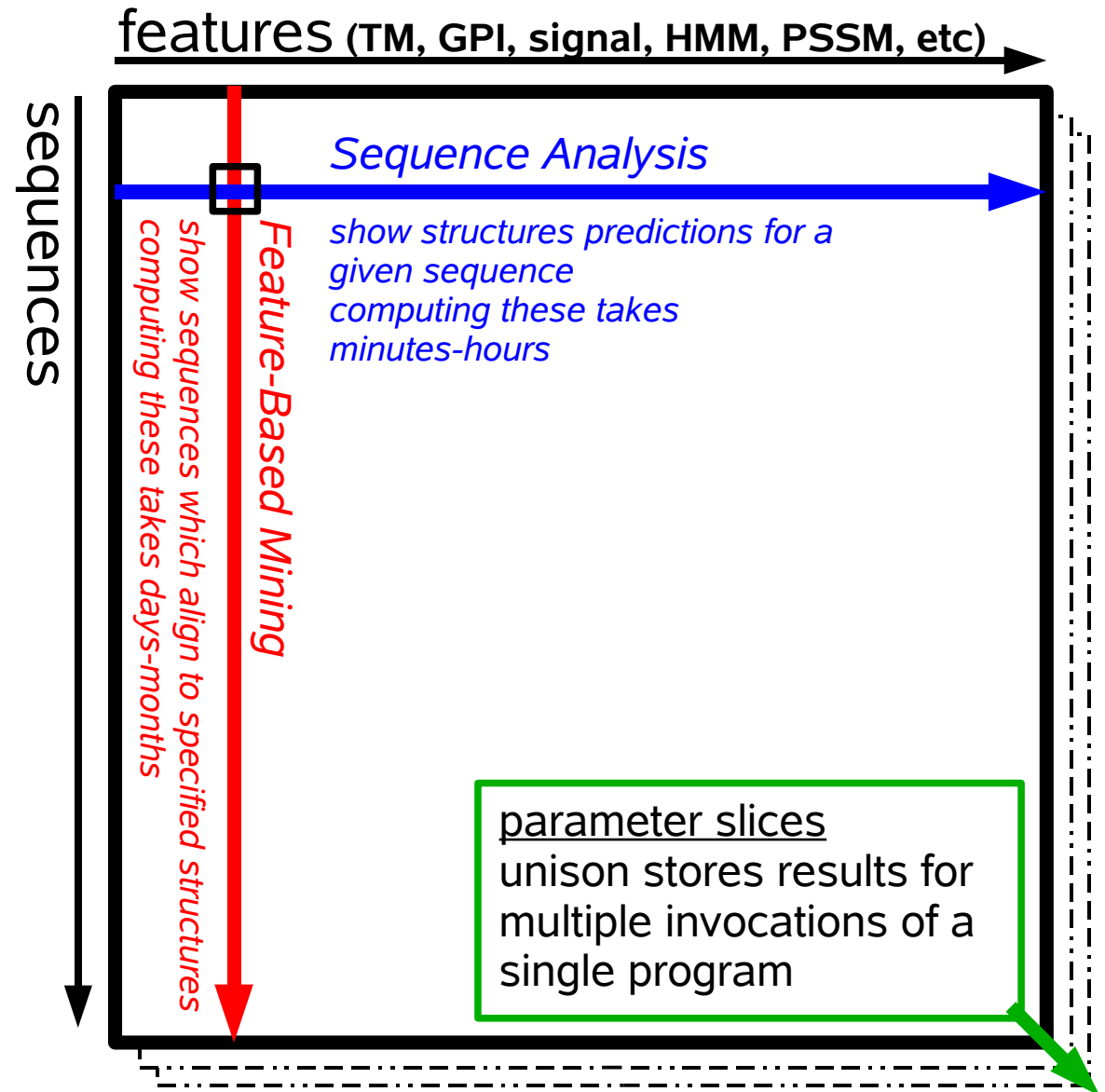


Unison: Integrated Feature-Based Mining for Target Discovery

Reece Hart
Genentech, Inc.
rkh@gene.com
<http://unison-db.org/>



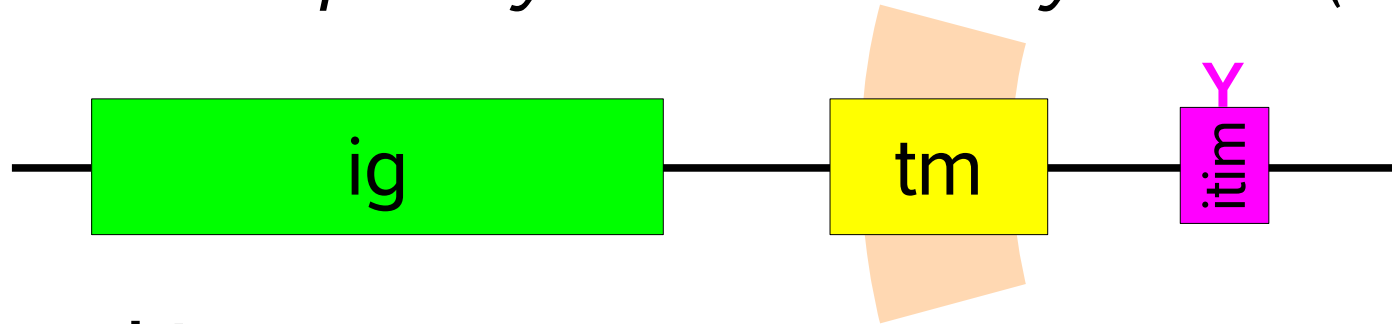
Unison in a Nutshell

- >6M non-redundant sequences of varying qualities from >20 sources, all species
- precomputed results from HMM, Prospect threading*, TM detection, signal prediction, cellular localization, genomic localization, regular expression motifs, and others
- integrated patent info*
- indexes designed for rapid hypothesis testing
- auxiliary data: SCOP, GO, HomoloGene, MINT, and others
- bookkeeping (run histories) and automated load/update procedures
- web front-end

* Prospect threading results and Derwent patent information is not available in the public release. The code to load and maintain these data *is* available.

Feature-Based Mining with Unison

Immunoreceptor Tyrosine Inhibitory Motifs (ITIMs)



SELECT

```
hmm.pseq_id,
```

```
hmm.start as "ig_start",hmm.stop as "ig_stop",
```

```
tm.start as "tm_start",tm.stop as "tm_stop",
```

```
itim.start as "itim_start",itim.stop as "itim_stop"
```

FROM

```
pahmm hmm
```

```
JOIN pftmdetect tm on hmm.pseq_id = tm.pseq_id
```

```
JOIN pfregexp itim on hmm.pseq_id = itim.pseq_id
```

WHERE

```
hmm.pmodel_id=pmodel_id('ig') AND hmm.eval <= 0.02
```

```
AND hmm.params_id=params_id('Pfam_fs 18.0')
```

```
AND tm.pftype_id in (5,6) AND tm.prob >= 0.5
```

```
AND itim.pmodel_id=pmodel_id('ITIM')
```

```
AND ig.stop < tm.start AND tm.stop < itim.start
```

Unison Interfaces

- <http://unison-db.org/>
 - public data, Unison tour, download, mailing list, etc.

Welcome to **unison**

integrated, precomputed proteomic predictions for rapid feature-based mining, sequence analysis, and hypothesis generation

Unison integrates many types of protein predictions to enable mining of sequences based on holistic protein feature criteria, synthesis of these predictions for the analysis of individual sequences, and refinement of hypotheses regarding the composition of protein families. The Unison schema, source code, web interface, and public data are [released under the Academic Free License](#).

Try Unison. There are two primary ways to use the Unison web interface:

- 1. Individual Sequence Analysis: Specify a sequence, retrieve features.**

Enter a sequence, sequence alias, md5 checksum:

An alias may be an accession or identifier from any database contained in Unison.
e.g., INFA_HUMAN, P01375, NP_000585.2, IP100001671.1, 60ada54e69e411bcf6b08e9dacff7a48

Be sure to click the Protein Analysis tabs to explore prediction details.
- 2. Feature-Based Mining: Specify features, retrieve matching sequences.**

Unison was designed for feature-based mining. Please try:

 - [searching by sequence properties](#)
 - [exploring curated models and sequence sets](#)
 - [browsing predefined, dynamic queries](#)

Learn More. Much more sophisticated queries are possible using the Perl API and the PostgreSQL interactive SQL interpreter. Please see the [Unison tour](#) for real-life examples and a demonstration of some of Unison's features.

Genentech Bioinformatics Public Share SOURCEFORGE.net PostgreSQL Powered open source

- unison-db.org:5432
 - direct db access (psql, Pg, odbc, jdbc)

protein features →

Analyze Search Browse

Summary Aliases Patents Features BLAST Prospect2 HMM PSSM Loc History

Unison:8602 Features Overview

current "best" annotation: (Human) Homo sapiens hypothetical protein

Unison:8602

Unison:8602: 333 AA:

detect

EMBOSS/seqcleave

EMBOSS/antigenic

reseqp motif

PSSM/SBP (top 4 hits of 93 w/evalC5)

HMM (top 4 hits of 18 w/evalC3)

prospect (top 5 hits of 38 w/evalC7)

\$Id: features.pm.v 1.1 2004/02/25 20:40:00 rkh Exp \$

aliases ↓

Aliases (7)

These are the aliases from the most reliable sources only; see

origin	alias	description
Swiss-Prot	P40200	T-cell surface protein tactile pre
Swiss-Prot	TACT_HUMAN	T-cell surface protein tactile pre
Proteome	NP_005807.1	[Human] NM_005816 Homo sap (TACTILE), mRNA. 5205 bp, mR
Refseq	NP_005807.1	CD96 antigen isoform 2 precurs surface protein tactile [Homo sa
GenenGenes	PRO36589	Human TACT
GenenGenes	PRO2460	Human TACT
GenenGenes	PRO25146	Human TACT Faux

genomic loci →

chr3:112576997-112694551

112590k 112600k 112610k 112620k 112630k 112640k 112650k 112660k 112670k 112680k 112690k

Unison:1184 (Swiss-Prot:TACT_HUMAN)

Unison:3186453 (Proteome:NP_937839.1)

Unison:138994 (Incyte:INCY:950994_FL1p)

Unison:2852837 (ProAnno_v1:gene3819_trans4)

Unison:7004 (Geneseq:AAB94459)

Unison:13591204 (TNF6F 2:chr6L144088201:1440886501-F0)

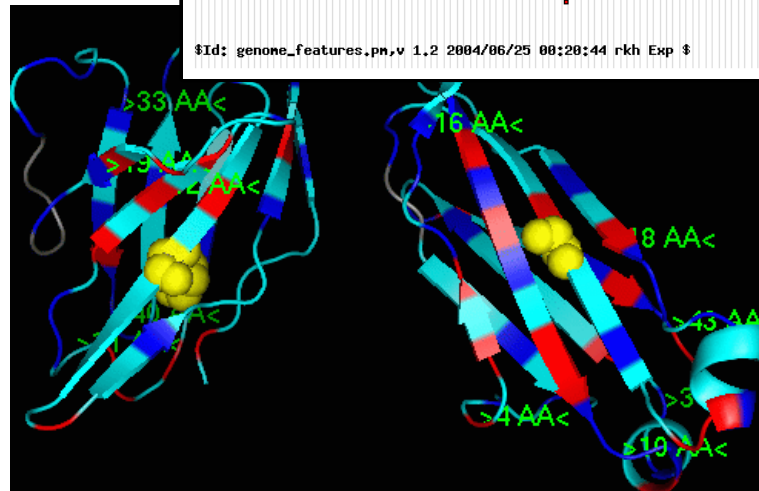
Unison:3418715 (TNF6F 2:chr2L88288201:882886501-F2)

Unison:2837809 (TNF6F:chr11L97180801:19)

Unison:3631284 (TNF6F 2:chr3L112579801:1125802501-R0)

Unison:139588 (Swiss-Prot:ZBE2_HUMAN)

\$Id: genome_features.pm.v 1.2 2004/06/25 00:20:44 rkh Exp \$



← PDB structures and homology models, with variant viewer