

# Bioinformatics of Protein Domains: New Computational Approach for the Detection of Protein Domains

Maricel Kann

*Assistant Professor*

*University of Maryland, Baltimore County*

*mkann@umbc.edu*

# The Human Genome Project

GATCCTCCATATACAACGGTATCTCCACCTCAGGTTTAGATCTCAACAACGGAACCATTGCCGACATGAGACAG  
GGTATCGTCGAGAGTTACAAGCTAAAACGAGCAGTAGTCAGCTCTGCATCTGAAGCCGCTGAAGTTCTACTAAG  
GGATAACATCATCCGTGCAAGACCAAGAACCGCCAATAGACAACATATGTAACATATTTAGGATATACCTCGAA  
ATAAACCGCCCACTGTCATTATTATAATTAGAAACAGAACGCAAAAATTATCCACTATATAATTCAAAGACGCGA  
AAAAAGAACAACGCGTCATAGAACTTTTGGCAATTCGCGTCACAAATAAATTTTGGCAACTTATGTTTCCTCTTC  
CAGTACTCGAGCCCTGTCTCAAGAATGTAATAATACCCATCGTAGGTATGGTTAAAGATAGCATCTCCACAACC  
AGCTCCTTGCCGAGAGTCGCCCTCCTTTGTCGAGTAATTTTCACTTTTCATATGAGAACTTATTTTCTTATTCTTT  
CTCACATCCTGTAGTGATTGACACTGCAACAGCCACCATCACTAGAAGAACAGAACAACTTAAATAGAAAAAT  
ATCTTCTCGAAGGCTAATCGATAACTGACGATTTCTGCTTCCAACATCTACGTATATCAAGAAGCATTCACTT  
ATGACACAGCTTCAGATTTTCAATTATTGCTGACAGCTACTATATCACTACTCCATCTAGTAGTGGCCACGCCCTAT  
GCATATCCTATCGGAAAACAATACCCCCAGTGGCAAGAGTCAATGAATCGTTTACATTTCAAATTTCCAATGAT  
TATAAATCGTCTGTAGACAAGACAGCTCAAATAACATAACAATTGCTTTCGACTTACCGAGCTGGCTTTCGTTGAC  
AGTTCTAGAACGTTCTCAGGTGACACTTCTTCTGACTTACTATCTGATGCGAACACCACGTTGTATTTCAATGTA  
TCGAGGGTACGGACTCTGCCGACAGCAGCTCTTTGAACAATACATACCAATTTGTTGTTACAAACCGTCCATCC  
CGCTATCGTCAGATTTCAATCTATTGGCGTTGTTAAAAAACTATGGTTATACTAACGGCAAAAACGCTCTGAAAC  
ATCCTAATGAAGTCTTCAACGTGACTTTTGACCGTTCAATGTTCACTAACGAAGAATCCATTGTGTCGTATTACG  
GTTCTCAGTTGTATAATGCGCCGTTACCCAATTGGCTGTTCTTCGATTCTGGCGAGTTGAAGTTTACTGGGACG  
CCGGTGATAAACTCGGCGATTGCTCCAGAAACAAGCTACAGTTTTGTCATCATCGCTACAGACATTGAAGGATT  
GCCGTTGAGGTAGAATTCGAATTAGTCATCGGGGCTCACCAGTAACTACCTCTATTCAAATAGTTTGATAATC  
GTTACTGACACAGGTAACGTTTCATATGACTTACCTCTAAACTATGTTTATCTCGATGACGATCCTATTTCTTCTG  
AATTGGGTTCTATAAACTTATTGGATGCTCCAGACTGGGTGGCATTAGATAATGCTACCATTTCGGGGTCTGTCC  
ATGAATTAAGTTCGGAAGAACTCCAATCCTGCCAATTTTCTGTGTCCATTTATGATACTTATGGTGATGTGATTTA  
CAACTTCGAAGTTGTCTCCACAACGGATTTGTTTGCCATTAGTTCTCTTCCAATATTAACGCTACAAGGGGTGA  
GTTCTCCTACTATTTTTGCTTCTCAGTTTACAGACTACGTGAATACAAACGTTTCATTAGAGTTTACTAATTCAA  
CAAGACCATGACTGGGTGAAATTCCAATCATCTAATTTAACATTAGCTGGAGAAGTGCCCAAGAATTTGACAAC  
TCATTAGGTTTGAAGCGAACCAAGGTTCACAATCTCAAGAGCTATATTTTAAACATCATTGGCATGGATTCAAAG  
ACTCACTCAAACACAGTGCGAATGCAACGTCCACAAGAAGTTCTCACCCTCCACCTCAACAAGTTCTTACAC  
TCTACTTACACTGCAAAAATTTCTTCTACCTCCGCTGCTGCTACTTCTTCTGCTCCAGCAGCGCTGCCAGCAGCC  
AAAAC TTCATCTCACAATAAAAAAGCAGTAGCAATTGCGTGCGGTGTTGCTATCCCATAGGCGTTATCCTAGTA  
CTCATTTGCTTCTAATATTCTGGAGACGCAGAAGGGGAAAATCCAGACGATGAAAAC TTACCGCATGCTATTAG  
ACCTGATTTGAATAATCCTGCAAATAAACCAAATCAAGAAAACGCTACACCTTTGAACAACCCCTTTGATGATGA  
TTCCTCGTACGATGATACTTCAATAGCAAGAAGATTGGCTGCTTTGAACACTTTGAAATTGGATAACCACTCTGC  
TGAATCTGATATTTCCAGCGTGGATGAAAAGAGAGATTCTCTATCAGGTATGAATACATAACAATGATCAGTTCCA  
CCAAAGTAAAGAAGAATTATTAGCAAAAACCCCGAGTACAGCCTCCAGAGAGCCCGTTCTTTGACCCACAGAATA  
CTTCTTCTGTGTATATGGATAGTGAACCAGCAGTAAATAAATCCTGGCGATATACTGGCAACCTGTCACCAGTCT  
ATAATTGTCAGAGACAGTTACGGATCACAAAAAACTGTTGATACAGAAAAAACTTTTCGATTTAGAAGCACCAGAGA  
AAAAACGTACGTCAAGGGATGTCACTATGTCTTCACTGGACCCTTGGAACAGCAATATTAAGCCGTTACCGTA  
AAATCAGTAACACCATCACCATATAACGTAACGAAGCATCGTAACCGCCACTTACAAAATATTCAAGACTCTCAA

READING THE ENTIRE  
GENOME CODE...

Listen to this  
part: 'AGGCTAATC  
CGCATAACTG'

Wow!



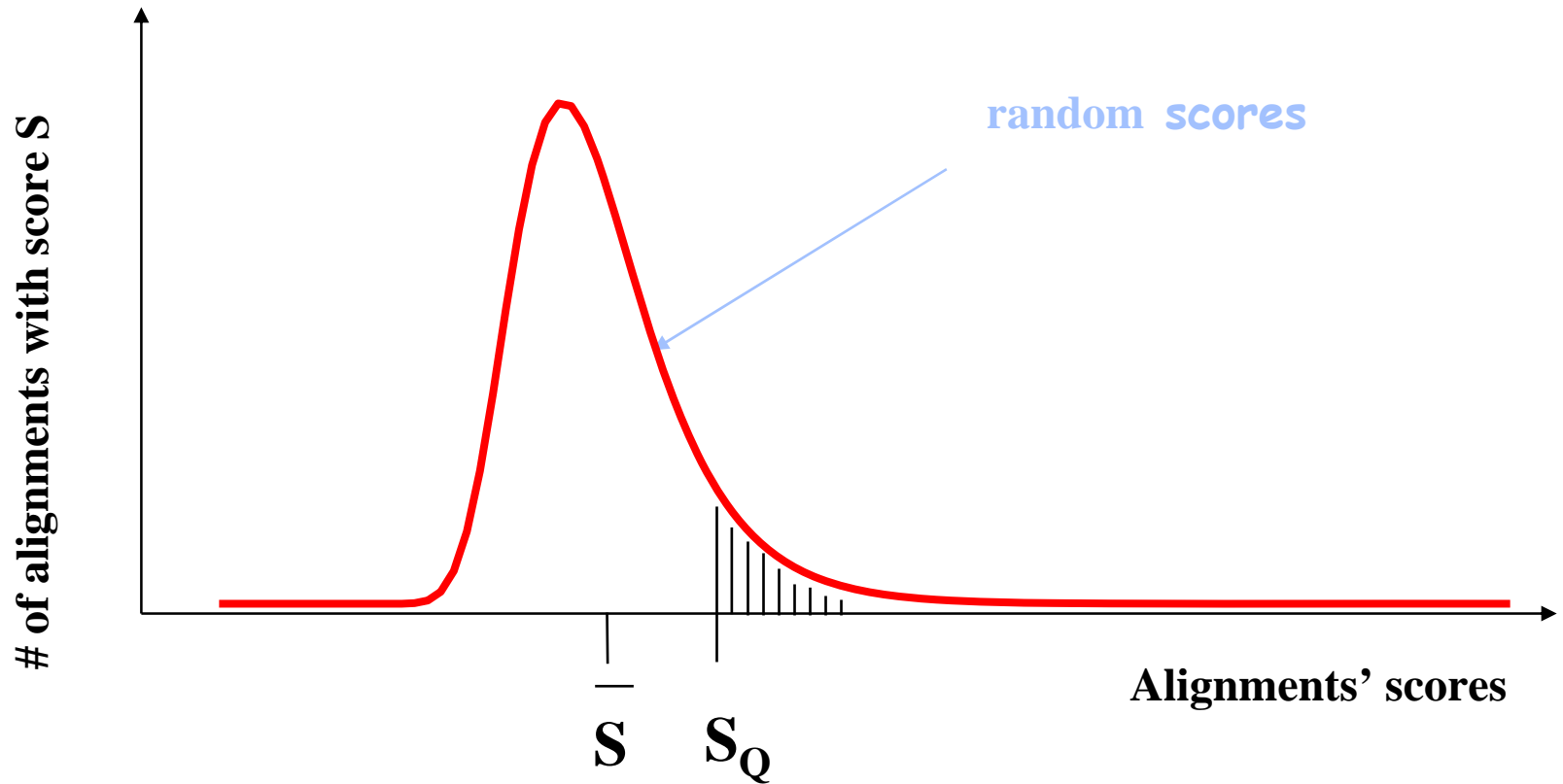


# Significance of a score

Estimated number of non-related sequences in the database that score higher than the query

$$E = p(S_Q < S_R)D$$

D= size of database

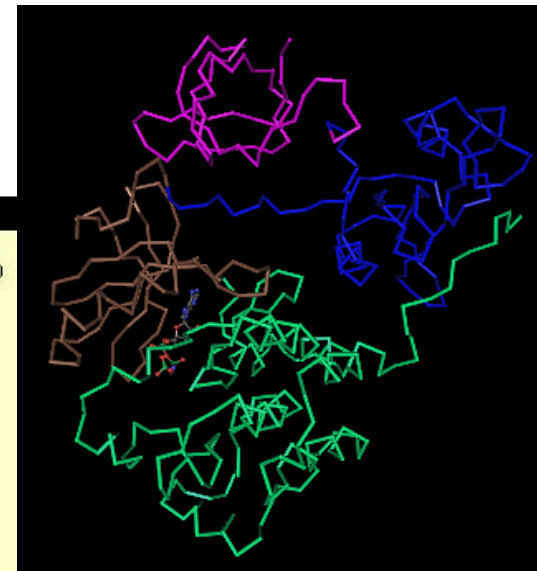
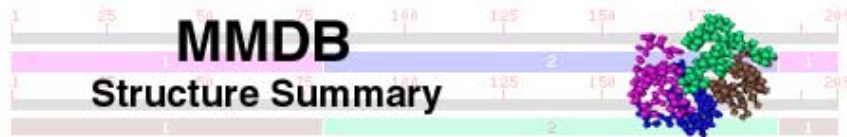


$$p(S_Q < S_R) = 1 - \exp[-KMNe^{-\lambda S_R}]$$

# Outline

A new computational approach for the detection of protein domains: *Semi-Global Alignment of Protein Domains with accurate statistics.*

- Introduction
  - Definition of protein domain.
  - Main features of the Conserved domain database (CDD)
  - Position specific scoring matrices (PSSM)
  - Classification of alignment methods
- Current methods for protein domain searches
- Our approach (Global Blocks Aligned Locally)
- Results

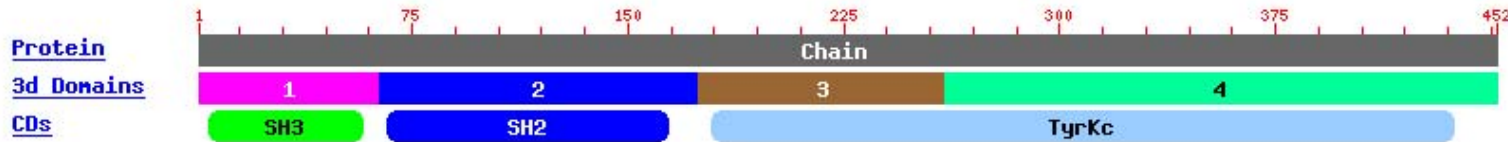


[PubMed](#)   [BLAST](#)   [Structure](#)   [Taxonomy](#)   [OMIM](#)   [Help?](#)

**Description:** Crystal Structure Of Human Tyrosine-Protein Kinase C-Src, In Complex With Amp-Pnp  
**Deposition:** Xu W, Doshi A, Lei M, Eck MJ, Harrison SC, 1998/12/29  
**Taxonomy:** [Homo sapiens](#)  
**Reference:** [PubMed](#)   **MMDB:** [11026](#)   **PDB:** [2SRC](#)   **Structure Neighbors:** [VAST](#)

of    [NEW Get Cn3D 4.1!](#)

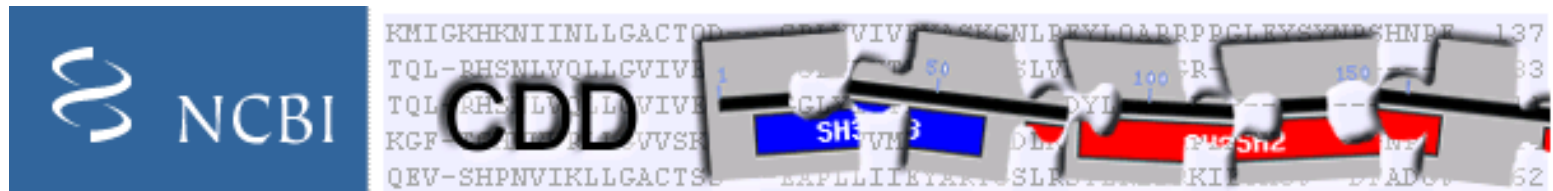
The graphics below indicate the individual chains, 3D domains and ligands identified, if present, in the MMDB structure. You may view them in [Cn3D](#) by clicking "View 3D Structure" above. You may also click each icon to get more information. [?](#)



The term protein domain (or *domain*) refers to a region of the protein with compact structure, usually with a hydrophobic core.

# Conserved Domains

- In 1974 Michael Rossmann recognized the NADH binding domain in several dehydrogenases (named after him).
- Conserved domains are determined by **sequence comparative analysis**.
- Molecular **evolution** uses such domains as **building blocks**
- They may be recombined in different arrangements to make proteins with different **functions**.
- Most proteins contain **multiple domains** (65% euk, 40% prok), giving rise to a variety of combinations of domains.



CDD: a collection of domain multiple alignments linked to protein 3D structure

### cd01040.2

#### Links:

- Source: Cdd
- Taxonomy: cellular organisms
- PubMed: 5 links
- Book: 3 book links
- Protein: cd01040 related architectures representatives
- Related CD: 2 links

#### Statistics:

- PSSM-Id: 29979
- Aligned: 203 rows
- Status: curated CD
- Created: 10-Jan-2006
- Updated: 10-Jan-2006

#### Structure:



### globin

Globins are heme proteins, which bind and transport oxygen. This family summarizes a diverse set of homologous protein domains, including: (1) tetrameric vertebrate hemoglobins, which are the major protein component of erythrocytes and transport oxygen in the bloodstream, (2) microorganismal flavohemoglobins, which are linked to C-terminal FAD-dependent reductase domains, (3) homodimeric bacterial hemoglobins, such as from *Vitreoscilla*, (4) plant leghemoglobins (symbiotic hemoglobins, involved in nitrogen metabolism in plant rhizomes), (5) plant non-symbiotic hexacoordinate globins and hexacoordinate globins from bacteria and animals, such as neuroglobin, (6) invertebrate hemoglobins, which may occur in tandem-repeat arrangements, and (7) monomeric myoglobins found in animal muscle tissue.

Feature List	Evidence	Feature Name	Labels in Sequence Alignment
	<input type="checkbox"/>	1: heme-binding site	<input type="checkbox"/>

cd01040 is part of a hierarchy of related CD models. Use the graphical representation to navigate this hierarchy.

#### cd01040 Sequence Cluster



#### Sub-family Hierarchy



Feature 1	#	#	##	#	#	##					
1ASH	1	ANKTRELCKMSL	[12]	.QDGIDLYKHMFFENY	[2]	.LRKYF	[16]	.FAKQGQKILLACHVLC	[13]	.ELDDRHR	99
1HDA_A	3	SAADKGNVKA	[8]	.EYGAEALERMFLSF	[2]	.TKTYF	[11]	.VKGHGAKVAAALTKAVE	[10]	.ELSDLHA	89
gi 422406	6	SDSEEKLV	[8]	.GTANTVFYNYLKKY	[2]	.NQDKF	[16]	.FKLIAGRIFTIFDNCVK	[13]	.DMSGPHVA	100
gi 1730834	167	TARPRKTKQ	[7]	.LFCSQFYDNLIAM	[2]	.LEEYF	[2]	.LKHQAVSFCKVLD	[13]	.KLGKRHSR	246
gi 2105139	206	TPHQIRDVQ	[8]	.ALVSSIFVKLFKET	[2]	.IQKFF	[16]	.YEKQIALVADRLDTMIS	[13]	.YMRYTHT	300

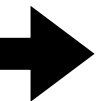
It combines information about **protein sequence**, their **conservation** patterns across evolution and the **protein structure** and provide useful **functional** annotation.

# Protein Classification

QUERY



Alignment Algorithm  
Scoring Function  
Accurate Statistics



Set of related sequences or protein family from database

PSSM can be derived from the MSA

```
      10      20      30      40      50      60      70      80
      .....*.....*.....*.....*.....*.....*.....*.....*.....*.....|
Feature 1      #      # # #      #
1ASH          1 ANKTRELCMKSLehakvd--tsnearQDGIDLYKHMFENYp-----pLRKYFksreeyta-----edvqndpfFAKQ 65
1HDA_A        3 SAADKGNVKAAWgktvg-----ghaaEYGAEALERMFLSFp-----tTKTYFphfdls-----hgsaqVKGH 58
gi 422406     6 SDSEEKLVRDAWapih-----gdlqGTANTVFYNYLKKYp-----sNQDKFetlkghpl-----devkdtanFKLI 66
gi 1730834    16TARPRKTKQRDNdnkv-----dtaLFCSQFYDNLIAMDp-----LLEEYFp-----sLKHQ 21
gi 2105139    20TPHQIRDVQRSWenir-----ndrnALVSSIFVKLFKETp-----rIQKFFakfanvav-----dslagnaeYEKQ 26
gi 17509143   16 KPEGRKADNQILnsy-----qkSIVRNAWRHMSQKGpsncgstiTRRMA
gi 17541936   16DKESCEVVADSWrlvesrssaaetsaCFGLVFQRVFSKIp-----mLRPLFg
gi 17558092   3VDDDFELARTHWiqlqk-----snkqgLAIRGCFLTMLEKYp-----qVRPIWg
gi 17570331   5SPEHQKLIKRSWnri-----pkaQFGRASLEAFITAaq-----vTHAIFv
gi 25155167   14SPYQQKLLVQCWpniyt----tgasgPFANSLYSTLSSRNa-----kAKELLa

      90      100     110     120     130
      .....*.....*.....*.....*.....*.....*.....*.....*.....*.....|
Feature 1      #      # # #      # # #
1ASH          66 GQKILLACHVLCAtyddret-fnaytrELLDRHARghvh---mppeVUTDFWKLF
1HDA_A        59 GAKVAAALTKAVEhlddl---pgalsELSDLHAHklrv---dpvNFKLLSHSL
gi 422406     6AGRIFTIFDNCVKnvgndkg-fqkviaDMSGPHVArpit---hgsYNDLRGVIY
gi 1730834    21SVSFCKVLDSAIDnlenvhv-lddyivKLGKRHSRilgi---ktvGFEVMGKAF
gi 2105139    26IALVADRLDTMISamddklq-llgninYHRYTHTErgi---praPWEDFSRLL
gi 17509143   78 NLQIVEFLQKVMQsldepdk-isklcqEIGQKHAKyrrekgmkidYUDKLGEAI
gi 17541936   22ARLFTSILHISVKnvdeleaqvaptvfKYGERHYRpditphmteeNVRVFCAQI
gi 17558092   10CASLQAALNMIIQnkddksg-mrrmlnEMGAHHFFyda---cepHFEVFQDSL
gi 17570331   10VKYFVDLVQSCVDnlenletgvpwldLIGRGHANFKi---tgkHWEKFGESL
gi 25155167   20CRVTVEILDTVIKnldndharitqyltEIGQKHRHlkaeg-lssaVUDDLGDTI
```

A PSSM, or Position-Specific Scoring Matrix (or profile), is a type of scoring matrix in which amino acid substitution scores are given separately for each position in a protein multiple sequence alignment.

# MSA contains conserved blocks

Other Related Conserved Domains: **COG1017** **pFan00042**

Reformat Sequence Alignment Format: **Hypertext** Row Display: **up to 10** Color Bits: **2.0 bits** Type Selection: **the most**

```

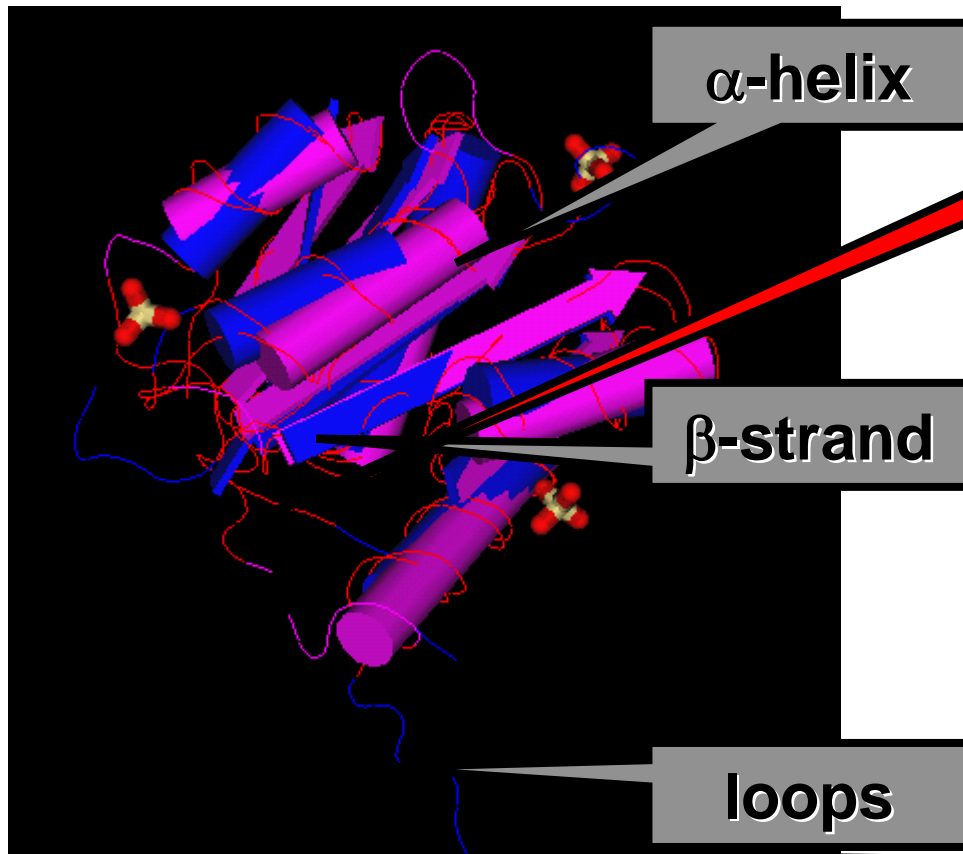
      10      20      30      40      50      60      70      80
.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|
Feature 1
1ASH      1 ANKTRELCMKSLehakvd--tsnearQDGIDLYKHMFENYp-----pLRKYFksreeyta-----edvqndpfFAKQ 65
1HDA_A    3 SAADKGNVKAAWgkvg-----ghaaEYGAEALERMFLSFp-----tTKTYFphfdls-----hgsaqVKGH 58
gi 422406  6 SDSEEKLVRDAWapih-----gdlqGTANTVFYNYLKKYp-----sNQDKFetlkghpl-----devkdtanFKLI 66
gi 1730834 167 TARPRKTKQRDNdnkv-----dtaLFCSQFYDNLIAMDP-----lLEEYFp-----sLKHQ 212
gi 2105139 206 TPHQIRDVQRSWenir-----ndrnALVSSIFVKLFKETp-----rIQKFFakfanvav-----dslagnaeYEKQ 266
gi 17509143 16 KPEGRKADNQILnsy-----qkSIVRNAWRMSQKGpsncgstiTRRMarkstig-----dildrstLDYH 77
gi 17541936 162 DKESCEVVADSWrlvesrssaaetsaCFGLFVFQRFSKIp-----mLRPLFglsesddv-----fdlpdnhpVRRH 228
gi 17558092 37 VDDDFELARTHWiqlqk----snkqgLAIRGCFLTMLEKYp-----qVRPIWgfgkriegrgdetwkpeivedfyFRHH 106
gi 17570331 58 SPEHQKLIKRSWnri-----pkaQFGRASLEAFITAAq-----vTHAIFvdk-----etENRH 105
gi 25155167 144 SPYQQKLLIVQCWpniyt----tgasgPFANSLYSTLSSRNa-----kAKELLakadgvav-----fsksdfdcSVMH 206

      90      100      110      120      130      140      150      160
.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|.....*.....|
Feature 1
1ASH      66 GQKILLACHVLCAtyddret--fnaytrELLDRHARdhvh--mppeVWTDFWKLFEEYLgkktt----ldeptKQAWHEIG 138
1HDA_A    59 GAKVAAALTKAVEhlddl----pgalsELSDLHAHklrv---dpvNFKLLSHSLLVTLashlps--dftpvHASLDKFL 129
gi 422406  67 AGRIFTIFDNCVKnvgndkg--fqkviaDMSGPHVArpit---hgsYNDLRGVIYDSMHlds-----thGAAWNKMM 133
gi 1730834 213 AVSFCKVLDSAIDnlenvhv--lddyivKLGRHSRilgi---ktvGFEVMGKAFMTTLqdrfgs--fltlelKNLWGQLY 286
gi 2105139 267 IALVADRLDTMISamddklq--llgninYMRYHTErgi---praPWEDFSRLLDVLGskgvst--ddldswKGLAVFV 340
gi 17509143 78 NLQIVEFLQKVMQsldepdk--isklcqEIGQKHAKYrrskgmkidYWDKLGEAITETIreyqgw--kihresLRAATVLV 154
gi 17541936 229 ARLFTSILHISVKnvdeleaqvaptvfKYGERHYRpditphmteeNVRVFCAQIVCTVfdflrdt--eatpkcAESWIELM 307
gi 17558092 107 CASLQAALNMIIQnkddksg--mrrmlnEMGAHHFFyda----cepHFEVFQDSLLESMklvlnnggdslddieQSWICAA 181
gi 17570331 106 VKYFVDLVQSCVDnlenletgvkpwldLIGRGHANfki----tgkHWEKFGESLLTTAtewngpg--rrhketVKAWMVMS 180
gi 25155167 207 CRVTVEITDTVIKnlnddharitqultFTGKOHRHlkaeg--laaVWDDIGDTMDCArryceav--rkkelRRAWIAIT 284

```

# Protein Sequence Conservation Occurs in Blocks with Intervening Gaps

## Protein Structure Alignment

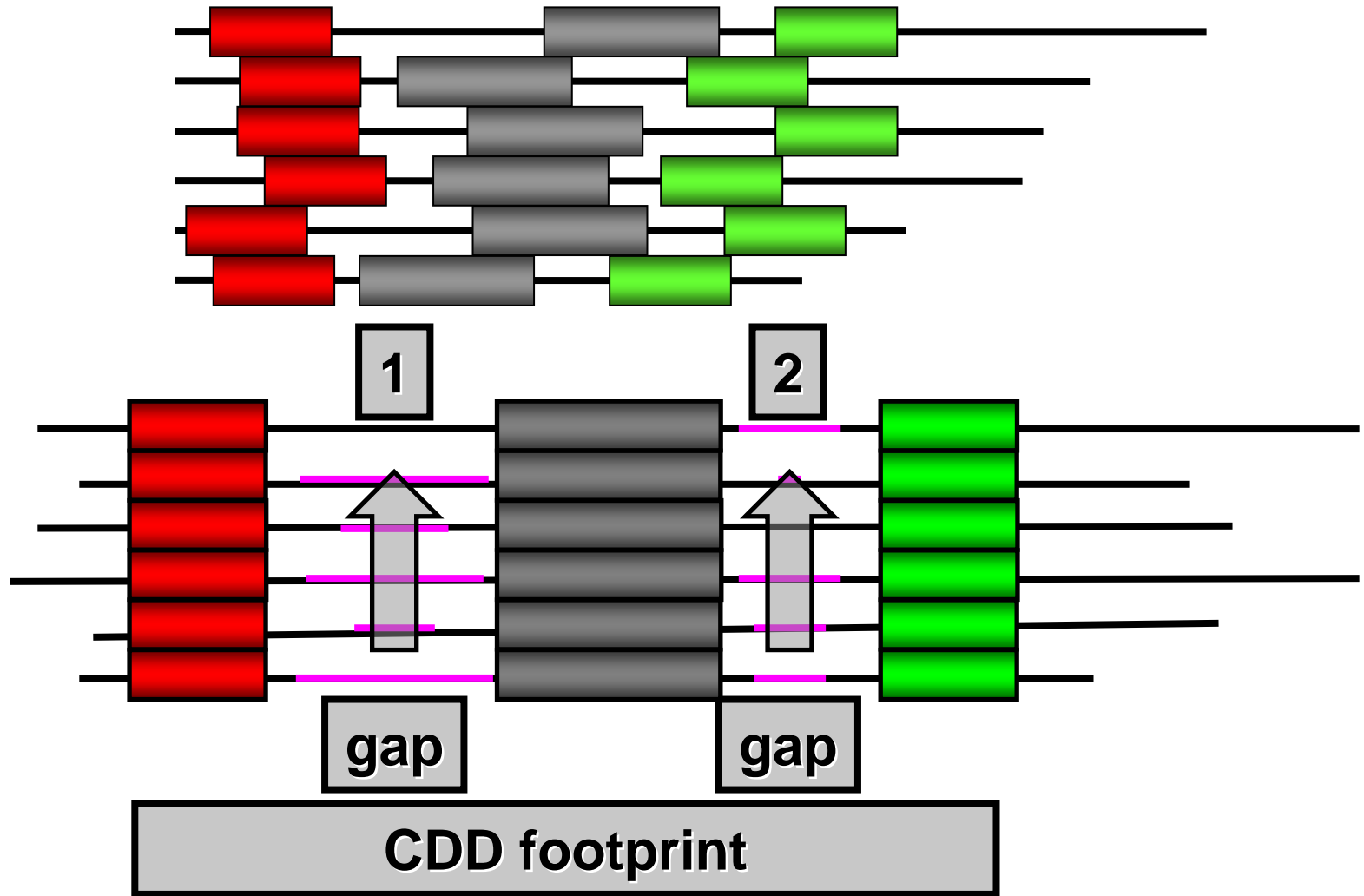


**red  
sequence**

**Subsequences corresponding to secondary structure elements (SSEs:  $\alpha$ -helices and  $\beta$ -strands) are more conserved than the intervening loops.**

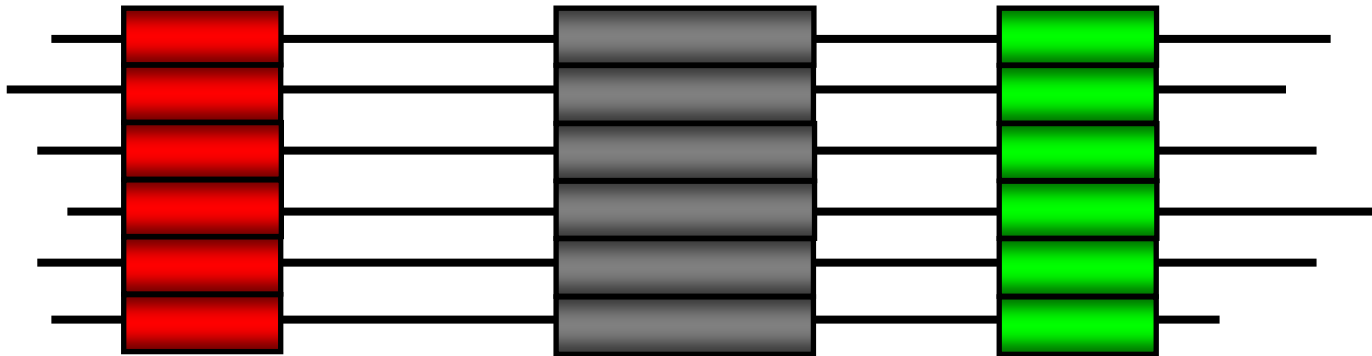
**blue  
sequence**

# CDD representation

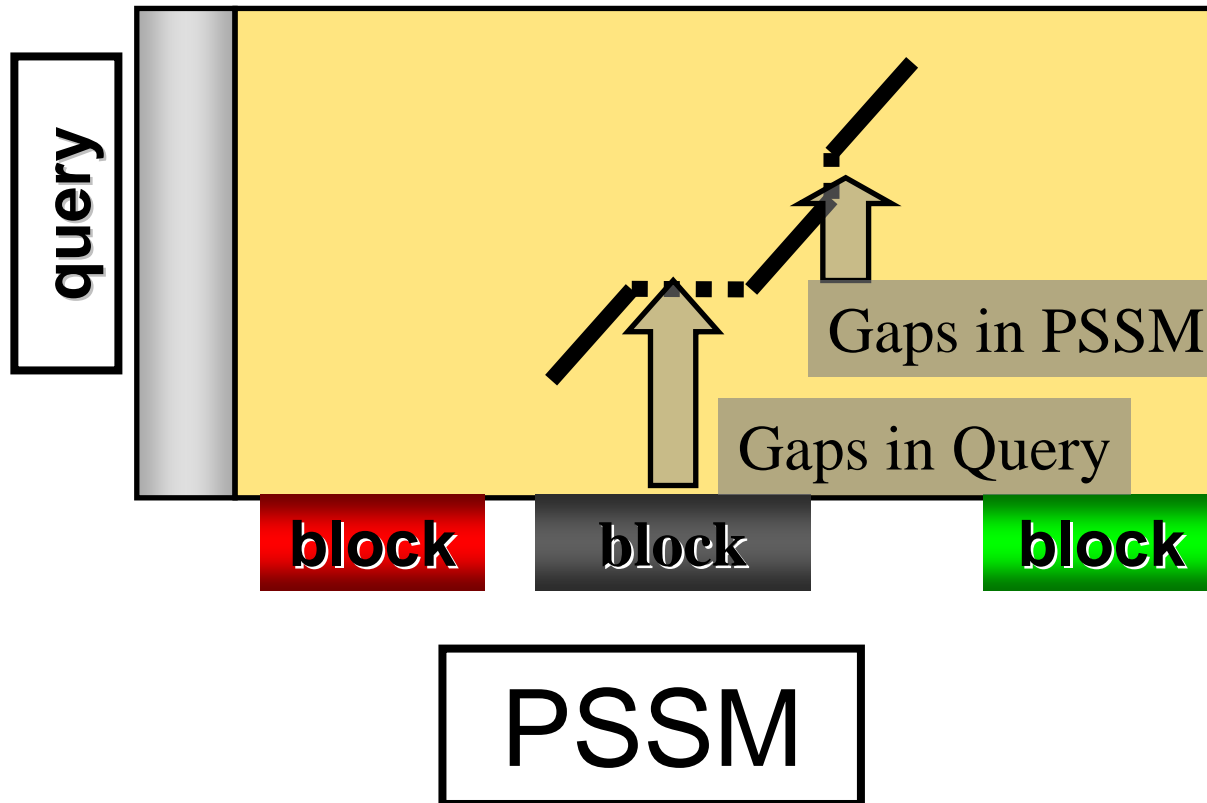


# Sequence-PSSM alignment

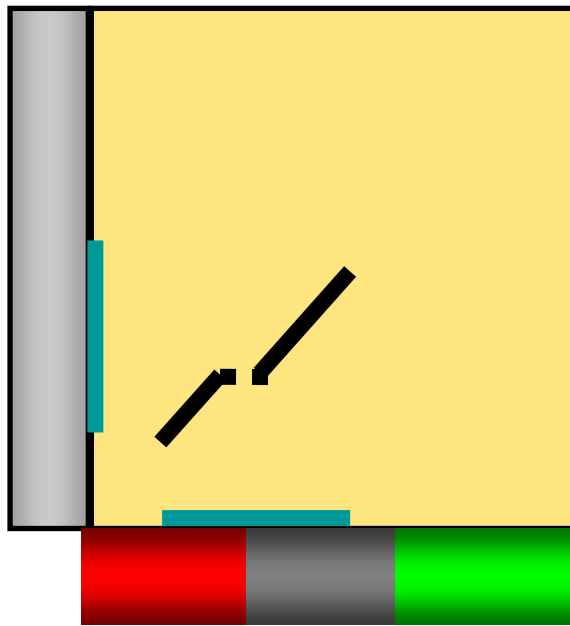
A L I G N M E N T



# Sequence-PSSM alignment

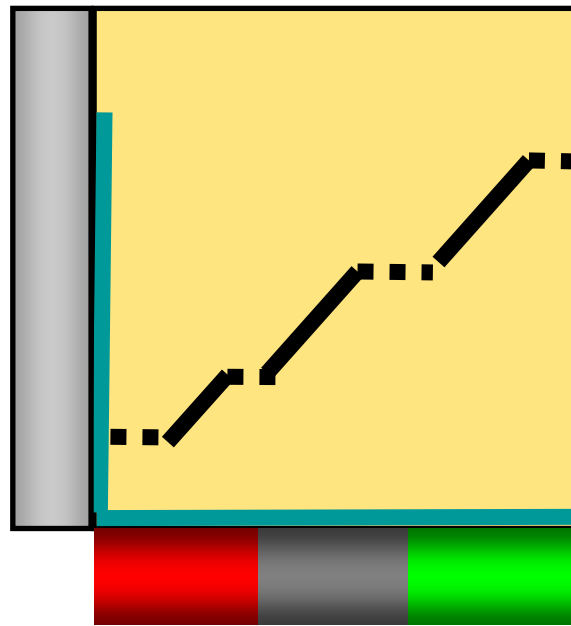


# Three Types of Sequence Alignments



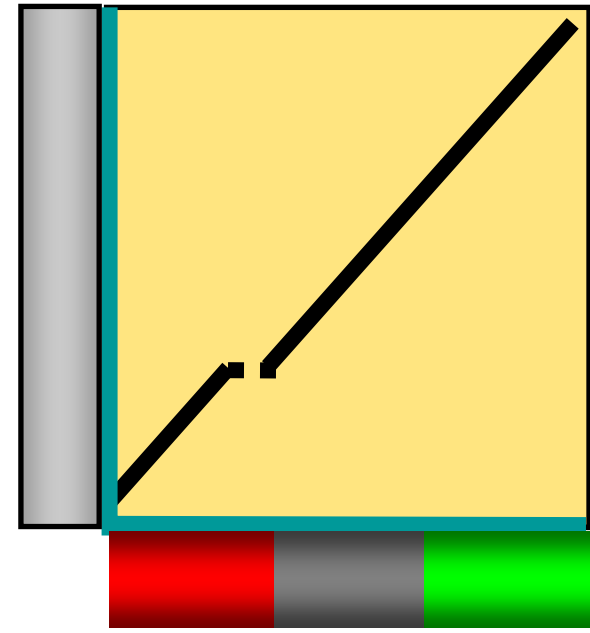
**Local  
Alignment**

**Subsequence  
To  
Subsequence**



**Semi-Global  
Alignment**

**Subsequence  
Onto  
Sequence**

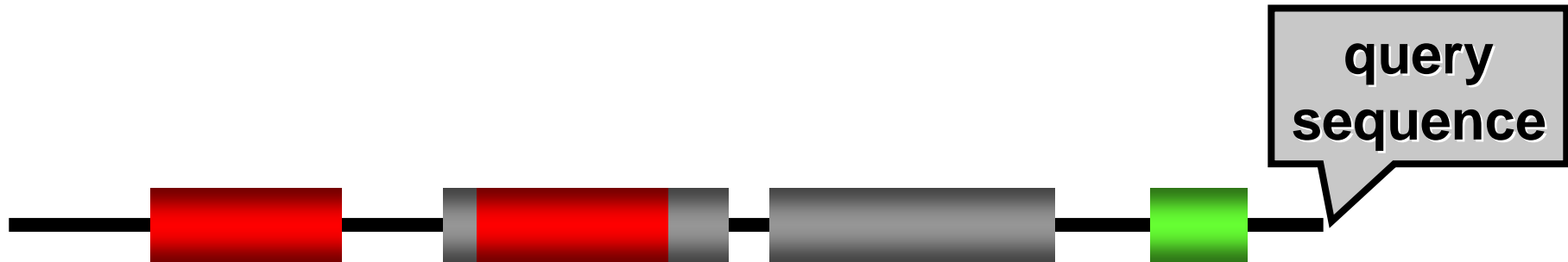


**Global Alignment**

**Sequence  
To  
Sequence**

# Semi-global Alignment

- Finding a complete domain in the query , semi-global, is the natural choice in the context of the protein structure, function and evolution

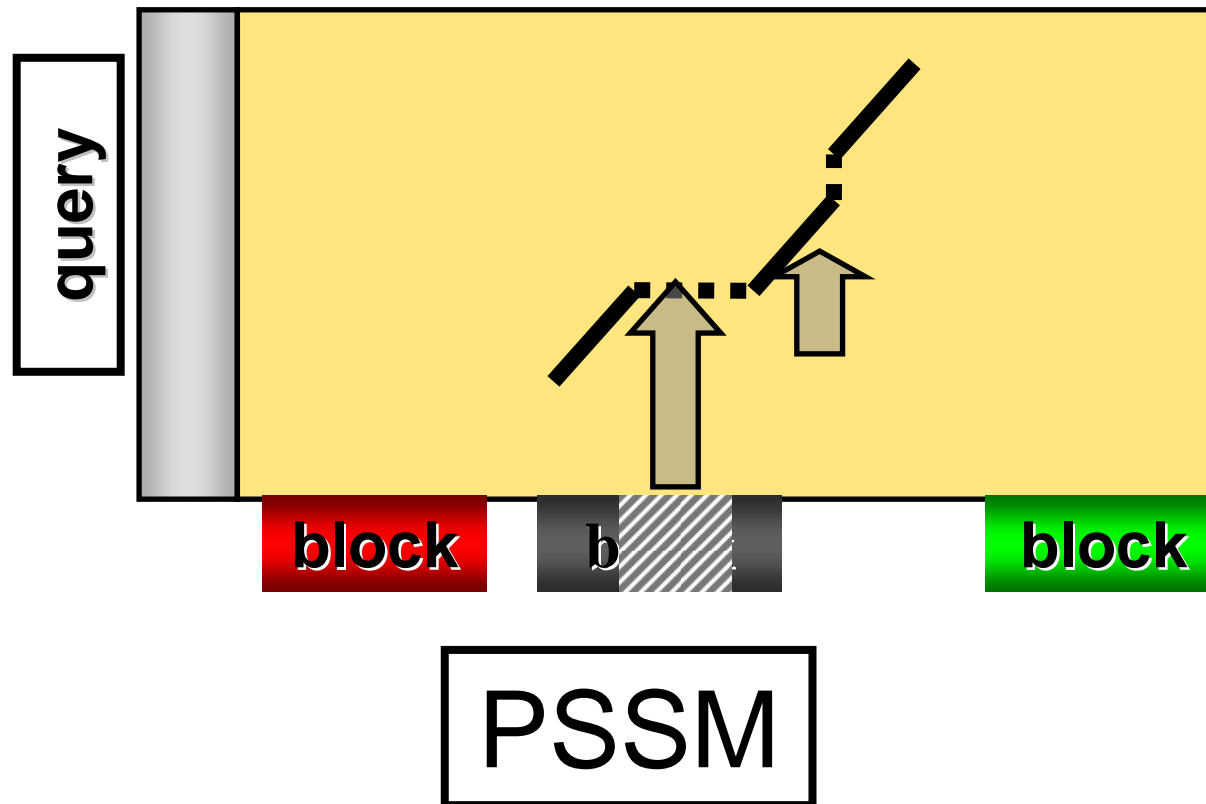


# Outline

A new computational approach for the detection of protein domains: *Semi-Global Alignment of Protein Domains with accurate statistics.*

- Introduction
- Current methods for protein domain searches
  - RPS-BLAST
  - HMMer
  - SALTO
- (Global Blocks Aligned Locally)
- Derivation of Statistics
- Results

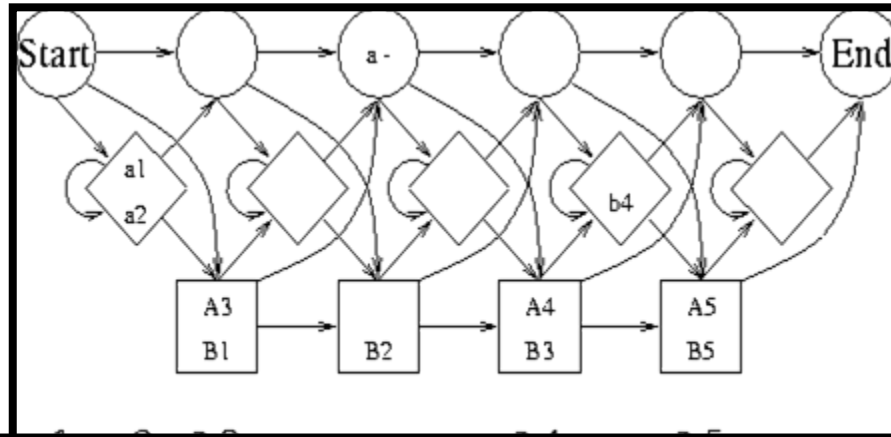
# Reverse Position-Specific BLAST(RPS-BLAST)



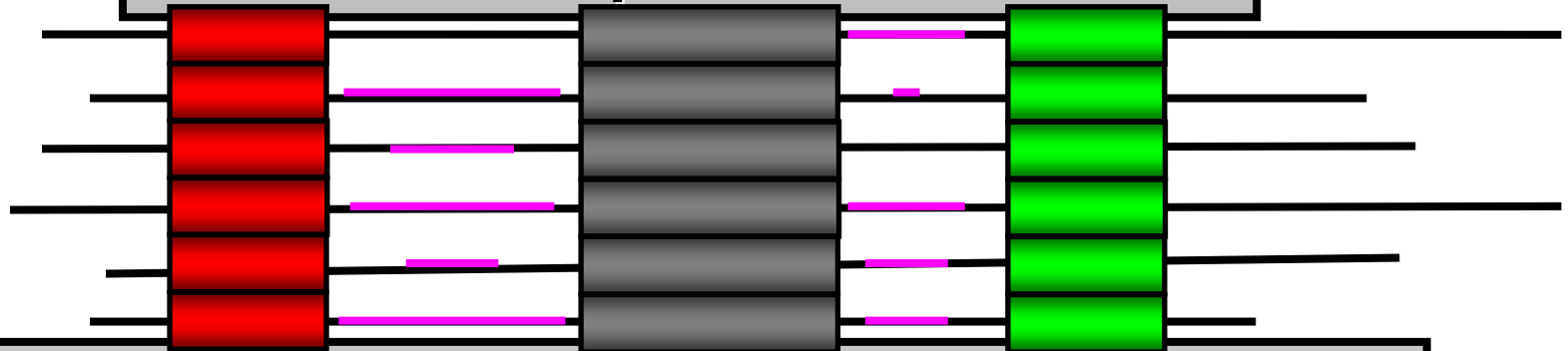
The role of the PSSM has changed from being the “query” in PSI-BLAST to “subject”, hence the term “reverse” in RPS-BLAST

(Reversed-Position Specific)  
**rpsBLAST doesn't incorporate the concept of “block”**

# HMM



**HMMer is trained on the CDD sequences.**



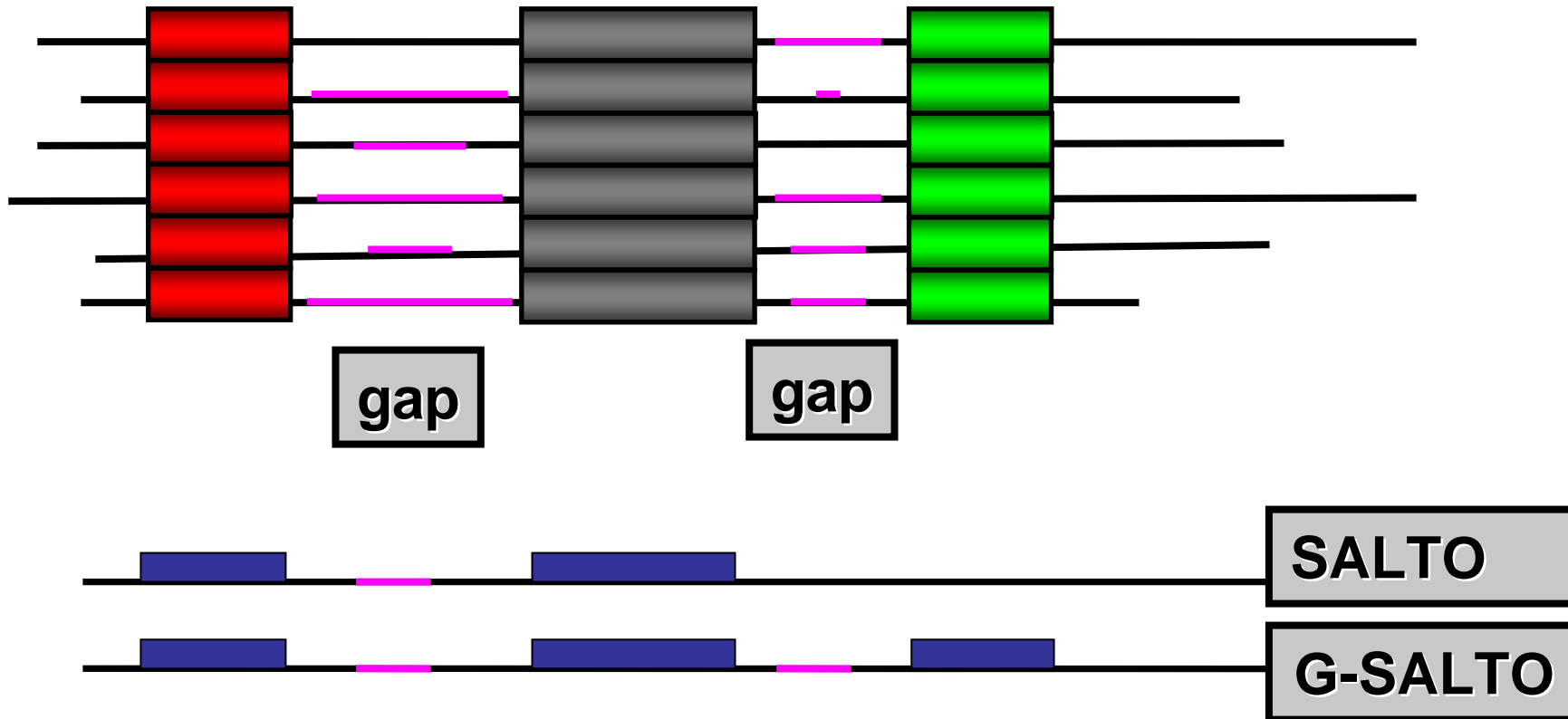
**HMMer does not specifically incorporate the concept of “block”.**

# HMMer's Statistics are a (Poor) Empirical Fit

- HMMer fits the EVD distribution parameters  $\lambda$  and  $K$  to simulated sequences with a Gaussian length distribution.
- HMMer\_semi-global Gumbel E-value approximation is sometimes very inaccurate.

# SALTO

## Structure-based **A**lignment **T**ool



# Properties of an Ideal Alignment Method

- **Semi-global alignment method is intrinsically the right tool for searching for domains within proteins.**
  - **Local alignment methods match only a portion of a domain against a query.**
    - **Reverse Position-Specific BLAST (rpsBLAST)**
- **Screening a database for matches needs to be fast.**
  - **HMMs have no intrinsic heuristics to speed computation.**
  - **The word heuristics in rpsBLAST speed screening and are available for any local alignment method.**
- **Accurate Statistics.**

# GLOBAL (Global Blocks Aligned Locally)

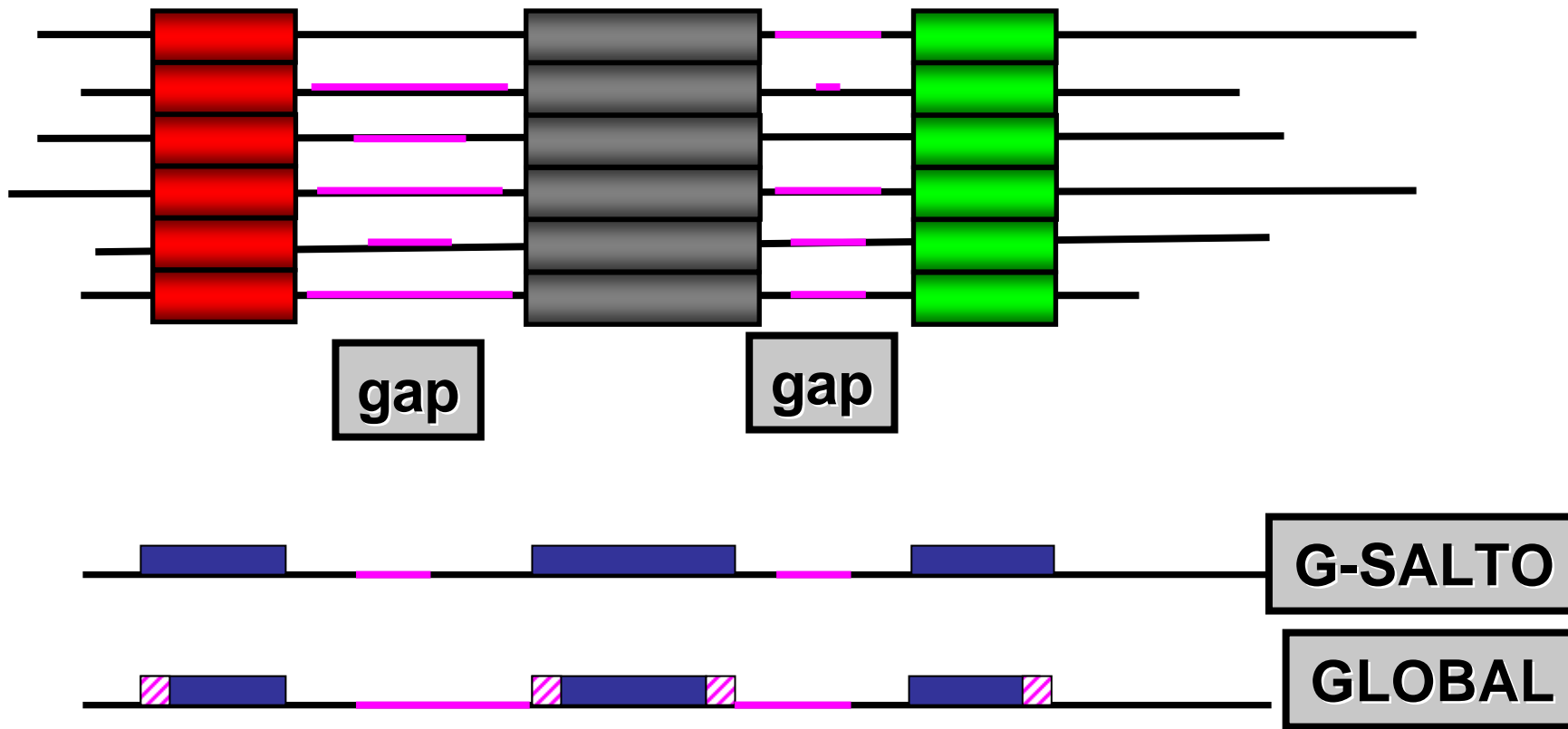
A semi-global Alignment Method for Querying  
A Database of Protein Domains  
with Accurate Statistics

# Outline

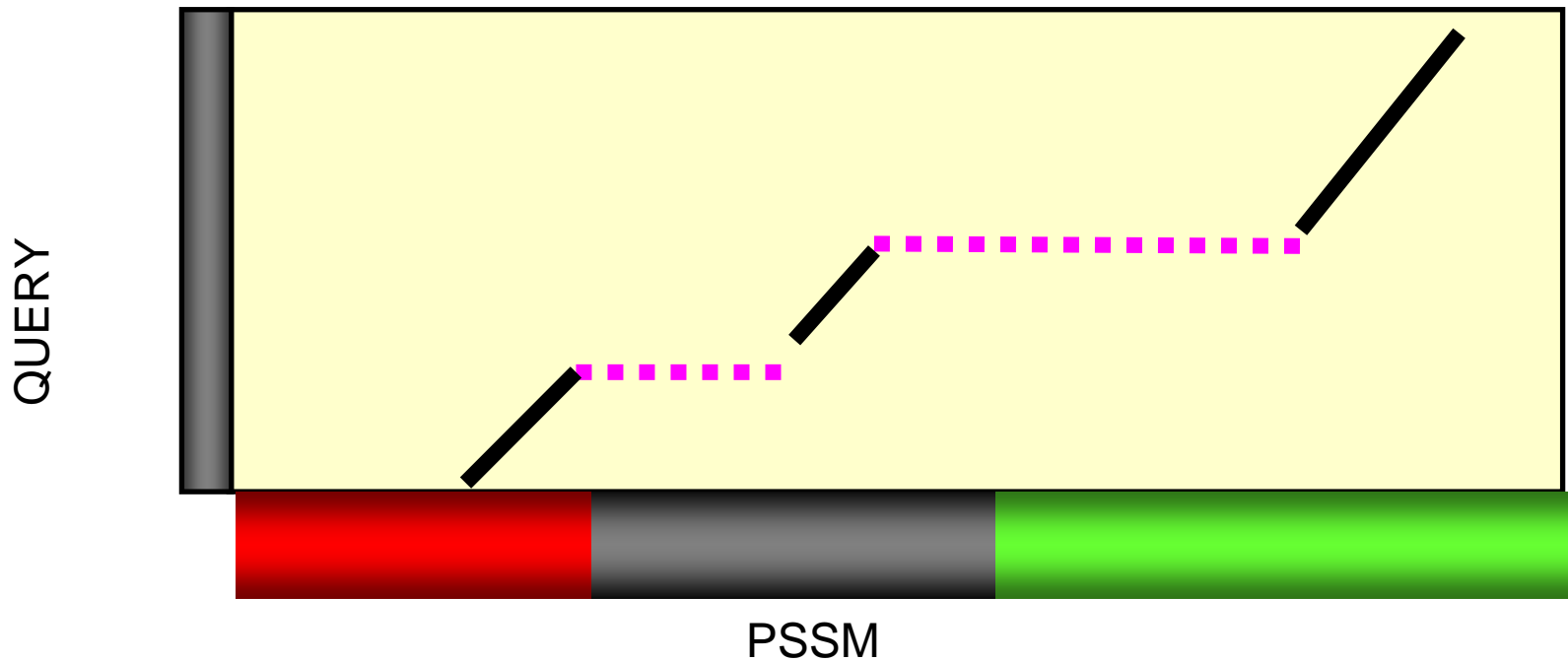
A new computational approach for the detection of protein domains: *Semi-Global Alignment of Protein Domains with accurate statistics.*

- Introduction
- Current methods for protein domain searches
- Method (Global Blocks Aligned Locally)
  - Algorithm and scoring scheme
  - Derivation of Statistics
- Results

# GLOBAL: aligns blocks locally



# GLOBAL

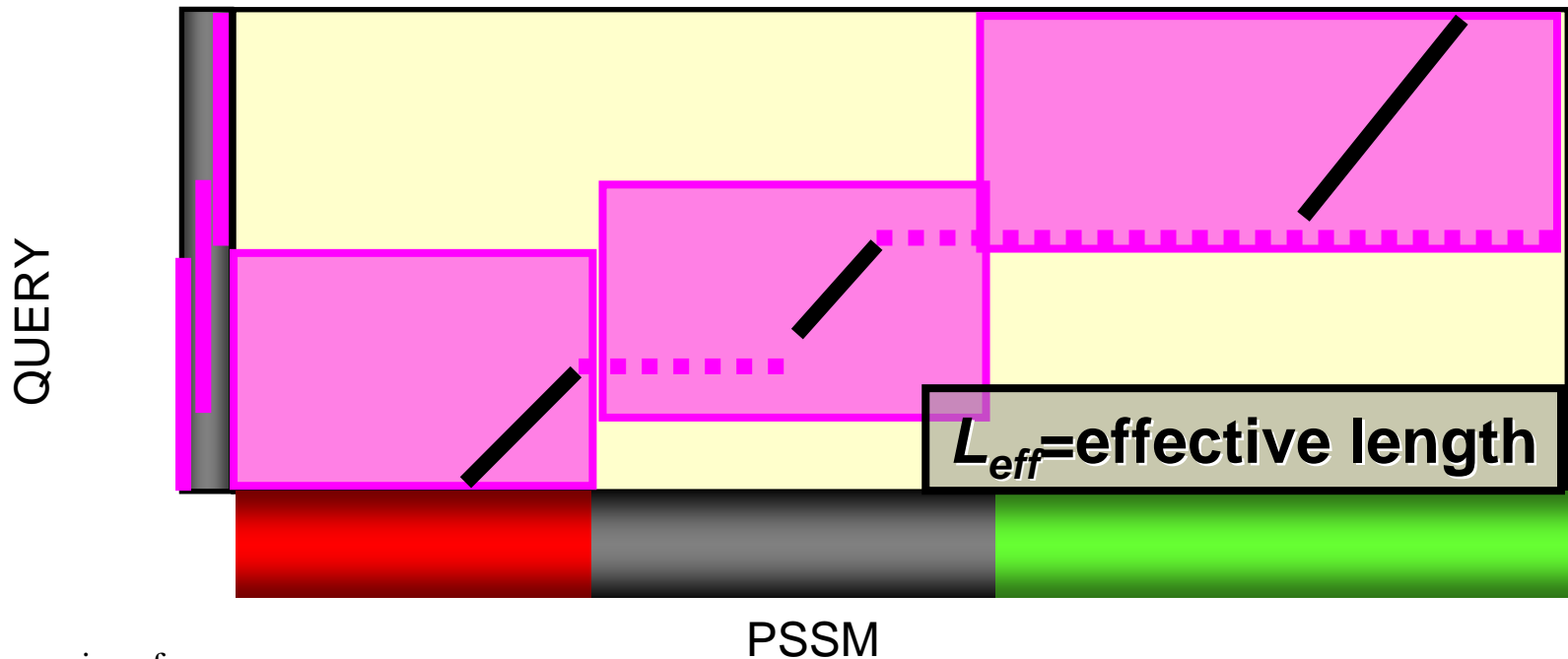


## Global Algorithm:

- Uses dynamic programming (DP) to find the alignment of a protein query sequence to all blocks of the PSSM (in order).
- Penalty=0 both for unaligned regions of the PSSM at the ends of the blocks and unaligned regions of the queries between blocks.

# GLOBAL: statistics for b blocks

For  $b$  blocks, the total alignment score  $T$  is: 
$$T = \sum_{i=1,b} \hat{M}_i(L_{eff})$$



$n$ =size of query

$b$ =number of blocks in the PSSM

Assuming the score for each block is independent of each other, GLOBAL estimates total the alignment score as: 
$$L_{eff} = \left[ \frac{(n+b-1)!}{(n-1)!b!} \right]^{1/b}$$
 e.g.,  $n=160$ ,  $b=3$ ,  $L_{eff}=89$

# Outline

A new computational approach for the detection of protein domains: *Semi-Global Alignment of Protein Domains with accurate statistics.*

- Introduction
- Current methods for protein domain searches
- Method (Global Blocks Aligned Locally)
- Results
  - Benchmarking database
  - ROC (L-ROC) curves
  - P-value Accuracy

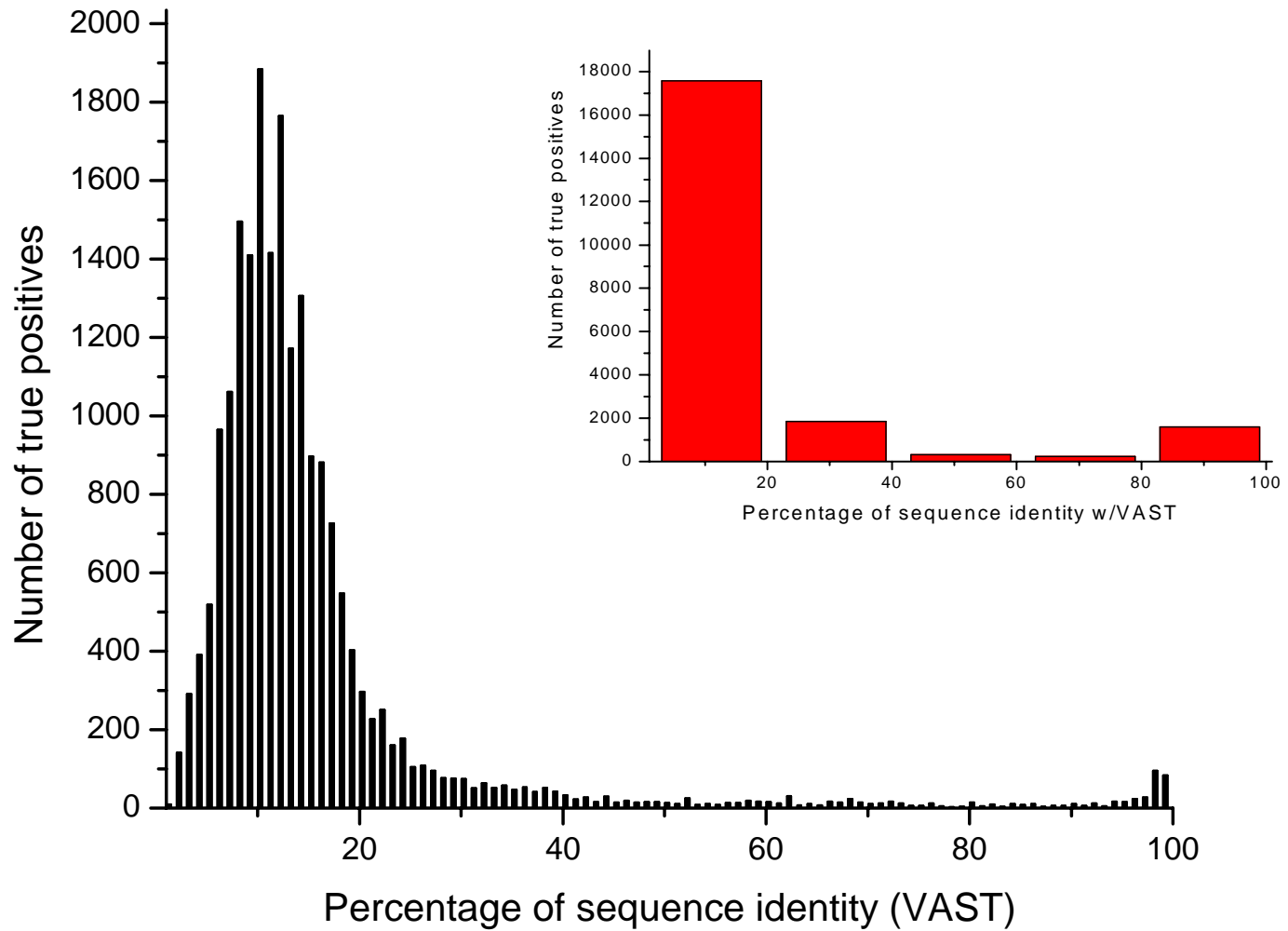
# Benchmarking test set

Database of queries: ~ 10,000 sequences with known structure (from MMDB database).

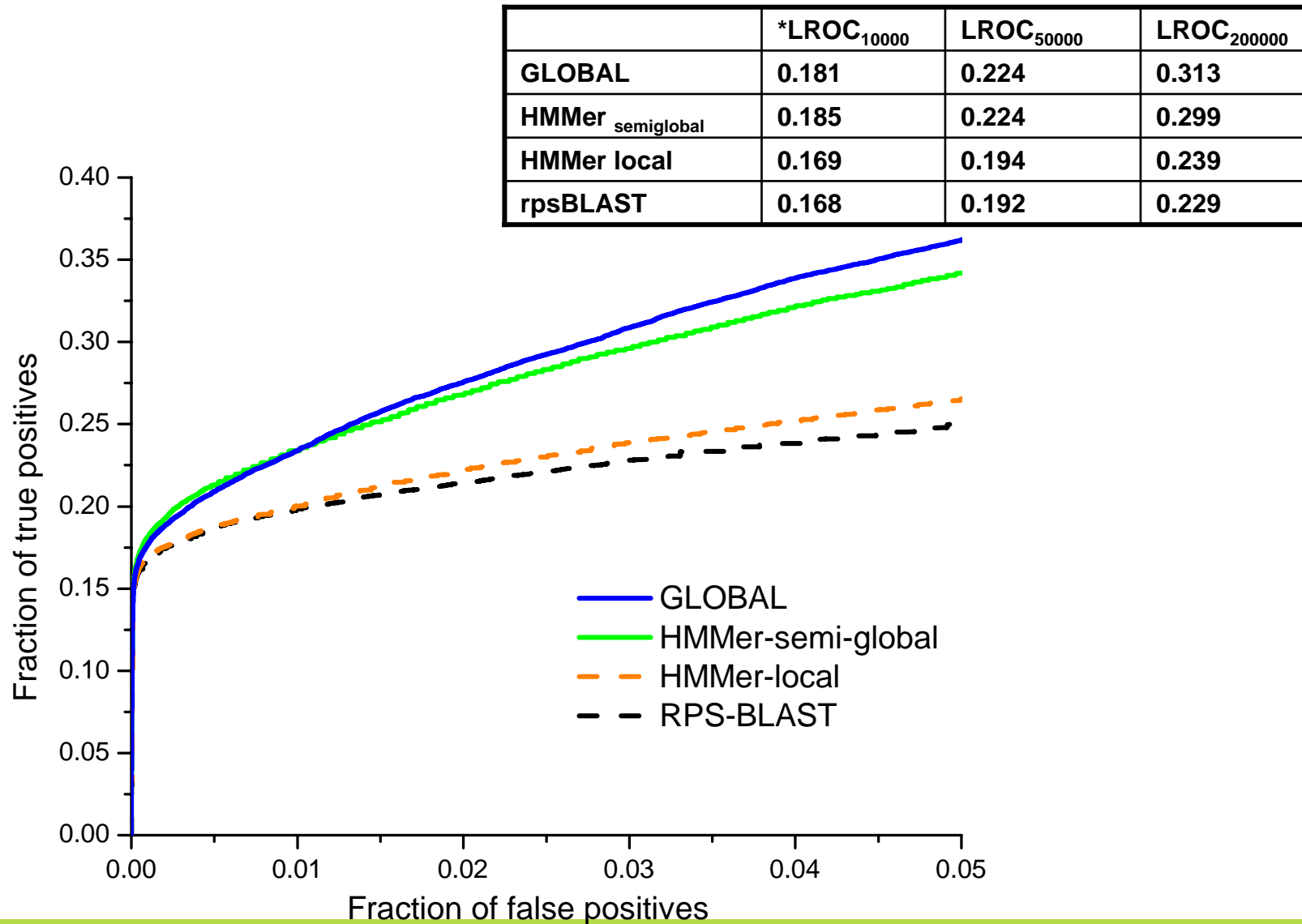
To define true relationships to a CDD entry a query sequence need to be a structure neighbor (using VAST) of a CD's protein from for which the structure is known

The resulting test has >300 families with almost 30,000 known true positives.

# Benchmarking test set

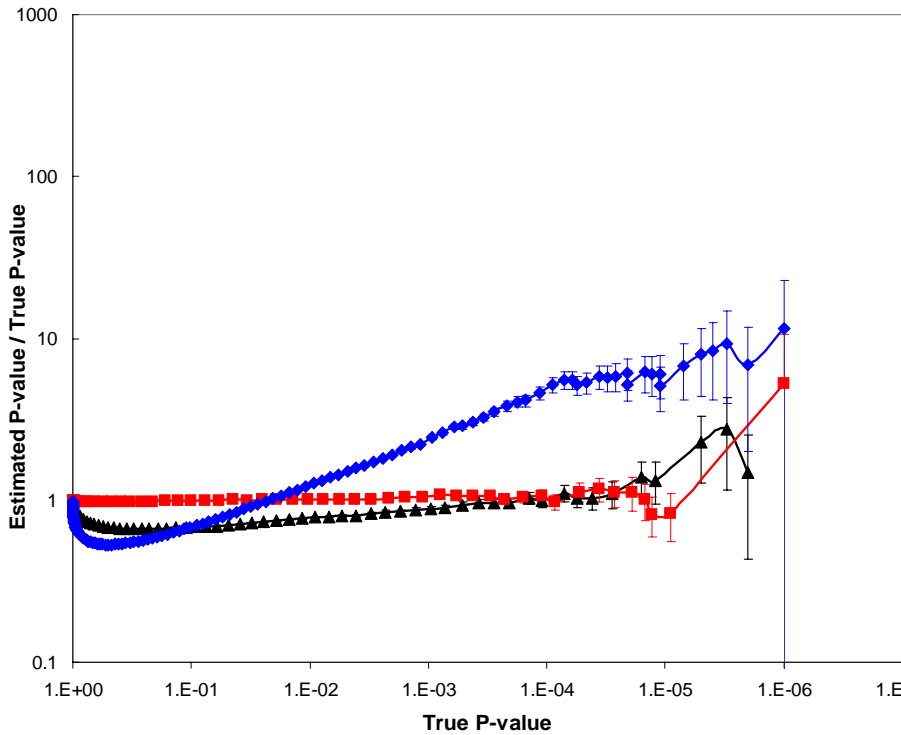


# ROC curve for GLOBAL

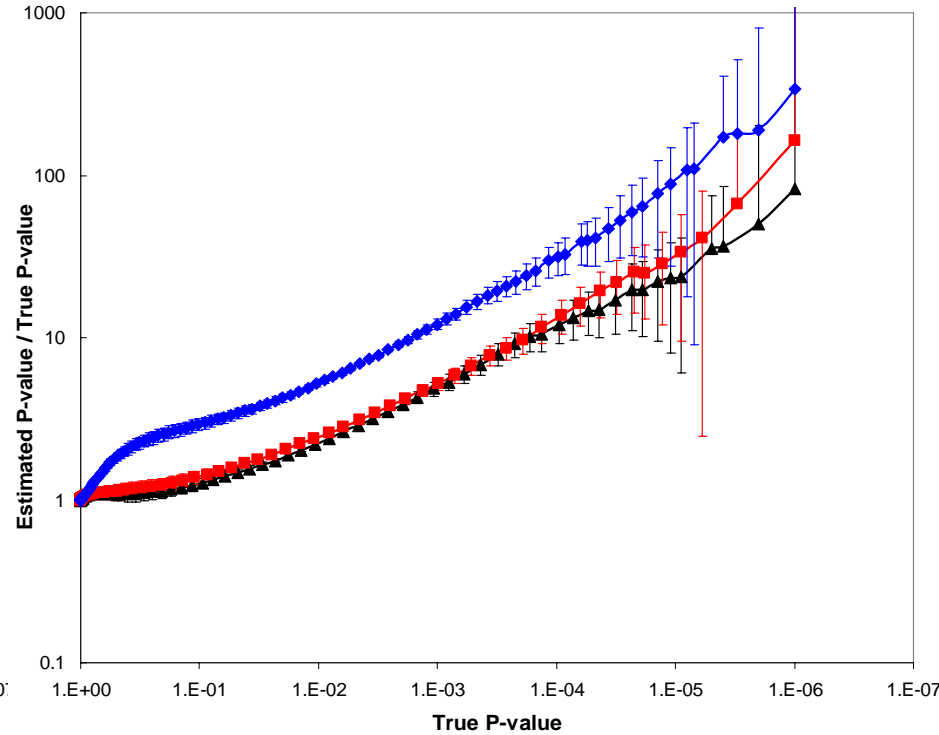


# P-value accuracy

## GLOBAL



## HMMer



- ▲ Cd00030
- Cd00083
- ◆ Cd00288

**1,000,000 simulations using  
random sequences of length 350**

# Conclusions

- The GLOBAL algorithm and p-value provides a flexible format for semi-global sequence alignments.
- GLOBAL respect block structure but adds flexibility at the ends of each block.
- The GLOBAL p-value is based on local alignment p-values. BLAST heuristics from local alignment therefore apply to GLOBAL.
- While the overall performance is similar to that of HMMer semi-global, GLOBAL has more accurate statistics and the possibility to implement heuristics similar to those used in local methods could make it orders of magnitude faster.

# Future work

- Implementation of GLOBAL:
  - “Blockalizer”: creates blocks within the MSA.
  - Heuristics to increase the speed.
- Optimization of domain discovery: Can we mix and match methods/CDs?

# Acknowledgments

## **SALTO:**

- Stephen Altschul, Anna Panchenko, Paul Thiessen and Steve Bryant.

## **GLOBAL:**

- John Spouge, Sergey Sheetlin and Yonil Park.

## **PROTEIN INTERACTIONS:**

*Predicting protein-protein interaction by searching evolutionary tree automorphism space*

- Teresa Przytycka and Raja Jothi.

*Predicting protein domain interactions from co-evolution of conserved regions: Teresa Przytycka, Praveen Cherukuri and Raja Jothi.*

- **UMBC Computational Biology lab team.**

# Kann's Computational Biology lab.

