

例題 3.4

1. PROSITE でモチーフがどのように記述されているかを、「**zinc finger C2H2-type domain signature and profile**」 ([PDOC00028](#)) を例にとって確認し、正規表現とプロフィールの表現法を比較せよ。
2. PROSITE でチロシンキナーゼ SRC の配列 (UniProtKB: [SRC HUMAN](#)) をクエリとして実行し、SH2、SH3、プロテインキナーゼの 3 つのドメインをもつことを確認せよ。また、プロテインキナーゼドメインの中で複数のモチーフがヒットすることを確認せよ。

解答

1. PROSITE のトップページの「Search」で「zinc finger」を入力する。

Expasy PROSITE Home | Contact

Home | ScanProsite | ProRule | Documents | Downloads | Links | Funding

prosite Database of protein domains, families and functional sites

New! SARS-CoV-2 relevant PROSITE motifs

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More... / References / Commercial users]. PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More...].

Release 2021_04 of 29-Sep-2021 contains 1895 documentation entries, 1311 patterns, 1326 profiles and 1338 ProRule.

Search

zinc finger e.g. PDOC00022, PS50089, SH3, zinc finger
Search

「zinc finger」と入力して、「Search」ボタンを押す

Browse

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hits

Quick Scan mode of ScanProsite

Quickly find matches of your protein sequences to PROSITE signatures (max. 10 sequences). [?] [Examples]

Enter UniProtKB accessions or identifiers or PDB identifiers or sequences in FASTA format

Scan Clear

Other tools

- PRATT - allows to interactively generate conserved patterns from a series of unaligned proteins.
- MyDomains - Image Creator - allows to generate custom domain figures.

Custom Images of Domains

さらに zinc finger C2H2 を検索

- PDOC51015 YDG domain profile
- PDOC51792 Yippee domain profile
- PDOC00883 YrdC-like domain profile
- PDOC51915 ZAD domain profile
- PDOC51036 Zinc finger A20-type profile
- PDOC51039 Zinc finger AN1-type profile
- PDOC50119 Zinc finger B-box-type profile
- PDOC50808 Zinc finger BED-type profile
- PDOC51113 Zinc finger Btk-type profile
- PDOC51799 Zinc finger C2H2 AKAP95-type profile
- **PDOC00028 Zinc finger C2H2-type domain signature and profile**
- PDOC51804 Zinc finger C2HC LYAR-type profile
- PDOC51803 Zinc finger C2HC RNF-type profile
- PDOC51807 Zinc finger C2HC baculovirus (BV)-type profile
- PDOC50103 Zinc finger C3H1-type profile
- PDOC51896 Zinc finger C4H2-type profile
- PDOC51810 Zinc finger CCHC FOG-type profile
- PDOC51811 Zinc finger CCHC HIVEP-type profile
- PDOC51801 Zinc finger CCHC NOA-type profile
- PDOC50158 Zinc finger CCHC-type profile
- PDOC51802 Zinc finger CCHC-type profile
- PDOC51800 Zinc finger CHHC U11-48K-type profile
- PDOC51266 Zinc finger CHY-type and CTCHY-type profiles
- PDOC51050 Zinc finger CW-type profile
- PDOC51058 Zinc finger CXXC-type profile
- PDOC51501 Zinc finger DNL-type profile
- PDOC51265 Zinc finger Dbf4-type profile

「zinc finger C2H2-type domain signature and profile」を選択

PDOC00028 のドキュメンテーションエントリは



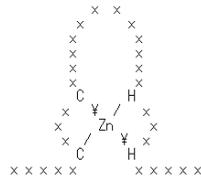
PROSITE documentation PDOC00028

Zinc finger C2H2-type domain signature and profile

Description Technical section References Copyright Miscellaneous

Description

'Zinc finger' domains [1,2,3,4,5] are nucleic acid-binding protein structures first identified in the *Xenopus* transcription factor TFIIIA. These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino-acid residues. There are two cysteine or histidine residues at both extremities of the domain, which are involved in the tetrahedral coordination of a zinc atom. It has been proposed that such a domain interacts with about five nucleotides. A schematic representation of a zinc finger domain is shown below:



Many classes of zinc fingers are characterized according to the number and positions of the histidine and cysteine residues involved in the zinc atom coordination. In the first class to be characterized, called C2H2, the first pair of zinc coordinating residues are cysteines, while the second pair are histidines. A number of experimental reports have demonstrated the zinc-dependent DNA or RNA binding property of some members of this class.

その下を見ると

Description Technical section References Copyright Miscellaneous

ZINC_FINGER_C2H2_2, PS50157; Zinc finger C2H2 type domain profile (MATRIX)

- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 2237
 - detected by PS50157: 2018 (true positives)
 - undetected by PS50157: 219 (219 false negatives and 0 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS50157: 9 false positives and 4 unknowns.
- Domain architecture view of Swiss-Prot proteins matching PS50157

プロフィール (MATRIX)

- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits: Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50157
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS50157
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS50157
- View ligand binding statistics of PS50157
- Matching PDB structures: 1A1F 1A1G 1A1H 1A1I ... [ALL]

ZINC_FINGER_C2H2_1, PS00028; Zinc finger C2H2 type domain signature (PATTERN)

- Consensus pattern: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H
The 2 C's and the 2 H's are zinc ligands
- Sequences in UniProtKB/Swiss-Prot known to belong to this class: 2241
 - detected by PS00028: 2120 (true positives)
 - undetected by PS00028: 121 (120 false negatives and 1 'partial')
- Other sequence(s) in UniProtKB/Swiss-Prot detected by PS00028: 238 false positives and 14 unknowns.
- Retrieve an alignment of UniProtKB/Swiss-Prot true positive hits: Clustal format, color, condensed view / Clustal format, color / Clustal format, plain text / Fasta format
- Retrieve the sequence logo from the alignment
- Taxonomic distribution of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00028
- Retrieve a list of all UniProtKB (Swiss-Prot + TrEMBL) entries matching PS00028
- Scan UniProtKB (Swiss-Prot and/or TrEMBL) entries against PS00028
- View ligand binding statistics of PS00028

正規表現 (PATTERN)

プロフィール PS50157 のエントリ

Entry: **PS50157**

General information about the entry

Entry name [info]	ZINC_FINGER_C2H2_2
Accession [info]	PS50157
Entry type [info]	MATRIX
Date [info]	01-DEC-2001 CREATED; 07-APR-2021 DATA UPDATE; 29-SEP-2021 INFO UPDATE.
PROSITE Doc. [info]	PDOC00028
Associated ProRule [info]	PRU00042

Name and characterization of the entry

Description [info]	Zinc finger C2H2 type domain profile.
Matrix / Profile [info]	<pre> /GENERAL_SPEC: ALPHABET='ABCDEFGHIJKLMNPQRSTVWYZ'; LENGTH=28; /DISJOINT: DEFINITION-PROTEO1: N1=3; N2=28; /NORMALIZATION: MODE=1; FUNCTION=LINEAR; R1=-0.6889000; R2=0.0207831; TEXT='NScore'; /NORMALIZATION: MODE=-1; FUNCTION=LINEAR; R1=1303.8804932; R2=0.9673913; PRIORITY=1; TEXT='Heuristic 5.0K'; /OUT_OFF: LEVEL=0; SCORE=442; H_SCORE=1737; N_SCORE=8.5; MODE=1; TEXT=''; /OUT_OFF: LEVEL=-1; SCORE=345; H_SCORE=1844; N_SCORE=8.5; MODE=1; TEXT=''; /DEFAULT: D=-20; I=-20; B1=-50; E1=-50; M1=-105; MD=-105; IM=-105; DM=-105; ... </pre>

[» more](#)

モチーフの配列パターン (プロファイル)

Post-processing [info]

正規表現 (PS00028) のエントリ

Entry: **PS00028**

General information about the entry

Entry name [info]	ZINC_FINGER_C2H2_1
Accession [info]	PS00028
Entry type [info]	PATTERN
Date [info]	01-APR-1990 CREATED; 01-JUN-1994 DATA UPDATE; 29-SEP-2021 INFO UPDATE.
PROSITE Doc. [info]	PDOC00028

Name and characterization of the entry

Description [info]	Zinc finger C2H2 type domain signature.
Pattern [info]	<code>C-x(2,4)-O-x(3)-[LIVMFYWQ]-x(8)-H-x(3,5)-H.</code>

モチーフの配列パターン (正規表現)

Numerical results [info]

Numerical results for UniProtKB/Swiss-Prot release **2021_04** which contains **565'928** sequence entries.

Total number of hits	13'633 in 2'372 different sequences
Number of true positive hits	13'274 in 2'120 different sequences
Number of 'unknown' hits	49 in 14 different sequences
Number of false positive hits	310 in 238 different sequences

プロファイルは、人間が見てわかりにくい、パターンを柔軟に表すことができる、ファミリーやドメインのような広範囲の類似性を表すのに適しているという利点があり、正規表現は、プロファイルに比べて人間が見てわかりやすい、簡潔にパターンを表すことができるという利点がある。なお、

PROSITE の「プロファイル」は、挿入や欠失の表現を可能にするなどの一般化が行われている。

2. TYRO3_HUMAN の FASTA 形式の配列は以下の通り

```
>sp|P12931|SRC_HUMAN Proto-oncogene tyrosine-protein kinase Src OS=Homo sapiens OX=9606
GN=SRC PE=1 SV=3
MGSNKSKPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPASADGHRGPSAAFAPAAAE
PKLFGGFNSSDTVTSPQRAGPLAGGVTTFFVALYDYESRTEITDLSFKKGERLQIVNTEGD
WVLAHSLSTGQTGYIPSNYVAPSDSIQAEWYFGKITRRESERLLLNAENPRGTFVRES
ETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSSGGFYITSRTOFNSLQQLVAYYSKHADGL
CHRLTTVCPTSKPQTQGLAKDAWEIPRESLRLEVKLGQGCFFGEVVMGTWNGTTRVAIKTL
KPGTMSPEAFLEAQVMKKLRHEKLVQLYAVVSEEPYIVTEYMSKGSLLDFLKGETGKY
LRLPQLVDMAAQIASGMAYVERMNYVHRDLRAANILVGENLVCKVADFLARLIEDNEYT
ARQGAKFPIKWTAPEAALYGRFTIKSDVWSFGILLTELTTRGRVYPYPMVNRVLDQVER
GYRMPCCPECPESLHDLMCQWRKEPEERPTFEYLQAFLEDYFTSTEPQYQPGENL
```

PROSITE の「Quick Scan mode of ScanProsite」に入力すると

The screenshot shows the PROSITE website interface. At the top, there is a navigation bar with 'ExPASy' logo, 'PROSITE', and links for 'Home' and 'Contact'. Below this, there are links for 'Home', 'ScanProsite', 'ProRule', 'Documents', 'Downloads', 'Links', and 'Funding'. The main heading is 'prosite Database of protein domains, families and functional sites'. A 'New!' badge highlights 'SARS-CoV-2 relevant PROSITE motifs'. A brief description of PROSITE and ProRule is provided, along with a release note for 2021_04. The interface is divided into four main sections: 'Search' (with a search box and example text), 'Browse' (with a list of search criteria), 'Quick Scan mode of ScanProsite' (with a text area for protein sequences and a 'Scan' button), and 'Other tools' (with links to PRATT and MyDomains - Image Creator).

以下のような結果を得る。

PS50002 SH3 Src homology 3 (SH3) domain profile :

84 - 145: score = 29.855
06VTFVALVDYERSTETDLSFKKGERLQI VNMTEGDWLLAHSLSLSTGOTGYI PSMYVAPS
DS
Predicted feature:
DOMAIN 84 145 SH3 [condition: none]

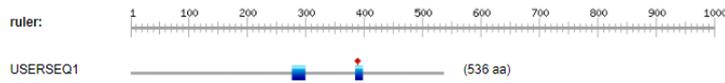
PS50001 SH2 Src homology 2 (SH2) domain profile :

151 - 248: score = 26.969
WVFGKI TRRESERLLLNAENPRTGFLVRESEETTKGAYCLSVSDFDnAKG INVKHYKIRKL
DSGGFYITSRTOFNSLQDLVAYYSKHADGLCHRLTVC
Predicted feature:
DOMAIN 151 248 SH2 [condition: none]

PS50011 PROTEIN_KINASE_DOM Protein kinase domain profile :

270 - 523: score = 43.425
LRLVWKLGGGCFGEVIMGTIP--NGSTRVAI KTLKP---GTMSPFAFLQEAQVMKLRHEK
LVQLVAVVSE-EPTIYI VTEYMSGSLDLFLKGETGKYLRLPOLVDMAAQI ASGMAYVERM
NYVHRDLRAANI LVGENLVCKVADFLARLIEDNEYTRAGAKFPI KWTAPAEAL-YGRF
TIKSDVWSFGILLTELTKGRVDPYP-MVNRVLDQVVERGYRMPCCP---ECPESLHOLM
CQWRKEPEERTFEYLGAFTIEDYF
Predicted features:
DOMAIN 270 523 Protein kinase [condition: none]
NP_BIND 276 284 ATP [condition: <30=K> or <Feature:PS00107>]
BINDING 298 ATP [condition: <30=K>]
ACT_SITE 389 Proton acceptor [condition: D and <FTTag:ATP>]

hits by patterns: [2 hits (by 2 distinct patterns) on 1 sequence]



PS00107 PROTEIN_KINASE_ATP Protein kinases ATP-binding region signature :

276 - 298: [confidence level: (0)] LGGGCFGEVImgtwngttr.....VAIK

PS00109 PROTEIN_KINASE_TYR Tyrosine protein kinases specific active-site signature :

385 - 397: [confidence level: (0)] YVHRDLRAANILV

prosite ScanProsite Results Viewer

Output format: Graphical view - this view shows ScanProsite results together with ProRule-based predicted intra-domain features [help].

Hits for all PROSITE (release 2021_04) motifs on sequence USERSEQ1 :

found: 5 hits in 1 sequence

USERSEQ1 (536 aa)

MGSNKSXPKDASRRRSLPEAENYHGAGGGAPPASQTPSKPASADGHRGPPSAFAPAAAEKLFGG
FNSSDITVTSRORAGPLAGGVTFVALVDYERSTETDLSFKKGERLQI VNMTEGDWLLAHSLSLSTGOT
GYI PSMYVAPSDSI QAEWYFGKI TRRESERLLLNAENPRTGFLVRESEETTKGAYCLSVSDFDnAK
GLNWKHYKIRKLDSSGFYITSRTOFNSLQDLVAYYSKHADGLCHRLTVCPTSKPTGDLAKDnWE
IPRESLRLVWKLGGGCFGEVIMGTIP--NGSTRVAI KTLKP---GTMSPFAFLQEAQVMKLRHEKLVQLV
VVSEEPYI VTEYMSGSLDLFLKGETGKYLRLPOLVDMAAQI ASGMAYVERMNYVHRDLRAANI LV
VGENLVCKVADFLARLIEDNEYTRAGAKFPI KWTAPAEALYGRFTIKSDVWSFGILLTELTKG
RVPYDGMVNRVLDQVVERGYRMPCCPESLHOLMCQWRKEPEERTFEYLGAFTIEDYFTSTEP
QVSPGENL

Legend:



Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes are not intended to indicate homology or shared function. For more information about how these graphical representations are constructed, go to https://prosite.expasy.org/mydomains/.

hits by profiles: [3 hits (by 3 distinct profiles) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



プロテインキナーゼドメイン PROTEIN_KINASE_DOM の中に ATP-binding region、Tyrosine protein kinases specific active-site のパターンが存在している。それらと機能との関係を ProRule が規定している。