

2009年1月27日

ライフサイエンス統合データベースプロジェクト 共通基盤技術開発チーム進捗状況報告 (CBRC)

担当業務: ライフサイエンス統合データベース開発運用
(共通基盤技術開発)

ワークフロー技術を用いた統合DB環境の構築

利用者が得たい情報(知識)を、パソコンなどの端末から要求すると、必要なデータ、解析手法などを、国内、海外から自動的に選び、データベースと解析ツールのワークフローを作成し、最適な計算資源を使って解析を行う統合DB環境を構築する。

CBRC: 浅井、野口、諏訪、
金、ホートン、広川、福井、藤渕、堀本、田代

2008年度の計画と現在までの進捗

・ ワークフローの開発・公開

Phase 1: アミノ酸配列から広範囲な立体構造に関する予測を行うワークフロー (8月末プロジェクト内公開 完了)

Phase 2: アミノ酸配列から予測とデータ取得によりタンパク質に関する網羅的な情報を得るワークフロー (12月末一般公開 完了)

Phase 3: アミノ酸配列からタンパク質の比較情報を提示し、保存部位、変異部位を推定するワークフロー (3月末一般公開 予定)

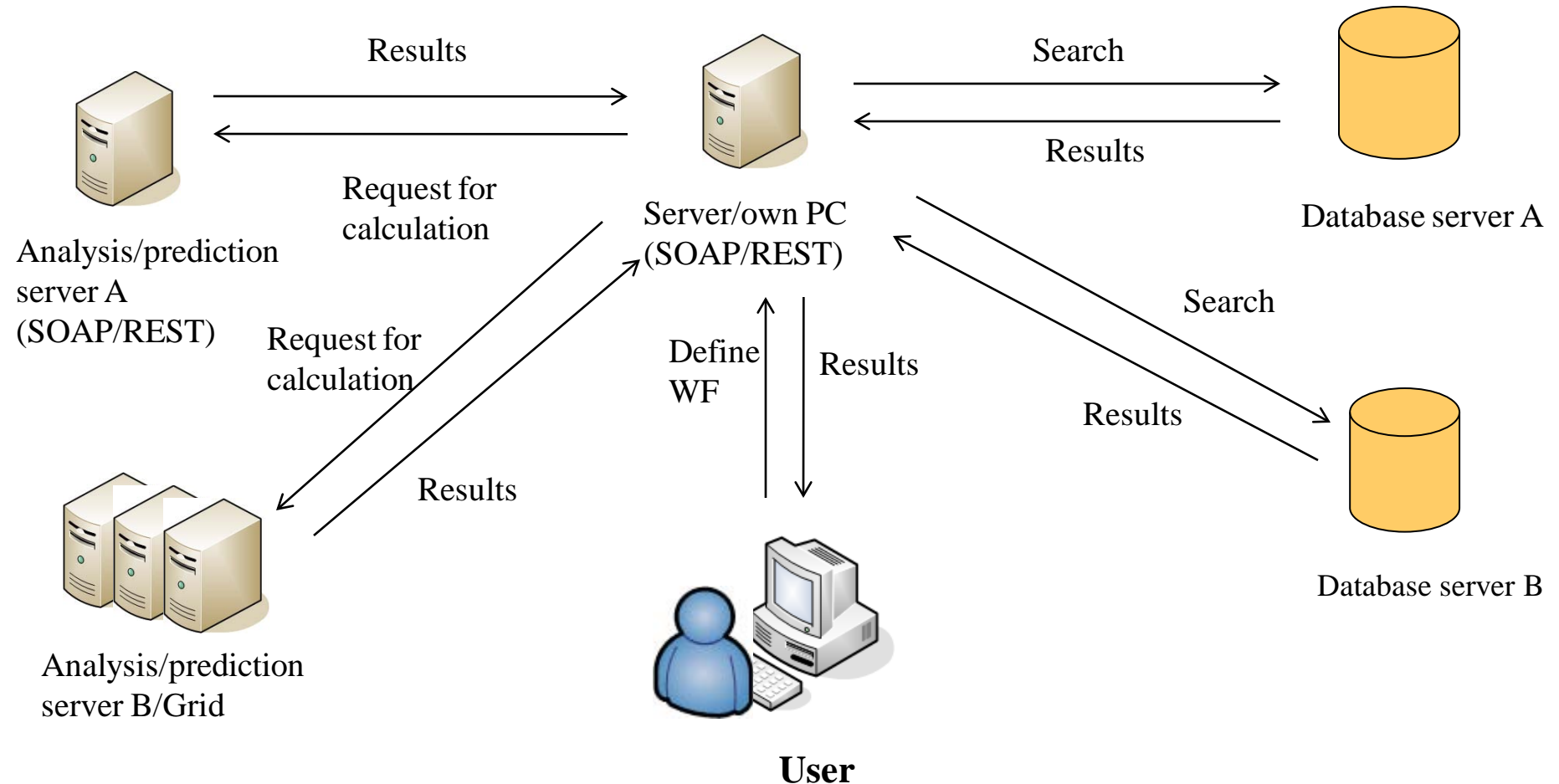
・ その他

ーワークフローの議論

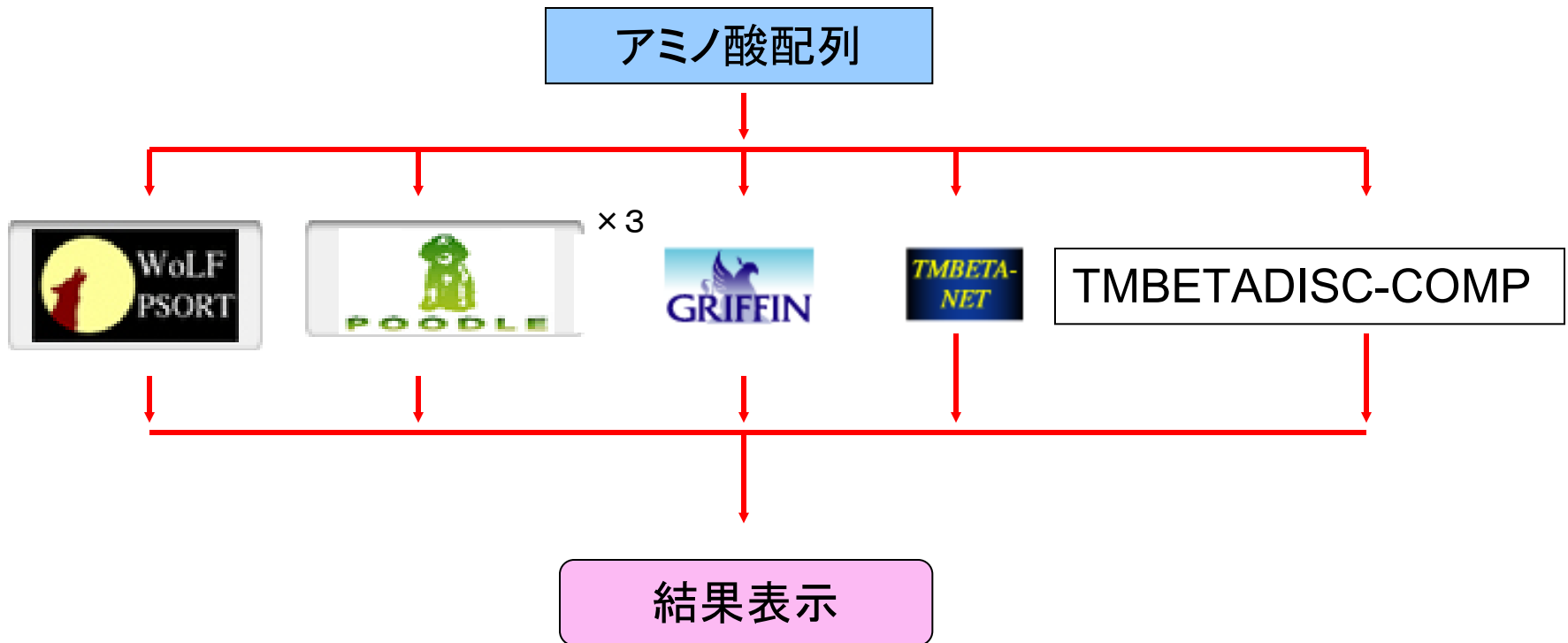
ー利用制限のあるソフトウェアの調査 (Modellerなど)

CBRC統合DBグリッドシステム

- ・計算の自動分散処理(完了)
- ・登録ユーザごとにサービス制限(検討中)



アミノ酸配列から広範囲な立体構造に関する予測を行うワークフロー (Phase 1)



Phase1 ワークフロー(プロジェクト内公開)

初期画面

Recommended browsers: Internet Explorer 6.0 or later/Firefox 2.0 or later/Safari 3.0 or later

Select Web Servers ?

- POODLE-L** [Link](#) **POODLE-S** [Link](#) **POODLE-W** [Link](#)
- TMBETA-NET** [Link](#) **TMBETADISC-COMP** [Link](#)
- GRIFFIN** [Link](#) **WoLF-PSORT** [Link](#)

Enter Your Sequence ?

Disorder Prediction

	10	20	30	40	50	
POODLE-S Confidence Level	PWATAEYDYG 0000000112	AAEDNQLTFV 2222111100	ENDKINIEF 0000000011	VDDDWLQEL 1111111111	EKDGSKGLFP 2222211110	SMY 000
POODLE-L Confidence Level	PWATAEYDYG 2221222211	AAEDNQLTFV 1111111111	ENDKINIEF 1111111100	VDDDWLQEL 0000000000	EKDGSKGLFP 0000000000	SMY 000
POODLE-T Confidence Level	PWATAEYDYG 1111000000	AAEDNQLTFV 0000000000	ENDKINIEF 0000000000	VDDDWLQEL 0000000000	EKDGSKGLFP 0000000000	SMY 011
POODLE-W	0.02482, Ordered					
Final Disorder Prediction Confidence Level	PWATAEYDYG 2222122211	AAEDNQLTFV 1111111111	ENDKINIEF 1111111000	VDDDWLQEL 0000000000	EKDGSKGLFP 0000000000	SMY 000

Colors of Amino Acids

Blue : "Disorder" (Confidence Level >= 5)

Black: "Order" (Confidence Level <= 4)

Confidence Level from 0 (low confidence) to 9 (high confidence)

Disorder 予測結果

生命科学データベース横断検索結果

Query: acetylcholine receptor

Result All (3394)

Score	[19823]
Database	[report]
Document	[公基研究生命システム情報 公基研究生命システム情報 生命システム情報 運動異常を示す変異体の単離と rat fish relative-related mutants have a ryanodine receptor defect, show slow swimming and provide a model of?] The zebrafish emul behavioral mutation disrupts acetylcholine receptor localization and motor axon stability. De]
Link	運動異常を示す変異体の単離と解析
Score	[19046]
Database	[jastart]
Document	[JP 377863 特許公報(B2) 20000310 1999021891 19990112 NF-κB発現する細胞 塩野義興株式会社] 0000 r proteins that activate the K B element in the IL-2 receptor α gene. Science, 241:1952-1955. [046] Bersud, kidney cell clones: epidermal growth factor (EGF) receptor s and the effect of mouse sarcoma virus transform no. M and Numa, S (1994) Expression of functional acetylcholine receptor from cloned cDNAs. Nature, 367:604-608, a signaling proteins for the tumor necrosis factor receptor family. Cell, 80, 389-399, Munger, K., Warness, B.]
Link	日本特許NF-κBを発現する細胞
Score	[17927]
Database	[lnc]
Document	[蛋白質情報誌 43 1 1999 35-41 総題 ムスカリン性アセチルコリン受容体からG-βγのシグナル伝達経路 蛋白質を介した経路 The Pathway Connecting Muscarinic Acetylcholine Receptor s to the Nucleus Involves Small GTP-Bindin 程度してはアタラクチン蛋白質であるGrb2 (growth factor receptor-bound protein 2) およびクアニタスクレオチナ交換因子]
Link	蛋白質情報誌 ムスカリン性アセチルコリン

Thank you for using TMBETA

Your prediction results are written below

Final results after refinements: Transmembrane segments have been shown in red color

1

M N T I S A P P A V S P N I T V L A P Q K D

41

V

81

S

121

R Y F S V T R P L S Y R A K R T P R

161

A P D C Y I D F L S D P I I T

201

V

241

S E R S D P D A E D S P E T P P D R C R C R A P R

281

D E D S M E S L T F S S E G E E P D S

321

S S P N T V K R P T K K D R D R A G K D K P R G K

361

F C K D S V F

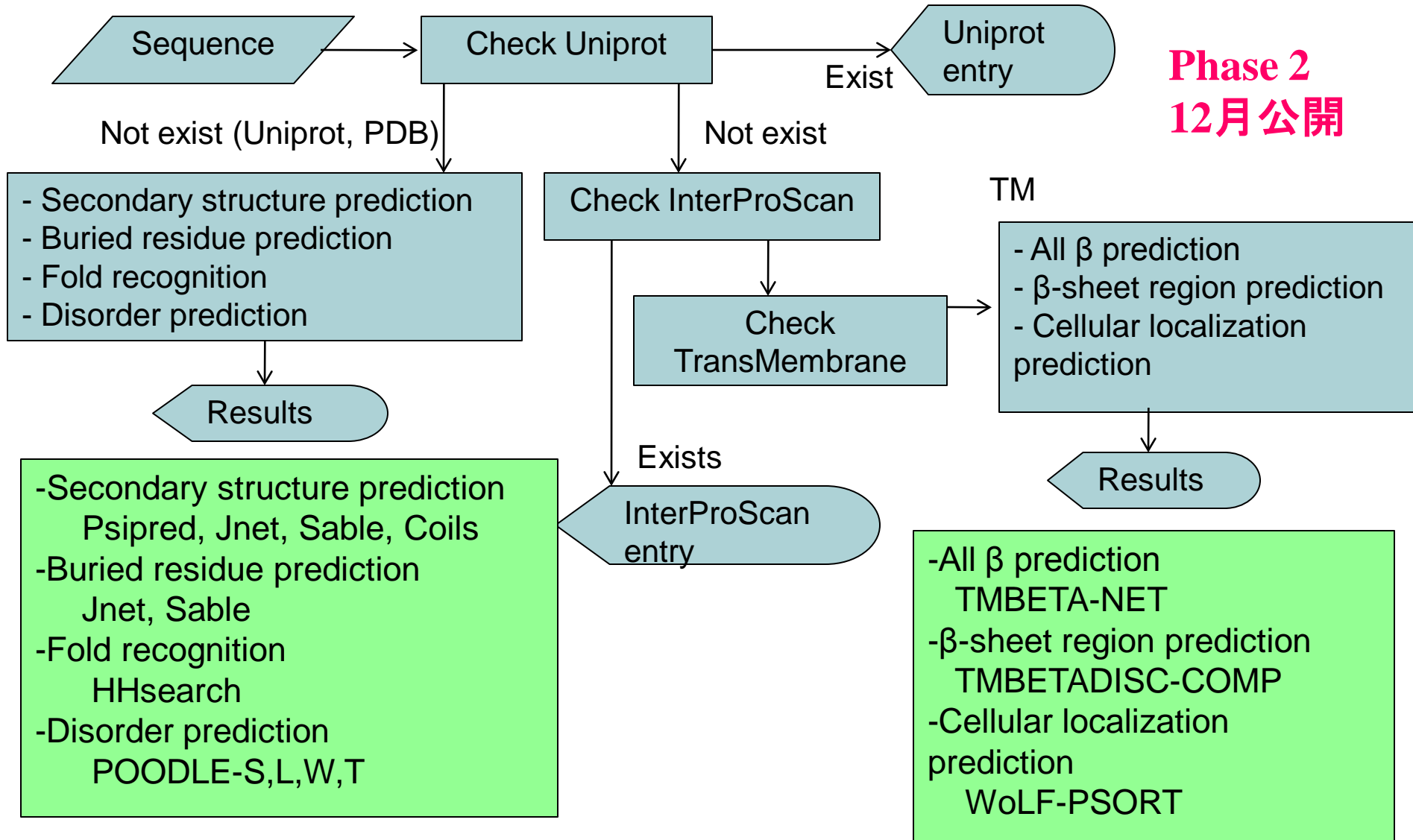
401

Y V N S T I N P M C Y A L G N K A

W D K R

TMBETA-NET 予測結果

アミノ酸配列から予測とデータ取得によりタンパク質に関する網羅的な情報を得るワークフロー



CBRC統合情報基盤サイト(一般公開)

CBRC 統合DB情報基盤サイト - Mozilla Firefox

ファイル(F) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)

産業技術総合研究所
生命情報工学研究センター

統合DB情報基盤サイト

概要 Summary

「統合DB情報基盤サイト」とは

「ワークフロー」技術開発について

ワークフロー Workflow

タンパク質アノテーションワークフロー Protein Annotation Workflow

お問合せ

- 統合データベースプロジェクトポータル
- 文科省統合データベースプロジェクト

http://www.aist.go.jp/index_ja.html

「統合DB情報基盤サイト」とは

CBRCは文部科学省のライフサイエンス分野の統合DB情報基盤の整備を担当しています。ここはその情報基盤に関する様々な情報を提供しています。

「統合データベースプロジェクト」

ライフサイエンスデータの利便性を高めるための詳細は、[文科省統合データベースプロジェクト](#)

プロジェクトにおけるCBRCの役割

CBRCは分散する解析ソフトウェアを統合し、ワークフロー及び分散処理を実現しています。

お問合せ

- 統合データベースプロジェクトポータル
- 文科省統合データベースプロジェクト

CBRC 統合DB情報基盤サイト | Protein Annotation Workflow | About - Mozilla Firefox

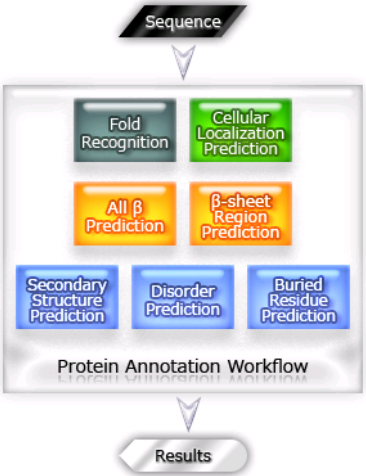
ファイル(F) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)

Protein Annotation Workflow

About EnterSequence SampleData SampleResult Help

Protein Annotation Workflowはアミノ酸配列から立体構造に関する様々な情報を提供することを目的としています。

利用方法や、使用プログラムの詳細については、[ヘルプ](#)を参照して下さい。



```
graph TD; Sequence --> FR[Fold Recognition]; Sequence --> CLP[Cellular Localization Prediction]; Sequence --> AP[All beta Prediction]; Sequence --> BRP[beta-sheet Region Prediction]; Sequence --> SSP[Secondary Structure Prediction]; Sequence --> DP[Disorder Prediction]; Sequence --> BRP2[Buried Residue Prediction]; FR --> Results; CLP --> Results; AP --> Results; BRP --> Results; SSP --> Results; DP --> Results; BRP2 --> Results;
```

完了

タンパク質アノテーション ワークフロー

CBRC 統合DB情報基盤サイト | Protein Annotation Workflow | Sample Data - Mozilla Firefox

ファイル(E) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)

産業技術総合研究所
生命情報工学研究センター

統合DB情報基盤サイト

概要 Summary

「統合DB情報基盤サイト」とは

「ワークフロー」技術開発について

ワークフロー Workflow

タンパク質アノテーションワークフロー
Protein Annotation Workflow

お問合せ

統合データベース
プロジェクトポータル

文科省統合データベース
プロジェクト

Protein Annotation Workflow

About EnterSequence SampleData SampleResult Help

You can use a sample sequence below for trying. Just enter your e-mail address, then click the submit button.

- Notice will be sent by e-mail upon completion of calculation.

```
>sample-sequence
ALRYGAINSKDVIIDAYS TDSSELIGYQLGILKDDKHLFFPYQGAPLLRQDTIKKYPGV
```

Perform transmembrane related prediction: No Yes

E-mail address:

submit clear

完了

Mozilla Firefox

ファイル(E) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)

PSIPRED Prediction Confidence

Joet Prediction Confidence

Sable Prediction Confidence

HHsearch Fold recognition (top list)

COILS Prediction

Joet Barred 25% prediction
Barred 5% prediction
Barred 0% prediction

Sable Barred level prediction

Barred confidence

POODLE-S Prediction Confidence

POODLE-L Prediction Confidence

POODLE-T Prediction Confidence

POODLE-W Prediction 0.02482, Ordered

BLAST search results for SWISS-PROT

InterProScan execution results

α-helix transmembrane region prediction

PHOBIUS, SignalP

Your sequence is NOT likely to have transmembrane regions

β-sheet transmembrane region prediction

TMBETADISC-COMP Details TMBETA-NET Details

Subcellular localization prediction

WoLF-PSORT

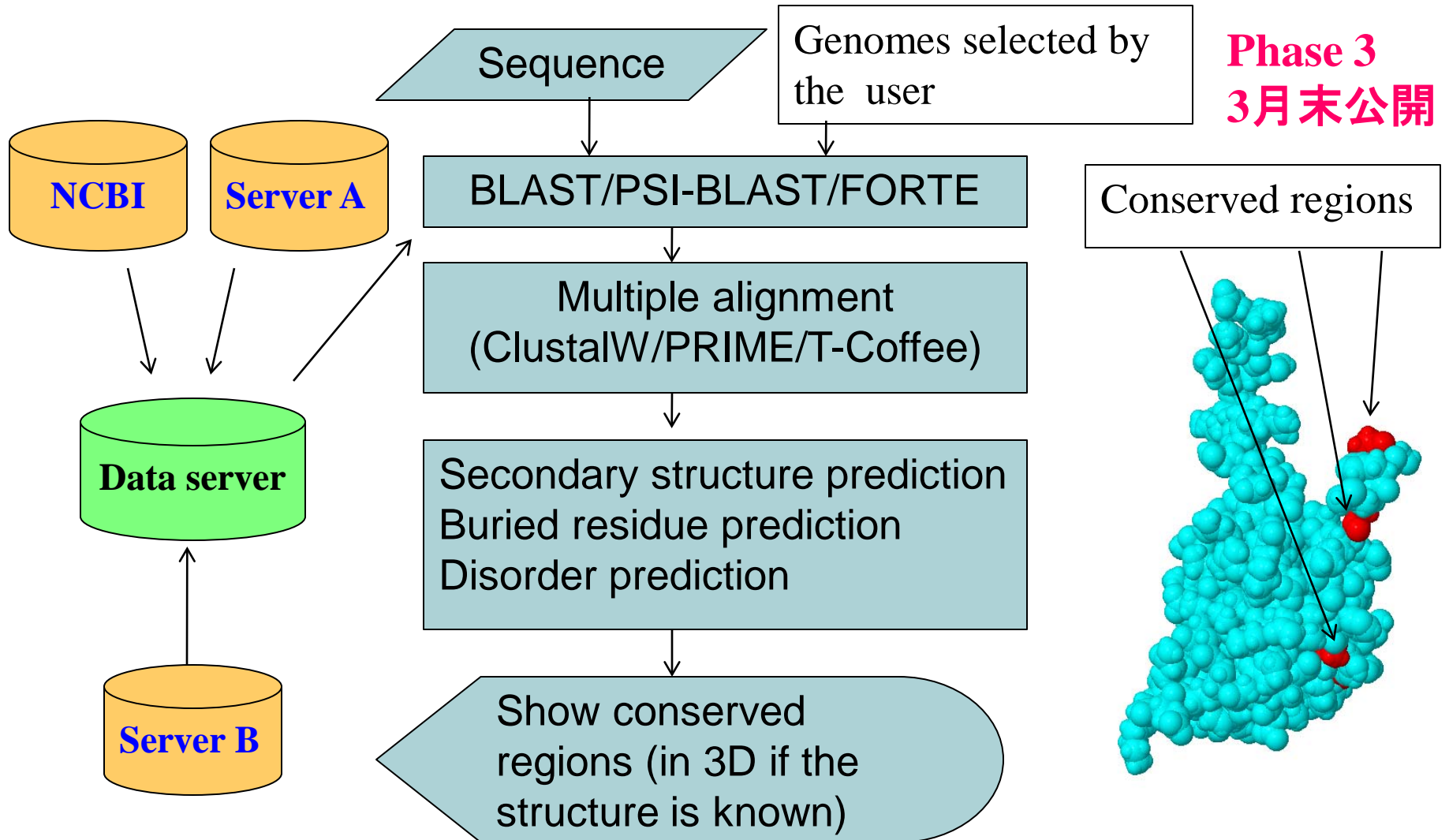
Animal cyto 22, cyto,nuc1 20, cytoplasm 14, 1007, nuc1 8.5

Plant cyto,nuc1 8.5, nuc1 3.5, cyto 4.5, mito 2

Fungi cyto 13.5, cyto,nuc1 13.5, nuc1 8.5, mito 5

完了

アミノ酸配列からタンパク質の比較情報を提示し、 保存部位、変異部位を推定するワークフロー



今後の計画

・ プロジェクト内連携強化

- 1) CBRC のサービスが DBCLS のサーバから利用可能
- 2) CBRC のサービスにDBCLSのサーバのサービスの組み込み
(横断検索の応用など)
- 3) 糖鎖DB、経産省統合DB (MEDALS)との連携

・ CBRC内のグリッド環境

- 1) DBCLSとグリッド環境の連携 (OpenIDとグリッドIDなど)
- 2) グリッド用計算サーバ整備

・ ワークフローのサービス

- 1) 他のサービスの検討・設計
 - ・タンパク質立体構造モデリング、RNA関連のワークフロー
 - ・ユーザニーズの調査に基づくワークフロー
- 2) ワークフローを構成するソフトウェアのモジュール化
(taveruna、KNIMEのイメージ)