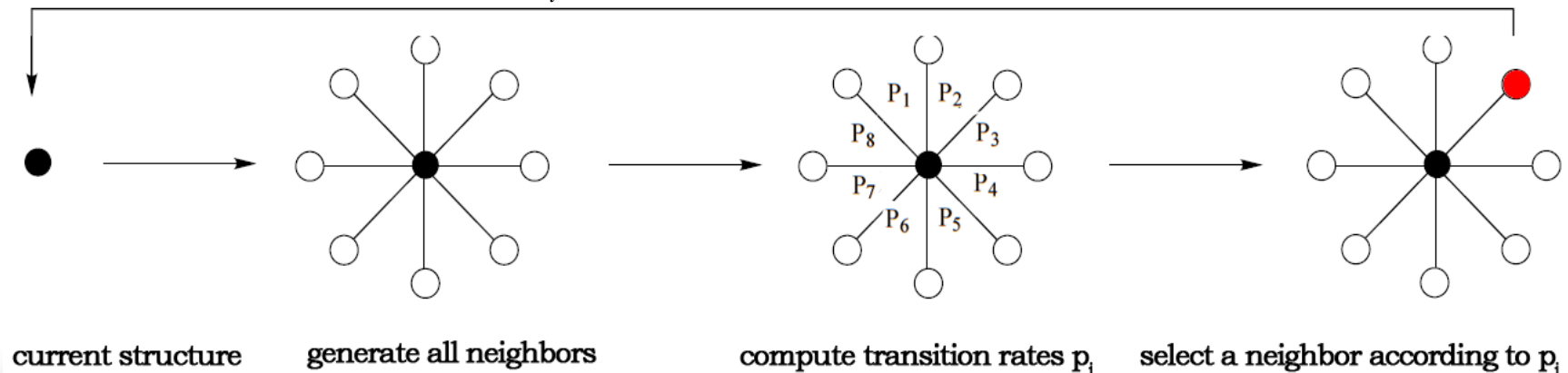
1.3 Pinfold simulation

- Move set
 - Set of rules
 - Moves have to be reversible
 - Results have to be in conformation space
 - Ergodicity
- describes topology of conformation space
- a) pivot move
- b) c) end move
- d) corner move
- e) crankshaft-move



1.3 Pinfold simulation

- Algorithm:
 1. Begin with start structure.
 2. Check terminating conditions
(max. time reached, stop structure found).
 3. Generate all valid neighbor structures with their energy.
 4. Compute transition rates. $p_i = \min\{1, \exp(-\frac{\Delta E}{kT})\}$
 5. Choose neighbor structure with probability proportional to p .
 6. Advance clock $\frac{1}{\sum_i p_i} \rightarrow$ Step 2



1.3 Pinfold simulation

- Output:
 - Sequence
 - Start structure
 - Chosen structure, Energy, Elapsed time
- **Our task:**
 - Visualize pinfold output
 - Support data analysis
 - Emphasize possible relationships
 - Provide simulation comparison

HHPHPHPPHPHPHPH		
FFLRLRFLLFFLRL		
FFFRLRFLLFFLRL	-4.00	0.056
FFFRLRFLLFFLRF	-3.00	0.735
FFFFRRFLLFFLRF	-2.00	2.520
FRFFLRFLLFFLRF	-4.00	2.849
FRFFLRFLLFFLRF	-4.00	3.521
FRFRLRFLLFFLRF	-3.00	3.689
FLFRLRFLLFFLRL	-4.00	4.444
FRFRLRFLLFFLRR	-3.00	4.587
FRFRLRFLLFFLRR	0.00	5.556
FRFRLRFLLFFLRR	-3.00	5.908
FLFRLRFLLFFLRL	-4.00	8.026
FFFRLRFLLFFLRL	-4.00	8.624
FFLRLRFLLFFLRF	-3.00	9.211
FFRRLRFLLFFLRL	-4.00	9.381
FFLRLRFLLFRLRL	-1.00	9.786
FFLRLRFLLRFLRL	0.00	11.068
FFLRFRLFFLRL	-2.00	12.552
...		

2. Information Visualization

- “to visualize”: form a mental vision, image, or picture of something not visible or present to sight, or an abstraction; to make visible to the mind or imagination. (The Oxford English Dictionary; 1989)
- Task:
 - Provide data analysis and exploration
 - Emphasize relationships
 - Uncover regularities
 - Expose the unseen (abstract)
 - Speedup cognition

2.1 Shneiderman's mantra

“Overview first, zoom and filter, details on demand.”

Idea:

- Start with an overview for a better orientation
- Let user decide to filter interesting data
- Show different details only on demand for selected data

2.2 Overview

“Overview first, zoom and filter, details on demand.”

Name: kh68_21to5
Date: Wed Sep 14 11:07:26 2005


	0	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5
Sequence	N	N	H	H	P	P	N	N	P	H	H	H	H	P	X	P
Start structure	F	L	F	F	L	F	F	L	L	F	R	R	F	L	L	
Stop structure	F	L	F	F	L	F	F	L	L	F	R	R	L	F	L	

Energy model:

Model: HPNX
dangle: 2
Temp: 37.0
Logml: logarithmic
Par: VRNA-1.4

Move set:

Lattice: SQ
NoShift: off
NoLP: off



Simulation:

Num: 1000
Time: 10000.00
Seed: clock
Mc: Metropolis
Fpt: on

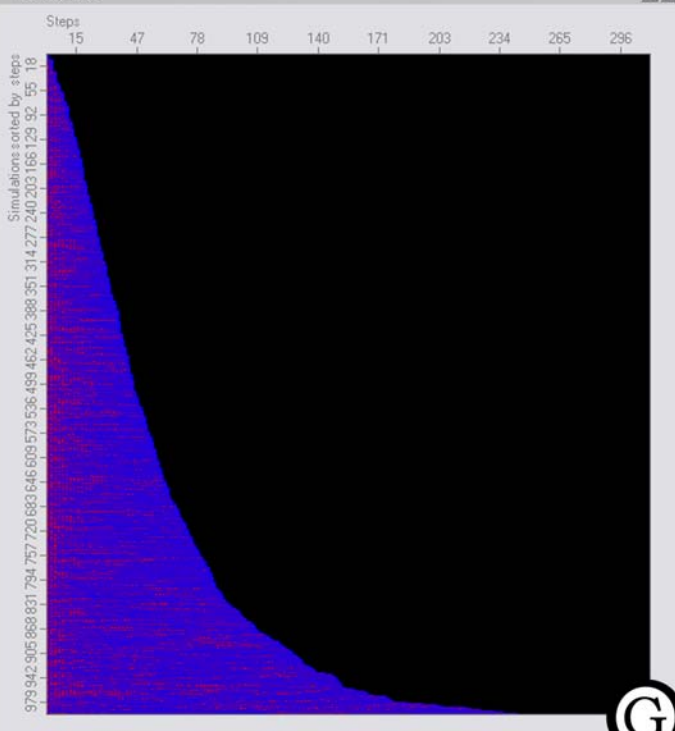
Output:

Log: kh68_21to5
Silent: on
Lmin: off
Cut: 20.00

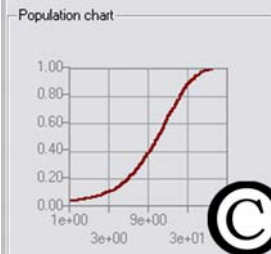
Overview table

Simulation Index	Time	Steps	Found	Energy Max	Energy Min
1	591	0.230	3	true	-11 -14
2	856	0.088	3	true	-11 -14
3	174	0.220	3	true	-11 -14
4	977	0.308	3	true	-11 -14
5	415	0.414	3	true	-11 -14
6	450	0.167	3	true	-11 -14
7	878	0.560	3	true	-11 -14
8	350	0.446	3	true	-11 -14
9	139	0.502	3	true	-11 -14
10	471	0.864	5	true	-3 -14
11	455	0.264	5	true	-11 -14
12	444	0.823	5	true	-11 -14
13	876	1.227	5	true	-11 -14
14	940	0.619	5	true	-11 -14
15	686	2.957	5	true	-3 -14
16	751	0.449	5	true	-11 -14
17	743	0.766	5	true	-11 -14
18	490	1.911	5	true	-11 -14
19	702	1.336	5	true	-3 -14
20	149	0.422	5	true	-3 -14
21	489	1.159	5	true	-3 -14
22	469	1.091	5	true	-3 -14
23	359	1.087	5	true	-3 -14
24	209	0.704	5	true	-3 -14
25	230	0.582	5	true	-11 -14
26	28	0.735	5	true	-11 -14
27	23	1.830	6	true	-11 -14
28	673	1.369	7	true	-10 -14
29	498	1.735	7	true	-3 -14
30	192	0.954	7	true	-11 -14

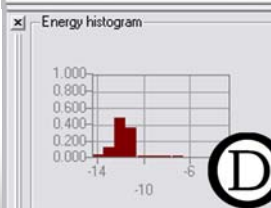
Energy map



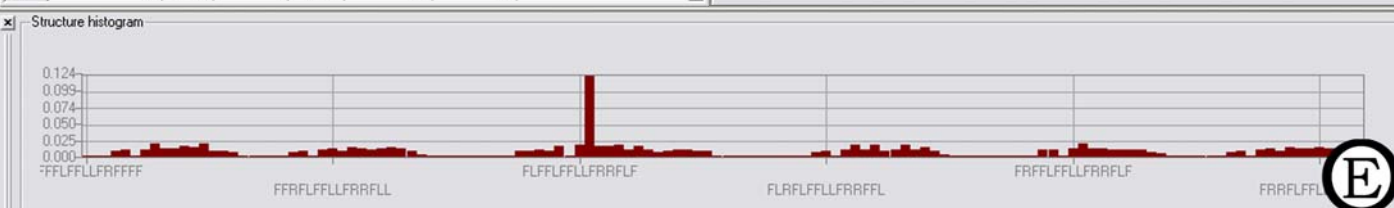
Population chart



Energy histogram



Structure histogram



2.2 Overview

“Overview first, zoom and filter, details on demand.”

Name: kh68_21to5 Date: Wed Sep 14 11:07:26 2005

Sequence: N N H H P P N N P H H H H P X P
Start structure: F L F F L F F L L F R R F L L
Stop structure: F L F F L F F L L F R R L F L

Energy model:
Model: HPNX
dangle: 2
Temp: 37.0
Logmi: logarithmic
Par: RNA-1.4

Move set:
Lattice: 2D
NoShift: off
NoLP: off

Simulation:
Num: 1000
Time: 10000.00
Seed: clock
Mc: Metropolis
Fpt: on

Output:
Log: kh68_21to5
Silent: on
Umin: off
Cut: 20.00

Population chart

Simulation Index	Time	Steps	Found	Energy Max	Energy Min
1	591	0.230	3	true	-11 -14
2	856	0.098	3	true	-11 -14
3	174	0.220	3	true	-11 -14
4	377	0.230	3	true	-11 -14
5	450	0.167	3	true	-11 -14
6	878	0.560	3	true	-11 -14
7	350	0.446	3	true	-11 -14
8	138	0.502	3	true	-11 -14
9	471	0.864	5	true	-3 -14
10	455	0.264	5	true	-11 -14
11	444	0.823	5	true	-11 -14
12	876	1.227	5	true	-11 -14
13	940	0.619	5	true	-11 -14
14	686	2.957	5	true	-3 -14
15	751	0.449	5	true	-11 -14
16	743	0.766	5	true	-11 -14
17	490	1.911	5	true	-11 -14
18	702	1.336	5	true	-3 -14
19	149	0.422	5	true	-3 -14
20	489	1.159	5	true	-3 -14
21	463	1.091	5	true	-3 -14
22	359	1.087	5	true	-3 -14
23	209	0.704	5	true	-3 -14
24	230	0.582	5	true	-11 -14
25	28	0.735	5	true	-11 -14
26	23	1.830	6	true	-11 -14
27	673	1.369	7	true	-10 -14
28	498	1.735	7	true	-3 -14
29	192	0.954	7	true	-11 -14
30					

Energy map

Energy histogram

Structure histogram

Annotations: A, B, C, D, E, F, G

2.2 Overview

“Overview first, zoom and filter, details on demand.”

Name: kh68_21to5
Date: Wed Sep 14 11:07:26 2005

```

Sequence: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
          G H H P P H H H H P X P
Start structure: F L F L F L F L F L F L F L F L
Stop structure: F L F L F L F L F L F L F L
                
```


A

Energy model:

Model: HPNX
dangle: 2
Temp: 37.0
Logml: logarithmic
Par: VRNA-1.4

Move set:

Lattice: SQ
NoShift: off
NoLP: off



Simulation:

Num: 1000
Time: 10000.00
Seed: clock
Mc: Metropolis
Fpt: on

Output:

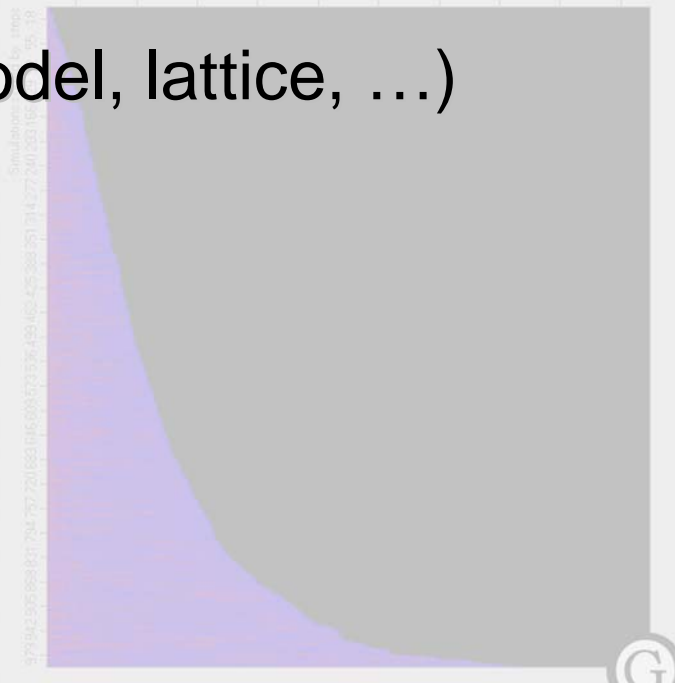
Log: kh68_21to5
Silent: on
Lmin: off
Cut: 20.00

B

Overview table

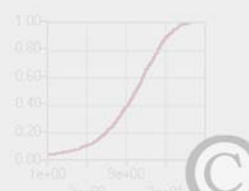
Simulation Index	Time	Steps	Found	Energy Max	Energy Min
1	591	0.230	3	true	-11 -14
2	856	0.098	3	true	-11 -14
3	174	0.220	3	true	-11 -14
6	450	0.167	3	true	-11 -14
7	878	0.560	3	true	-11 -14
8	350	0.446	3	true	-11 -14
9	139	0.502	3	true	-11 -14
10	471	0.864	5	true	-3 -14
11	455	0.264	5	true	-11 -14
12	444	0.823	5	true	-11 -14
13	876	1.227	5	true	-11 -14
14	940	0.619	5	true	-11 -14
15	686	2.957	5	true	-3 -14
16	751	0.449	5	true	-11 -14
17	743	0.766	5	true	-11 -14
18	490	1.911	5	true	-11 -14
19	702	1.336	5	true	-3 -14
20	149	0.422	5	true	-3 -14
21	489	1.159	5	true	-3 -14
22	463	1.091	5	true	-3 -14
23	359	1.087	5	true	-3 -14
24	209	0.704	5	true	-3 -14
25	230	0.582	5	true	-11 -14
26	28	0.735	5	true	-11 -14
27	23	1.830	6	true	-11 -14
28	673	1.369	7	true	-10 -14
29	498	1.735	7	true	-3 -14
30	192	0.954	7	true	-11 -14

Energy map



G

Pinfold parameters (model, lattice, ...)


Population chart


C

Energy histogram


D

Structure histogram


E

F

2.2 Overview

“Overview first, zoom and filter, details on demand.”

The screenshot displays a simulation software interface with the following components:

- Name:** kh68_21to5
- Date:** Wed Sep 14 11:07:26 2005
- Sequence:** H H H P P P H H H H P X P
- Start structure:** F L F F L F F L L F H L L L
- Stop structure:** F L F F L F F L L F H L L L
- Energy model:** HPNX, dangle 2, Temp 37.0, LogMf logarithmic, Par VRNA-1.4
- Move set:** SQ, Lattice, NoShift off, NoLP off
- Simulation:** Num 1000, Time 10000.00, Seed clock, Mc Metropolis, Fpt on
- Output:** Log kh68_21to5, Silent on, Lmin off, Cut 20.00
- Overview table:** A table with columns: Simulation Index, Time, Steps, Found, Energy Max, Energy Min. It lists 30 simulation steps.
- Energy map:** A plot showing energy vs. simulation index, with a purple shaded area under a curve.
- Population chart:** A graph showing a sigmoidal curve from 0.00 to 1.00 over time.
- Energy histogram:** A bar chart showing the distribution of energy values.
- Structure histogram:** A bar chart showing the distribution of different structures, with a prominent peak for a specific structure.

Annotations A through G are placed on the interface:

- A:** Top right corner.
- B:** Output settings panel.
- C:** Population chart.
- D:** Energy histogram.
- E:** Structure histogram.
- F:** Overview table.
- G:** Energy map.

Stop structure population

2.2 Overview

“Overview first, zoom and filter, details on demand.”

Name: kh68_21to5 Date: Wed Sep 14 11:07:26 2005

Sequence: H H H P P P H H H H P X P
Start structure: F L F F L F F L L F L L L L
Stop structure: F L F F L F F L L F L L L L

Energy model:
Model: HPNX
dangle: 2
Temp: 37.0
LogM: logarithmic
Par: VRNA-1.4

Move set:
Lattice: SQ
NoShift: off
NoLP: off

Simulation:
Num: 1000
Time: 10000.00
Seed: clock
Mc: Metropolis
Fpt: on

Output:
Log: kh68_21to5
Silent: on
Umin: off
Cut: 20.00

Population chart

Simulation Index	Time	Steps	Found	Energy Max	Energy Min
1	591	0.230	3	true	-11 -14
2	856	0.098	3	true	-11 -14
3	174	0.220	3	true	-11 -14
4					
5					
6	450	0.167	3	true	-11 -14
7	878	0.560	3	true	-11 -14
8	350	0.446	3	true	-11 -14
9	138	0.502	3	true	-11 -14
10	471	0.864	5	true	-3 -14
11	455	0.264	5	true	-11 -14
12	444	0.823	5	true	-11 -14
13	876	1.227	5	true	-11 -14
14	940	0.619	5	true	-11 -14
15	686	2.957	5	true	-3 -14
16	751	0.449	5	true	-11 -14
17	743	0.766	5	true	-11 -14
18	490	1.911	5	true	-11 -14
19	702	1.336	5	true	-3 -14
20	149	0.422	5	true	-3 -14
21	489	1.159	5	true	-3 -14
22	463	1.091	5	true	-3 -14
23	359	1.087	5	true	-3 -14
24	208	0.704	5	true	-3 -14
25	230	0.582	5	true	-11 -14
26	28	0.735	5	true	-11 -14
27	23	1.830	6	true	-11 -14
28	673	1.369	7	true	-10 -14
29	498	1.735	7	true	-3 -14
30	192	0.954	7	true	-11 -14

Energy map

Energy and structure histograms

Energy histogram

Structure histogram

The screenshot displays a simulation software interface with several panels. At the top, the simulation name 'kh68_21to5' and date 'Wed Sep 14 11:07:26 2005' are shown. Below this is a sequence of nucleotides and their corresponding start and stop structures. The main area is divided into several sections: 'Energy model' (HPNX, dangle: 2, Temp: 37.0, LogM: logarithmic, Par: VRNA-1.4), 'Move set' (SQ, NoShift: off, NoLP: off), 'Simulation' (Num: 1000, Time: 10000.00, Seed: clock, Mc: Metropolis, Fpt: on), and 'Output' (Log: kh68_21to5, Silent: on, Umin: off, Cut: 20.00). A 'Population chart' shows a sigmoidal curve. The 'Overview table' is a central table with columns for Simulation Index, Time, Steps, Found, Energy Max, and Energy Min, containing 30 rows of simulation data. To the right, the 'Energy map' shows a heatmap of energy values over simulation steps. At the bottom, there are two histograms: 'Energy histogram' showing a distribution of energy values, and 'Structure histogram' showing the frequency of different RNA structures. A large text overlay 'Energy and structure histograms' is centered over the Overview table and Energy map. Several circular callouts (A-G) are placed around the interface to highlight specific features.

2.3 Zoom and filter

“Overview first, zoom and filter, details on demand.”

Name: kh68_21to5 Date: Wed Sep 14 11:07:26 2005

Sequence: H H H P P P H H H H P X P
 Start structure: F L F L F L F L L F L L L L L
 Stop structure: F L F L F L F L L F L L L L L

Energy model:
 Model: HPNX
 dangle: 2
 Temp: 37.0
 LogM: logarithmic
 Par: VRNA-1.4

Move set:
 Lattice: SQ
 NoShift: off
 NoLP: off

Simulation:
 Num: 1000
 Time: 10000.00
 Seed: clock
 Mc: Metropolis
 Fpt: on

Output:
 Log: kh68_21to5
 Silent: on
 Umin: off
 Cut: 20.00

Population chart

Simulation Index	Time	Steps	Found	Energy Max	Energy Min
1	591	0.230	3	true	-11 -14
2	856	0.088	3	true	-11 -14
3	174	0.220	3	true	-11 -14
4	977	0.308	3	true	-11 -14
5	415	0.414	3	true	-11 -14
6	450	0.167	3	true	-11 -14
7	878	0.560	3	true	-11 -14
8	350	0.446	3	true	-11 -14
9	139	0.502	3	true	-11 -14
10	471	0.864	5	true	-3 -14
11	455	0.264	5	true	-11 -14
12	444	0.823	5	true	-11 -14
13	876	1.227	5	true	-11 -14
14	940	0.619	5	true	-11 -14
15	686	2.957	5	true	-3 -14
16	751	0.449	5	true	-11 -14
17	743	0.766	5	true	-11 -14
18	490	1.911	5	true	-11 -14
19	702	1.336	5	true	-3 -14
20	149	0.422	5	true	-3 -14
21	489	1.159	5	true	-3 -14
22	469	1.091	5	true	-3 -14
23	359	1.087	5	true	-3 -14
24	209	0.704	5	true	-3 -14
25	230	0.582	5	true	-11 -14
26	28	0.735	5	true	-11 -14
27	23	1.830	6	true	-11 -14
28	673	1.369	7	true	-10 -14
29	498	1.735	7	true	-3 -14
30	192	0.954	7	true	-11 -14

Energy map

Steps: 15, 47, 78, 109, 140, 171, 203, 234, 265, 296

Simulation sorted by steps: 979 942 905 868 831 794 757 720 683 646 609 573 536 499 462 425 388 351 314 277 240 203 166 129 92 55 18

Energy histogram

Structure histogram

Energy histogram: 1000, 800, 600, 400, 200, 0.000

Structure histogram: 0.074, 0.050, 0.025, 0.000

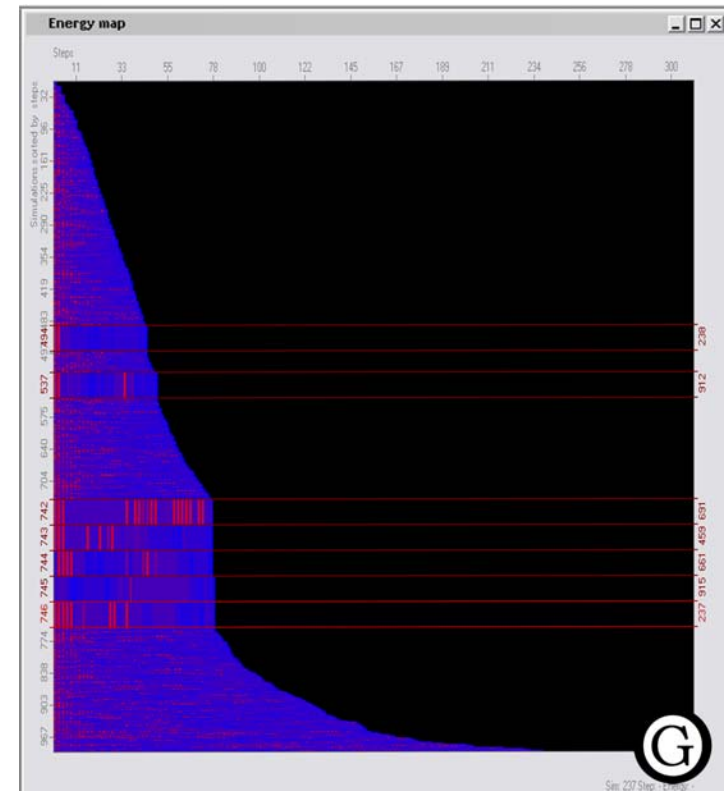
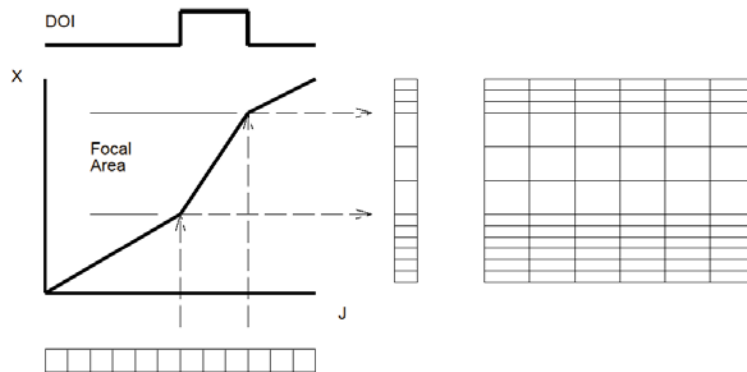
Structure histogram labels: FFLFLFLFLFLFL, FFALFLFLFLFLFL, FLFLFLFLFLFLFL, FLALFLFLFLFLFL, FFALFLFLFLFLFL, FFLFLFLFLFLFL

Data selection (overview table, energy map)

2.3 Zoom and filter

“Overview first, zoom and filter, details on demand.”

- Energy map
- Focus+context technique
- Huge data sets ↔ limited screen size
- Details and overview in one view
- Hold up users attention as long as possible



2.4 Detail views

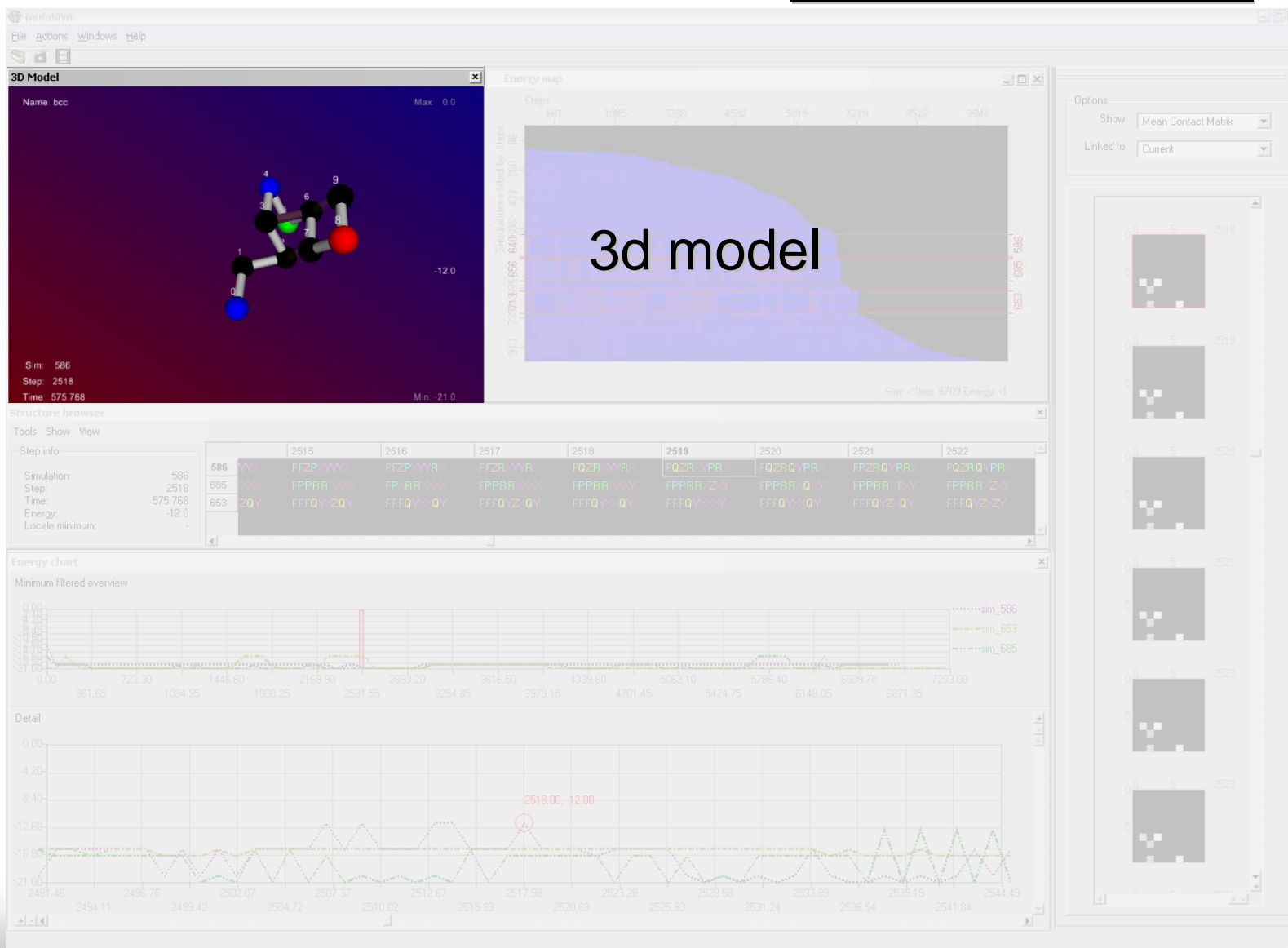
“Overview first, zoom and filter, details on demand.”

■ Multiple Views

- Two or more distinct views
- Support the investigation of a single conceptual entity
- Views can differ in data or visual representation
- Different perspectives
- Allows direct comparison in real-time
- Linking several views emphasizes relationships

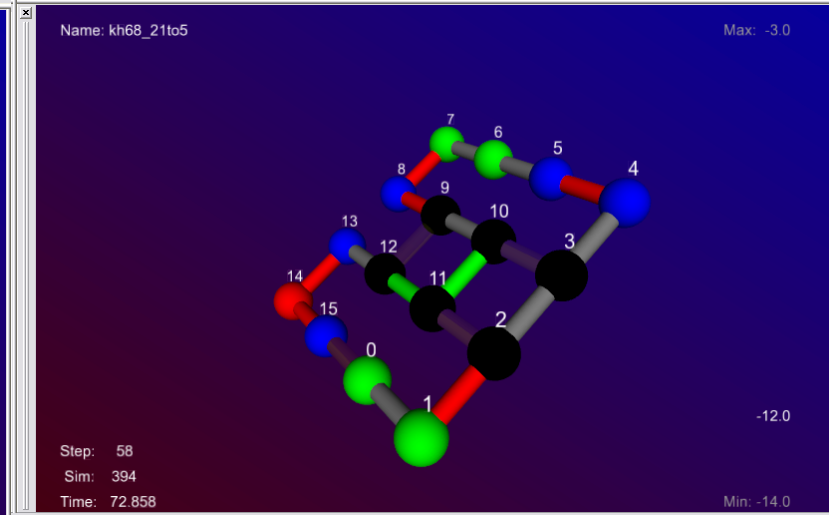
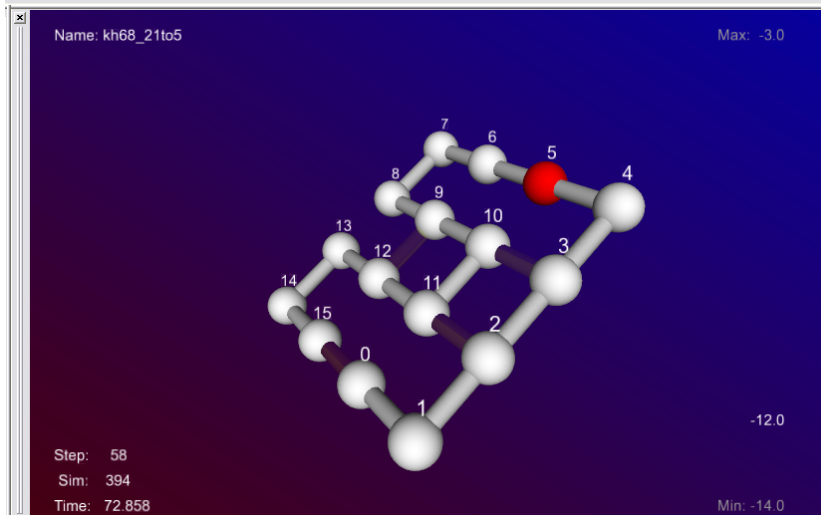
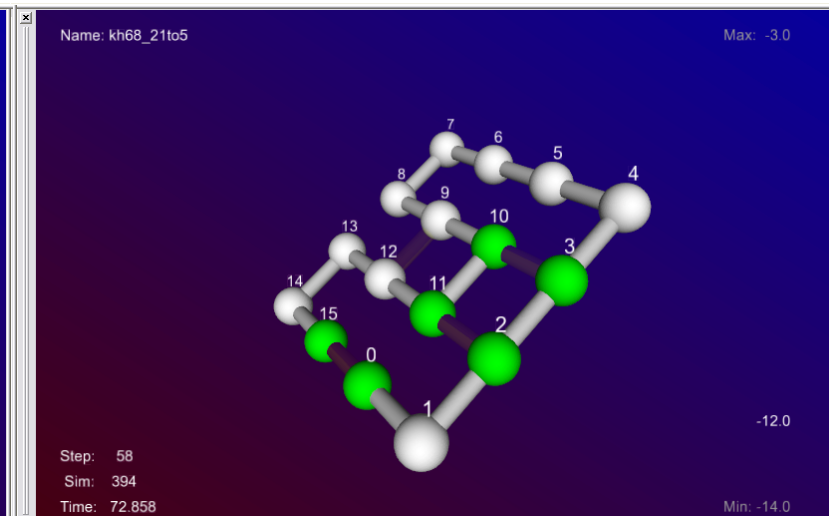
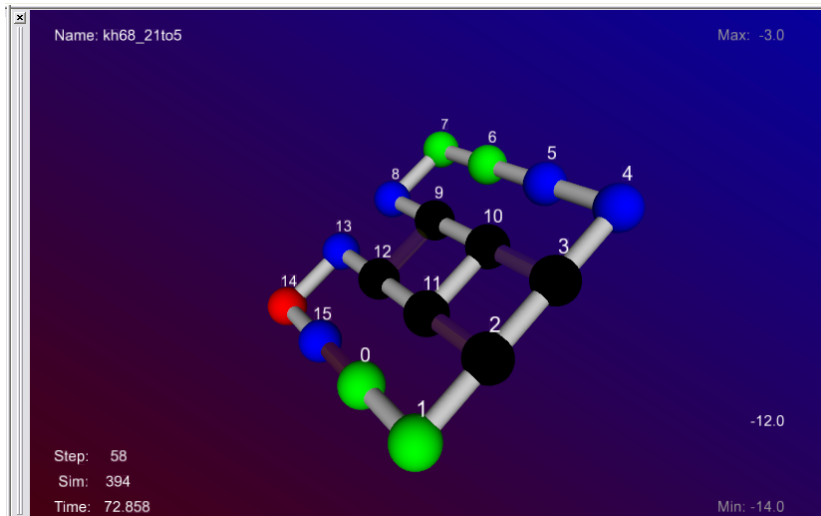
2.4 Detail views

“Overview first, zoom and filter, details on demand.”



2.4 Detail views

- 3D structure viewer



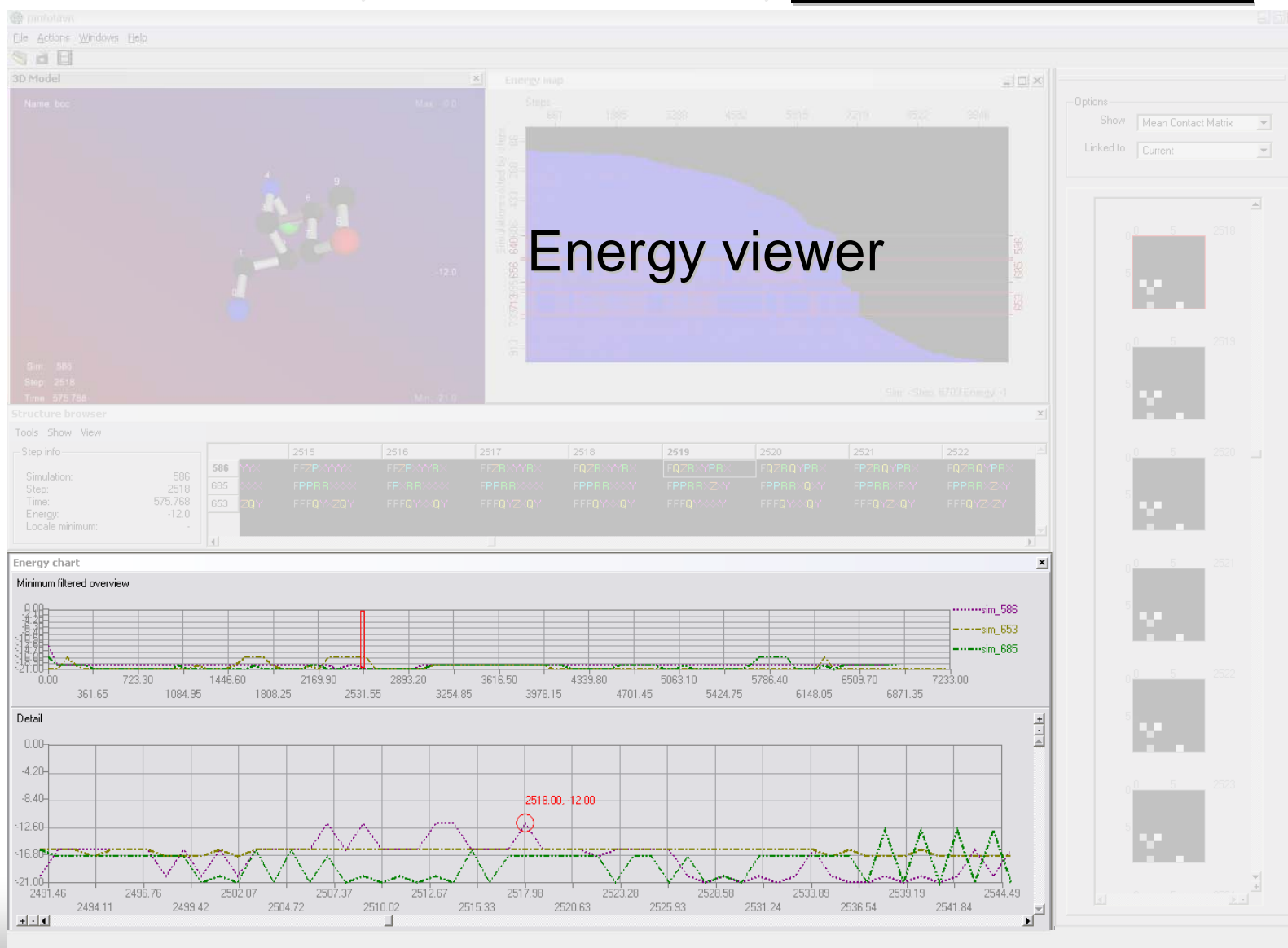
2.4 Detail views

“Overview first, zoom and filter, details on demand.”



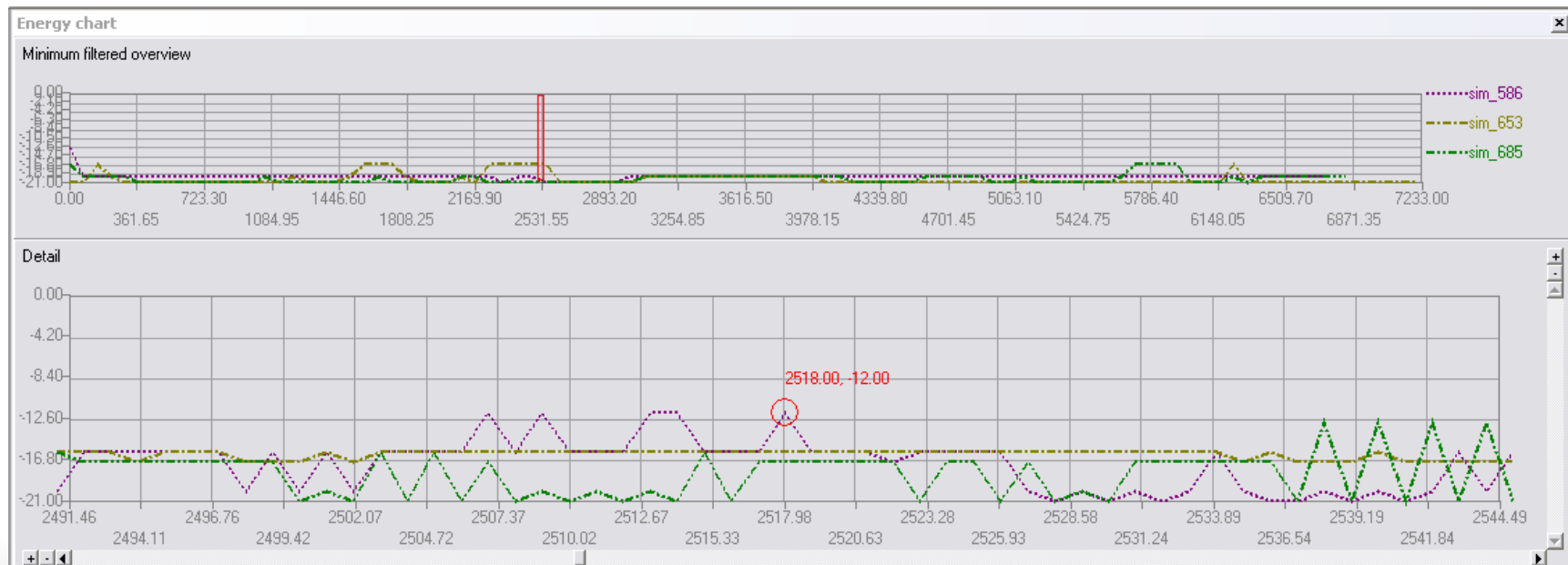
2.4 Detail views

“Overview first, zoom and filter, details on demand.”



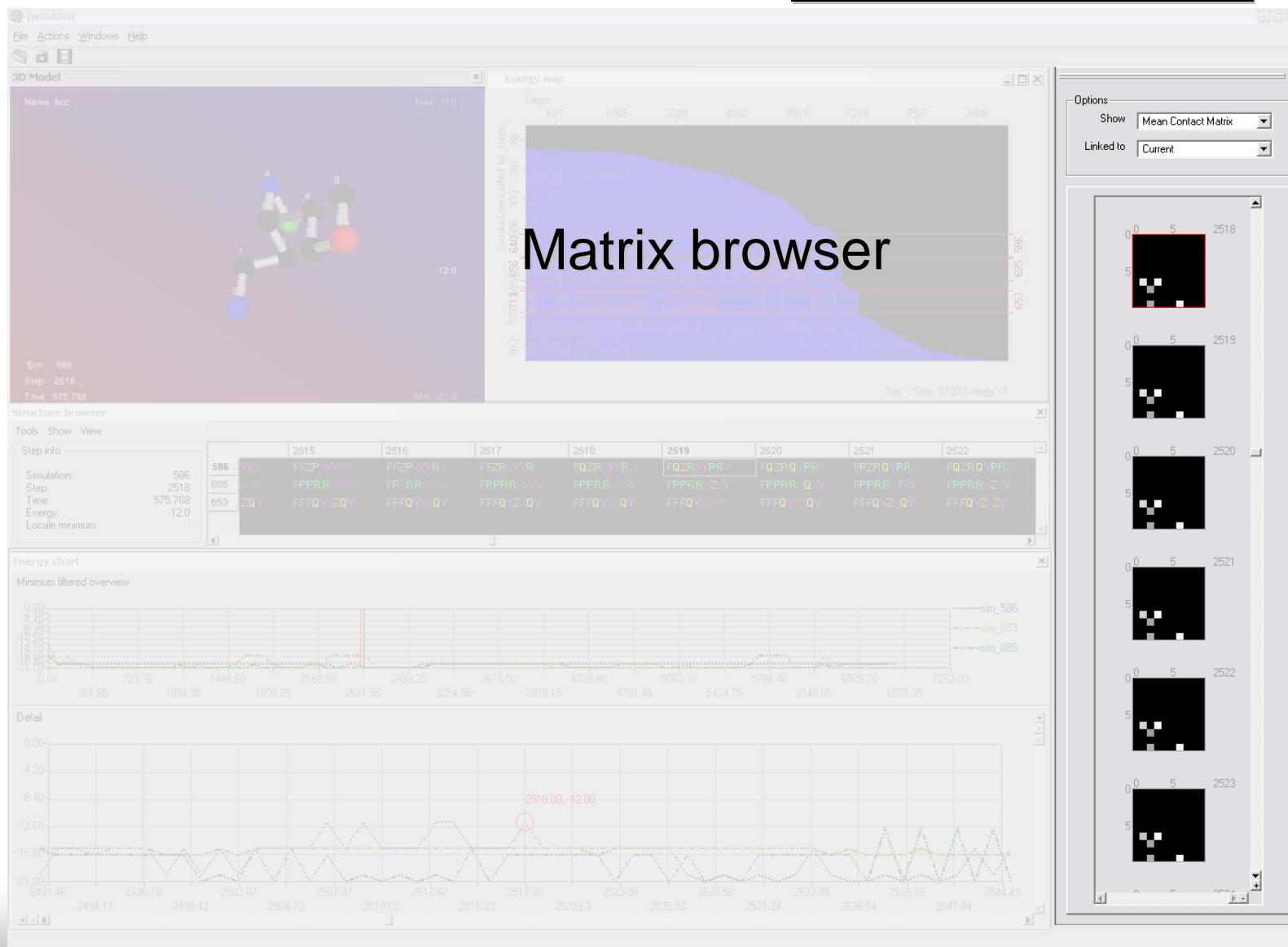
2.4 Detail views

- Energy viewer
- Overview+detail technique
 - Preserve the overview while exploring details
 - Space multiplexing or time multiplexing
 - Geometric zoom or semantic zoom



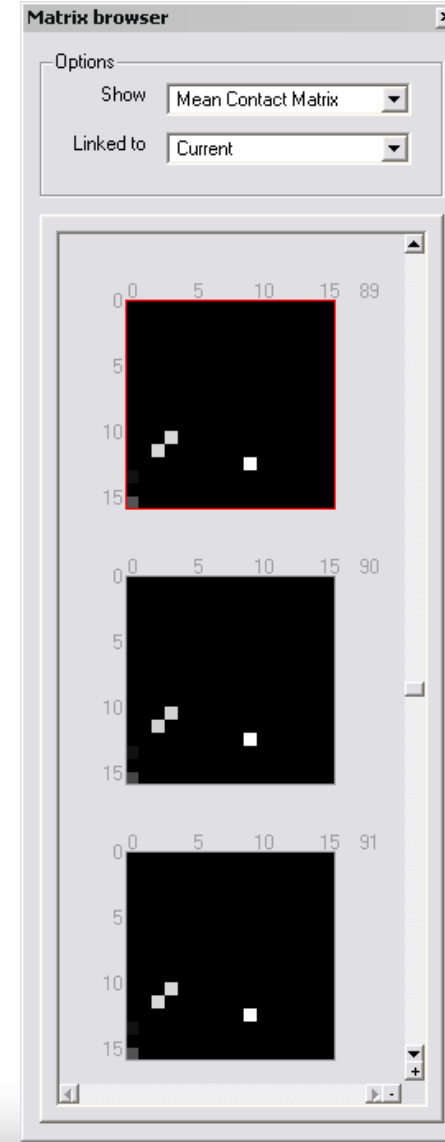
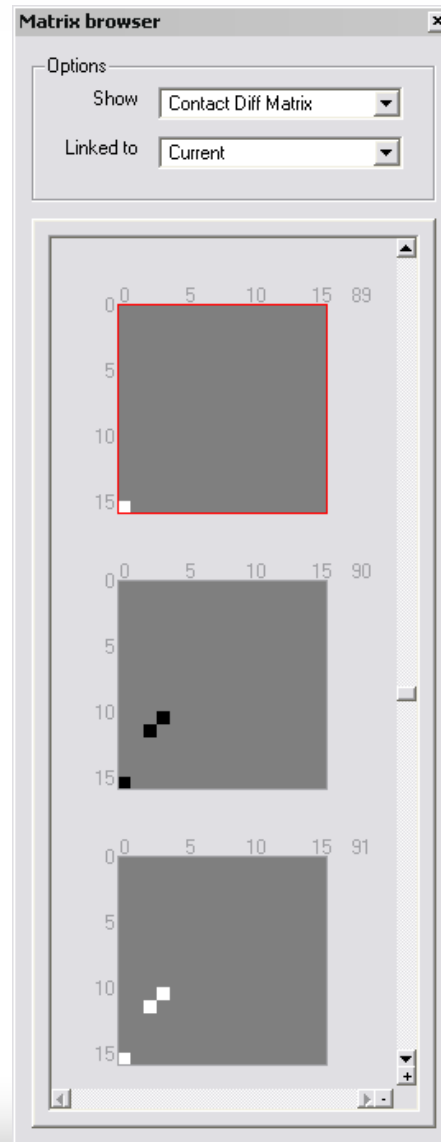
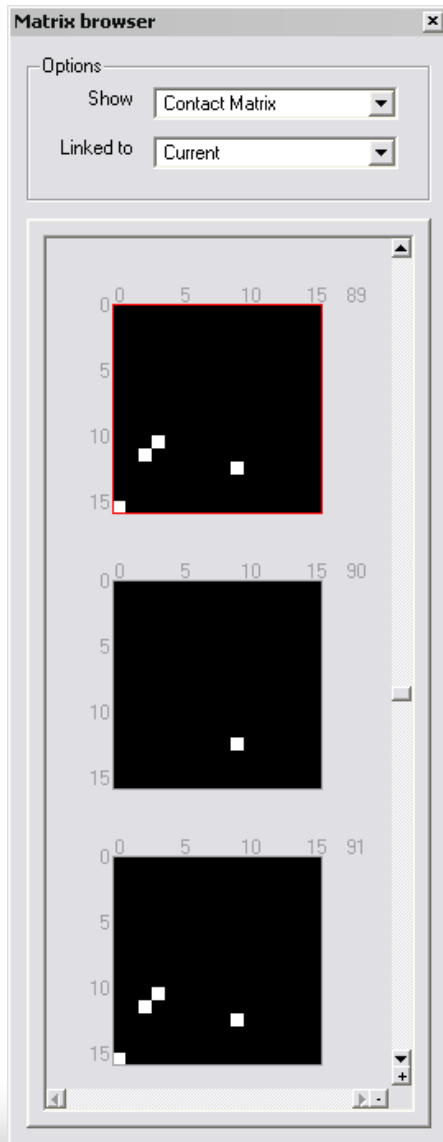
2.4 Detail views

“Overview first, zoom and filter, details on demand.”



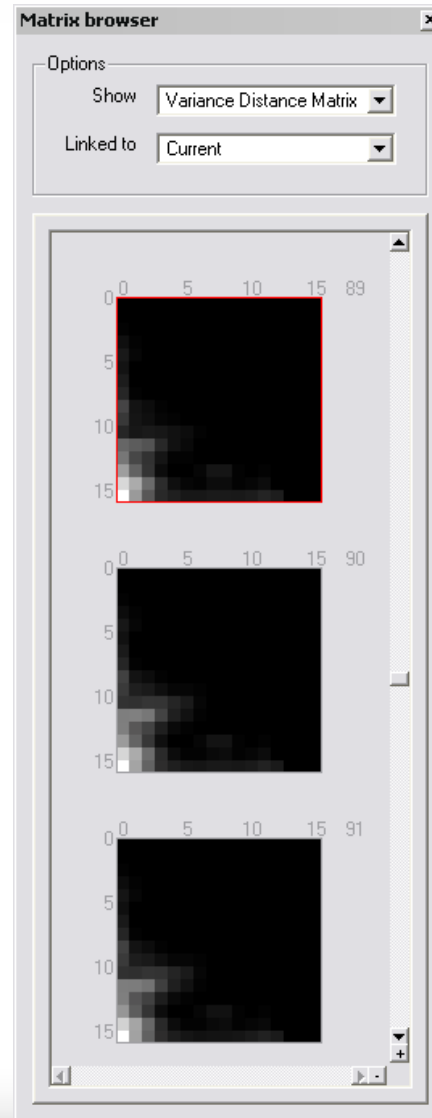
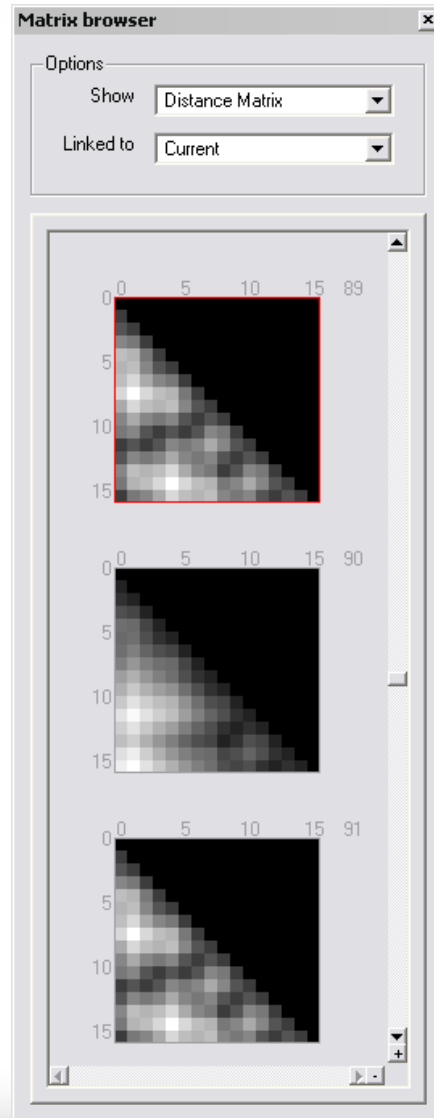
2.4 Detail views

- Matrix browser



2.4 Detail views

- Matrix browser



3. Pinfoldvis Demo

- Tool demonstration

4. Bibliography

- *Baldonado, M. Q. W., Woodruff, A., and Kuchinsky, A. Guidelines for using multiple views in information visualization. In Advanced Visual Interfaces (2000), pp. 110–119.*
- *Dill, K. A. Polymer principles and protein folding. Protein Science 8 (1999), 1166–1180.*
- *Card, S. K., Mackinlay, J. D., and Shneiderman, B., Eds. Readings in Information Visualization — Using Vision to Think. Morgan Kaufmann, 1999.*
- *Will, S. Exact, constraint-based structure prediction in simple protein models. PhD thesis, Universität Jena, 2005.*
- *Wolfinger, M. Energy Landscapes of Biopolymers. PhD thesis, Fakultät für Chemie, Universität Wien, 2004.*