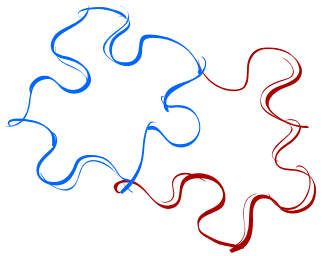
http://sarst.life.nthu.edu.tw/iSARST

SARST – Structural similarity search Aided by Ramachandran Sequential Transformation

Lo WC, Chang CH, Huang PJ, Lyu PC. ***Protein structural similarity search by Ramachandran codes***. BMC Bioinformatics 2007, 8:307.

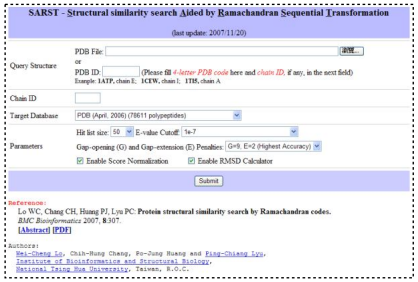
呂平江

清華大學 生命科學系



20100831

Progress High light



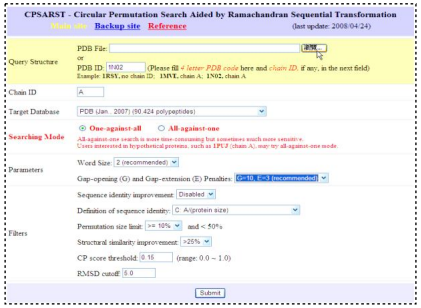
✓ SARST: Structural similarity search Aided by Ramachandran Sequential Transformation.

W. C. Lo, C. H. Chang, P. J. Huang, P. C. Lyu*

BMC Bioinformatics,

2007, 8:307

http://sarst.life.nthu.edu.tw/sarst/



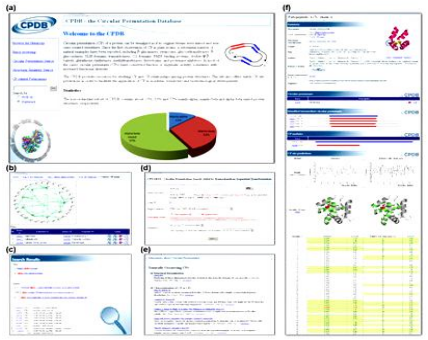
✓ CPSARST – An efficient circular permutation search tool applied to the detection of novel protein structural relationships.

W. C. Lo, and P. C. Lyu*.

Genome Biology 9, R11

(2008).

http://sarst.life.nthu.edu.tw/cpsarst/



✓ CPDB: a database of circular permutation in proteins.

W. C. Lo, C. C. Lee, C. Y. Lee, P. C. Lyu*

Nucleic Acids Research, doi:10.1093/nar/gkn679

(Database Issue)

(2009).

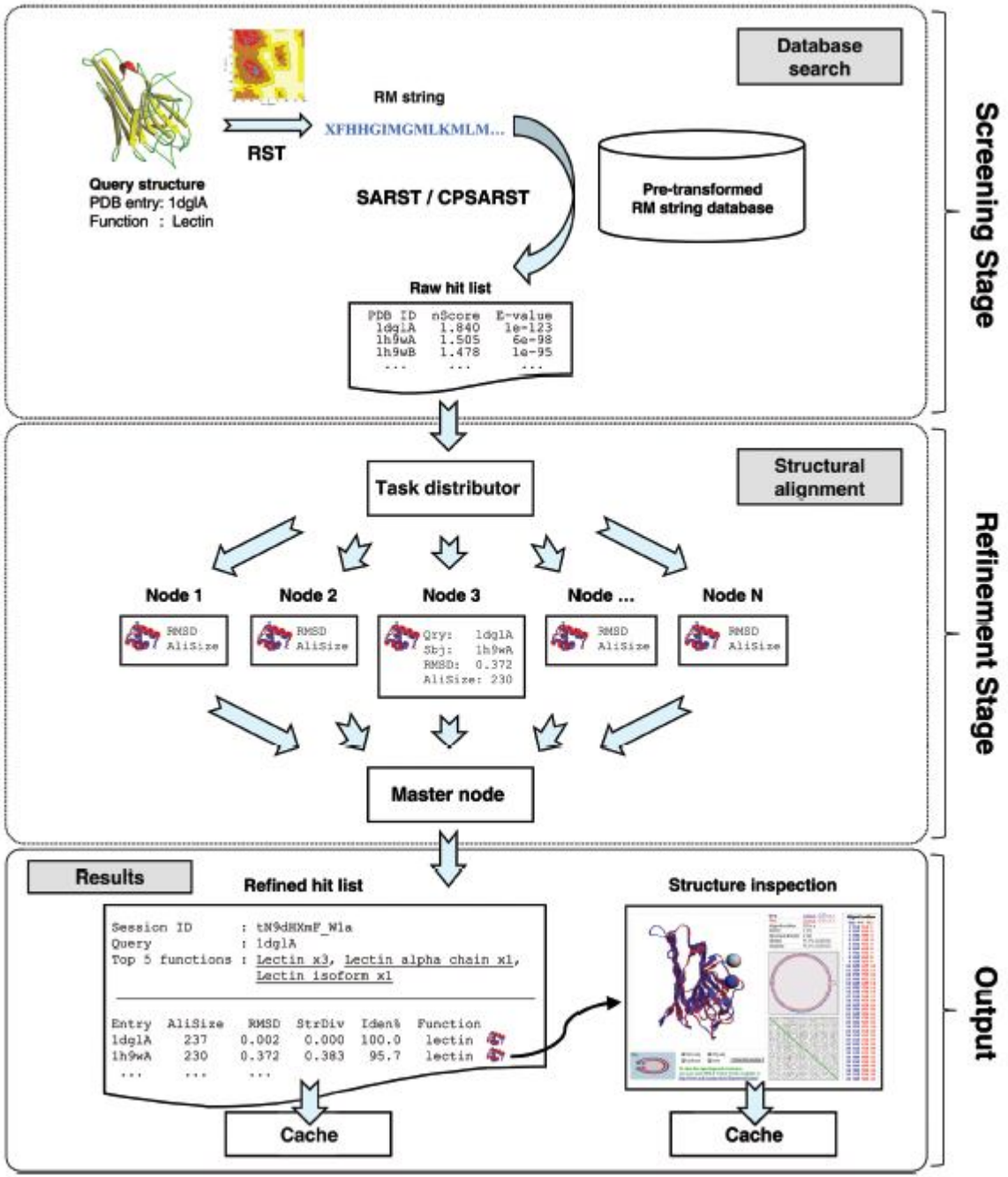
http://sarst.life.nthu.edu.tw/cpdb/

iSARST: an integrated SARST web server for rapid protein structural similarity searches

W. C. Lo, C. Y. Lee, C. C. Lee, P. C. Lyu*

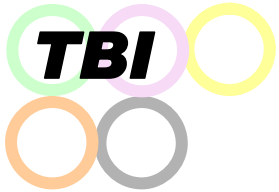
Nucleic Acids Research, 37(Web Server issue):W545-51 (2009)

http://sarst.life.nthu.edu.tw/isarst/



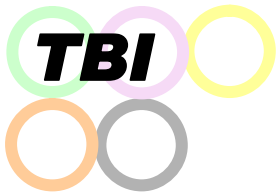
Introduction to SARST

- SARST transforms 3D protein structures into 1D text sequences and recruit blast to perform protein structural alignment searches
- Features
  - high speed
  - reasonable compromise of accuracy
  - giving statistically meaningful results



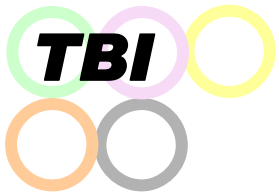
Structural Comparisons – Why?

- Protein is the functional unit of biological systems.
- The function of a protein is basically determined by its structure.
- Proteins sharing similar structures usually have similar functions.



Structural Comparisons – How?

- Two categories of current methods
  - By amino acid sequence alignments.
  - By 3D structural alignments.

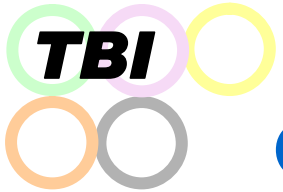


Classical Sequence Alignment Methods

- BLAST
  - Basic Local Alignment Search Tool
- FASTA
  - FAST-All, reflecting that it can be used for fast protein comparisons

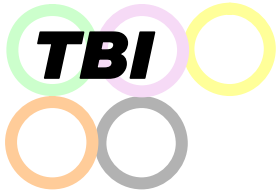
Performance: Rapid but inaccurate*

* Kolodny *et al.* (2005) *J Mol Biol.* 346:1173-1188



Conventional Structural Alignment Methods

- Double Dynamic Programming – SSAP
- Distance Alignment Tools – DALI
- Combinatorial Extension – CE
- Vector Alignment Search Tool – VAST
- Fast Alignment Search Tool – FAST
- MAtching Molecular Models Obtained from Theory – MAMMOTH



The Basic Algorithm of Structural Alignments

- Based on distances or relations among vectors of backbone atoms
- Try to match as many residues and achieve as small RMSDs as possible
- RMSD
  - Root Mean Square Distance

Performance: Accurate but slow

CE takes 2.5 days to search one protein against PDB

Speed vs. Accuracy: Incompatible?

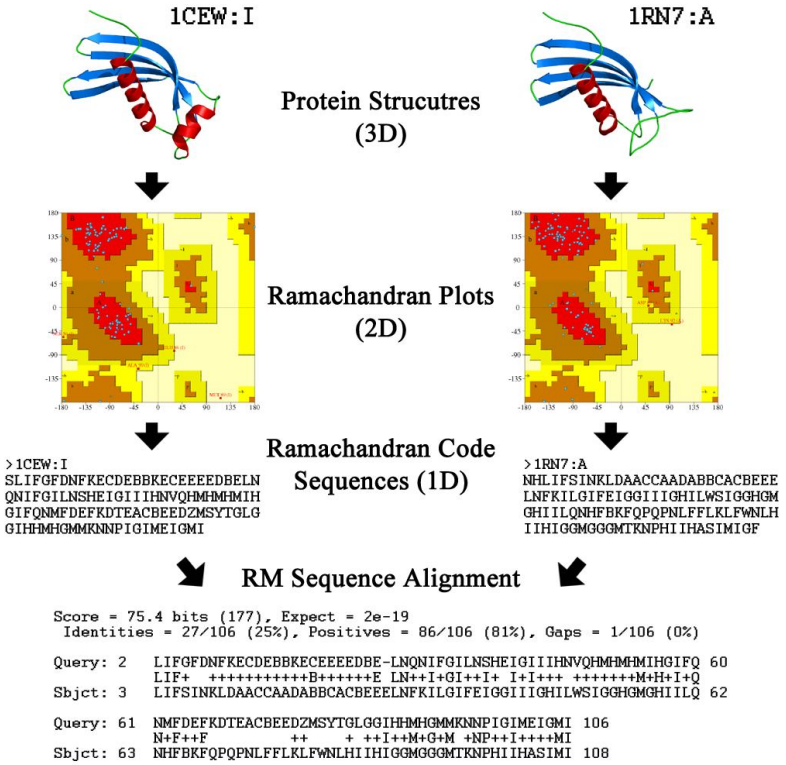
- Possible solution: the linear encoding method

  3D structure ➔ 1D text sequence

- Example:

  SARST

  – Structure Alignment by Ramachandran Search Tool

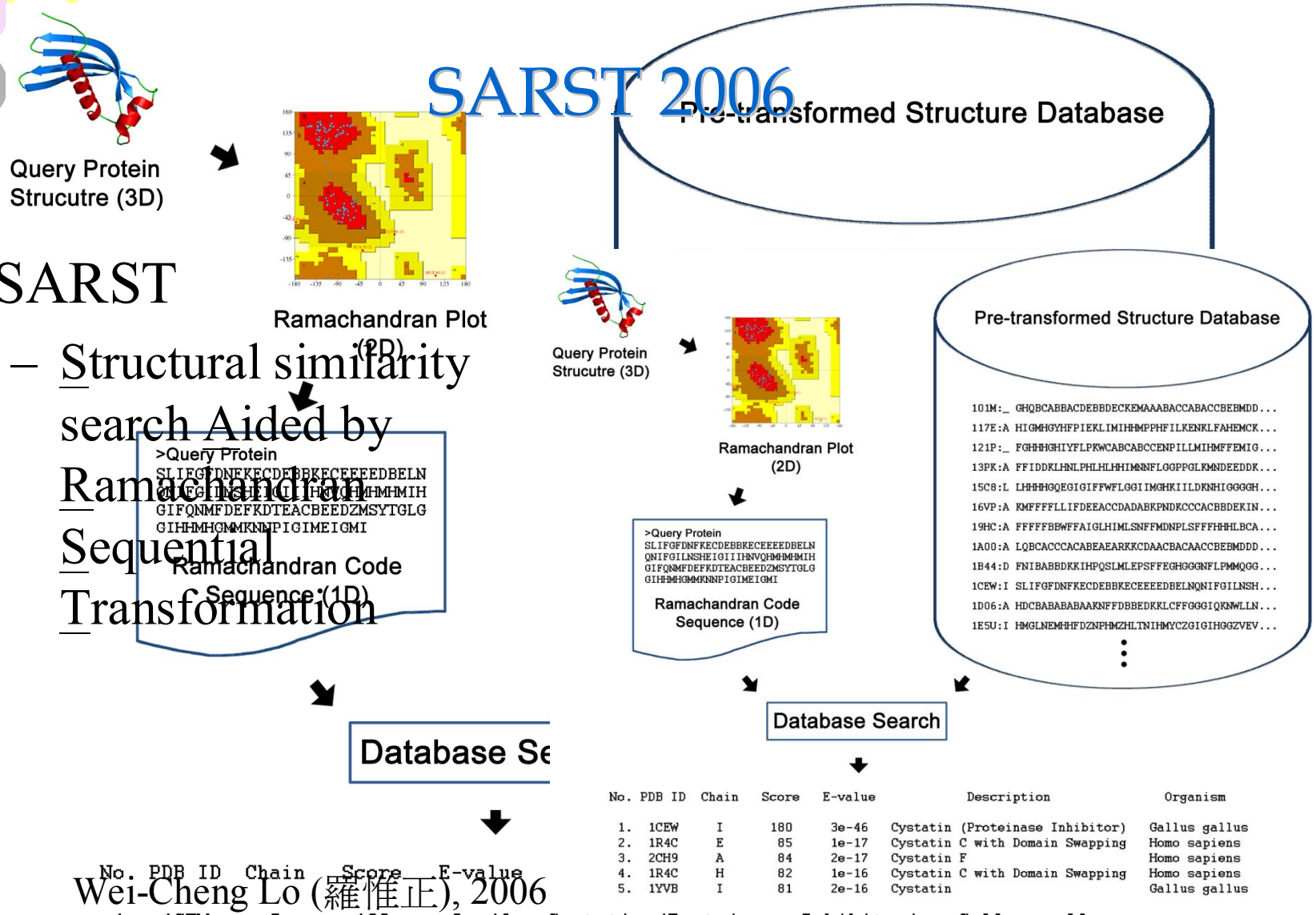


Po-Jung Huang (黃柏榕), 2002

Chih-Hung Chang (張志宏), 2004

SARST 2006

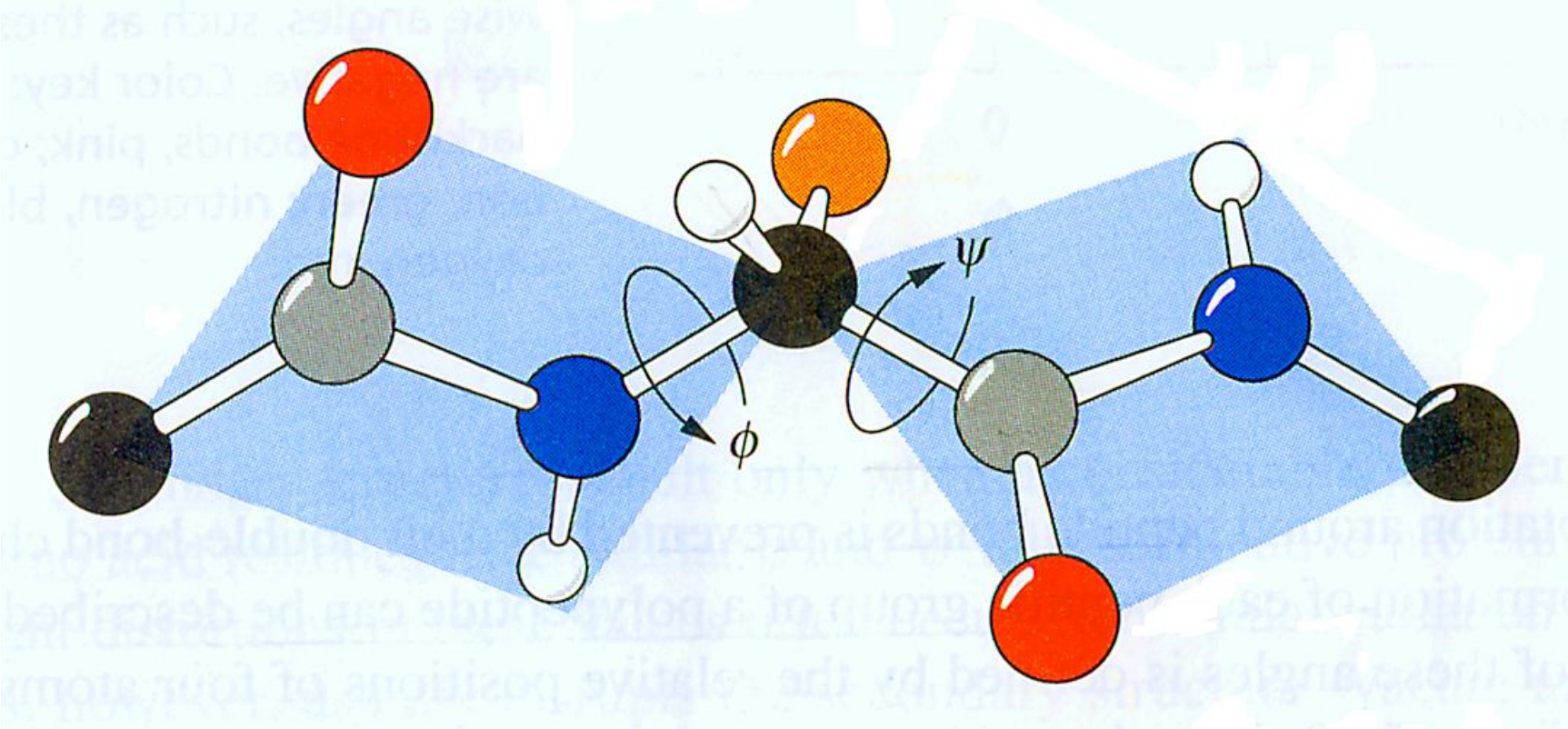
- SARST
  - Structural similarity search Aided by Ramachandran Sequential Transformation



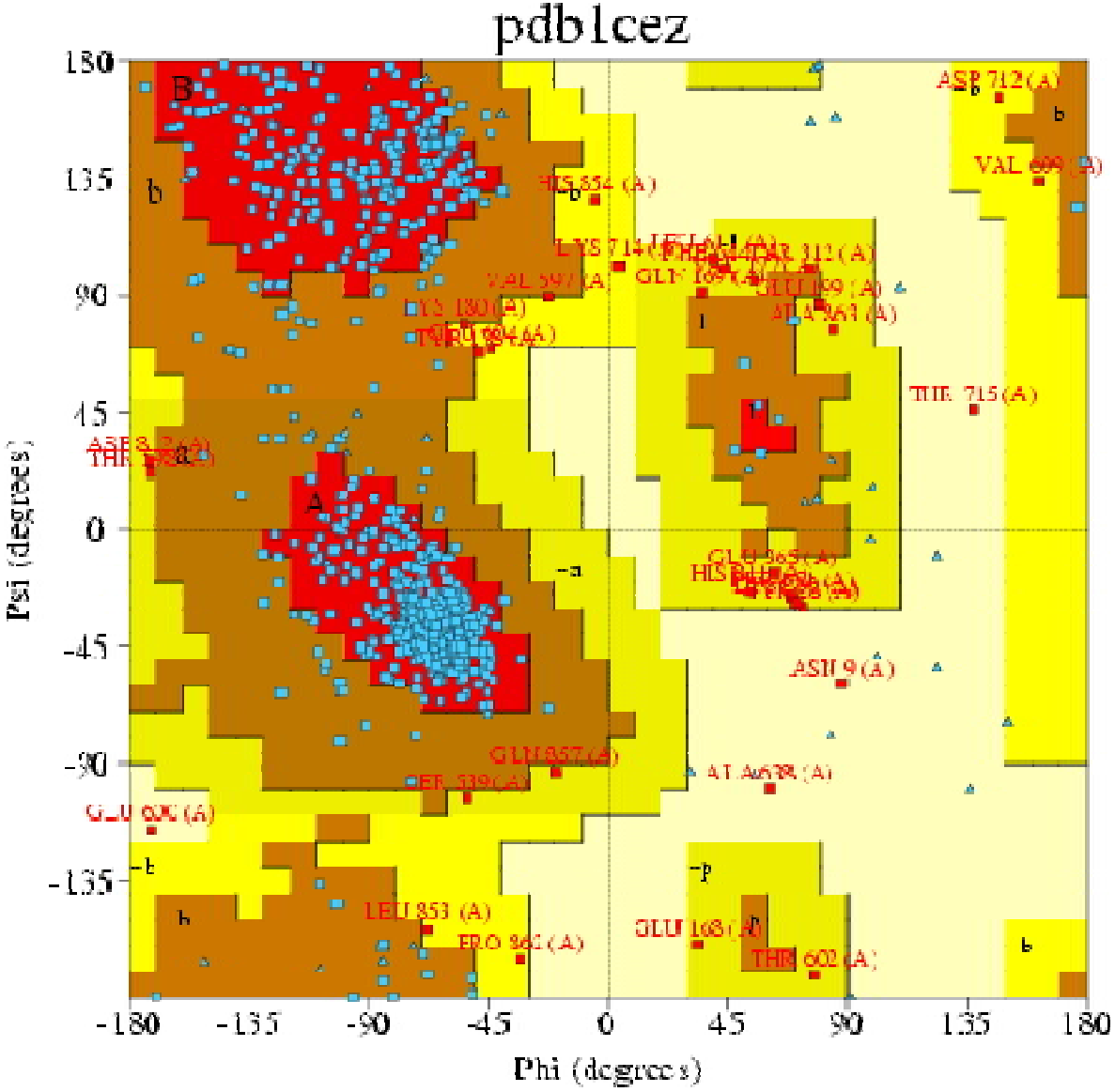
No. PDB ID Chain Score E-value Description Organism
1. 1CEW I 180 3e-46 Cystatin (Proteinase Inhibitor) Gallus gallus
2. 1R4C E 85 1e-17 Cystatin C with Domain Swapping Homo sapiens
3. 2CH9 A 84 2e-17 Cystatin F Homo sapiens
4. 1R4C H 82 1e-16 Cystatin C with Domain Swapping Homo sapiens
5. 1YVB I 81 2e-16 Cystatin Gallus gallus

Wei-Cheng Lo (羅惟正), 2006

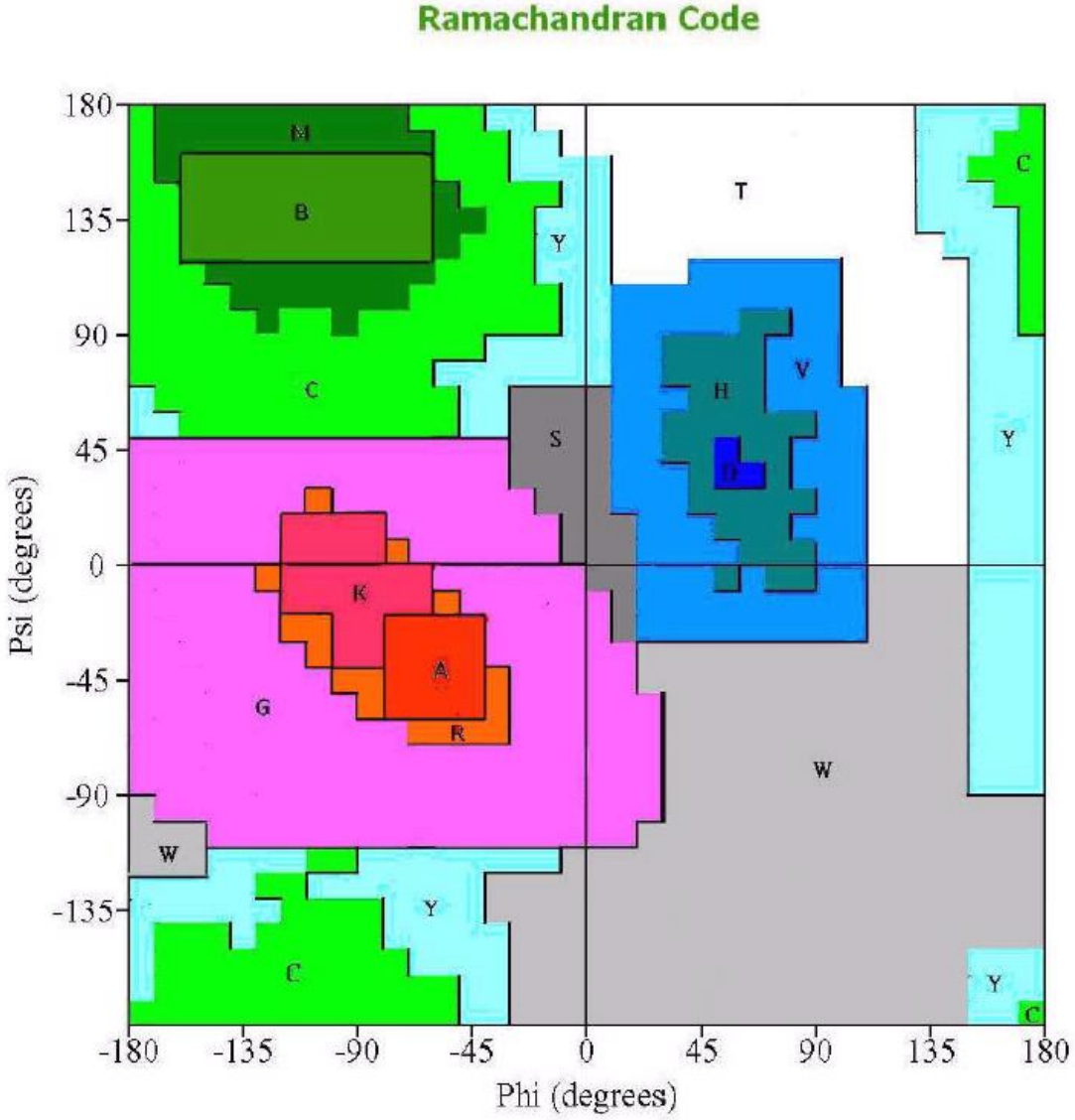
Phi (φ) and Psi (ϕ) Angles



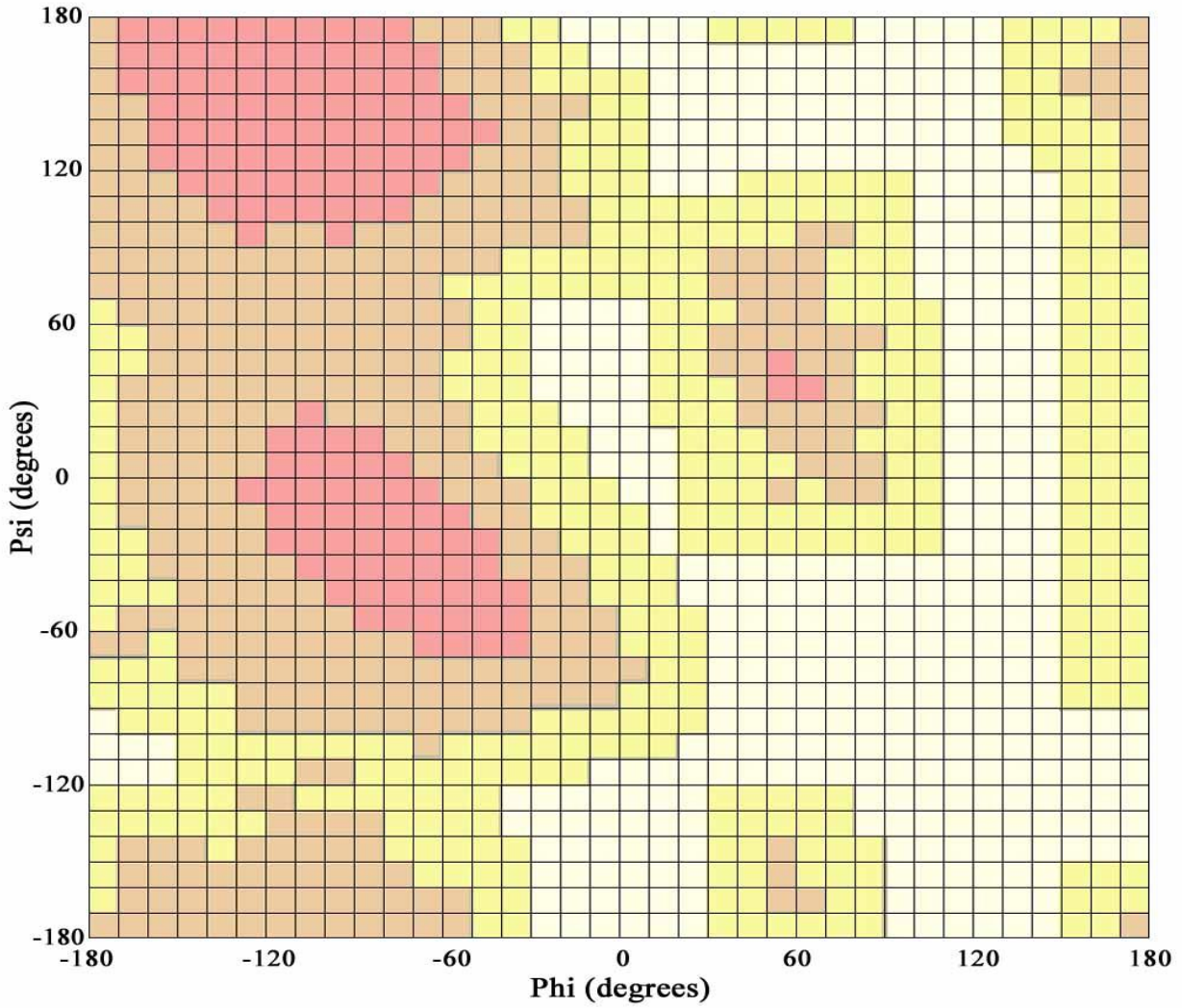
The Ramachandran Map

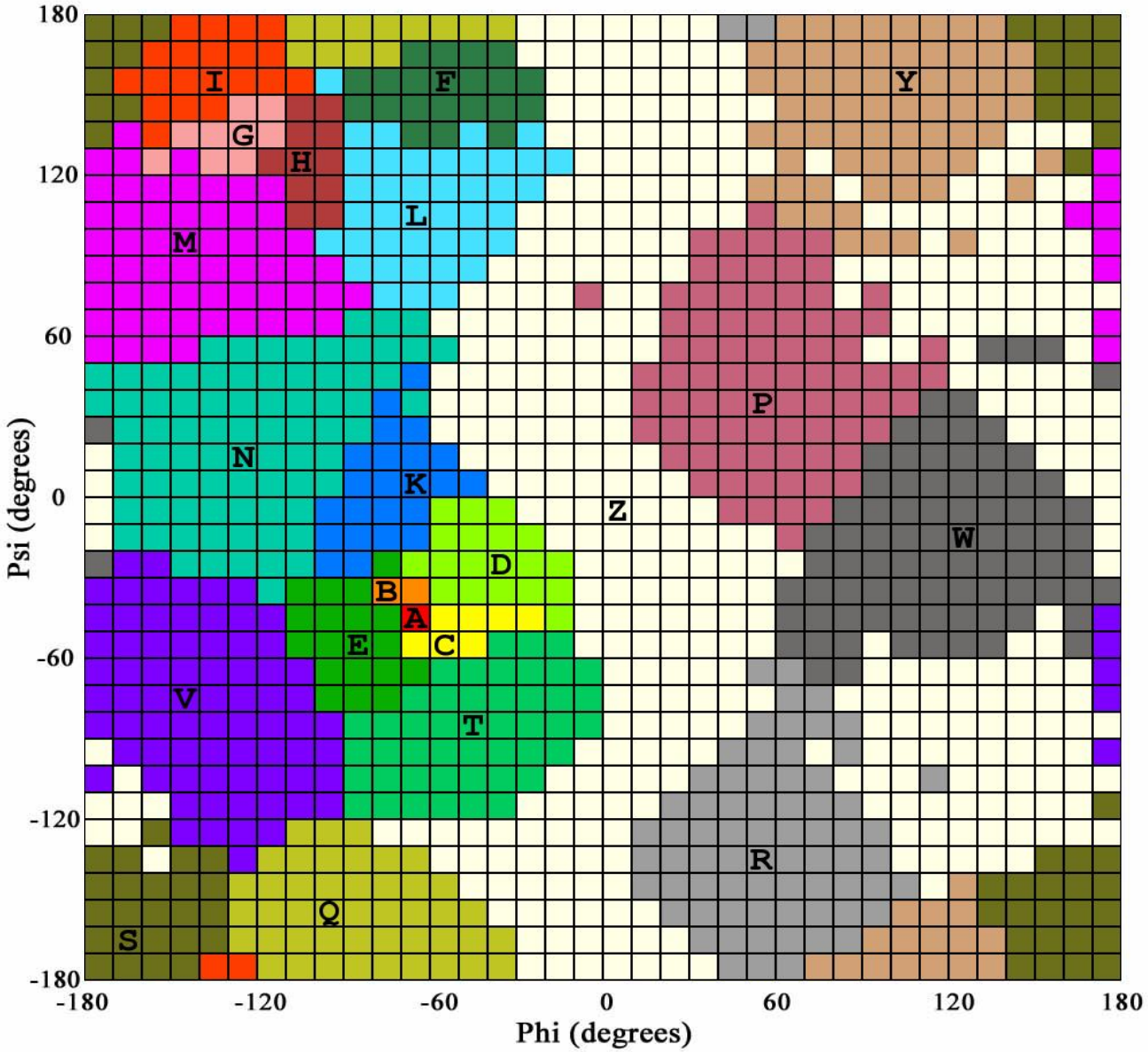


An Example of Organized Ramachandran Map



Dissection of the Ramachandran plot





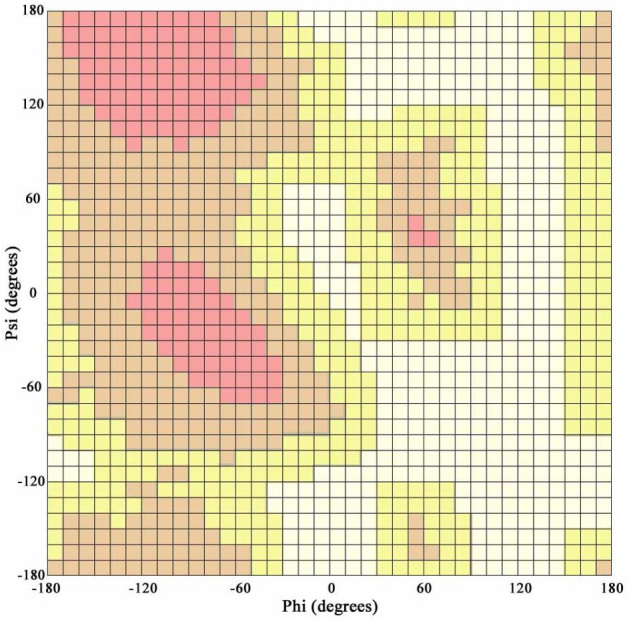
Ramachandran (RM) Sequential Transformation

RM Seq: I L L P C

- Algorithm: Nearest-neighbor clustering
- 1,296 cells were clustered into 22 groups
- Each group was assigned with an English letter, that is, a Ramachandran code

W.C. Lo, et al. **Protein structural similarity search by Ramachandran codes.** *BMC Bioinformatics* 2007, 8:307.

Determine the Distance Among Cells



$$RSAD = \sqrt{(\Delta\varphi)^2 + (\Delta\psi)^2}$$

How to Evaluate Similarities?

AAAAWWW
AAAAAWW

WWWWAAA
WWWWWAA

Are they equally similar?
Score A:A = ?
Score W:W = ?
Score A:W and W:A = ?

The Scoring Matrix of SARST

	A	B	C	D	E	T	K	V	N	F	G	H	I	L	M	Q	S	Y	R	P	W	Z	X
A	3	2	2	1	1	0	−2	−3	−3	−8	−11	−11	−13	−8	−8	−9	−14	−9	−7	−8	−7	−4	0
B	2	2	2	1	1	1	0	−1	−2	−6	−12	−10	−10	−7	−7	−6	−10	−8	−5	−6	−4	−6	0
C	2	2	2	1	1	3	−1	−2	−3	−6	−13	−11	−9	−7	−8	−7	−9	−10	−2	−7	−5	−3	0
D	1	1	1	3	1	2	2	−1	−1	−4	−9	−7	−8	−4	−6	−5	−7	−4	1	−3	−4	−2	0
E	1	1	1	1	3	1	2	3	1	−5	−7	−6	−7	−4	−4	−4	−7	−2	−1	−5	−3	−1	0
T	0	1	3	2	1	5	−1	2	−1	−2	−6	−6	−4	−4	−5	−2	−4	−4	2	−1	−1	3	0
K	−2	0	−1	2	2	−1	4	1	3	−3	−6	−6	−5	−3	−3	−2	−5	−2	−2	0	0	−1	0
V	−3	−1	−2	−1	3	2	1	9	3	−3	−4	−4	−2	−2	−2	0	0	3	2	−1	3	4	0
N	−3	−2	−3	−1	1	−1	3	3	5	−2	−4	−4	−3	−2	0	−2	−3	−2	−1	1	1	1	0
F	−8	−6	−6	−4	−5	−2	−3	−3	−2	5	−1	1	0	3	0	3	0	2	0	−2	−2	1	0
G	−11	−12	−13	−9	−7	−6	−6	−4	−4	−1	4	3	3	0	2	0	1	−3	−5	−5	−6	−2	0
H	−11	−10	−11	−7	−6	−6	−6	−4	-4	1	3	4	1	2	2	0	−1	−2	−4	−3	−5	−1	0
I	−13	−10	−9	−8	−7	−4	−5	−2	−3	0	3	1	4	0	1	2	4	0	−1	−4	−7	−2	0
L	−8	−7	−7	−4	−4	−4	−3	−2	−2	3	0	2	0	4	1	1	−1	0	0	−1	−2	1	0
M	−8	−7	−8	−6	−4	−5	−3	−2	0	0	2	2	1	1	4	0	1	−1	−4	−2	−2	1	0
Q	−9	−6	−7	−5	−4	−2	−2	0	−2	3	0	0	2	1	0	6	1	3	1	−3	−3	1	0
S	−14	−10	−9	−7	−7	−4	−5	0	−3	0	1	−1	4	−1	1	1	7	5	2	−3	−3	3	0
Y	−9	−8	−10	−4	−2	−4	−2	3	−2	2	−3	−2	0	0	−1	3	5	10	7	2	2	7	0
R	−7	−5	−2	1	−1	2	−2	2	−1	0	−5	−4	−1	0	−4	1	2	7	11	3	0	7	0
P	−8	−6	−7	−3	−5	−1	0	−1	1	−2	−5	−3	−4	−1	−2	−3	−3	2	3	8	7	4	0
W	−7	−4	−5	−4	−3	−1	0	3	1	−2	−6	−5	−7	−2	−2	−3	−3	2	0	7	9	5	0
Z	−4	−6	−3	−2	−1	3	−1	4	1	1	−2	−1	−2	1	1	1	3	7	7	4	5	6	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Building the Scoring Matrix (SM)

- A "regenerative approach" was developed to build SM for SARST based on the BLOSUM algorithm*:

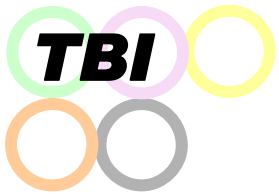
$$Score_{ij} = f_s \times \log_2(q_{ij} / e_{ij})$$

$$\frac{q_{ij}}{e_{ij}} = \frac{95.5\%\%}{1.3\%\times 1.3\%}$$

* Henikoff and Henikoff. (1992) *Proc Natl Acad Sci USA.* **89:**10915-10919

The Scoring Matrix of SARST

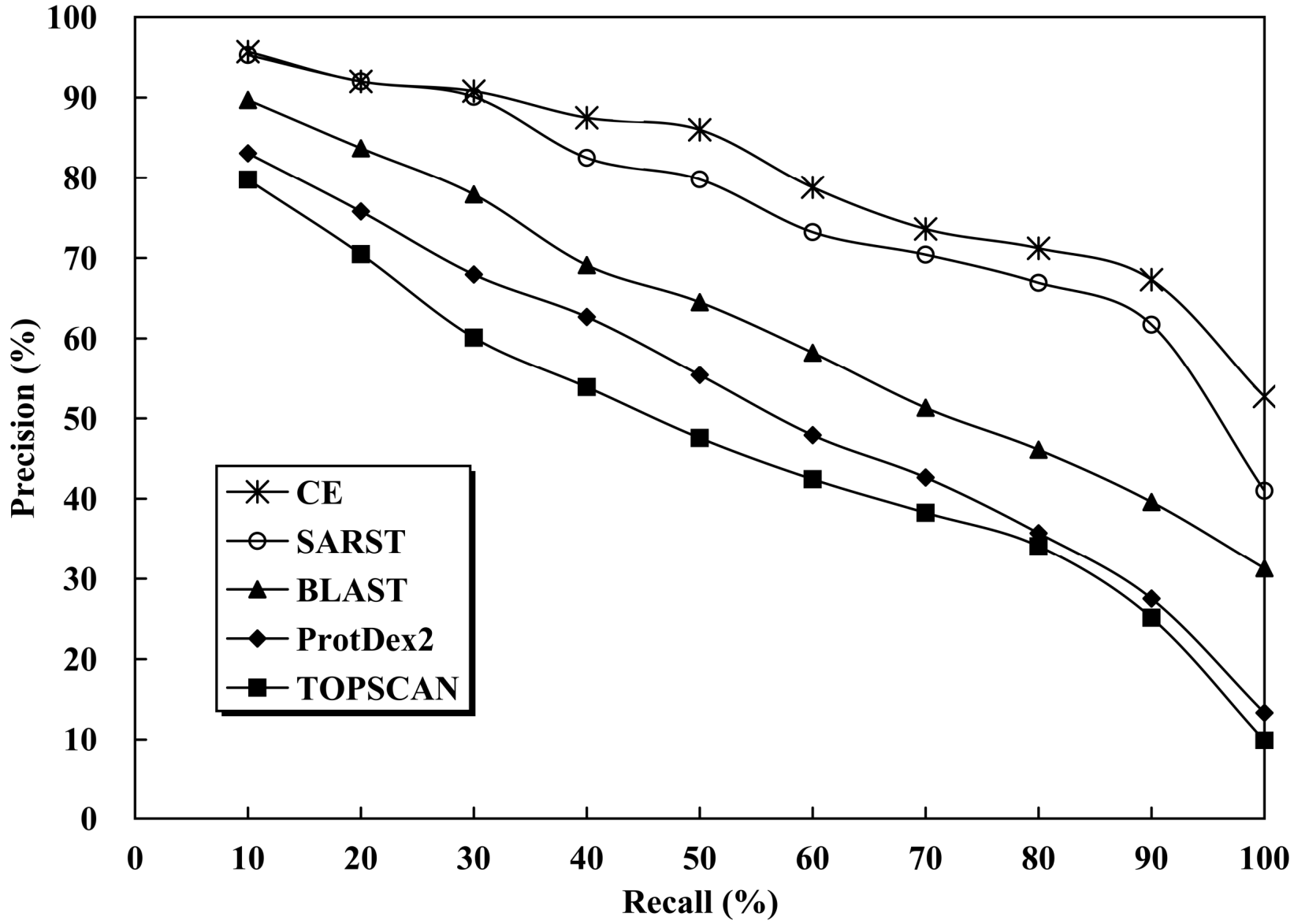
	A	B	C	D	E	T	K	V	N	F	G	H	I	L	M	Q	S	Y	R	P	W	Z	X
A	3	2	2	1	1	0	−2	−3	−3	−8	−11	−11	−13	−8	−8	−9	−14	−9	−7	−8	−7	−4	0
B	2	2	2	1	1	1	0	−1	−2	−6	−12	−10	−10	−7	−7	−6	−10	−8	−5	−6	−4	−6	0
C	2	2	2	1	1	3	−1	−2	−3	−6	−13	−11	−9	−7	−8	−7	−9	−10	−2	−7	−5	−3	0
D	1	1	1	3	1	2	2	−1	−1	−4	−9	−7	−8	−4	−6	−5	−7	−4	1	−3	−4	−2	0
E	1	1	1	1	3	1	2	3	1	−5	−7	−6	−7	−4	−4	−4	−7	−2	−1	−5	−3	−1	0
T	0	1	3	2	1	5	−1	2	−1	−2	−6	−6	−4	−4	−5	−2	−4	−4	2	−1	−1	3	0
K	−2	0	−1	2	2	−1	4	1	3	−3	−6	−6	−5	−3	−3	−2	−5	−2	−2	0	0	−1	0
V	−3	−1	−2	−1	3	2	1	9	3	−3	−4	−4	−2	−2	−2	0	0	3	2	−1	3	4	0
N	−3	−2	−3	−1	1	−1	3	3	5	−2	−4	−4	−3	−2	0	−2	−3	−2	−1	1	1	1	0
F	−8	−6	−6	−4	−5	−2	−3	−3	−2	5	−1	1	0	3	0	3	0	2	0	−2	−2	1	0
G	−11	−12	−13	−9	−7	−6	−6	−4	−4	−1	4	3	3	0	2	0	1	−3	−5	−5	−6	−2	0
H	−11	−10	−11	−7	−6	−6	−6	−4	-4	1	3	4	1	2	2	0	−1	−2	−4	−3	−5	−1	0
I	−13	−10	−9	−8	−7	−4	−5	−2	−3	0	3	1	4	0	1	2	4	0	−1	−4	−7	−2	0
L	−8	−7	−7	−4	−4	−4	−3	−2	−2	3	0	2	0	4	1	1	−1	0	0	−1	−2	1	0
M	−8	−7	−8	−6	−4	−5	−3	−2	0	0	2	2	1	1	4	0	1	−1	−4	−2	−2	1	0
Q	−9	−6	−7	−5	−4	−2	−2	0	−2	3	0	0	2	1	0	6	1	3	1	−3	−3	1	0
S	−14	−10	−9	−7	−7	−4	−5	0	−3	0	1	−1	4	−1	1	1	7	5	2	−3	−3	3	0
Y	−9	−8	−10	−4	−2	−4	−2	3	−2	2	−3	−2	0	0	−1	3	5	10	7	2	2	7	0
R	−7	−5	−2	1	−1	2	−2	2	−1	0	−5	−4	−1	0	−4	1	2	7	11	3	0	7	0
P	−8	−6	−7	−3	−5	−1	0	−1	1	−2	−5	−3	−4	−1	−2	−3	−3	2	3	8	7	4	0
W	−7	−4	−5	−4	−3	−1	0	3	1	−2	−6	−5	−7	−2	−2	−3	−3	2	0	7	9	5	0
Z	−4	−6	−3	−2	−1	3	−1	4	1	1	−2	−1	−2	1	1	1	3	7	7	4	5	6	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

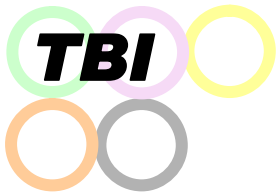


Speed Evaluation

Method	Average time per query (sec)	Relative to SARST
CE	82,789.20	243,497.65
TOPSCAN	85.08	250.24
ProtDex2	0.76	2.24
BLAST	0.30	0.88
SARST	0.34	1.00
SARST (2 CPUs)	0.16	0.47

Accuracy Evaluation





Information Retrieval Techniques

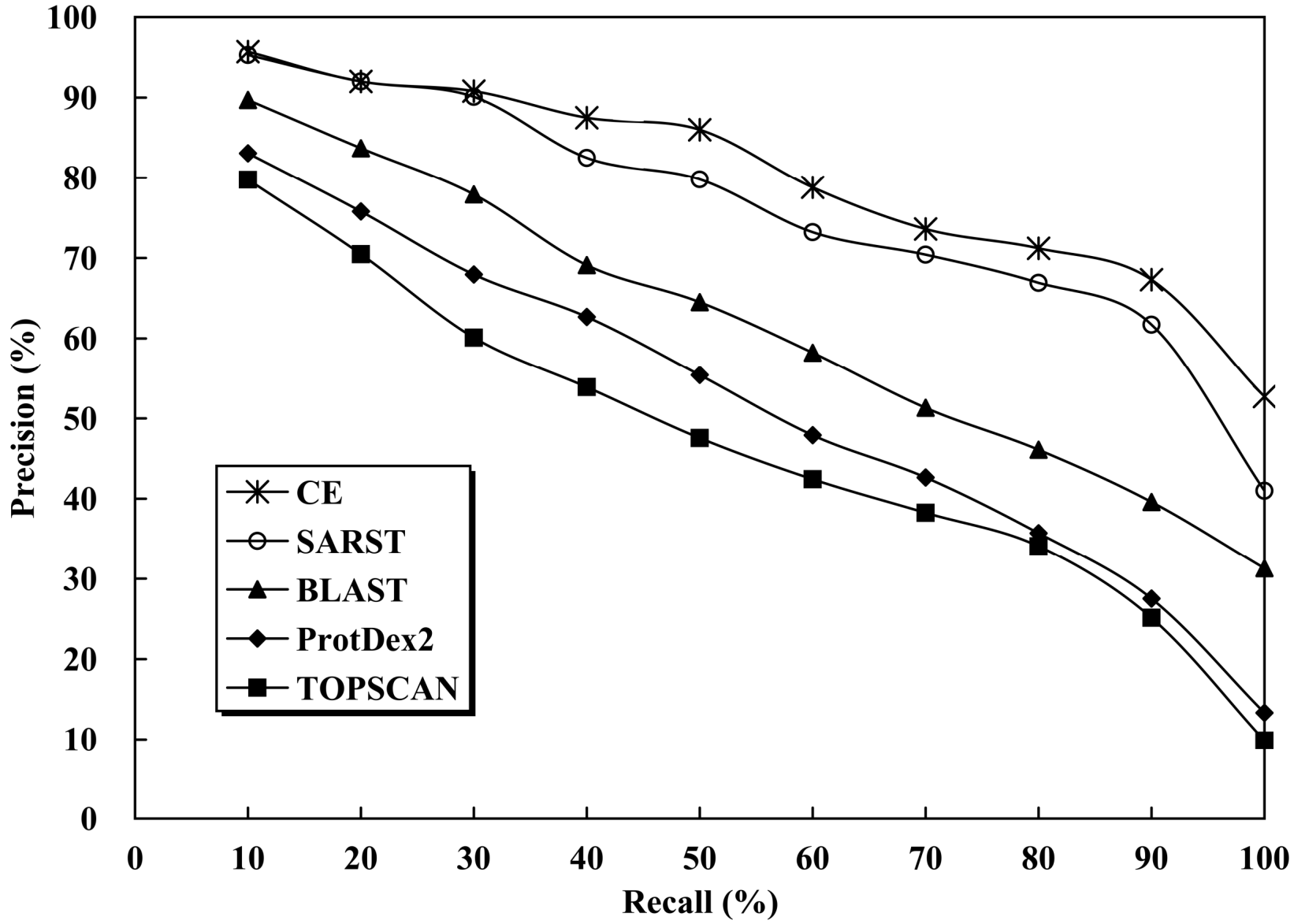
- Recall

  ➔ the ability to extract answers

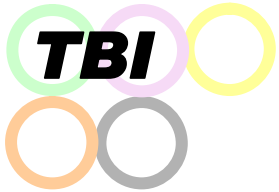
- Precision

  ➔ the ability to give correct answers

Accuracy Evaluation



Next…



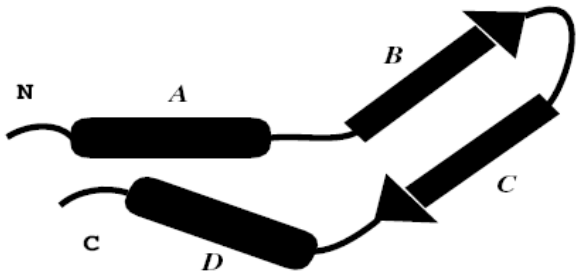
http://sarst.life.nthu.edu.tw/iSARST

CPSARST - Circular Permutation Search Aided by Ramachandran Sequential Transformation

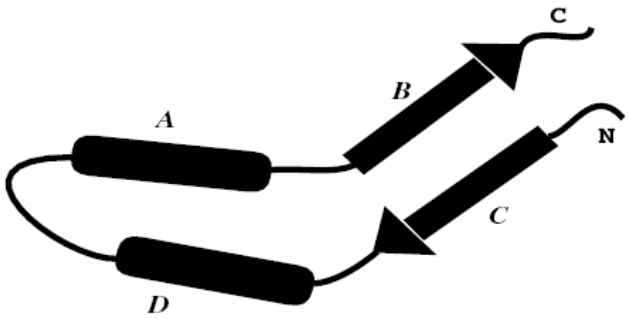
Lo WC, Lyu PC: ***CPSARST: an efficient circular permutation search tool applied to the detection of novel protein structural relationships.*** Genome Biology 2008,9:R11.

Circular Permutation (CP)

- Circular permutation of a protein can be visualized as if the original N- and C-termini were linked and new ones created elsewhere[1].
- In most of the cases, naturally occurring CPs have similar 3D structures and conserved biological functions[2].
- Efficient CP search tool is not available yet.

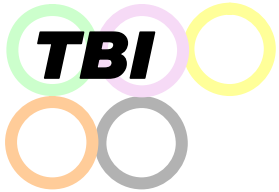


The sequence: ..A..B..C..D..



The sequence ..C..D..A..B..

1. Uliel S et al.: **A simple algorithm for detecting circular permutations in proteins.** *Bioinformatics* 1999,**15**:930-936.
2. Lindqvist Y, Schneider G: **Circular permutations of natural protein sequences: structural evidence.** *Curr Opin Struct Biol* 1997,**7**:422-427.

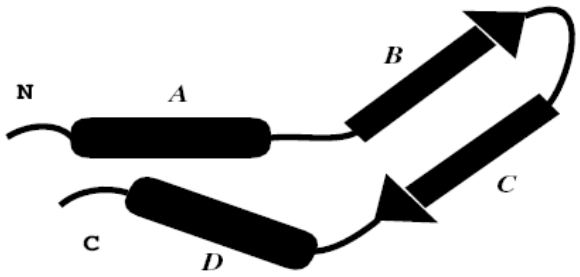


Natural Circular Permutants

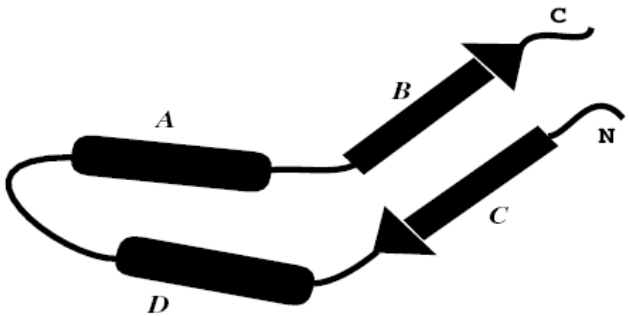
- Plant lectins
- Transaldolases
- DNA and other methyltransferases
- Ferredoxins
- Proteinase inhibitors
- Bacterial β-glucanases
- Swaposins
- Glucosyltransferases
- β-glucosidases
- SLH domains
- C2 domains
- FMN-binding proteins
- Double-φ β-barrels
- Glutathione synthetases

Circular Permutation (CP)

- Circular permutation of a protein can be visualized as if the original N- and C-termini were linked and new ones created elsewhere[1].
- In most of the cases, CPs have similar 3D structures and conserved biological functions[2].
- Efficient CP search tool is not available yet.



The sequence: ..A..B..C..D..



The sequence ..C..D..A..B..

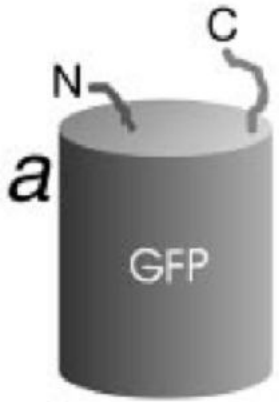
1. Uliel S et al.: **A simple algorithm for detecting circular permutations in proteins.** *Bioinformatics* 1999,**15**:930-936.
2. Lindqvist Y, Schneider G: **Circular permutations of natural protein sequences: structural evidence.** *Curr Opin Struct Biol* 1997,**7**:422-427.

Applications of Circular Permutation

- Folding researches.
- Determination of structurally and functionally important segments[1,2].
- Modification (enhancement) of the activity and/or stability[3-5].
- Creation of novel fusion proteins, the tethered sites of which are not confined to the native termini[5,6].

1. Anand.B. et al. Nucleic Acid Res 2006;34:2196-2205.
2. Gebhard.LG. et al. J Mol Biol 2006;358:280-288.
3. Qian.Z., Lutz.S. J Am Chem Soc 2005;127:13466-13467.
4. Schwartz.TU. et al. Protein Sc 2004;13:2814-2818.
5. Kojima.M. et al. J Biosci Bioeng 2005;100:197-202
6. Baird.GS. et al. Proc Natl Acad Sci USA 1999;96:11241-11246.

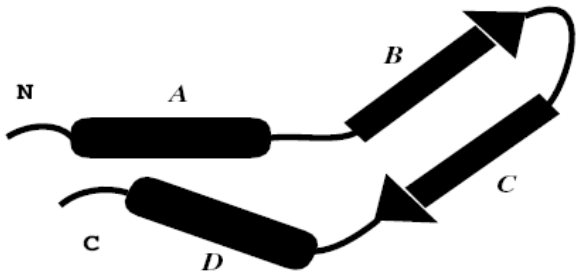
Fluorescent Calcium Sensor with CP



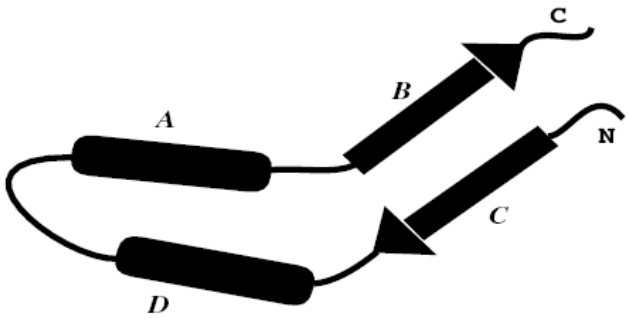
G.S. Baird, et al. **Circular permutation and receptor insertion within green fluorescent proteins.** *PNAS* 1999;96:11241-11246

Circular Permutation (CP)

- Circular permutation of a protein can be visualized as if the original N- and C-termini were linked and new ones created elsewhere[1].
- In most of the cases, naturally occurring CPs have similar 3D structures and conserved biological functions[2].
- **Efficient CP search tool is not available yet.**



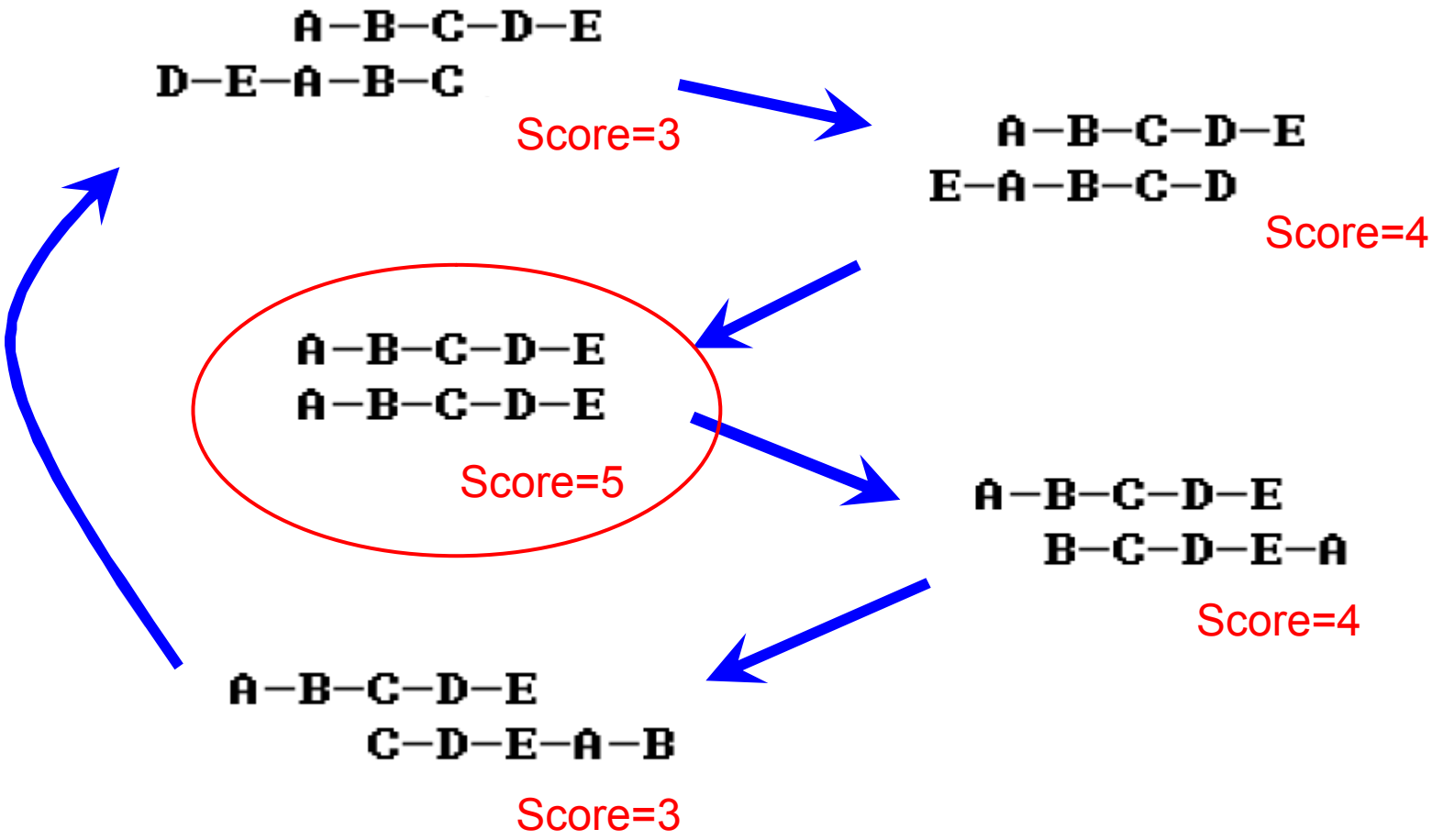
The sequence: ..A..B..C..D..



The sequence ..C..D..A..B..

1. Uliel S et al.: **A simple algorithm for detecting circular permutations in proteins.** *Bioinformatics* 1999,**15**:930-936.
2. Lindqvist Y, Schneider G: **Circular permutations of natural protein sequences: structural evidence.** *Curr Opin Struct Biol* 1997,**7**:422-427.

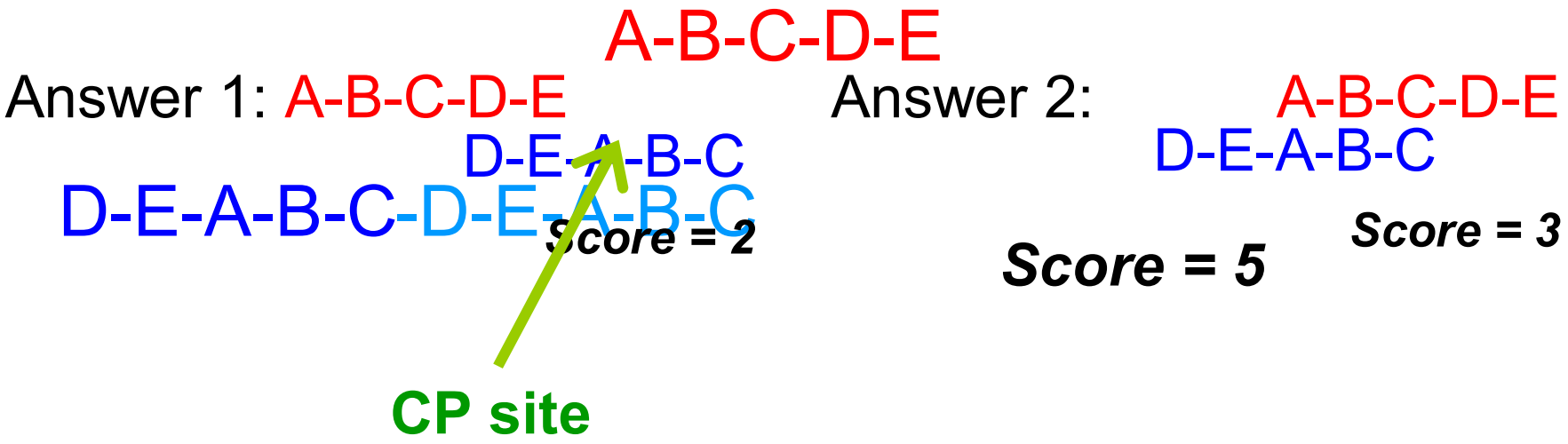
Basic Approach to the Detection of CP



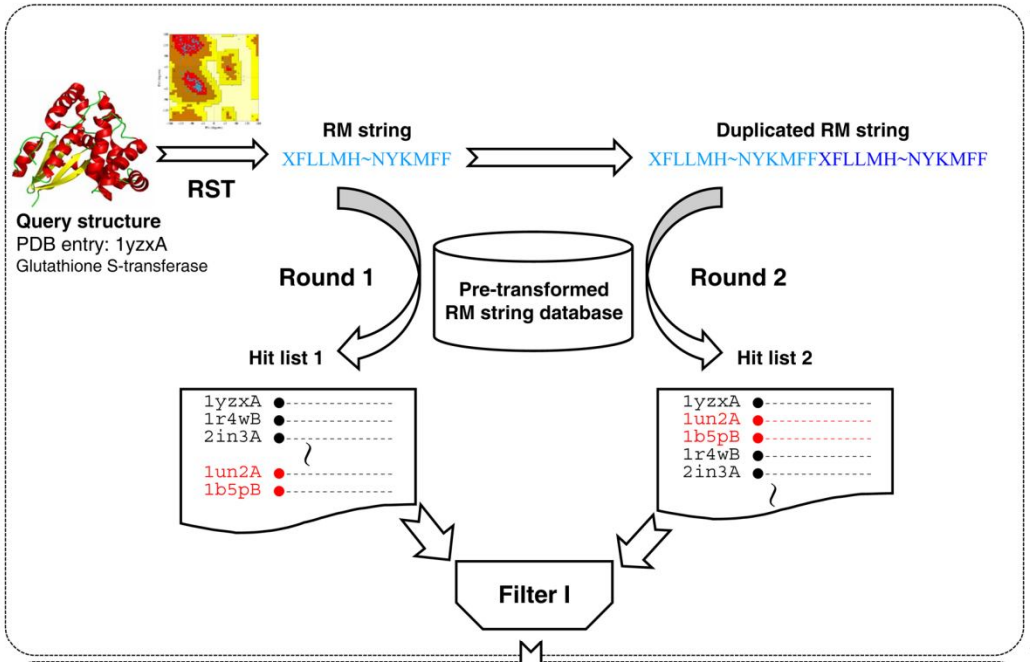
The Basic Idea of CPSARST

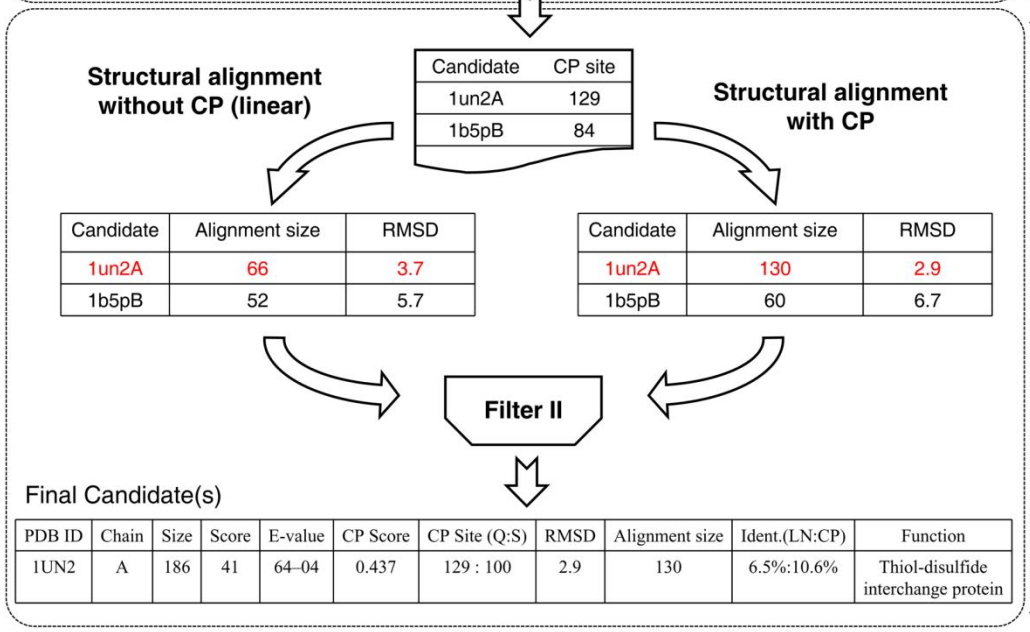
Target: A-B-C-D-E

Query: D-E-A-B-C



The Double Filter-and-Refine Strategy



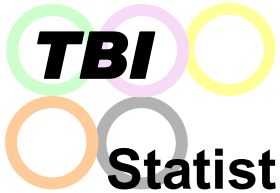


Candidate	CP site
1un2A	129
1b5pB	84

Candidate	Alignment size	RMSD
1un2A	66	3.7
1b5pB	52	5.7

Candidate	Alignment size	RMSD
1un2A	130	2.9
1b5pB	60	6.7

PDB ID	Chain	Size	Score	E-value	CP Score	CP Site (Q:S)	RMSD	Alignment size	Ident.(LN:CP)	Function
1UN2	A	186	41	64–04	0.437	129 : 100	2.9	130	6.5%:10.6%	Thiol-disulfide interchange protein



Statistics of protein structural database searches by CPSARST

Database			nrPDB-90	nrSCOP-90
No. of proteins			14,422	11,688
No. of candidate pairs	1. Detected by amino acid sequence		5,020	1,802
	2. Detected only by Ramachandran string		252,287	196,533
	3. Confirmed after the refinement stage	Total	2,911	4,228
		Symmetric CP	682	1,161
Total No. of protein pairs			208.0×10^6	136.6×10^6
Total running time (minutes)			3,942	1,974
No. of protein pairs scanned per minute			52,764	69,204

Speed Advantage of CPSARST

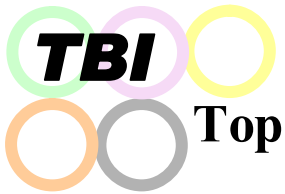
- 4 times faster than UFAU (sequence-based)
  - Uliel S et al.: **A simple algorithm for detecting circular permutations in proteins.** *Bioinformatics* 1999,**15**:930-936.
- 8,824 times faster than SAMO (structure-based)
  - Chen L et al.: **Revealing divergent evolution, identifying circular permutations and detecting active-sites by protein structure comparison**. *BMC Struct Biol* 2006, **6**:18.
- CPSARST requires only 1.7 minute to scan the current PDB (~90,000 polypeptides).

Performance of pair-wise comparisons for natural candidate CP pairs over various sequence identities

$$\frac{\text{RMSD}}{\left(\frac{\text{Alignment size}}{\text{Average protein size}}\right)^{1.5}}$$

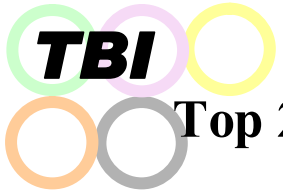
Identity (%)	No. of candidate CP pairs	Structural diversity		
		CPSARST	SHEBA	SAMO
≤ 10	823	6.309	11.180	4.396
10 ~20	152	5.864	13.881	4.994
20 ~ 30	11	3.581	4.506	3.363
30 ~ 40	33	1.868	3.284	2.210
40 ~ 50	40	1.755	3.096	1.544
> 50	9	1.385	2.247	1.520

Lu G: **Top: A new method for protein structure comparisons and similarity searches.** *J Appl Cryst* 2000,**33**:176-183.



Top 20 homologs retrieved from nrPDB by DALI for hypothetical protein YlqF

No.	PDB entry / Size	Function
1	1pujA / 261	Conserved hypothetical protein YlqF
2	1u0lA / 278	Probable GTPase
3	1ctqA / 166	p21h-Ras-1 fragment
4	1ejjA / 508	Phosphoglycerate mutase (isomerase)
5	1gpmA / 501	Amidotransferase, GMP synthetase
6	1efcA / 386	Elongation factor Eftu (RNA binding)
7	1hrkA / 359	Ferrochelatase fragment (lyase)
8	1ni5A / 428	Putative cell cycle protein Mesj
9	1dpgA / 485	Glucose 6-phosphate reductase
10	2hjgA / 390	GTP-binding protein engA
11	1veeA / 134	Unknown function proline-rich protein
12	1cqxA / 403	Flavohemoprotein (lipid binding)
13	2p8zT / 813	Elongation factor 2
14	1mkyA / 400	Probable GTP-binding protein
15	1dar / 615	Elongation factor G (translational GTPase)
16	1kk1A / 397	Eif2gamma mutant
17	1hurA / 180	Human ADP-ribosylation factor 1
18	1fdr / 244	Flavodoxin reductase
19	2clsA / 179	Rho-related GTP-binding protein
20	1wcwA / 254	Uroporphyrinogen III synthase
21	1ak1 / 308	Ferrochelatase



Top 20 circular permutants detected from nrPDB by CPSARST for hypothetical protein YlqF

No.	PDB entry / Size	Function
1	1ZBD / 203	Rabphilin-3A
2	1KY2 / 182	GTP-binding
3	2F7S / 217	Ras-related protein Rab-27B protein YPT7P
4	2NZJ / 175	GTP-binding protein REM 1
5	1T91 / 207	Ras-related protein Rab-7
6	1X3S / 195	Ras-related protein Rab-18
7	1YU9 / 175	GTP-binding protein, GTPase domain
8	2EW1 / 201	Ras-related protein Rab-30
9	2GF9 / 189	Ras-related protein Rab-3D
10	1YVD / 169	Ras-related protein Rab-22A
11	1PUI / 210	Probable GTP-binding protein engB
12	2O52 / 200	Ras-related protein Rab-4B
13	1U8Y / 168	Ras-related protein Ral-A
14	1HUQ / 164	Rab5C, GTPase domain
15	2HUP / 201	Ras-related protein Rab-43
16	1FZQ / 181	ADP-ribosylation factor-like protein 3
17	2OCB / 180	Ras-related protein Rab-9B
18	1OIV / 191	Ras-related protein Rab-11A
19	2FN4 / 181	Ras-related protein R-Ras
20	1Z0F / 179	Rab14, member Ras oncogene family

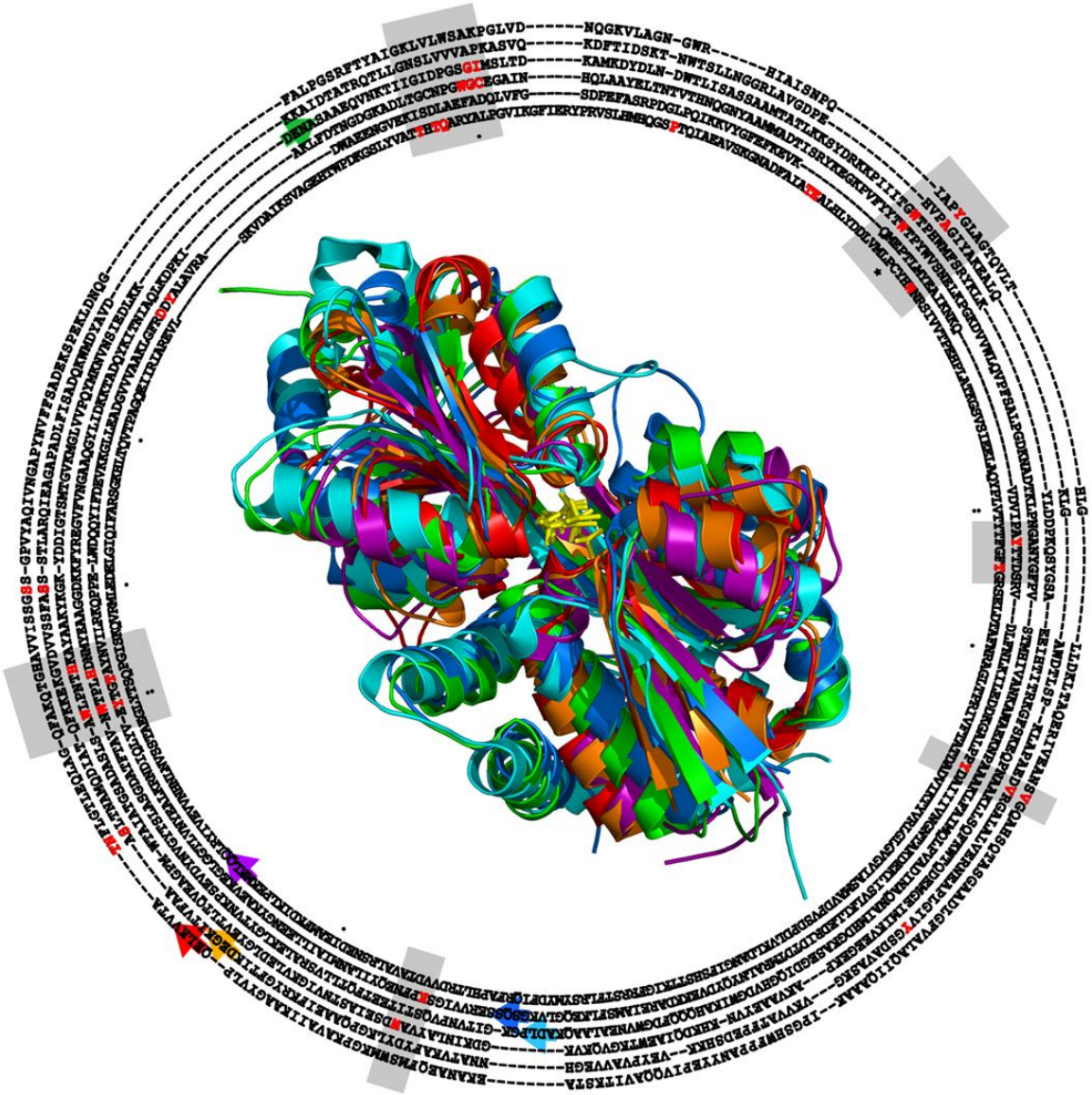
Multiple Alignment of Raw Sequences

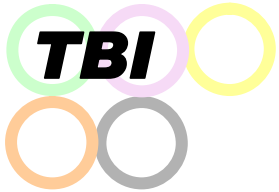
```
1atg    ---------------------------------------------ELKVVTATNFLGTLEQLAGQFAKQTGHAVVISSGSSGPVYAQIVNGAPYNVFFSADEKSPEKLDN----QGFALPG 72
1amf    ---------------------------------------------DEGKITVFAAASLTNAMQDIATQFKKEKGVDVVSSFASSSTLARQIEAGAPADLFISADQKWMDYAVD----KKAIDTA 75
1al3    MKLQQLRYIVEVVNHNLNVSSTAEGLYTSQPGISKQVRMLEDELGIQIFARSGKHLTQVTPAGQEIIRIAREVLSKVDAIKSVAGEHIWPDKGSLYVATTHTQARYALPGV-IKGFIERY 119
1sw1A   ------------GSQSSERVVIGSKPFNEQYILANMIAILLEENGYKAEVKEGLGGTLVNYEALKRNDIQLYVEYTGTAYNVILRKQPPELWDQQYIFDEVKKGLLEADG--VVVAAKLG 106
2b41A   ----------------------DENASAAEQVNKTIIGIDPGSGIMSLTDKAMKDYDLNDWTLISASSAAMTATLKKSYDRKKPIIITGWTPHWMFSRYKLKYLDDPKQSYGSAEEIHTI 98
1r91A   ---------------ADLPGKGITVNPVQSTITEETFQTLLVSRALEKLGYTVNKPSEVDYNVGYTSLASGDATFTAVNWTPLHDNMYEAAGGDKKFYREGVFVNGAAQGYLIDKKTADQ 105
                                                                                                .  .

1atg    SRFTYAIGKLVLWSAKPGLVDNQGKVLAGNGWR--------------HIAISNPQIAPYGLAGTQVLTHLGLLD----------------------KLTAQERIVEANSVGQAHSQTASGA 157
1amf    TRQTLLGNSLVVVAPKASVQKDFT-IDSKTNWTSLLN--------GGRLAVGDPEHVPAGIYAKEALQKLGAWD----------------------TLSP--KLAPAEDVRGALALVERNE 163
1al3    PRVSLHMHQGSPTQIAEAVSKGNADFAIATEALHLYDDLVMLPCYHWNRSIVVTPEHPLATKGSVSIEELAQYP----------------------LVTYTFGFTGRSELDTAFNRAGLTP 218
1sw1A   FRDDYALAVRADWAEENGVEKISDLAEFADQLVFGSD---------PEFASRPDGLPQIKKVYGFEFKEVKQME----------------------PTLMYEAIKNKQVDVIPAYTTDSRV 196
2b41A   TRKGFSKEQPNAAKLLSQFKWTQDEMGEIMIKVEEGE---------KPAKVAAEYVNKHKDQIAEWTKGVQKVK----------------------GDKINLAYVAWDSEIASTNVIGKVL 188
1r91A   YKITN-IAQLKDPKIAKLFDTNGDGKADLTGCNPGWGCEGAINHQLAAYELTNTVTHNQGNYAAMMADTISRYKEGKPVFYYTWTPYWVSNELKPGKDVVWLQVPFSALPGDKNADTKLP 224
          :                .                                                :                                .

1atg    ADLGFVALAQIIQAAAKIPGSHWFPPANYYEPIVQQAVITKST--------------AEKANAEQFMSWMK--GPKAVAIIKAAGYVLPQ---------------- 231
1amf    APLGIVYGSDAVASKG-VKVVATFPEDSHKK--VEYPVAVVEG--------------HNNATVKAFYDYLK--GPQAAEIFKRYGFTIK---------------- 233
1al3    RIVFTATDADVIKTYVRLGLGVGVIASMAVDPVSDPDLVKLDANGIFSHSTTKIGFRRSTFLRSYMYDFIQRFAPHLTRDVVDTAVALRSNEDIEAMFKDIKLPEK 324
1sw1A   DLFNLKILEDDKGALPPYDAIIIVNGNTAKDEKLISVLKLLEDR----------IDTDTMRALNYQYDVEKKDAREIAMSFLKEQGLVK---------------- 275
2b41A   EDLGYEVTLTQVEAGPMWTAIATGSADASLSAWLPNTHKAYAAKYKG---------KYDDIGTSMTGVKMGLVVPQYMKNVNSIEDLKK---------------- 268
1r91A   NGANYGFPVSTMHIVANKAWAEKNPAAAKLFAIMQLPVADINAQNAIMHDG----KASEGDIQGHVDGWIKAHQQQFDGWVNEALAAQK---------------- 309
                                                                                                .  .
```

Multiple Alignment of Circularly-Permutated Sequences



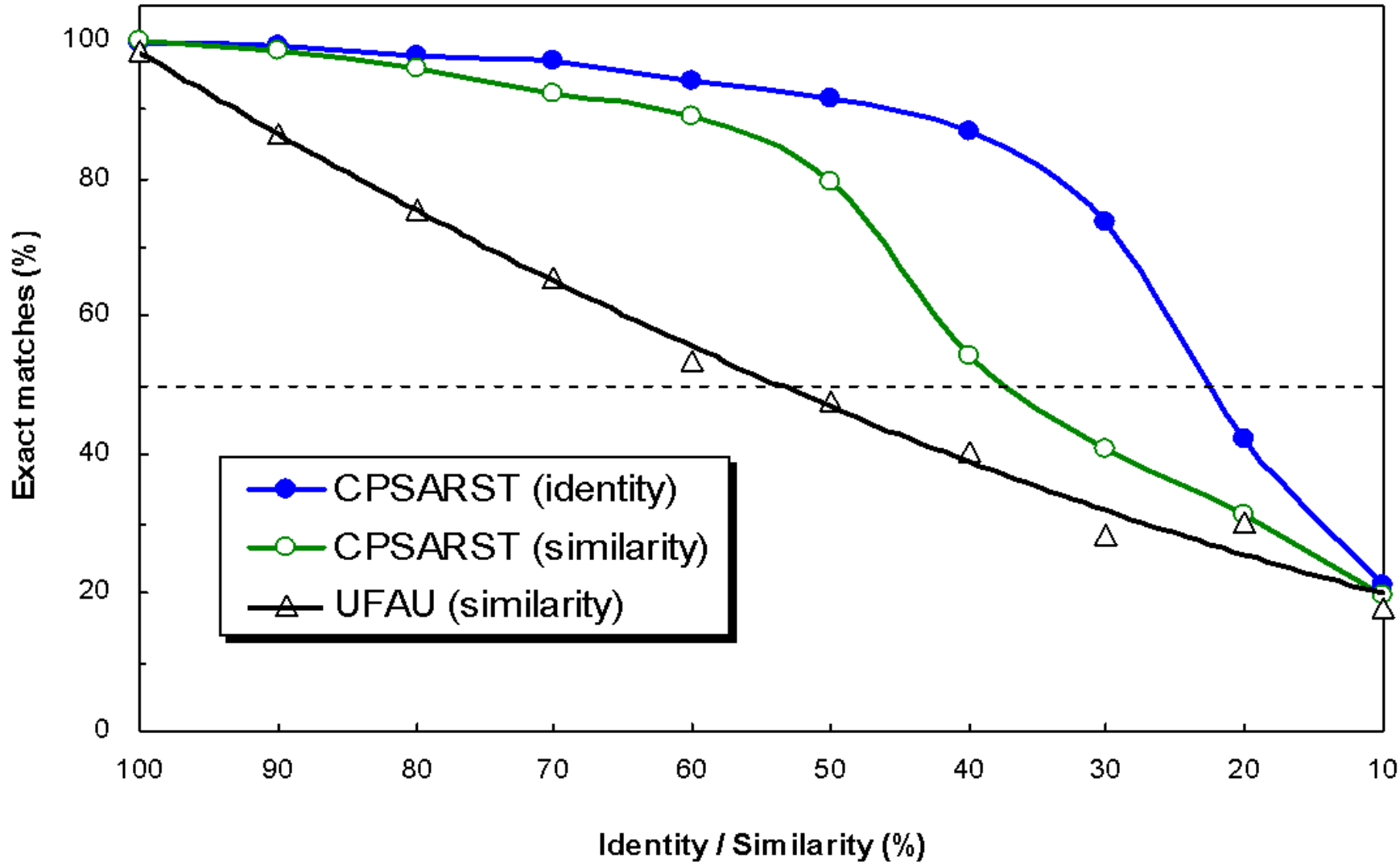


Possible Applications of CPSARST

- Bank-against-bank searches are achievable.
- Develop automated procedures such as the functional assignment system for novel hypothetical proteins
- Construct CP database

TBI

(a)


Exact matches (%)
100
80
60
40
20
0
100
90
80
70
60
50
40
30
20
10
Identity / Similarity (%)
CPSARST (identity)
CPSARST (similarity)
UFAU (similarity)

http://sarst.life.nthu.edu.tw/iSARST/

Home | Downloads | References | Tutorial

Welcome to *i*SARST

Tutorial

Currently 85 threads are running on this PC-cluster.
A typical search along with superimposing **100 structures** takes only **3~5 seconds**.
Circular permutants can be identified, even when the **sequence identity** is **<10%** (➢*Example pair / family*).

Please enjoy the speed, accuracy and convenience brought about by *i*SARST!

Query

○ 5-letter PDB or 7-letter PDB entry(s):

* Format: PDB id + chain ID. If there is no chain ID, simply use the 4-letter PDB id or use '_' to represent the chain.
Multiple entries are acceptable (batch mode), please use the comma to separate them.
Example: **1atpE, 1cewI, 1ti5A, 1JUL, 1HEL_, d1swya_, d1oxda_**

○ Local PDB file: 瀏覽...

& Chain ID in this file:

* If there is no chain ID, please leave it blank or use '_' to represent it.
You can also use '*' as the chain ID and then every chain will be used to search the database.

○ Compressed PDB collection: 瀏覽...

File type: ○ .zip ○ .rar ○ .tar.gz ○ .tar

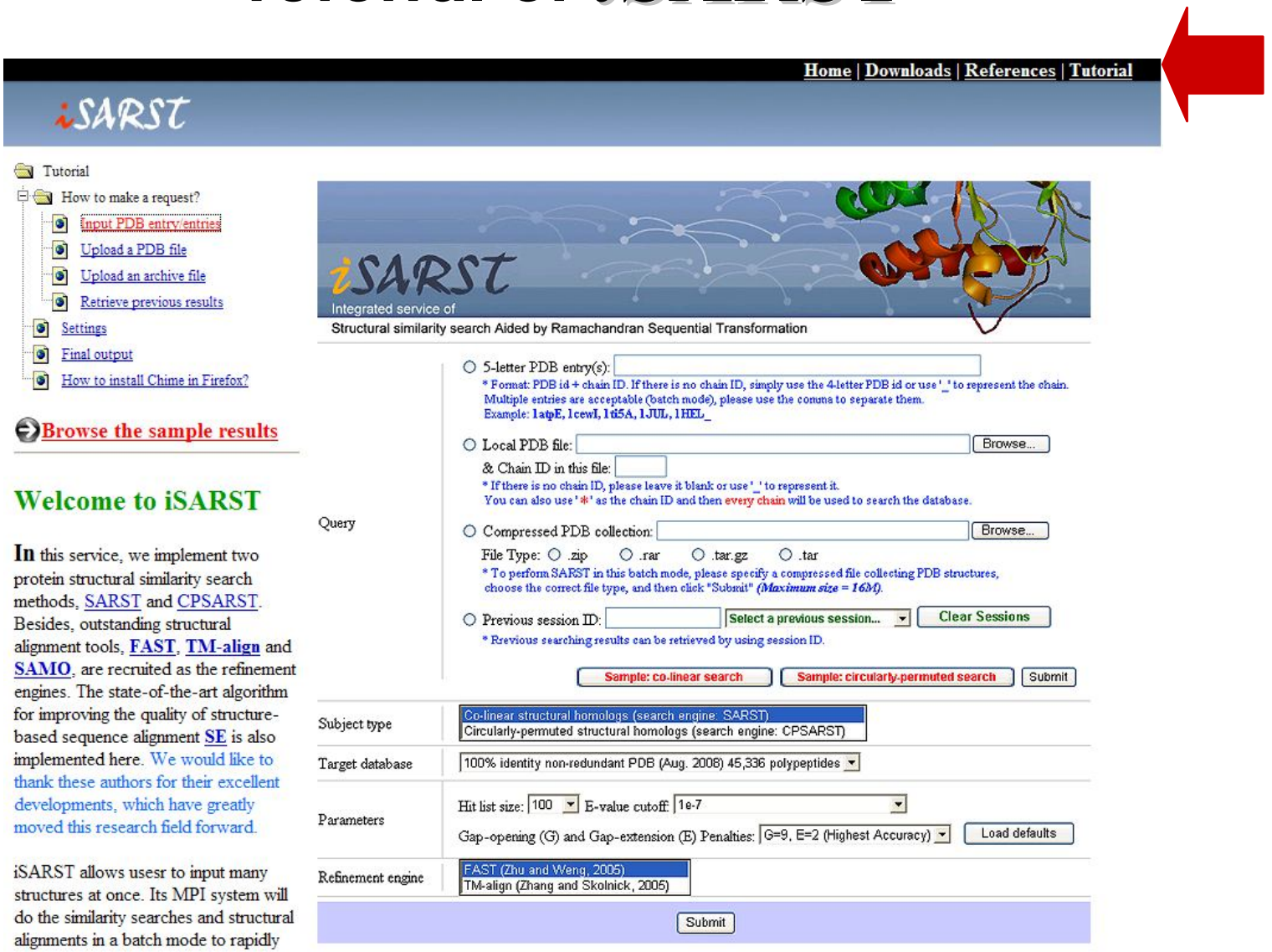
* To perform SARST in this batch mode, please specify a compressed file collecting PDB structures, choose the correct file type, and then click "Submit" ***(Maximum size = 16M)***.

○ Previous session ID:

* Rrevious searching results can be retrieved by using session ID.

Sample: co-linear search | Sample: circularly-permuted search | Submit

Tutorial of *i*SARST



http://sarst.life.nthu.edu.tw/iSARST/hlp/tutorial.php