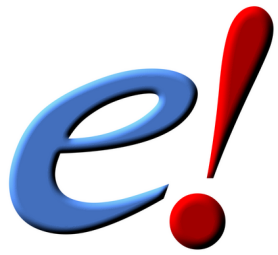


Overview of DAS

Andreas Kähäri

andreas.kahari@ebi.ac.uk

das@ebi.ac.uk

EMBL
European Bioinformatics Institute
Wellcome Trust Genome Campus,
Hinxton, United Kingdom

25th April 2006

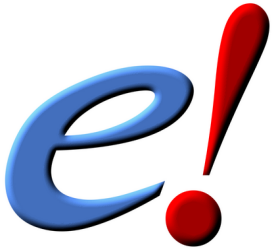
The idea of DAS

The original idea of Lincoln Stein (CSHL) and his collaborators was for the Distributed Annotation System (DAS) in 1999/2000 was to provide a GFF-based web service.

In this way, applications would be able to access genomic annotations from multiple remote locations using the Internet.

The idea of DAS (cont.)

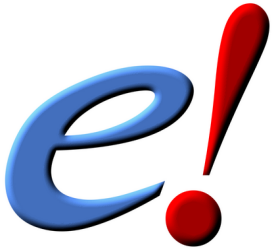
DAS shows its roots in the more or less 1:1 mapping between the DAS XML schema for annotation features and the GFF file format.



The DAS specification

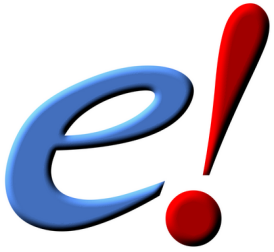
The DAS specification is a specification of a communication protocol for

- a DAS client requesting annotation data from a DAS server, and
- a DAS server sending annotation data to a DAS client.



The DAS specification (cont.)

A *protocol* is an agreed way of how to go about doing things.

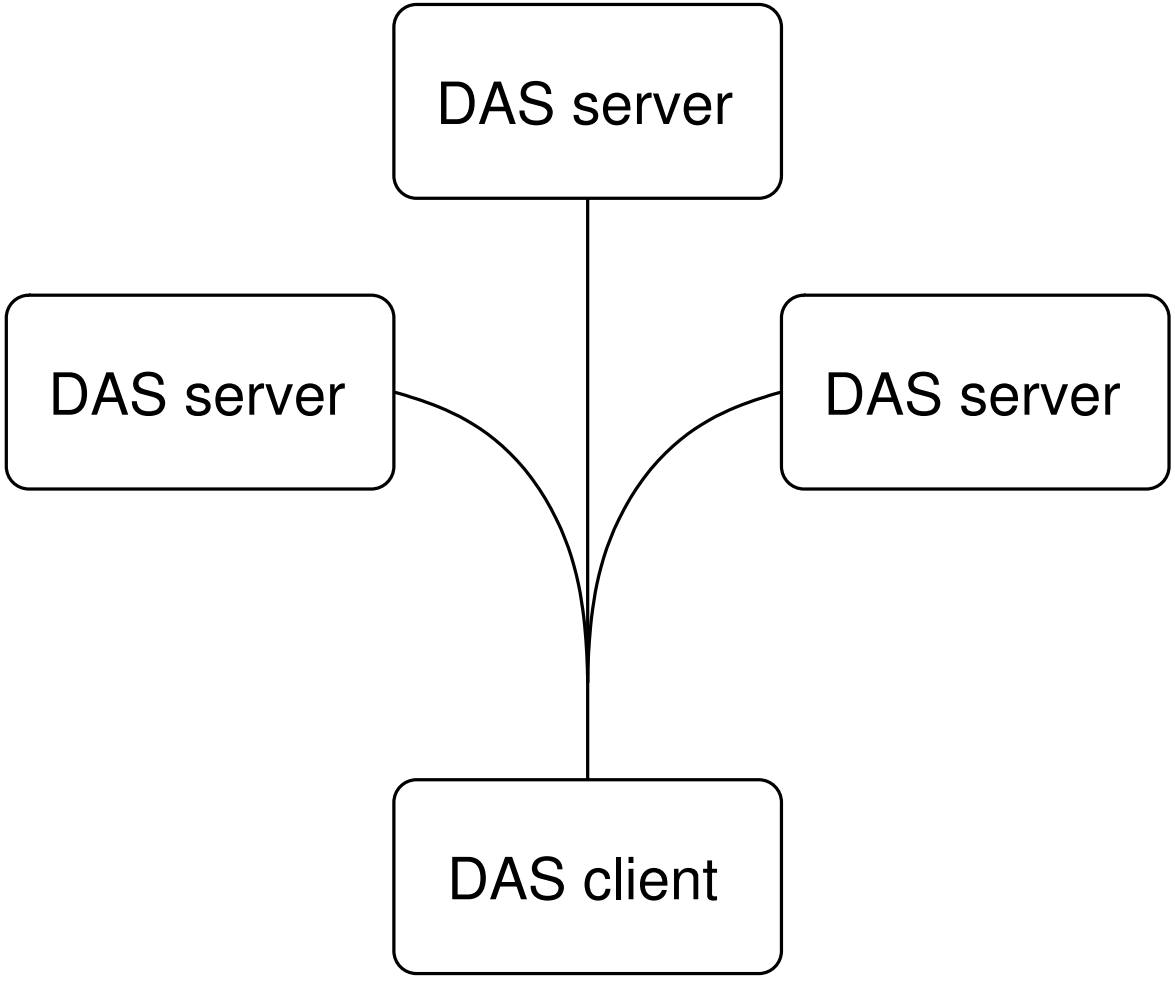


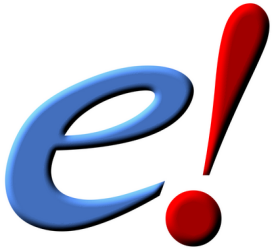
The DAS specification (cont.)

A *communication* protocol, in a computer context, is a set of rules that formalises communication between two or more *communication end-points*.

In the context of DAS, the communication end-points consist of the DAS server and the DAS client.

The DAS specification (cont.)

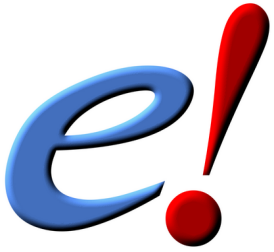




DAS servers

A *server*, in general, is a program that provides (serves) data.

A DAS server is made up of a collection of one or more *data sources*, each serving a set of annotations.



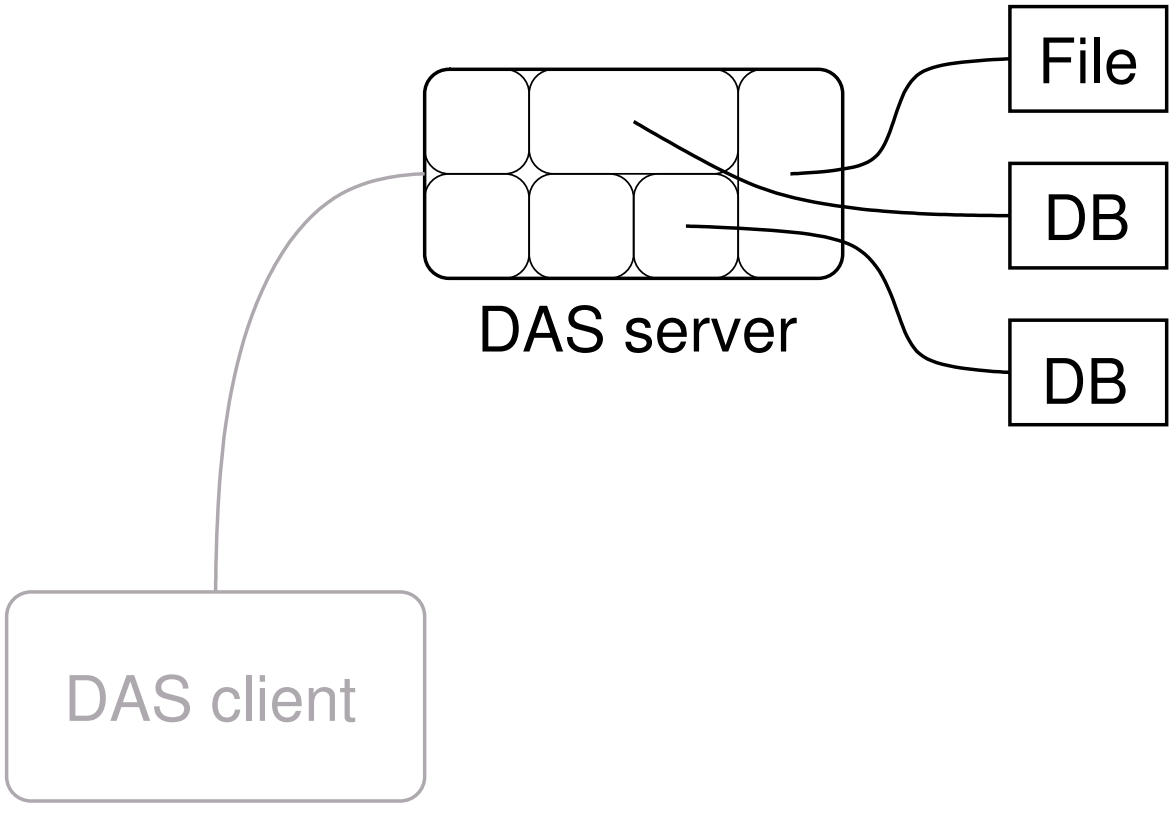
DAS servers (cont.)

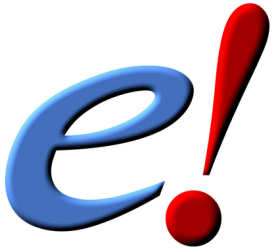
Usually, but not always, a DAS data source is connected to a database which contains some sort of annotations.

Some DAS data sources might also serve data directly from GFF files or from other types of files or storage.

Dalec (by Tom Oinn's group at EBI) provides a *compute on demand* backend to DAS.

DAS servers (cont.)





DAS servers (cont.)

Examples of DAS server implementations:

Dazzle: a server written by Thomas Down (Sanger) in Java, running on Tomcat or Resin.

ProServer: a lightweight server written by Roger Pettett (Sanger) in Perl.

LDAS: the original server implementation by Lincoln Stein.

Different types of DAS servers

There are three main categories of DAS servers. This categorization depends on the annotation domain:

Genomic DAS:

Annotating positional features on chromosomes, super-contigs, contigs or scaffolds.

Non-positional DAS ("GeneDAS"):

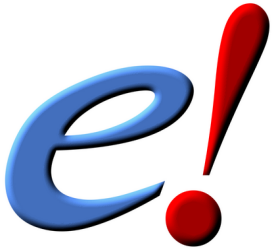
Attaching annotations to gene IDs or protein IDs. The annotation concerns the object as a whole.

Protein DAS:

Annotating positional features on proteins.

Querying a DAS server

If you know how to construct a DAS URL, you can query a DAS server with a web browser, without using a special client.



Querying a DAS server (cont.)

A DAS query is an URL which follows a certain pattern:

`http://`**`prefix`**`/das/`**`source`**`/`**`command`**`?`**`args`**

Querying a DAS server (cont.)

The first part, "`http://`**`prefix`**`/das`", denotes a specific DAS server.

The **`source`** is the name of a data source on the specified server. This is also commonly known as the DSN, the Data Source Name.

The **`command`** and **`args`** part of the query determines what kind of data will be sent back from the server.

Querying a DAS server (cont.)

The DAS specification describes a number of commands, most of which are optional.

The most commonly used commands are:

`features:` Given a *segment* as argument, features on that segment are returned.

`types:` Returns information about the annotation types that this server uses. This command also takes an optional segment as argument.

Querying a DAS server (cont.)

A *segment* is a region on a reference sequence and is specified as "`segment=`**`REF:start,stop`**".

`REF:` The name of the reference object. Depending on the server, this is a name of a chromosome, contig, a protein, or of a gene.

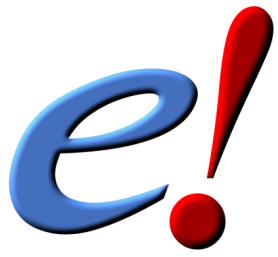
`start,stop:` These are optional and mostly only used when querying genomic DAS servers. The start and stop positions are always specified on the positive strand (if applicable) on the reference object `REF`.

Querying a DAS server (cont.)

There is also a **`dsn`** command that allows a client to determine what data sources are available on a DAS server.

The format of this command differs a bit from that of the other DAS commands:

`http://`**`prefix`**`/das/`**`dsn`**



Example queries

Firefox Community Edition

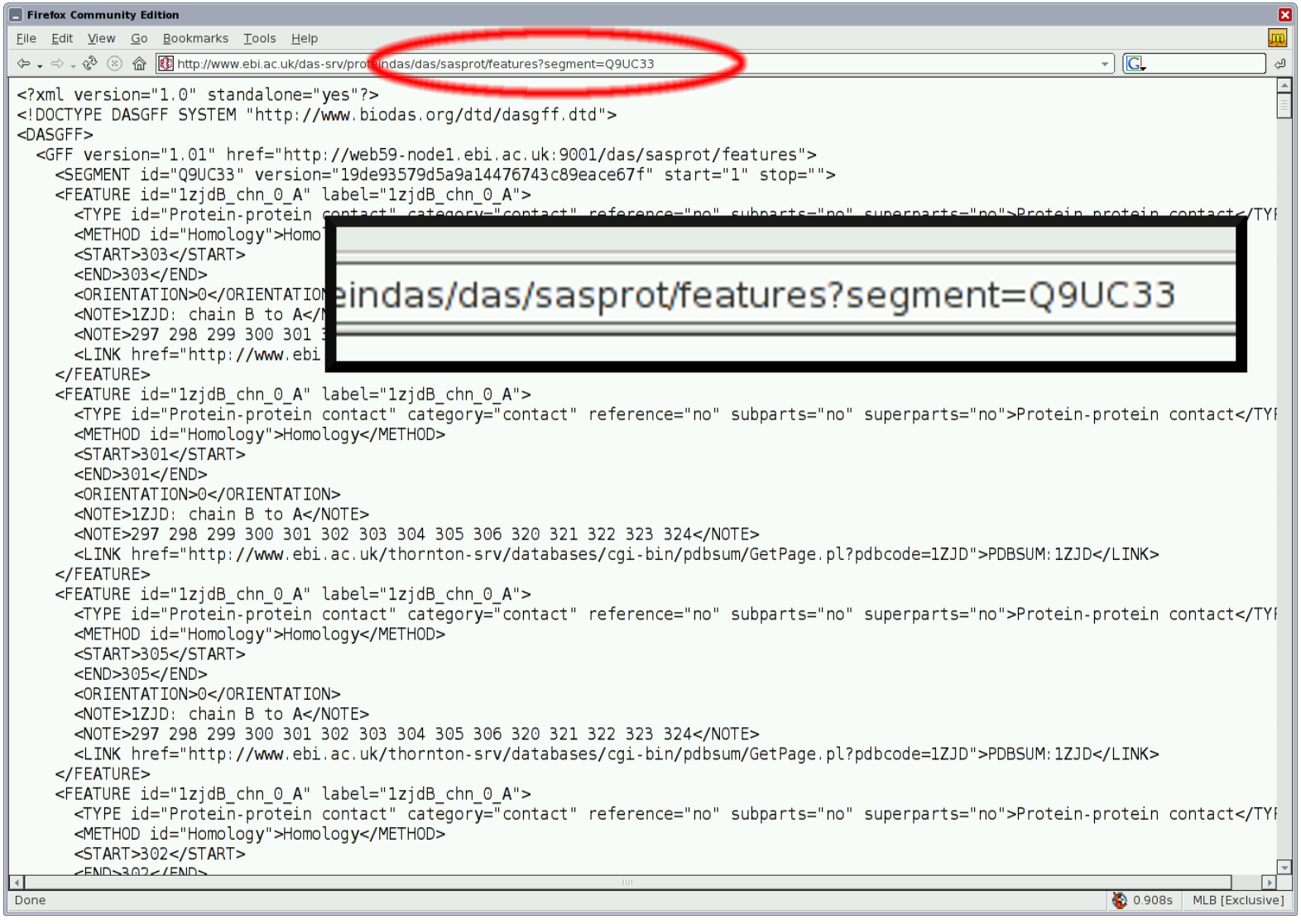
File Edit View Go Bookmarks Tools Help

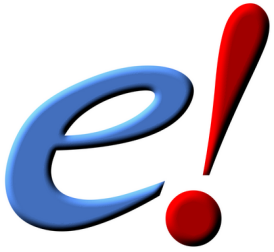
http://www.ebi.ac.uk/das-srv/proteindas/das/dsn

```
<?xml version="1.0" standalone="no"?>
<!DOCTYPE DASDSN SYSTEM 'http://www.biodas.org/dtd/dasdsn.dtd' >
<DASDSN>
  <DSN>
    <SOURCE id="gencodealign
    <MAPMASTER>http://www.ebi
    <DESCRIPTION>Serving UniP
  </DSN>
  <DSN>
    <SOURCE id="sasdna" versi
    <MAPMASTER>http://www.ebi
    <DESCRIPTION>Putative protein-DNA binding sites by homology with those observed in crystallized protein structures.</DESCRIPTION>
  </DSN>
  <DSN>
    <SOURCE id="gencodepfam" version="1.0">gencodepfam</SOURCE>
    <MAPMASTER>http://www.ebi.ac.uk/das-srv/proteindas/das/gencode</MAPMASTER>
    <DESCRIPTION>Putative domains predicted using Pfam HMMs on ENCODE peptides</DESCRIPTION>
  </DSN>
  <DSN>
    <SOURCE id="csaextended" version="1.0">csaextended</SOURCE>
    <MAPMASTER>http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle</MAPMASTER>
    <DESCRIPTION>Catalytic Site Atlas (http://www.ebi.ac.uk/thornton-srv/databases/CSA/) literature annotations extended to putative
  </DSN>
  <DSN>
    <SOURCE id="gencodealign-t" version="1.0">gencodealign-t</SOURCE>
    <MAPMASTER>http://www.ebi.ac.uk/das-srv/proteindas/das/gencode</MAPMASTER>
    <DESCRIPTION>Serving UniProt and UniParc alignment on GENCODE peptides (TESTING SERVER)</DESCRIPTION>
  </DSN>
  <DSN>
    <SOURCE id="sasprot" version="1.0">sasprot</SOURCE>
    <MAPMASTER>http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle</MAPMASTER>
    <DESCRIPTION>Putative protein-protein binding sites by homology with those observed in crystallized protein structures.</DESCRIP
  </DSN>
  <DSN>
    <SOURCE id="gencode-t" version="1.0">gencode-t</SOURCE>
    <MAPMASTER>http://www.ebi.ac.uk/das-srv/proteindas/das/gencode</MAPMASTER>
    <DESCRIPTION>GENCODE peptide reference server (TESTING SERVER)</DESCRIPTION>
  </DSN>
  <DSN>
    <SOURCE id="pdbsas" version="1.0">pdbsas</SOURCE>
```

Done 1.560s MLB [Exclusive]

Example queries (cont.)





Example queries (cont.)

For users, it is fortunately never necessary to look at the XML directly. There are clients (“viewers”) that makes life easier.

DAS clients

A *client*, in general, is a program that needs data.

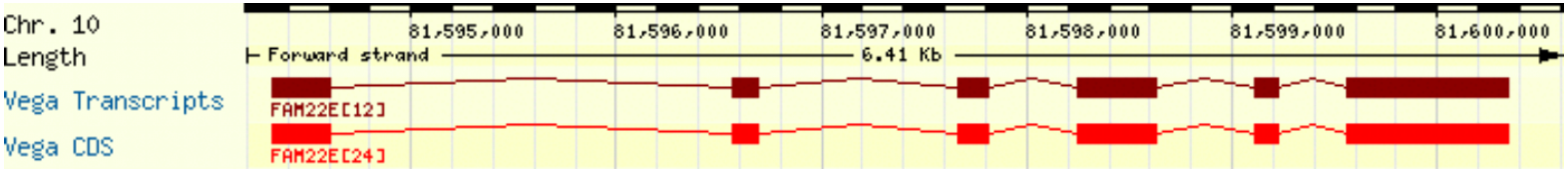
Usually, a DAS client needs annotation data to construct some sort of graphical or textual display for a user who is interested in a particular region of a chromosome (or protein or gene).

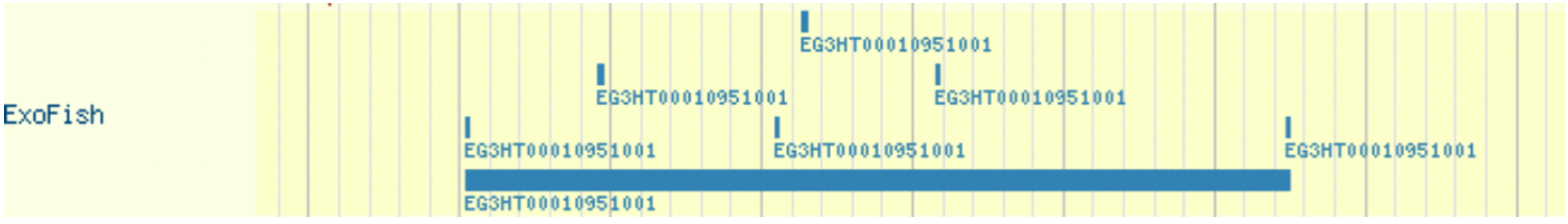
DAS clients (cont.)

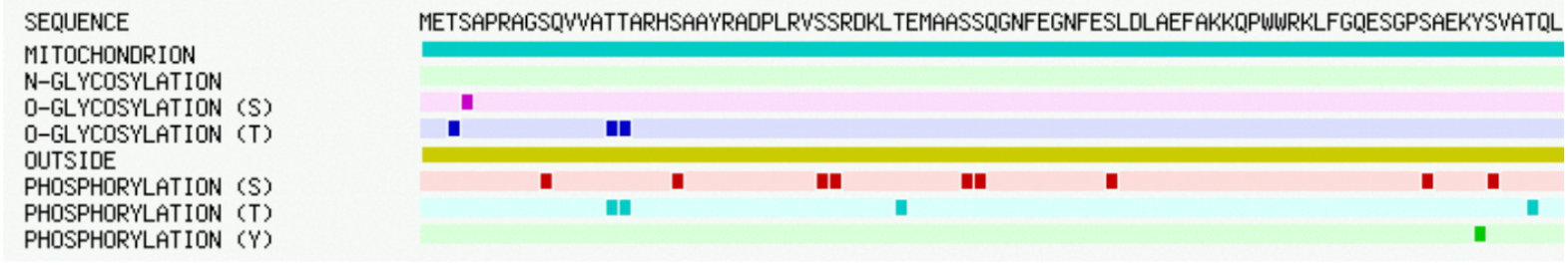
The DAS client may represent the data in any way it wants. The DAS specification does not impose any restrictions on presentation.

For graphical displays, the most common way to represent a feature as a simple box, spanning the length of the annotation.

DAS clients (cont.)

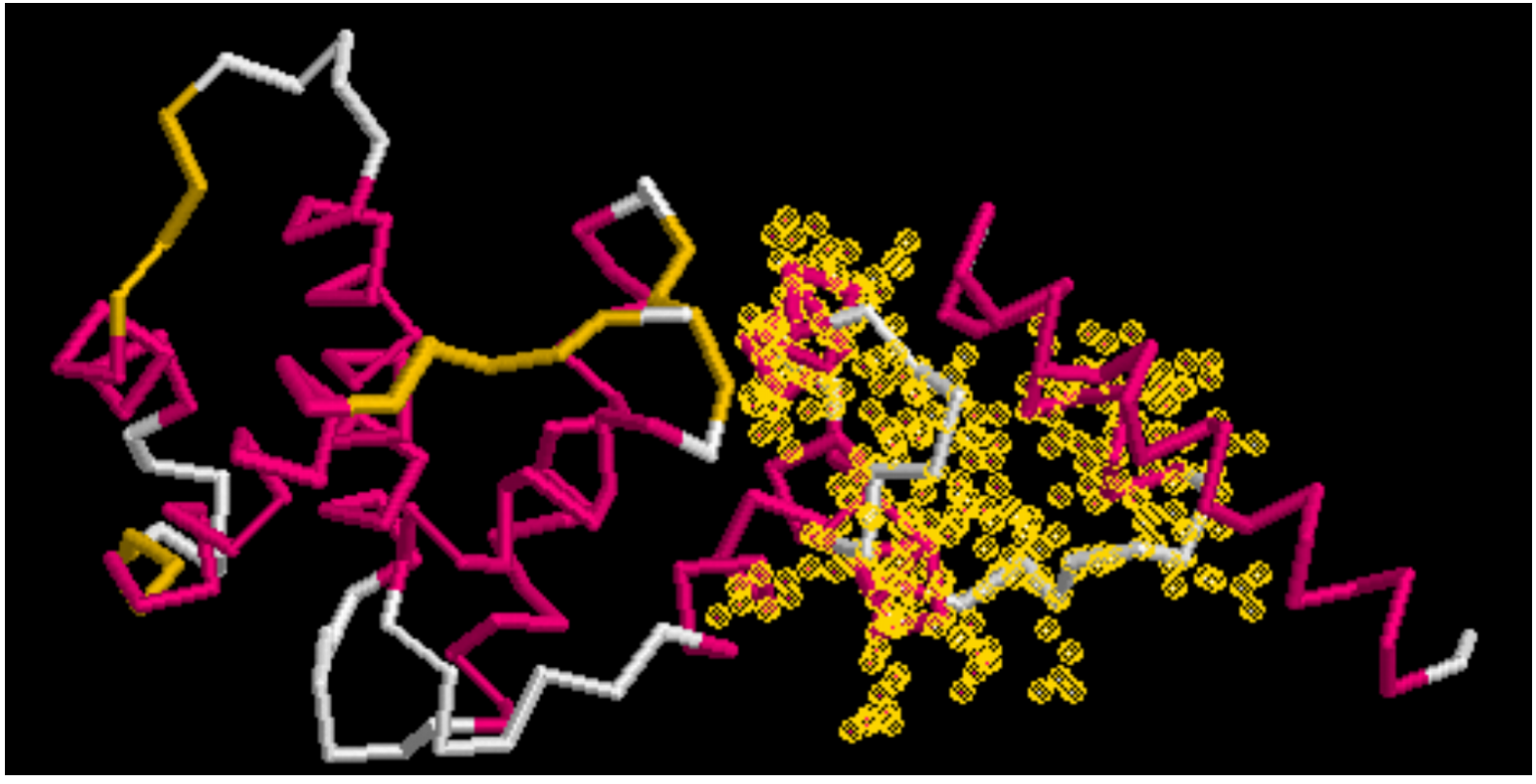






P69905	reference	Braunitzer G., Gehring-Muller R., Hilschmann N., Hilse K., Hobom G., Rudloff V., Wittmann-Liebold B. (1961). "The constitution of normal adult human haemoglobin.", Hoppe-Seyler's Z. Physiol. Chem. 325:283-286
P69905	reference	Brimhall B.J., Duerst M., Hollan S.R., Stenzel P., Szelenyi J., Jones R.T. (1974). "Structural characterizations of hemoglobins J-Buda (alpha 61 (E10) Lys-to-Asn) and G-Pest (alpha 74 (EF3) Asp-to-Asn).", Biochim. Biophys. Acta 336:344-360
P69905	reference	Flint J., Higgs D.R. (1997). Submitted to database EMBL/GenBank/DDBJ databases

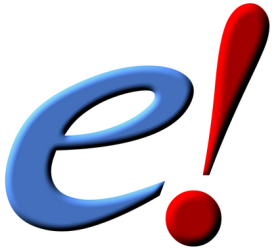
DAS clients (cont.)



DAS clients (cont.)

Examples of DAS clients:

- Parts of Ensembl ContigView (“classic” genomic DAS).
- Parts of Ensembl ProtView (protein DAS).
- Parts of Ensembl GeneView (gene DAS, or “non-positional” DAS).
- Dasty (protein DAS and gene DAS).
- Spice (protein DAS, with structure and alignment extensions).
- ProView (protein DAS).



General DAS client functionality

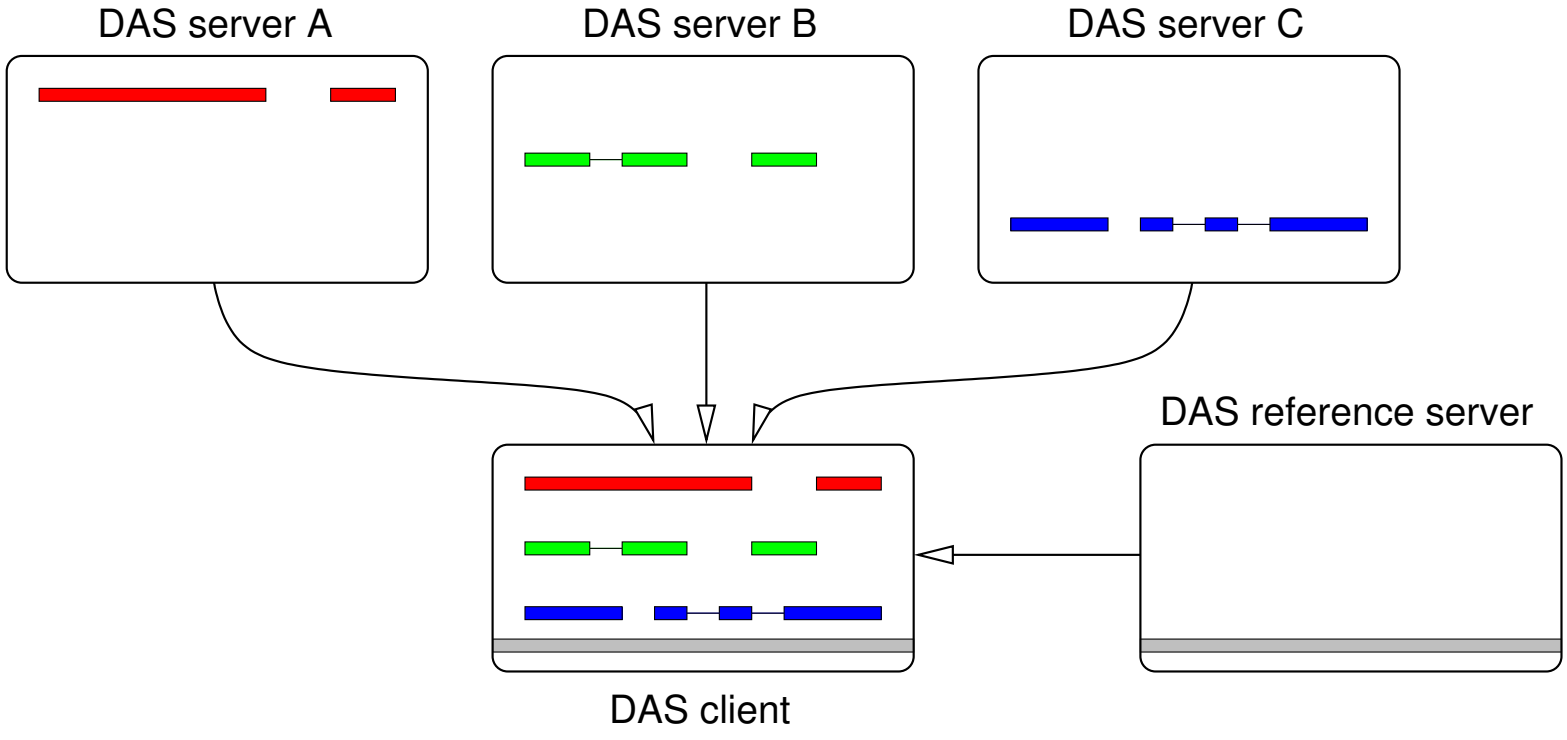
Clients (“viewers”) for the Distributed Annotation System (DAS) comes in different shapes and forms, but they all operate according to the same principles.

General DAS client functionality (cont.)

Given a list of DAS servers,

1. query all servers,
2. display the annotations on the common reference sequence.

General DAS client functionality (cont.)



General DAS client functionality (cont.)

Important things to remember:

- A DAS server annotates a reference sequence.
- Unless the client is very smart, all attached DAS servers must annotate *the same* reference sequence.

General DAS client functionality (cont.)

The following will *not* make sense unless the client is very smart:

- Attaching a DAS server that annotates human NCBI 35 to a client that shows mouse NCBI m35.
- Attaching a DAS server that annotates human NCBI 35 to a client that shows human NCBI 36.
- Attaching a DAS server that annotates UniProt proteins to a client that shows ENCODE proteins.