

# インターネットセミナー

## 配列検索入門



**JAICI**  
化学情報協会

1

## 本日の内容

1. STN の配列データベース
2. 配列検索の種類
3. 検索例 【DEMO】



**JAICI**  
化学情報協会

2

# 1. STN の配列データベース

STN には、核酸・タンパク質配列を検索できるファイルが 4 つ搭載されています

- REGISTRY ファイル
- DGENE ファイル
- PCTGEN ファイル
- USGENE ファイル

各ファイルの特長を理解しましょう

## REGISTRY ファイル

CAplus/CA ファイルで索引された配列を収録！

製作者	Chemical Abstracts Service (CAS)
収録源	- CAplus/CA ファイルに収録されている ベーシック特許および雑誌論文 - GenBank
収録期間	1957 年 -
更新頻度	毎日
特長	- 雑誌論文に記載された配列情報も収録 - 収録源 (文献情報) を確認したいときは CAplus ファイルへクロスオーバー

## DGENE ファイル

WPI ファイルで索引された配列を収録！

製作者	Thomson Reuters
収録源	WPI ファイルのベーシック特許
収録期間	1981 年 -
更新頻度	隔週
特長	<ul style="list-style-type: none"><li>- 配列に焦点を当てた独自抄録を収録</li><li>- 収録源 (特許) 情報も収録</li></ul>

## PCTGEN ファイル

PCT 出願に記載された配列を収録！

製作者	WIPO, FIZ Karlsruhe
収録源	PCT 出願
収録期間	2001 年 -
更新頻度	毎週
特長	<ul style="list-style-type: none"><li>- 速報性に優れている (最短タイムラグは公報発行後 1 日)</li></ul>

# USGENE ファイル

米国特許に記載された配列を収録！

製作者	SequenceBase Corporation
収録源	INSCD, NCBI/EMBL-EBI, USPTO PSOPS, 米国特許の Sequence Listing
収録期間	1982 年 -
更新頻度	毎週
特長	<ul style="list-style-type: none"><li>- 米国登録特許由来の配列情報も収録</li><li>- 速報性に優れている (タイムラグは公報発行後 3 日以内)</li><li>- 著者抄録と全クレームを収録</li></ul>

## 2. STN の配列検索

STN の配列検索は 3 種類

### ① 完全配列検索

質問式と完全に一致する配列を検索

### ② 部分配列検索

質問式を一部に含む配列を検索

### ③ ホモロジー検索

質問式と類似した配列を検索

## ① 完全配列検索

質問式 GCCCAAGCTGGCATCCGTCA



質問式と完全に一致する配列を検索

回答例 GCCCAAGCTGGCATCCGTCA

## ① 完全配列検索の検索方法

=> 検索コマンド コード/検索タイプ



S (REGISTRY ファイル)

RUN GETSEQ (その他のファイル)



SQEN (核酸)

SQEP (タンパク質)

\* コードの入力ルール

核酸は「5' 末端 → 3' 末端」、タンパク質は「N 末端 → C 末端」の順で入力

## ② 部分配列検索

質問式      GCCCAAGCTGGCATCCGTCA



質問式を一部に含む配列を検索

回答例      UTCGCCCAAGCTGGCATCCGTCAGT

## ② 部分配列検索の検索方法

=> 検索コマンド    コード/検索タイプ



S (REGISTRY ファイル)

RUN GETSEQ (その他のファイル)



SQSN (核酸)

SQSP (タンパク質)

### ③ ホモロジー検索

質問式

GCCCAAGCTGGCATCCGTCA



質問式と類似した配列を検索

回答例

UTCGCCCAAGCTGGUTTCCGTCAGT



### ③ ホモロジー検索の検索方法

#### ◆ REGISTRY ファイル

- 専用ソフトウェアを利用する

#### ◆ DGENE, PCTGEN, USGENE ファイル

- コマンド検索 (RUN コマンド)
- Assistants 機能 (検索補助機能) を利用

### 3. 検索例 【DEMO】

#### ◆ 検索例 1

DGENE ファイルで、核酸配列 ACCGGCCGGT  
を検索する

- ① 完全配列検索
- ② 部分配列検索

それぞれの回答の違いを確認する

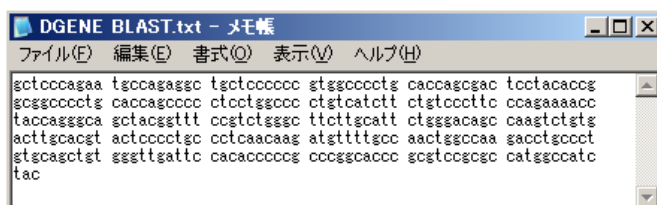


### 3. 検索例 【DEMO】

#### ◆ 検索例 2

DGENE ファイル, REGISTRY ファイルで  
BLAST ホモロジー検索を行う

\* 長い質問式の場合は、事前にテキストファイルを作成しておく





## ホモロジー検索のプログラム

	BLAST	GETSIM
処理速度	速い	遅い (バッチ検索が望ましい)
類似性の高い配列	○	○
類似性の低い配列	△	○
比較方法	短い配列を比較 ギャップはあまり考慮されない	配列全体を比較 ギャップも考慮される
感受性	デフォルト設定では低い	高い
ファイル	REGISTRY DGENE, PCTGEN, USGENE	DGENE, PCTGEN, USGENE

## 参考資料

### ◆ 核酸・タンパク質配列検索法テキスト



定期講習会にぜひご参加ください！

[https://www.jaici.or.jp/seminar/smnr\\_stn.php](https://www.jaici.or.jp/seminar/smnr_stn.php)

### ◆ STN on the Web 配列検索ガイド

[http://www.jaici.or.jp/stn\\_web/webseq.pdf](http://www.jaici.or.jp/stn_web/webseq.pdf)

### ◆ STNews Vol.28 No.2 (2012年3・4月号)

<http://www.jaici.or.jp/newslet/stnews.htm>