

### El caso

Os han llamado para ayudar en la investigación de una escena de un crimen: se ha encontrado el cuerpo sin vida de un turista americano en el aeropuerto. Parece que ha padecido fuertes convulsiones y una hemorragia interna. Los investigadores que se encuentran en la escena del crimen han encontrado un envase que contiene restos de un líquido que parece leche. Esto podría ser una pista.

El líquido se ha analizado en el laboratorio y os han mandado una lista de los componentes encontrados en la bebida. Se encontraron algunas moléculas pequeñas, como azúcares, pero también se encontraron cuatro proteínas desconocidas. Vuestro trabajo es analizar estas proteínas para intentar descubrir que causó la muerte del turista. Usad vuestro ordenador para buscar y analizar información referente a estas proteínas. Algunas de las potentes herramientas y bases de datos utilizadas en bioinformática os ayudaran durante la investigación.

### Identificación de las proteínas sospechosas

La lista que contiene las secuencias de aminoácidos de las cuatro proteínas (nombradas sospechosa 1 a 4) están en el archivo llamado P1\_seq\_fa.doc.

Las secuencias de aminoácidos de las proteínas se dan con el código de una letra que usan la mayoría de científicos, si no conocéis la correspondencia de códigos podéis consultar este artículo de la Wikipedia en el apartado "Clasificación".

Ahora ya tenéis suficiente información para empezar con vuestra investigación. Para cada una de las proteínas sin identificar debéis contestar a las siguientes preguntas:

- 1 ¿De que proteína se trata?
- 2 ¿En que organismo se encuentra?
- 3 ¿Cual es la función de esta proteína?
- 4 ¿Es esta la proteína culpable? ¿Puede ser la causante de la muerte del turista? ¿Porque? ¿O Porque no?
- 5 ¿Tiene la proteína alguna característica remarcable?

**Nota:** Tenéis una plantilla para anotar las respuestas.

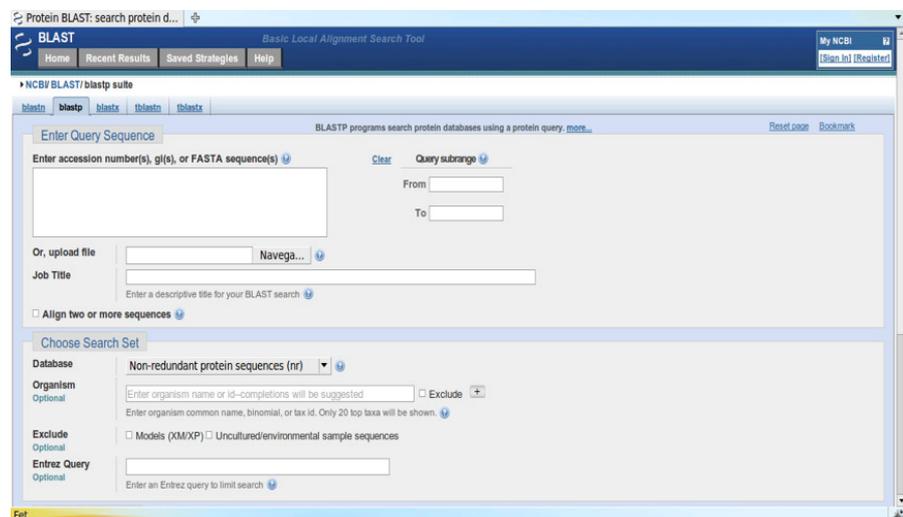
BLAST es un programa usado para la comparación de una determinada secuencia proteica contra todas las proteínas presentes en una de las bases de datos biológicos disponibles. En nuestro caso usaremos la base de datos SwissProt. Esta base de datos contiene proteínas anotadas manualmente para las que se conoce la función, estructura y/o cambios post-traduccionales. Nos será de utilidad para encontrar la función de nuestras proteínas. Hay diferentes tipos de blast. En nuestro caso utilizaremos blastp, que hace la búsqueda en una base de datos proteicos usando una secuencia proteica. Encontraremos el blastp en el servidor del "The National Center for Biotechnology" (NCBI).

Para ayudaros en vuestra búsqueda cogeremos la proteína sospechosa 1 y os guiaremos a lo largo del proceso. Luego podréis hacer lo mismo con las proteínas restantes.

#### Paso 1

Abrid la pagina de BLAST y escoged la opción "protein blast".

Llegareis a una página como esta:

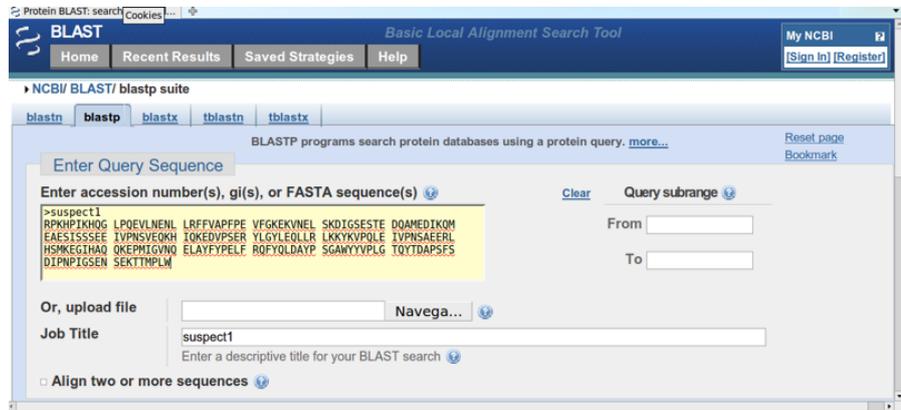




# Asesinato en el aeropuerto

## Paso 2

Copiad ahora la secuencia des del archivo al cuadro nombrado "Enter accession numbers...". Como podéis ver en el ejemplo de abajo, que contiene la proteína sospechosa 1, la primera línea debe empezar con este símbolo: >, y vendrá seguido del nombre de la proteína. Después de un salto de línea, vendrá la secuencia. Este formato se llama formato FastA.



En el apartado "Choose Search Set", en el campo "Database" debeis escoger la base de datos de SwissProt ("UniProtKB/Swiss-Prot protein sequences (swissprot)")

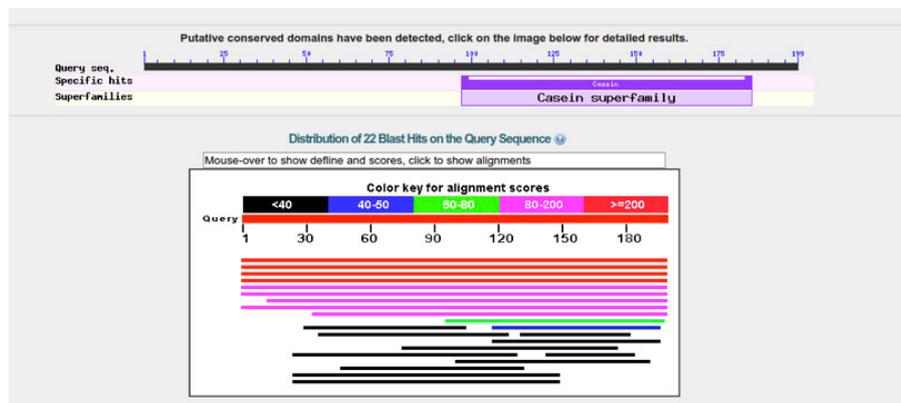
Ahora estáis preparados para lanzar vuestro primer blast. Apresad el botón BLAST situado a la izquierda de la ventana.

Esto os mandara a una página de espera mientras se carguen vuestros resultados. Debéis considerar que mucha gente está usando el BLAST a la vez y que la base de datos contiene cientos de miles de secuencias proteicas, así que quizás tendréis que esperar un poco para obtener vuestros resultados.

## Paso 3

Una vez la búsqueda ha terminado veréis una ventana con los resultados. Moveos por la página sin miedo, está llena de información.

Si miráis la tabla de resultados veréis esto:



Sequences producing significant alignments:

Description	Max score	Total score	Query cover	E value	Ident	Accession
ReName: FullAlpha-S1 casein; AltName: A1casein; Bos d.8. Contain; ReName: FullArcticoidant peptide; Flaga; Precursor (Bos taurus)	409	409	100%	5e-145	100%	F52852.2
ReName: FullAlpha-S1 casein; Flaga; Precursor (Bubalus bubalis)	363	363	100%	8e-127	95%	O62823.2
ReName: FullAlpha-S1 casein; Flaga; Precursor (Ovis aries)	355	355	100%	7e-124	87%	F51653.3
ReName: FullAlpha-S1 casein; ShortAlpha-S1-CN; Flaga; Precursor (Capra hircus)	352	352	100%	1e-122	87%	F13626.2
ReName: FullAlpha-S1 casein; Flaga; Precursor (Bos indicus)	159	159	100%	4e-39	40%	F30555.1
ReName: FullAlpha-S1 casein; Flaga; Precursor (Camelus dromedarius)	109	109	100%	1e-27	42%	O37432.2
ReName: FullAlpha-S1 casein; ShortAlpha-casein; Flaga; Precursor (Oryzopsis carolinensis)	92.4	92.4	93%	2e-21	38%	F30115.1
ReName: FullAlpha-S1 casein; (Tricus asiaticus)	91.7	91.7	100%	2e-21	40%	F98372.1
ReName: FullAlpha-S1 casein; AltName: FullCasein-B; Flaga; Precursor (Civita porcellina)	85.5	85.5	83%	3e-19	46%	F94856.1
ReName: FullAlpha-S1 casein; Contain; ReName: FullCasein-D; Flaga; Precursor (Bos taurus)	56.2	56.2	51%	5e-09	37%	F47121.1
ReName: FullAlpha-S1 casein; ShortAlpha-casein; Flaga; Precursor (Mus musculus)	44.3	44.3	39%	2e-04	34%	F13228.1
ReName: Full-Dox; transport protein; Zebf (Photobacterium ambrosiense subsp. leuconis) T101	34.7	34.7	44%	0.39	25%	O76253.1
ReName: Full-AMP; GMP-dependent 2,3-bisphosphoglycerate kinase 7; ANName: FullPhosphodiesterase 7; ShortPDE7; Flaga; Precursor	32.7	32.7	26%	1.6	37%	Q5H10.1

**Ident:** Muestra como de parecidas son las dos proteínas. 100% indica que la proteína encontrada es idéntica a nuestra proteína sospechosa.

**Query cover:** Señala que parte de la proteína se corresponde con la nuestra. Si por ejemplo el valor es de 50% indica que solo la mitad de la proteína encontrada se parece a la nuestra.

Si bajamos a la parte de la página que se encuentra debajo de la tabla, podremos ver el alineamiento de nuestra secuencia sospechosa con cada uno de los resultados e información sobre este alineamiento.

una idea de cuánto podemos confiar en nuestros resultados.

## Asesinato en el aeropuerto

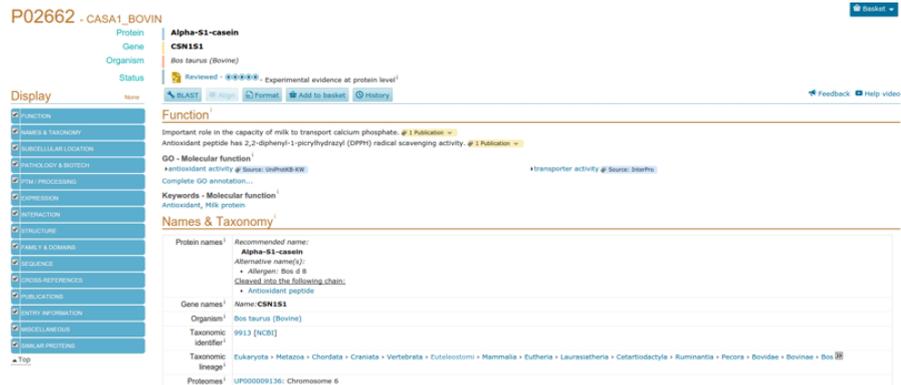
pág. 3/3

### Paso 4

Apretad sobre el "accession number" de la primera secuencia de la lista. Como podréis ver os llevara a una página distinta donde encontrareis mucha información sobre la proteína como el nombre, nombres alternativos, organismo al que pertenece la proteína, la secuencia o en que artículos aparece citada la proteína.

De todos modos, si esta manera de ver la información no os es familiar, podéis ver la misma información pero de una manera mas clara si vais al apartado "DBSOURCE" (base de datos de donde se ha obtenido la proteína) y apretáis sobre el accesoión "number de UniProt".

De esta manera veréis las características de la proteína en la interficie de UniProt que tiene un formato más intuitivo.



The screenshot shows the UniProt entry for P02662, CAS1\_BOVIN. The entry is for the protein Alpha-S1-casein (CSN1S1) from Bos taurus (Bovine). It includes a list of accession numbers, a definition, and a detailed description of its function as an antioxidant peptide. The 'Function' section states: 'Important role in the capacity of milk to transport calcium phosphate. Antioxidant peptide has 2,2-diphenyl-1-picrylhydrazyl (DPPH) radical scavenging activity.' The 'GO - Molecular function' section lists 'antioxidant activity' and 'transporter activity'. The 'Names & Taxonomy' section shows the protein name 'Alpha-S1-casein', gene name 'CSN1S1', and taxonomic classification: Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi, Mammalia, Eutheria, Laurasiatheria, Cetartiodactyla, Ruminantia, Pecora, Bovidae, Bovinae, Bos.

Ahora ya tenéis la información sobre la proteína más parecida a la sospechosa 1 y podéis contestar las preguntas que os plateábamos al principio

- 1 ¿De qué proteína se trata?
- 2 ¿En qué organismo se encuentra?
- 3 ¿Cuál es la función de esta proteína?
- 4 ¿Es esta la proteína culpable? ¿Puede ser la causante de la muerte del turista? ¿Porque? ¿O Porque no?
- 5 ¿Tiene la proteína alguna característica remarcable?

Si aún no sabéis donde encontrar la respuesta a cada una de las preguntas aquí tenéis una pequeña ayuda:

- Que información da el campo "Protein"
- Mirad en "Organism", si no sabéis de que organismo se trata no dudéis en copiar el nombre científico y usar google.
- En el apartado de "Function" podréis encontrar lo que se sabe referente a la función de la proteína. También os Puede dar pistas sobre si la proteína pudo estar implicada en la muerte del turista o si tiene alguna característica remarcable.

¿Qué proteína creéis que causó la muerte del turista?

### Paso 5

Usad BLAST para repetir el mismo proceso con las demás proteínas y contestad a las preguntas.

¿Cuál es vuestra conclusión final sobre el asesinato? ¿Cómo murió la víctima?