Structural alphabets as tools for the analysis of protein structures

Pr Bernard OFFMANN

Unité Fonctionnalité et Ingénierie des Protéines

CNRS FRE 3478 - Université de Nantes

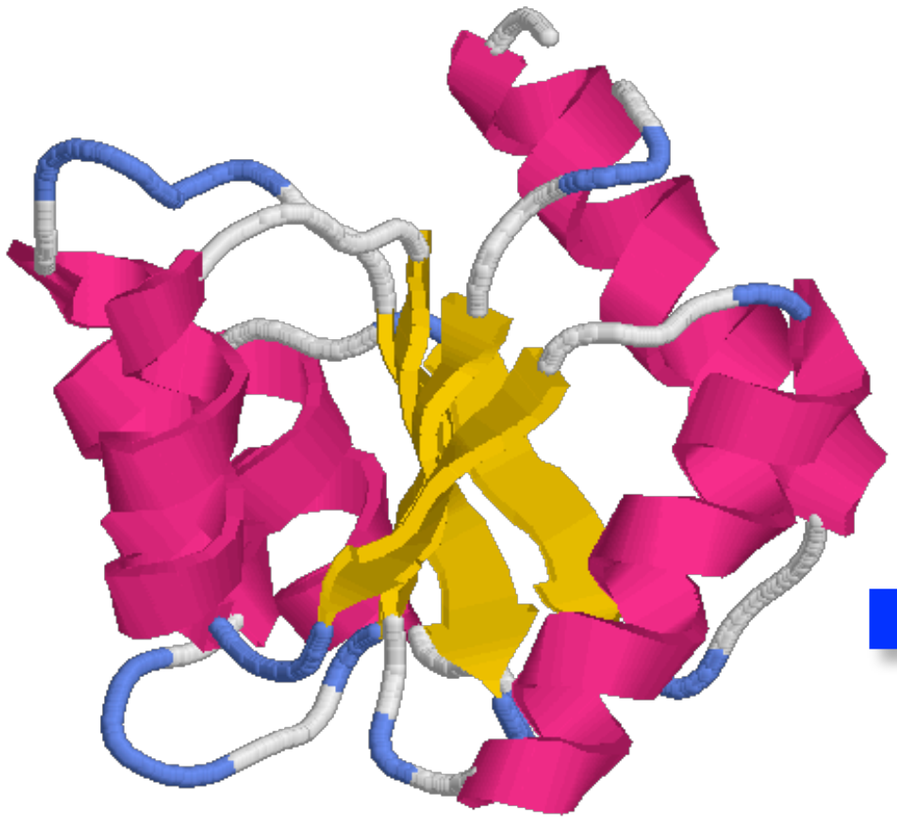
Synopsis

UNIVERSITÉ DE NANTES

- ✓Introduction
  - Classical backbone description
  - Structural alphabets
  - A structural alphabet : Protein Blocks
- ✓ Structure analysis using structural alphabets
- ✓ Mining protein structures
- ✓ Analysis of structural diversity of pentapeptides in protein structures
- ✓Fold recognition

Protein Blocks

Encoding PDB structures into sequences of protein blocks



3D structure

Amino acid sequence

```
SNISRQAYADMFGPTVGDKVRLADTELWIEVEDDLTTYGEEVKF
GGGKVIRDGMGQGQMLAADCVDLVLTNALIVDHWGIVKADIGVK
DGRIFAIGKAGNPDIDIQPNVTIPIGAATEVIAAEGKIVTAGAS
```

PB sequence

```
ZZKBCCDDDDDFBFKLMMMMMMMMMMNOPABDCDDFBFKL
MMMMMMNGOIABDCDDFBDGHILMLMMMMMMMMPMKL
MMPCCDDDDFBDCFKLMMMMMMNOPABDCDDDDDFKL
MMMMMMMMMMMNOZZ
```

PBs when combined together gives back regular structures of a protein and also highlight variable regions present between regular structure elements

Problem overview

UNIVERSITÉ DE NANTES

SNISRQAYADMFGPTVGDKVRLADTELWI
EVEDDLTTYGEEVKFGGGKVIRDGMGQGQ
MLAADCVDLVLTNALIVDHWGIVKADIGV
KDGRIFAIGKAGNPDIDIQPNVTIPIGAA
TEVIAAEGKIVTAGAS

Amino acid sequence

3D structure

Functional analysis

Structure analysis & comparison

de Brevern et al (*Proteins,* 2000)
de Brevern et al (ISB, 2004)
de Brevern et al (BMC Bioinf, 2004)
Etchebest et al, (Proteins, 2005)

de Brevern,
(BBA, 2005)

Sequences of PBs

ZZKBCCDDDDFBFKLMMMMMMMMMMNOPABDCD
DFBFKLMMMMMMNGOIABDCDDFBDGHILMLMM
MMMMMMPMKLMMPCCDDDDFBDCFKLMMMMMMN
OPABDCDDDDDFKLMMMMMMMMMMNOZZ

Main hypotheses

- ✓ PB sequences contain enough information to be able to discriminate protein folds
- ✓ Strong sequence to structure relationships should exist in some situations at a very local level

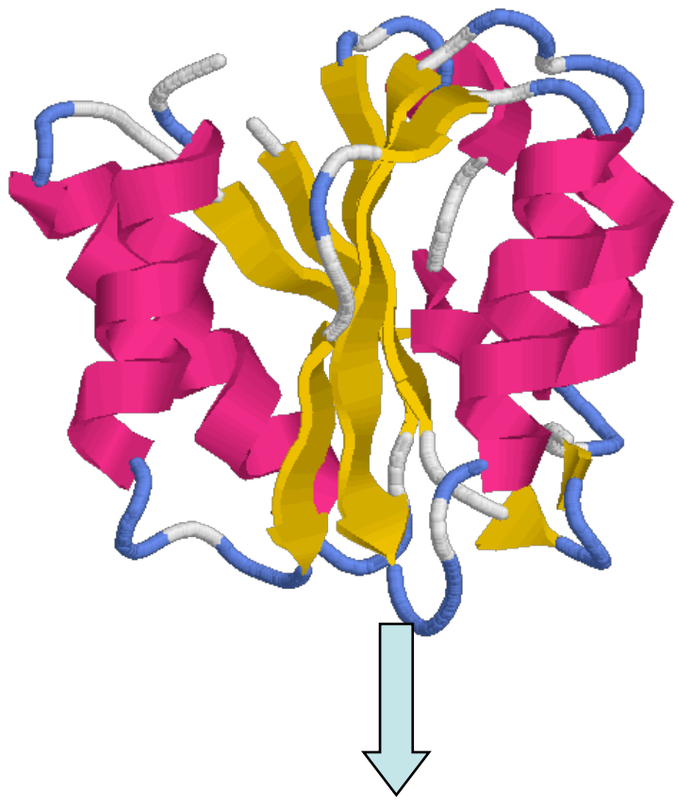
What can we do with PB sequence ?

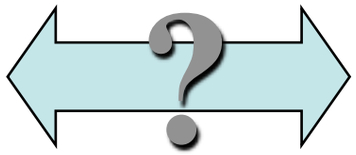
- ✓PB sequence comparison to identify equivalent regions
- ✓Structure comparison based on sequence alignment algorithm
- ✓Identification of conformational change or rigid body displacement/shift in proteins
- ✓Extension of above to study active & inactive states of enzymes
- ✓Mining structure databases for similar folds
- ✓Fold recognition
- ✓Protein design

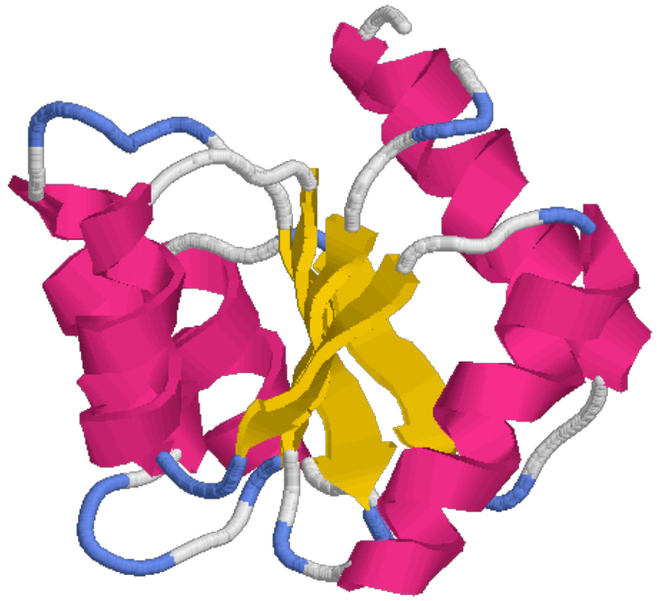
What can we do with PB sequence ?

UNIVERSITÉ DE NANTES

3chy







2fox

KBCCDDDDFBFKLMMMMMMMMMMNOPABDCDDFBF
KLMMMMMMNGOIABDCDDFBDGHILMLMMMMMMMM
PMKLMMPCCDDDDFBDCFKLMMMMMMNOPABDCDD
DDDFKLMMMMMMMMMMNO

DDDDFKOMMMMMMMMMMMMMMMMNOPACDDDDDFKLP
CFKLMMMPFBDCDDDDDEHJLPACFKLMMMMMMMM
MMMMGHIACDDEHIAFKLNOMMMMMMMMMMMMNOP
ACFBACDDDEHJACKLMMMMMMMMMMMM

Structure comparison using PBs ?

Our approach

Sequence of PBs

PB-ALIGN

Algorithm : Needleman & Wunch

Alignment method : local or global

Substitution matrix

Scoring function

Alignment of PB sequences

Mining for similar structures

Structure comparison

Conformational changes & movements

Performance evaluation

Generating a PB substitution matrix

Substitution matrix calculation

Database of aligned sequences derived from structural superimposition of homologous structures

Transformation of PALI alignments into aligned PB sequences

Alignments of PB seq.

PB substitution counts in structurally equivalent regions

Normalized scores

PB substitution matrix (16x16)

Tyagi M, Venkataraman SG, Srinivasan N, de Brevern AG, Offmann B. A substitution matrix for structural alphabet based on structural alignment of homologous proteins and its applications. *Proteins* 65(1):32-9, (2006).

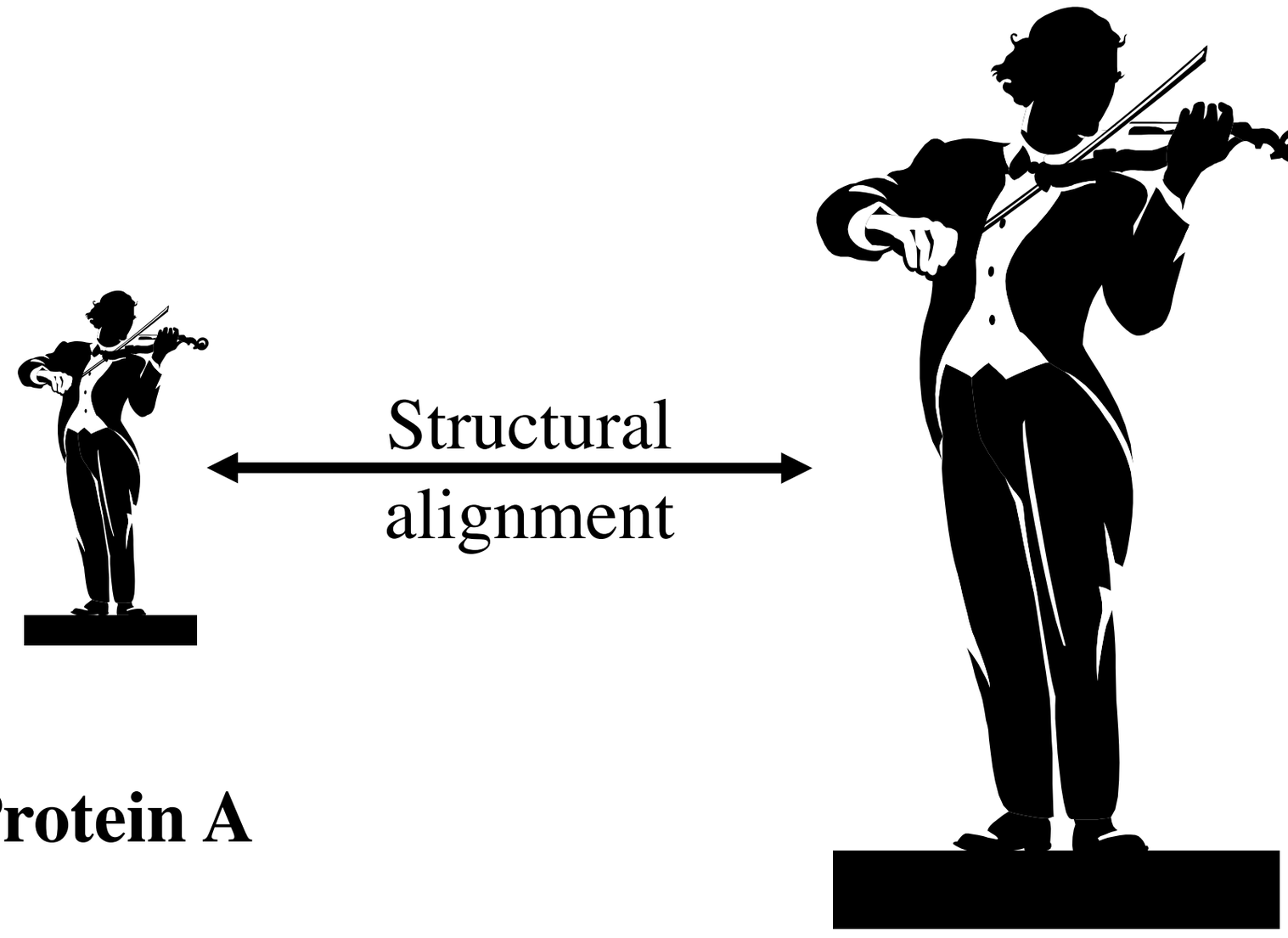
PB substitution matrix

Protein blocks	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*	*n*	*o*	*p*
a	2.28															
b	-0.12	2.49														
c	0.54	-0.21	1.69													
d	-0.29	-0.44	0.17	1.35												
e	-1.59	-0.48	-1.10	-0.36	3.05											
f	-0.54	-1.53	-0.39	-0.49	0.75	2.21										
g	0.31	-0.73	0.18	-1.29	1.37	-0.33	3.25									
h	-1.14	0.20	-1.63	-1.20	0.66	-0.34	-0.74	3.07								
i	0.39	0.24	-1.11	-1.12	-1.15	-1.07	-0.19	-0.92	3.37							
j	-1.15	0.32	-1.03	-0.92	-0.76	-0.34	-0.51	1.18	1.54	3.74						
k	-1.75	-0.03	-2.45	-2.63	-0.38	-0.04	-1.39	0.51	-0.15	0.07	2.52					
l	-0.60	0.04	-2.21	-1.56	-1.76	-0.33	-0.74	-0.36	-0.22	-0.12	0.19	2.24				
m	-2.40	-2.98	-2.70	-5.20	-4.75	-2.14	-1.10	-2.93	-3.15	-2.00	-1.02	-0.68	1.06			
n	-1.40	-0.83	-1.68	-3.07	-0.58	-1.99	1.07	-1.07	-0.97	-0.44	-0.56	-0.27	-0.77	3.65		
o	-0.54	-0.55	-0.65	-2.66	-2.48	-1.41	-0.01	0.96	-0.89	-0.48	-1.71	0.06	-1.26	0.26	3.36	
p	-0.36	0.33	-0.01	-2.10	-2.22	-1.91	0.47	-1.81	1.32	0.60	-1.35	-1.23	-1.10	0.36	0.24	2.83

Tyagi M, Venkataraman SG, Srinivasan N, de Brevern AG, Offmann B. A substitution matrix for structural alphabet based on structural alignment of homologous proteins and its applications. *Proteins* 65(1):32-9, (2006).

PB alignment for structure comparison

Comparing protein structures



Protein A

Protein B

Not a trivial problem…

PB alignment for structure comparison

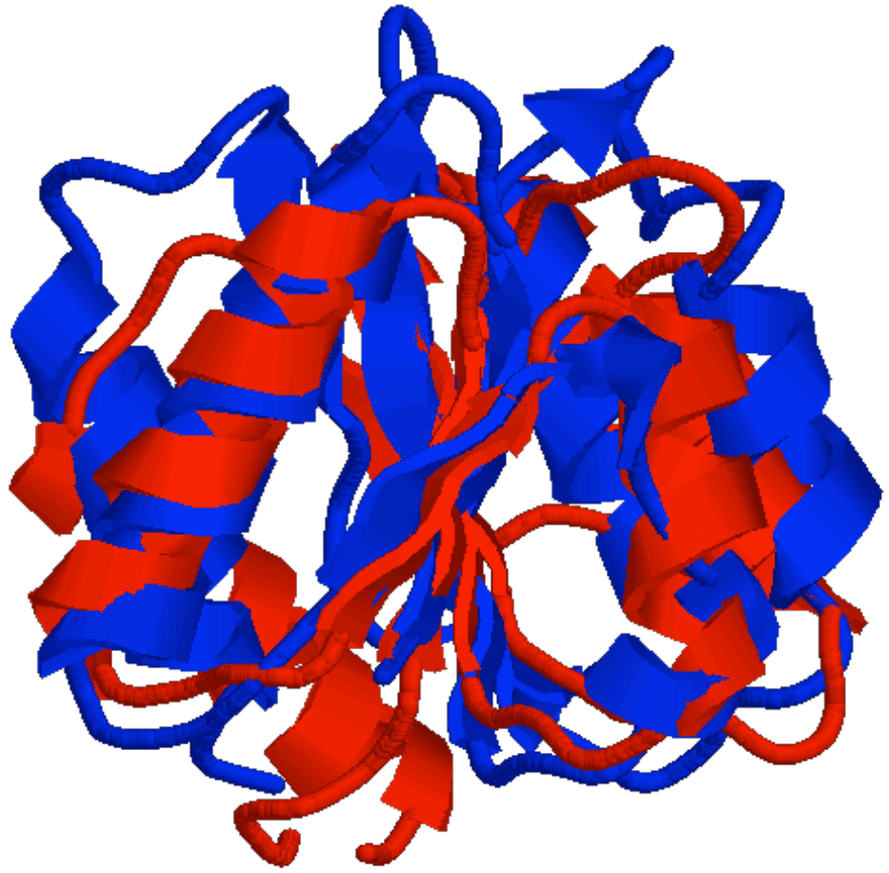
PB alignment using dynamic programming and substitution matrix

```
3chy_:  KBCCDDDDFBF--KLMMMMMMMMMMNOPABDCDDFBFKLMMMMMMNGOIABDCDDFBDGHI
2fox_:  ---DDDDFKOMMMMMMMMMMMMMMMNOPACDDDD--FKLPCFKLMMM-PFBDCDDDDDEHJ

3chy_:  LM--LMMMMMMMMPMKLMMPCCDDDDFBDCFKL-----MM-MMM-M-NOPA--BDCDDDD-
2fox_:  LPACFKLMMMMMMMMMMMMGHIACDDEHIAFKLNOMMMMMMMMMMMMNOPACFBACDDDEH

3chy_:  -DFKLMMMMMMMMMMMNO
2fox_:  JACKLMMMMMMMMMMMM
```

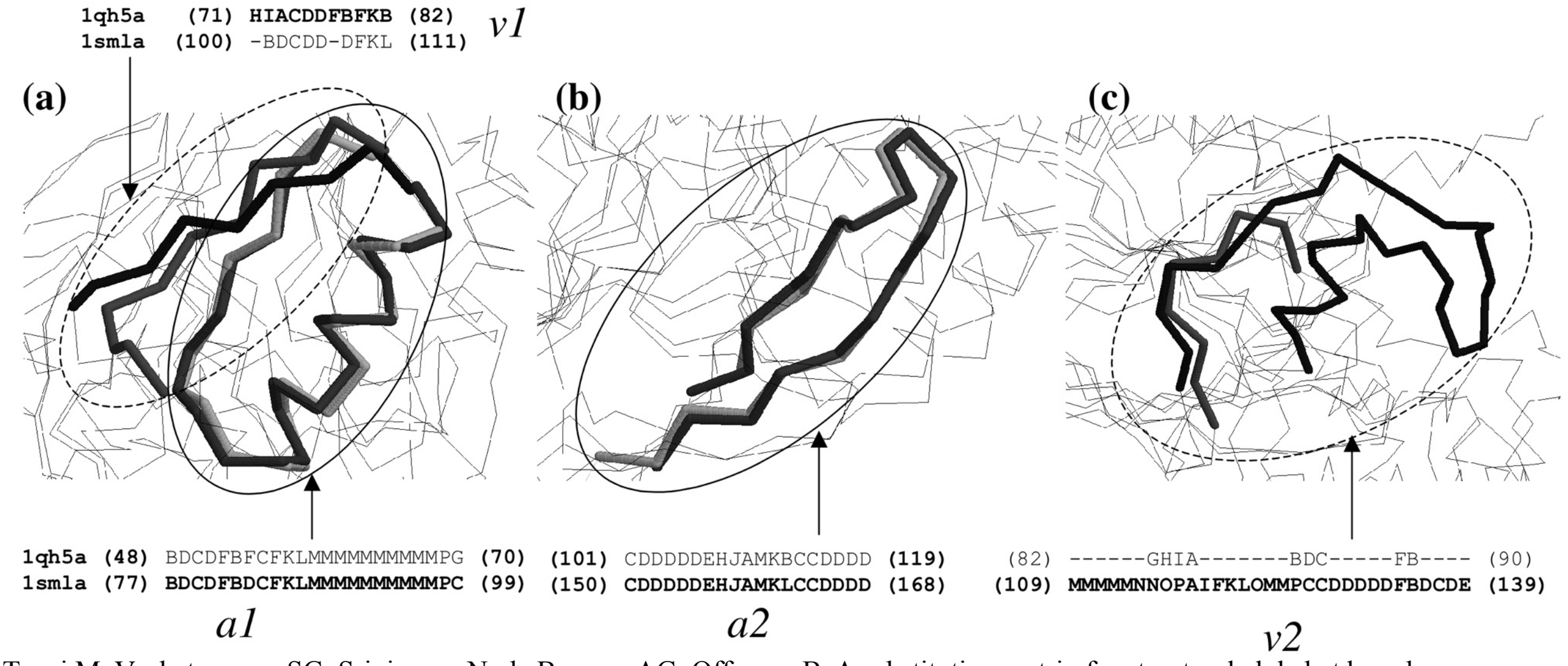
PB-ALIGN



PB alignment for structure comparison

UNIVERSITÉ DE NANTES

Metallohydrolase superfamily members 1qh5:a and 1sml:a of equivalent length

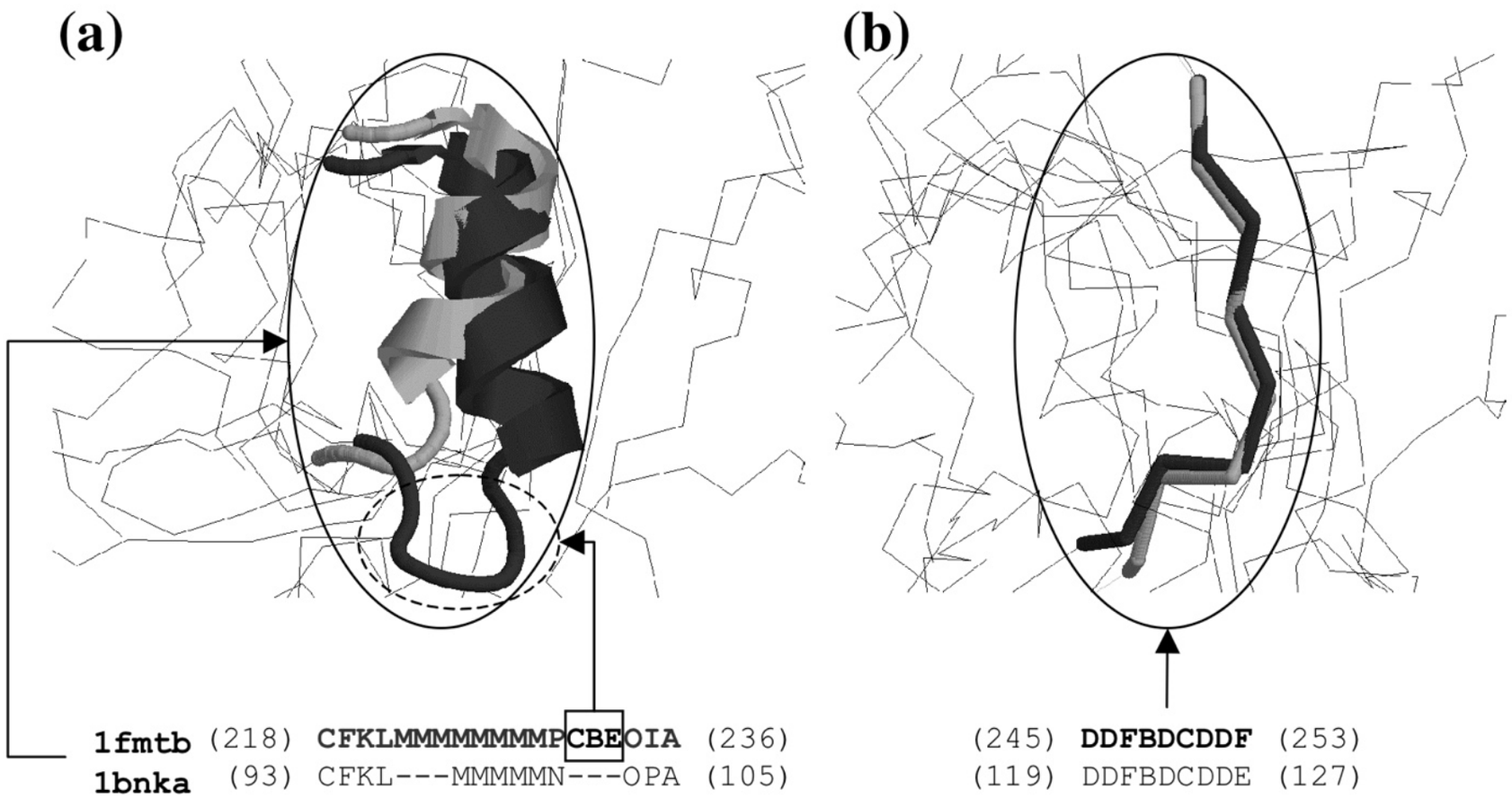


Tyagi M, Venkataraman SG, Srinivasan N, de Brevern AG, Offmann B. A substitution matrix for structural alphabet based on structural alignment of homologous proteins and its applications. *Proteins* 65(1):32-9, (2006).

PB alignment for structure comparison

UNIVERSITÉ DE NANTES

- ✓1bnk:a & 1fmt:b from all beta class FMT C-terminal domain like superfamily



Tyagi M, Venkataraman SG, Srinivasan N, de Brevern AG, Offmann B. A substitution matrix for structural alphabet based on structural alignment of homologous proteins and its applications. *Proteins* 65(1):32-9, (2006).

PB alignment for structure comparison

UNIVERSITÉ DE NANTES

PB-ALIGN as efficient as FATCAT for the **flexible** alignment of structures

Protein1	Protein2	FATCAT	PB-ALIGN
1fxiA	1ubq_	63 (3.01)	59 (2.6)
1ten_	3hhrB	87 (1.9)	82 (4.1)
3hlaB	2rhe_	79 (2.81)	67 (2.4)
2azaA	1paz_	87 (3.01)	79 (2.3)
1cewI	1molA	83 (2.44)	74 (2.5)
1cid_	2rhe_	100 (3.11)	87 (2.2)
1crl_	1ede_	269 (3.55)	179 (2.3)
2sim_	1nsbA	286 (3.07)	262 (2.4)
1bgeB	2gmfA	100 (3.19)	90 (2.4)
1tie_	4fgf_	117 (3.05)	105 (2.2)

Tyagi M., de Brevern A., Srinivasan N., Offmann B. Protein structure mining using a structural alphabet. *Proteins*, (2007).

Further analysis…

Distinguishing between conformational changes and rigid body movement in superimposed structures

Structure comparison using PBs

Structure comparison using PBs

- ✓Rigid body superimposition methods e.g. STAMP report sequence alignment based on structurally equivalent and variable regions

Structure comparison using PBs

- ✓Rigid body superimposition methods e.g. STAMP report sequence alignment based on structurally equivalent and variable regions
- ✓A cutoff/threshold residue residue rmsd value is used to define variable regions

Structure comparison using PBs

- ✓Rigid body superimposition methods e.g. STAMP report sequence alignment based on structurally equivalent and variable regions
- ✓A cutoff/threshold residue residue rmsd value is used to define variable regions
- ✓Are these high rmsd values, due to difference in conformations or rigid body displacement of equivalent regions ?

Structure comparison using PBs

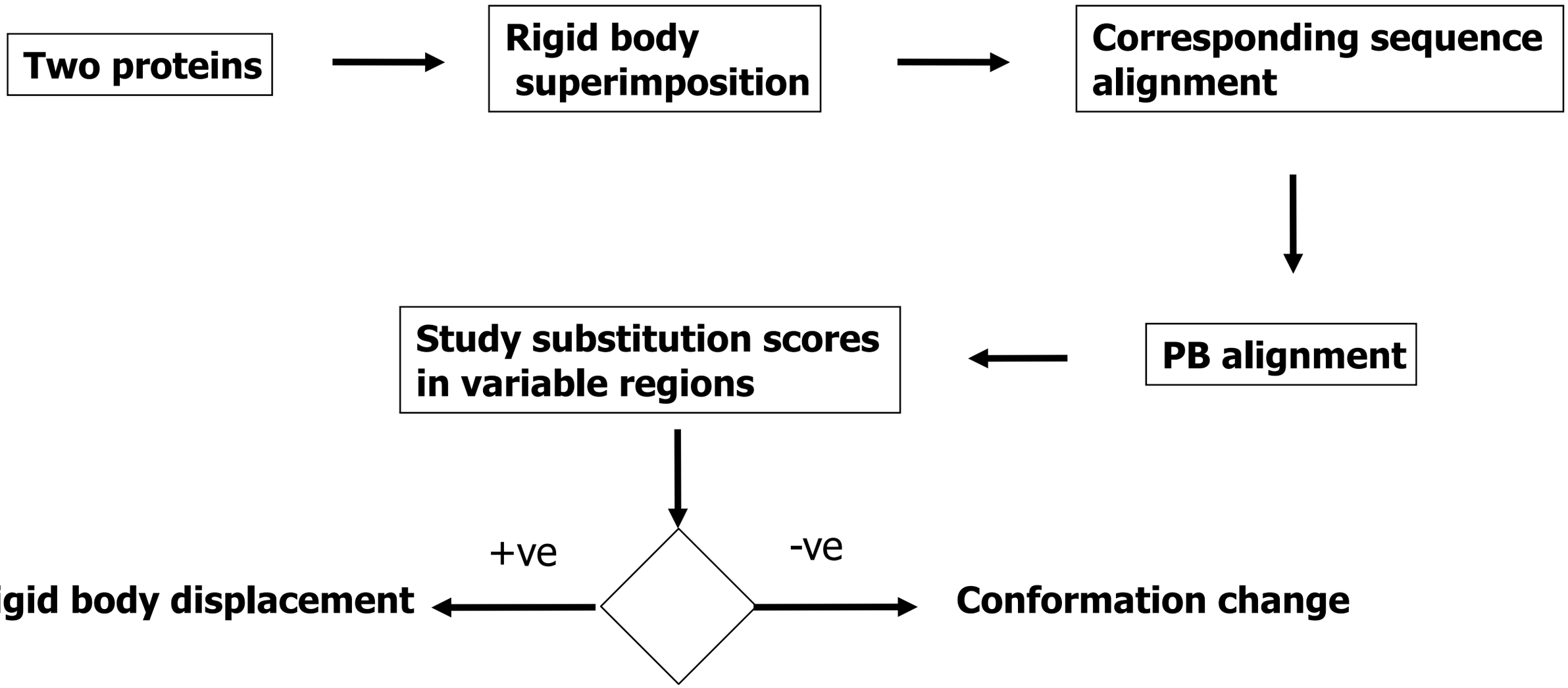
- ✓Rigid body superimposition methods e.g. STAMP report sequence alignment based on structurally equivalent and variable regions
- ✓A cutoff/threshold residue residue rmsd value is used to define variable regions
- ✓Are these high rmsd values, due to difference in conformations or rigid body displacement of equivalent regions ?

Structure comparison using PBs

- ✓Rigid body superimposition methods e.g. STAMP report sequence alignment based on structurally equivalent and variable regions
- ✓A cutoff/threshold residue residue rmsd value is used to define variable regions
- ✓Are these high rmsd values, due to difference in conformations or rigid body displacement of equivalent regions ?

We don’t know from simple rigid body structure alignment

Structure comparison using PBs



Tyagi M, Venkataraman SG, Srinivasan N, de Brevern AG, Offmann B. A substitution matrix for structural alphabet based on structural alignment of homologous proteins and its applications. *Proteins* 65(1):32-9, (2006).

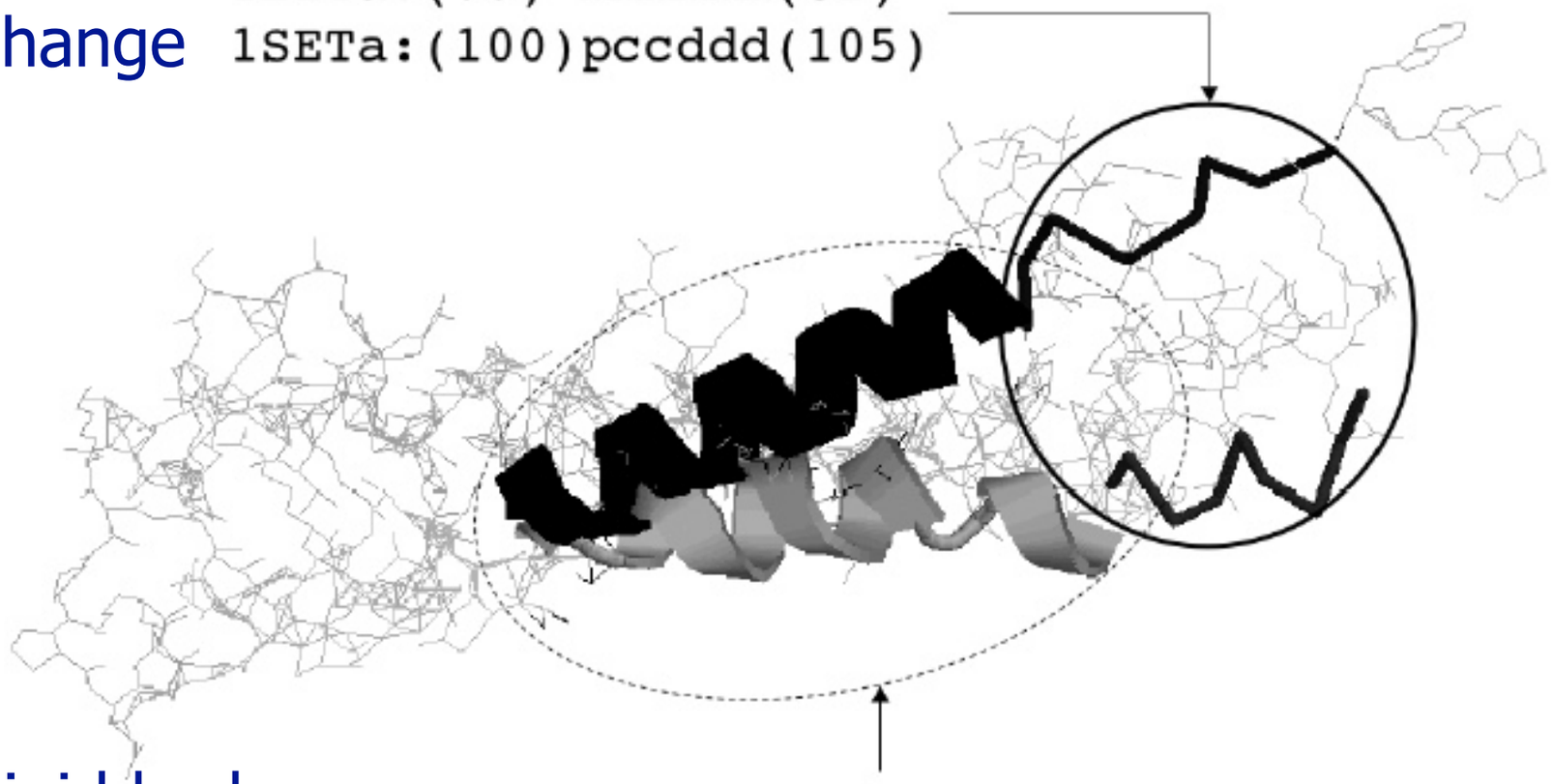
Structure comparison using PBs

UNIVERSITÉ DE NANTES

Comparing two distantly related tRNA synthetases

Conformational change

```
1EIYa:(77) mmmmmm(82)
1SETa:(100)pccddd(105)
```



Rigid body movement

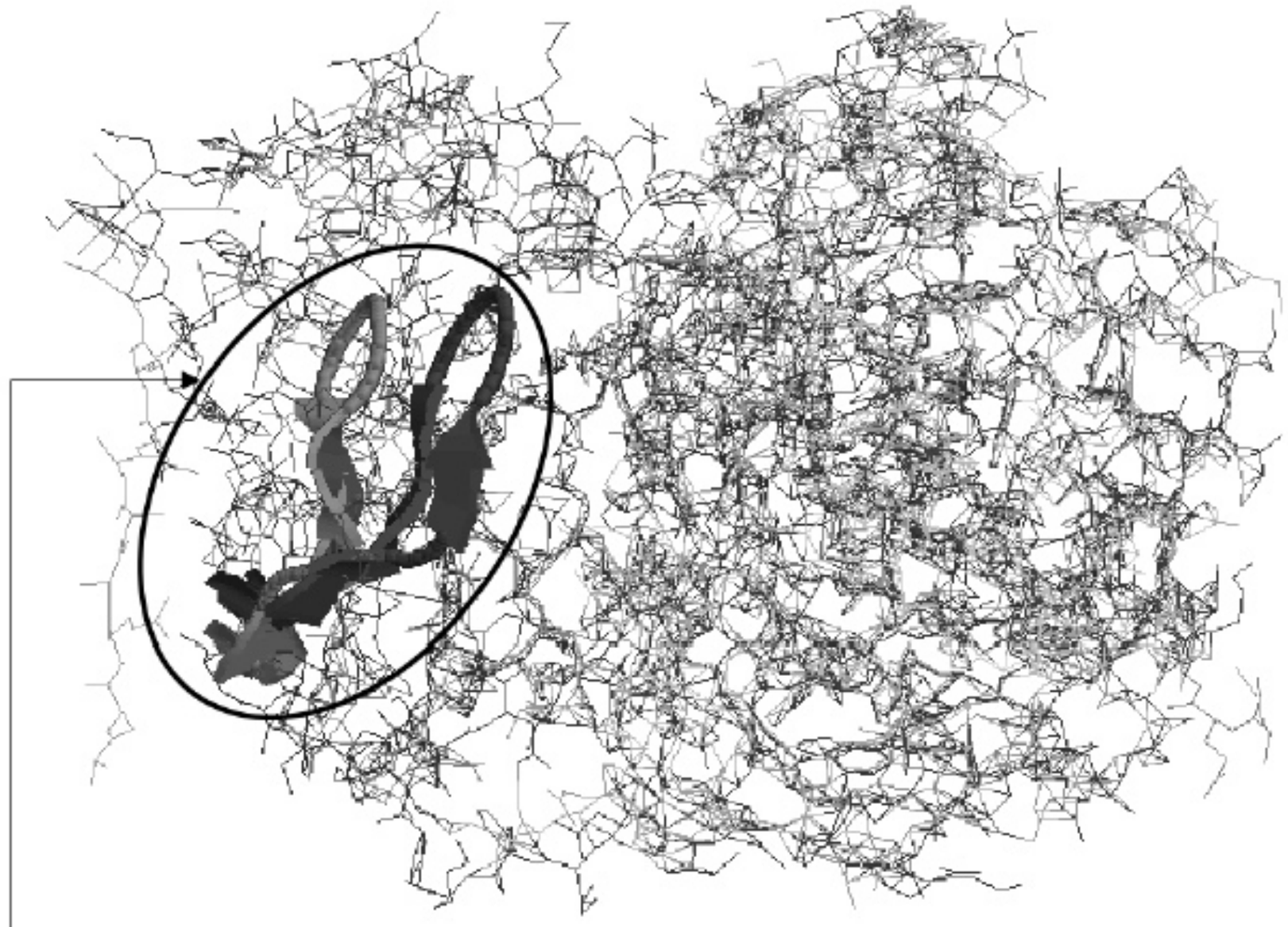
```
1EIYa:(59)mmmmmmmmmmmmm-mmmmm(76)
1SETa:(81)mmmmmmmmmmmmmmmmmmmm(99)
```

Tyagi M, Venkataraman SG, Srinivasan N, de Brevern AG, Offmann B. A substitution matrix for structural alphabet based on structural alignment of homologous proteins and its applications. *Proteins* 65(1):32-9, (2006).

Structure comparison using PBs

UNIVERSITÉ DE NANTES

Rigid body movement upon enzyme activation in cyclic AMP dependent protein kinases



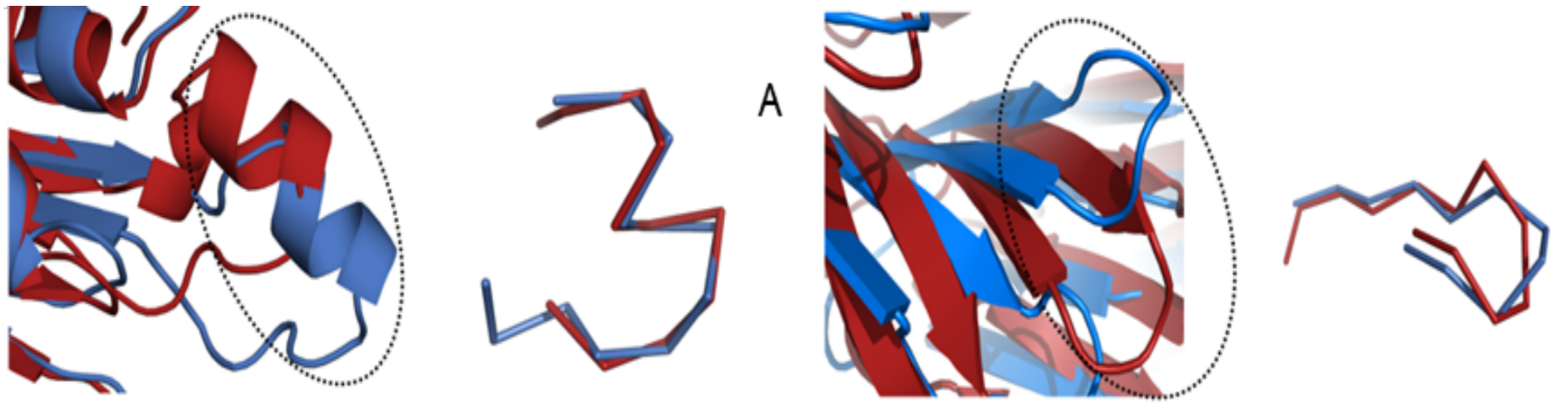
```
1ATPe: (44)CFBdcbdFKLMMCCD (58)
1J3Ha: (44)CFBaghpFKLMMCCD (58)
```

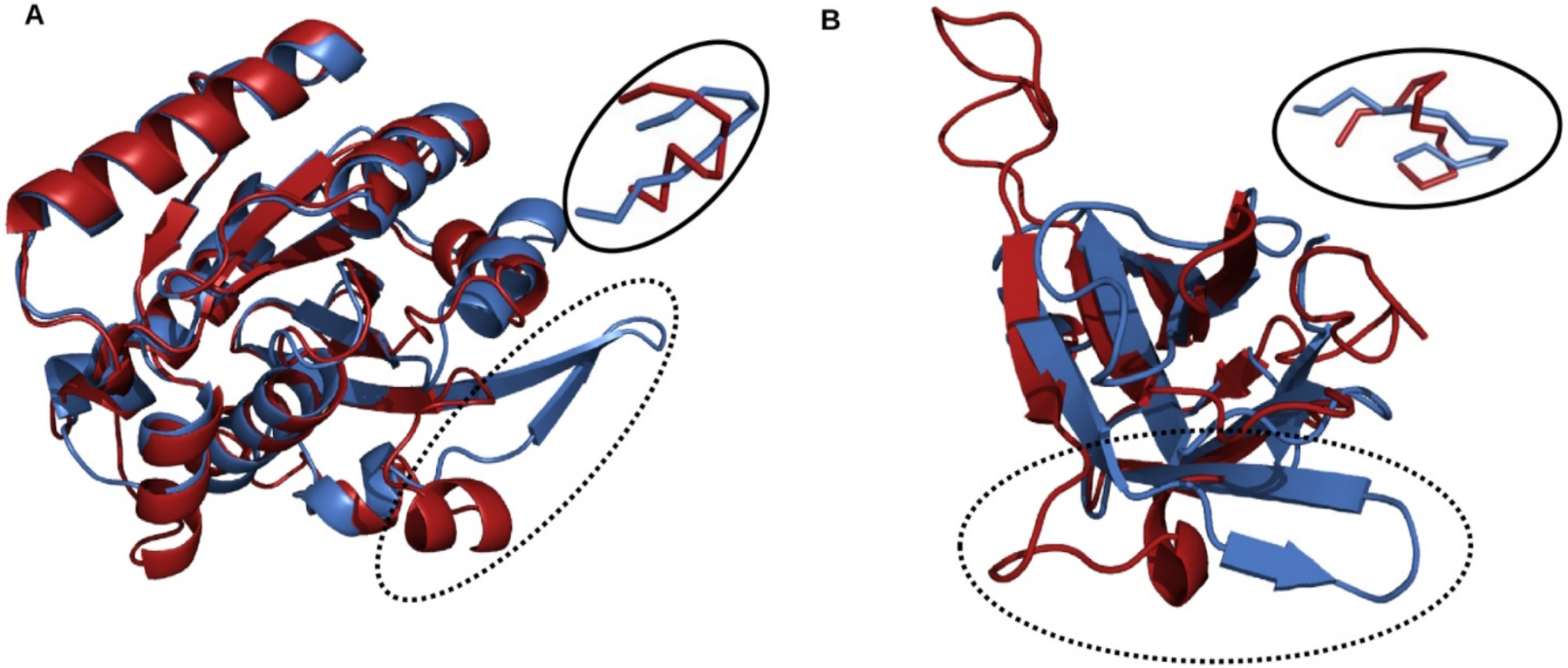
Structurally Variable Regions (SVRs)

Cα- Cα RMSD > 3 Å

Structurally Variable Regions (SVRs)

Cα- Cα RMSD > 3 Å



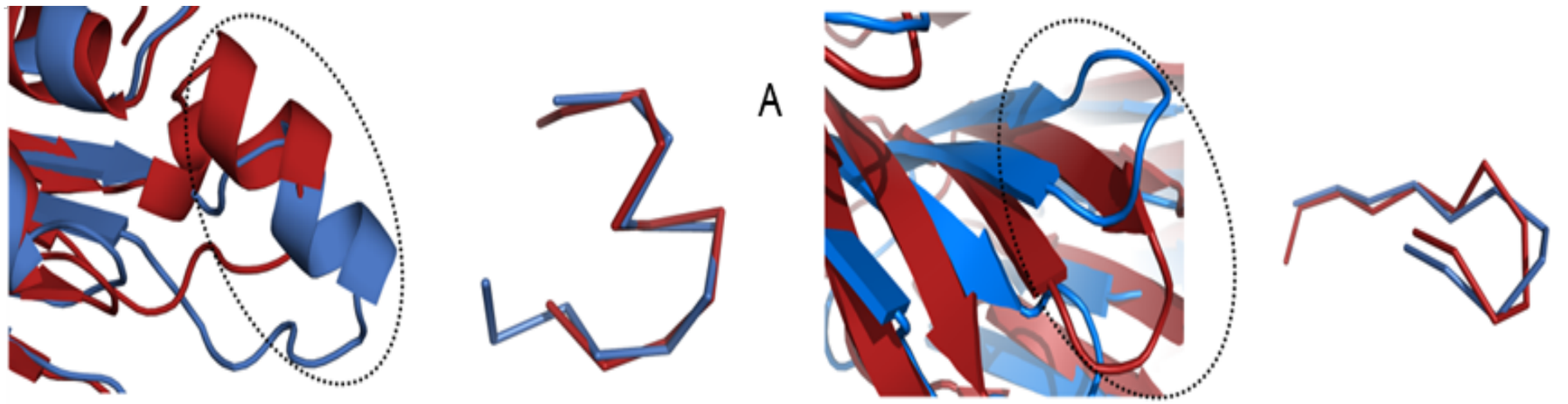


Agarwal, **Mahajan** et al, (2011) Identification of Local Conformational Similarity in Structurally Variable Regions of Homologous Proteins Using Protein Blocks. *PLoS ONE*, 6, e17826.

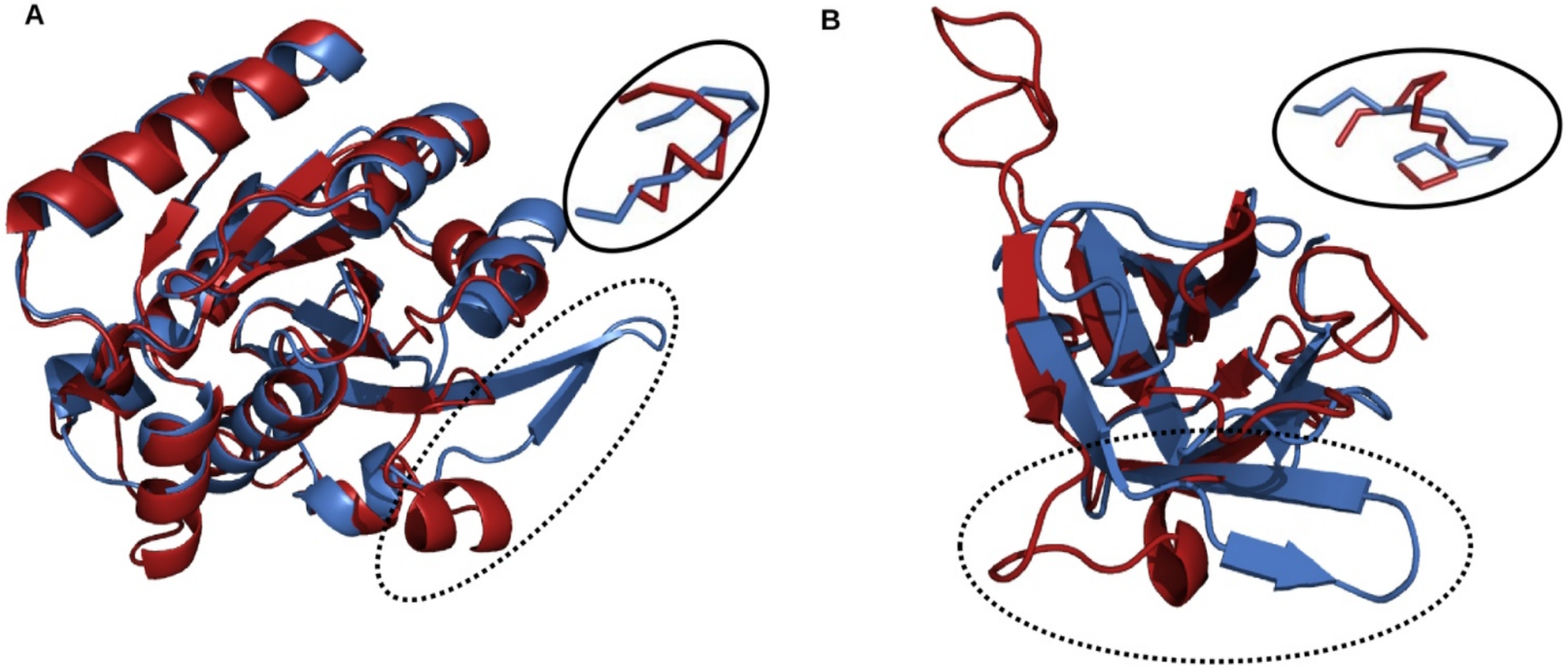
Structurally Variable Regions (SVRs)

Cα- Cα RMSD > 3 Å

Conformationally similar SVRs



Conformationally dissimilar SVRs



Agarwal, **Mahajan** et al, (2011) Identification of Local Conformational Similarity in Structurally Variable Regions of Homologous Proteins Using Protein Blocks. *PLoS ONE*, 6, e17826.

DoSA: Database of Structural Alignments

6,420 domains
1,867 protein domain families in PALI.
62,730 pairwise alignments
542,610 SCRs (structurally conserved regions)
347,062 SVRs

159,780 (74%) conformationally similar SVRs
56,140 (26%) conformationally dissimilar SVRs

DoSA: Database of Structural Alignments

6,420 domains
1,867 protein domain families in PALI.
62,730 pairwise alignments
542,610 SCRs (structurally conserved regions)
347,062 SVRs

159,780 (74%) conformationally similar SVRs
56,140 (26%) conformationally dissimilar SVRs

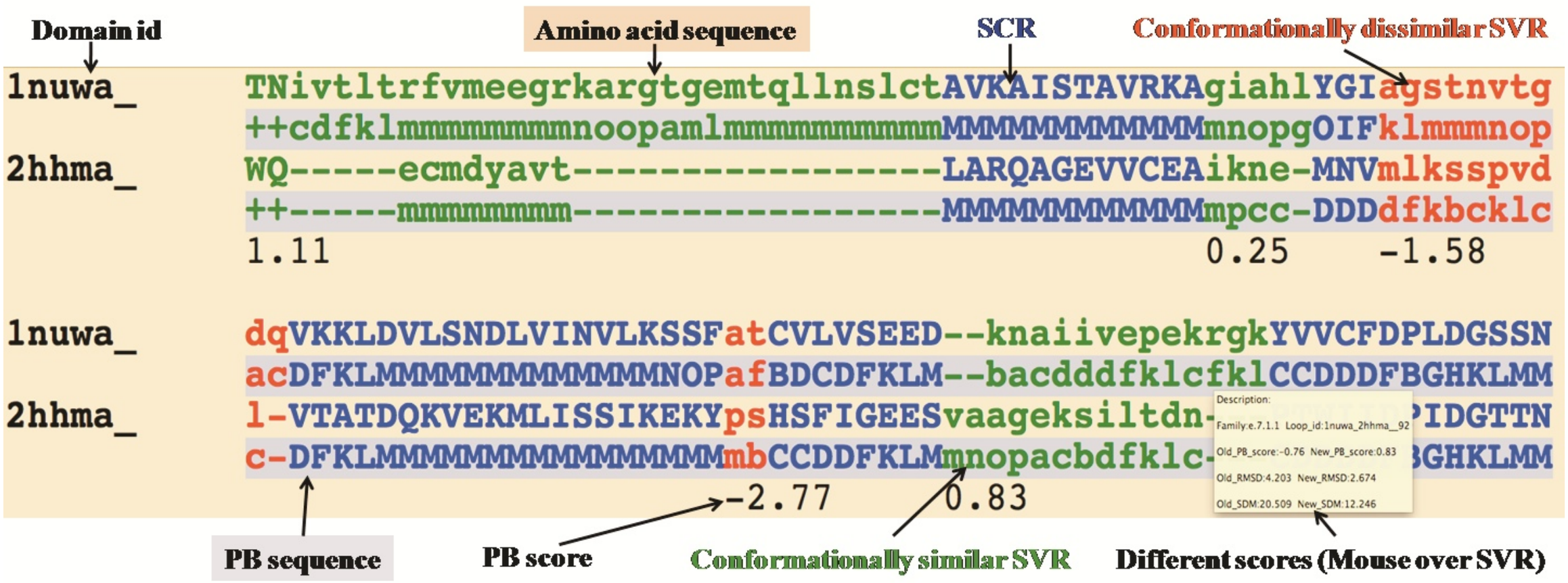
Mahajan et al, (2013) DoSA: Database of Structural Alignments. Database (oxford), 2013:bat048–bat048.

DoSA: Database of Structural Alignments

6,420 domains
1,867 protein domain families in PALI.
62,730 pairwise alignments
542,610 SCRs (structurally conserved regions)
347,062 SVRs

159,780 (74%) conformationally similar SVRs
56,140 (26%) conformationally dissimilar SVRs

http://www.bo-protscience.fr/dosa/

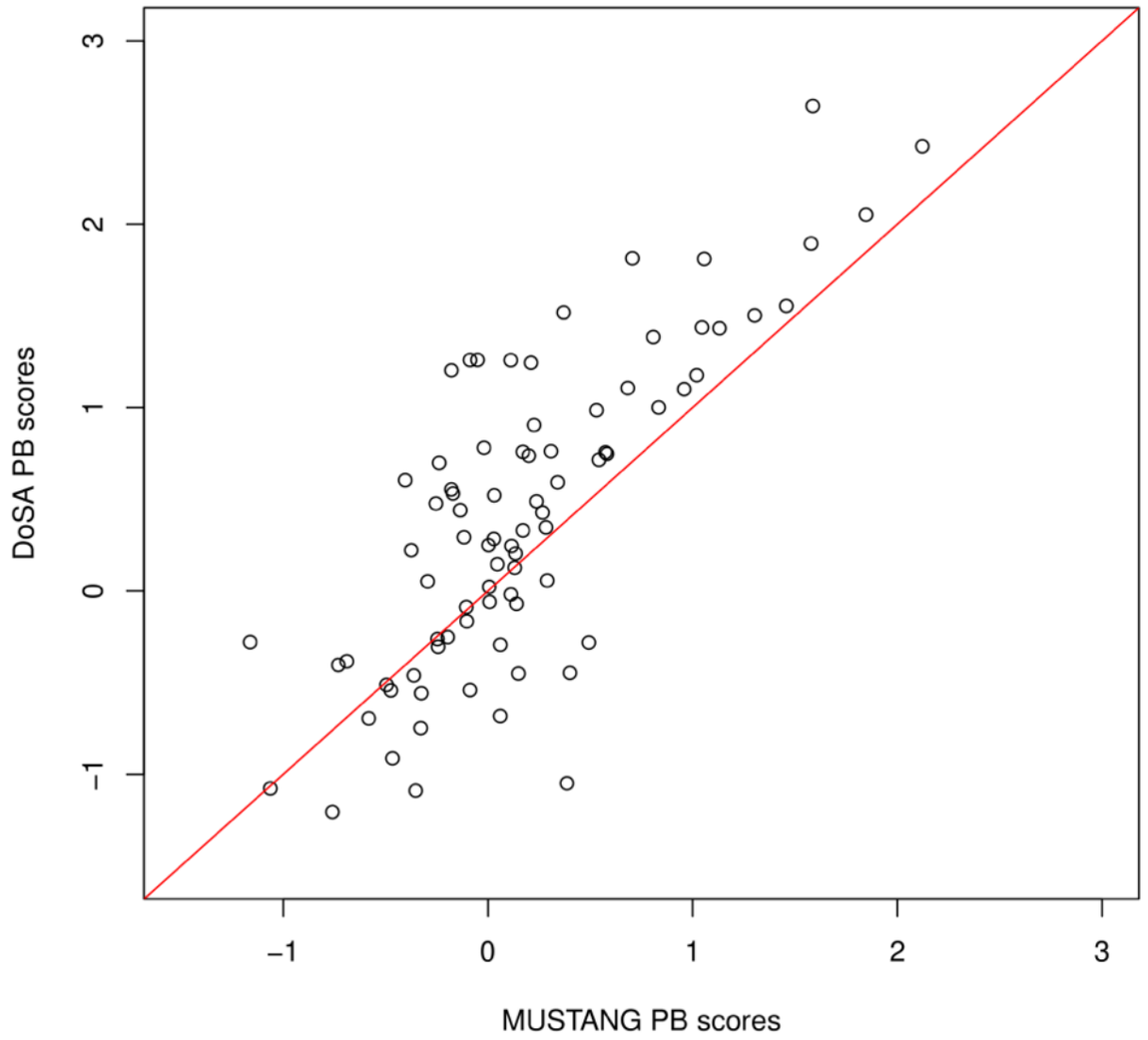


Mahajan et al, (2013) DoSA: Database of Structural Alignments. Database (oxford), 2013:bat048–bat048.

DoSA: Database of Structural Alignments

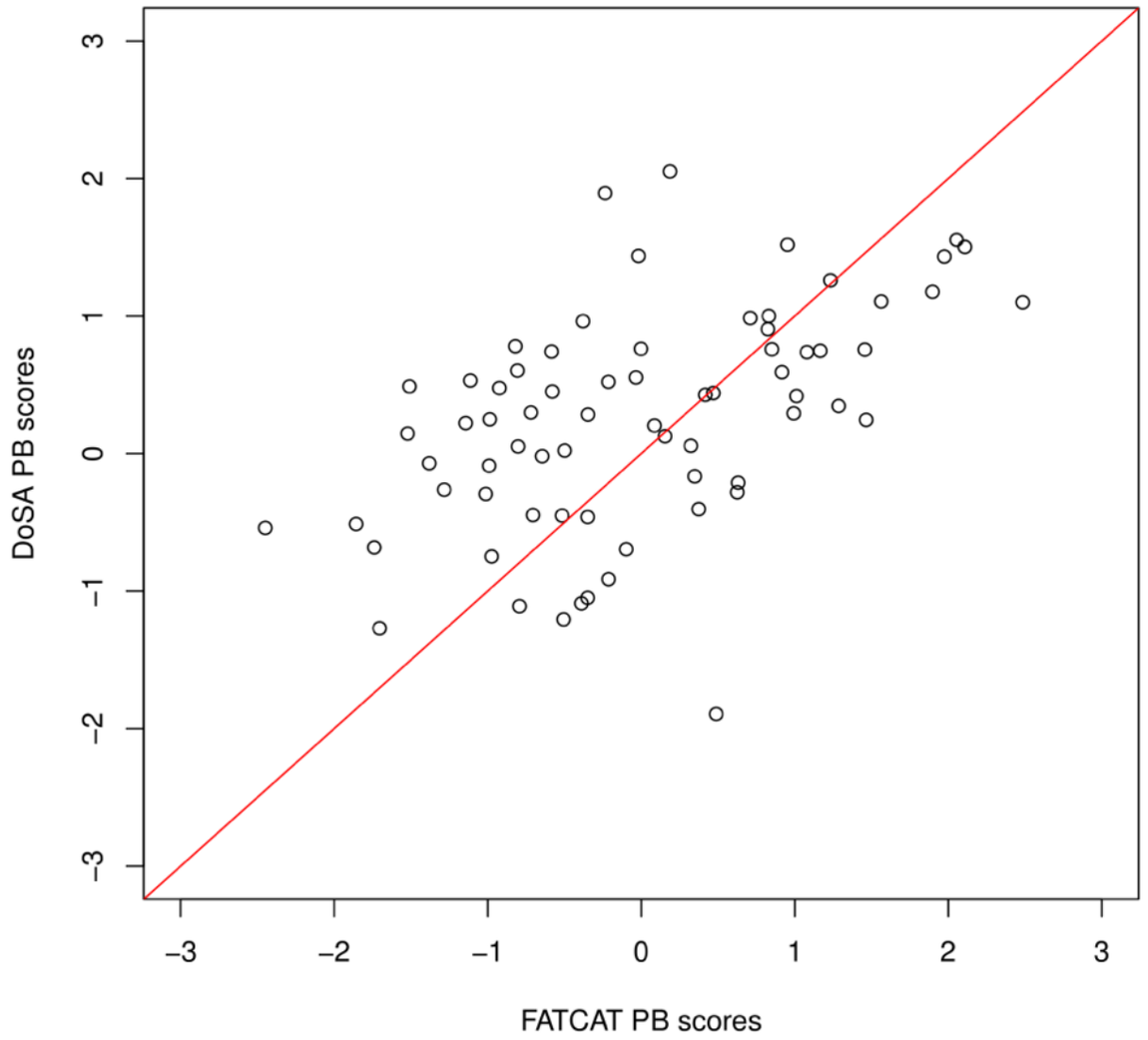
MUSTANG vs DoSA

67.9% of alignments with better PB scores



FATCAT vs DoSA

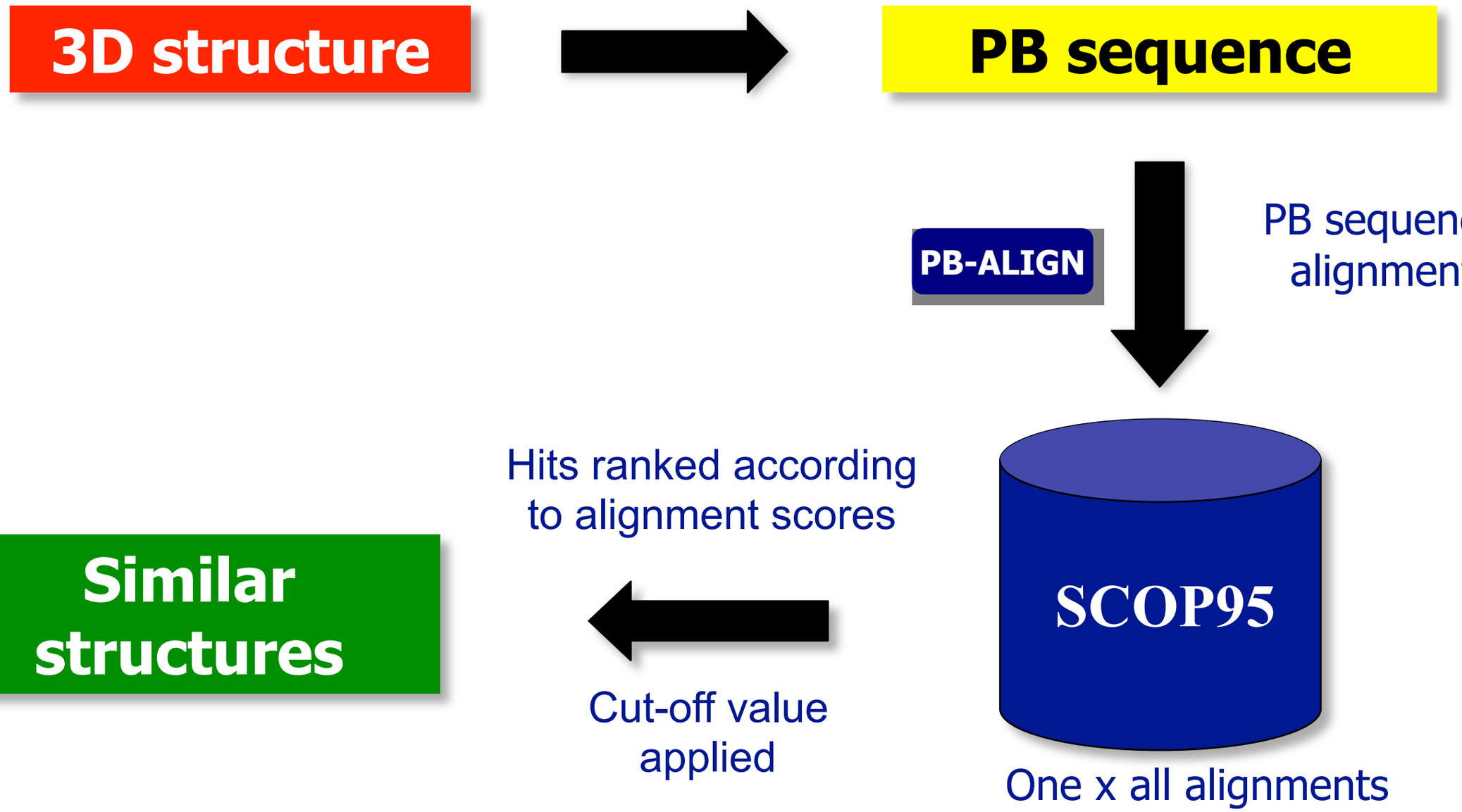
57.8 % of alignments with better PB scores



Mahajan et al, (2013) DoSA: Database of Structural Alignments. Database (oxford), 2013:bat048–bat048.

Mining protein structures using PBs

Mining protein structures using PBs



Tyagi M., de Brevern A., Srinivasan N., Offmann B. Protein structure mining using a structural alphabet. *Proteins*, (2007).

Mining protein structures using PBs

UNIVERSITÉ DE NANTES

3D structure → **PB sequence**

Similar to Fasta and Blast approaches…

PB-ALIGN

PB sequence alignment

SCOP95

One x all alignments

Hits ranked according to alignment scores

Cut-off value applied

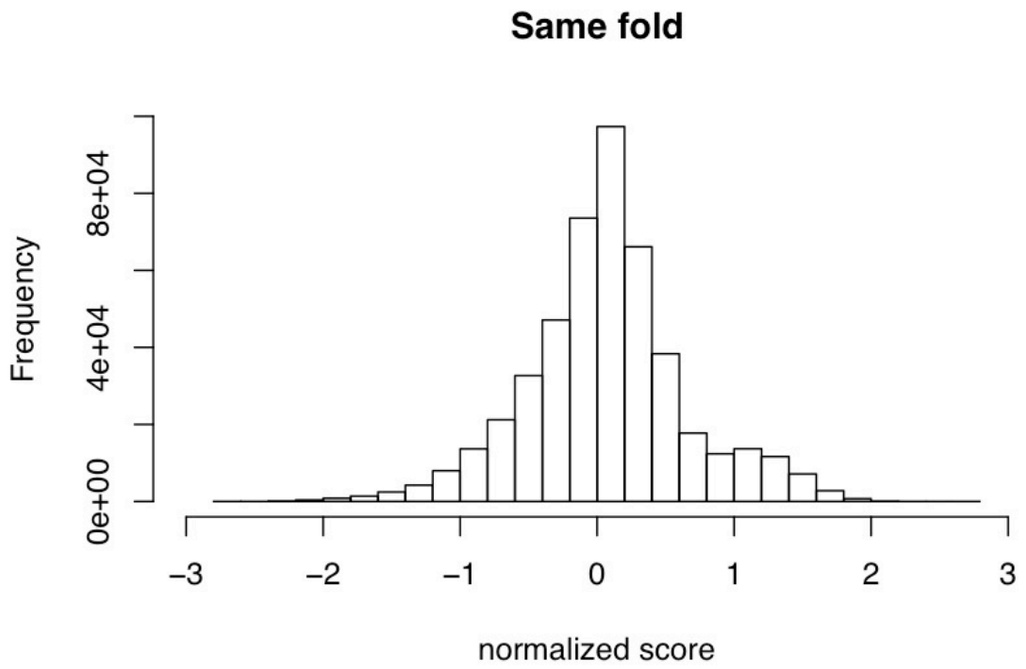
Similar structures

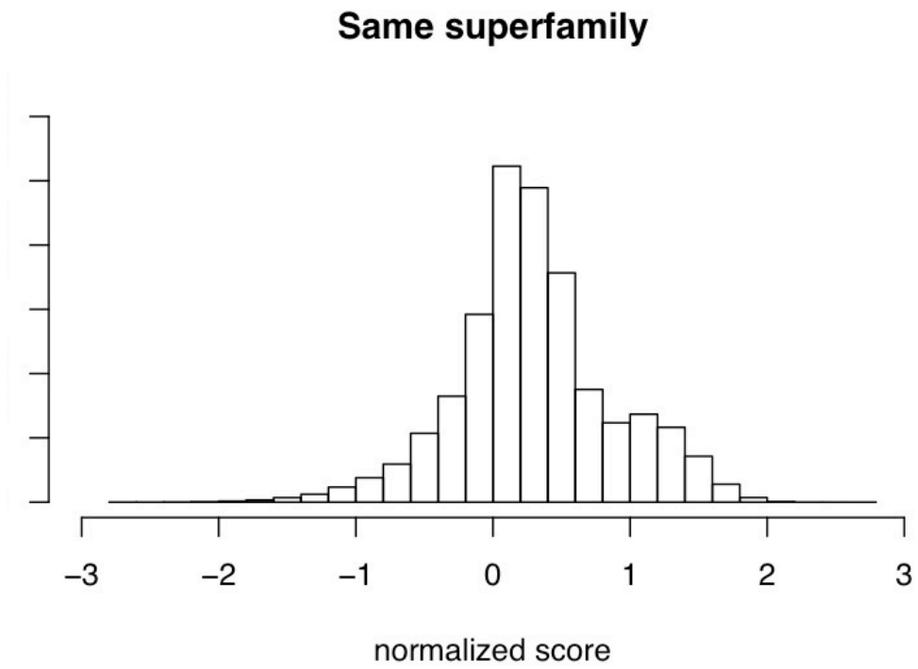
Tyagi M., de Brevern A., Srinivasan N., Offmann B. Protein structure mining using a structural alphabet. *Proteins*, (2007).

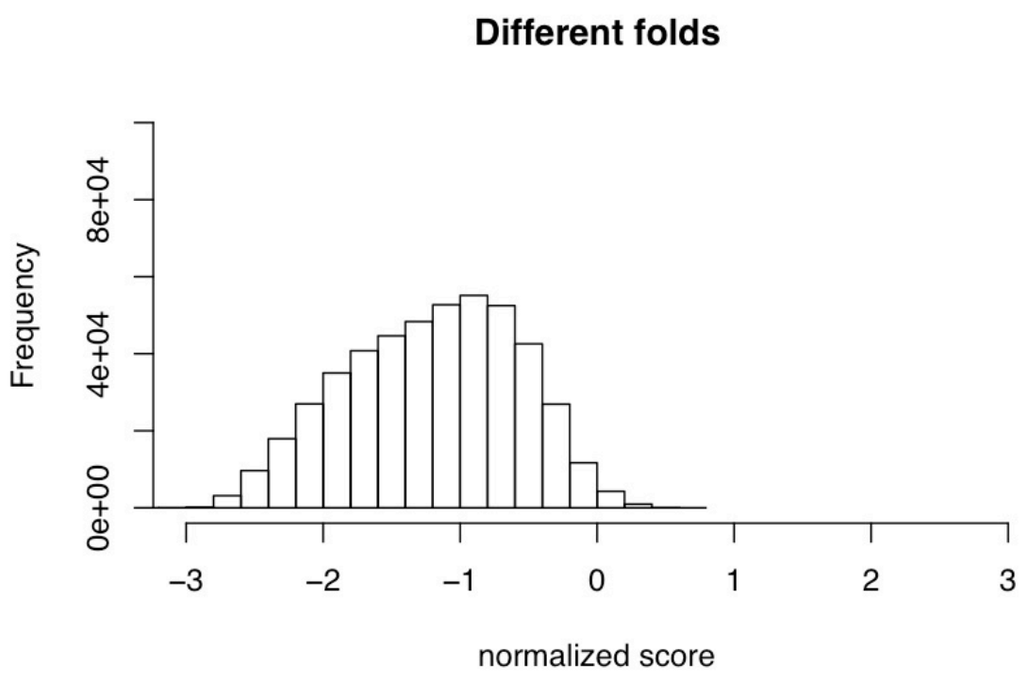
Mining protein structures using PBs

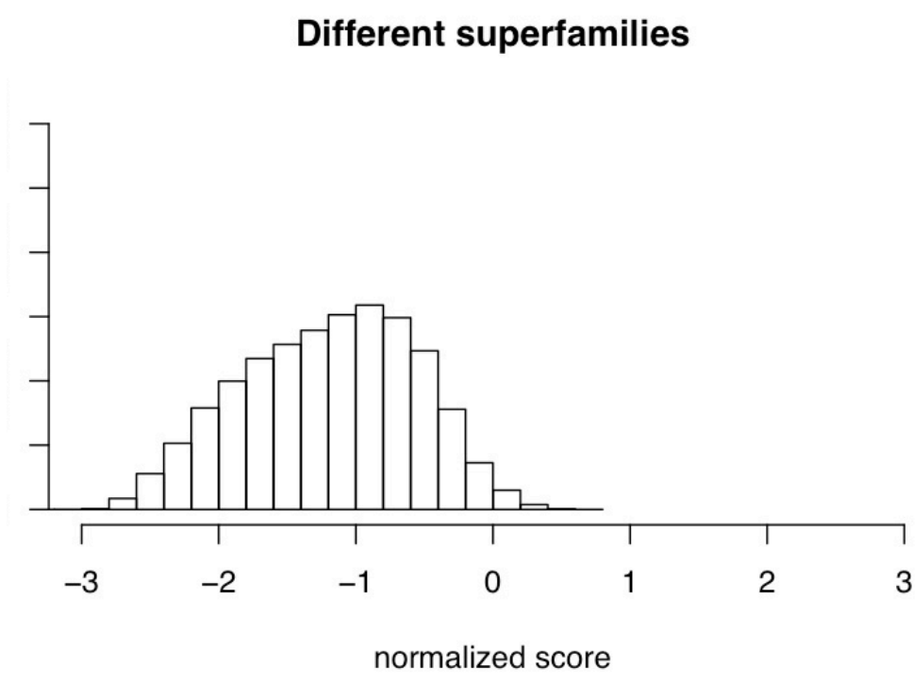
UNIVERSITÉ DE NANTES

7259 x 7259 pairwise structure comparisons using PB alignment



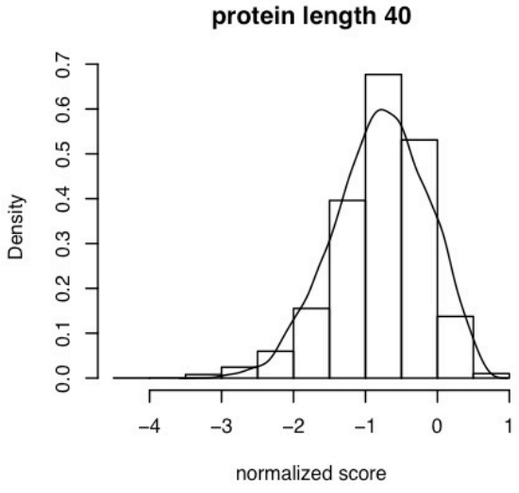


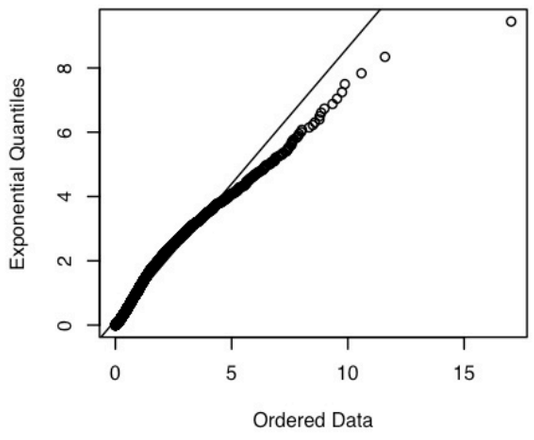




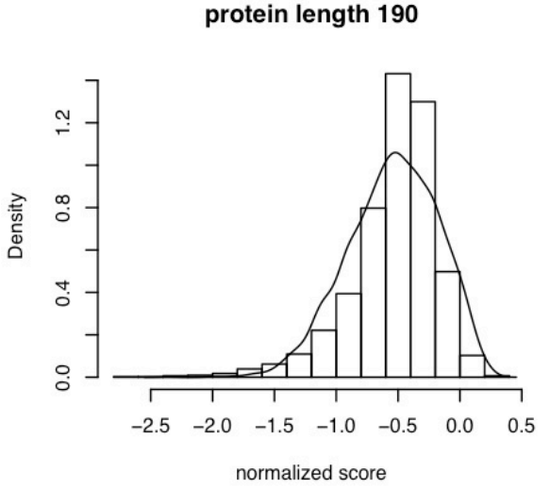
Mining protein structures using PBs

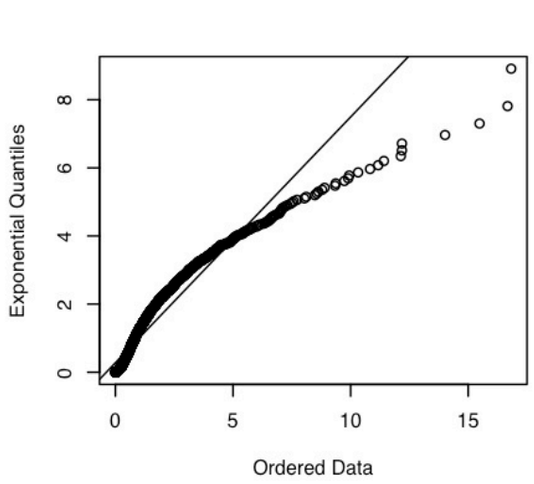
UNIVERSITÉ DE NANTES



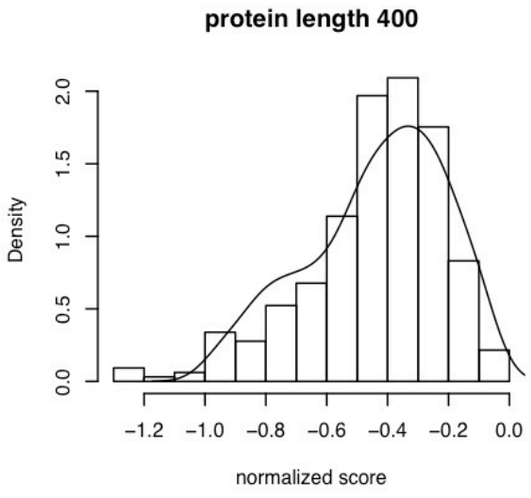


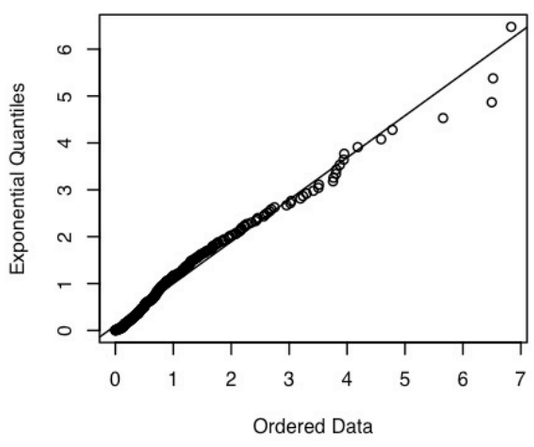
Assessing robustness of alignment scores (query against unrelated random sequences)





Testing for EVD distribution

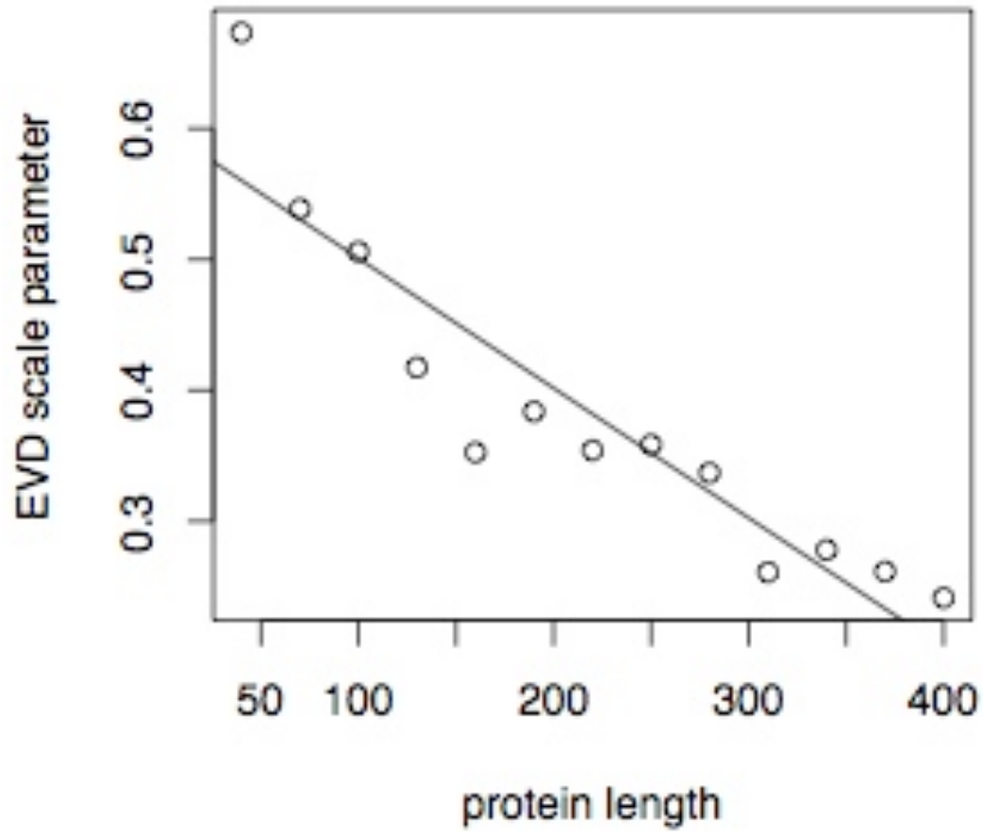




Mining protein structures using PBs

UNIVERSITÉ DE NANTES

Assessing robustness of alignment scores (query against unrelated random sequences)



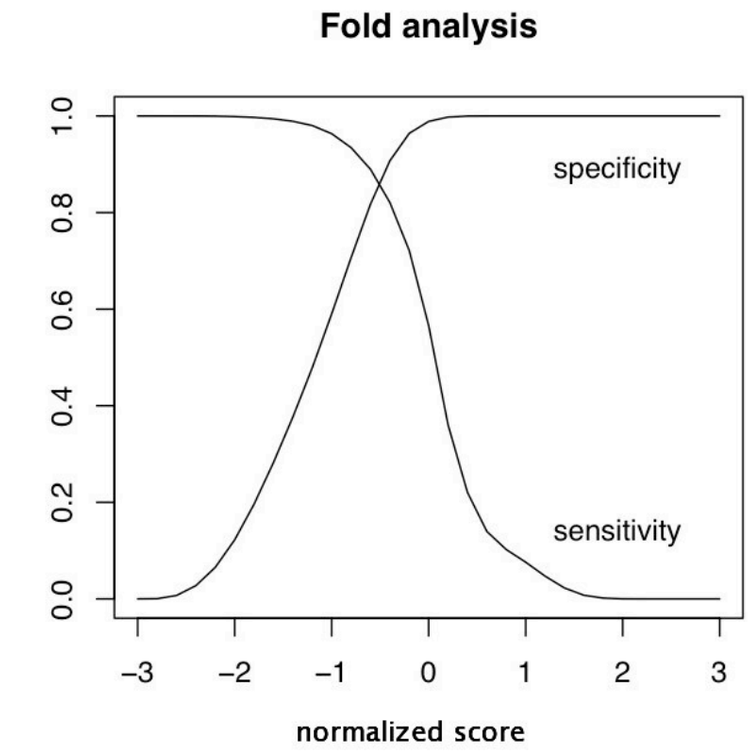
➡ Follows EVD distribution

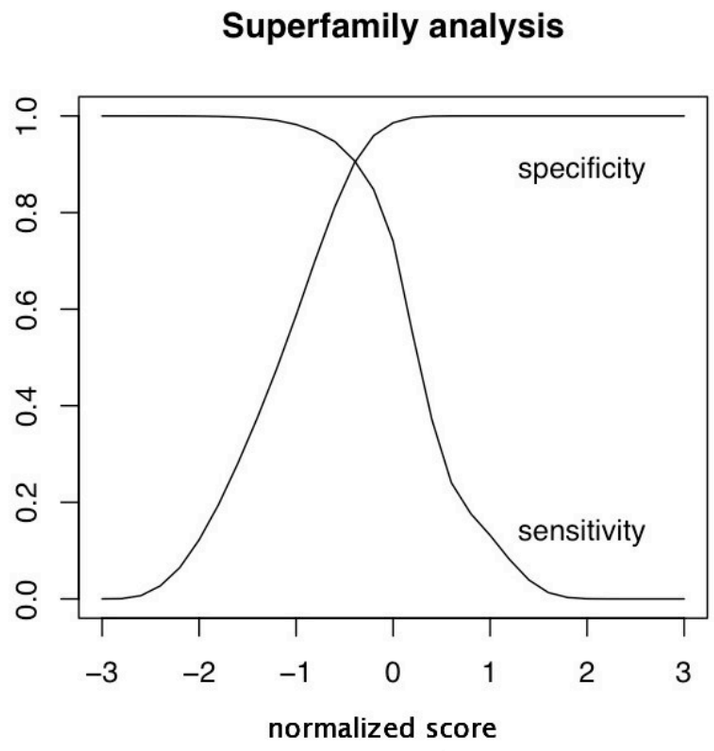
Tyagi M., de Brevern A., Srinivasan N., Offmann B. Protein structure mining using a structural alphabet. *Proteins*, (2007).

Mining protein structures using PBs

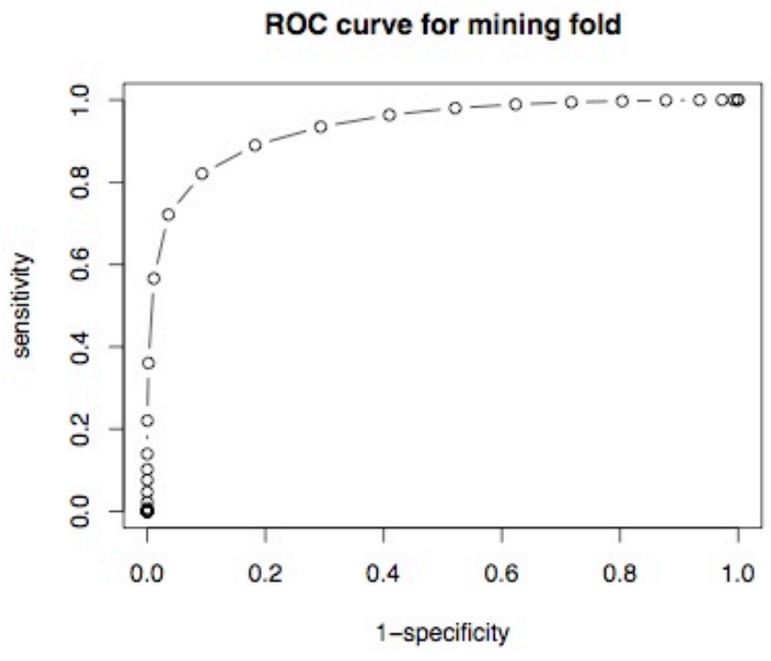
UNIVERSITÉ DE NANTES

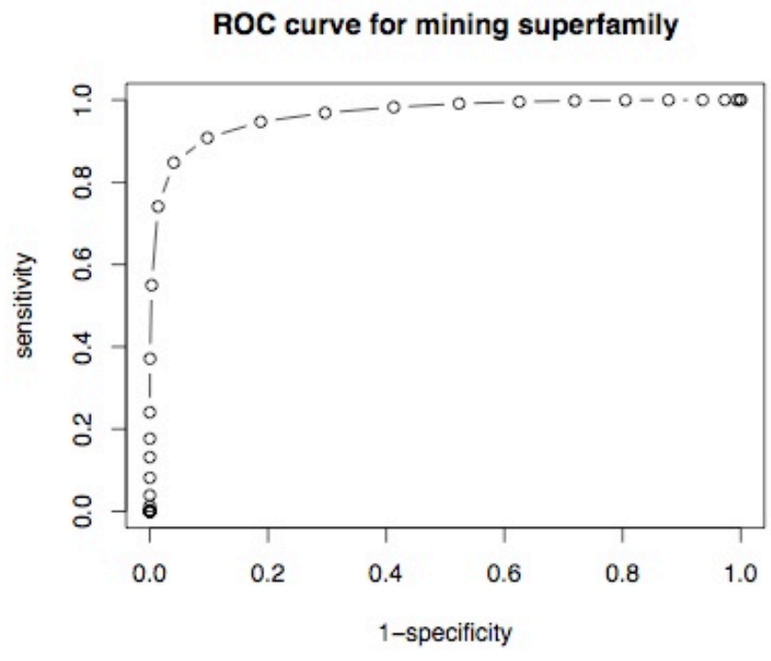
Defining a score threshold for decision





Optimal normalized score cut-off : -0.25





Mining protein structures using PBs

UNIVERSITÉ DE NANTES

Efficiency rate of mining proteins at various SCOP classification levels

	Only query domain is removed from databas e			Whole family related to query is removed from database		
SCOP level	Top10	Top5	Top1	Top10	Top5	Top1
Class	99.1	96.8	93.1	92.5	88	76.1
Fold	87.4	85.6	81.3	62.6	57.5	47.4
Super Family	84.3	82.8	79.0	65.1	60.8	53.0
Family	80.0	78.7	75.1	n/a	n/a	n/a

Results based on 7259 queries (from SCOP95 version 1.65)

Tyagi M., de Brevern A., Srinivasan N., Offmann B. Protein structure mining using a structural alphabet. *Proteins*, in press (2007).

Mining protein structures using PBs

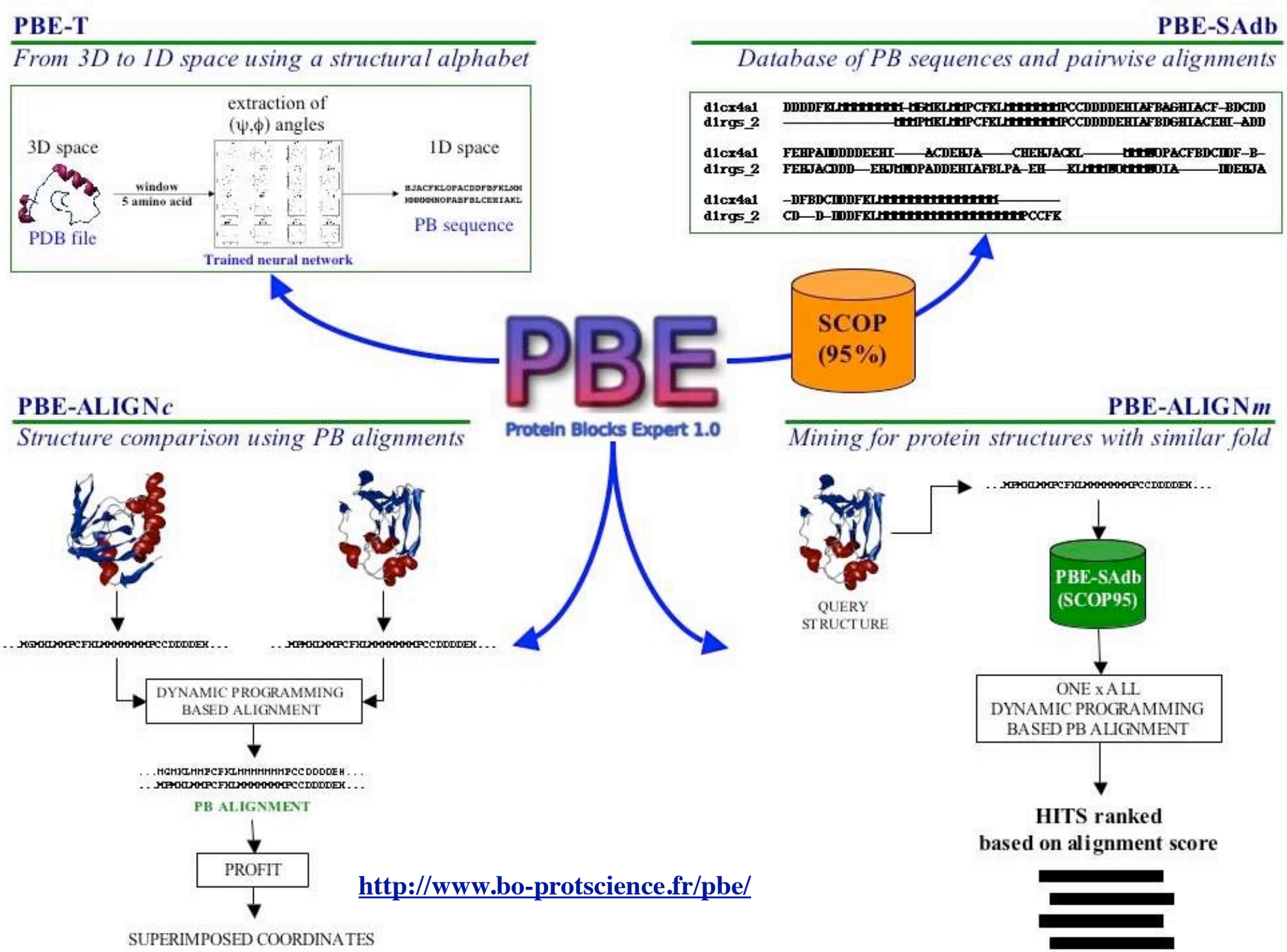
UNIVERSITÉ DE NANTES

Comparing PB-ALIGN with other methods (61 non-trivial cases from Carpentier et al, 2005)

Program	Mainly α (19)	Mainly β (19)	Mixed αβ (15)	Few SSEs (8)	Total (%)
PB-ALIGN	18^{+}	17^{*}	14	8	96.6
YAKUSA	17	19	14	8	95
CE	17	19	13	8	93
DALI	14	19	14	8	90
MATRAS	11	19	14	8	85
VAST	12	17	15	7	84
TOP	14	18	12	7	84
DEJAVU	14	19	9	4	75
TOPSCAN	15	12	9	7	70
TOPS	2	15	14	7	62
PRIDE	14	14	7	3	62
LOCK	0	14	11	8	54
SSM	5	13	10	5	54

A dedicated server for Protein Blocks

UNIVERSITÉ DE NANTES



http://www.bo-protscience.fr/pbe/

M. Tyagi, P. Sharma, C.S. Swamy, F. Cadet, N. Srinivasan, A.G. de Brevern and B. Offmann.
Protein Block Expert (PBE): A web-based protein structure analysis server using a structural alphabet.
Nucleic Acids Res., 34:W119-W123, (2006).