

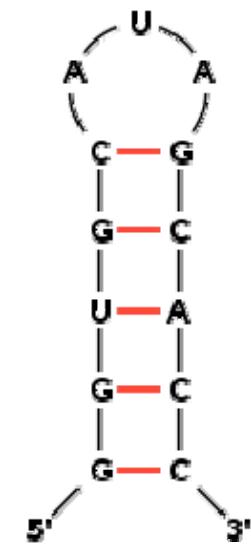
Proteins, remote homologies, ...

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

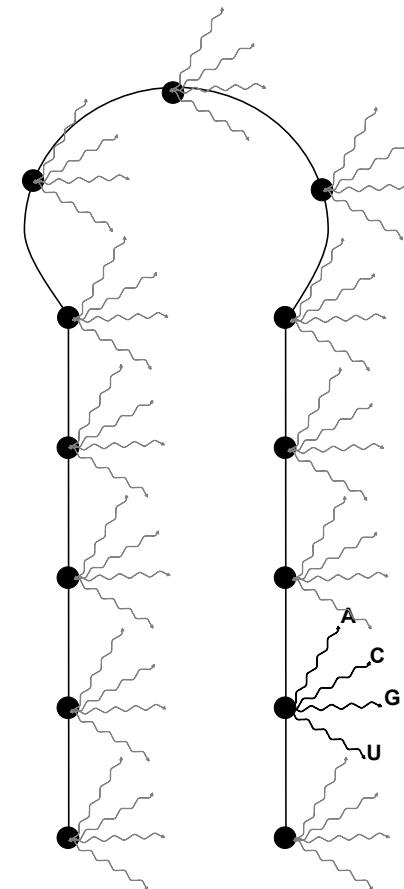
- Unusual spaces and dynamics
- Design molecules
 - useful structure (ribozyme, riboswitch, ...)
 - find a new sequence which preserves structure
- energy functions ? Literature conventions + tricks
- particles can be A, C, G, U
 - search space = $4 \times 4 \times 4 \dots = 4^N$
 - discrete problem ? maybe not



RNA design

- particles can be of mixed types
- treat as coordinates
- dynamics in this space
- will it work ?

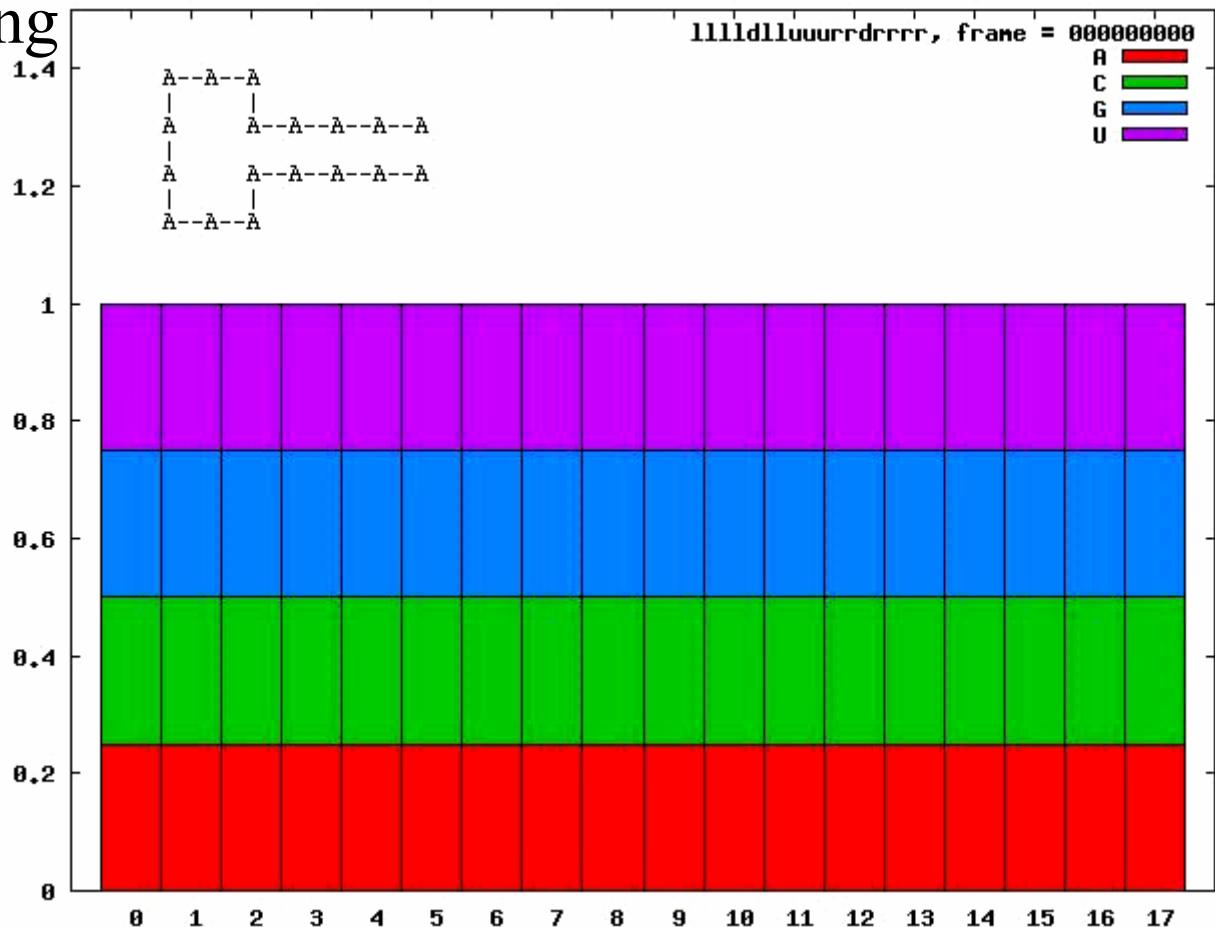
Stefan Bienert
Marco Mathies



Sequence Dynamics

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

The diagram illustrates a protein sequence 'A--A--A' at the top. Below it are four horizontal bars representing energy levels. The bottom bar is red, the second is green, the third is blue, and the top bar is magenta. The y-axis on the left is labeled with values 0.2, 0.4, 0.6, 0.8, 1, 1.2, and 1.4.



Proteins – prediction and similarities

- goals – structure predictions, similarities

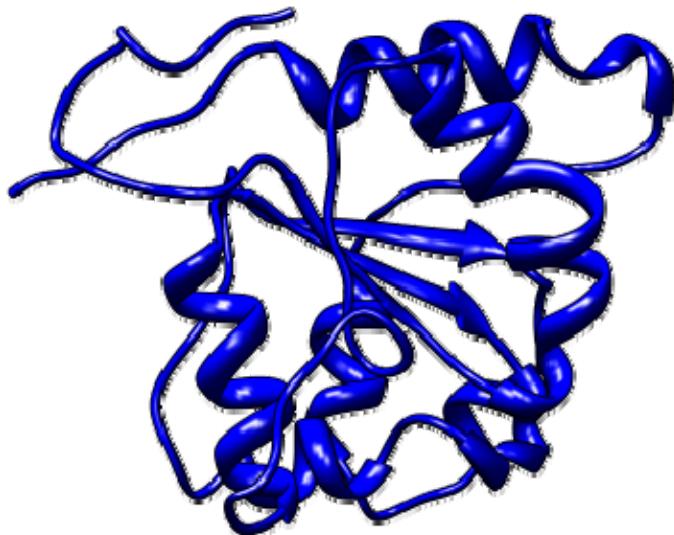
Similarities – what is my protein related to ?

- 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



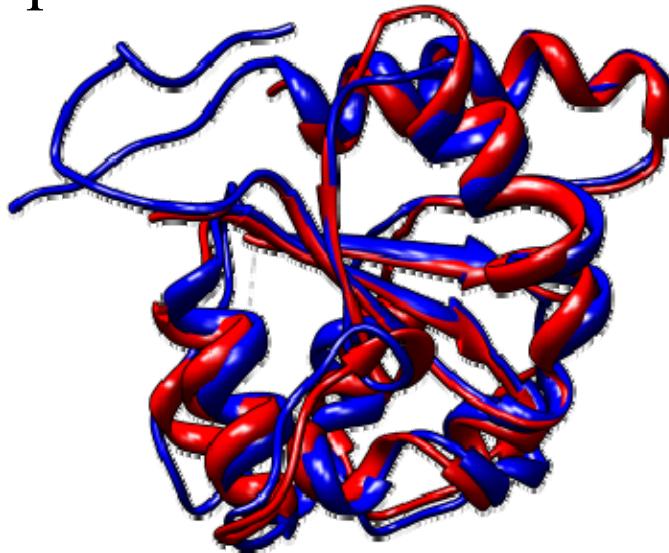
Gundolf Schenk
Thomas Margraf

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

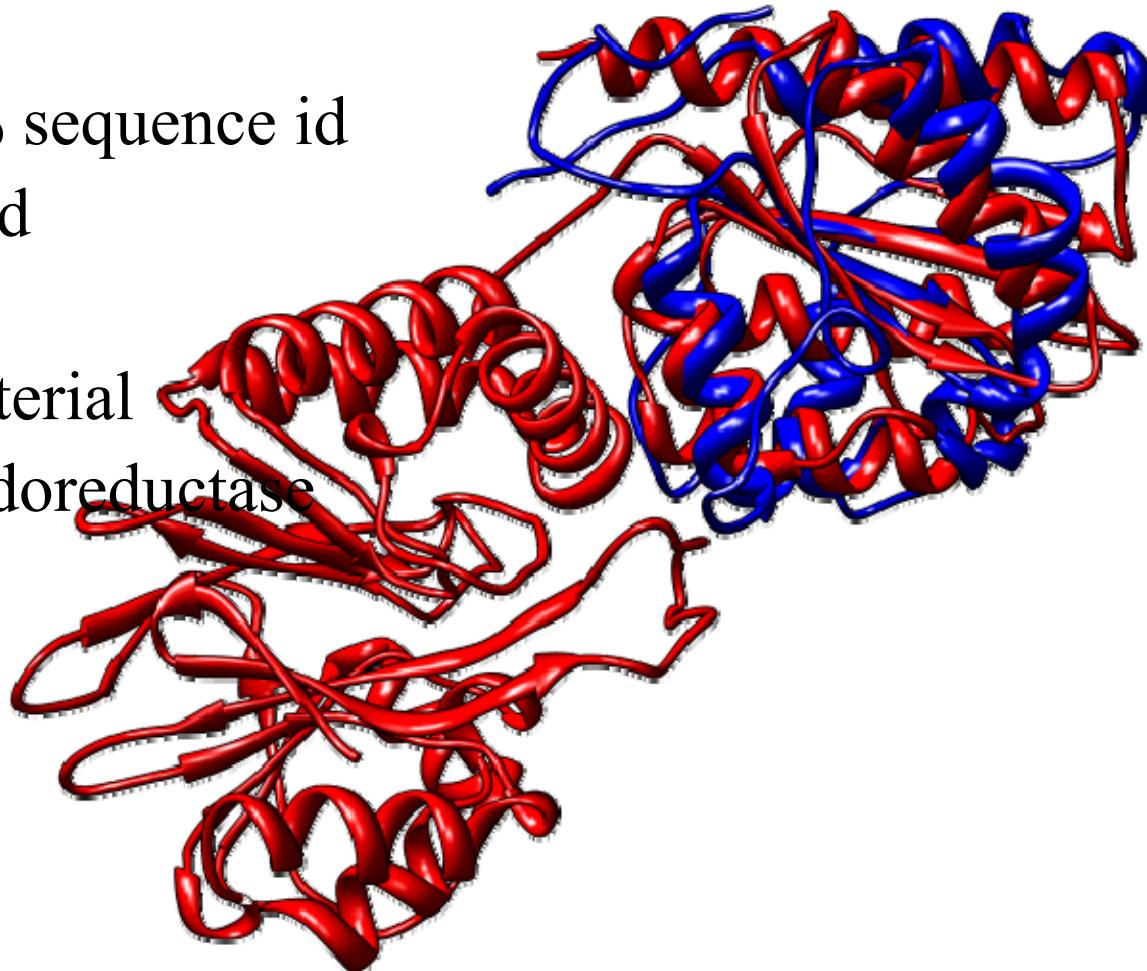


Structure / sequence similarity

- TLR / toll-like receptors

9 % sequence id
1e5d

bacterial
oxidoreductase

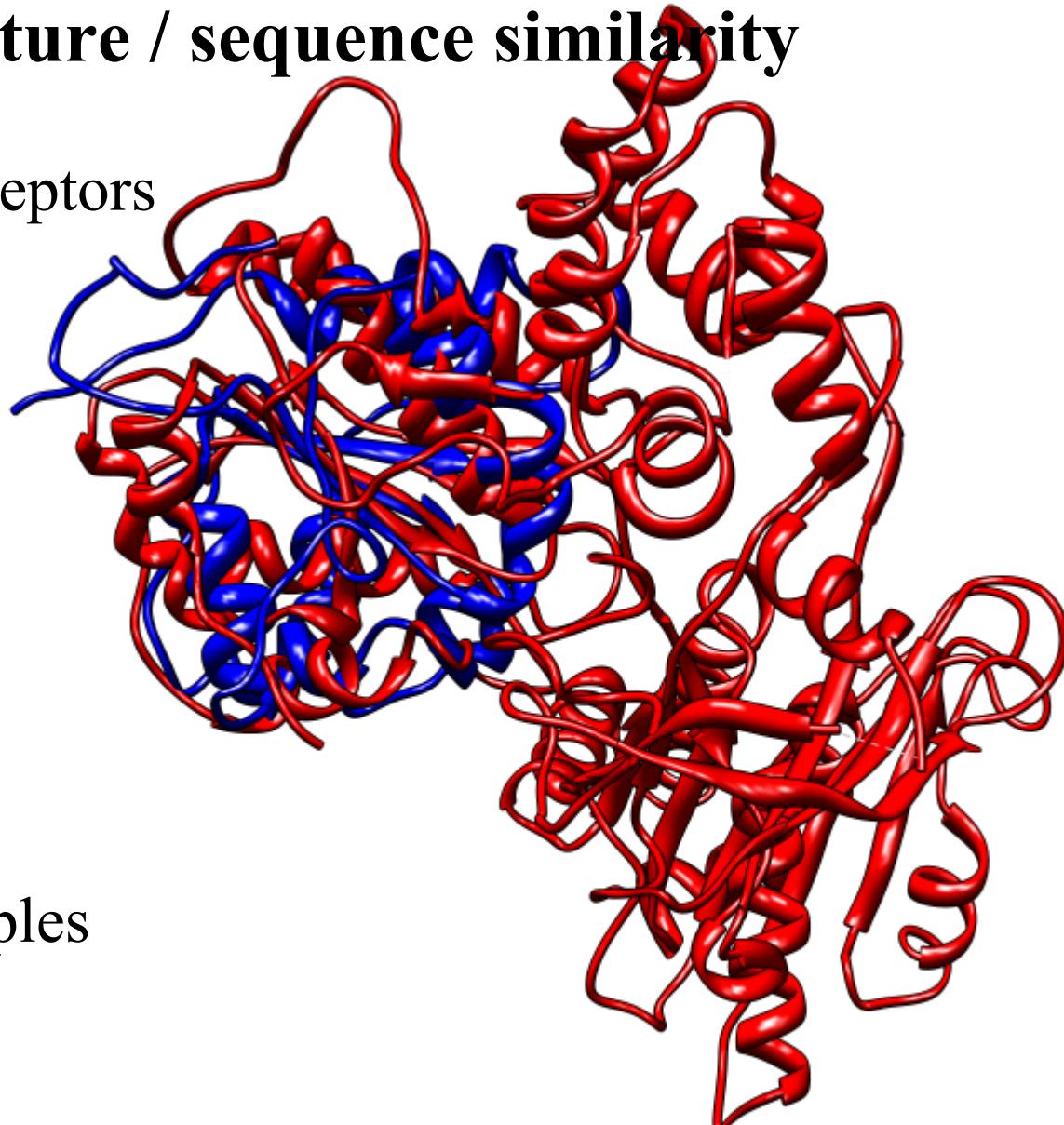


Structure / sequence similarity

- TLR / toll-like receptors

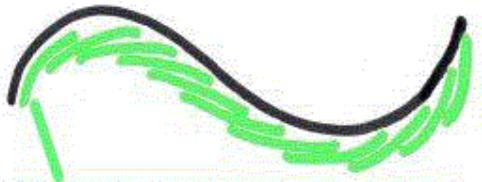
7 % sequence id
1ja0

rat
oxidoreductase



- 100's more examples
- how are they calculated ?

Calculating alignments



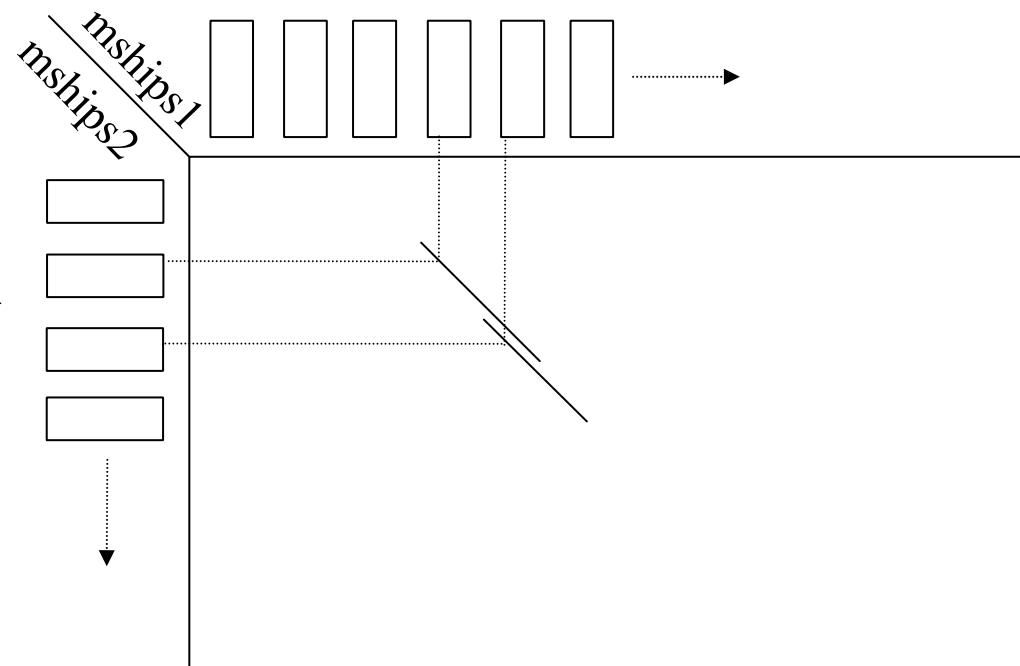
	1	2	3	4	5	6	7	8	9	...	N-6+1
1	15	0	3	100	54	34	68	19	3	...	73
2	22	3	80	40	5	20	34	78	5	...	19
3	1	50	50	50	64	34	35	20	58	...	34
4	38	65	80	87	75	3	68	20	...	75	
5	47	80	34	34	34	68	67	40	...	34	
6	34	80	77	68	57	51	54	78	...	57	
7	1	96	45	34	22	80	93	22	3	...	53
8	47	15	68	93	72	65	40	27	...	72	
9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



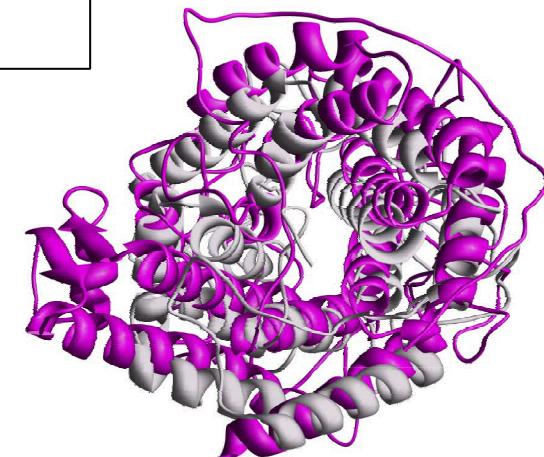
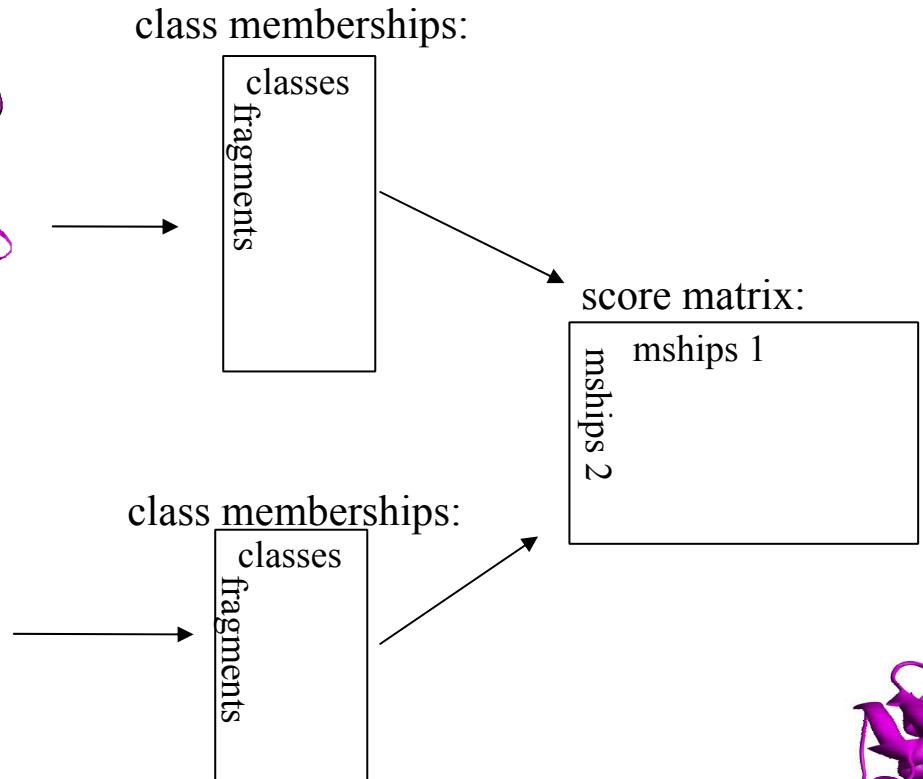
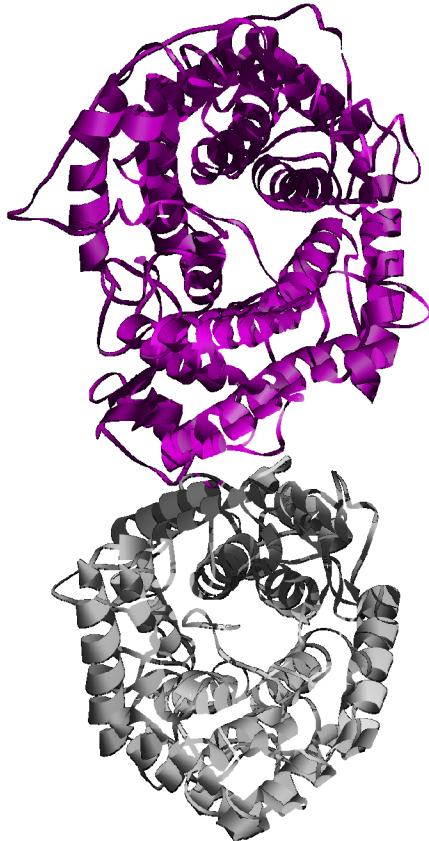
Martin Mosisch

	1	15	0	3	100	54	34	68	19	3	...	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	323	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	
	1	1	1	1	1	1	1	1	1	1	1	
308	0	35	22	57	57	21	13	27	...	33		

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

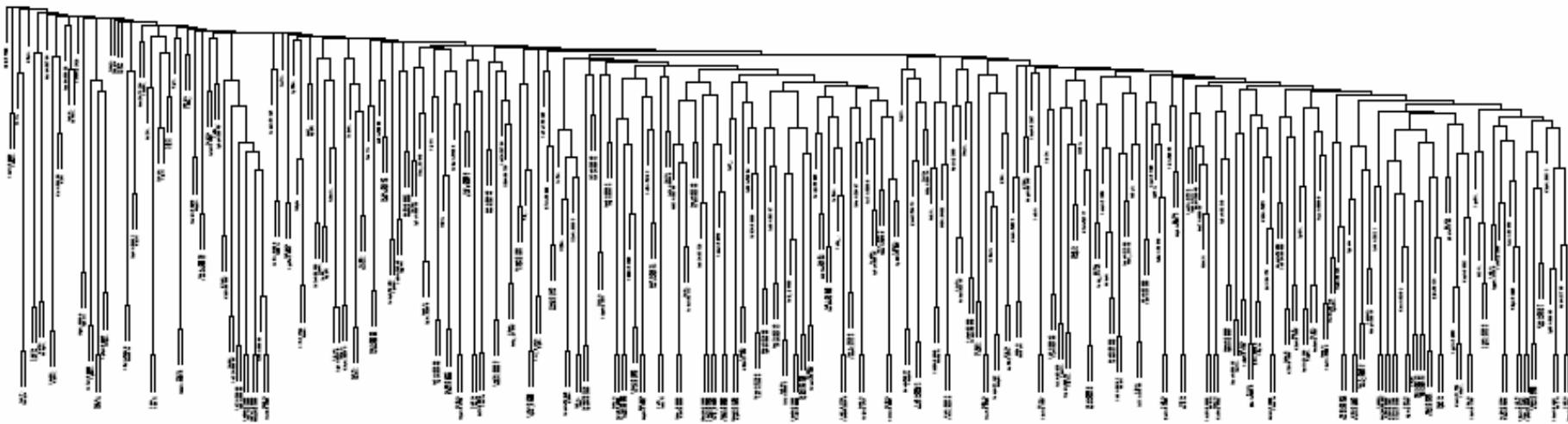


Calculating alignments



- if one can do pairs of proteins swiftly..

Structural Phylogeny

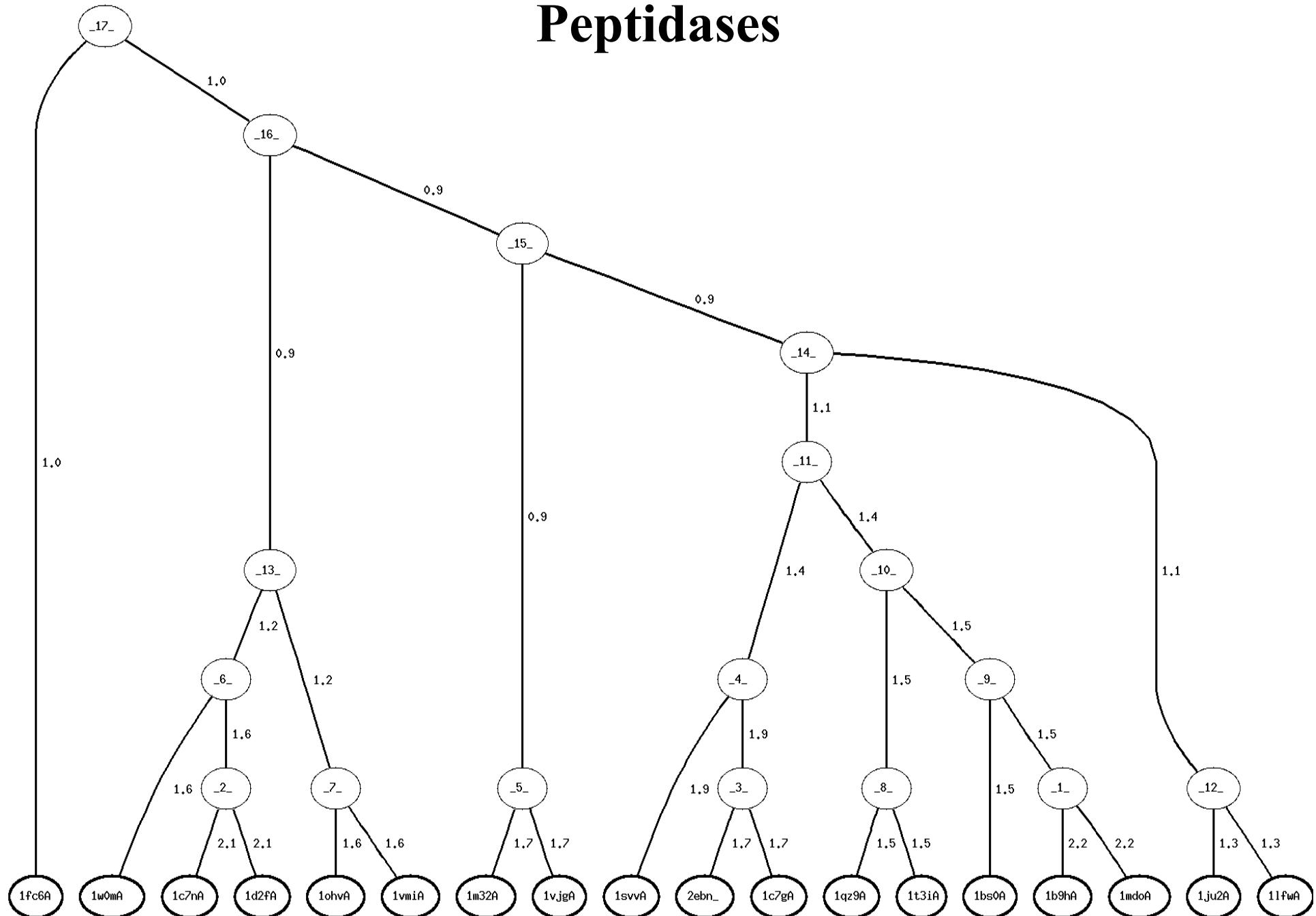


- a bigger alignment

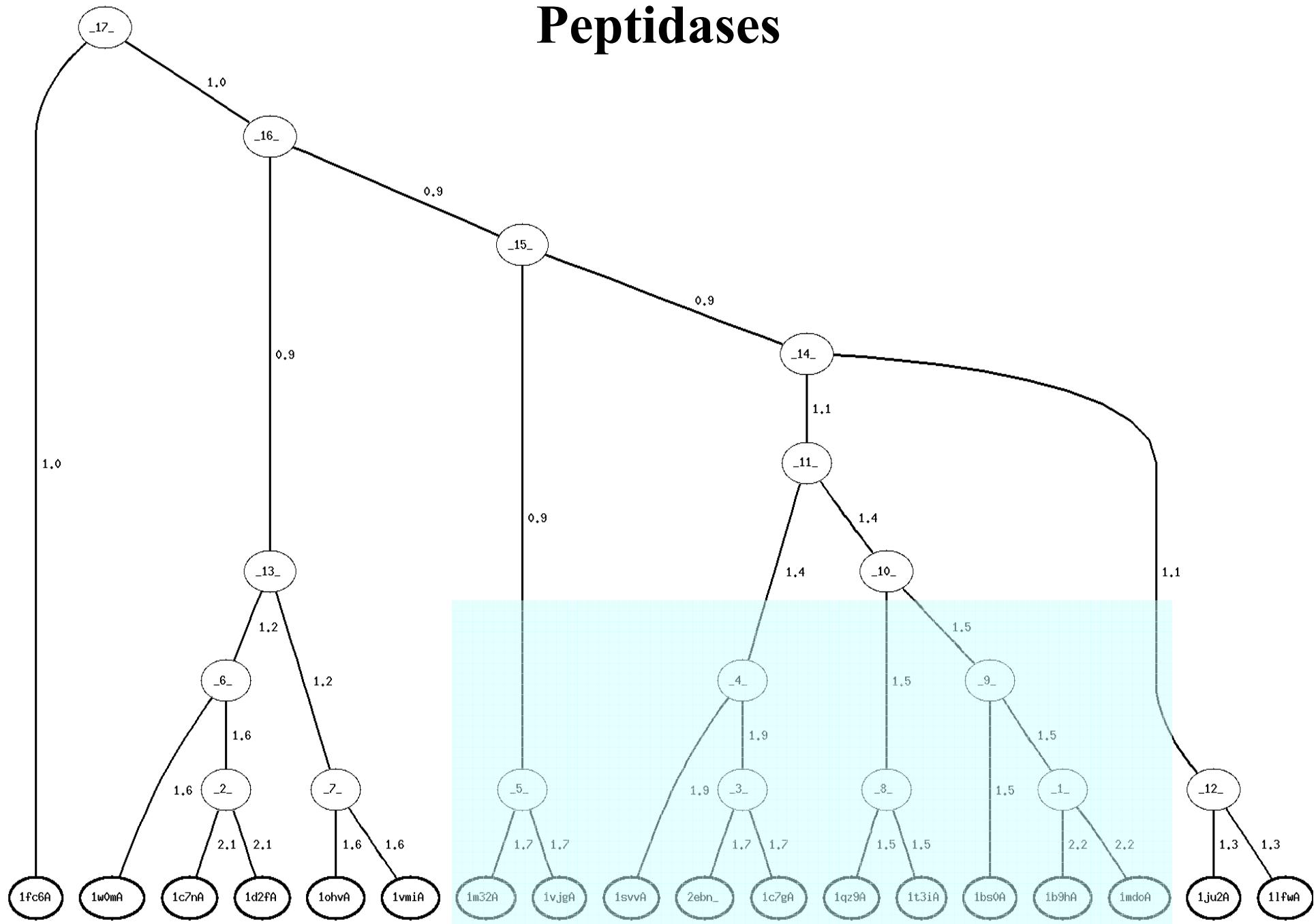
Structural Phylogeny



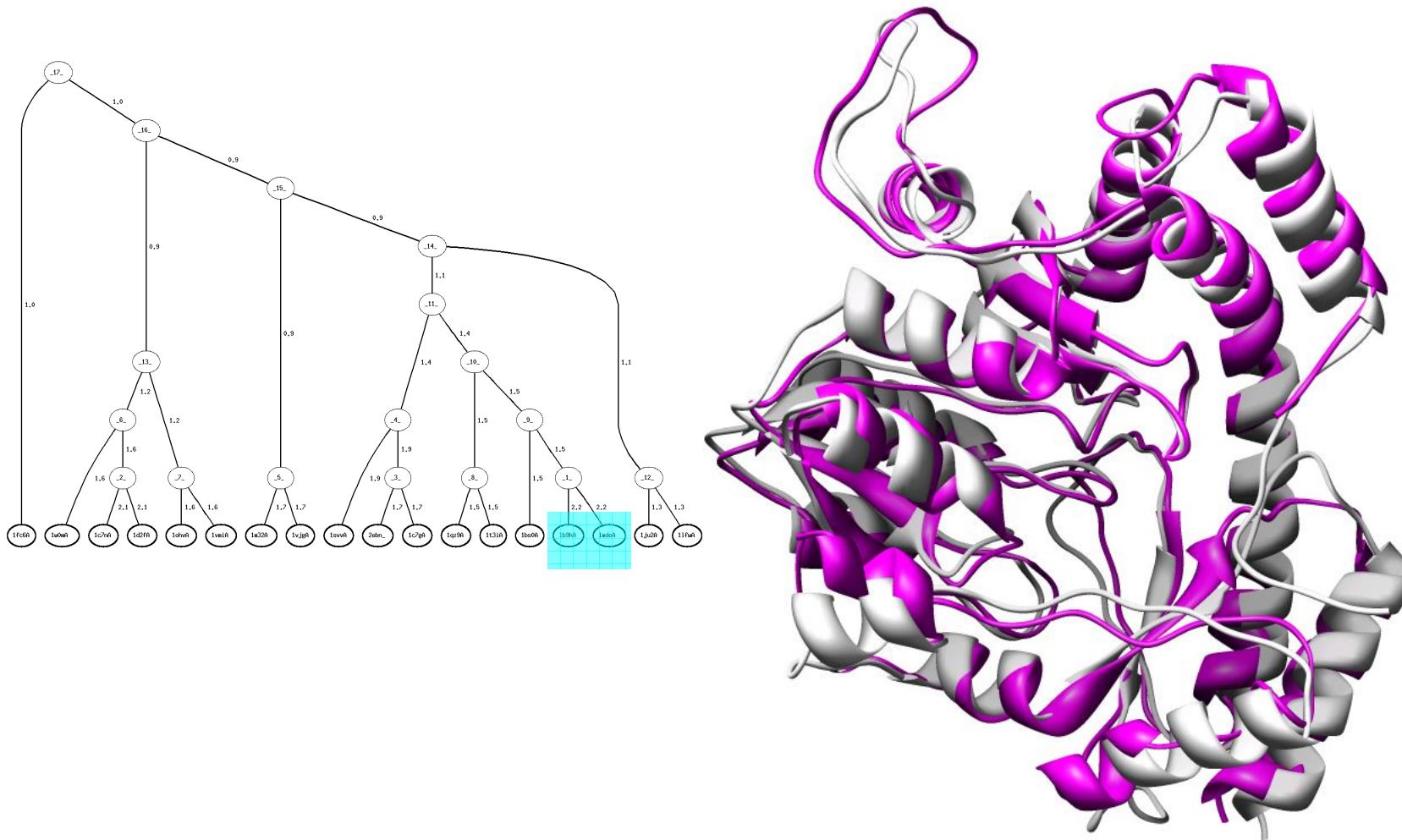
Peptidases



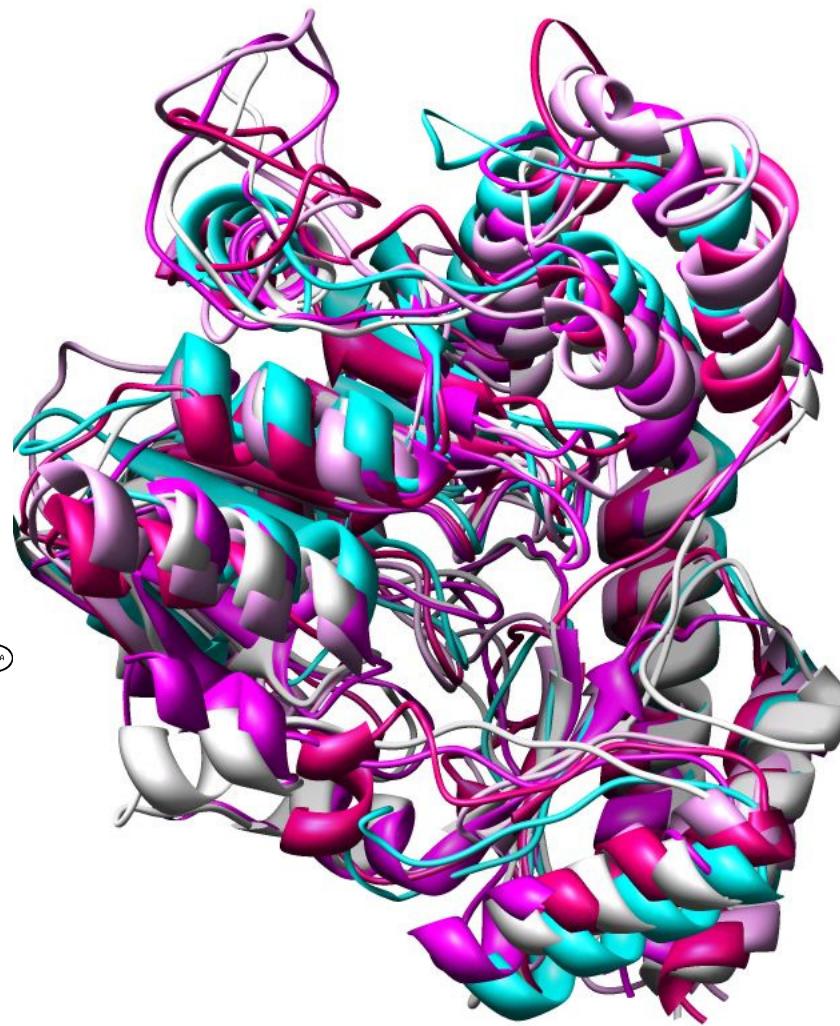
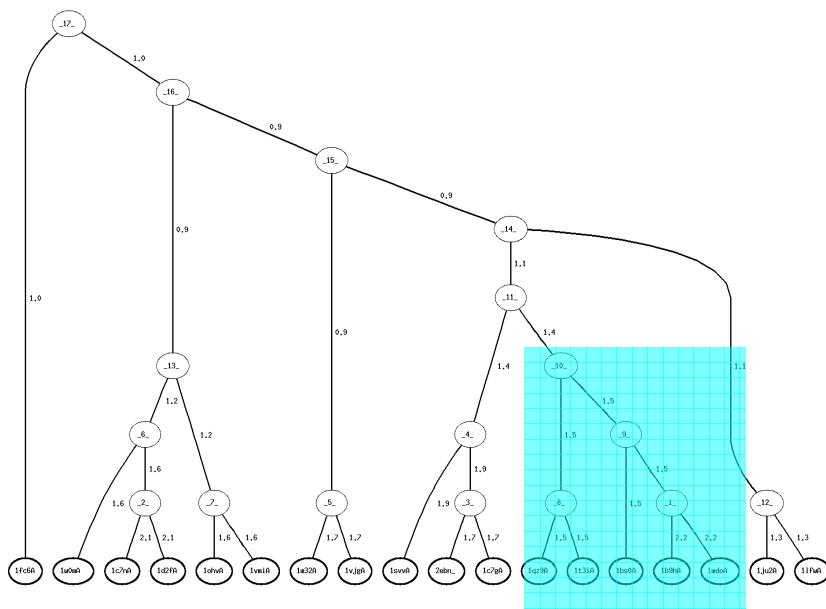
Peptidases



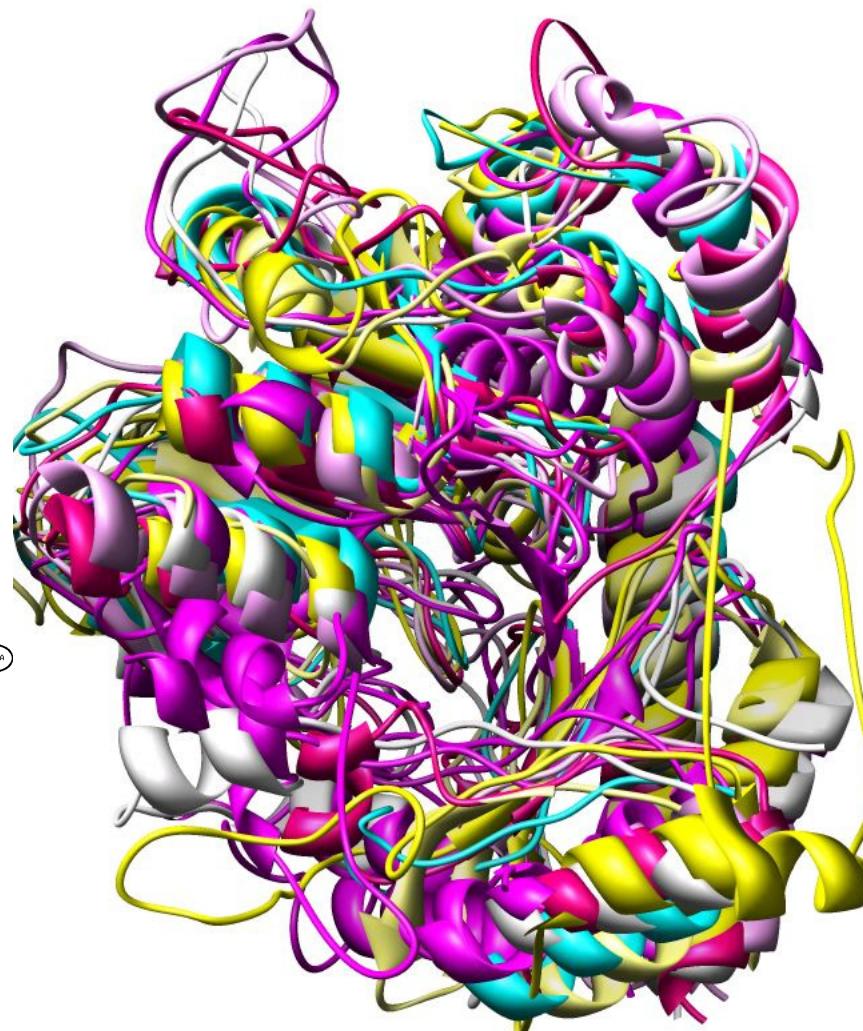
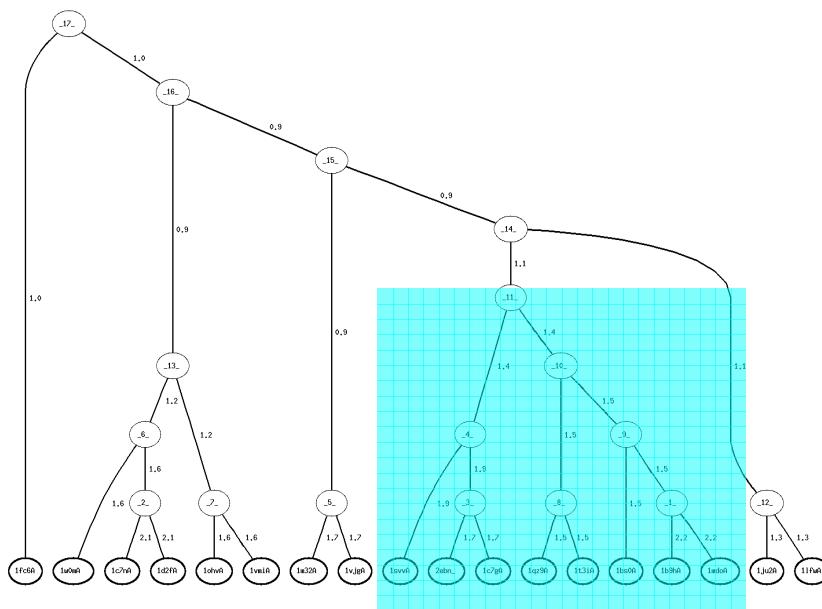
Peptidases



Peptidases

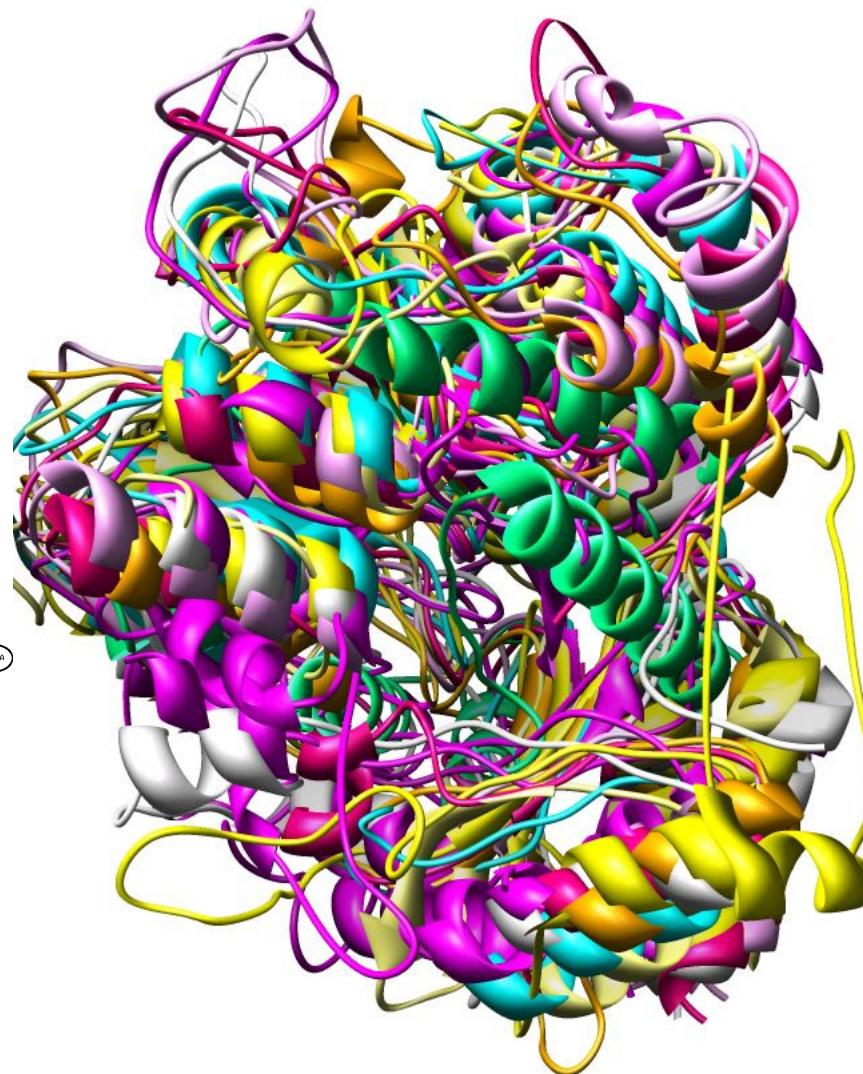
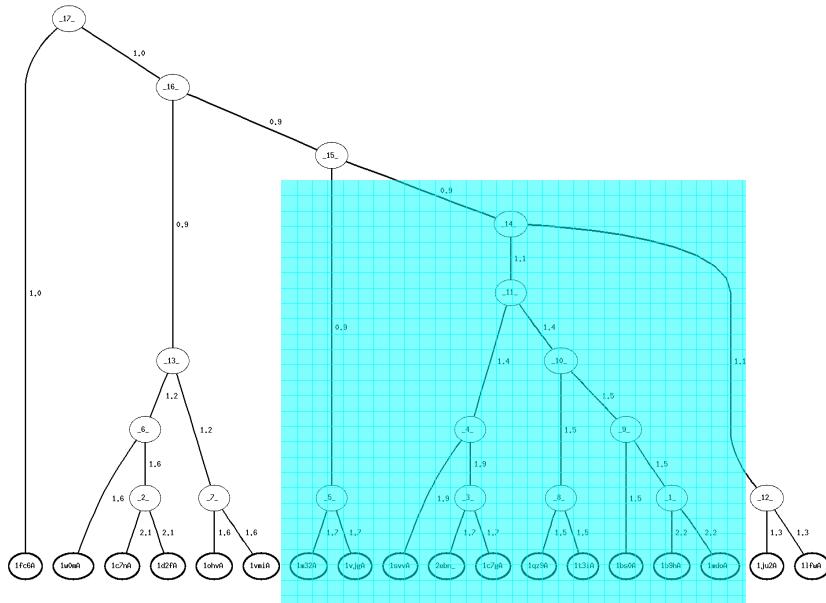


Peptidases



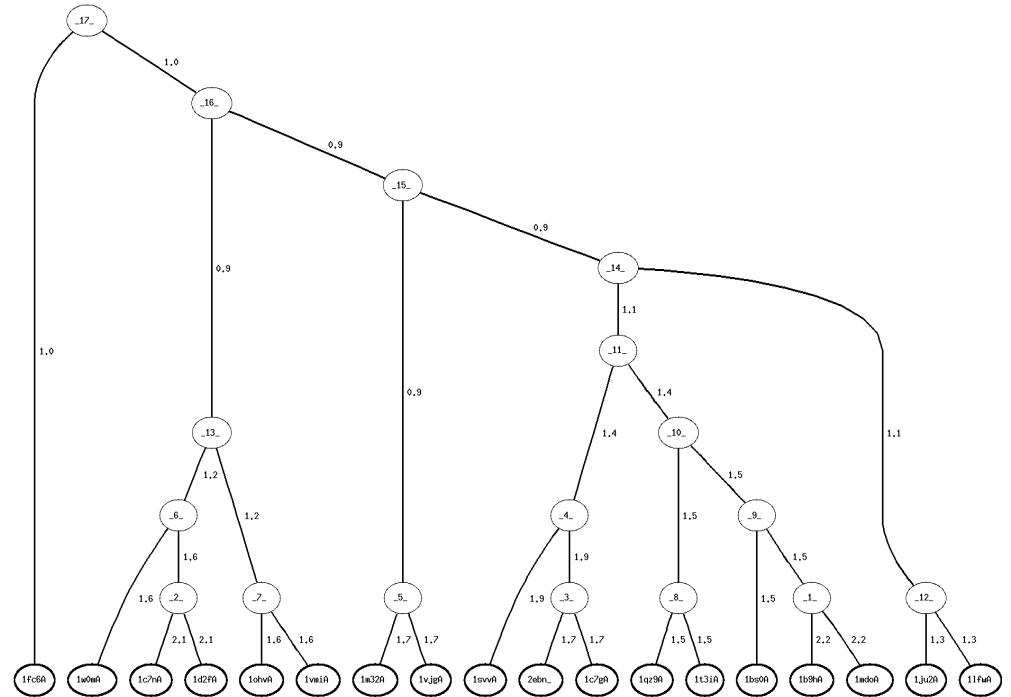
- includes TIM barrel

Peptidases



Methodology

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



- structure comparing ?

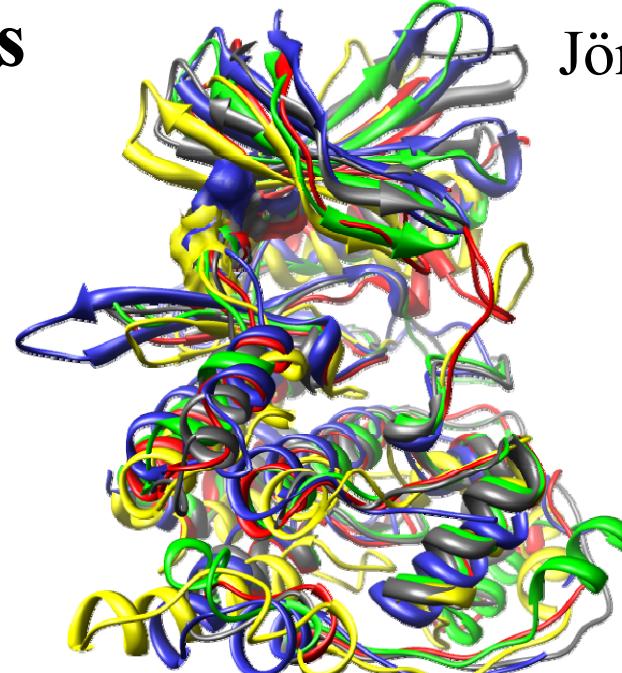
cardigan.zbh.uni-hamburg.de/salami/

- who cares ?

Kinases

Jörn Lenz

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



PDB ID	Accession	Sequence	Length
1omwA	. . TM . .	N . D F S V H . .	181
2srcA	. . PR . .	E . S L R I E . .	202
1lp4A	W G E Q . .	D . D Y E V V . .	51
1jklA	. . V D . .	D . Y Y D T G . .	30
1luFA	. .	I . .	33
1phKA	. .	V R D I G E G A F G R . .	26
1f3mC	. . KK . .	V F Q Q A R . .	42
1cdKA	. . K E . .	L G R G V S S V V R R . .	37
1kwPA	. . SG . .	C I R K P . .	26
1howA	. . K D . .	I K G Q G A S G T V Y . .	42
1b6cB	. . I A . .	T A M D V . .	33
1gngA	. . Q E . .	W E N P A Q N T A H L . .	49
1a06A	. . F S . .	D Q Q . .	39
1csna	. . N V V G V . .	K I G K C R E G E V . .	44
1h1wA	. . P Q P . .	W R . .	24
1hckA	. . M E N . .	G K . .	26
1ir3A	. . V S . .	G E S . .	22
1tkia	. . L Y . .	T V . .	35
1ia8A	. . P F . .	A R . .	27
1muOA	. . V E D W D . .	K A . .	29
1o6yA	. .	R . .	27
1jnkA	. . V L . .	L . .	32
1m14A	. . K E . .	G . .	44
1gjoA	. . D P . .	F . .	44
1o61A	. . T M . .	R . .	35
		N . D F D Y L . .	30
		K L L G K G T F G K V I L V R E K A . .	
		T G . .	

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
- result
 - from known (sequence)
 - to unknown (structure)
- via known structures (threading)
www.zbh.uni-hamburg.de/wurst/
- completely new (Monte Carlo like methods)

