**BLASTP de 1a6m con PDB para obtener mayor cantidad de proteínas:**

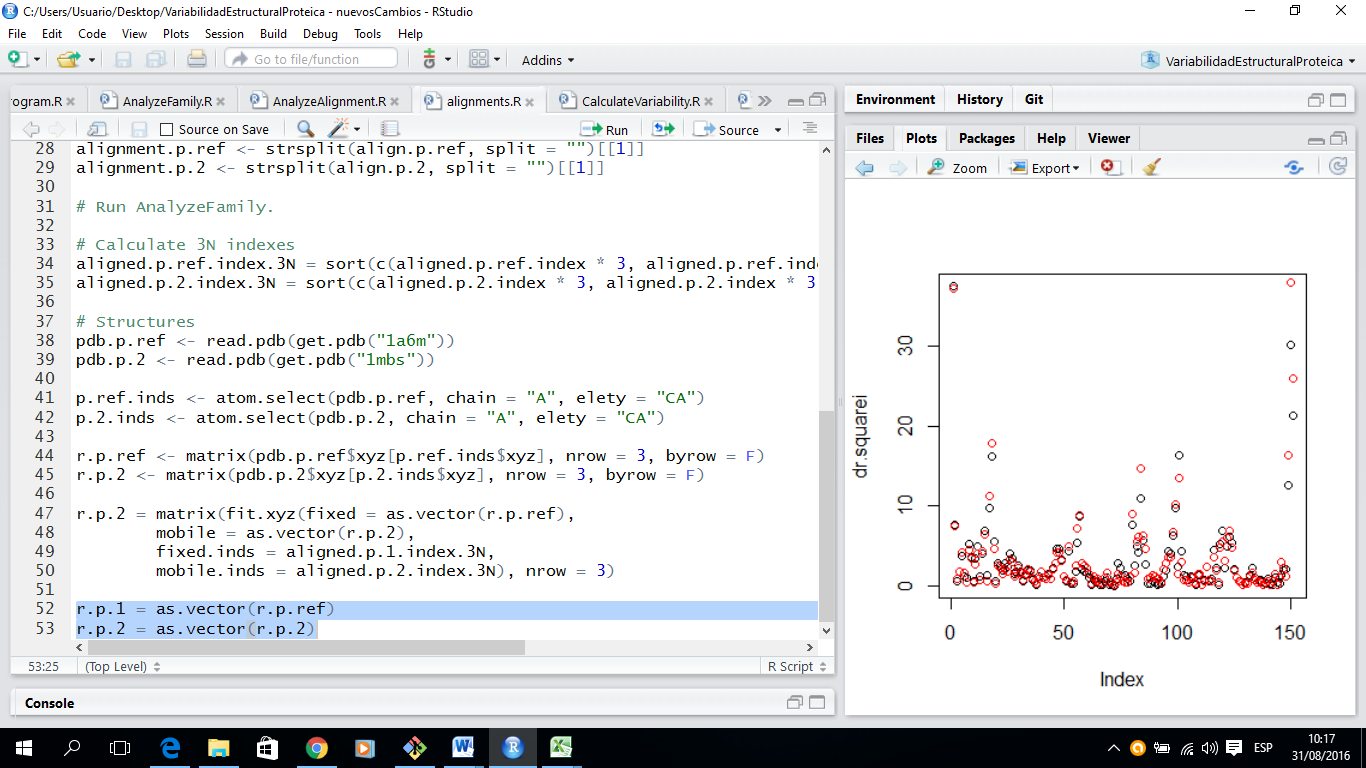
La mayor parte de las proteínas alineadas con alta identidad secuencial (aprox. 70 – 80 %) por BLASTP son mutantes ingenierizadas o proteínas iguales pero expuestas a Co o con otras variantes durante la cristalización. Mucha mioglobina de corazón de caballo con sus variantes.

BLASTP encuentra a la mayor parte de las proteínas con alta %id del conjunto de Homstrad. Encuentra solo unas pocas con menor % id ya que no muestra resultados con %id <25 (aprox.).

Usando psi-BLAST si fui encontrando a todas ya que este algoritmo permite encontrar porteínas más divergentes.

**Alienamientos:**

R permite alinear con MUSCLE con biocLite("muscle"). MUSCLE hace alineamiento secuencial, no estructural, pero para proteínas con alto % id (aprox. 80, lo que buscamos) obtengo el mismo alineamiento que da Homstrad. Más aún, después las roto usando la función de r fit.xyz() y obtengo resultados muy parecidos:



Dr.square i para 1a6m vs 1mbs (%id = 83%).

Obtengo el mismo alineamiento para ambas.

Negro: alineadas con muscle y rotadas con fit.xyz

Rojo: alineadas por Homstrad y no re-rotadas.

También hay en R una función para alinear primero secuencialmente (usando MUSCLE) y después estructuralmente. La función es struct.aln(). No me funciona en mi R. Sin embargo, no se hasta qué punto hace alinemientos estructurales.

* Agregar nuevas familias. EN CURSO.
  + - Plastocianina. (OXIDORESUCTASA). Falta p.ref, CM , correr y análisis
    - **Cistein proteasas. (HIDROLASA).**
    - **Serin proteases. (HIDROLASA).**
    - **Fosfolipasa (HIDROLASA).**
    - Ferredoxinas. (OXIDOREDUCTASA). FALTA ANÁLISIS
    - Glutatión S Trasferasa (TRANSFERASA). FALTA ref CMS Y CORRER. FALTA ANÁLISIS
    - Lactato/malato deshidrogenasa (LIASA). Corriendo. FALTA ANÁLISIS
    - Aspartic proteinasa. (HIDROLASA).
    - Piridin Nucleótido disulfuro oxidoreductasa. (OXIDOREDUCTASA).
    - Triosa fosfato isomerasa. (ISOMERASA).

**En función de su acción catalítica específica, las enzimas se clasifican en 6 grandes grupos o clases:**

* Clase 1: OXIDORREDUCTASAS. (Tengo 2)
* Clase 2: TRANSFERASAS. (Tengo 1)
* Clase 3: HIDROLASAS. (Tengo 4)
* Clase 4: LIASAS. (Tengo 1)
* Clase 5: ISOMERASAS. (Tengo 1)
* Clase 6: LIGASAS. (No tengo ninguna)

Comparaciones en coordenadas cartesianas:

* Con rotación global.
* Con rotación local. Ventana de 15 sitios.
* Empaquetamiento sitio. R0 = 7.5.
* Modificar reporte. Listo
* Agregar empaquetamiento como función. NO
* Agregar nuevas proteínas. Falta
* Agregar nuevas medidas a mainCA. NO
* Agregar MSF y nH a programa de variabilidad. Función escrita, falta implementar.
* Función para hacer blast en camino.

BLAST: con función BLAST.PDB obtengo muy pocos hits para 1a6m y de bajo score.

Nuevo set de proteínas:

* Ingresar pdbid
* Blast
* Filtrar
* Alinear
* Correr programa

Programa de correlaciones múltiples

Archivos que agreguen columna con correlaciones a un archivo.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Family 1 | Family 2 | Family … |
| Exp – mut |  |  |  |
| Exp - strong |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Family 1 | Family 2 | Family … |
| CA |  |  |  |
| CM |  |  |  |

Core

* Sitios sin gaps. No hay demasiados cambios.
* Sitios sin gaps y con vecinos sin gaps (6 o 3, me quedo sin sitios para analizar). Se mantienen las tendencias. Podría ser que algún sitio activo se conserve más… pero solo es marcado en globinas.