質問

Q1　ホモロジー検索におけるスコア値（Score）と同一性（Identities）の関係は？

Q2　アライメント情報に表示されるハイフン（-), プラス（+), コロン（:）などの意味を知りたい.

Q3　BLAST ホモロジー検索の結果で, 自分が作成した配列質問式に含めていない XXX や NNN が表示されることがあります. なぜですか?
(低分子領域とは?)

Q4　tBLASTn, tBLASTx, BLASTx を実行* すると, 結果のアライメント情報に Frame 1, Frame 2 などが表示されます. この Frame の意味は?
また, 一つの回答に一つの Frame しかない場合と Frame 1, Frame 3 など複数の Frame が表示される場合があります. これはなぜですか?

* tBLASTn は REGISTRY ファイル（BLAST), DGENE/PCTGEN/USGENE ファイル（BLAST, GETSIM）で利用可能.
tBLASTx と BLASTx は REGISTRY ファイル（BLAST）のみで利用可能.

Q5　ホモロジー検索でホモロジー検索で曖昧コードを使用した場合と使用しなかった場合のスコア値は同じですか?
例えば, DGENE ファイルの曖昧コード R は A または G を意味しますが, コード R で検索した時の結果は, (コード A の結果 OR コード G の結果）と同じになりますか

Q6　基となる配列（質問式 A）と, その一部のコードを切り取った配列（質問式 B）の部分配列検索の結果は, 下記の関係が成立します.

質問式 A の回答件数 ≦ 質問式 B の回答件数

同じ関係が, ホモロジー検索の結果でも成立しますか?

Q7　環状の核酸・タンパク質を検索する方法は？

参考

下記サイトには, 上記以外の DGENE/PCTGEN/USGENE ファイルの BLAST および GETSIM ホモロジー検索に関する多くの質問が記載されています（英語資料).
ぜひご覧ください.

http://www.stn-international.com/fileadmin/be_user/STN/pdf/search_materials/Bioseq uence_Searching/dgenefaq.pdf

Q1　ホモロジー検索におけるスコア値（Score）と同一性（Identities）の関係は？

A1　ホモロジー検索では，質問式配列と回答配列が一致した局所領域について類似性を，スコア値および同一性パーセントで表しています．

スコア値（Score）

・ 配列質問式および結果のコードの一致，不一致，類似性により，点数が決まっており，その点数を加算していくことでスコア値が算出されます．

- BLAST：核酸

```
Query: 67 ggagacgtagattcggagg  ← 質問式
          |||||||||||||||| ||
Sbjct: 44 ggagacgtagattcggcgg  ← 回答
```

デフォールト設定の場合
コードが一致　：1 点
コードが不一致　：-3 点

- BLAST：タンパク質

選択した置換行列により，コードの点数が異なります．
下記は，デフォールト BLOSUM 62 におけるマトリックスチャートです．

BLOSUM 62 Matrix (the BLAST default)

* column uses minimum score
BLOSUM Clustered Scoring Matrix in 1/2 Bit Units
Cluster Percentage: = 62
Entropy = 0.6979, Expected = -0.5209

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

```
Query: 22  KPIKLVELINIHV  ← 質問式
           KPIK+VEL  IHV
Sbjct: 565 KPIKIVELGKIHV  ← 回答
```

K-K でー致　：5 点
L+I でファミリーー致　：2 点
I G で不ー致　：-4 点

・ 以上より，ー致しているコード，ファミリーでー致しているコードが多いほど，高いスコア値になります．

同一性パーセント（Identitties）

・ 局所領域において塩基やアミノ酸コードが一致している割合を示します.

```
Identities = 29/30 (96%)
Query:   16 caggttccctggtgcaggcagcgctgactc 45
            ||||||||||||| ||||||||||||||||
Subject: 65 caggttccctggtccaggcagcgctgactc 94
```

局所領域の配列長　　：30
一致しているコード数　：29
⇒ 同一性パーセント　：29/30（96%）

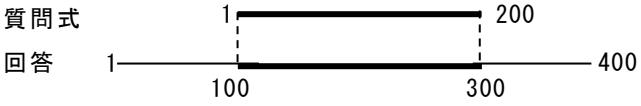
スコア値（Score）と同一性パーセント（Identities)の関係

例）質問式の配列長が 200 である場合
（200 コードが完全に一致した時のスコア値を 400 と仮定し，100 コードが一致した時のスコア値を 300 と仮定します.）

1. 質問式の配列と完全に同じ配列（配列長 200）のレコード

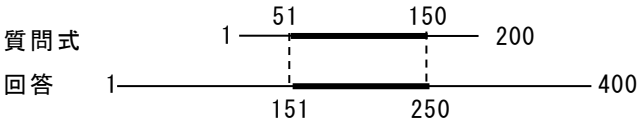
200 コードが完全に一致する
スコア値は 400
同一性パーセント：100%（200/200）

2. 全配列長が 400 で，質問式の配列を完全に含むレコード



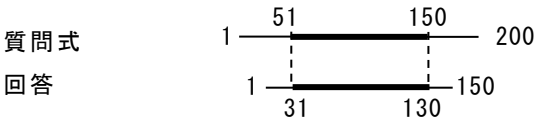
200 コードが完全に一致する
スコア値は 400
同一性パーセント：100%（200/200）

3. 全配列長が 400 で，質問式の一部の配列と完全に一致する配列を有するレコード



100 コードが一致する
スコア値は 300
同一性パーセント：100%（100/100）

4. 全配列長が 150 で，質問式の一部の配列と完全に一致する配列を有するレコード



100 コードが一致する
スコア値は 300
同一性パーセント：100%（100/100）

・ 1. と 2. は回答の配列長に関わらず，局所領域で一致したコードの数が同じなので，スコア値が同じです.（最大スコア値）

・ 3. と 4. は同一性パーセント（局所領域においてコードが一致している割合）が 100% と高いですが，1., 2. に比べて一致したコードが半分のため，スコア値が低くなります.

Q2　アライメント情報に表示されるハイフン（-），プラス（+），コロン（:）などの意味を知りたい．

A2　アライメント情報に表示される記号は，核酸，タンパク質および BLAST，GETSIM ホモロジー検索で異なります．

・ 核酸 – REGISTRY BLAST および DGENE/PCTGEN/USGENE ファイルの BLAST

```
Query: 67 ggagacgtagattcggaggcggccaggcggcg  ← 質問式
          ||||||||||||||| ||||||  |||||||
Sbjct: 44 ggagacgtagattcggcggcggc--ggcggcg  ← 回答
```

|	：一致
空欄	：不一致
-	：ギャップ（GAP）

・ 核酸 – DGENE/PCTGEN/USGENE ファイルの GETSIM

```
29 na overlap starting at 546
aggagugguaggucuuagccagcuguaau  ← 質問式
:: : .  .:  .  .:::::::.  ::.
agtattcatatttactagccagct--aat  ← 回答
```

:	：一致
空欄	：不一致
.	：U と T の一致
-	：ギャップ（GAP）

・ タンパク質 – REGISTRY BLAST および DGENE/PCTGEN/USGENE ファイルの BLAST

```
Query: 22  QVGKGSVSPNAALVAEKISARSGAE
           QV + S+  +AAL        RS  +
Sbjct: 565 QVQQNSLHRDAAL--------RSKLQ
```

コード	：一致
空欄	：不一致
+	：アミノ酸ファミリーの一致
-	：ギャップ（GAP）

・ タンパク質 – DGENE/PCTGEN/USGENE ファイルの GETSIM

```
31 aa overlap starting at 569
qlkkflkialetparicp_inysllasllpk
::.::::.:..::.r. : :. : :.: .::
qlrkflklaiktpvwlnpsitlsslgs_fpk
```

:	：一致
空欄	：不一致
.	：アミノ酸ファミリーの一致
-	：ギャップ（GAP）

ギャップ（GAP）は質問式または回答の配列に挿入されることがあり，ギャップが挿入されると，スコア値は下がります．

Q3　BLAST ホモロジー検索の結果で，自分が作成した配列質問式に含めていない XXX や NNN が表示されることがあります．なぜですか?

A3　REGISTRY BLAST で，Low Complete Filtering にチェックが付いた情報で検索，または DGENE/PCTGEN/USGENE ファイルの BLAST で，デフォールト（オプションで -F F を付与しない）検索すると，配列質問式の低複雑度領域* にマスクフィルタリングが行われます．
フィルタリングが行われるのは，質問式に対してのみで，データベースの配列には行われません．
フィルタプログラムにより見つけられた配列質問式の低複雑度領域は，塩基配列の場合は N，アミノ酸配列の場合は X の文字で置き換えられます．

例）アミノ酸の配列質問式

NLDDNKNTGIFIISARGGIEGLQQKLWTGISIAIAQAAAALEGLRIAATLQGDNQ

```
Query:    NLDDNKNTGIFIISARGGIEGLQQKLWTGXXXXXXXXXXXXXXXXXXXTLQGDNQ
          NLDDNKNTGIFIISARGGIEGLQQKLWTG                   TLQGDNQ
Subject:  NLDDNKNTGIFIISARGGIEGLQQKLWTGISIAIAQAAAALEGLRIAATLQGDNQ
```

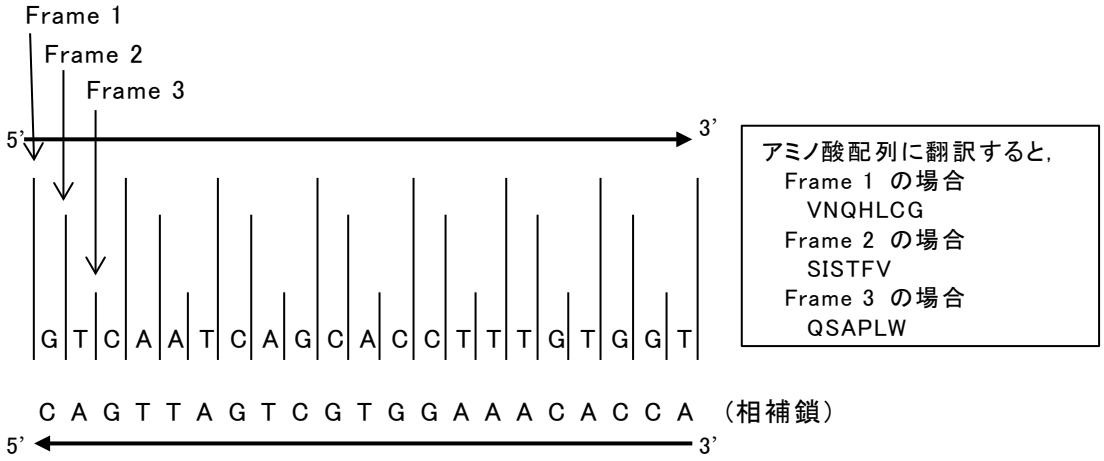
＊ 低複雑領域とは，ホモポリマー領域，短周期リピート，特定残基への偏りなど，偏った組成を持った配列領域のことです．（例 ：proline-rich 領域，poly A tails など）
一般に，低複雑度領域は構造的な偏りが反映され，BLAST プログラムでは非常に高いスコアがつく傾向があります．このような配列は，統計的には有意（スコアが高い）であっても生物学的には類似度は高くないということになります．そのため，フィルタリングを行うことにより，このような低複雑度領域に対する一致など，有意でない一致を結果から除くことができます．

Q4-1 tBLASTn, tBLASTx, BLASTx を実行* すると, 結果のアライメント情報に Frame 1, Frame 2 などが表示されます. この Frame の意味は?

A4-1 3 つの核酸が 1 つのアミノ酸に翻訳されます.
tBLASTn, ではデータベース中の核酸配列を, tBLASTx, BLASTx では質問式中の核酸配列をアミノ酸配列に翻訳する際に, 3 つの核酸をどこから括り始めるかにより翻訳されるアミノ酸にバリエーションが生じます.
この際の読み枠を Frame (フレーム, 読み枠) と呼びます.

Frame (フレーム, 読み枠)

- Frame をずらせば, 一本の核酸配列から三本のアミノ酸配列を作成できます
  また核酸には相補鎖もあるので, 相補鎖からも, 三本のアミノ酸配列が作成されます.
  よって, 合計 6 通りのアミノ酸配列が作成され, 検索に利用されます.



- Universal Genetic Code Table (遺伝子コード表): 核酸をアミノ酸に翻訳する際に使用される表

Symbol	3-letter	Codons[1]
A	Ala	GCT GCC GCA GCG
B	...	...
C	Cys	TGT TGC
D	Asp	GAT GAC
E	Glu	GAA GAG
F	Phe	TTT TTC
G	Gly	GGT GGC GGA GGG
H	His	CAT CAC
I	Ile	ATT ATC ATA
K	Lys	AAA AAG
L	Leu	TTG TTA CTT CTC CTA CTG
M	Met	atg
N	Asn	AAT AAC
P	Pro	CCT CCC CCA CCG
Q	Gln	CAA CAG
R	Arg	CGT CGC CGA CGG AGA AGG
S	Ser	TCT TCC TCA TCG AGT AGC
T	Thr	ACT ACC ACA ACG
V	Val	GTT GTC GTA GTG
W	Trp	TGG
X	Xxx	
Y	Tyr	TAT TAC
Z	...	...
*	STOP	TAA TAG TGA

1: 表中の T は T または U を表現しています

例）下記のアミノ酸配列を REGISTRY ファイルの tBLASTn で検索する.

TVDQHLCGSHLVEALYSVWVHEAKGLPRAAAGAPGVRAELWLDGALLARTAPRAGPG
QLFWAERFHFEALPPARRLSLRLRGLGPGSAVLGRVALALEELDAPRAPAAGLERWF

- ヒットしたホモロジー検索の回答例

```
RN 903919-84-6

Length = 351
Score = 35.0  Expect = 7.1
Identities = 15/17 (88%)  Positives = 16/17 (94%)
Frame = +3
Query:     2 VDQHLCGSHLVEALYSV  18
             V+QHLCGSHLVEALY V
Subject:  93 VNQHLCGSHLVEALYLV 143
```

・ 対応する RN 903919-84-6 の全配列長を REGISTRY ファイルで表示（SQD 表示形式）

```
RN   903919-84-6  REGISTRY
FS   NUCLEIC ACID SEQUENCE
SQL  351
NA   65 a   93 c   100 g   93 t

PATENT ANNOTATIONS (PNTE):
Sequence |Patent
Source   |Reference
=========+==============
Not Given|IN190003
         |unclaimed
         |7

SEQ        1 aattcatggg cctatggatc cgtctactgc ctctgatcgc gctgctgatc
          51 ctctggggac cggatccagc tgcggccgaa ttccggatgt ttgtcaatca
         101 gcacctttgt ggttctcacc tggtggaggc tctgtacctg gtgtgtgggg
         151 aacgtggttt cttctacaca cccaagaccc gtcgtgaagc tgaagacctt
         201 caagtgggtc aaa
         251 acctt
         301 gctgc
         351 g
```

ここから核酸配列の翻訳を開始したので，Frame = +3

Frame 3 から核酸のアミノ酸への翻訳を開始すると，塩基コード 93 番目から 143 番目で翻訳されたアミノ酸（VNQHLCGSHLVEALYLV）が質問式のアミノ酸配列と類似していた

```
**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
```

Q4-2 また，一つの回答に一つの Frame しかない場合と Frame 1, Frame 3 など複数の Frame が表示される場合があります. これはなぜですか?

A4-2 Frame が一つしか表示されない回答は，それ以外の Frame とは類似性が無いことを意味しています.
複数の Frame で類似性が生じた場合は，その類似性のあるヒットした Frame がすべて表示されます.

Q5　ホモロジー検索で曖昧コードを使用した場合と使用しなかった場合のスコア値は同じですか?
例えば, DGENE ファイルの曖昧コード R は A または G を意味しますが, コード R で検索した時の結果は, (コード A の結果 OR コード G の結果) と同じになりますか?

A5　いいえ, スコア値は異なります.
コード R で検索した結果は, (コード A の結果 OR コード G の結果) の結果とは異なります.

理由 : 使用するコードにより, スコア値を算出するための点数が異なるためです.

・ GETSIM の核酸検索のマトリックス

	A	B	C	D	G	H	K	M	N	R	S	T	U	V	W	X	Y
A	5	-2	-4	1	-4	1	-1	2	-1	2	-1	-4	-4	1	2	-1	-1
B	-2	1	1	-1	1	-1	1	-1	-1	-1	1	1	1	-1	-1	-1	1
C	-4	1	5	-2	-4	1	-1	2	-1	-1	2	-4	-4	1	-1	-1	2
D	1	-1	-2	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1
G	-4	1	-4	1	5	-2	2	-1	-1	2	2	-4	-4	1	-1	-1	-1
H	1	-1	1	-1	-2	1	-1	1	-1	-1	-1	1	1	-1	1	-1	1
K	-1	1	-1	1	2	-1	2	-1	-1	1	1	2	2	-1	1	-1	1
M	2	-1	2	-1	-1	1	-1	2	-1	-1	1	-1	-1	1	1	-1	-1
N	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
R	2	-1	-1	1	2	-1	1	-1	-1	2	1	-1	-1	1	1	-1	-2
S	-1	1	2	-1	2	-1	1	1	-1	1	2	-1	-1	1	-1	-1	1
T	-4	1	-4	1	-4	1	2	-1	-1	-1	-1	5	5	-2	2	-1	2
U	-4	1	-4	1	-4	1	2	-1	-1	-1	-1	5	5	-1	2	-1	2
V	1	-1	1	-1	1	-1	-1	1	-1	1	1	-2	-1	1	-1	-1	-1
W	2	-1	-1	1	-1	1	1	1	-1	1	-1	2	2	-1	2	-1	1
X	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
Y	-1	1	2	-1	-1	1	1	-1	-1	-2	1	2	2	-1	1	-1	2

- 質問式中のコード R が回答のコード A または G でヒットした時は 2 点
- 質問式中のコード A が回答のコード A でヒットした時は 5 点
- 質問式中のコード G が回答のコード G でヒットした時は 5 点

点数が異なればスコア値にも影響します

Q6　基となる配列（質問式 A）と，その一部のコードを切り取った配列（質問式 B）の部分配列検索の結果は下記のようになりますが，同じ関係がホモロジー検索の結果でも成立しますか?

質問式 A の回答件数 ≦ 質問式 B の回答件数

A6　いいえ，成立しません．

部分配列検索の結果は，質問式の配列を含むすべての配列が得られますが，ホモロジー検索では，質問式の配列と局所領域において類似した配列が得られます．そのため，通常は配列長の長い質問式 A の回答件数の方が，質問式 B の回答件数より多くなります．

（検証）下記の 2 つのアミノ酸配列を REGISTRY ファイルで検索する．

・ 質問式 A（配列長 250）

KYCLNWRHQS VKLFARSLDR LFGLDHAFSW IHVRLTNSTM YVADPFNPPD
SDACTNLDDN KNTGIFIISA RGGIEGLQQK LWTGISIAIA QAAAALEGLR
IAATLQGDNQ VLAITKEFMT PVPEDVIHEQ LSEAMSRYKR TFTYLNYLMG
HQLKDKETIQ SSDFFVYSKR IFFNGSILSQ CLKNFSKLTT NATTLAENTV
AGCSDISSCI ARCVENGLPK DAAYIQNIIM TRLQLLLDHY YSMHGGINSE

・ 質問式 B（配列長 50）：質問式 A の一部（上記の網掛け部分）の配列

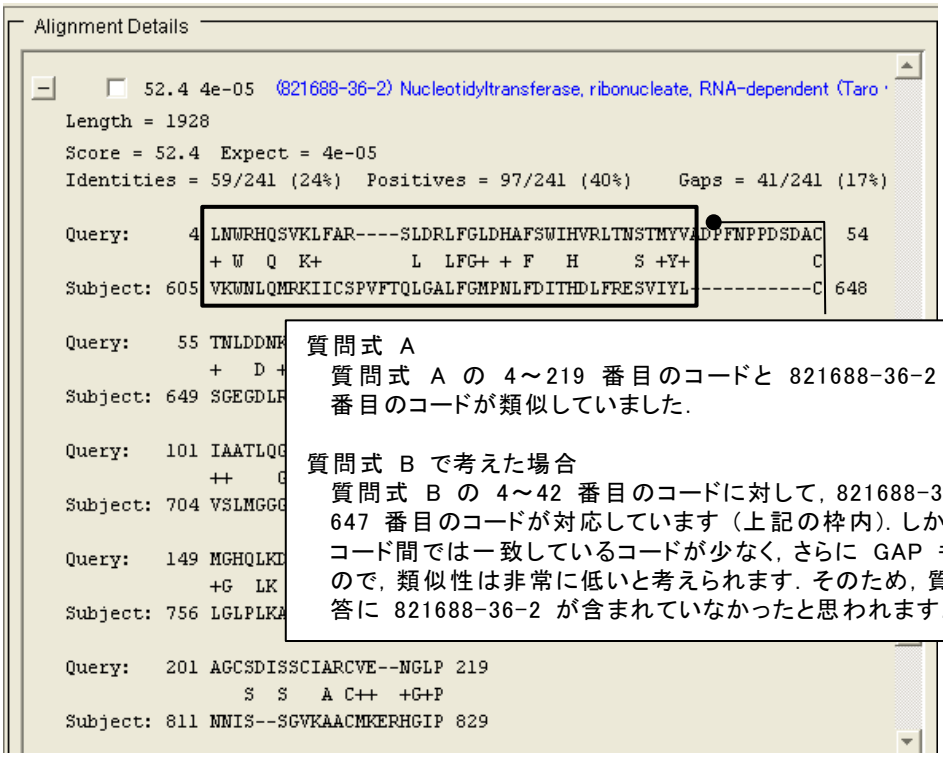
KYCLNWRHQS VKLFARSLDR LFGLDHAFSW IHVRLTNSTM YVADPFNPPD

（検索）

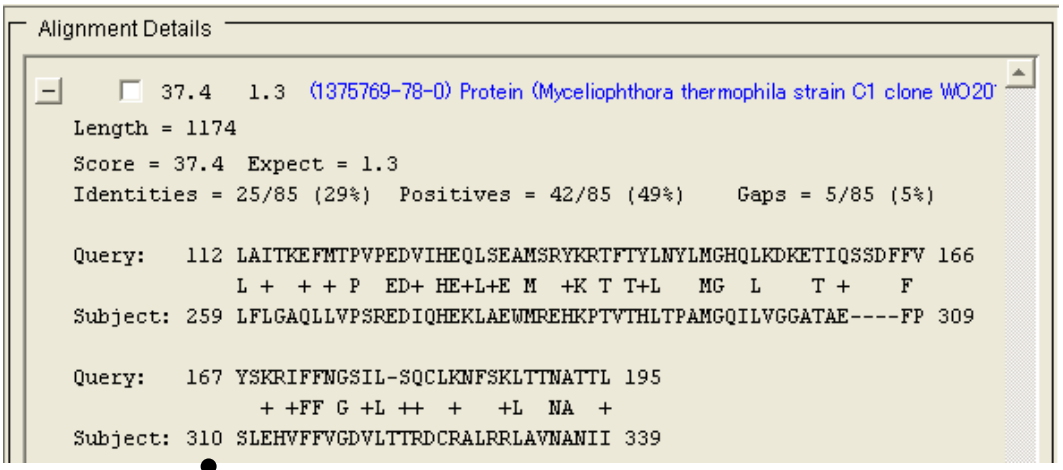
質問式	部分配列検索の結果（件数）	BLAST ホモロジー検索の結果（件数）
A	1	800
B	3	452

BLAST ホモロジー検索の結果で，質問式 A のみで得られた回答例

① CAS 登録番号 821688-36-2



② CAS 登録番号 1375769-78-0



質問式 A の 112～195 番目のコードと 1375769-78-0 の 259～339 番目のコードが類似していました.
つまり, 質問式 A の 1～50 番目のコード (= 質問式 B) に対しては, 1375769-78-0 のコードは類似性がありませんでした.

Q7　環状の核酸・タンパク質を検索する方法は？

A7　REGISTRY ファイル

・ 完全な環状であれば，特別な登録処理がされており，いずれの位置から始めても検索出来るようになっています．

・ NTE フィールドに環化（CYCLIC）と収録されます．

⇩

検索式 : => S コード/SQEP AND CYCLIC/NTE

DGENE/PCTGEN/USGENE ファイル

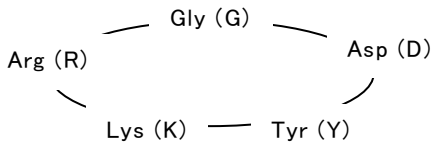
・ ある位置で切断され，鎖状で登録されているため，開始位置を考慮する必要があります．

・ 環化の場合は，FEAT フィールドや標題・抄録中に CYCLIC のキーワードが収録されている場合が多いので，必要に応じて環化のキーワード/FEAT で検索します．

⇩

検索式 : => RUN GETSEQ コード/SQEP AND CYCL?/FEAT（または CYCL?）

<検索例 : 下記の環状ペプチドを調査する>



REGISTRY ファイル

・ 上記の環状ペプチドは完全な環状なので，GDYKR/SQEP，DYKRG/SQEP，YKRGD/SQEP，KRGDY/SQEP RGDYK/SQEP,の何れの検索式でもヒットします．

```
=> FILE REGISTRY

L1          326 S GDYKR/SQEP      ← G から始める
L2        54121 S CYCLIC/NTE      ← 環化で限定するためのキーワード
L3          319 S L1 AND L2
```

（検証）1 個ずつコードをずらして検索します．

```
L4          321 S DYKRG/SQEP      ← D から始める
L5          319 S L4 AND L2       ← 環化されているレコードに限定
L6          324 S YKRGD/SQEP      ← Y から始めます
L7          319 S L6 AND L2
L8          322 S KRGDY/SQEP      ← K から始めます
L9          319 S L8 AND L2
L10         322 S RGDYK/SQEP      ← G から始めます
L11         319 S L10 AND L2
L12         319 S L3 OR L5 OR L7 OR L9 OR L11 ← 何れの場合も結果は同じです
```

=> D L3 SQIDE 1　　← L3 の 1 件目を SQIDE 表示形式で表示

L3　ANSWER 1 OF 319　REGISTRY　COPYRIGHT 2012 ACS on STN
RN　1380037-49-9　REGISTRY
CN　Cyclo[L-arginylglycyl-L-α-aspartyl-D-tyrosyl-N6-(1-oxooctadecyl)-L-lysyl]　(CA INDEX NAME)
FS　PROTEIN SEQUENCE; STEREOSEARCH
SQL　5
NTE　cyclic
　　modified (modifications unspecified)

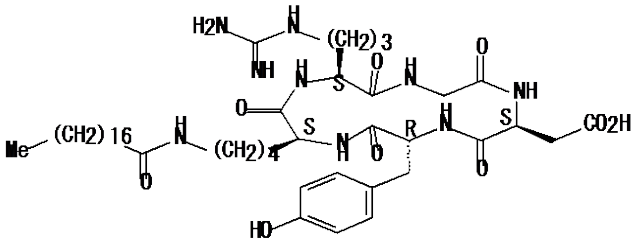
type	location		description
modification	Lys-5	-	undetermined modification

SEQ　　1 RGDYK
　　=====
HITS AT:　1, 2-5

質問式 GDYKR/SQEP と SEQ フィールドの並びは異なっていますが，完全な環化の場合は，開始位置に関わらずヒットします．
HITS AT フィールドでヒット位置が表示されます

RELATED SEQUENCES AVAILABLE WITH SEQLINK
MF　C45 H75 N9 O9
SR　CA
LC　STN Files:　CA, CAPLUS, TOXCENTER
DT.CA　CAplus document type:　Patent
RL.P　Roles from patents:　PREP (Preparation); PROC (Process)

Absolute stereochemistry.



PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

1 REFERENCES IN FILE CA (1907 TO DATE)
1 REFERENCES IN FILE CAPLUS (1907 TO DATE)

参考 : 一部が環状化された配列

REGISTYR ファイルでは，一部が環状化された配列については，特別な登録処理がされていないため，開始位置を考慮する必要があります

DGENE ファイル

- REGISTRY ファイルと異なり，環状の配列はある位置で切断された鎖状の形式で登録されているため，開始位置を 1 個ずつずらして検索するしかありません．

```
=> FILE DGENE

=> RUN GETSEQ GDYKR/SQEP            ← G から始めます
L1            0 GDYKR/SQEP

=> RUN GETSEQ DYKRG/SQEP            ← D から始めます
L2            2 DYKRG/SQEP

=> RUN GETSEQ YKRGD/SQEP            ← Y から始めます
L3            1 YKRGD/SQEP

=> RUN GETSEQ KRGDY/SQEP            ← K から始めます
L4            1 KRGDY/SQEP

=> RUN GETSEQ RGDYK/SQEP            ← R から始めます
L5           64 RGDYK/SQEP

=> S CYCL?/FEAT
L6        26800 CYCL?/FEAT

=> S L2 AND L6
L7            0 L2 AND L6

=> S L3 AND L6
L8            0 L3 AND L6

=> S L4 AND L6
L9            0 L4 AND L6

=> S L5 AND L6
L10          38 L5 AND L6
```

特徴表（FEAT）に環化のキーワードが収録されているレコードに限定します
（特徴表でヒットしない場合は，標題・抄録中に環化があるレコードまで広げます（(CYCL?) で検索））

```
=> D L10 SQIDE 1   ← RGDYK/SQEP でヒットした回答を SQIDE 表示形式で表示

L10   ANSWER 1 OF 38  DGENE  COPYRIGHT 2012 THOMSON REUTERS on STN
AN    AZV31409  peptide       DGENE
AA    0 A; 1 R; 0 N; 1 D; 0 B; 0 C; 0 Q; 0 E; 0 Z; 1 G; 0 H;  0 I; 0 L; 1 K; 0
      M; 0 F; 0 P; 0 S; 0 T; 0 W; 1 Y; 0 V;  0 Others
SQL   5
SEQ
        1 rgdyk
          =====
HITS AT:  1-5

FEATURE TABLE:
Key           |Location|Qualifier|
==============+========+=========+==========================
Modified-site|1       |note     |"This residue is condensed
             |        |         |onto residue 5 to form  a
             |        |         |cyclic peptide"
Modified-site|5       |note     |"This residue is condensed
             |        |         |onto residue 1 to form  a
             |        |         |cyclic peptide"
```

1 位と 5 位のアミノ酸が結合し，環を形成しています

完全な環状配列の調査は，DGENE ファイルよりも REGISTRY ファイルの検索方法が簡単です