Proteins, remote homologies, …

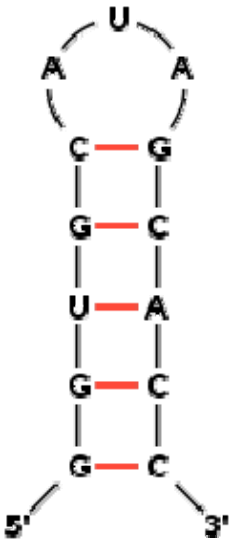
Andrew Torda, ZBH

- What do we do in Bundesstrasse ?
  - numerical methods
    - simulations - proteins, RNA, evolution
    - design of molecules – RNA
    - structure prediction – RNA, proteins
    - MD simulations (DNA + protein)

- Today mostly proteins
  - bit of RNA

RNA – design and structure

- Unusual spaces and dynamics
- Design molecules
  - useful structure (ribozyme, riboswitch, …)
  - find a new sequence which preserves structure
- energy functions ? Literature conventions + tricks



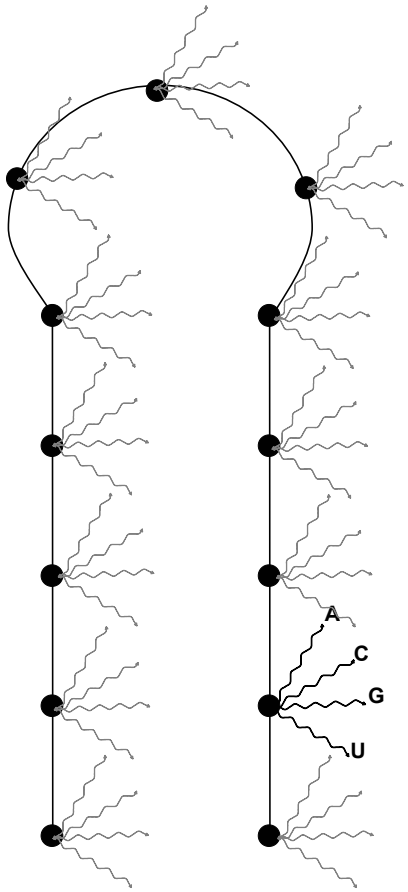
- particles can be A, C, G, U
  - search space = $4 \times 4 \times 4 \ldots = 4^N$
  - discrete problem ? maybe not

RNA design

- particles can be of mixed types

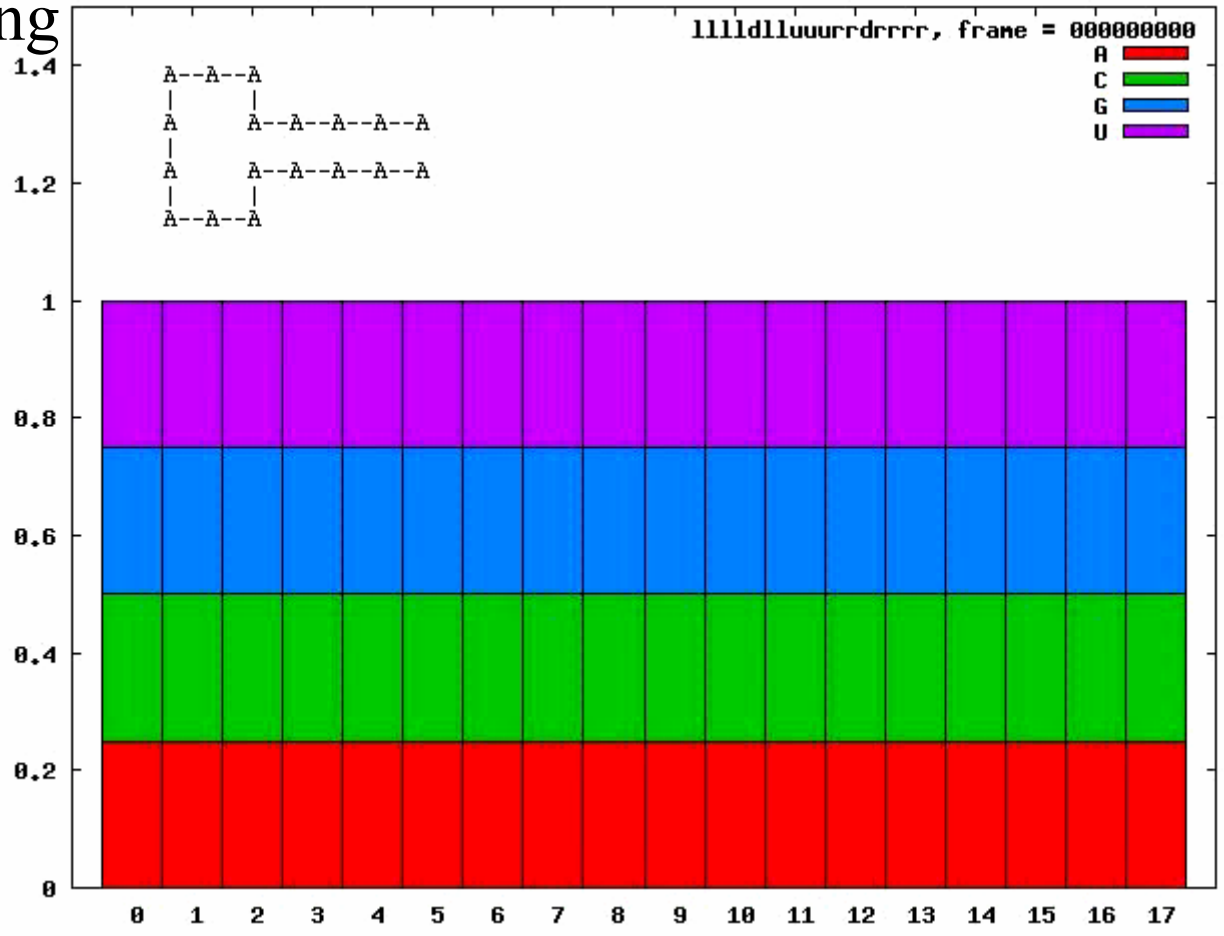
Stefan Bienert
Marco Mathies

- treat as coordinates
- dynamics in this space
- will it work ?



Sequence Dynamics

- Classic Newtonian dynamics / fictitious space
- Energy functions
  - literature +
  - prevent mis-folding
- Toy system
- read out sequence at end
- real tests..
- proteins



Proteins – prediction and similarities

- goals – structure predictions, similarities

Similarities – what is my protein related to ?

- sequences
  - what function do I have ?
  - what structure do I have ?

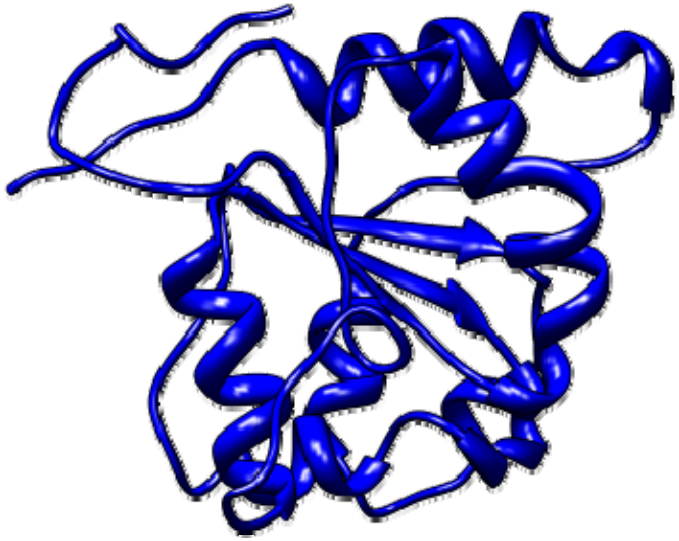
- structures
  - able to find more remote similarities
  - hopeless running time (NP complete)

- sequence changes faster than structure

Structure / sequence similarity

- TLR / toll-like receptors

1fyv



Gundolf Schenk
Thomas Margraf

Structure / sequence similarity

- TLR / toll-like receptors

48 % sequence id
1o77

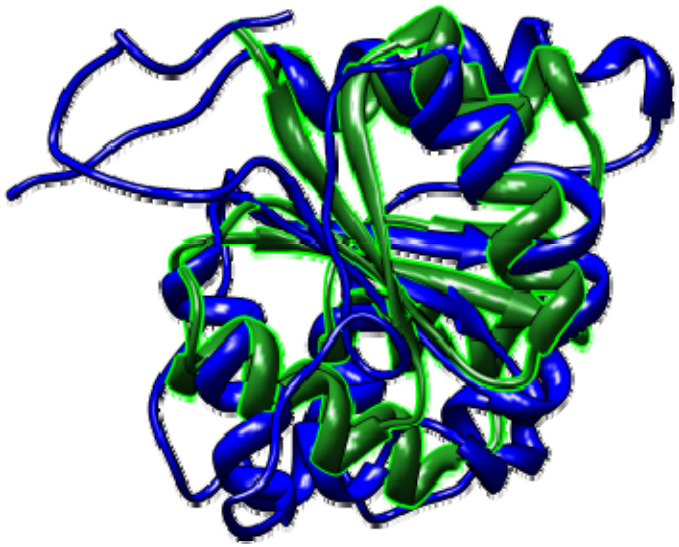
another TLR

Structure / sequence similarity

- TLR / toll-like receptors

6 % sequence id
2qxy

response regulator
T. Maritima

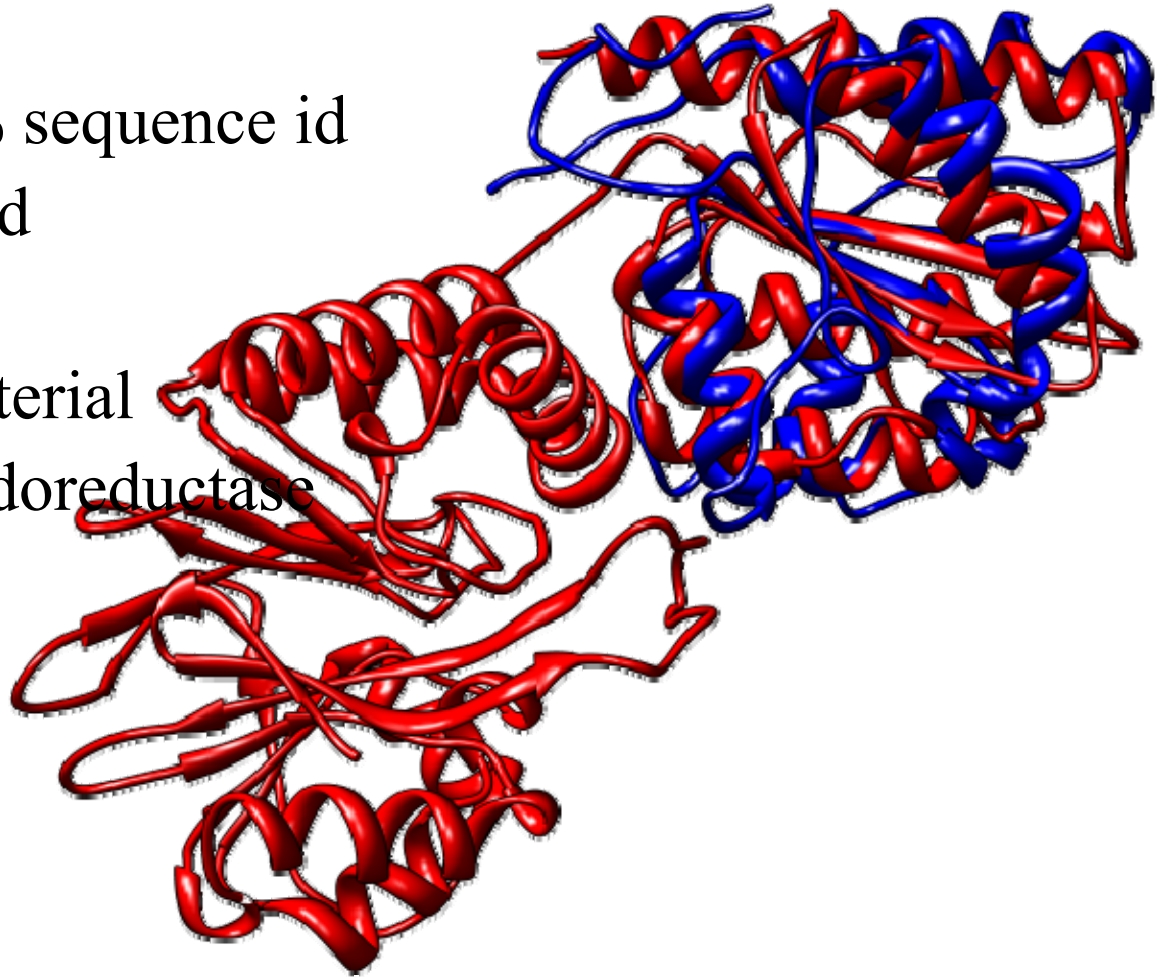


Structure / sequence similarity

- TLR / toll-like receptors

9 % sequence id
1e5d

bacterial
oxidoreductase

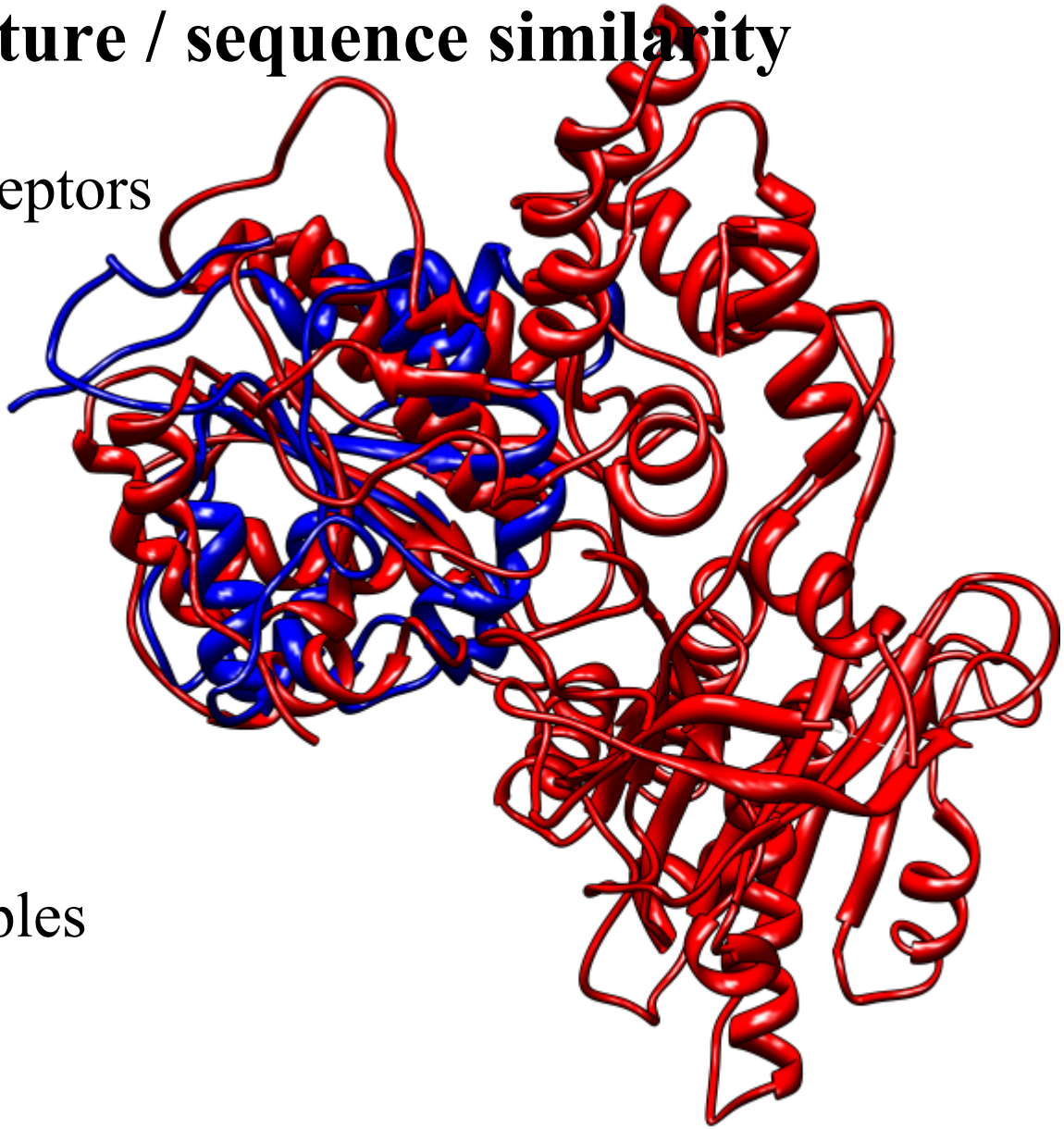


Structure / sequence similarity

- TLR / toll-like receptors

7 % sequence id
1ja0

rat
oxidoreductase



- 100’s more examples

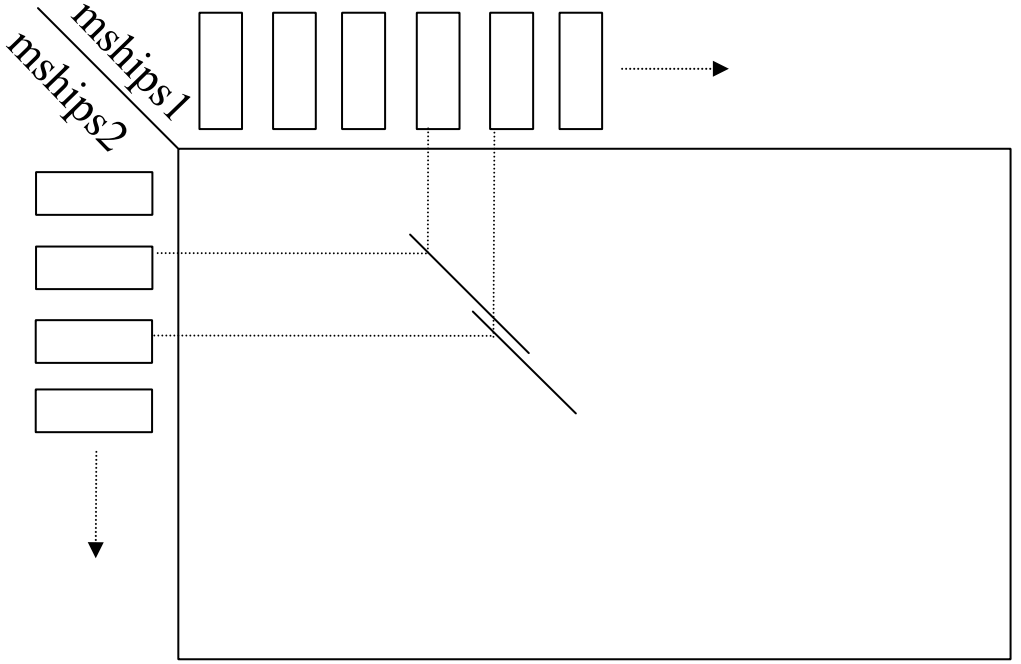
- how are they calculated ?

Calculating alignments

Martin Mosisch

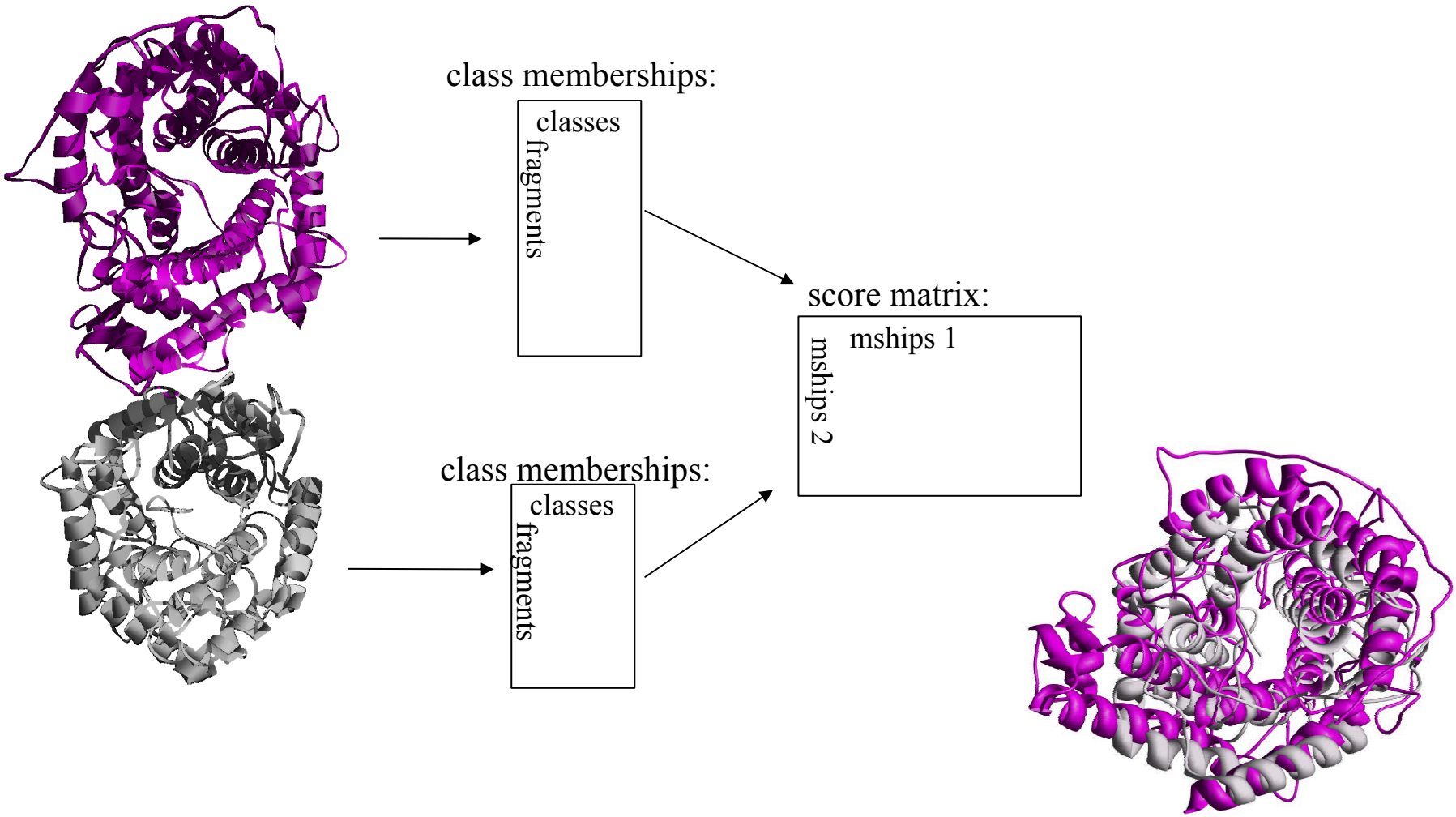
	1	2	3	4	5	6	7	8	9	...	N-6+1
1	15	0	3	100	54	34	68	19	3	...	73
2	22	3	80	40	5	20	34	78	5	...	19
3	1	50	50	50	64	34	35	20	58	...	34
4	22	38	65	80	87	75	3	68	20	...	75
5	3	47	80	34	34	34	68	67	40	...	34
6	38	34	80	77	68	57	51	54	78	...	57
7	1	96	45	34	22	80	93	22	3	...	53
8	22	47	15	68	93	72	65	40	27	...	72
9	54	123	83	1	22	83	76	96	88	...	90
10	12	34	62	54	44	40	32	15	7	...	60
11	38	47	22	63	12	1	5	40	68	...	38
12	83	78	15	3	54	30	22	47	71	...	20
13	0	5	20	40	38	53	83	100	93	...	53
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
308	19	35	22	57	57	57	21	13	27	...	33

	1	2	3	4	5	6	7	8	9	...	N-6+1
1	15	0	3	100	54	34	68	19	3	...	73
2	22	3	80	40	5	20	34	78	5	...	19
3	1	50	50	50	64	34	35	20	58	...	34
4	22	38	65	80	87	75	3	68	20	...	75
5	3	47	80	34	34	34	68	67	40	...	34
6	38	34	80	77	68	57	51	54	78	...	57
7	1	96	45	34	22	80	93	22	3	...	53
8	22	47	15	68	93	72	65	40	27	...	72
9	54	123	83	1	22	83	76	96	88	...	90
10	12	34	62	54	44	40	32	15	7	...	60
11	38	47	22	63	12	1	5	40	68	...	38
12	83	78	15	3	54	30	22	47	71	...	20
13	0	5	20	40	38	53	83	100	93	...	53
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
308	19	35	22	57	57	57	21	13	27	...	33



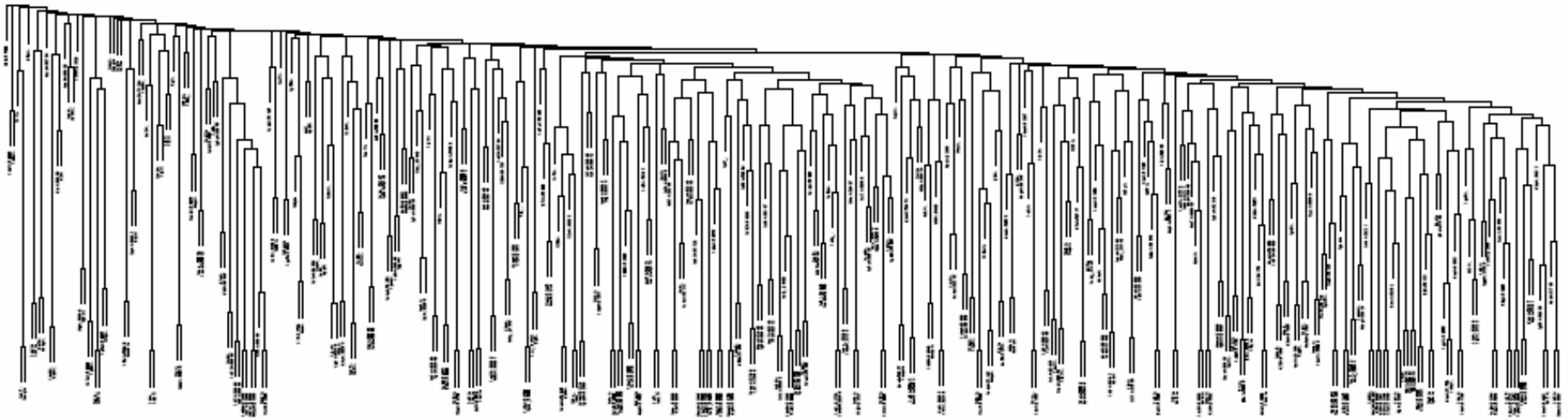
- coordinates to vectors of structure properties
- fill score matrix
- find best path

Calculating alignments



- if one can do pairs of proteins swiftly..

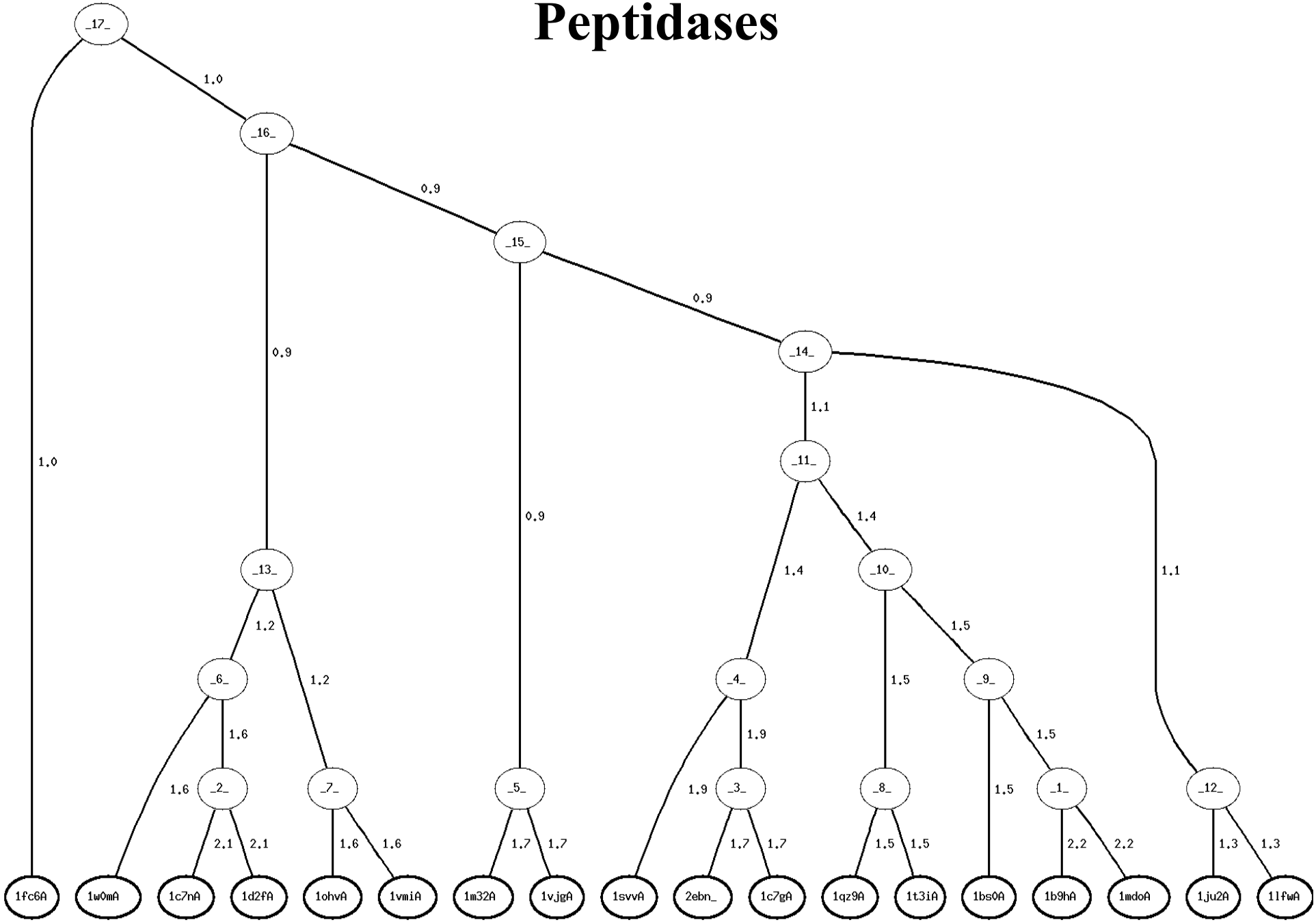
Structural Phylogeny



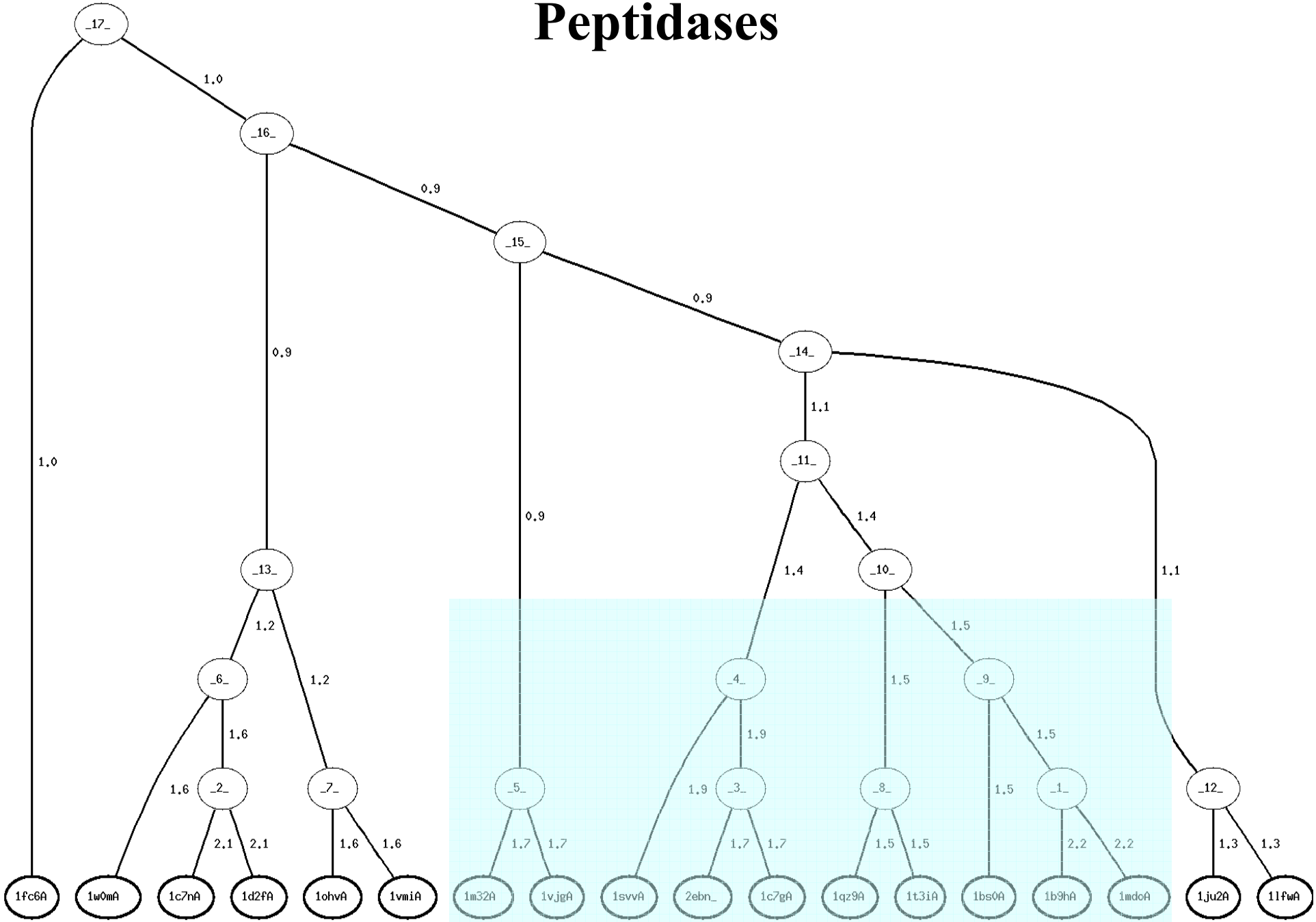
- a bigger alignment

Structural Phylogeny

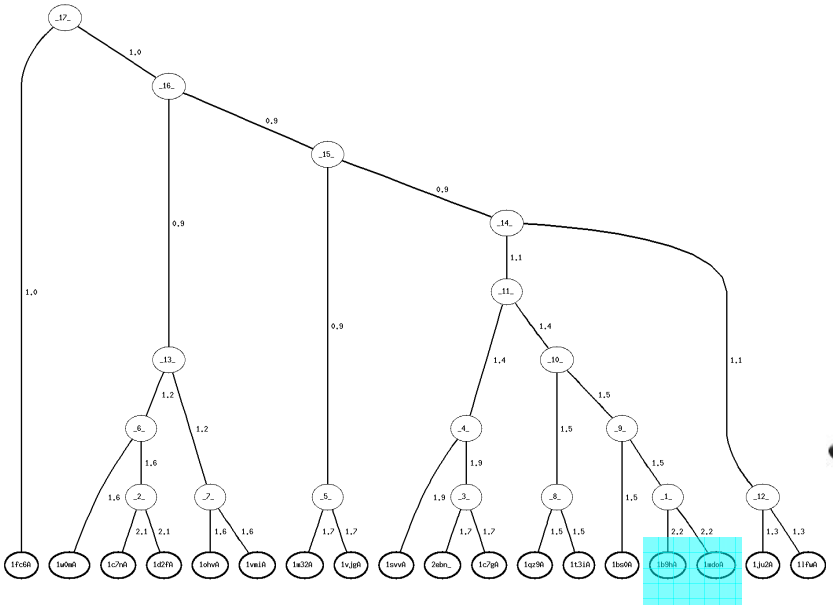
Peptidases

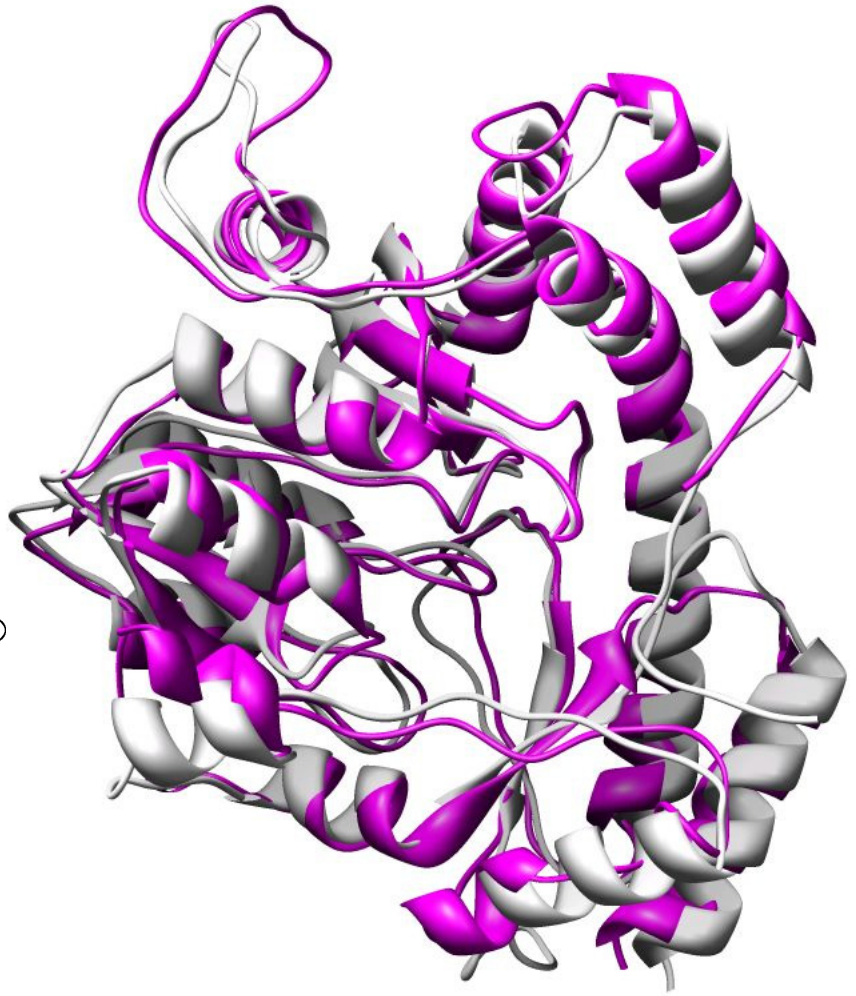

_17_
1.0
_16_
0.9
_15_
0.9
_14_
1.0
0.9
1.1
_11_
0.9
1.4
1.4
_10_
1.1
_13_
1.2
1.5
_6_
1.2
_4_
1.5
_9_
1.6
1.9
1.5
1.6
_2_
_7_
_5_
1.9
_3_
_8_
1.5
_1_
_12_
2.1
2.1
1.6
1.6
1.7
1.7
1.7
1.7
1.5
1.5
2.2
2.2
1.3
1.3
1fc6A
1w0mA
1c7nA
1d2fA
1ohvA
1vmiA
1m32A
1vjgA
1svvA
2ebn_
1c7gA
1qz9A
1t3iA
1bs0A
1b9hA
1mdoA
1ju2A
1lfwA

Peptidases

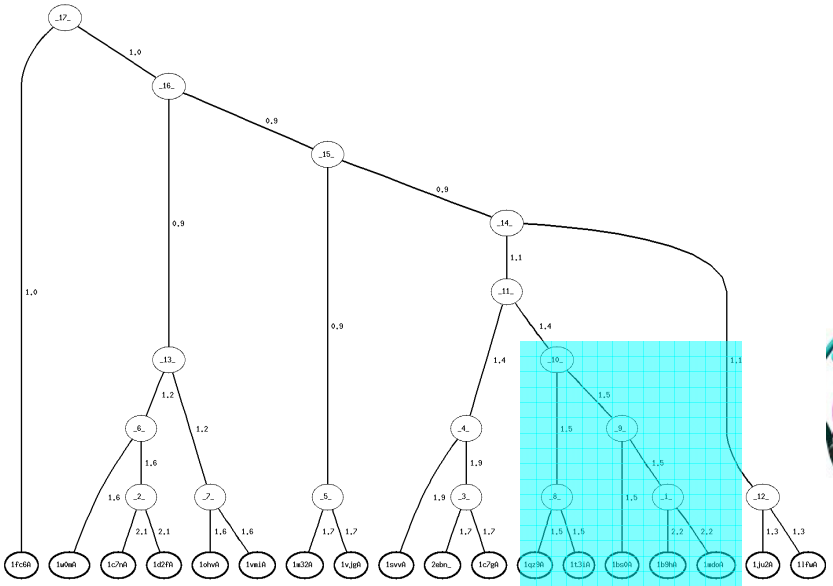

_17_
1.0
_16_
0.9
_15_
0.9
_14_
1.0
0.9
1.1
_11_
0.9
1.4
1.4
_10_
1.1
_13_
1.5
1.2
_4_
1.5
_9_
1.2
_6_
1.9
1.5
1.6
1.6
_5_
1.9
_3_
1.5
_8_
1.5
_1_
_12_
_2_
_7_
1.7
1.7
1.7
1.7
1.5
1.5
2.2
2.2
1.3
1.3
2.1
2.1
1.6
1.6
1fc6A
1w0mA
1c7nA
1d2fA
1ohvA
1vmiA
1m32A
1vjgA
1svvA
2ebn_
1c7gA
1qz9A
1t3iA
1bs0A
1b9hA
1mdoA
1ju2A
1lfwA

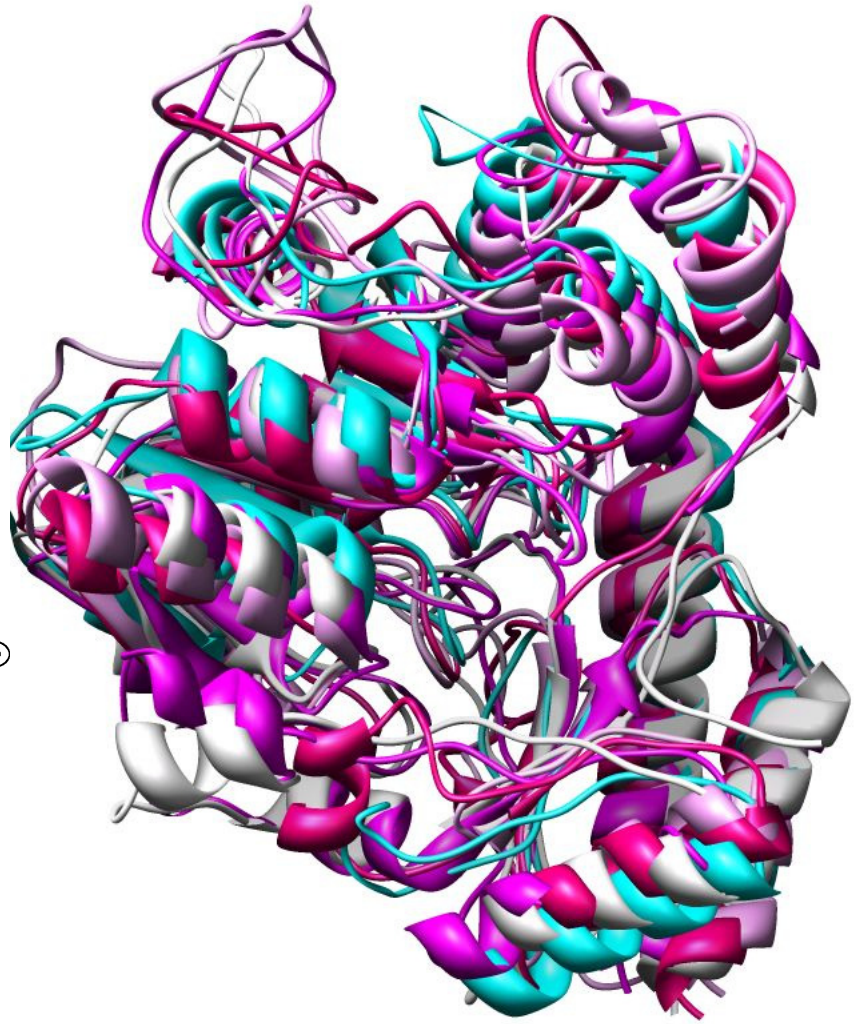
Peptidases


_17_
_16_
_15_
_14_
_13_
_11_
_10_
_6_
_4_
_9_
_2_
_7_
_5_
_3_
_8_
_1_
_12_
1fc6A
1w0wA
1c7vA
1d2fA
1ohvA
1vmlA
1m32A
1vjgA
1svvA
2ebn_
1c7gG
1qz9A
1t31A
1bs0A
1b9hA
1wdoA
1ju2A
11fwA

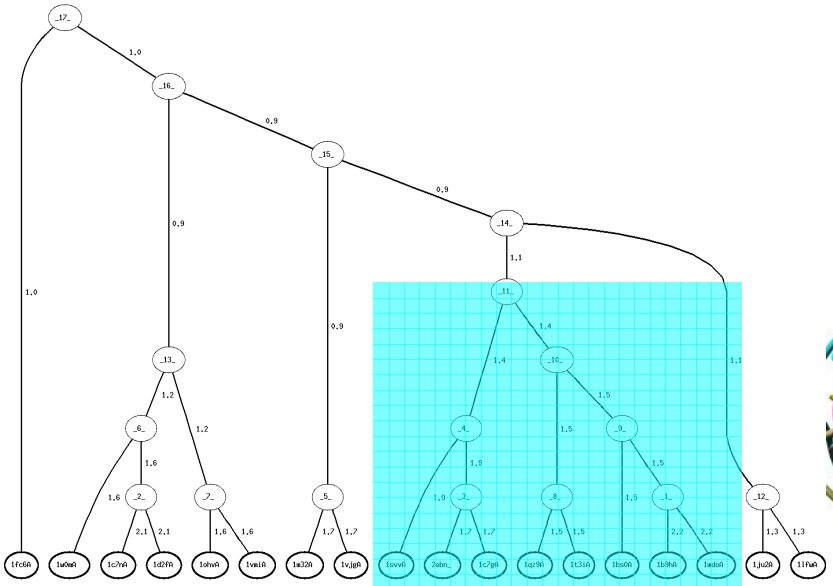


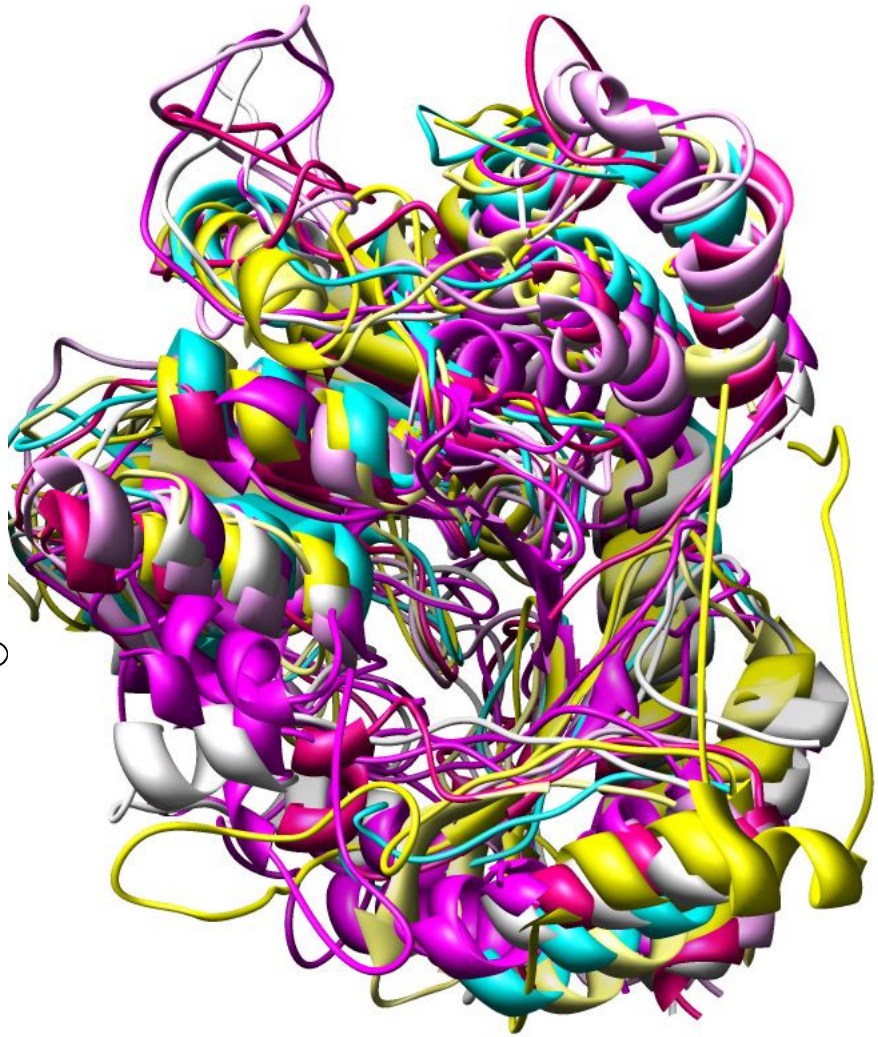
Peptidases


_17_
_16_
_15_
_14_
_11_
_13_
_10_
_6_
_4_
_9_
_2_
_7_
_5_
_3_
_8_
_1_
_12_
1fc6A
1w0wA
1c7vA
1d2fA
1ohvA
1vmlA
1m32A
1vjgA
1svvA
2ebn_
1c7gB
1qz3A
1t3iA
1bs0A
1b9hA
1wdoA
1ju2A
11fwA



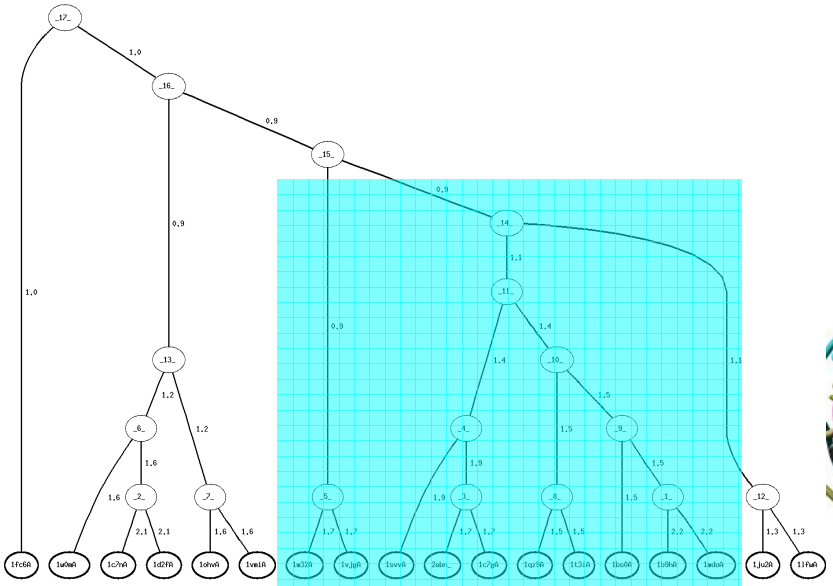
Peptidases

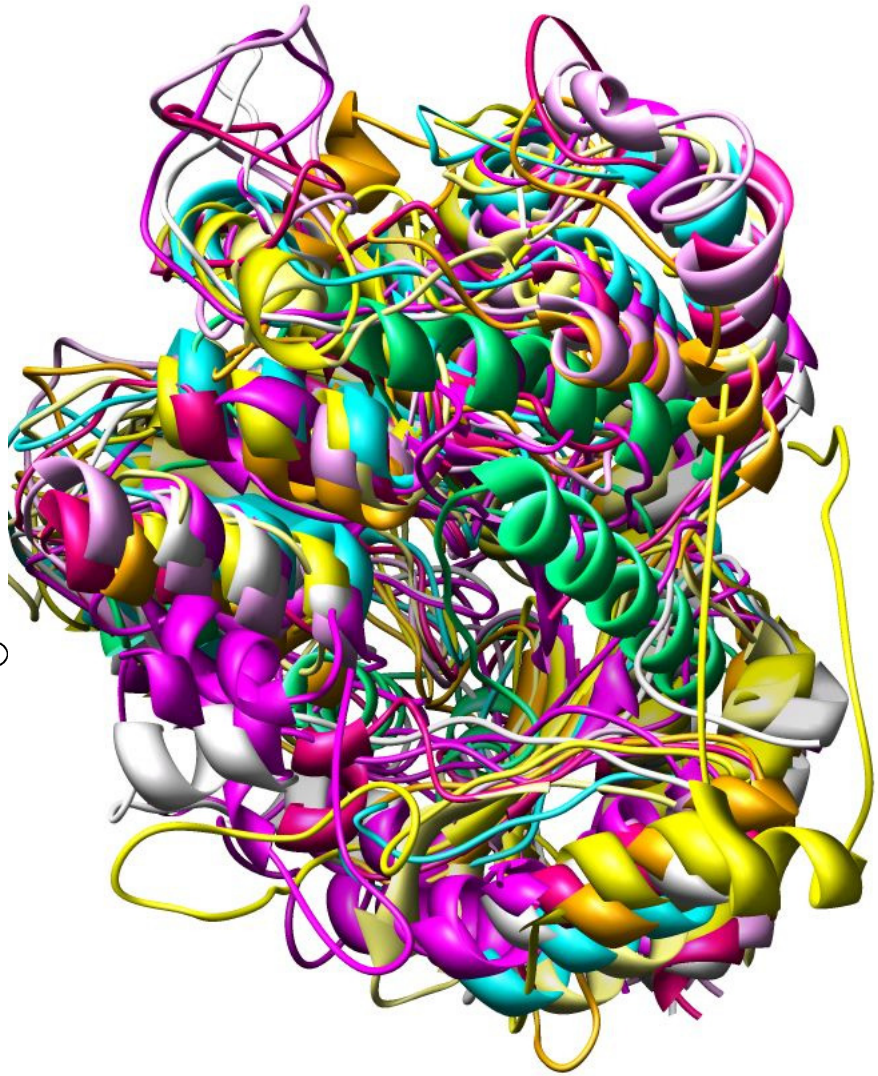




- includes TIM barrel

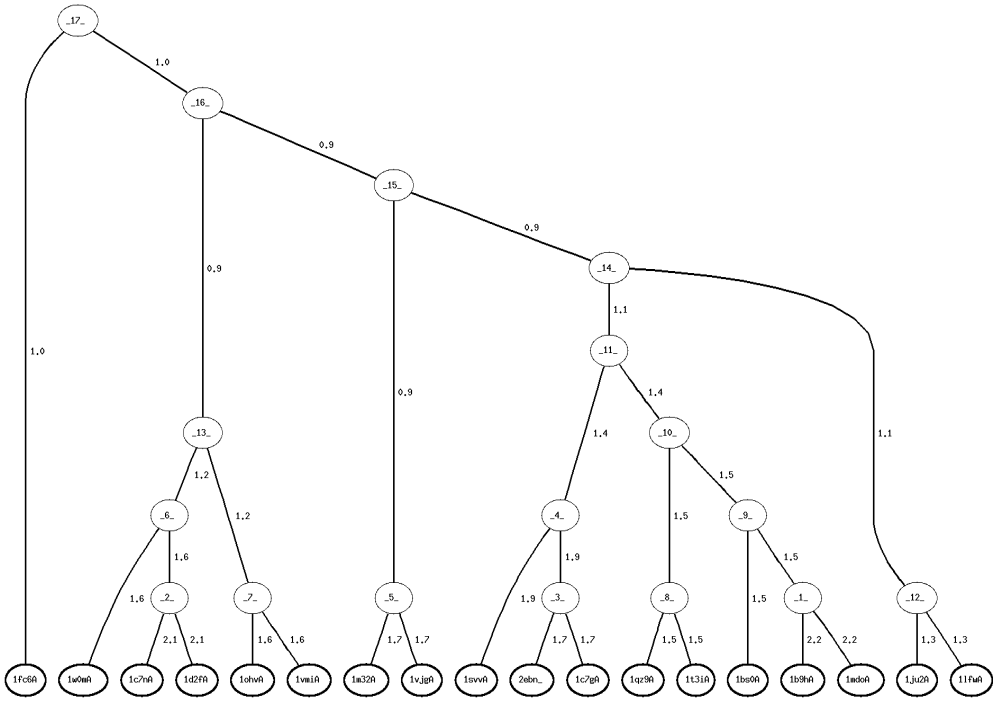
Peptidases





Methodology

- mostly classic phylogeny – some unique features
  - numerical approach
    - nodes are imaginary structures



- structure comparing ?

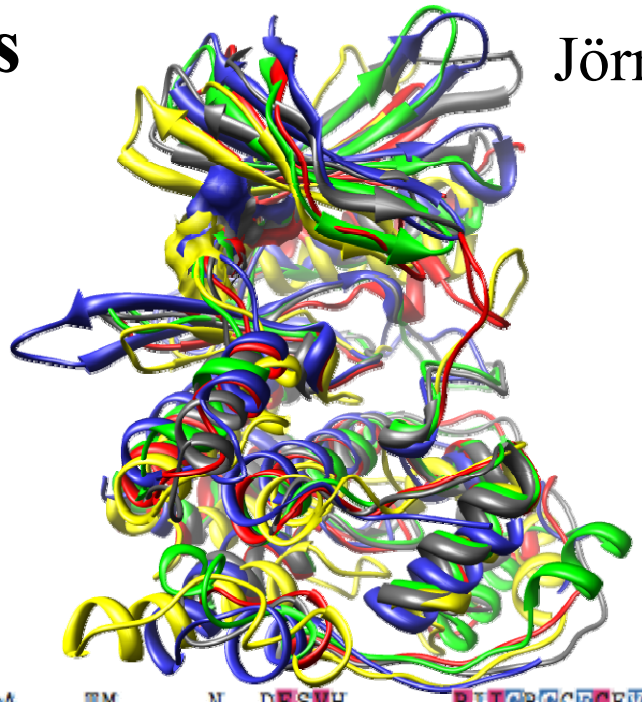
`cardigan.zbh.uni-hamburg.de/salami/`

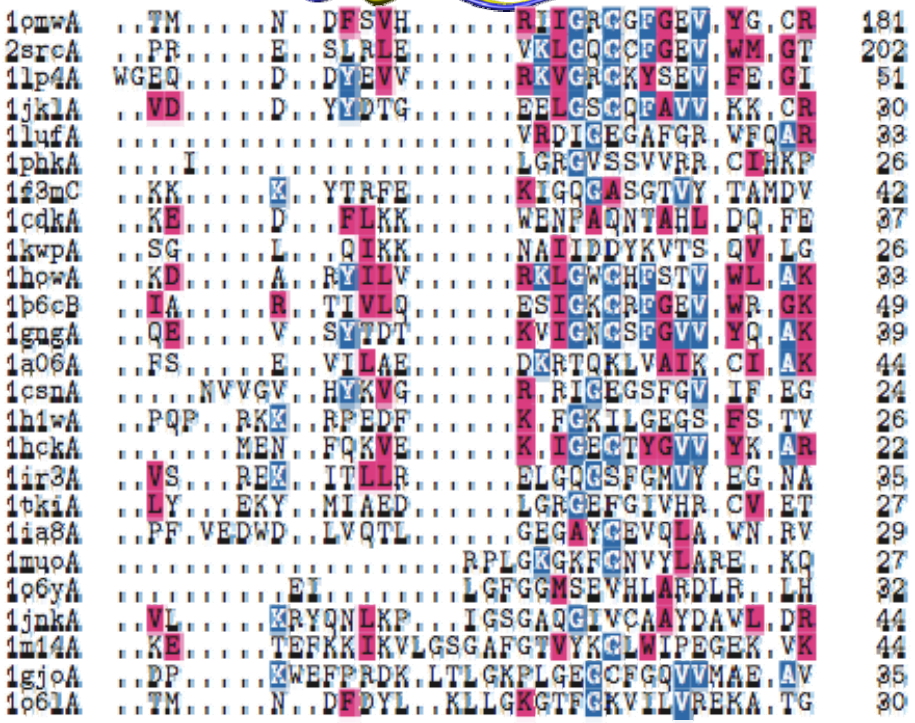
- who cares ?

Kinases

Jörn Lenz

- kinases
  - structure based methods
  - accurate alignments
    - active / conserved residues
  - reliable classification
  - needs structure ?
    - mostly
    - interplay of sequence and structure





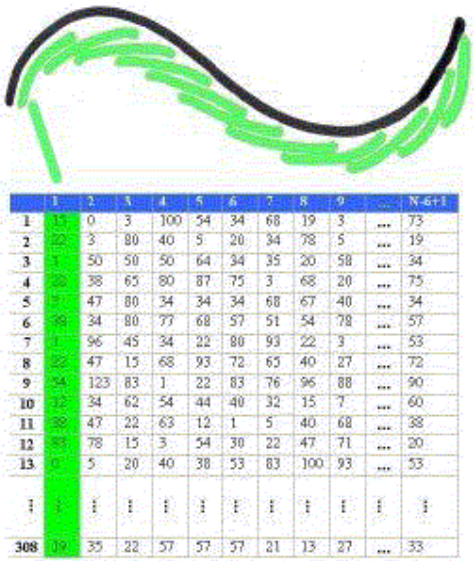
Sequence versus structure

- close homology
  - use sequence – models, classification, function
  - easy

- remote homology
  - more speculative
  - needs structural information

From classification to prediction

- previously
  - vectors of structural properties
- now
  - mix sequence and structure properties



	1	2	3	4	5	6	7	8	9	...	N-6+1
1	17	0	3	100	54	34	68	19	3	...	73
2	82	3	80	40	5	20	34	78	5	...	19
3	5	50	50	50	64	34	35	20	58	...	34
4	28	38	65	80	87	75	3	68	20	...	75
5	1	47	80	34	34	34	68	67	40	...	34
6	39	34	80	77	68	57	51	54	78	...	57
7	1	96	45	34	22	80	93	22	3	...	53
8	22	47	15	68	93	72	65	40	27	...	72
9	94	123	83	1	22	83	76	96	88	...	90
10	12	34	62	54	44	40	32	15	7	...	60
11	38	47	22	63	12	1	5	40	68	...	38
12	83	78	15	3	54	30	22	47	71	...	20
13	0	5	20	40	38	53	83	100	93	...	53
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
308	19	35	22	57	57	57	21	13	27	...	33

- result
  - from known (sequence)
  - to unknown (structure)

- via known structures (threading)

  `www.zbh.uni-hamburg.de/wurst/`

- completely new (Monte Carlo like methods)