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

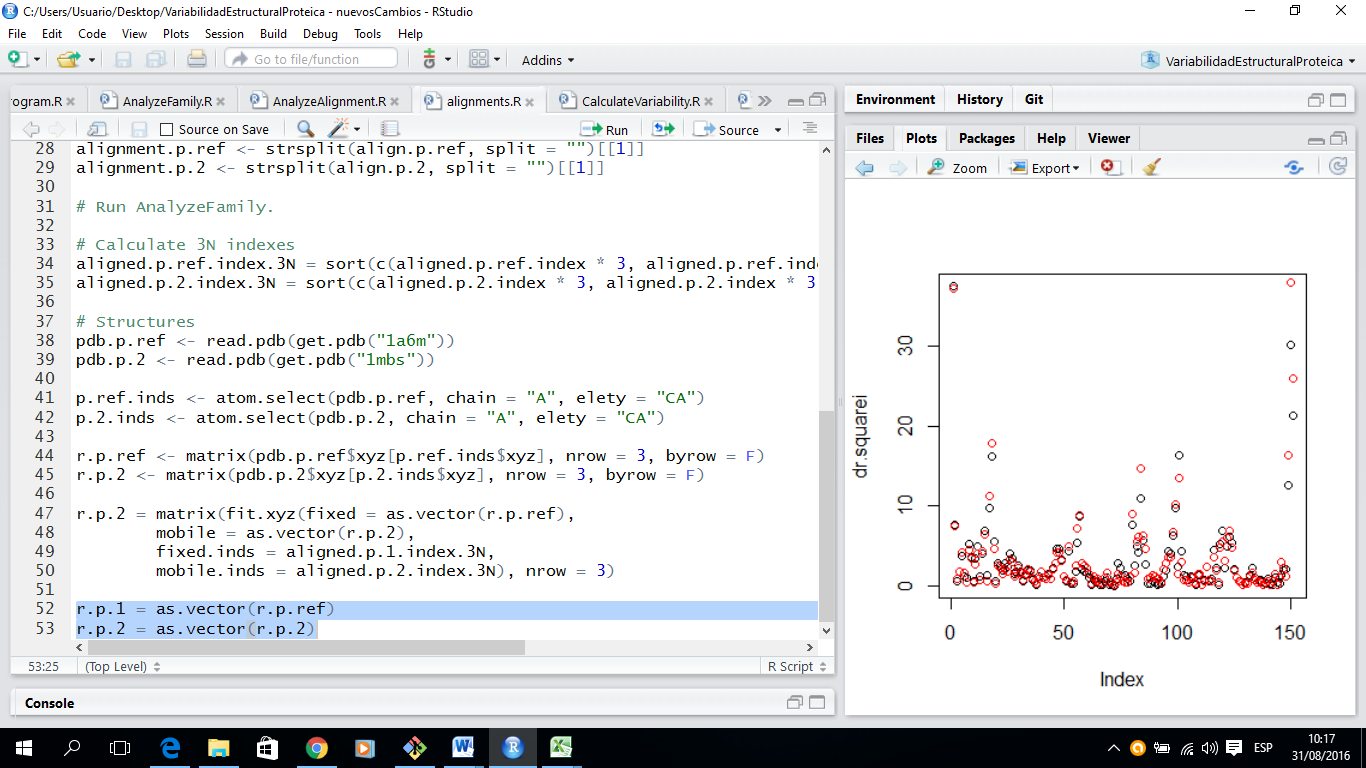
La mayor parte de las proteínas alineadas con alta identidad secuencial (aprox. 70 – 80 %) por BLASTP son mutantes ingenierizadas o proteínas iguales pero expuestas a Co o con otras variantes durante la cristalización. Mucha mioglobina de corazón de caballo con sus variantes.

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.

**Otras familias:**

Ya tengo información de sitios activos. Faltan familias q no son enzimas, el toxinas o de unión a ADN. No encuentro que sitios son importantes funcionalmente.

Mirar alineamientos en búsqueda de más secuencias.

* Volver a correr ANM CA y CM (terminar de escribir el input. listo)
* Informes con sitio activo

Hay q adaptarlos los informes:

* Los archivos que no dicen “globins” o “1a6m” deben decirlo para cuando vengan resultados de otras familias. listo
* Función para cálculo m.da ya está adaptada. Loop listo.
* Escribir archivo con sitios activos de todos. Listo.
* correr el programa de m.da. cambiar de donde lee las p.ref
* En el principal del reporte, hacer un loop para distintas familias q especifique la familia. Listo.
* Escribir los archivos necesarios (faltan 2/familia, lista y p.ref). listo.
* Adaptar los nombres a familias y p.refs. listo.
* Cambiar glicinas a alaninas en chimera para CM otras familias Listo
* Agregar archivos de pdb de pref de cada familia en out/out\_subset\_CA-CM con las gicinas transformadas a alaninas. Agregar también al archivo de coordenadas múltiples. Listo
* Calcular m.da para cada familia. Listo
* Correr Main CM. listo
* Correr reportes.listo
* En min.da sacar hemo-listo
* En reporte cambiar nombres de matrices min.da.listo

Rrm y sh3 listo

Falta snakes toxin sin active info

Cys listo

Fabplisto

Serin prot no corre CM