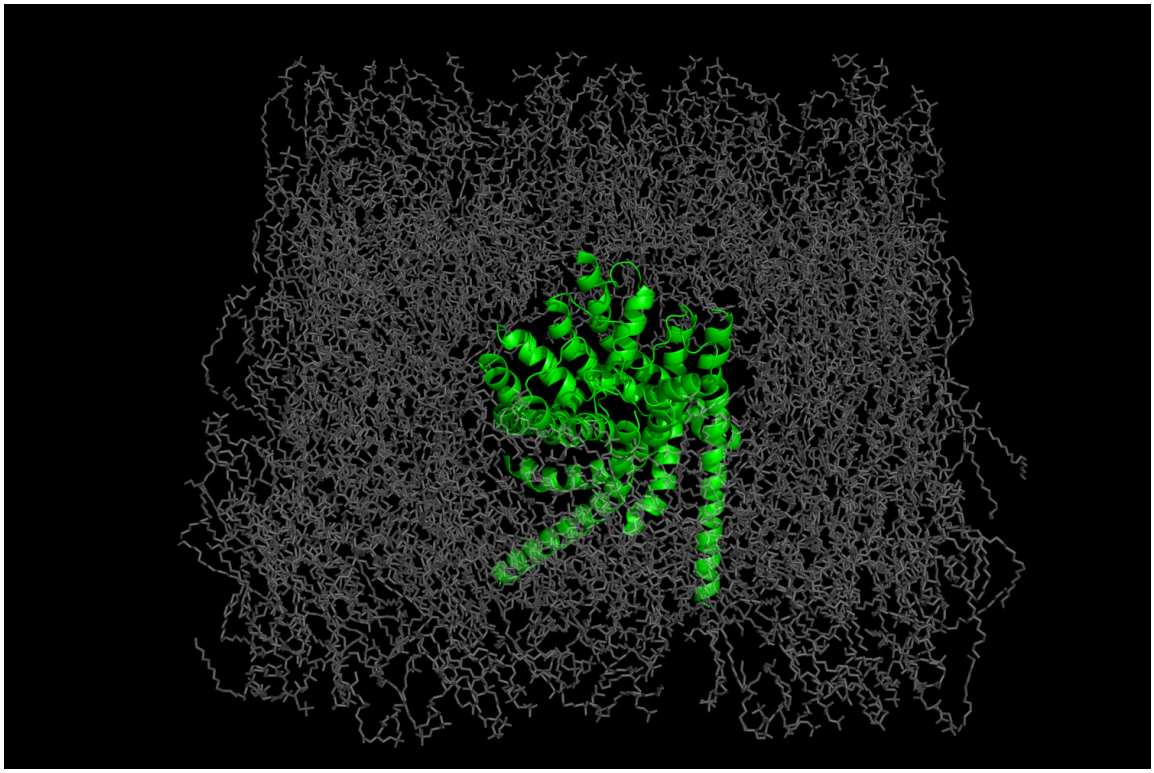
Biología Estructural

PEC 2- Segunda Prueba de Evaluación Continua

Author: Ramón Tamarit Agusti

Titulo: Predicción de la estructura de proteínas Diseño de fármacos



http://www.pcgg.de/report_pictures/Schneider_hergbox.png

Ejercicio 1 (10%): Conociendo la proteína de estudio

La especie Escherichia coli

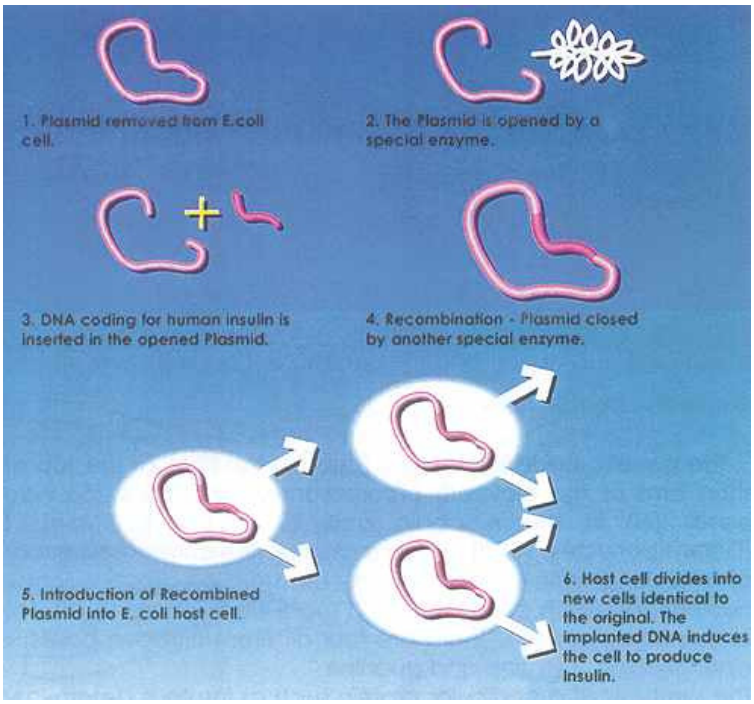
En este ejercicio vamos a estudiar una proteína de la especie *Escherichia coli*. Leed la siguiente entrada de la Wikipedia y resumid los aspectos mas relevantes según vuestra opinión:
http://en.wikipedia.org/wiki/Escherichia_coli

La Escherichia coli (o simplemente E. coli) es uno de los muchos grupos de bacterias que viven en los intestinos de los humanos y en la mayoría de los animales de sangre caliente. Esta bacteria ayuda a mantener el equilibrio de la flora intestinal contra las bacterias nocivas y sintetiza o produce algunas vitaminas. No obstante, existen cientos de tipos o cepas de bacterias E-coli. Las distintas cepas de E. coli tienen diferentes características distintivas, y algunas pueden incluso crearnos graves problemas de salud.

Lo más relevante a efectos de nuestro trabajo es:

- Dada su disponibilidad y facilidad de cultivo, hace que sea una bacteria utilizada frecuentemente en experimentos de genética y biotecnología molecular. "*En algún momento, antes de 1982, un virus desconocido que ataca a las bacterias pasó una parte de su código genético a la E. Coli, permitiendo que algunas cepas produzcan la toxina Shiga. Este veneno mortal causa la grave infección alimentaria que da lugar a los terribles síntomas descritos.*".

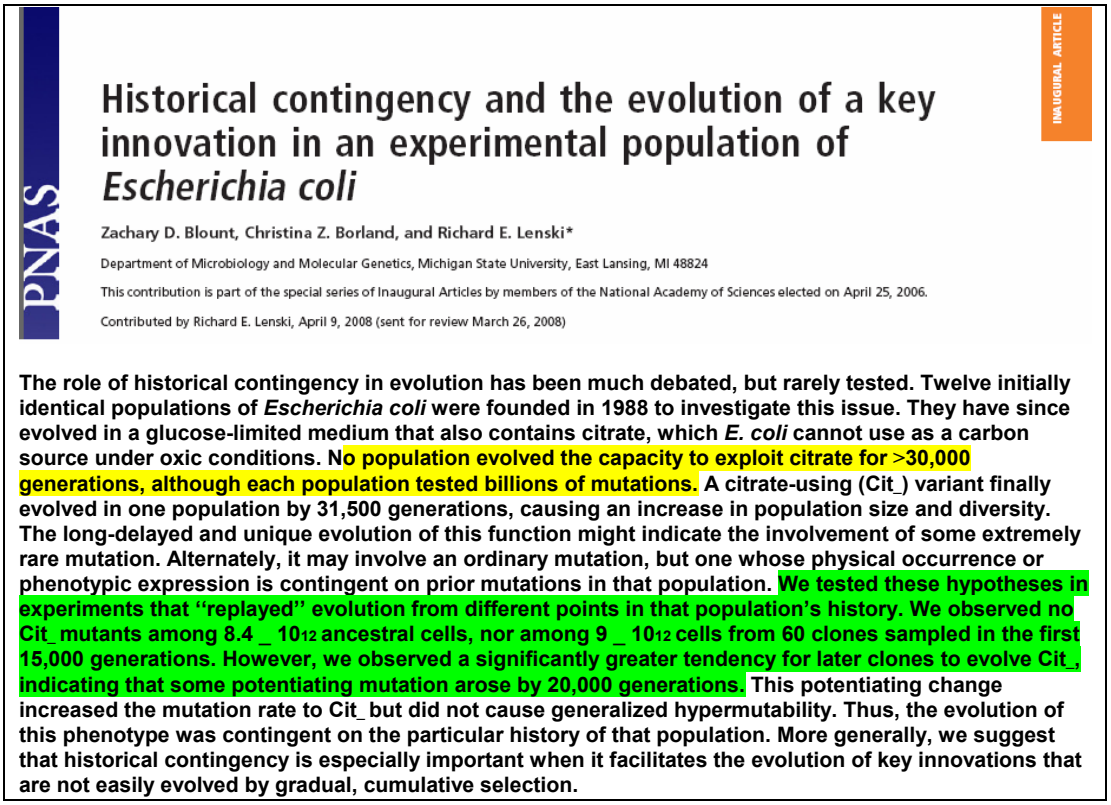
http://www.cmdr.ubc.ca/bfindlaypict.jpg



- E. Coli como organismo modelo Esta bacteria nos puede servir para introducir mutaciones en genes o simplemente como fábrica de proteínas. Las cepas cultivadas (por ejemplo, E. coli K12) se adaptan bien al medio ambiente del laboratorio, y, a diferencia de las cepas de tipo salvaje, han perdido su capacidad de crear infecciones en el intestino.

- La E. coli forma parte integrante de los primeros experimentos para entender la genética fagos, Seymour Benzer, utilizo E. coli y fagos T4 para explicar la topografía de la estructura de genes.

- Con cepas de E. coli es posible realizar experimentos que han permitido la observación directa de cambios evolutivos en el laboratorio.



PNAS

INAUGURAL ARTICLE

Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*

Zachary D. Blount, Christina Z. Borland, and Richard E. Lenski*

Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected on April 25, 2006.

Contributed by Richard E. Lenski, April 9, 2008 (sent for review March 26, 2008)

The role of historical contingency in evolution has been much debated, but rarely tested. Twelve initially identical populations of *Escherichia coli* were founded in 1988 to investigate this issue. They have since evolved in a glucose-limited medium that also contains citrate, which *E. coli* cannot use as a carbon source under oxic conditions. No population evolved the capacity to exploit citrate for >30,000 generations, although each population tested billions of mutations. A citrate-using (Cit_) variant finally evolved in one population by 31,500 generations, causing an increase in population size and diversity. The long-delayed and unique evolution of this function might indicate the involvement of some extremely rare mutation. Alternately, it may involve an ordinary mutation, but one whose physical occurrence or phenotypic expression is contingent on prior mutations in that population. We tested these hypotheses in experiments that "replayed" evolution from different points in that population's history. We observed no Cit_ mutants among 8.4 _ 10^{12} ancestral cells, nor among 9 _ 10^{12} cells from 60 clones sampled in the first 15,000 generations. However, we observed a significantly greater tendency for later clones to evolve Cit_, indicating that some potentiating mutation arose by 20,000 generations. This potentiating change increased the mutation rate to Cit_ but did not cause generalized hypermutability. Thus, the evolution of this phenotype was contingent on the particular history of that population. More generally, we suggest that historical contingency is especially important when it facilitates the evolution of key innovations that are not easily evolved by gradual, cumulative selection.

Interpro

INTERPRO es un portal centralizado que recoge la información de la mayoría de bases de datos conocidas de familias de proteínas. Leed la documentación (**http://www.ebi.ac.uk/interpro/tutorial.html**) y resumid en pocas líneas cuáles son sus funciones principales:
http://www.ebi.ac.uk/interpro/

http://nar.oxfordjournals.org/cgi/content/abstract/37/suppl_1/D211

¿Qué es interpro?

InterPro es una base de datos de anotaciones con información sobre la función y la secuencia de la anotación. Siendo un recurso integrado de familias de proteínas, dominios y sitios.

Agrupación de las secuencias

Las secuencias se clasifican en grupos por huellas o métodos (traducción literal de "protein signaturas" or "methods"). Los grupos representan superfamilias, familias o subfamilias de secuencias o grupos de secuencias que tienen una o más secuencias características en común. Los grupos pueden definirse (o tipificarse) como familias, dominios, regiones, repeticiones o sitios.

La función biológica de las secuencias dentro de cualquier grupo puede limitarse a: un único proceso biológico; una amplia gama de funciones; o el grupo puede ser funcionalmente no caracterizado.

Bases de datos miembro

Combina un número de bases de datos que utiliza diferente metodología y diversos grados de información biológica de proteínas bien caracterizadas, provenientes de firmas de proteínas. Las bases de datos que la componen son Gene3D, PANTHER, Pfam, PIRSF, PRINTS, ProDom, PROSITE, SMART, SUPERFAMILY y TIGRFAMs. Las entradas son revisadas manualmente e integradas dentro de InterPro en un formato unificado.

Creación de las entradas de InterPro

Los registros de InterPro (IPR) se crean a partir de las nuevas huellas de proteína proporcionadas por las bases de datos participantes.

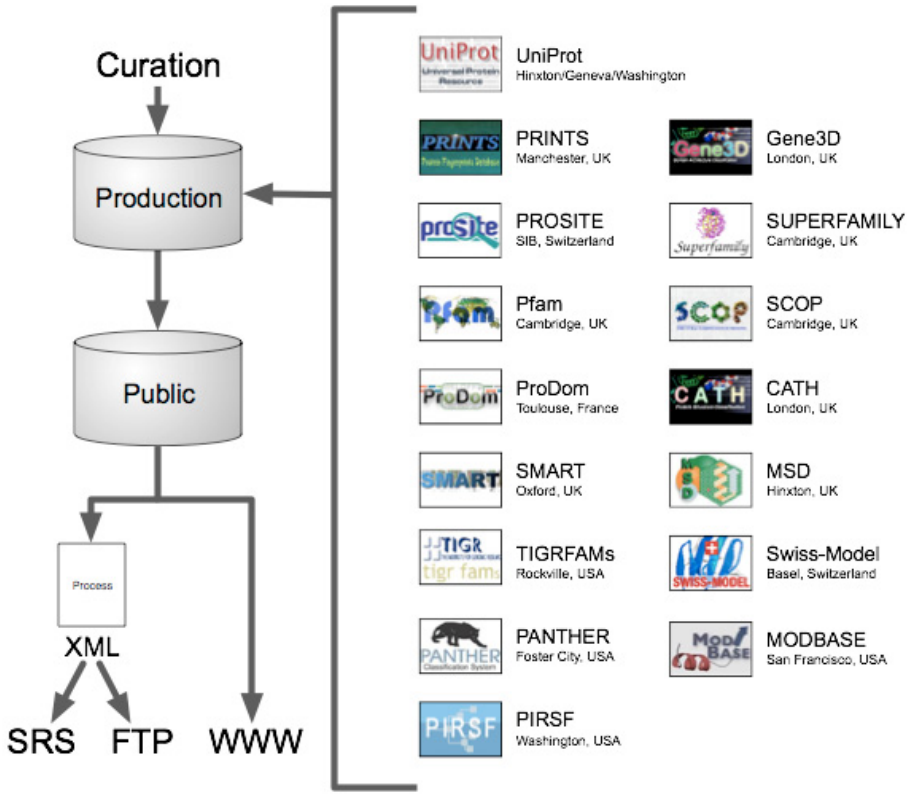
Primero se comparan a todas las entradas en UniProtKB. Las huellas que describen la misma familia de proteínas o dominio proteico se agrupan en una misma entrada de InterPro. Cada entrada tiene un código único, y se genera un resumen que describe las características de las proteínas asociadas con la entrada, las referencias bibliográficas y enlaces a la base de datos de miembro.

- Si la nueva firma de proteína es único, en otras palabras, identifica a un grupo de secuencias que no están actualmente en InterPro, se le asigna un nuevo número de InterPro y se creará una nueva entrada,
- Si la nueva firma se identifica con secuencias de proteínas existentes en InterPro pero no se solapa con las firmas correspondientes de las proteínas, se creará igualmente una nueva entrada InterPro.

Si la nueva firma proteína se encuentra asociada o solapada con una o más firmas que, se reclasificará como:

- Si se solapa con al menos el 75% de las secuencias se fusionará su entrada con el resto, y no se creará una nueva entrada.
- Sino, pasara a ser PARENT o CHILD de una entrada existente (creándose una nueva entrada) o bien CONTAINS o FOUND IN de entradas existentes (también se crea una nueva entrada)

Esquema del proceso de creación de las entradas de InterPro



Fuente: http://www.ebi.ac.uk/interpro/project_outlines.html

I've read through your papers and I don't understand how the signatures and relationship fields are created. For example, do you have an automatic algorithm that groups together signatures from multiple databases into a family? Or is it that you manually curate the signatures and relationships?

1) We receive the signatures from the member databases and the process of grouping them into families or domains is entirely manually curated.
2) We use the matches of the new signatures against UniProt proteins to determine whether the matches overlap in position on the sequence and the signatures match the same set of proteins, which is the criterion for grouping signatures into the same InterPro entries.
3) Where one signature matches a subset of another, then we curate relationships between the entries. There are some automatic methods for producing the matches and overlap files, but the ultimate decision for grouping and for determining relationships comes from a biologist.

Fuente: http://www.ebi.ac.uk/interpro/User-FAQ-InterPro.html

Los tipos de InterPro

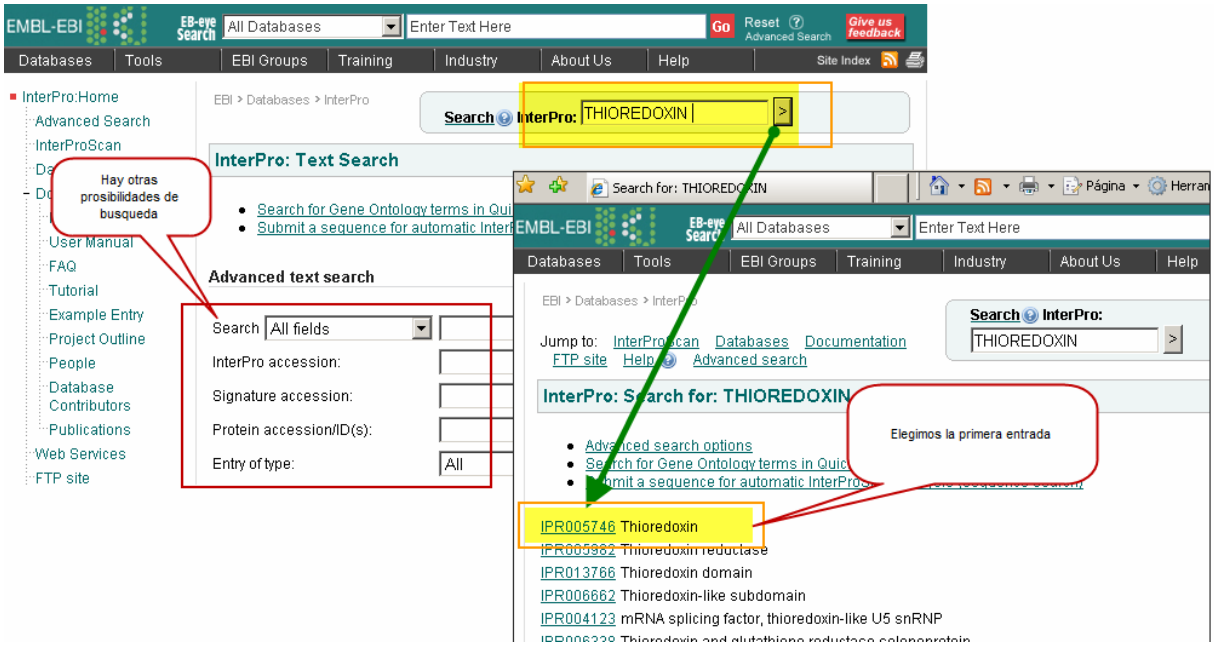
El tipo define a cada entrada como: de familia, de dominio, de región, o de sitio. Los sitios se subclasifican en: bien conservados Sitios, sitios activos y sitios de unión o PTMs (modificaciones post-translacionales).

- Familia: es un grupo de proteínas relacionadas evolutivamente y que tienen uno o más dominios o repeticiones en común. Una familia puede contener un 'motivo' que la defina. Las familias se pueden dividir según relaciones de 'padres' e 'hijos', en familias y subfamilias (PARENT/CHILD).
- *Dominio*: un dominio es una unidad estructural independiente. Una entrada de InterPro del tipo dominio se puede usar como diagnóstico de la presencia de un dominio, pero no tiene por qué definir correctamente los límites del dominio. En las entradas de dominios, hay un campo de CONTAINS / FOUND IN ' para indicar si alguna familia ha sido caracterizada por este dominio.
- *Repetición*: es una corta región que no tiene entidad estructural independiente, es decir, se requiere la presencia de varias repeticiones para dar lugar a un dominio.
- *Sitios de unión Modificación post-transduccional (PTM)*: se refiere a motivos de secuencia que son reconocidos en la célula para que se produzcan PTMs sobre la proteína. Las proteínas que son agrupadas de acuerdo a un PTM no tienen por qué compartir un origen evolutivo.
- Sitios de unión: son sitios responsables de la unión de ligandos que por sí mismos no son sustratos de reacción enzimática.
- Sitios activos: son sitios de unión en los que se produce una reacción enzimática.

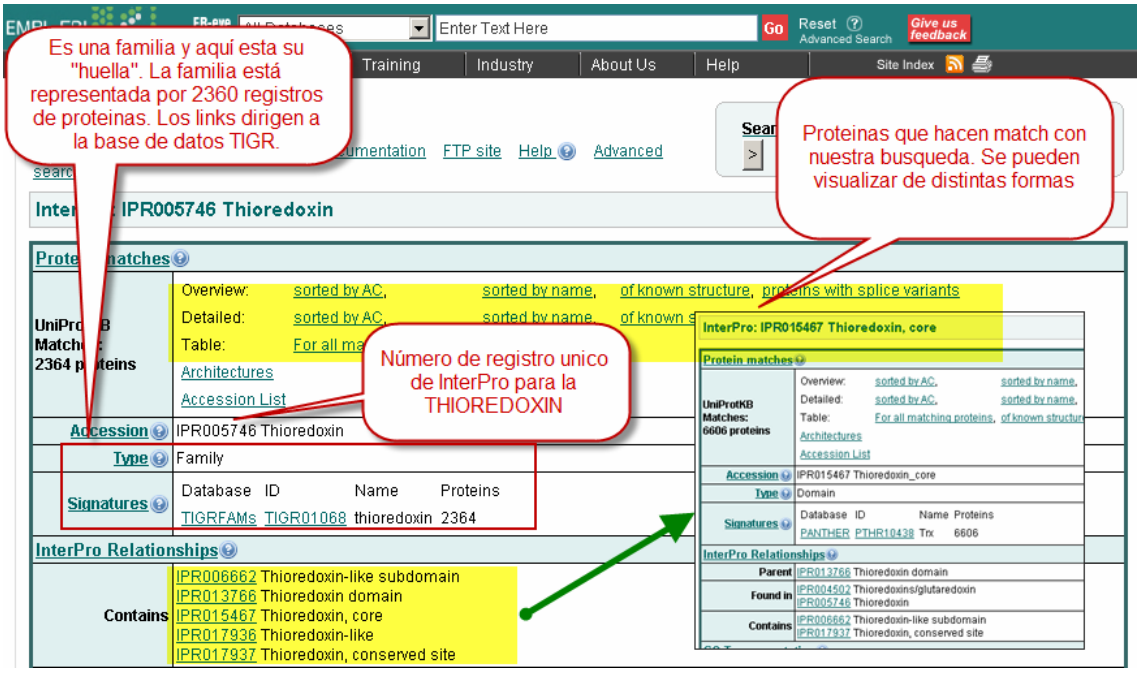
THIOREDOXIN

La proteína THIOREDOXIN en *E.coli* se usará en este ejercicio para practicar diferentes modelos de predicción computacional de estructura. Buscad la proteína THIOREDOXIN en INTERPRO y resumid la información en esta entrada que sea más relevante:

Desde la pagina de inicio de InterPro se pueden iniciar búsquedas de distintos tipos, en concreto iniciamos una búsqueda general por texto.



El resultado de la búsqueda nos devuelve un listado de entradas, seleccionamos la más genérica, y el link nos lleva a la página de la entrada.

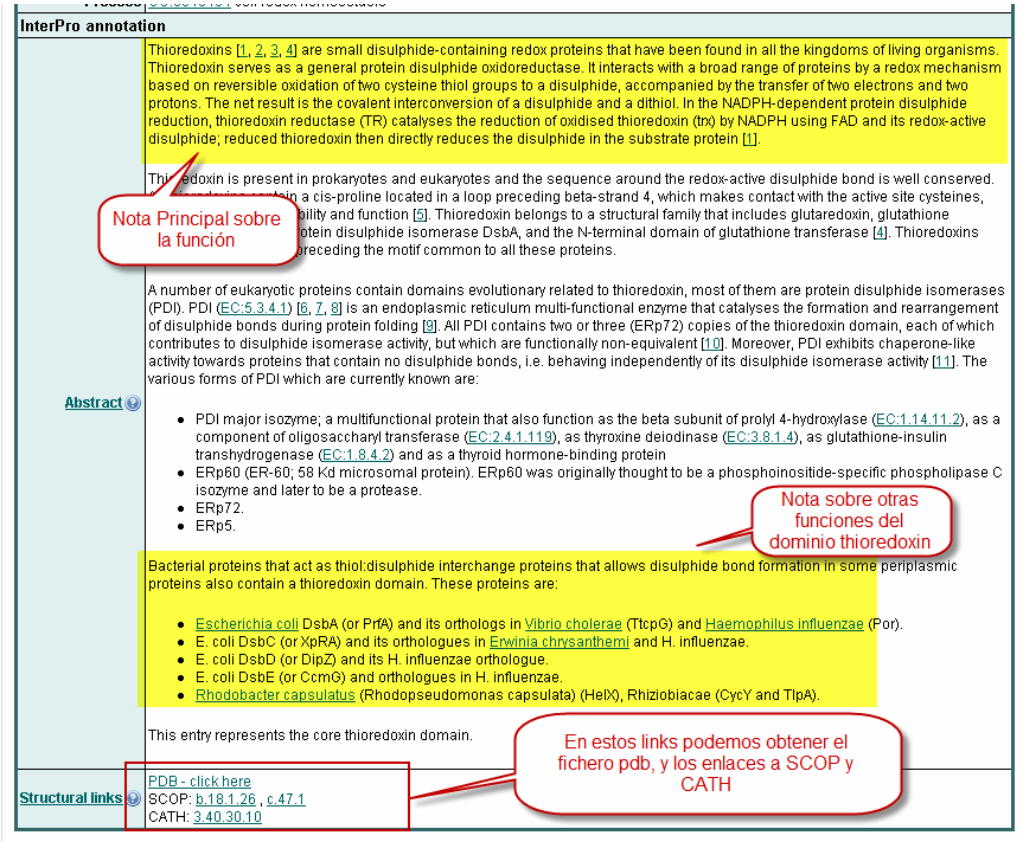


Fácilmente se visualizan en la cabecera: El tipo, en este caso es una familia, los dominios, subdominios y sitios contenidos (están bajo el epígrafe de "Contains"). Los links nos llevan siempre a paginas con más información, el link de "Thiredoxin core" nos lleva al pagina descriptiva del dominio (observamos que en este caso hay entradas "parent", Found in" y Contanins").

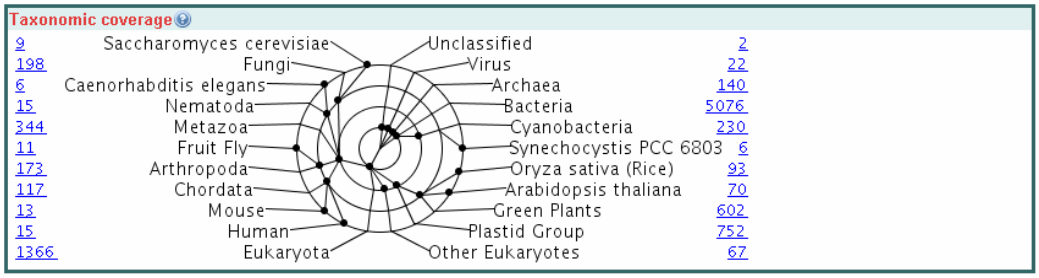
El apartado de “signatures” nos proporciona los liks a las bases de datos de las cuales ha salido las huella.

En el apartado de “Interpro annotatión”, tenemos el “abstract” que resume las funciones biológicas principales y comentarios sobre los dominios evolutivos.

Los “structural links” son enlaces al fichero de coordenadas pdb y a las bases de datos de clasificación estructural.



La cobertura taxonómica resume la distribución por especies de las entradas de secuencia relacionadas con esta familia.

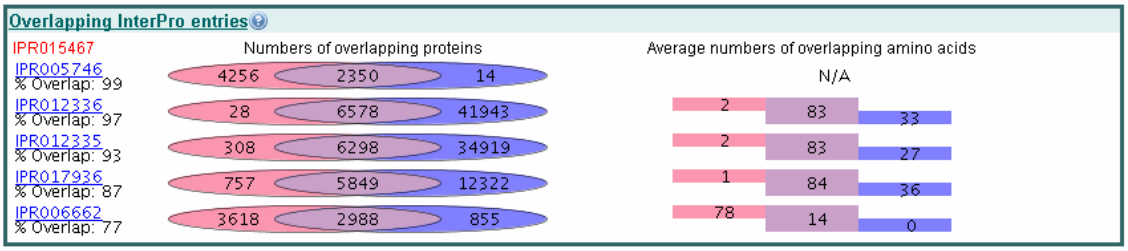


Taxonomy Coverage

The Taxonomy Coverage aims to provide 'at a glance' view of the taxonomic range of the sequences associated with each InterPro entry and the number of sequences associated with each lineage. The taxonomic lineages are 'clickable' and provide a pop-up, which displays the tax-ID, the taxonomy and taxonomic subgroup(s)/species having matches to proteins, the protein match counts and a FASTA link. Clicking on the taxonomy or taxonomic subgroup(s)/species links to the protein overview matches for the selected taxonomy. Clicking on the FASTA box will download the complete set of FASTA sequences for the selected taxonomy of the entry.

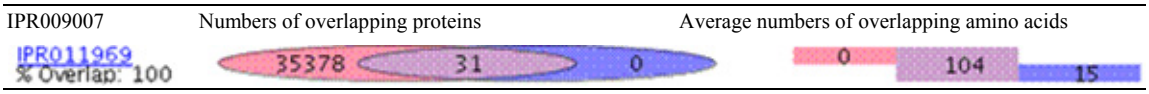
The lineages were carefully selected to provide a view of the major groups of organisms. The circular display has the taxonomy-tree root as its centre. Selected model organisms populate the outer most circle. Nodes of the taxonomy-tree are placed on the inner circles. Radial lines lead to the description for each node. No significance is attached to the position of the node on a particular inner-circle, other than convenience, though some attempt has been made to group nodes. The nodes themselves are either true taxonomy nodes and have a NCBI taxonomy number or are artificial nodes created for this display; of which there are three: Unclassified, Other Eukaryotes (Non-Metazoa) and the Plastid Group.

Como ya comentamos, el solapamiento nos ofrece una idea de cual es la compartición de registros de secuencia (de proteína) con otras entradas.



Overlapping InterPro Entries

This section displays entries that share more than 70% of their proteins. Such overlaps define PARENT/CHILD and CONTAINS/FOUND IN relationships between InterPro entries.



In the above example, InterPro entry IPR011969 contains proteins which are also found in IPR009007 as a result of the protein signatures of the two entries overlapping.

The two entries have been compared firstly by counting the number of proteins which are common to both, the results of which are displayed in the Venn diagram on the left, and secondly by calculating the average overlap of the protein signatures, in amino acids, with the results displayed in the bar diagram on the right.

Venn diagram display of the overlap of proteins common to both entries:

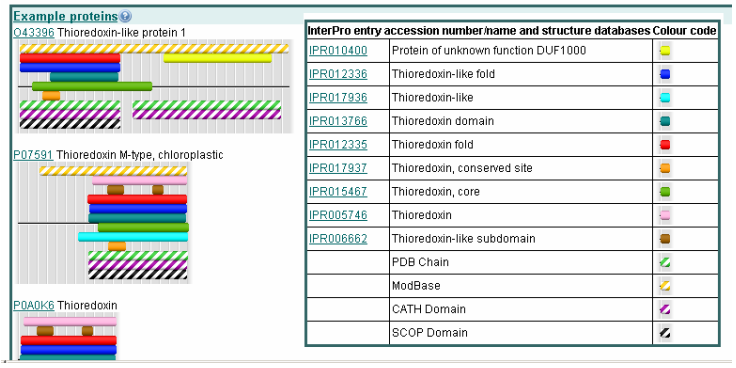
- The purple intersection contains the number of overlapping proteins common to both IPR009007 and IPR011969, which is 31 in this case.
- The pink section on the left is the number of proteins found in IPR009007 but not IPR011969, which is 35378.
- The blue section on the right is the number of proteins found in IPR011969 but not IPR009007, which is 0; i.e. all proteins associated with IPR011969 occur in IPR009007.

Bar diagram display of the average amino acid overlap between the protein signatures:

- The average number of amino acids overlapping in the sequences of the 31 proteins common to both entries is then calculated, with the results displayed in the bar diagram on the right. The bar diagram display is only shown for 'Domain - Domain' relationships.
- The purple segment in the middle shows the average number of amino acids overlapping between IPR009007 and IPR011969 for the 31 proteins, in this case 104.
- The pink segment shows the average number of amino acids found in IPR009007, but not IPR011969, for the 31 proteins, which is 0.
- The blue segment shows the average number of amino acids found in IPR011969, but not IPR009007, for the 31 proteins, which is 15.

The results of these comparisons are used to calculate the percentage overlap score, with all scores greater than 70% displayed on the InterPro pages. In this example, since all proteins found in IPR011969 are also found in IPR009007, and all the amino acids from IPR009007 overlap with those from IPR011969, the percentage overlap score is 100%.

Las “Example proteins” son links a las entradas de proteína que pueden servir como “ejemplo”. Los códigos de colores nos muestran las zonas de la proteína que corresponden a las clasificaciones.



Por último tenemos dos apartados con enlaces a PubMed de artículos relacionados y lecturas adicionales.

Publications

1. Holmgren A.
   Thioredoxin.
   Annu. Rev. Biochem. 54 237-71 1985 [PubMed: 3896121]
   http://dx.doi.org/10.1146/annurev.bi.54.070185.001321
2. Holmgren A.
   Thioredoxin and glutaredoxin systems.
   J. Biol. Chem. 264 13963-6 1989 [PubMed: 2668278]
   http://intl.jbc.org/cgi/reprint/264/24/13963.pdf
3. Holmgren A.
   Thioredoxin structure and mechanism: conformational changes on oxidation of the active-site sulfhydryls to a disulfide.
   Structure 3 239-43 1995 [PubMed: 7788289]
   http://dx.doi.org/10.1016/S0969-2126(01)00153-8
4. Martin JL.
   Thioredoxin--a fold for all reasons.

Additional Reading

- Wahl MC, Irmler A, Hecker B, Schirmer RH, Becker K.
  Comparative structural analysis of oxidized and reduced thioredoxin from Drosophila melanogaster.
  J. Mol. Biol. 345 2005 1119-30 [PubMed: 15644209]
  http://dx.doi.org/10.1016/j.jmb.2004.11.004
- Brieba LG, Kokoska RJ, Bebenek K, Kunkel TA, Ellenberger T.
  A lysine residue in the fingers subdomain of T7 DNA polymerase modulates the miscoding potential of 8-oxo-7,8-dihydroguanosine.
  Structure 13 2005 1653-9 [PubMed: 16271888]
  http://dx.doi.org/10.1016/j.str.2005.07.020
- Song J, Tyler RC, Wrobel RL, Frederick RO, Vojtek FC, Jeon WB, Lee MS, Markley JL.
  Solution structure of At3g04780.1-des15, an Arabidopsis thaliana ortholog of the C-terminal domain of human thioredoxin-like protein.
  Protein Sci. 14 2005 1059-63 [PubMed: 15741346]
  http://dx.doi.org/10.1110/ps.041246805
- Coudevylle N, Thureau A, Hemmerlin C, Gelhaye E, Jacquot JP, Cung MT.
  Solution structure of a natural CPPC active site variant, the reduced form of thioredoxin h1 from poplar.
  Biochemistry 44 2005 2001-8 [PubMed: 15697225]

Clase estructural de THIOREDOXIN. CATH y SCOP

Buscad en la página anterior, los enlaces de esta proteína a las bases de datos CATH y SCOP (vistas en la primera prueba) y decidid a qué clase estructural pertenece:

Structural Classification of Proteins

Superfamily: Thioredoxin-like

Superfamily

Lineage:

1. Root: scop
2. Class: Alpha and beta proteins (a/b) [51349]
   Mainly parallel beta sheets (beta-alpha-beta units)
3. Fold: Thioredoxin fold [52832]
   core: 3 layers, a/b/a; mixed beta-sheet of 4 strands, order 4312; strand 3 is antiparallel to the rest
4. Superfamily: Thioredoxin-like [52833]
   Superfamily

Families:

1. Thioltransferase [52834] (31)
2. PDI-like [52849] (6)
   duplication: contains two tandem repeats of this fold
3. Calsequestrin [52855] (1)

Homologous Superfamily: 3.40.30.10

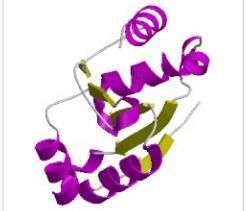
Glutaredoxin

Classification Lineage (3.40.30.10)

CATH Code	Level Description	Links
3	Alpha Beta	
3.40	3-Layer(aba) Sandwich	
3.40.30	Glutaredoxin	
3.40.30.10	Glutaredoxin	Gene3D

Summary of Child Nodes

-	-	-	-	85	133	187	340	1137



Representative domain: 2fwhA00

Es una alfa/beta proteína:

Las alfa/beta proteínas contienen hélices alfa y láminas beta (habitualmente paralelas). Las hélices y las hebras de la lámina se alternan a lo largo de la secuencia. Existen algunos subtipos clásicos:

- Con lámina cerrada (barriles alfa/beta). Láminas de 7 u 8 hebras que forman un barril sobre cuya cara externa se empaquetan las hélices. Si hay más hebras, aparece un hueco central que se rellena con un bucle o una hélice. Las hebras son esencialmente apolares. Muchas son enzimas, en cuyo caso el sitio activo está en los bucles del lado carboxilo de las hebras.
- Con lámina abierta. Con frecuencia tienen menos de 7 hebras pero a veces tienen más. Casi siempre dan lugar a tres capas que suelen ser alfa/beta/alfa, aunque existen disposiciones de tipo beta/beta/alfa. A veces sólo hay dos capas (alfa/beta), como en las herraduras alfa/beta.

Ejercicio 2 (70%): Predicción de estructura (homología)

Aunque actualmente ya se conoce su estructura, vamos a usar un método de predicción de estructura por homología para obtener cuál podría ser el modelo tridimensional más adecuado para la proteína THIOREDOXIN en *E.coli*. Esta metodología sería aplicable a aquellas proteínas cuya estructura se desconoce experimentalmente.
Generalmente, los programas de predicción por homología buscan mediante alineamiento de secuencias cual es la proteína más parecida a la que tenemos para inferir su estructura a partir de este patrón. En nuestro caso, vamos a usar la proteína humana para ver si podemos reconstruir la estructura de la proteína homologa bacteriana.
Esta es la secuencia de aminoácidos de la proteína THIOREDOXIN en *E. coli* (adjunta también al enunciado, trabajaremos con una de las dos subunidades):
>THIOREDOXIN|E.coli SDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKL TVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDAN LA

La proteína 3TRX. Fichero de secuencia y estructura en pdb.

La estructura de la proteína humana es conocida y se puede acceder en la base de datos PDB: buscad la entrada 3TRX en PDB y rescatad la secuencia de aminoácidos de dicha proteína humana (formato FASTA). Guardad también esta entrada PDB en un fichero .pdb para visualizarla más tarde.

La entrada de PDB nos advierte de las diferencias y parecidos de las dos thioredoxinas, reproduzco parte de la entrada:

Fuente: http://www.ncbi.nlm.nih.gov/pubmed/2001356?dopt=Abstract

> The overall structure is similar to the crystal and NMR structures of oxidized [Katti, S. K., LeMaster, D. M., & Eklund, H. (1990) J. Mol. Biol. 212, 167-184] and reduced [Dyson, J. H., Gippert, G. P., Case, D. A., Holmgren, A., & Wright, P. (1990) Biochemistry 29, 4129-4136] Escherichia coli thioredoxin, respectively, despite the moderate 25% amino acid sequence homology. Several differences, however, can be noted. The human alpha 1 helix is a full turn longer than the corresponding helix in E. coli thioredoxin and is characterized by a more regular helical geometry. The helix labeled alpha 3 in human thioredoxin has its counterpart in the 3(10) helix of the E. coli protein and is also longer in the human protein. In contrast to these structural differences, the conformation of the active site loop in both proteins is very similar, reflecting the perfect sequence identity for a stretch of eight amino acid residues around the redox-active cysteines.

Como previsión, esto nos lleva a una pequeña búsqueda en PubMed. Algunas referencias que nos puedan ayudar. En principio supondremos que aunque ambas proteínas tienen un índice de homología en la secuencia de aminoácidos del 25%, al menos el centro activo estará “intacto”.

En http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=10049365 encontramos que el centro activo esta formado por un puente disulfuro entre dos cisteinas.

En esta otra cita http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=19181668 hay una referencia concluyente:

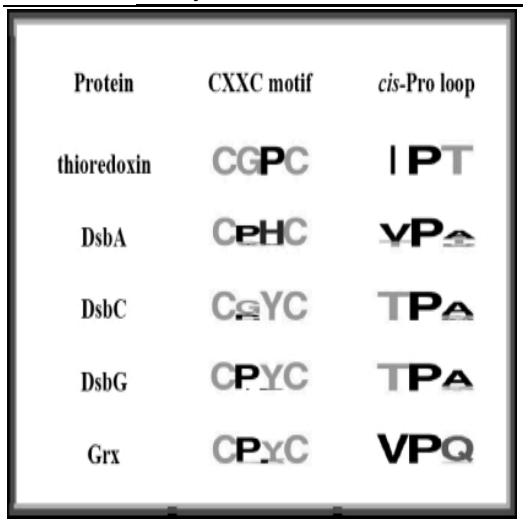
Members of the thioredoxin superfamily share two features in common: they contain a short sequence motif that includes a Cys-X_1-X_2-Cys sequence (the active site) and an overall structure containing this motif that corresponds to what is called a thioredoxin-like fold (29). The latter structural features have been determined directly by X-ray crystallography for some members of the family and by structural modelling in others (29). The actual role of each of these proteins in the cell is partly determined by the redox potential of the protein and partly by the direction of the electron transport pathway it participates in, showing the importance of both kinetic and thermodynamic factors. Thioredoxin 1 of *E. coli*, with a redox potential of −270 mV, is a major reductant in the cytoplasm; it is important for the reduction of such cytoplasmic enzymes as ribonucleotide reductase (26). DsbA, a periplasmic protein with a redox potential of −122 mV, is highly oxidizing; it is required for disulfide bond formation in the cell envelope (6).

Esta otra cita, es mucho más explicita con la información, reproduzco unos párrafos e imágenes interesantes.

Fuente: http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=19181668

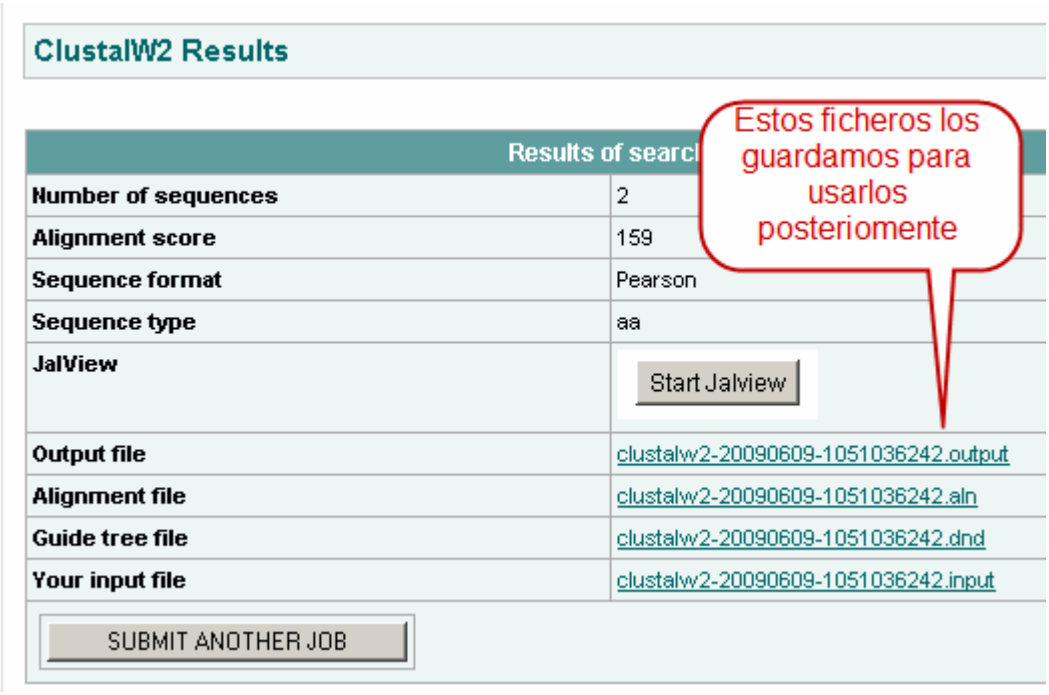
The ubiquitous thioredoxin fold proteins catalyze oxidation, reduction, or disulfide exchange reactions depending on their redox properties. They also play vital roles in protein folding, redox control, and disease. Here, we have shown that a single residue strongly modifies both the redox properties of thioredoxin fold proteins and their ability to interact with substrates. This residue is adjacent in three-dimensional space to the characteristic *CXXC* active site motif of thioredoxin fold proteins but distant in sequence. This residue is just N-terminal to the conservative *cis*-proline. It is isoleucine 75 in the case of thioredoxin. Our findings support the conclusion that a very small percentage of the amino acid residues of thioredoxin-related proteins are capable of dictating the functions of these proteins

The thioredoxin fold is the core scaffold of numerous proteins that control disulfide redox activity in the cell (1-3). These redox proteins share very little sequence homology, but all of them incorporate the four-stranded β-sheet, three flanking α-helices, and the redox-active *CXXC* motif of the TRX[5] fold (Fig. 1*A*). The archetype of the family is thioredoxin (4), a disulfide reductase that maintains a reducing cytosolic environment. Other TRX fold redox proteins include the Dsb proteins (1), which regulate the formation of disulfide bonds in prokaryotes, and protein-disulfide isomerase (5), which catalyzes the oxidation and shuffling of disulfides in the endoplasmic reticulum of eukaryotic cells.



Sequence of the CXXC motif and cisPro loop
The Sequences of each protein family surrounding CXXC and cis-proline were compiled and used to generate sequence logos. The size of the amino acid single-letter code is proportional to the occurrence of that amino acid at that position. To determine the conservation of the residues in the CXXC and cis-proline loop in these proteins, we analyzed the proteins from genomes that are as divergent as possible; however, we wanted to avoid comparing proteins that have been evolving over different evolutionary time frames. Thioredoxin and glutaredoxin, for instance, are present in eukaryotes, bacteria, and archaea, and therefore have been evolving for at least 3.8 billion years (53), whereas DsbC and DsbG are restricted to proteobacteria and have probably been evolving for at least ~0.5 billion years (23, 53). Thus we restricted our comparison to genomes that contain an orthologue to DsbC. We used our previous alignment of all species-specific DsbC sequences available in GenBank™ (23) to obtain the list of organisms that contain DsbC. We then obtained the sequence of the individual thioredoxin, DsbA, DsbG, and glutaredoxin orthologues present in these individual genomes using blast by searching with the E. coli homologue.

Antes de ejecutar la predicción por homología, utilizad el programa CLUSTALW para alinear las dos proteínas (THIOREDOXIN en *E.coli* y en *H.Sapiens*). Explorad el alineamiento y decidid si son dos proteínas más o menos parecidas
http://www.ebi.ac.uk/clustalw/



El alineamiento con CLUSTALW usando los valores por defecto resulta en el siguiente alineamiento:

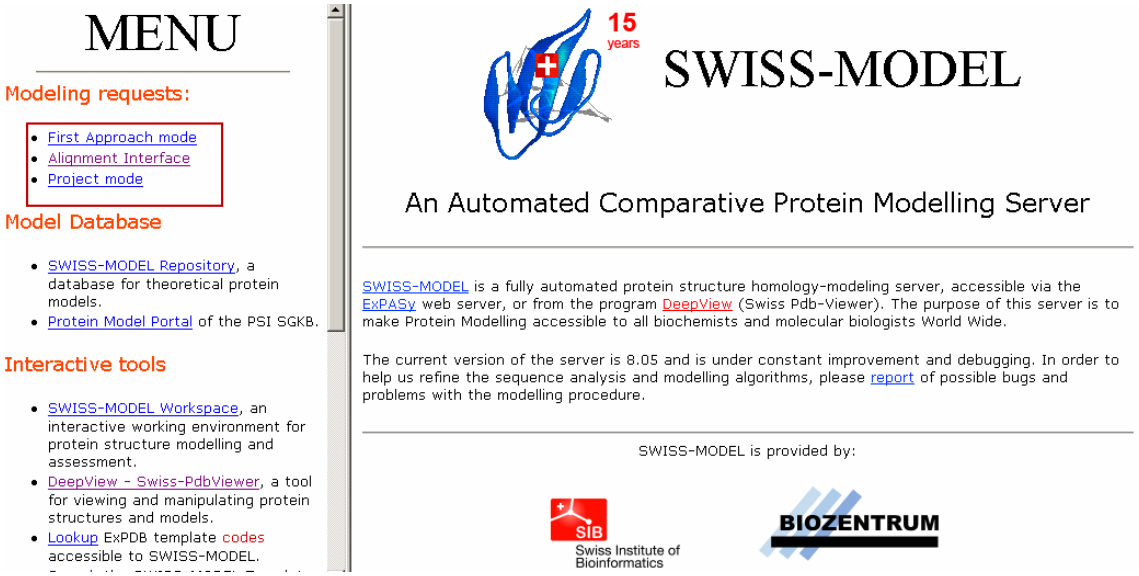
```
CLUSTAL 2.0.10 multiple sequence alignment

THIOREDOXIN|E.coli      SDKIIHLTDDSFDTDVLKADG--AILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKL 58
3TRX_A|HUMAN            --MVKQIESKTAFQEALDAAGDKLVVVDFSATWCGPCKMIKPFFHSLSEKYS-NVIFLEV 57
                           : :: ..:   :.*.* *   ::*** * ******** *::..::::*. :: . ::

THIOREDOXIN|E.coli      NIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLA--- 108
3TRX_A|HUMAN            DVDDCQDVASECEVKCTPTFQFFKKG----QKVGEFS-GANKEKLEATINELV 105
                        ::*:  ..*.:  ::  **: :**:*     *** :* *  ** *:*.:
```

Como ya esperábamos ambas proteínas parecen homologas, pero no hay mucha conservación. De todas formas este alineamiento quizás no sea la mejor forma de decidir si esta mucho o poco conservada (deberíamos hacer una búsqueda de homólogos con "genome browser" y comparar).

En amarillo he marcado los sitios activos conservados de una secuencia a otra. Comprobamos que el motivo CXXC esta conservado en ambas secuencias, y el motivo XPT (EL CIS-proline loop) aparece como IPT en un caso y como TPT en otro.

(Ver la figura 1 de esta cita: http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=19181668)

Salvad el alineamiento de CLUSTALW en un fichero de texto (.txt sin formato, no HTML). Ahora vamos a usar el programa SWISS MODEL, que realiza predicciones de estructura de proteína a partir de homología con proteínas cuyas estructuras están resueltas.
http://swissmodel.expasy.org//SWISS-MODEL.html

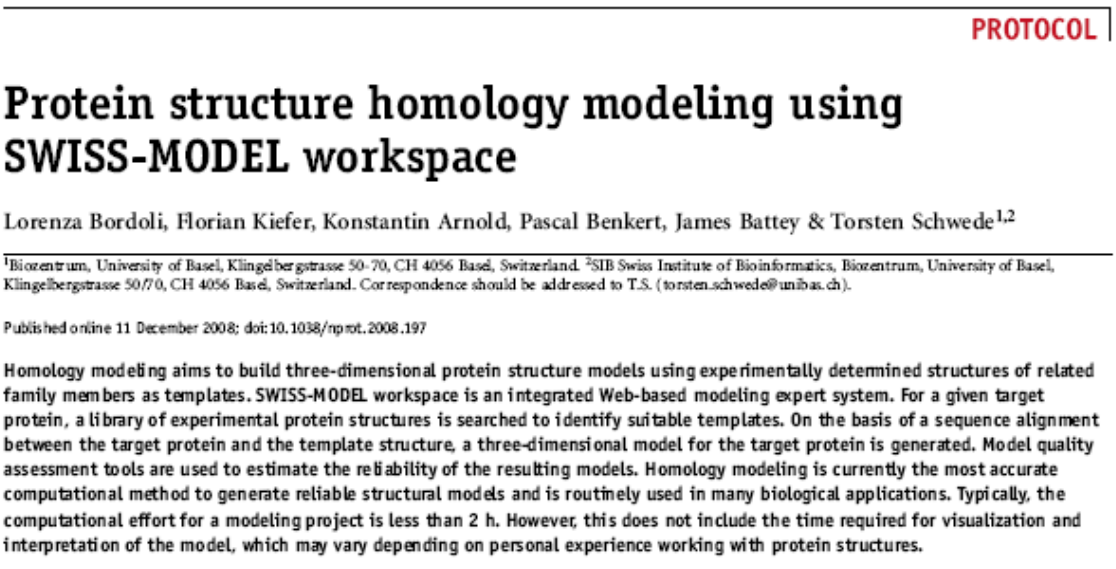


En nuestro caso, vamos a utilizar la aplicación (*Alignment interface*) de SWISS MODEL que nos permite seleccionar la proteína patrón (*template*) que queremos usar sobre nuestra secuencia problema (*target*). Sinó, el programa buscaría por defecto la proteína más adecuada en su base de datos. Introducid el fichero que contiene el alineamiento y seleccionad CLUSTALW como formato:

Swiss-model dispone de tres formas de trabajo, la que nosotros hemos elegido por sugerencia de la PEC, debería estas basada en una serie de asunciones, y comprobaciones posteriores, siendo la más importante: Comprobar con distintas proteínas homologas la estructura resultante. De todas formas, con los datos que hemos acumulado sobre los centros activos seguramente podamos llegar a dar algunas conclusiones.

El protocolo de cálculo está perfectamente descrito en este artículo:

http://www.nature.com/nprot/journal/v4/n1/abs/nprot.2008.197.html



PROTOCOL

Protein structure homology modeling using SWISS-MODEL workspace

Lorenza Bordoli, Florian Kiefer, Konstantin Arnold, Pascal Benkert, James Battey & Torsten Schwede[1,2]

[1]Biozentrum, University of Basel, Klingelbergstrasse 50-70, CH 4056 Basel, Switzerland. [2]SIB Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Klingelbergstrasse 50/70, CH 4056 Basel, Switzerland. Correspondence should be addressed to T.S. (torsten.schwede@unibas.ch).

Published online 11 December 2008; doi:10.1038/nprot.2008.197

Homology modeling aims to build three-dimensional protein structure models using experimentally determined structures of related family members as templates. SWISS-MODEL workspace is an integrated Web-based modeling expert system. For a given target protein, a library of experimental protein structures is searched to identify suitable templates. On the basis of a sequence alignment between the target protein and the template structure, a three-dimensional model for the target protein is generated. Model quality assessment tools are used to estimate the reliability of the resulting models. Homology modeling is currently the most accurate computational method to generate reliable structural models and is routinely used in many biological applications. Typically, the computational effort for a modeling project is less than 2 h. However, this does not include the time required for visualization and interpretation of the model, which may vary depending on personal experience working with protein structures.

El protocolo de SWISS-MODEL queda descrito en su artículo base:

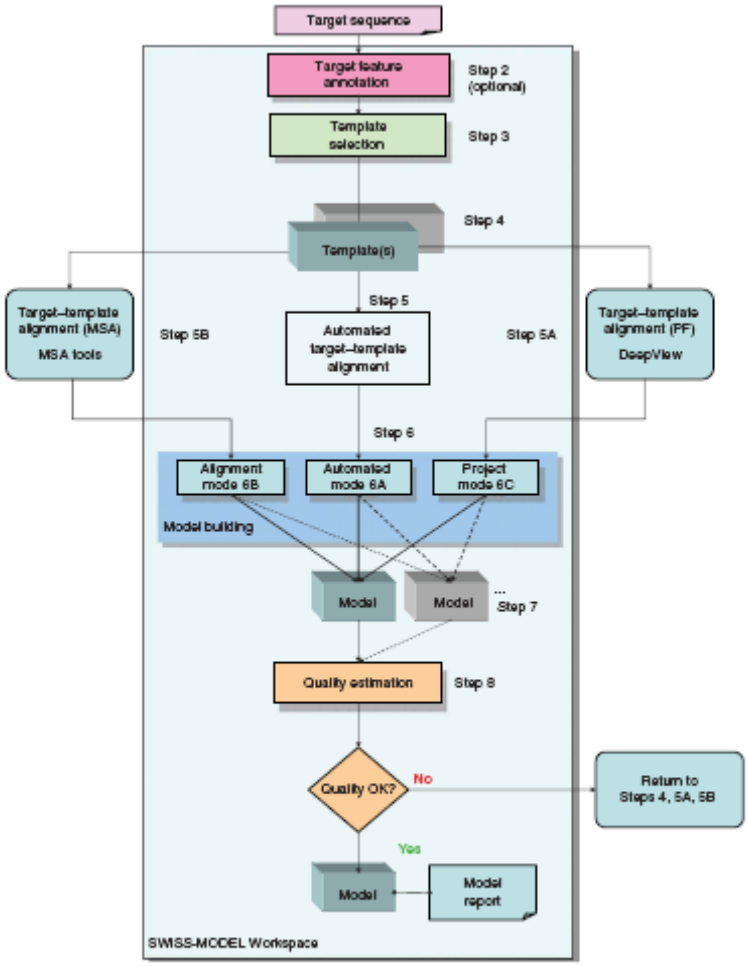
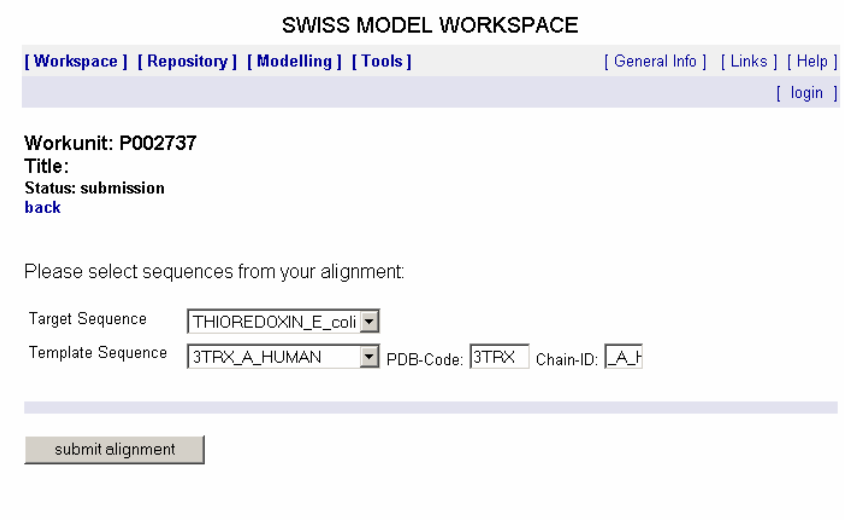


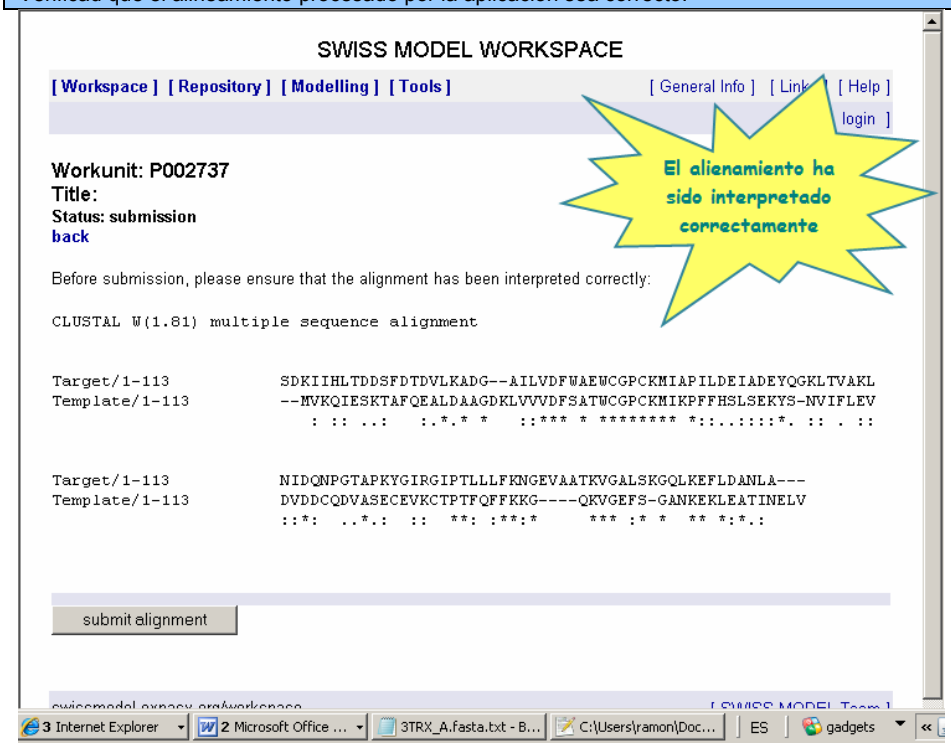
Figure 2 | Workflow of comparative protein structure modeling using SWISS-MODEL workspace. Starting from the amino-acid sequence of a 'Target' protein, three alternative routes for model building are provided—depending on the difficulty of the modeling task. Individual steps are described in detail in PROCEDURE.

Seleccionad el *Target* y el *Template* adecuadamente (fijaos en el ejemplo):



Ojo¡ , hay que poner una A en Chain-ID

Verificad que el alineamiento procesado por la aplicación sea correcto:

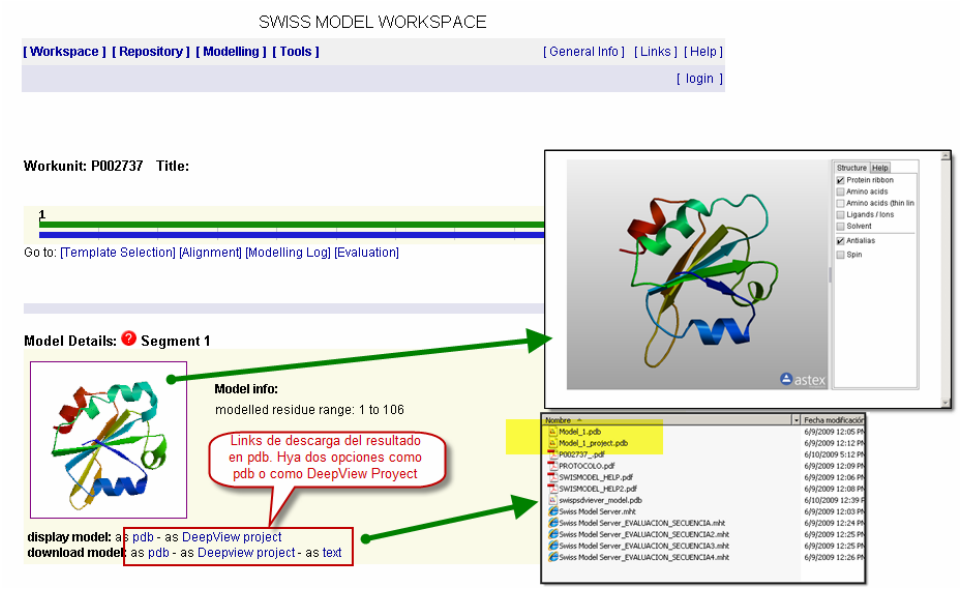


Analizad los resultados que habéis obtenido (calidad del alineamiento, estructuras secundarias, evaluación del modelo) y enumerad las partes más importantes de la salida de este programa:

En unos pocos segundos la "pipeline" nos devuelve el output del cálculo. Este tiene tres partes

A) Partes del modelo.

Desde aquí tenemos una vista preliminar del modelo, un visor con funcionalidades básicas de visualización, y enlaces de descarga de los ficheros pdb con los resultados de la estructura.



Descargamos ambos ficheros, posteriormente comprobaremos que el formato DeepView (**Swiss-PdbViewerDeepView**) es bastante útil:

DeepView
The program DeepView (Swiss-PdbViewer, *Guex et al.*) can be used to generate, display, analyze and manipulate modeling project files for the SWISS-MODEL workspace.
Project files contain the superposed template structures, and the alignment between the target and template. The user has therfor full control over essential modelling parameters, i.e. the choice of template structures, the correct alignment of residues, and the placement of insertions and deletions in the context of the three-dimensional structure.
Project files can be generated inside DeepView, by the workspace template identification tools, and are also the default output format of the modeling pipeline. This allows analyzing and iteratively improving the output of the different modeling tools.
DeepView allows to visualize the model and the templates, and to analyse certain structural features e.g. Ramachandran plots or electrostatic properties. Moreover, it allows adjusting manually the placement of insertions and deletions in the alignment on which the initial modelling process was based on. The project with the modified alignment can then be re-submitted to the SWISS-MODEL workspace for model building.
DeepView can be downloaded at: http://www.expasy.org/spdbv/
DeepView does not require administrator privileges for installation. E.g. under MS windows, simply uncompress the distributed archive at any location you like (e.g. c:\spdbv or on your desktop) and start working by starting the spdbv.exe application.

B) Alineamiento

El output también nos proporciona el alineamiento con algunas indicaciones extra:

Las h y s representan la estructura secundaria resuelta: las "h" se refieren a las hélices y las "s" a las laminas (sheet),

```
                    1          10         20         30         40
TARGET     3        KIIHLT DDSFDTDVLK AD--GAILVD FWAEWCGPCK MIAPILDEIA
3trxA      1        MVKQIE SKTAFQEALD AAGDKLVVVD FSATWCGPCK MIKPFFHSLS

TARGET                sss   hhhhhhhh       sssss ss     h    h   hhhhh
3trxA                 sss   hhhhhhhhh h    sssss ss     h    h   hh

                47 50         60         70         80         90
TARGET     47    DEYQGKLTVA KLNIDQNPGT APKYGIRGIP TLLLFKNGEV AATKVGALSK
3trxA      47    EKYS-NVIFL EVDVDDCQDV ASECEVKCTP TFQFFKKG-- --QKVGEFS-

TARGET          hh     sss ssss     h hhhh       sssssss     sssssss
3trxA           hh     sss ssss   hhh hhhh     s ssssss       sssssss

                97 100
TARGET     97    GQLKEFLDAN LA --
3trxA      91    GANKEKLEAT INELV

TARGET               hhhhh hh
3trxA           s   hhhhhh hhh
```

Comentarios al alineamiento:

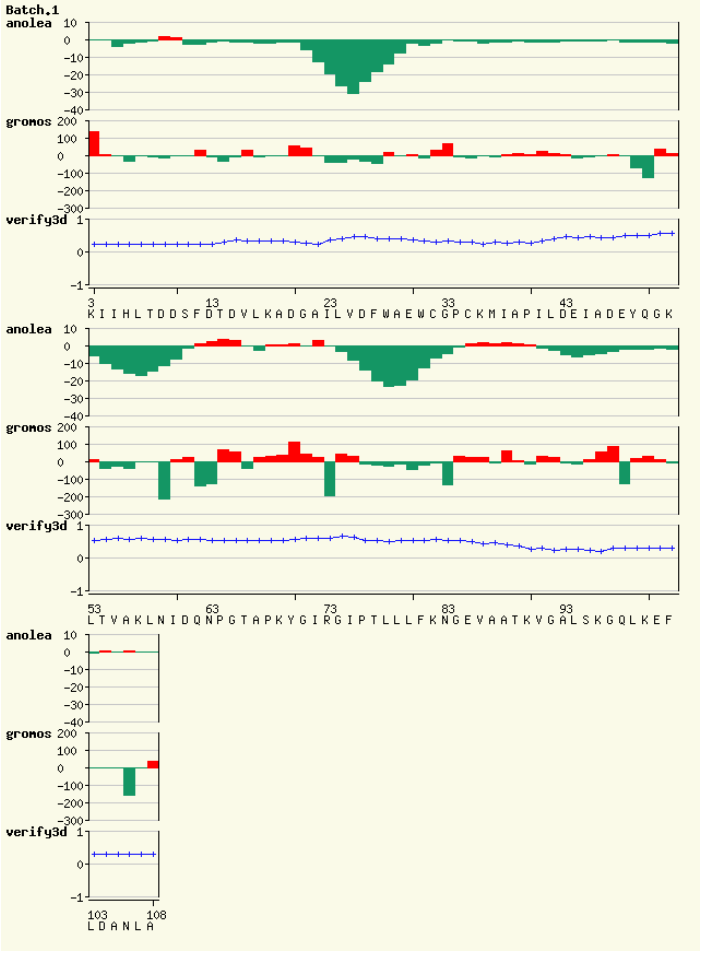
a. Las inserciones y deleciones las he señalado en amarillo. Corresponden a zonas de la secuencia que no son ni hélices ni láminas.
b. Comprobamos que en principio que zonas muy conservadas no tienen por que coincidir con estructuras secundarias. Están señaladas con verde. En concreto ha quedado señalado en verde el centro activo.
c. En rojo he señalado algunas (no todas) zonas con baja conservación que pertenecen a estructuras secundarias

C) Evaluación del modelo

En este caso la ayuda de SWISS-MODEL es bastante explicita (la reproduzco aquí ya que facilita mucho la interpretación.).

Protein Structure & Model Assessment Tools
Evaluation of model quality is a crucial step in homology modeling. While the performance of the automated SWISS-MODEL (*Schwede et al.*) pipeline in general is continuously evaluated by the EVA project (*Koh et al.*), the quality of individual models can vary significantly. Therefore, graphical plots of Anolea mean force potential (*Melo et al.*),GROMOS empirical force field energy (*van Gunsteren et al.*) and ProQres (*Wallner et al.*) are provided to enable the user to estimate the local quality of the predicted structure. The stereochemistry of protein models and template structures can be analysed with Whatcheck (*Hooft et al.*) and Procheck (*Laskowski et al.*). In order to be able to rank alternative models of the same target, pseudo energies for the entire model as calculated by QMEAN (*Benkert et al.*) and DFIRE (*Zhou et al.*) are provided as well. To facilitate the description of template and model structures, DSSP (*Kabsch et al.*) and Promotif (*Hutchinson et al.*) can be invoked to classify structural features.

Anolea
The atomic empirical mean force potential ANOLEA (*Melo et al.*) is used to assess packing quality of the models. The program performs energy calculations on a protein chain, evaluating the "Non- Local Environment" (NLE) of each heavy atom in the molecule.
The y-axis of the plot represents the energy for each amino acid of the protein chain. Negative energy values (in green) represent favourable energy environment whereas positive values (in red) unfavourable energy environment for a given amino acid.



Verify3D
The Verify3D (*Eisenberg et al.*) method assess protein structures using three-dimensional profiles. This program analyzes the compatibilty of an atomic model (3D) with its own amino acid sequence (1D). Each residue is assigned a structural class based on its location and environment (alpha, beta, loop, polar, apolar etc). Then a database generated from good structures is used to obtain a score for each of the 20 amino acids in this structural class.
The vertical axis in the plot represents the average 3D-1D profile score for each residues in a 21-residue sliding window. The scores ranges from -1 (bad score) to +1 (good score).

Gromos
GROMOS (*van Gunsteren et al.*) is a general-purpose molecular dynamics computer simulation package for the study of biomolecular systems and can be applied to the analysis of conformations obtained by experiment or by computer simulation.
The y-axis of the plot represents the energy for each amino acid of the protein chain. Negative energy values (in green) represent favourable energy environment whereas positive values (in red) unfavourable energy environment for a given amino acid.

El output nos proporciona tres gráficos para evaluar el modelo. Los gráficos de Anolea y Gromos, se basan en cálculos de dinámica molecular. En ambos casos el significado es el mismo. Las zonas verdes indican estabilidad molecular (un mínimo local de la energía Conformacional).

El grafico nos muestra 4 zonas de mayor estabilidad intercaladas entre otras de menos estabilidad. Como es posible que el escalado de energía no sea correcto no tendremos en cuenta la profundidad de estas zonas.

- La primera zona, entre los aminoácidos 20-28, coincide con la primera deleccion del TARGET (ver alineamiento).

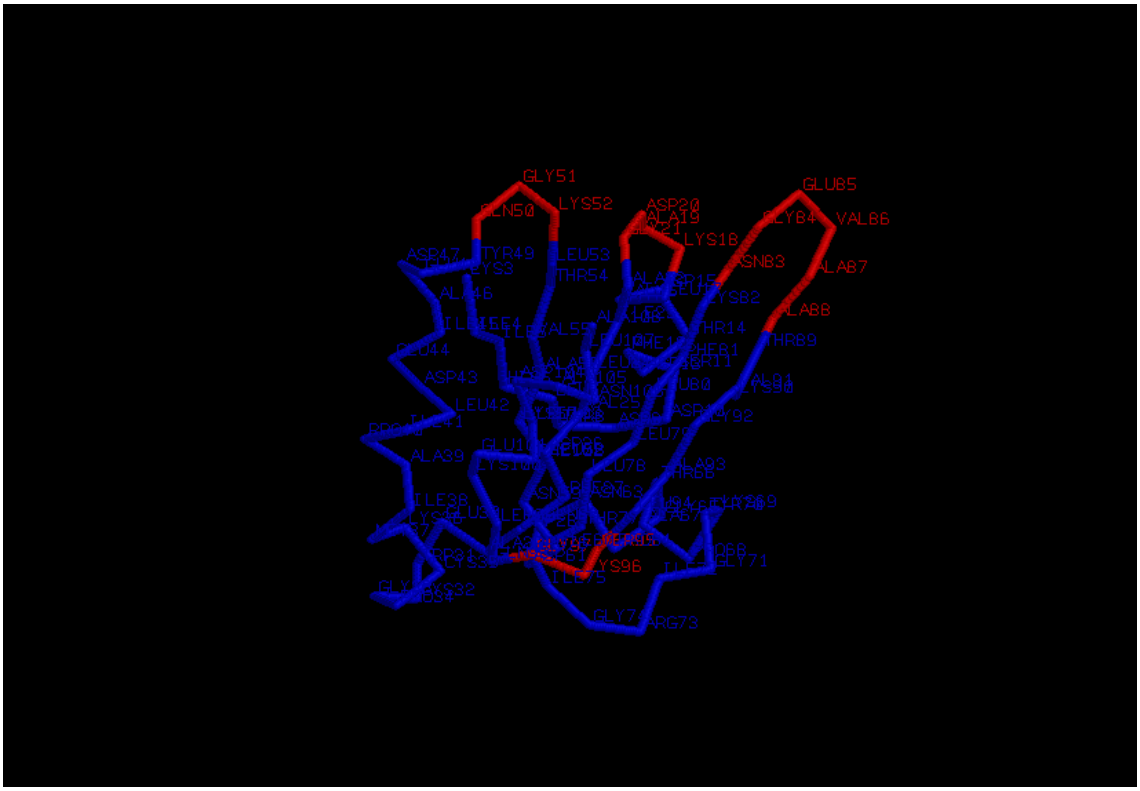
- Vemos como las zonas de Gromos coinciden con las de Anorea

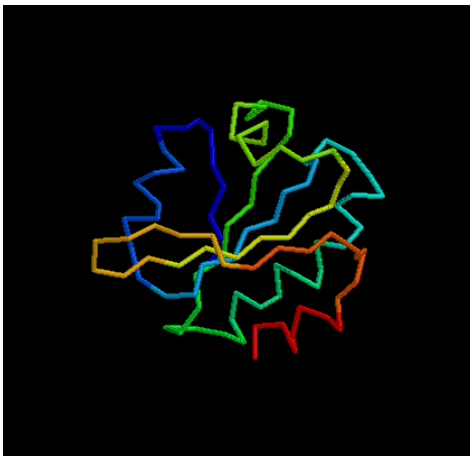
La grafica de Veryfy3D, determina 2 zonas con mayor y menor estabilidad la frontera entre ambas esta más o menos en el residuo 90.

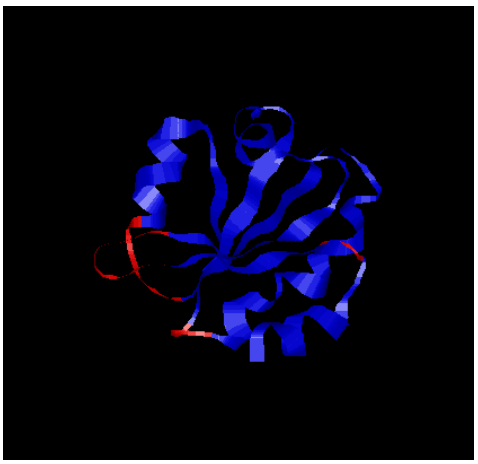
Desde la pagina de Swiss model se puede hacer una evaluación más detallada del modelo. Posponemos para más adelante este análisis.

Podéis visualizar moléculas en vuestro ordenador usando por ejemplo el programa RASMOL (para su instalación ver la sección de Tutoriales en el Fórum de la asignatura).
http://www.openrasmol.org/

Desde Rasmol podemos hacer una visualización por temperatura del modelo teórico. Esta visualización corresponderá más o menos con los modelos anolea y gromos, se aprecian con claridad las cuatro zonas rojas de mayor energía.

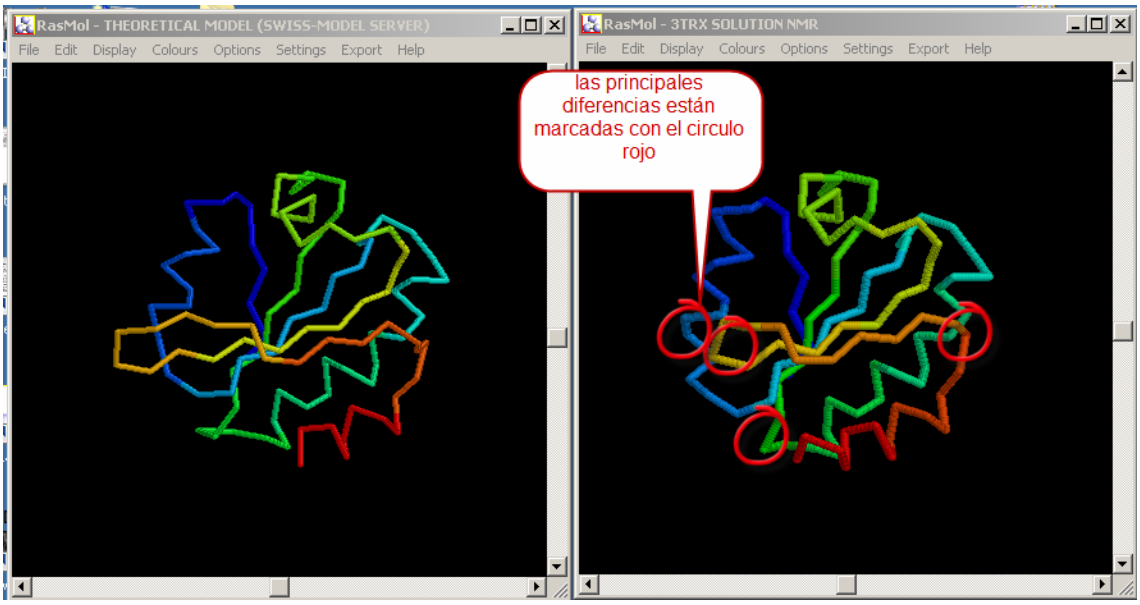




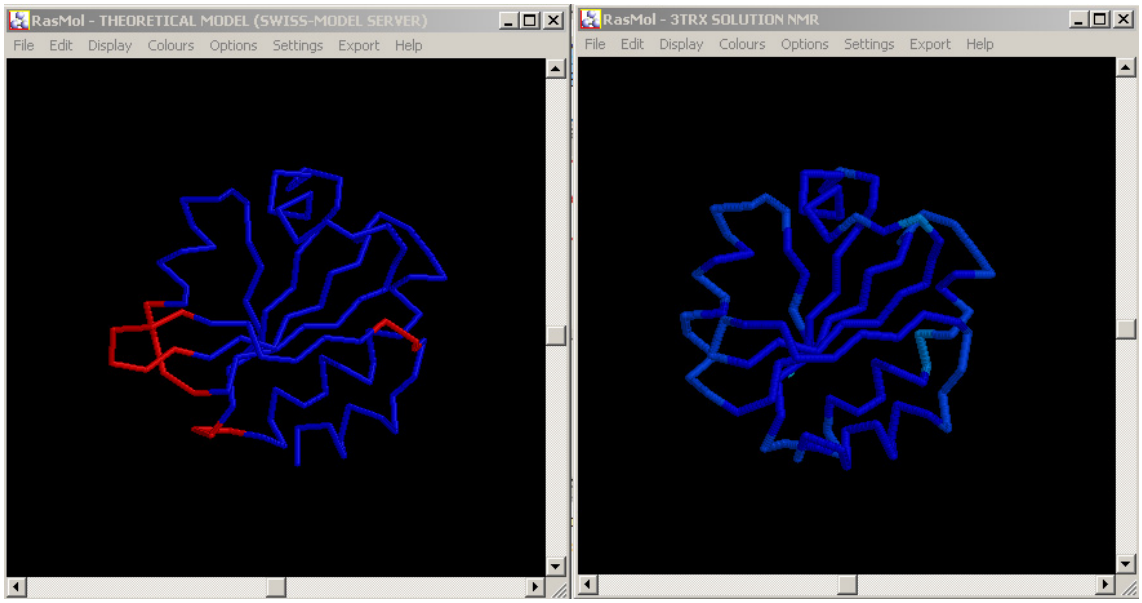


La imagen de la izquierda esta coloreada por estructura secundaria.

Extraed el fichero en formato PDB con el modelo predicho por SWISS MODEL para nuestra proteína bacteriana. Repetid lo mismo con la entrada de la proteína humana 3TRX que guardásteis antes. Comparad visualmente ambas proteínas con RASMOL (¿detectáis alguna diferencia?, ¿podéis encontrar los extremos de ambas proteínas?):

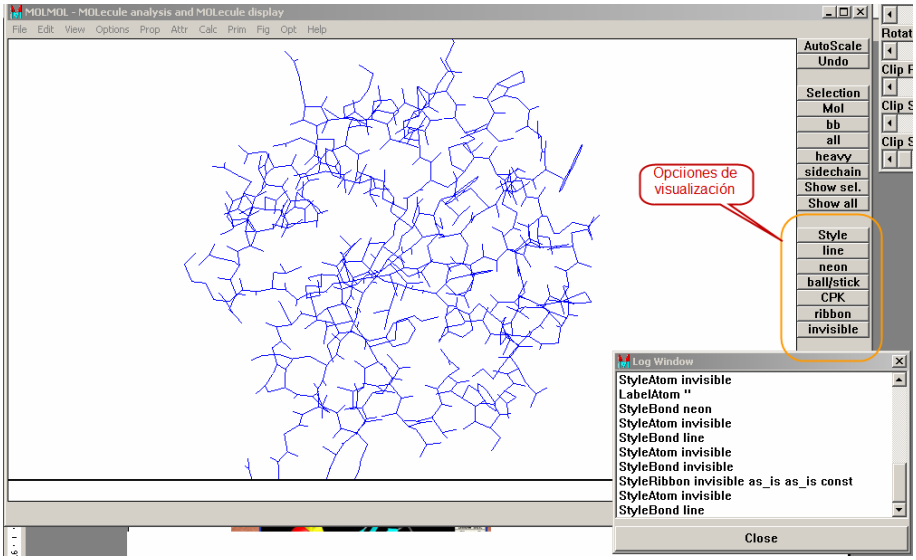


Si comprobamos las estructuras con la coloración por temperatura (abajo), encontramos que curiosamente corresponden a las zonas rojas. (las teóricamente más inestables)

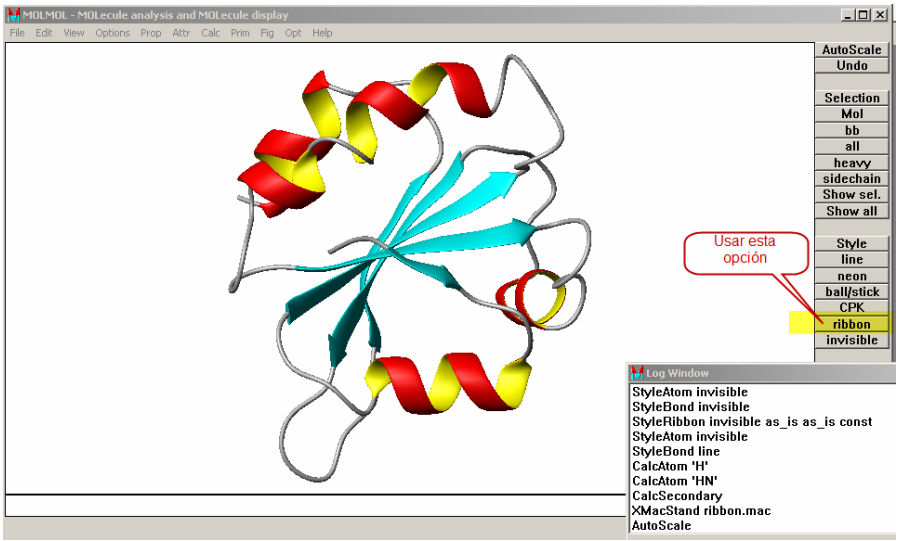


Visualización con MOLMOL

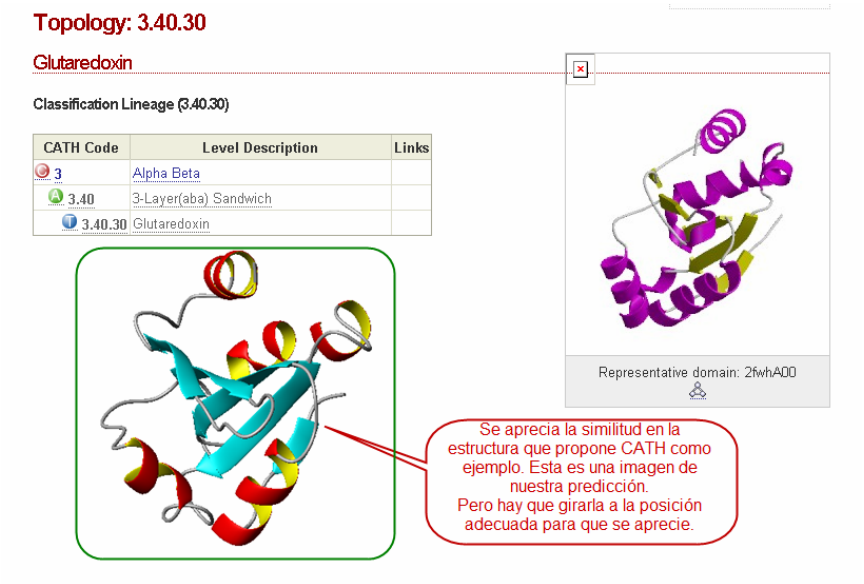
Existen programas que permiten la comparación gráfica y cuantitativa de dos o más estructuras para evaluar la calidad de las predicciones. El programa **MOLMOL** es una potente aplicación que vamos a usar para comparar la predicción de la estructura de la proteína THIOREDOXIN de *E.coli* con su homóloga de *H.Sapiens,* que habíamos usado para crear la predicción:
Ejecutad el programa MOLMOL. Cargad la prediccion en formato PDB obtenida con el programa SWISS MODEL. Para cargar una molécula debéis ir al menú File -> ReadMol -> PDB (ahí seleccionáis vuestro fichero PDB-predicción guardado anteriormente).



Buscad las opciones de visualización para tener la vista de estructura secundaria (imagen inferior). Analizad si la clase estructural definida en CATH para ésta molécula coincide con lo que véis aqui:

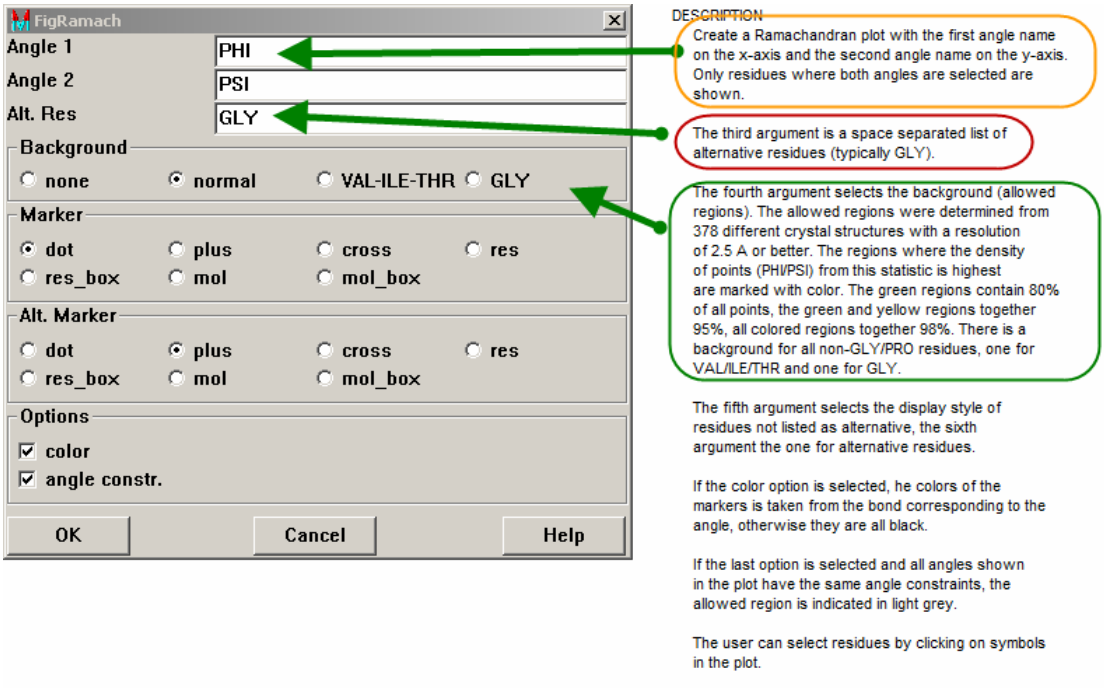


Si miramos a primera vista nuestra imagen y la que sale en CATH, pensaremos "no se parecen en nada", pero si giramos un poco nuestra imagen, no cuesta mucho apreciar un gran parecido. Respecto a la clase no tenemos nada que objetar es una alfa/beta y bien pude ser "3-layer(aba) sándwich" (no soy un experto¡¡¡).

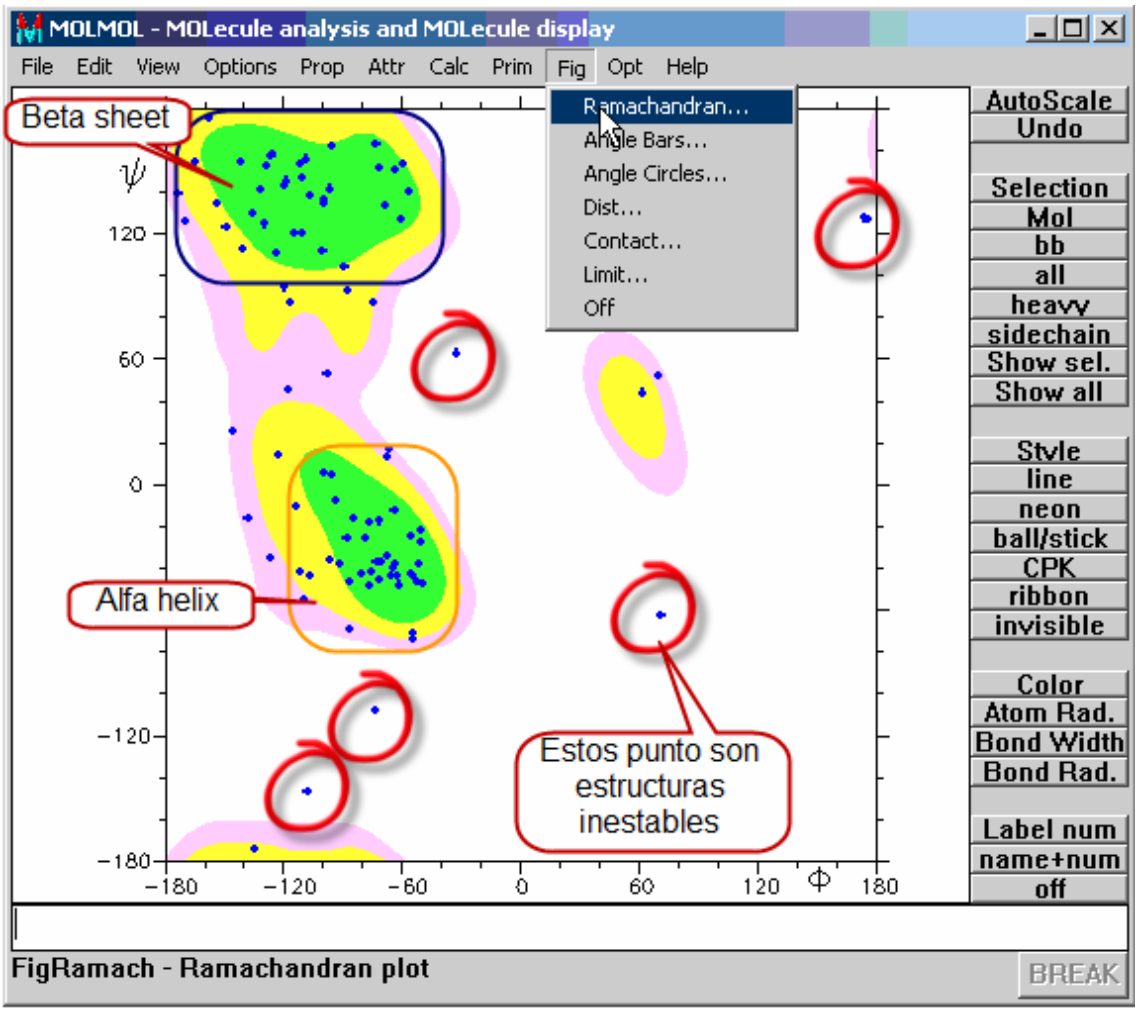


Buscad la opción que os permite generar un diagrama de *Ramachandran*, analizad qué significa cada *cluster* de puntos claramente definido y cómo encajan en la definición estructural de esta proteína:
http://en.wikipedia.org/wiki/Ramachandran_plot

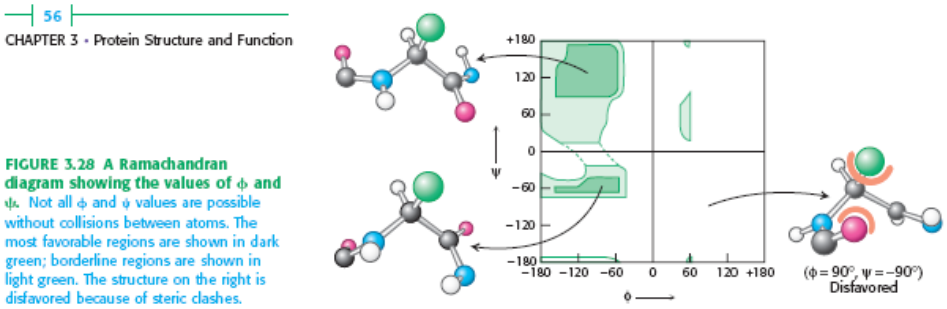
El diagrama de Ramachadran se obtiene desde el menú "Fig", y tiene algunas opciones que están perfectamente explicadas en la ayuda:



El diagrama que obtenemos con las opciones por defecto es el siguiente:



Los ángulos por los que pueden girar los residuos dentro de la proteína están representados por φ,ψ.



La situación conformacional del esqueleto peptídico queda entonces reducido al estudio de las parejas de los ángulos diedros (φ,ψ). Esta información se expresa gráficamente en el mapa de Ramachandran.

Las geometrías más probables están indicadas como zonas de color. Fuera de estas zonas, las estructuras presentan problemas estéricos y por lo tanto, representan geometrías muy poco probables.

Cada zona marcada del mapa es característica de un tipo de estructura secundaria.

Fuente: http://mail.efn.uncor.edu/departamentos/quimica/Clases%20QuimBiol/Clase%203%20Proteínas%202009.pdf

DIAGRAMA DE Ramachandran

Correlación entre las combinaciones de valores de psi y phi con conformaciones en las estructuras secundarias de proteínas

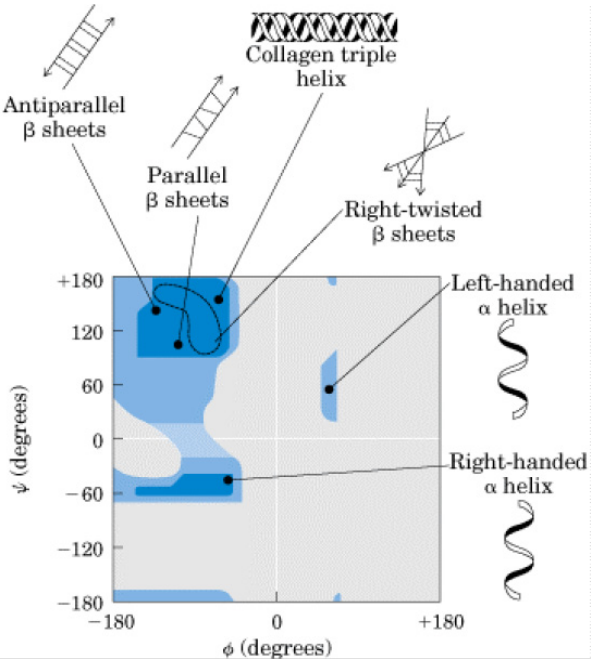


DIAGRAMA DE Ramachandran

Ubicación de las α-hélices

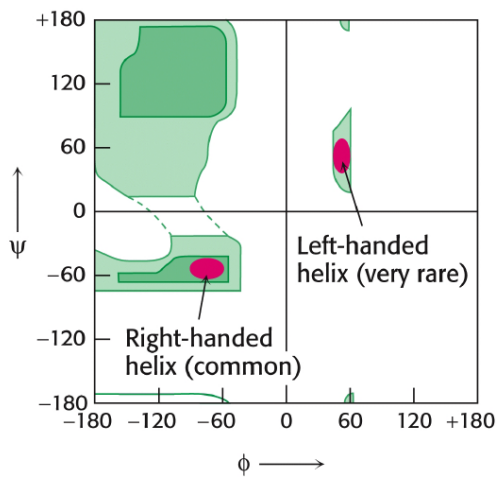
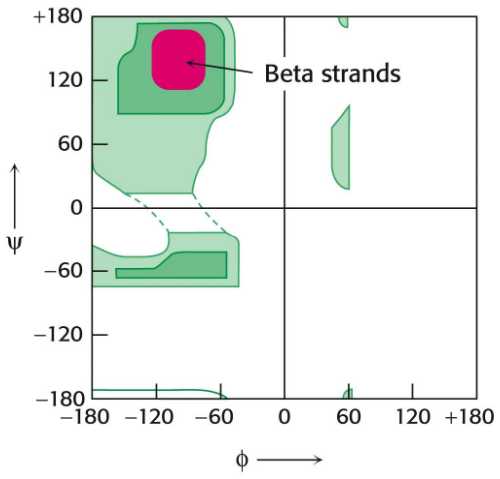
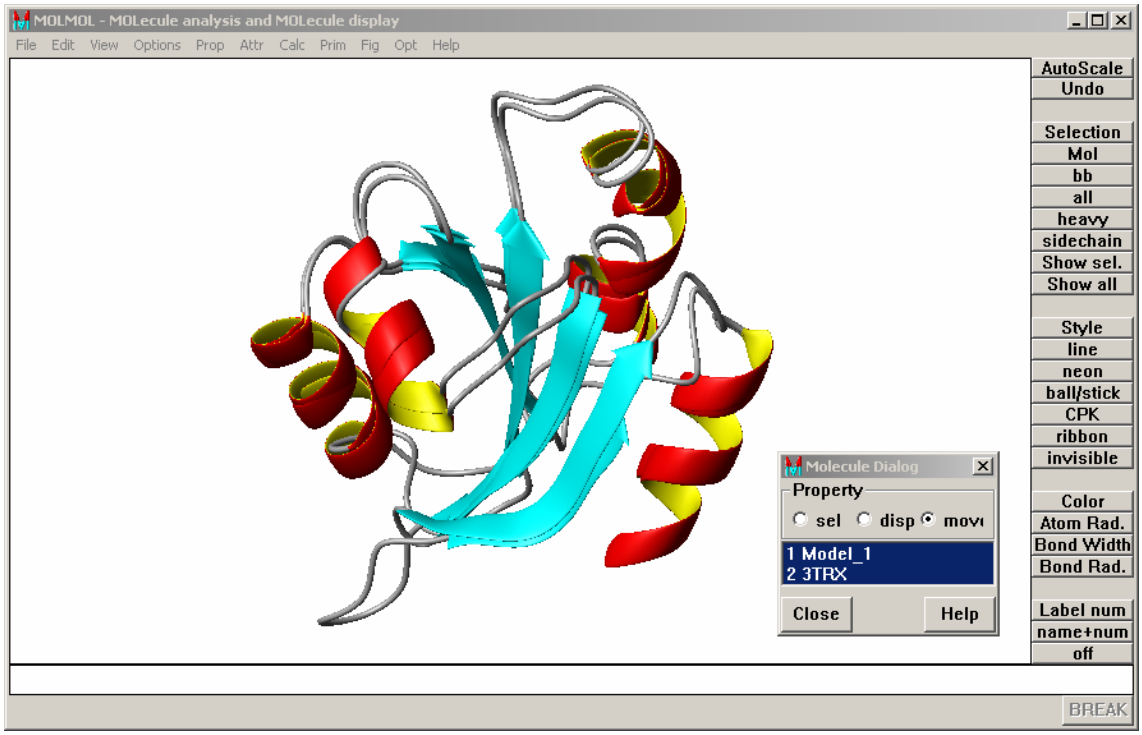


DIAGRAMA DE Ramachandran

Ubicación de las cadenas beta.



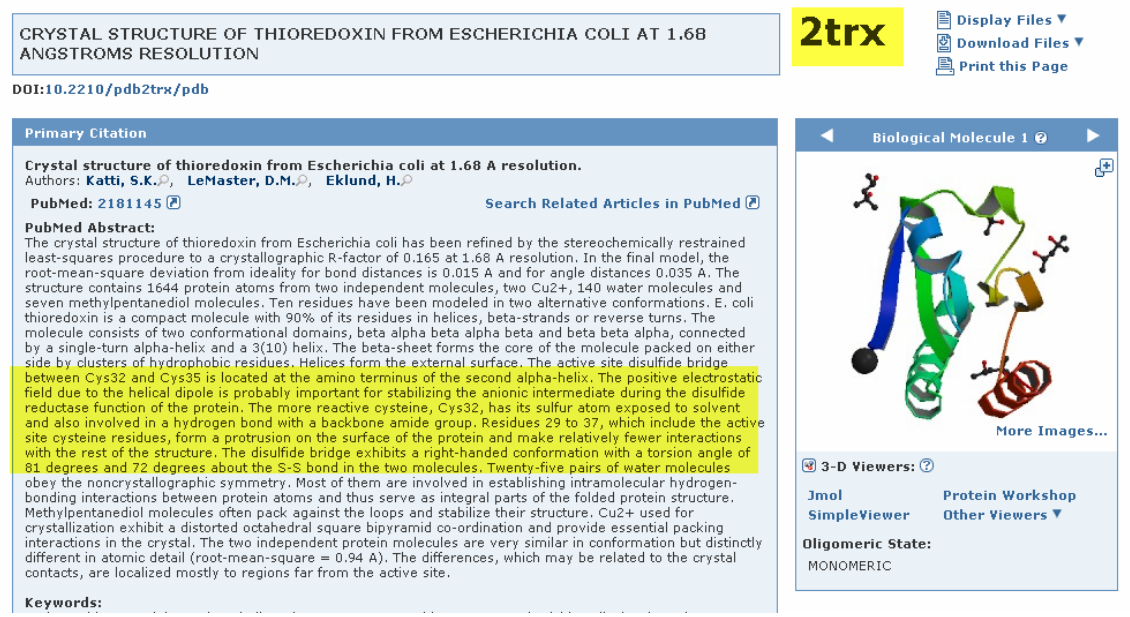
Ahora, superponed la proteína humana que fue usada para obtener la predicción (PDB: **3TRX**). Intentad reproducir la imagen inferior con el programa MOLMOL y explicad qué estáis viendo y si os parece que la predicción en *E.coli* se ajusta al template humano (subjetivamente hablando). Comparad lo que estáis viendo con el nivel de similaridad que observásteis entre las secuencias de las dos proteínas:



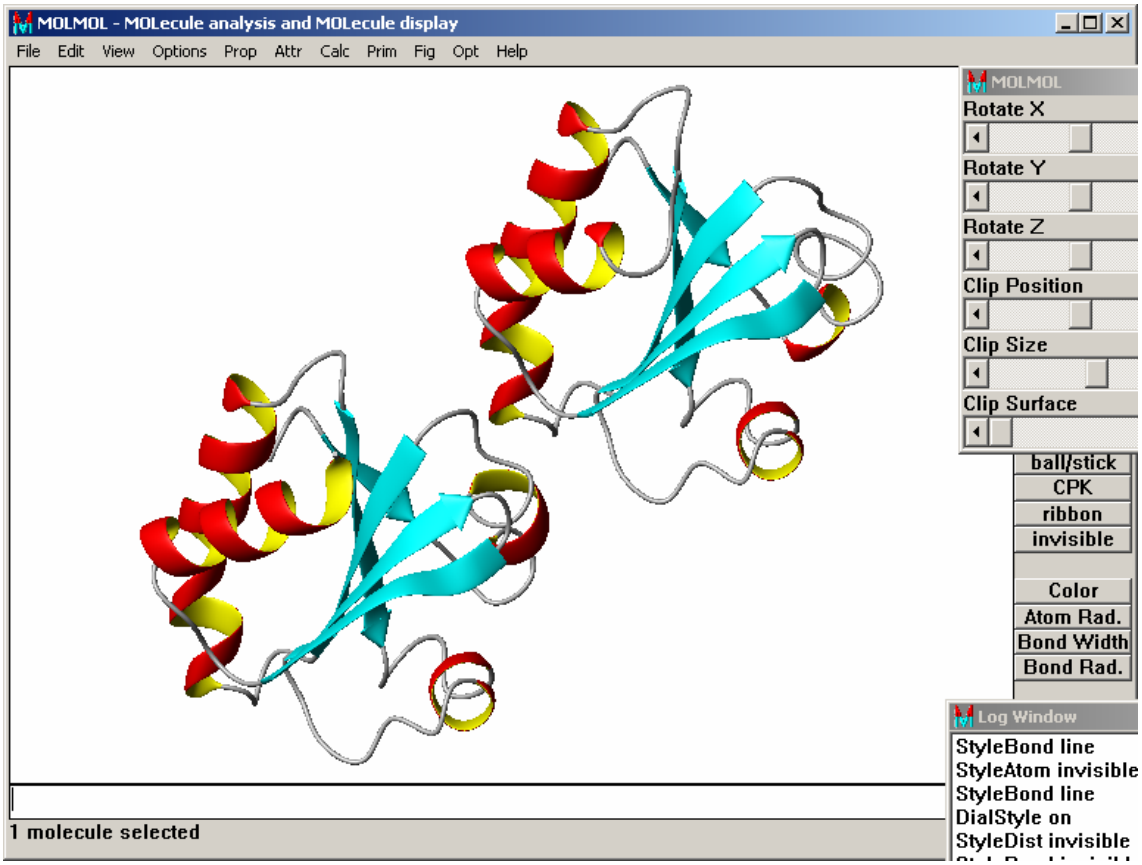
Lo que vemos es que se han copiado prácticamente las estructuras secundarias una encima de otra. No debemos asombrarnos, como hemos visto la estructura resultante tiene muchas "tensiones", ahora de debería optimizar la estructura mediante mecánica molecular, y posteriormente rehacer la estructura secundaria.

De hecho, acabáis de reproducir la cubierta de este libro:
Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, 3rd Edition Andreas D. Baxevanis (Editor), B. F. Francis Ouellette (Editor) -- ISBN: 978-0-471-47878-2
http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471478784.html

Ahora eliminad estas dos moléculas del visor (*remove molecule*) y cargad otra vez la predicción de la estructura en *E.coli* pero ahora comparadla con la verdadera estructura de la proteína (busca la entrada en PDB: **2TRX**). Analizad lo que veis. Pensad que el ajuste de ambas estructuras se debe realizar manualmente utilizando heurísticos y reglas que sobrepasan los límites de este curso (aunque podéis encontrar esas opciones en el programa MOLMOL).



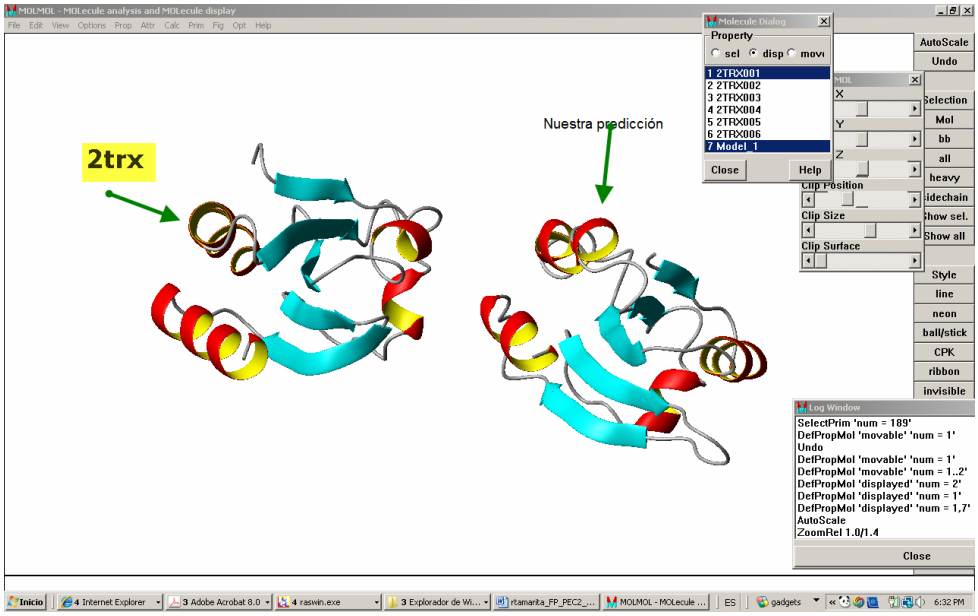
Si visualizamos el pdb con el molmol, lo que vemos es esto:



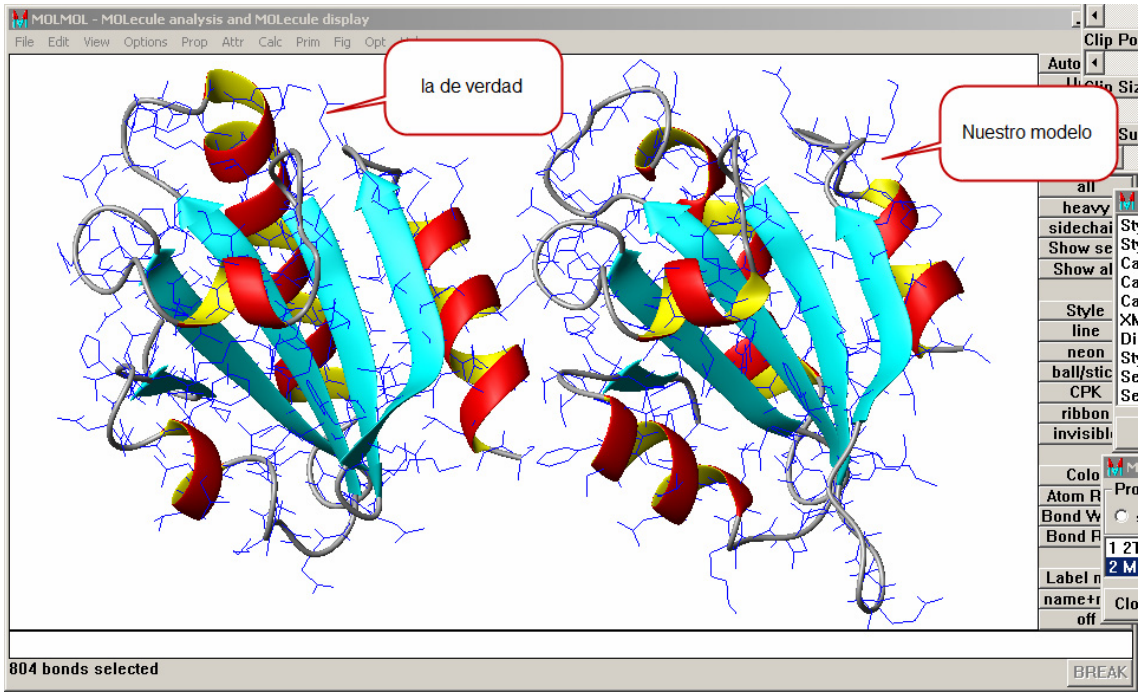
En realidad lo que nos acabamos de descargar es un dímero con algunos ligandos (seguramente del disolvente) y el complejo de coordinación con el Zn. Los

seleccionamos y borramos dejando una sola estructura, así se puede comparar mejor.

Cargamos ahora nuestra predicción y podemos comparar:



En esta otra vista, girándolas un poco se puede apreciar mejor el parecido:

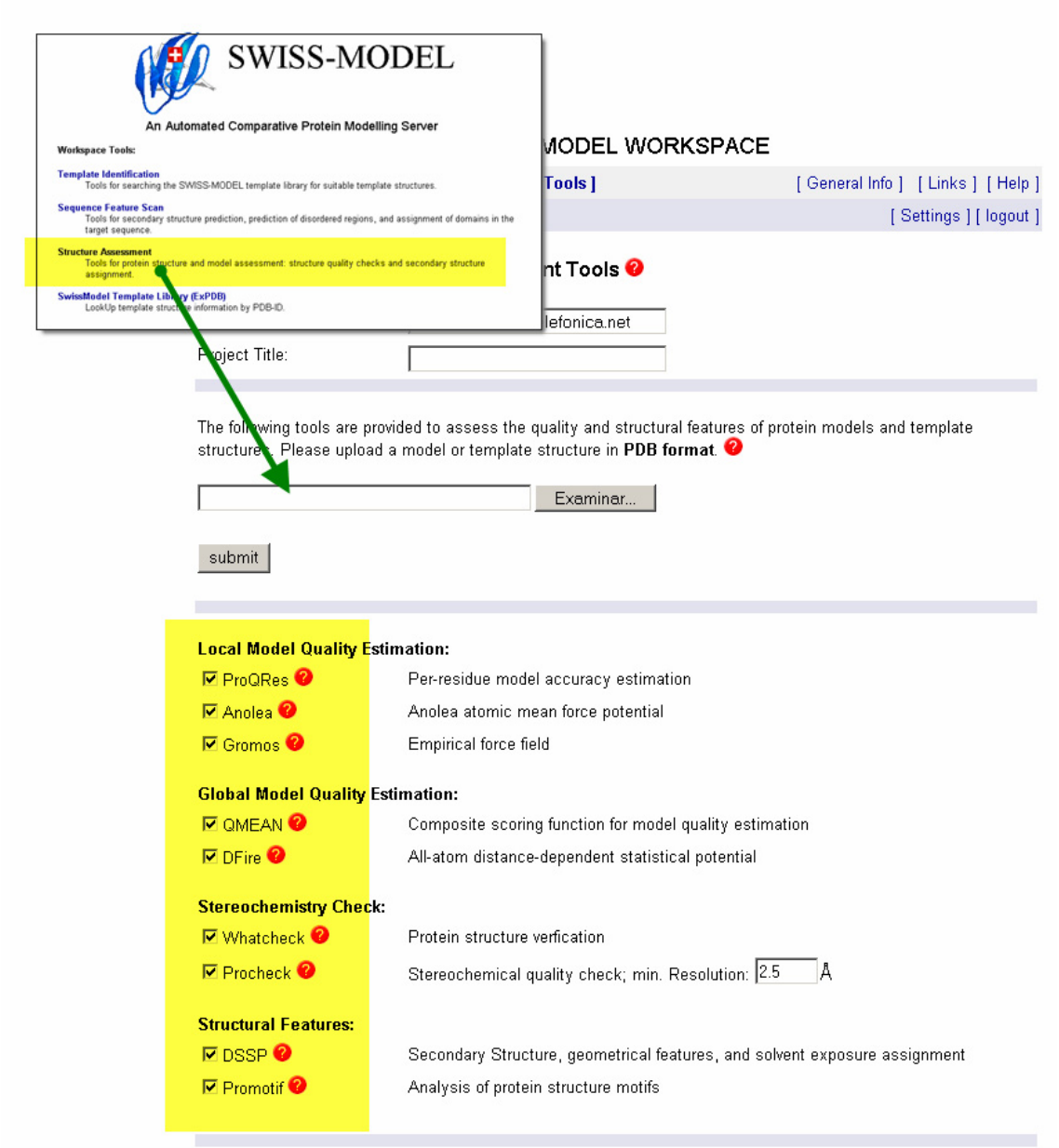


Bastante curioso, ambas estructuras, espacialmente se corresponden bastante bien, pero fallan lo siguiente;

- Los inicios y fin de las estructuras secundarias no están bien definidas en nuestro modelo. ¡¡¡ Cosa que ya habíamos deducido!!!
- Los ligandos de los residuos no acoplan (para eso aparecen en las imágenes).
- El esqueleto principal si que tiene un gran parecido.

Evaluación detallada de los modelos.

Para evaluar un poco más detalladamente el modelo calculado, he usado el "structure assesment" que ofrece swiss-model:



En concreto he calculado una evaluación con todas las opciones para, nuestra TRX calculada, la TRX de e.coli (2trx) y la TRX humana (3trx). Entre las diversas opciones de la evaluación, la que más interesante me ha parecido es la "procheck", que en su sumario permite evaluar de forma comprada algunos parámetros.

TRX calculada

```
+----------<<<  P  R  O  C  H  E  C  K     S  U  M  M  A  R  Y  >>>----------+
 |                                                                            |
 | input_atom_only.pdb   2.5                                  106 residues    |
 |                                                                            |
*| Ramachandran plot:   74.4% core   23.3% allow    2.2% gener    0.0% disall |
 |                                                                            |
*| All Ramachandrans:   11 labelled residues (out of 104)                     |
*| Chi1-chi2 plots:      2 labelled residues (out of  65)                     |
 | Main-chain params:    6 better     0 inside      0 worse                   |
 | Side-chain params:    5 better     0 inside      0 worse                   |
 |                                                                            |
*| Residue properties: Max.deviation:     5.5             Bad contacts:    3 |
*|                     Bond len/angle:    5.7    Morris et al class:  2  1  2 |
+|     3 cis-peptides                                                         |
 | G-factors           Dihedrals:  -0.39  Covalent:   0.22    Overall:  -0.13 |
 |                                                                            |
 | M/c bond lengths:100.0% within limits   0.0% highlighted                   |
 | M/c bond angles:  97.5% within limits   2.5% highlighted                   |
*| Planar groups:    81.2% within limits  18.8% highlighted       1 off graph |
 |                                                                            |
 +----------------------------------------------------------------------------+
   + May be worth investigating further.  * Worth investigating further.
```

TRX E. Coli (2trx)

```
+----------<<<  P  R  O  C  H  E  C  K     S  U  M  M  A  R  Y  >>>----------+
 |                                                                            |
 | input_atom_only.pdb   2.5                                  108 residues    |
 |                                                                            |
 | Ramachandran plot:   96.7% core    3.3% allow    0.0% gener    0.0% disall |
 |                                                                            |
 | All Ramachandrans:    0 labelled residues (out of 106)                     |
 | Chi1-chi2 plots:      0 labelled residues (out of  68)                     |
 | Main-chain params:    6 better     0 inside      0 worse                   |
 | Side-chain params:    5 better     0 inside      0 worse                   |
 |                                                                            |
+| Residue properties: Max.deviation:     4.2             Bad contacts:    0 |
+|                     Bond len/angle:    4.9    Morris et al class:  1  1  2 |
+|     1 cis-peptides                                                         |
 | G-factors           Dihedrals:   0.08  Covalent:  -0.45    Overall:  -0.11 |
 |                                                                            |
 | M/c bond lengths: 97.5% within limits   2.5% highlighted                   |
 | M/c bond angles:  86.4% within limits  13.6% highlighted                   |
 | Planar groups:   100.0% within limits   0.0% highlighted                   |
 |                                                                            |
 +----------------------------------------------------------------------------+
   + May be worth investigating further.  * Worth investigating further.
```

TRX Human (3trx)

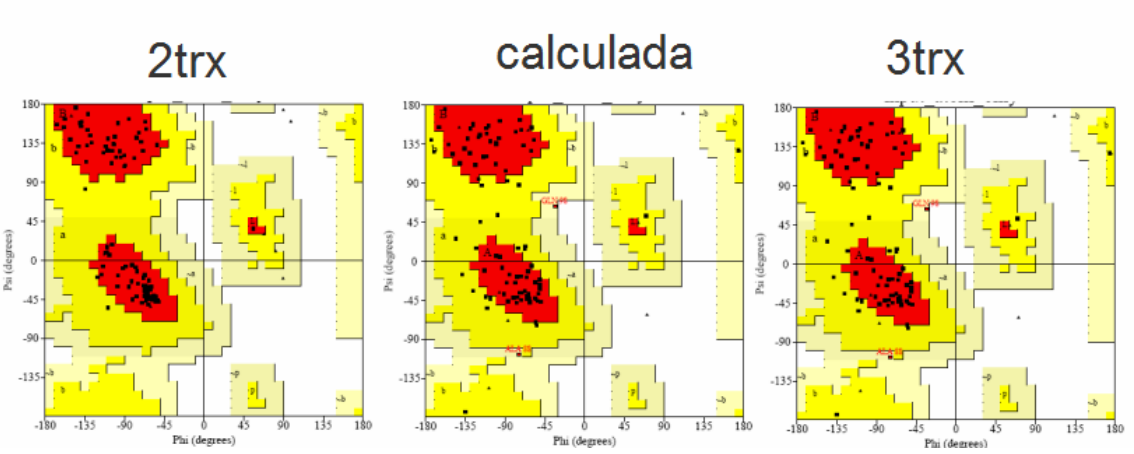
```
+----------<<<  P  R  O  C  H  E  C  K     S  U  M  M  A  R  Y  >>>----------+
 |                                                                            |
 | input_atom_only.pdb   2.5                                  105 residues    |
 |                                                                            |
*| Ramachandran plot:   73.7% core   26.3% allow    0.0% gener    0.0% disall |
 |                                                                            |
*| All Ramachandrans:    8 labelled residues (out of 103)                     |
*| Chi1-chi2 plots:      2 labelled residues (out of  61)                     |
 | Main-chain params:    6 better     0 inside      0 worse                   |
 | Side-chain params:    5 better     0 inside      0 worse                   |
 |                                                                            |
+| Residue properties: Max.deviation:     4.1             Bad contacts:    6 |
 |                     Bond len/angle:    2.8    Morris et al class:  2  3  2 |
+|     1 cis-peptides                                                         |
 | G-factors           Dihedrals:  -0.13  Covalent:   0.10    Overall:  -0.07 |
 |                                                                            |
 | M/c bond lengths: 96.0% within limits   4.0% highlighted                   |
 | M/c bond angles:  99.6% within limits   0.4% highlighted                   |
 | Planar groups:   100.0% within limits   0.0% highlighted                   |
 |                                                                            |
 +----------------------------------------------------------------------------+
```

Los resultados destacables son:

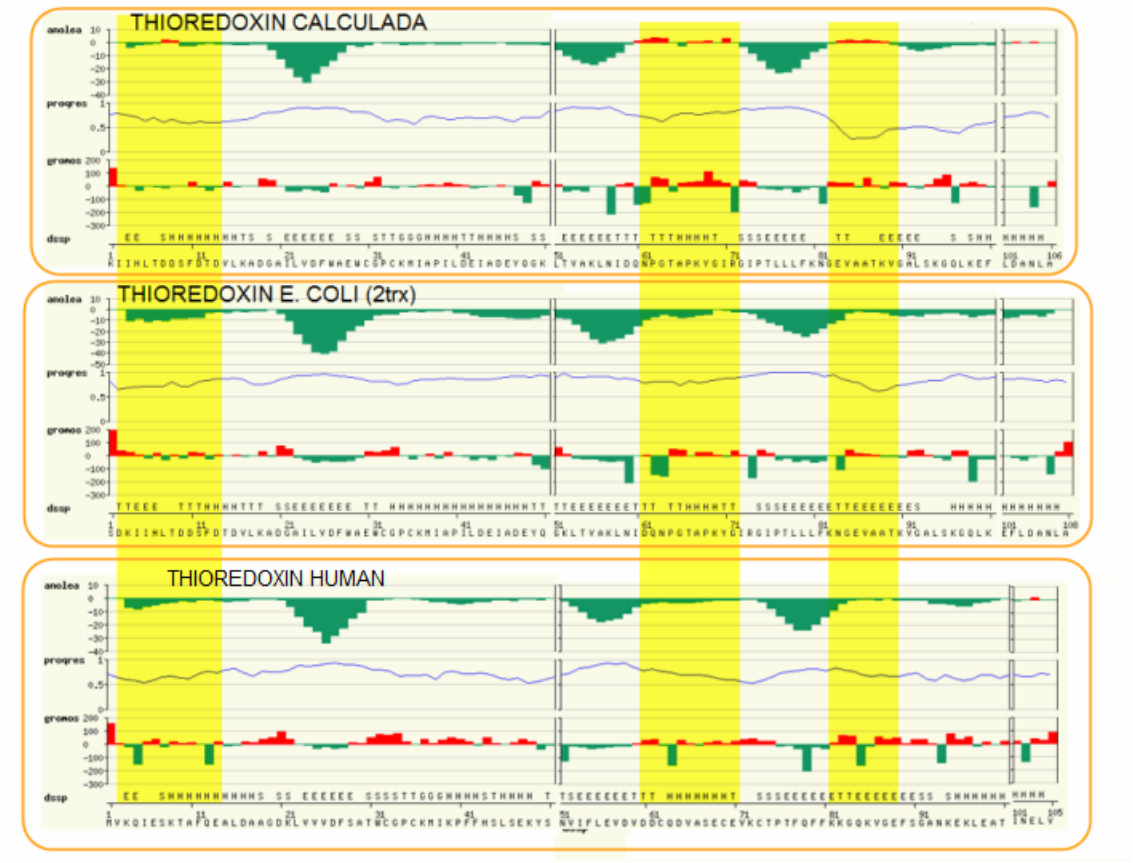
- En el ramachadran plot hay un gran parecido entre TXR Calculada y TXR humana, sin embargo, el Ramachadran plot de TXR e.coli tiene menos puntos fuera de las zonas favorables.

- Tanto en TXR Calculada como en la TXR humana hay “bad-contacts”
- En la estructura calculada hay un 19% de “planar groups”.

La conclusión inmediata es que nuestra estructura calculada tiene muchos enlaces forzados. Igualmente comparando la evaluación de la estructura de la TXR humana y la de e. coli, podríamos pensar en que la primera no es muy acertada (error experimental?¿?). En el cluster de Ramachadran de TRX de e.coli se aprecia con claridad la mayor agrupación de los clusters.

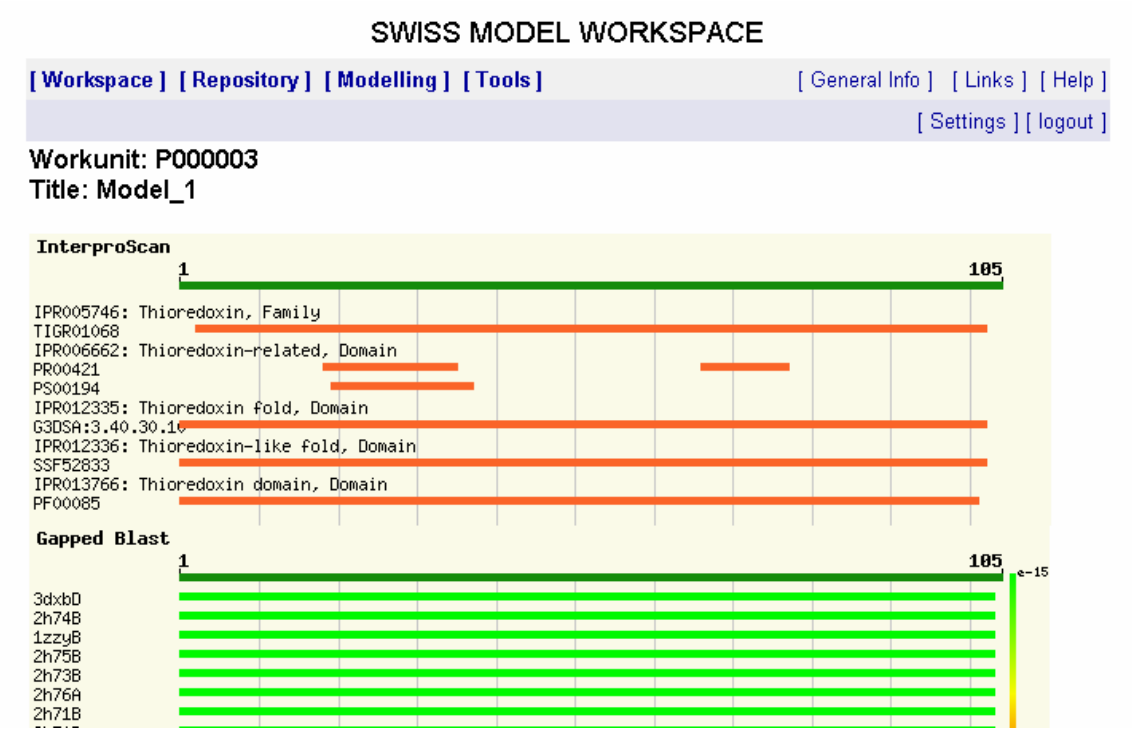


Si comparamos los gráficos de anolea y gromos de las tres estructuras comprobamos que las zonas de estabilidad e inestabilidad son similares, pero sin duda la estructura experimental de e. coli es la que mejor estabilidad estructural ofrece:



Usando otra estructura con mayor homología

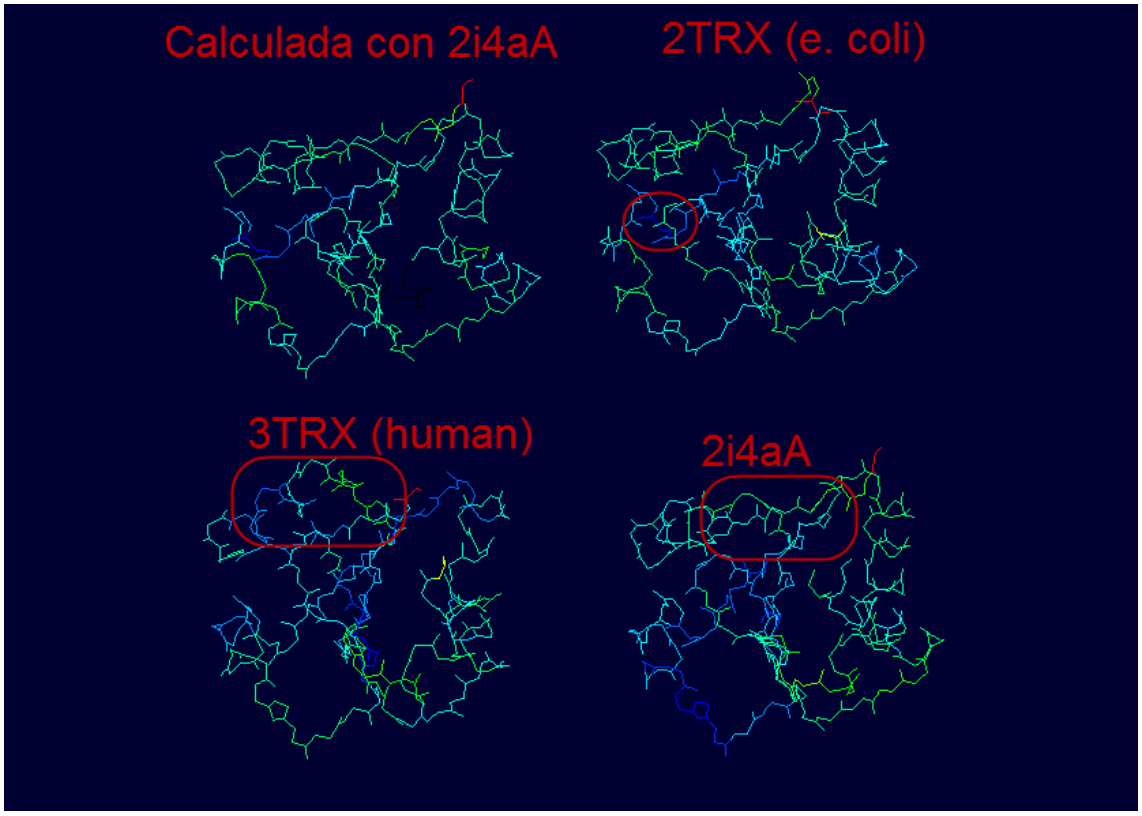
Si repetimos el calculo de la estructura con otro patrón con mayor homología, por ejemplo de otra bacteria, podemos comprobar como los resultados mejoran. Desde Swiss model, se puede realizar un alineamiento para seleccionar otra estructura (e incluso usando DeepView en el modo proyecto varias…. ¡¡¡¡ no hay tiempo)



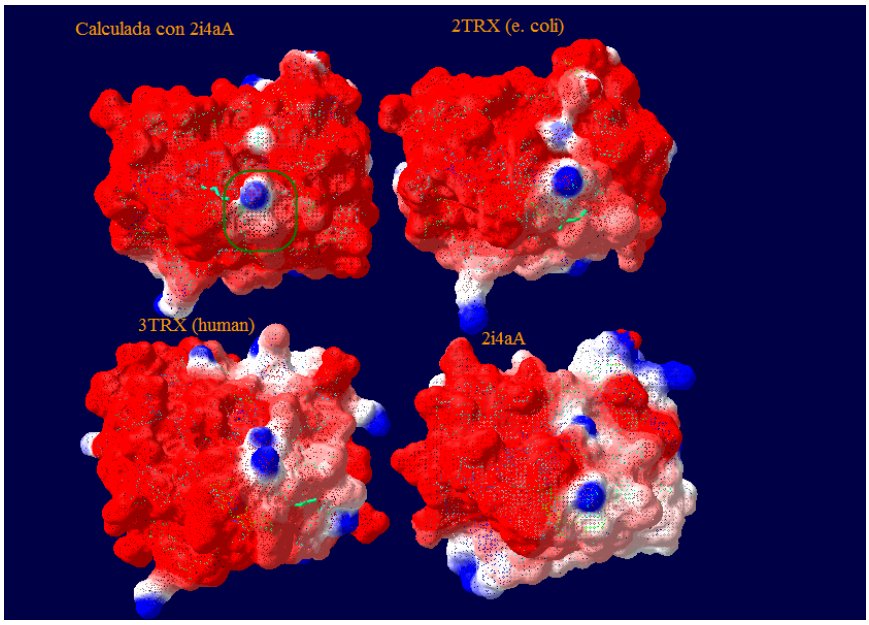
Seleccionamos la de esta bacteria y repetimos el cálculo.

```
>[Template]|2i4aA|1|Crystal structure of thioredoxin from the acidophile
           Acetobacter aceti
          Length = 107


 [Display Alignment in DeepView]
 Score =  128 bits (322), Expect = 6e-31
 Identities = 60/99 (60%), Positives = 75/99 (75%)

Query: 3   LTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGT 62
           ++D SFD DVLKA G +LVDFWAEWCGPCKMI P L EI  E+ GK+TVAK+NID NP T
Sbjct: 7   VSDSSFDQDVLKASGLVLVDFWAEWCGPCKMIGPALGEIGKEFAGKVTVAKVNIDDNPET 66

Query: 63  APKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDA 101
              Y +R IPTL+L ++G+V   KVGAL K QLK ++++
Sbjct: 67  PNAYQVRSIPTLMLVRDGKVIDKKVGALPKSQLKAWVES 105


[top]
```

El resultado final, comparando varias de las estructuras:



En la estructura calculada con 2i4aA he marcado en negro los residuos del motivo CXXC. Se puede apreciar que es una zona muy conservada, y por tanto se reproduce bastante bien en el modelo.(si comparamos con el modelo calculado con 3TRX ocurre lo mismo). El patrón de color usado es de energía. Si visualizamos la superficie molecular se aprecia el centro activo con más claridad.



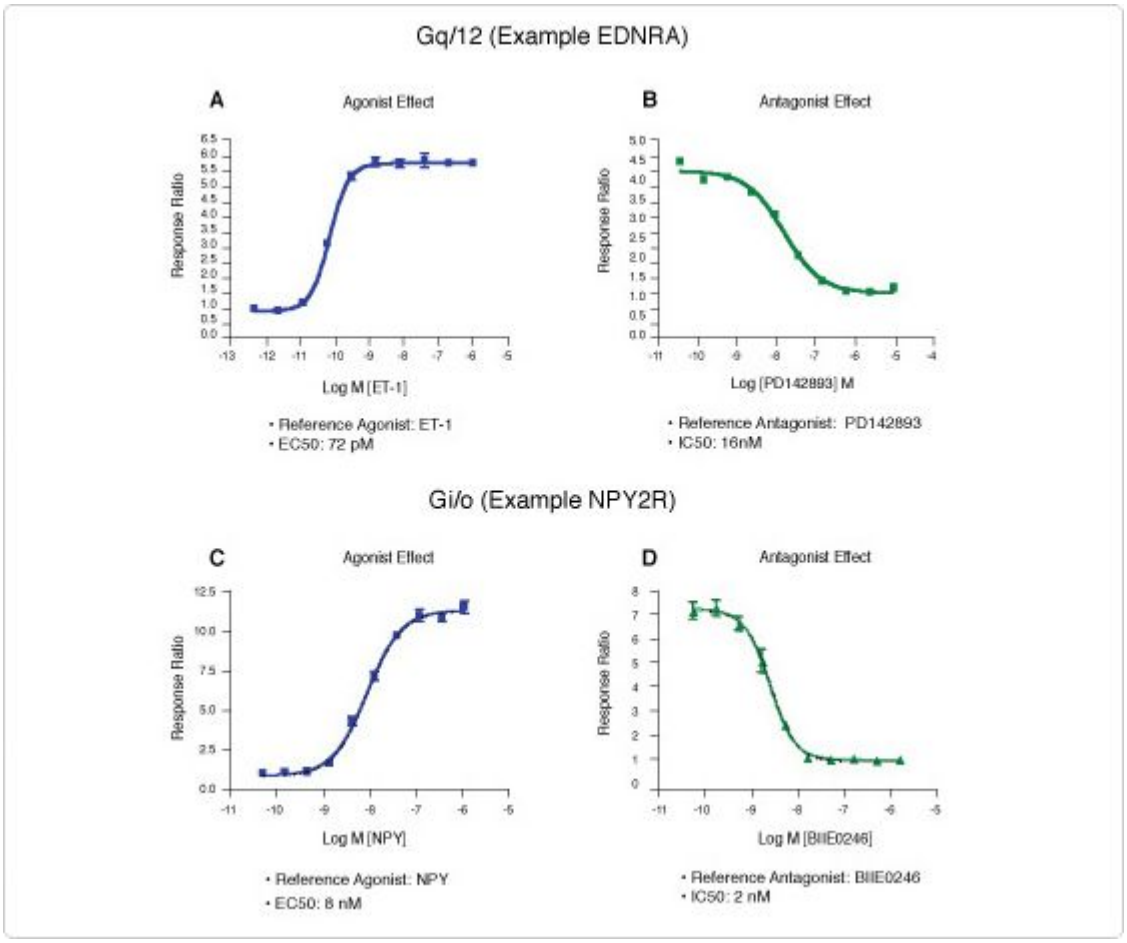
Esta metodología es de utilidad para determinar la reactividad, por ejemplo de un fármaco.

NOTA: Aquí tenéis un listado de varios programas de predicción de estructura de proteínas:
http://en.wikipedia.org/wiki/Protein_structure_prediction_software

Ejercicio 3 (20%): Diseño computacional de fármacos

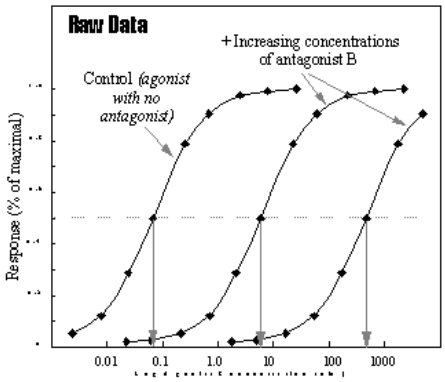
hERG es un gen que codifica para una proteína que es un canal de iones potasio. Este canal de potasio es el mejor conocido por su actividad eléctrica en el corazón y coordina su latido.
http://en.wikipedia.org/wiki/HERG
Cuando el canal pierde su habilidad para conducir la corriente eléctrica a través de la membrana celular debido a algún fármaco o alguna mutación en el gen se producen arítmias e incluso la muerte. Hay un gran número de buenos fármacos en el mercado que inhiben este canal, cosa que supone un grave efecto secundario para los pacientes. La indústria farmacéutica pone un gran interés en detectar *a priori* qué estructuras son suceptibles de bloquear el canal hERG.
Se han publicado varios trabajos de modelos estructurales de fármacos que inhiben este canal, entre ellos éste (q ue encontraréis adjunto a la PEC2)
Cavalli A, Poluzzi E, De Ponti F, Recanatini M. Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K^+ channel blockers. *J Med Chem* 2002; 45: 3844–53.
En el *chart 1* del artículo (pág 2,3 y 4) podéis observar una serie de fármacos (antibióticos, antiinflamatorios,...). Todos ellos inhiben el canal hERG de potasio con una cierta actividad, medida con el valor IC50 (observad la tabla 1). Buscad información sobre este valor IC50 (por ejemplo, en la wikipedia) y explicad qué representa. ¿Cúal de los fármacos que nos nombra *Cavalli* en el artículo bloquea con una actividad mayor al canal?

Las curvas dosis-respuesta hacen una correlación entre la dosis del fármaco y su efecto medido, que puede ser expresado de forma lineal (curva hiperbólica) o logarítmica (curva sigmoidea). Ojo, por que las graficas de agonistas y antagonistas son distintas.



El IC50 es la concentración que produce el 50% de inhibición de un parámetro o respuesta. En escala logarítmica se denomina pIC50.(como el pH o pK). Es el equivalente al EC50 en fármacos agonistas. (Véanse las dos figura de arriba)

En la siguiente grafica vemos los efectos de agonista y antagonistas mezclados, y como se puede interpolar de forma aproximada el IC50. (por ejemplo un fármaco que inhibe la respuesta de otra sustancia activa). Mediante la forma y evolución de las curvas de concentración respuesta se puede estudiar el mecanismo de reacción de un fármaco.



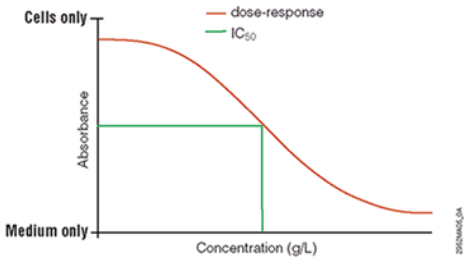


Table 1. Observed and Calculated HERG K+ Channel Blocking Activity of Compounds 1−31

compound	IC_{50} (nM)	$pIC_{50,calcd}$	$pIC_{50,fit}$[a]	Δ
astemizole (1)	0.9[b]	9.04	8.53	0.51
cisapride (2)	6.5[b]	8.19	7.96	0.23
E-4031 (3)	7.7[b]	8.11	7.85	0.26
dofetilide (4)	9.5−15[b]	7.91	7.67	0.24
sertindole (5)	14[b]	7.85	8.04	−0.19
pimozide (6)	18[b]	7.74	7.80	−0.06
haloperidol (7)	28.1[b]	7.55	7.58	−0.03
droperidol (8)	32.2[b]	7.49	7.82	−0.33
thioridazine (9)	35.7[b]	7.45	7.23	0.22
terfenadine (10)	56−204[b]	6.89	7.22	−0.33
verapamil (11)	143[b]	6.84	7.05	−0.21
domperidone (12)	162[b]	6.79	6.88	−0.09
loratadine (13)	173[b]	6.76	5.83	0.93
halofantrine (14)	196.9[c]	6.70	6.81	−0.11
mizolastine (15)	350[b]	6.45	6.65	−0.20
bepridil (16)	550[d]	6.26	6.30	−0.04
azimilide (17)	560[c]	6.25	6.15	0.10
mibefradil (18)	1430[d]	5.84	5.75	0.09
chlorpromazine (19)	1470[c]	5.83	5.68	0.15
imipramine (20)	3400[c]	5.47	5.98	−0.51
granisetron (21)	3730[b]	5.42	5.64	−0.22
dolasetron (22)	5950[b]	5.22	4.99	0.23
perhexiline (23)	7800[b]	5.11	5.18	−0.08
amitriptyline (24)	10000[b]	5.00	5.66	−0.66
diltiazem (25)	17300[b]	4.76	5.02	−0.26
sparfloxacin (26)	18000−34400[c]	4.58	4.39	0.19
glibenclamide (27)	74000[e]	4.13	4.07	0.06
grepafloxacin (28)	50000−104000[c]	4.11	4.35	−0.24
sildenafil (29)	100000[b]	4.00	3.50	0.50
moxifloxacin (30)	103000−129000[c]	3.93	3.82	0.11
gatifloxacin (31)	130000[c]	3.89	4.16	−0.27

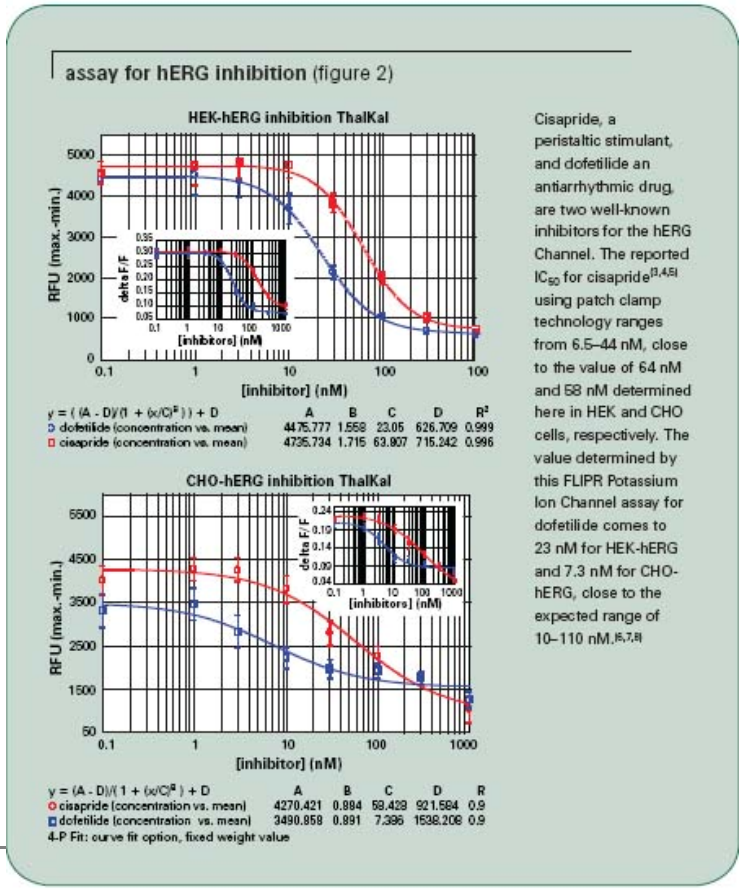
[a] Calculated from the non-cross-validated CoMFA model. [b] In human embryonic kidney (HEK) cells. [c] In Chinese hamster ovary (CHO) cells. [d] In African green monkey kidney derived cell line COS-7. [e] In neuroblastoma cells.

alignment. The template chosen for this study was the crystal structure of astemizole (1), because this molecule is one of the most potent long QT-inducing drugs, and HERG channel blockers (IC_{50} = 0.9 nM, Table 1). The astemizole crystal structure was directly retrieved from

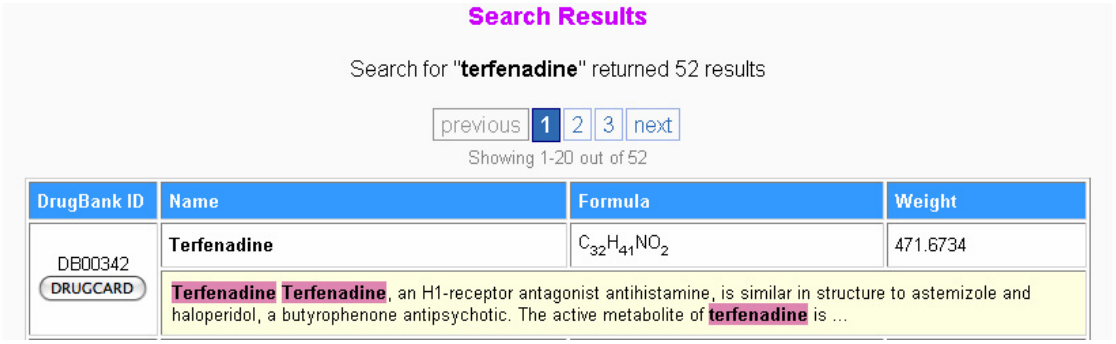
El fármaco con mayor potencia de bloqueo es el astemizol. (Menor IC50)

En el mercado hay dispositivos estándar para medir el IC50 del hERG.

http://www.moleculardevices.com:8080/pages/reagents/f_potassium_kit.html



Drugbank es una base de datos que proporciona datos sobre fármacos. Cada registro contiene datos sobre la estructura química y sobre las dianas terapéuticas del fármaco.
http://www.drugbank.ca/
Escojamos por ejemplo el fármaco *terfenadine*. Al clicar en la primera entrada obtenemos características del fármaco en cuestión. Es un antihistamínico. ¿Qué podemos decir acerca de los efectos secundarios que tiene el fármaco? ¿En qué año se retiró del mercado? ¿Qué fármaco le sustituyó?



Indication	For the treatment of allergic rhinitis, hay fever, and allergic skin disorders.
Pharmacology	Terfenadine, an H1-receptor antagonist antihistamine, is similar in structure to astemizole and haloperidol, a butyrophenone antipsychotic. The active metabolite of terfenadine is fexofenadine.
Mechanism of Action	Terfenadine competes with histamine for binding at H1-receptor sites in the GI tract, uterus, large blood vessels, and bronchial muscle. This reversible binding of terfenadine to H1-receptors suppresses the formation of edema, flare, and pruritus resulting from histaminic activity. As the drug does not readily cross the blood-brain barrier, CNS depression is minimal.
Absorption	On the basis of a mass balance study using 14C labeled terfenadine the oral absorption of terfenadine was estimated to be at least 70%
Toxicity	Mild (e.g., headache, nausea, confusion), but adverse cardiac events including cardiac arrest, ventricular arrhythmias including torsades de pointes and QT prolongation have been reported. LD_{50}=mg/kg (orally in mice)
Protein Binding	70%

Efectos secundarios

Los efecto secundarios son variados, pero sobre todo la lista de interacciones con otros medicamentos es elocuente.

Absorption	On the basis of a mass balance study using 14C labeled terfenadine the oral absorption of terfenadine was estimated to be at least 70%
Toxicity	Mild (e.g., headache, nausea, confusion), but adverse cardiac events including cardiac arrest, ventricular arrhythmias including torsades de pointes and QT prolongation have been reported. LD_{50}=mg/kg (orally in mice)
Protein Binding	70%
Biotransformation	Hepatic
Half Life	3.5 hours
Dosage Forms	Form: Tablet; Route: Oral
Patient Information	Show
Contraindications	Show
Interactions	Show

Drug	Interaction
Acetophenazine	Increased risk of cardiotoxicity and arrhythmias
Amiodarone	Increased risk of cardiotoxicity and arrhythmias
Amitriptyline	Increased risk of cardiotoxicity and arrhythmias
Amoxapine	Increased risk of cardiotoxicity and arrhythmias
Amprenavir	Increased risk of cardiotoxicity and arrhythmias
Aprepitant	Increased risk of cardiotoxicity and arrhythmias
Bepridil	Increased risk of cardiotoxicity and arrhythmias
Chlorpromazine	Increased risk of cardiotoxicity and arrhythmias

En http://www.drugs.com/sfx/terfenadine-side-effects.html tenemos un listado de sus efectos secundarios:

By body system

Cardiovascular side effects

Cardiovascular toxicity has been associated with the use of terfenadine. Reported effects include dizziness, syncopal episodes, palpitations, ventricular arrhythmias (including torsades de pointes), cardiac arrest, and cardiac death.

Terfenadine use may cause prolongation of the QT interval. Most cardiovascular events related to terfenadine occur in patients taking more than the recommended dose of 60 mg twice a day, in patients with higher-than-normal terfenadine serum concentrations, and in patients who are at risk for cardiac events. Patients with liver disease are also at risk of cardiovascular toxicity due to potential accumulation of the drug. Other predisposing factors for cardiovascular toxicity include congenital forms of QT interval prolongation, coronary artery disease, and electrolyte disorders including hypokalemia and hypomagnesemia. Although rare, arrhythmias have been reported in patients on recommended doses without apparent risk factors.

Nervous system side effects

Headaches have been reported in approximately 6% of treated patients. Terfenadine has not been demonstrated to cause significant drowsiness, sedation, or impaired psychomotor skills.

Gastrointestinal side effects

Gastrointestinal effects of terfenadine are rare and include nausea and dry mouth.

Genitourinary side effects

Urinary retention has been reported rarely.

A study of the effects of terfenadine on the urination of eight healthy male volunteers and 11 males with benign prostatic hypertrophy was not able to confirm a consistent effect on voiding characteristics.

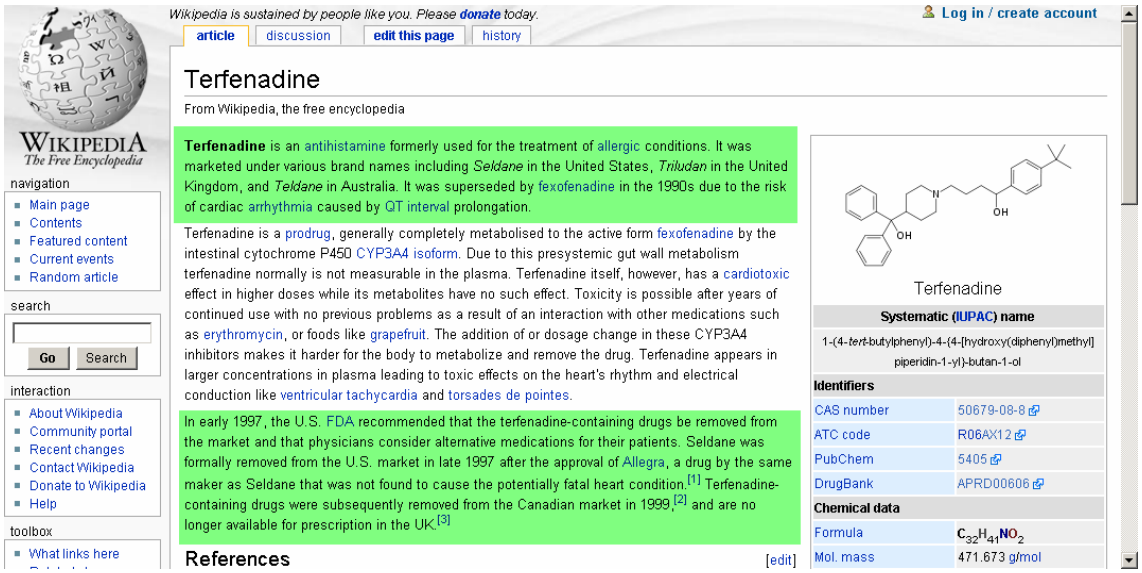
Hepatic side effects

Acute hepatitis, cholestatic hepatitis, and jaundice have been reported rarely in patients taking terfenadine. Hepatic dysfunction has been reversible upon discontinuation of the drug.

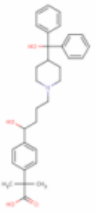
Añó de retirada

Name	Terfenadine
Drug Type	• Approved • Small Molecule • Withdrawn
Description	In the U.S., Terfenadine was superseded by fexofenadine in the 1990s due to the risk of cardiac arrhythmia caused by QT interval prolongation.
Synonyms	1. Ternadin
	1. Aldaban

Fue retirado por fexofenadine en los años 90 (en años distintos es cada lugar).



Podemos buscar el fármaco que sustituyó a *terfenadine.* Podemos ver que este fármaco no produce arrítmias. Observando las características farmacocinéticas (http:// en.wikipedia.org/wiki/ADME), este nuevo fármaco, ¿tiene mayor o menor absorción oral que el *terfenadine*?

Description	Fexofenadine hydrochloride (Allegra) is an antihistamine drug used in the treatment of hayfever and similar allergy symptoms. It was developed as a successor of and alternative to terfenadine. Fexofenadine, like other second and third-generation antihistamines, does not readily pass through the blood-brain barrier, and so causes less drowsiness than first-generation histamine-receptor antagonists.
Synonyms	1. Carboxyterfenadine 2. Fexofenadine hydrochloride 3. Fexofendine 4. Terfenadine acid metabolite 5. Terfenadine carboxylate 6. Terfenadine-COOH
Brand Names	1. Allegra
Brand Mixtures	1. Allegra-D (Fexofenadine hydrochloride + Pseudoephedrine hydrochloride)
Chemical IUPAC Name	2-[4-[1-hydroxy-4-[4-[hydroxy-di(phenyl)methyl]piperidin-1-yl]butyl]phenyl]-2-methylpropanoic acid
Chemical Formula	$C_{32}H_{39}NO_4$
Chemical Structure	
CAS Registry Number	83799-24-0

Absorption	33%

Tiene menor absorción que el terfenadine (aprox 70%).

Ya que actualmente se conoce la estructura tridimensional del canal de potasio, ¿qué tipo de estudios computacionales haríamos (directos / indirectos) para descubrir si un fármaco inhibe el canal? ¿En qué consisten este tipo de estudios?

Responder a estas dos preguntas no es obvio, y seguramente podría ser el tema de una tesis doctoral. Intentaré con la información que ya conocía y con la que he podido recopilar estos días, responder de forma resumida, pero completa y coherente. Lo inmediato es reformular las preguntas, para que sean coherentes con la explicación

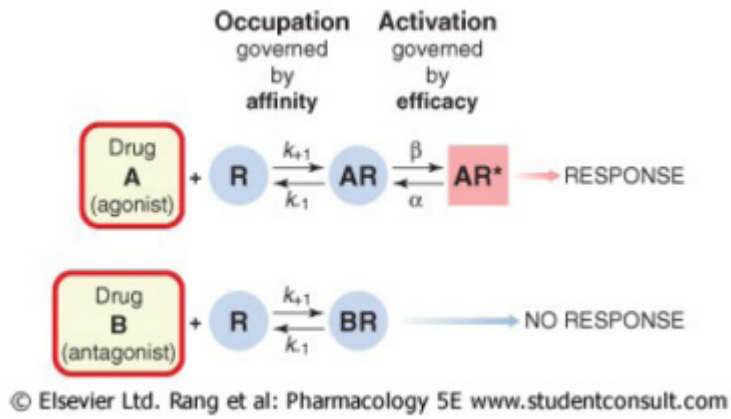
- ¿Qué tipos de técnicas computacionales existen para descubrir si un fármaco inhibe un proceso biológico, y en que consisten?
- ¿Qué técnicas podríamos aplicar en el caso de la inhibición por fármacos del canal herg?.

Vayamos paso a paso:

Técnicas computacionales de diseño de fármacos.

Si bajamos nivel molecular, un fármaco es una molécula que interacciona (de forma reversible o no reversible) con otra molécula con una función biológica determinada. A la molécula que lleva a cabo la función se le denomina "diana" o "receptor". Es evidente que la acción farmacológica puede ser benigna, (por ejemplo, detener el crecimiento bacteriano) o por el contrario maligna para el organismo o inhibitoria de un proceso biológico. En nuestro caso hemos visto en el artículo de Cavalleri, como algunos fármacos inhiben el canal de potasio provocando arritmias.

Por tanto, un fármaco es una sustancia capaz de modificar la actividad celular. El fármaco no origina mecanismos o reacciones desconocidos por la célula hasta entonces, sino que se limita a estimular o a inhibir los procesos propios de la célula. Para ello, el fármaco primero debe asociarse a las dianas con las cuales, y en razón de sus respectivas estructuras moleculares, pueda generar enlaces de unión que casi siempre son reversibles (Si la unión es muy intensa o el fármaco provoca grandes modificaciones en la molécula diana el enlace puede ser irreversible).

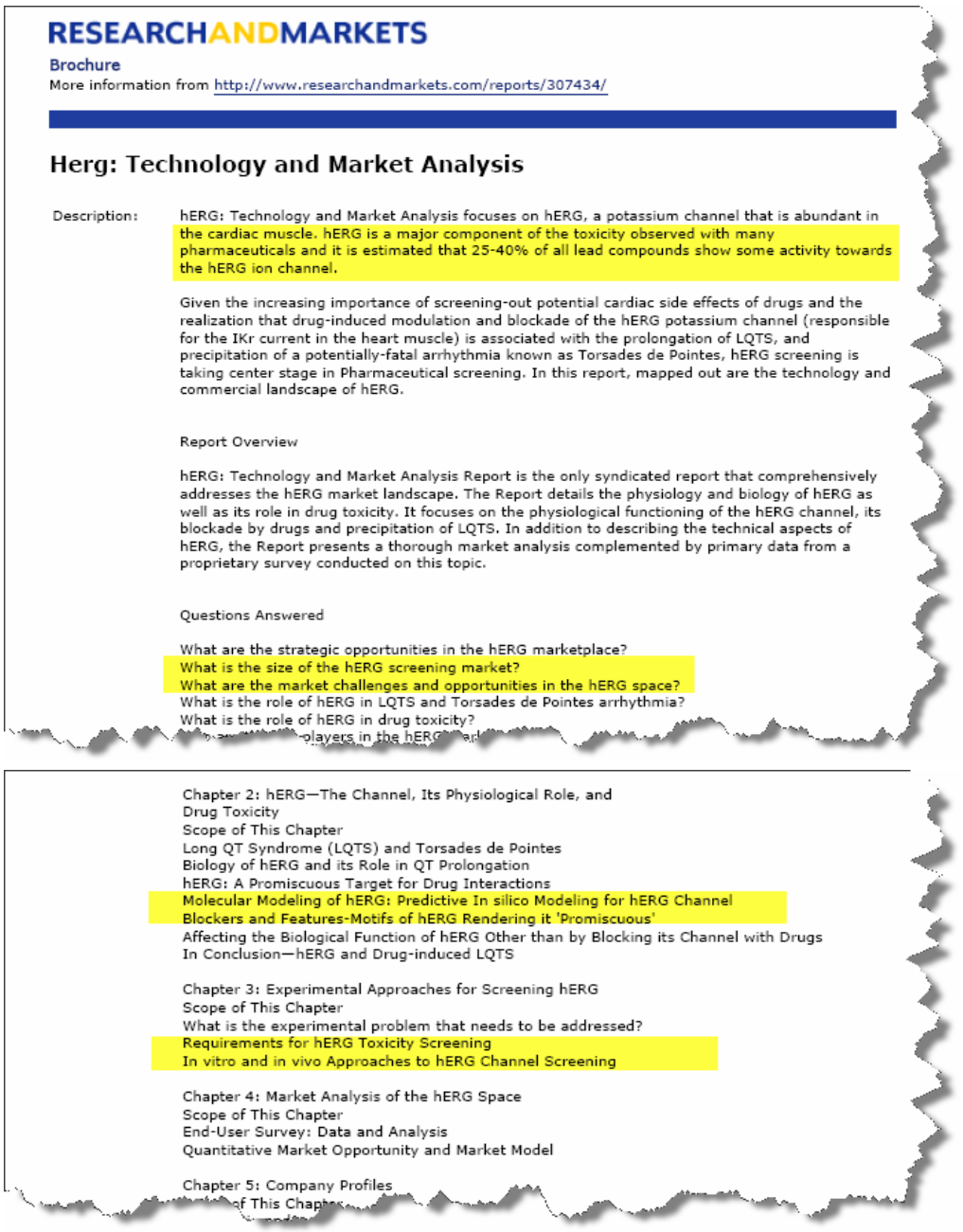


La caracterización y predicción de las constantes de equilibrio de los enlaces fármaco – diana (o ligando – receptor), es uno de los objetivos del "Diseño de fármacos por computador". Tradicionalmente, las constantes de equilibrio o sus "equivalentes" EC50 e IC50 se han determinado experimentalmente por medio de las curvas Respuesta – Dosis, como ya hemos visto.

El siguiente link (http://www.researchandmarkets.com/reports/307434/) lleva al "brochure" de un estudio de mercado y de tecnología sobre el Herg. ¿Por qué es tan importante el estudio farmacologico del herg?

- Entre el 25 y el 45% de los fármacos "cabezas de serie" muestran alguna actividad hacia el canal de potasio.

- Un fármaco tarda entre 5 o 6 años en desarrollarse, ¿Se puede perder el tiempo en múltiples y costosos experimentos de screening?



RESEARCHANDMARKETS

Brochure

More information from http://www.researchandmarkets.com/reports/307434/

Herg: Technology and Market Analysis

Description: hERG: Technology and Market Analysis focuses on hERG, a potassium channel that is abundant in the cardiac muscle. hERG is a major component of the toxicity observed with many pharmaceuticals and it is estimated that 25-40% of all lead compounds show some activity towards the hERG ion channel.

Given the increasing importance of screening-out potential cardiac side effects of drugs and the realization that drug-induced modulation and blockade of the hERG potassium channel (responsible for the IKr current in the heart muscle) is associated with the prolongation of LQTS, and precipitation of a potentially-fatal arrhythmia known as Torsades de Pointes, hERG screening is taking center stage in Pharmaceutical screening. In this report, mapped out are the technology and commercial landscape of hERG.

Report Overview

hERG: Technology and Market Analysis Report is the only syndicated report that comprehensively addresses the hERG market landscape. The Report details the physiology and biology of hERG as well as its role in drug toxicity. It focuses on the physiological functioning of the hERG channel, its blockade by drugs and precipitation of LQTS. In addition to describing the technical aspects of hERG, the Report presents a thorough market analysis complemented by primary data from a proprietary survey conducted on this topic.

Questions Answered

What are the strategic opportunities in the hERG marketplace?
What is the size of the hERG screening market?
What are the market challenges and opportunities in the hERG space?
What is the role of hERG in LQTS and Torsades de Pointes arrhythmia?
What is the role of hERG in drug toxicity?

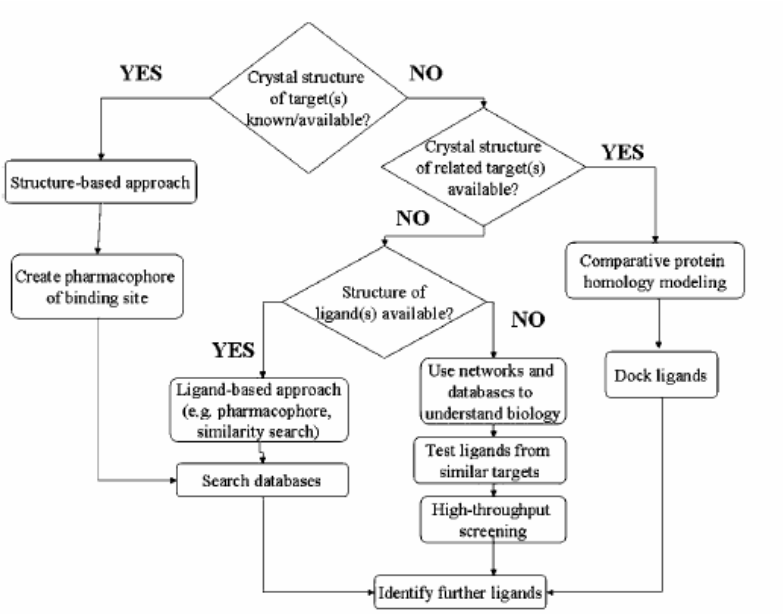
Chapter 2: hERG—The Channel, Its Physiological Role, and Drug Toxicity
Scope of This Chapter
Long QT Syndrome (LQTS) and Torsades de Pointes
Biology of hERG and its Role in QT Prolongation
hERG: A Promiscuous Target for Drug Interactions
Molecular Modeling of hERG: Predictive In silico Modeling for hERG Channel Blockers and Features-Motifs of hERG Rendering it 'Promiscuous'
Affecting the Biological Function of hERG Other than by Blocking its Channel with Drugs
In Conclusion—hERG and Drug-induced LQTS

Chapter 3: Experimental Approaches for Screening hERG
Scope of This Chapter
What is the experimental problem that needs to be addressed?
Requirements for hERG Toxicity Screening
In vitro and in vivo Approaches to hERG Channel Screening

Chapter 4: Market Analysis of the hERG Space
Scope of This Chapter
End-User Survey: Data and Analysis
Quantitative Market Opportunity and Market Model

Chapter 5: Company Profiles

Tradicionalmente las técnicas de diseño de fármacos por ordenador se dividen en dos clases dependiendo de si se conoce o no la estructura de la diana.

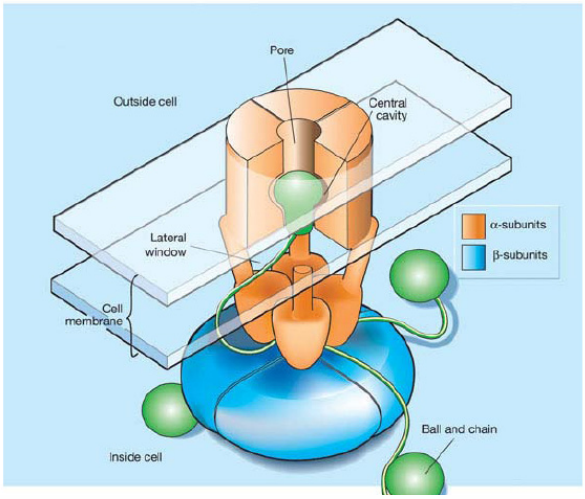


- Por una parte tenemos los métodos indirectos o basados en el ligando, cuando no conocemos la estructura de la diana. En este caso el método más utilizado actualmente es el 3D QSAR (no entraré en muchos detalles ya que en la bibliografía está ampliamente descrito). Este es el método empleado en el artículo de cavalli. El método consiste en construir un "farmacoforo" a partir de la superposición de múltiples estructuras de fármacos con actividad conocida, partiendo de un farmacoforo ideal (normalmente el que mayor actividad experimental presenta). El ajuste de la estructura tridimensional y de los grupos activos (ya sean hidrofobos, hidroxi, aminas, etc) con el valor experimental de la constante de equilibrio (o cualquier otra magnitud que pueda ser util, en el caso de Cavalli es el IC50), proporciona una ecuación de calibrado. Mediante la ecuación de calibrado se puede determinar la acción farmacológica de un ligando problema.

- Y los métodos directos, en los que si que conocemos la estructura de la diana. Dentro de este grupo tenemos los métodos de dockign y screeninig virtual, y luego los basados en Dinámica y mecánica molecular. Estos métodos se basan en simular el movimiento intramolecular mediante "bolas y muelles" (campos de fuerza), y el movimiento intermolecular con mecánica estadística (Monte-Carlo). El uso de uno u otro depende del problema a tratar. Si lo que se está buscando es un uso "farmacéutico" (determinar si un ligando es compatible), es preferible el docking (por rapidez y versatilidad, y porque está implementado en el software comercial usado por las farmacéuticas). Si se esta estudiando la biología del problema, se elegiría alguno de los segundos. No obstante, los métodos directos no son totalmente ab-initio, ya que siempre se necesita un "training" del método para que los resultados observables se adapten a los requerimientos del problema (por ejemplo, no es inmediato calcular un IC50 a partir de un docking, hay que parametrizar en base a valores experimentales)

Esta subdivisión no es azarosa. Mediante los "métodos de la química computacional" (y no entrare en más detalles sobre los mismos …..) se pueden determinar a priori las constantes de equilibrio que regulan la acción farmacológica, el problema es que cuanto mejor conozcamos las estructura tridimensional de la diana y

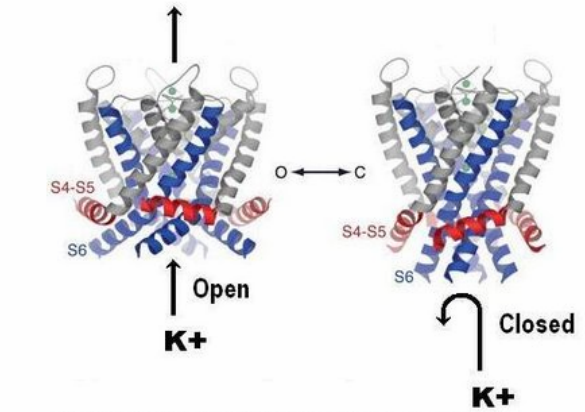
del ligando más fiables y precisos son. Iría mas lejos, no solo la estructura, sino también el mecanismo de acción biológica. Otro problema adicional es que normalmente las dianas son proteínas, y en muchos casos es "imposible" conocer su estructura tridimensional en el medio.



Como es el canal herg y como funciona?.

La estructura del canal de potasio es bien conocida. Es un tetrámero de proteína que está insertado en una bicapa lipídica. Esto que parece una parrafada en realidad es muy importante: Cualquier cálculo o simulación que hagamos debería tener en cuenta este hecho.

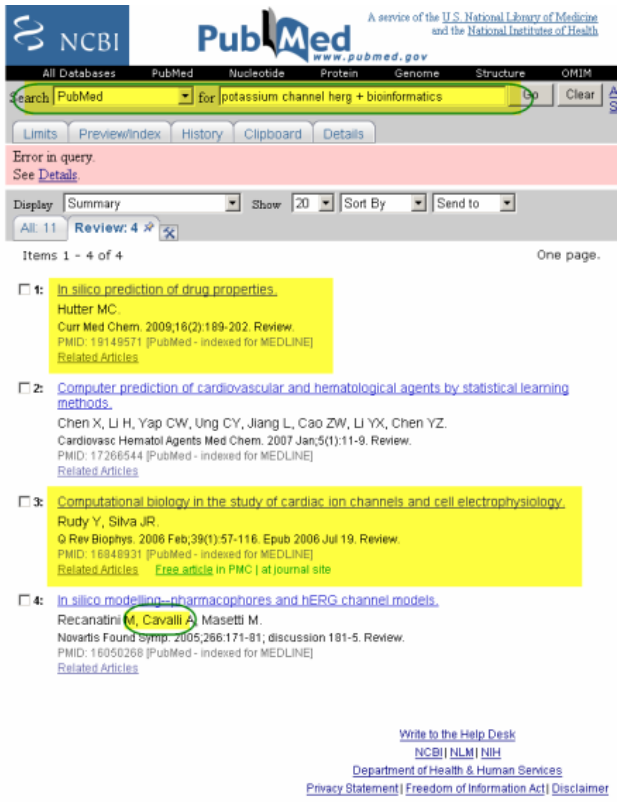
Otra cuestión es su actividad. El canal se abre y cierra por los dos lados a diferentes velocidades, impidiendo o facilitando el paso de los iones potasio a ambos lados de la membrana. Como sabemos, este pequeño interruptor proteico se activa por diferencias de potencial (acumulación de cargas positivas y negativas) entre ambos lados de la membrana. Cuando dispara cambia de conformación. En resumidas cuentas, el canal tiene distintas conformaciones tridimensionales, y por tanto si elegimos un método directo habrá que tener en cuenta las dos conformaciones. (en este link hay una animación que facilita la comprensión del fenómeno http://www.hergchannel.com/index.html).



K+ channel in the open (left) and closed (right) conformations. (figure modified from Long et al. 2005)

Pequeña revisión bibliografica sobre el herg.

Una breve revisión en PubMed para comprobar el estado del arte me hace responder instintivamente a la pregunta, ¡¡Que problema exactamente queremos solucionar¡¡. Veamos que están haciendo los investigadores en los últimos años (es una selección de articulos a simple vista, no es un estudio bibliográfico completo): Resaltare en cada caso lo que pueda resultar curioso. Los artículos que considere más interesantes estarán señalados en rosa, los que considere accesorios, pero necesarios para explicar las afirmaciones de este trabajo, simplemente tendrán las partes interesantes del resumen en amarillo o verde. Al final si me da tiempo¡¡¡¡ intentaré resumir lo aprendido en unas pocas imágenes¡¡¡ que es más rápido.



http://circ.ahajournals.org/cgi/content/abstract/82/6/2235

Channel specificity in antiarrhythmic drug action. Mechanism of potassium channel block and its role in suppressing and aggravating cardiac arrhythmias

TJ Colatsky, CH Follmer and CF Starmer

Division of Experimental Therapeutics, Wyeth-Ayerst Research, Princeton, New Jersey 08543.

Although work on class III antiarrhythmics remains at an early stage, these agents still appear to possess greater efficacy and less proarrhythmia than conventional class I agents in those experimental arrhythmia models considered to be most representative of the clinical situation. Although prolongation of repolarization carries with its own tendency for pause-dependent arrhythmogenesis (i.e., torsade de pointes), available data suggest that this may be a function of nonspecificity in potassium channel block rather than a general characteristic of class III activity. The availability of new and more selective blockers of specific cardiac potassium channels under development as class III agents have already helped to clarify basic questions about the ionic mechanism of repolarization in the heart, and one hopes that a growing clinical data base will eventually determine the relative safety and efficacy of these agents in preventing symptomatic and life-threatening arrhythmias.

http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B94RW-4TYB577-7&_user=10&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=e3ddae0d24297fd4c01946c1d7689b5d

Structure and dynamics of K channel pore-lining helices: a comparative simulation study.

I H Shrivastava, C E Capener, L R Forrest, and M S Sansom

Biophys J. 2000 January; 78(1): 79–92.

PMCID: PMC1300619

Abstract

Isolated pore-lining helices derived from three types of K-channel have been analyzed in terms of their structural and dynamic features in nanosecond molecular dynamics (MD) simulations while spanning a lipid bilayer. The helices were 1) M1 and M2 from the bacterial channel KcsA (Streptomyces lividans), 2) S5 and S6 from the voltage-gated (Kv) channel Shaker (Drosophila melanogaster), and 3) M1 and M2 from the inward rectifier channel Kir6.2 (human). In the case of the Kv and Kir channels, for which x-ray structures are not known, both short and long models of each helix were considered. Each helix was incorporated into a lipid bilayer containing 127 palmitoyloleoylphosphatidylcholine molecules, which was solvated with 4000 water molecules, yielding 20,000 atoms in each system. Nanosecond MD simulations were used to aid the definition of optimal lengths for the helix models from Kv and Kir. Thus the study corresponds to a total simulation time of 10 ns. The inner pore-lining helices (M2 in KcsA and Kir, S6 in Shaker) appear to be slightly more flexible than the outer pore-lining helices. In particular, the Pro-Val-Pro motif of S6 results in flexibility about a molecular hinge, as was suggested by previous in vacuo simulations (Kerr et al., 1996, Biopolymers. 39:503–515). Such flexibility may be related to gating in the corresponding intact channel protein molecules. Analysis of H-bonds revealed interactions with both water and lipid molecules in the water/bilayer interfacial region. Such H-bonding interactions may lock the helices in place in the bilayer during the folding of the channel protein (as is implicit in the two-stage model of membrane protein folding). Aromatic residues at the extremities of the helices underwent complex motions on both short (<10 ps) and long (>100 ps) time scales.

http://www.pnas.org/content/97/22/11683.extract

Channel structure and drug-induced cardiac arrhythmias

1. Robert S. Kass and . Candido Cabo

Electrical activity underlies the control of the frequency, strength, and duration of contraction of the heart. During the cardiac cycle, a regular rhythmic pattern must be established in time-dependent changes in ionic conductances to ensure events that underlie normal cardiac function. Electrical impulses, originating at the sino atrial node are conducted throughout the atria until they converge at the atrio-ventricular node, pass through the bundle of His and the Purkinje fiber conducting system, and eventually excite the working myocardial cells of both ventricles. Current flow through a large number of ion channels, exchange mechanisms, and pumps underlies and coordinates these electrical signals and alteration of the critical balance of these multiple current pathways can lead to disruptive, often fatal, rhythm disturbances: the cardiac arrhythmias. Although the cardiac arrhythmias form a complicated and diverse group reflecting the complexities of the ionic mechanisms underlying the electrical activity of the human heart, a surprisingly large number of rhythm disturbances are caused either directly or indirectly by mechanisms that prolong the duration of the action potential of the working myocardium (1, 2). Whereas prolongation of the cardiac ventricular action potential under controlled conditions can be, in principle, an effective mechanism to prevent certain types of re-entrant arrhythmias (3), excessive prolongation can be fatal. In this issue …

http://www.cell.com/biophysj/abstract/S0006-3495(01)76138-3

Brownian Dynamics Simulations of Interaction Between Scorpion Toxin Lq2 and Potassium Ion Channel

Meng Cui*, Jianhua Shen*, James M. Briggs , Xiaomin Luo*, Xiaojian Tan*, Hualiang Jiang*, , , Kaixian Chen* and Ruyun Ji*

Abstract

The association of the scorpion toxin Lq2 and a potassium ion (K+) channel has been studied using the Brownian dynamics (BD) simulation method. All of the 22 available structures of Lq2 in the Brookhaven Protein Data Bank (PDB) determined by NMR were considered during the simulation, which indicated that the conformation of Lq2 affects the binding between the two proteins significantly. Among the 22 structures of Lq2, only 4 structures dock in the binding site of the K+ channel with a high probability and favorable electrostatic interactions. From the 4 candidates of the Lq2-K+ channel binding models, we identified a good three-dimensional model of Lq2-K+ channel complex through triplet contact analysis, electrostatic interaction energy estimation by BD simulation and structural refinement by molecular mechanics. Lq2 locates around the extracellular mouth of the K+ channel and contacts the K+ channel using its -sheet rather than its -helix. Lys27, a conserved amino acid in the scorpion toxins, plugs the pore of the K+ channel and forms three hydrogen bonds with the conserved residues Tyr78(A C) and two hydrophobic contacts with Gly79 of the K+ channel. In addition, eight hydrogen-bonds are formed between residues Arg25, Cys28, Lys31, Arg34 and Tyr36 of Lq2 and residues Pro55, Tyr78, Gly79, Asp80, and Tyr82 of K+ channel. Many of them are formed by side chains of residues of Lq2 and backbone atoms of the K+ channel. Thirteen hydrophobic contacts exist between residues Met29, Asn30, Lys31 and Tyr36 of Lq2 and residues Pro55, Ala58, Gly79, Asp80 and Tyr82 of the K+ channel. These favorable interactions stabilize the association between the two proteins. These observations are in good agreement with the experimental results and can explain the binding phenomena between scorpion toxins and K+ channels at the level of molecular structure. The consistency between the BD simulation and the experimental data indicates that our three-dimensional model of Lq2-K+ channel complex is reasonable and can be used in further biological studies such as rational design of blocking agents of K+ channels and mutagenesis in both toxins and K+ channels.

http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.pharmtox.43.100901.140245

K+ channel structure-activity relationships and mechanisms of drug-induced qt prolongation

Colleen E. Clancy, Junko Kurokawa, Michihiro Tateyama, Xander H.T. Wehrens, and Robert S. Kass

▪ Abstract Pharmacological intervention, often for the purpose of treating syndromes unrelated to cardiac disease, can increase the vulnerability of some patients to life-threatening rhythm disturbances. This may be due to an underlying propensity stemming from genetic defects or polymorphisms, or structural abnormalities that provide a substrate allowing for the initiation of arrhythmic triggers. A number of pharmacological agents that have proven useful in the treatment of allergic reactions, gastrointestinal disorders, and psychotic disorders, among others, have been shown to reduce repolarizing K+ currents and prolong the QT interval on the electrocardiogram. Understanding the structural determinants of K+ channel blockade may provide new insights into the mechanism and rate-dependent effects of drugs on cellular physiology. Drug-induced disruption of cellular repolarization underlies electrocardiographic abnormalities that are diagnostic indicators of arrhythmia susceptibility.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1994938&tool=pmcentrez

Computational biology in the study of cardiac ion channels and cell electrophysiology

Yoram Rudy* and Jonathan R. Silva

The cardiac cell is a complex biological system where various processes interact to generate electrical excitation (the action potential, AP) and contraction. During AP generation, membrane ion channels interact nonlinearly with dynamically changing ionic concentrations and varying transmembrane voltage, and are subject to regulatory processes. In recent years, a large body of knowledge has accumulated on the molecular structure of cardiac ion channels, their function, and their modification by genetic mutations that are associated with cardiac arrhythmias and sudden death. However, ion channels are typically studied in isolation (in expression systems or isolated membrane patches), away from the physiological environment of the cell where they interact to generate the AP. A major challenge remains the integration of ion-channel properties into the functioning, complex and highly interactive cell system, with the objective to relate molecular-level processes and their modification by disease to whole-cell function and clinical phenotype. In this article we describe how computational biology can be used to achieve such integration. We explain how mathematical (Markov) models of ion-channel kinetics are incorporated into integrated models of cardiac cells to compute the AP. We provide examples of mathematical (computer) simulations of physiological and pathological phenomena, including AP adaptation to changes in heart rate, genetic mutations in SCN5A and HERG genes that are associated with fatal cardiac arrhythmias, and effects of the CaMKII regulatory pathway and β-adrenergic cascade on the cell electrophysiological function.

(ES UNA PENA NO DISPONER DE UNA COPIA DE ESTE ARTICULO)

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1978280&tool=pmcentrez

In silico modelling--pharmacophores and hERG channel models.

Recanatini M, Cavalli A, Masetti M.

In computational drug design, modelling studies are undertaken following two main strategies that depend on which information is available. If experimental data exist only for the molecules displaying the biological property of interest, a so-called ligand-based approach is taken; if information is available on the macromolecular target(s) of the compounds (e.g. proteins' 3D structures), target-based studies can be carried out. Recently, in the field of hERG K+-channel blocking drugs, pharmacophoric (ligand-based) studies started appearing aimed at determining the physicochemical features associated with the channel block, and also at predicting the hERG blocking potential of compounds. However, partial homology models (target-based) of the hERG channel have also been built and used as working tools to interpret electrophysiological and mutagenesis studies. Here, we review some of the ligand- and target-based in silico studies carried out on hERG, focusing on both their main characteristics and their meaning. In addition, we discuss some methodological aspects of the computational work that in our opinion should be considered, in view of the construction of reliable models possibly able to predict the functional behaviour of the channel system and the blocking potential of drugs.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=16848931

Computational biology in the study of cardiac ion channels and cell electrophysiology

ESTE ARTICULO ES IMPORTANTE POR QUE DEMUESTRA DE LA NECESIDAD D ALAS TECNICAS IN-SILICO

Yoram Rudy* and Jonathan R. Silva

The cardiac cell is a complex biological system where various processes interact to generate electrical excitation (the action potential, AP) and contraction. During AP generation, membrane ion channels interact nonlinearly with dynamically changing ionic concentrations and varying transmembrane voltage, and are subject to regulatory processes. In recent years, a large body of knowledge has accumulated on the molecular structure of cardiac ion channels, their function, and their modification by genetic mutations that are associated with cardiac arrhythmias and sudden death. However, ion channels are typically studied in isolation (in expression systems or isolated membrane patches), away from the physiological environment of the cell where they interact to generate the AP. A major challenge remains the integration of ion-channel properties into the functioning, complex and highly interactive cell system, with the objective to relate molecular-level processes and their modification by disease to whole-cell function and clinical phenotype. In this article we describe how computational biology can be used to achieve such integration. We explain how mathematical (Markov) models of ion-channel kinetics are incorporated into integrated models of cardiac cells to compute the AP. We provide examples of mathematical (computer) simulations of physiological and pathological phenomena, including AP adaptation to changes in heart rate, genetic mutations in SCN5A and HERG genes that are associated with fatal cardiac arrhythmias, and effects of the CaMKII regulatory pathway and β-adrenergic cascade on the cell electrophysiological function.

Thai, K. M. and G. F. Ecker (2008). "Classification models for HERG inhibitors by counter-propagation neural networks." Chem Biol Drug Des 72(4): 279-89.

Counter-propagation neural networks were used to develop computational models for classification and prediction of human ether-a-go-go-related-gene (hERG) potassium channel blockers. The data set used includes 285 compounds taken from literature sources and two sets of 2D molecular descriptors, one is based on 32 P_VSA descriptors derived from moe and the other comprises 11 descriptors retrieved by a feature selection method. The counter-propagation neural networks with a 3-dimensional output layer combined with a set of 11 hERG relevant descriptors showed best performance, especially in classifying compounds in the middle-activity class (hERG IC(50) = 1-10 microm). The total accuracy values obtained for training and test sets are 0.93-0.95 and 0.83-0.85, respectively. In each activity class (low, medium, high), 'Goodness of Hit lists' GH scores archived range from 0.89 to 0.97 for the training set and from 0.74 to 0.87 for the test set. This model thus provides possible strategies for improving the performance of predicting and classifying compounds having hERG IC(50) in the range of 1-10 microm.

(ES UNA PENA NO DISPONER DE UNA COPIA DE ESTE ARTICULO)

Masetti, M., A. Cavalli, et al. (2008). "Modeling the hERG potassium channel in a phospholipid bilayer: Molecular dynamics and drug docking studies." J Comput Chem 29(5): 795-808.

The hERG potassium channel has recently been a matter of extensive studies both at experimental and computational levels, because of its possible involvement in the potentially lethal drug-induced long QT syndrome. In this context, in the absence of an experimentally determined 3D structure, to acquire a detailed description of the channel pore and an understanding of atomic determinants for drug binding is of enormous interest at both academic and industrial levels. To contribute to this aim, we first built the open and closed states of the channel by homology modeling techniques, and then submitted both channel models to ns-time-scale MD simulations in explicit membrane. An in-depth analysis of the dynamical behavior of the channel pore with particular attention to the cavity volume was carried out. Finally, using hERG conformations coming from MD simulations, docking experiments were performed to identify a possible binding mode for the most potent hERG channel blocker so far known, the antihistaminic drug astemizole. We show that the combined use of MD and docking is suitable to identify a possible binding mode of drugs in a fairly good agreement with experiments. Moreover, the exploitation of MD snapshots in the docking experiments allowed us to capture some induced-fit effects related to the side chain conformations of Tyr652 and Phe656, which are residues playing a pivotal role in the hERG drug binding.

Inanobe, A., N. Kamiya, et al. (2008). "In silico prediction of the chemical block of human ether-a-go-go-related gene (hERG) K+ current." J Physiol Sci 58(7): 459-70.

A variety of compounds with different chemical properties directly interact with the cardiac repolarizing K(+) channel encoded by the human ether-a-go-go-related gene (hERG). This causes acquired forms of QT prolongation, which can result in lethal cardiac arrhythmias including torsades de pointes one of the most serious adverse effects of various therapeutic agents. Prediction of this phenomenon will improve the safety of pharmacological therapy and also facilitate the process of drug development. Here we propose a strategy for the development of an in silico system to predict the potency of chemical compounds to block hERG. The system consists of two sequential processes. The first process is a ligand-based prediction to estimate half-maximal concentrations for the block of compounds inhibiting hERG current using the relationship between chemical features and activities of compounds. The second process is a protein-based prediction that comprises homology modeling of hERG, docking simulation of chemical-channel interaction, analysis of the shape of the channel pore cavity, and Brownian dynamics simulation to estimate hERG currents in the presence and absence of chemical blockers. Since each process is a combination of various calculations, the criterion for assessment at each calculation and the strategy to integrate these steps are significant for the construction of the system to predict a chemical's block of hERG current and also to predict the risk of inducing cardiac arrhythmias from the chemical information. The principles and criteria of elemental computations along this strategy are described.

Imai, Y. N., S. Ryu, et al. (2009). "Docking model of drug binding to the human ether-a-go-go potassium channel guided by tandem dimer mutant patch-clamp data: a synergic approach." J Med Chem 52(6): 1630-8.

To characterize drug binding to the human ether-a-go-go related gene (hERG) channel, a synergic approach interplaying patch-clamp experiments and a docking study was developed. Mutations were introduced into concatenated dimers of the hERG channel that were assembled into a heterotetramer with mutated diagonal subunits. The binding affinities of three drugs (cisapride, terfenadine, and N-[4-[[1-[2-(6-methyl-2-pyridinyl)ethyl]-4-piperidinyl]carbonyl]phenyl]met hanesulfonamide dihydrochloride (E-4031, 1)) to a set of mutant channels were examined electrophysiologically to assess the involved residues, their number, and relative positions. Cisapride and 1 interacted with Tyr652 residues on adjacent subunits, while terfenadine interacted with Tyr652 residues on diagonal, but not on adjacent, subunits. Phe656 was involved in the binding of all three drugs, and Ser624 was found to be only involved in cisapride and 1. The docking models demonstrated that pi-pi and CH-pi interactions rather than cation-pi interaction play a key role in drug binding to the hERG channel.

Du, L., M. Li, et al. (2009). "The interactions between hERG potassium channel and blockers." Curr Top Med Chem 9(4): 330-8.

The human ether-a-go-go related gene (hERG) potassium channel is critical to the QT interval in the human heart measured by the electrocardiogram (ECG). The blockade of hERG would induce undesired lethal arrhythmia, named torsades de pointes (TdP), a rare but life-threatening symptom. Although a large number of experimental studies on hERG have been conducted so far, knowledge of how known ligands bind to hERG still remains sketchy and has been a major hindrance in the effort to designing novel medicinal molecules devoid of hERG activity in the hope of improving drug safety. This review summarizes several studies on ligand-hERG interactions by in silico receptor-based and ligand-based modeling approaches during recent years. These efforts could aid tremendously in understanding the determinants of ligand binding to hERG channel and the molecular basis of hERG channel blockade, and offer a more rational approach for the prediction of QT-prolongation liability and for the development of novel and safe non-cardiac agents.

Recanatini, M., A. Cavalli, et al. (2008). "Modeling HERG and its interactions with drugs: recent advances in light of current potassium channel simulations." ChemMedChem 3(4): 523-35.

The hERG K(+) channel is responsible for the rapid delayed rectifier current in cardiac myocytes, and a block of its functioning may be related with the (inherited or drug-induced) long QT syndrome. For this reason, in recent times, some interest has arisen around computational studies aimed at developing hERG/drug models for the prediction of drug binding (docking) modes, in view of the assessment of the hERG blocking potential. On the other hand, voltage-gated K(+) channels have been the subject of molecular simulations for several years, and rigorous protocols for studying the main aspects of their functions (permeation, gating, voltage sensing) have been published. In this article, we briefly introduce these classical computational works on K(+) channels, and then review in depth the reports on the latest advanced modeling studies on hERG. The aim is to put the hERG modeling work in the more general context of the ion channel simulations field, to show the peculiarity of hERG on the one side, and also to indicate some possible new avenues in the use of modeling techniques to increase our knowledge of this important channel.

On the Potential Functions used in Molecular Dynamics Simulations of Ion Channels

Benoît Rouxa and Simon Bernèchea

The determination of the structure of the KcsA K+ channel represents an extraordinary opportunity for understanding biological ion channels at the atomic level. In principle, molecular dynamics (MD) simulations based on detailed atomic models can complement the experimental data and help to characterize the microscopic factors that ultimately determine the permeation of ions through KcsA. A number of MD studies, broadly aimed at analyzing the dynamical motions of water molecules and ions in the KcsA channel, have now been reported (Guidoni et al., 1999; Allen et al., 1999; Shrivastava and Sansom, 2000; Åqvist and Luzhkov, 2000; Bernèche and Roux, 2000; Biggin et al., 2001; Luzhkov and Åqvist, 2001; Crouzy et al., 2001). The potential functions that were used to calculate the microscopic interatomic forces and generate the dynamical trajectory are listed in Table 1, where they can be seen to differ significantly. In particular, the atomic partial charges and the Lennard–Jones radii, which are at the heart of the potential function, varied widely. Furthermore, some include all atoms (AMBER and CHARMM PARAM22), whereas others are extended-atom models that treat only the polar hydrogens able to form hydrogen bonds explicitly (CHARMM PARAM19 and GROMOS). How these differences affect the results of MD calculations is an important concern of all scientists involved in investigations of ion channels, theoreticians and experimentalists alike. It is the goal of this short letter to discuss important aspects of potential functions related to MD studies of ion permeation.

COMP 65 Accurate prediction of binding modes and binding affinities of protein-ligand complexes

Richard A. Friesner, Department of Chemistry, Columbia University, 3000 Broadway, MC 3110, New York,

ESTA ES UNA DEMOSTRACION DE QUE SE NECESITA A LA QUIMICA COMPUTACIONAL PARA SEGUIR AVANZANDO

Over the past several years, we have developed novel methods and models for the prediction of binding modes and binding affinities of protein-ligand complexes. These methods include qualitatively improved empirical scoring functions for hydrogen bonding and hydrophobic interactions, the use of polarized charges in lead docking calculations, and rapid, robust computational methods for induced fit calculations. The talk will focus on the optimization and validation of these methods using large test suites of protein-ligand complexes (including a wide variety of series generated by medicinal chemistry), as well as challenging individual applications such as modeling of the binding of ligands to the HERG ion channel. Preliminary results indicate that the new methods are able to address not only virtual screening and enrichment, but also lead optimization, in an effective fashion.

http://pubs.acs.org/doi/abs/10.1021/jp013069h

The Potassium Ion Channel: Comparison of Linear Scaling Semiempirical and Molecular Mechanics Representations of the Electrostatic Potential

ESTA ES UNA DEMOSTRACION MAS DE QUE SE NECESITA A LA QUIMICA COMPUTACIONAL PARA SEGUIR AVANZANDO

AbstractFull Text HTMLHi-Res PDF[98 KB]Andrey A. Bliznyuk,*† Alistair P. Rendell,‡ Toby W. Allen,§ and Shin-Ho Chung§

The molecular electrostatic potential inside the potassium channel protein from Streptomyces lividans has been investigated using linear scaling semiempirical quantum chemical method, for a variety of geometries, with and without solvating water molecules. The results are compared with those given by a number of popular molecular mechanics force-fields. The difference between the quantum and molecular mechanics electrostatic potentials due to the protein exceeds 30 kcal/mol within the narrow selectivity filter of the channel and is attributed to the neglect of

electronic effects, e.g., polarization, in the molecular mechanics force-fields. In particular, mutual electronic interactions between four threonine residues in the selectivity filter are found to have a large effect on the electrostatic potential. Calculations in the presence of water molecules suggest that molecular mechanics methods also overestimate the stabilization of the cation inside the ion channel. The molecular electrostatic potentials computed by molecular mechanics force-fields expressed relative to bulk water, however, reveal a much smaller error

Conclusiones

(no señalo nada por que todo es importante, no esta por orden de importancia)

1. La animación de http://www.hergchannel.com/index.html realmente es verídica y parece que explica el funcionamiento del canal de potasio. En mayor o menor medida la debemos a las técnicas de mecánica y dinámica molecular (a las dos).
2. Los mecanismos moleculares y biofísicos del canal de potasio están perfectamente delimitados. En concreto resaltaría el uso de pulpos como organismo modelo para los ensayos, y las modelizaciones incluyendo la doble capa y simulación de disolvente
3. Las técnicas directas sirven a la industria farmacéutica para "afinar" las indirectas (véanse los últimos artículos de Cavalli sobre el herg). Si queremos explicar el problema a fondo necesitamos técnicas directas.
4. Aunque por homología esté establecida la estructura y funcionamiento del herg humano, eso no sirve a las autoridades…. Primero hay que seguir haciendo ensayos y luego ….. hay que cristalizar la proteína del canal humano, y luego intentar en micelas (con la membrana lipídica incluida) reconstruir el canal herg, y de esta forma por RMN (como se explica en uno de los artículos) intentar reconstruir la estructura y mecanismo de acción. Los métodos mecano-cuánticos puros serán de vital importancia para interpretar la información de salida (CNDO, MNDO etc…).
5. Tenemos múltiples ejemplos de dockign y simulación para el herg, frente a un fármaco nuevo y sin "parecido" con otros ya existentes, la elección es clara frente a los costosos experimentos in-vivo. Véase por ejemplo el artículo del veneno de la serpiente y escorpión. Parece que los centros activos del canal de potasio están cada vez mejor definidos.
6. Con los actuales avances en tiempo de cálculo y disponibilidad de software, cada vez es más creíble el hecho de calcular propiedades farmacológicas mediante modulado molecular.

Respuesta a que métodos utilizaríamos???.

- Pues eso¡¡, depende del ligando? …. si hay alguno parecido en la bibliografía mejor¡¡¡.
- Si es similar a los fármacos existentes, usaremos indirectos (3D QSAR . Existe software comercial disponible que hace el proceso muy automático), si no hay nada parecido, método directo (docking).

Cosas interesantes.

RESEARCH ARTICLES

The Structure of the Potassium Channel: Molecular Basis of K^+ Conduction and Selectivity

Declan A. Doyle, João Morais Cabral, Richard A. Pfuetzner, Anling Kuo, Jacqueline M. Gulbis, Steven L. Cohen, Brian T. Chait, Roderick MacKinnon*

The potassium channel from *Streptomyces lividans* is an integral membrane protein with sequence similarity to all known K^+ channels, particularly in the pore region. X-ray analysis with data to 3.2 angstroms reveals that four identical subunits create an inverted teepee, or cone, cradling the selectivity filter of the pore in its outer end. The narrow selectivity filter is only 12 angstroms long, whereas the remainder of the pore is wider and lined with hydrophobic amino acids. A large water-filled cavity and helix dipoles are positioned so as to overcome electrostatic destabilization of an ion in the pore at the center of the bilayer. Main chain carbonyl oxygen atoms from the K^+ channel signature sequence line the selectivity filter, which is held open by structural constraints to coordinate K^+ ions but not smaller Na^+ ions. The selectivity filter contains two K^+ ions about 7.5 angstroms apart. This configuration promotes ion conduction by exploiting electrostatic repulsive forces to overcome attractive forces between K^+ ions and the selectivity filter. The architecture of the pore establishes the physical principles underlying selective K^+ conduction.

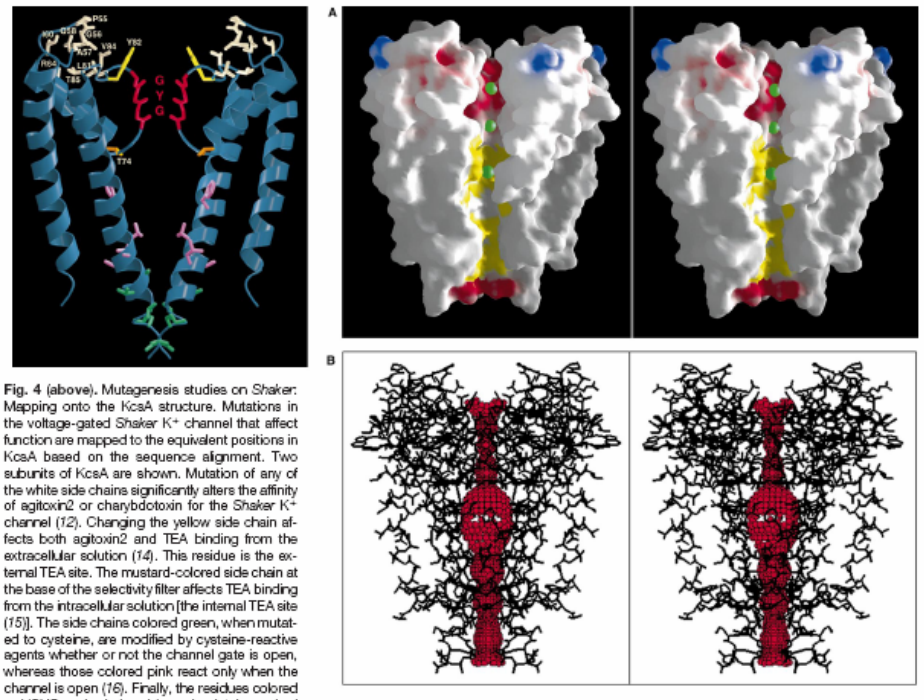


Fig. 4 (above). Mutagenesis studies on *Shaker*: Mapping onto the KcsA structure. Mutations in the voltage-gated *Shaker* K^+ channel that affect function are mapped to the equivalent positions in KcsA based on the sequence alignment. Two subunits of KcsA are shown. Mutation of any of the white side chains significantly alters the affinity of agitoxin2 or charybdotoxin for the *Shaker* K^+ channel (*12*). Changing the yellow side chain affects both agitoxin2 and TEA binding from the extracellular solution (*14*). This residue is the external TEA site. The mustard-colored side chain at the base of the selectivity filter affects TEA binding from the intracellular solution [the internal TEA site (*15*)]. The side chains colored green, when mutated to cysteine, are modified by cysteine-reactive agents whether or not the channel gate is open, whereas those colored pink react only when the channel is open (*16*). Finally, the residues colored red (GYG, main chain only) are absolutely required for K^+ selectivity (*4*). This figure was prepared with MOLSCRIPT and RASTER-3D. **Fig. 5 (right).** Molecular surface of KcsA and contour of the pore. (**A**) A cutaway stereoview displaying the solvent-accessible surface of the K^+ channel colored according to physical properties. Electrostatic potential was calculated with the program GRASP, assuming an ionic strength equivalent to 150 mM KCl and dielectric constants of 2 and 80 for protein and solvent, respectively. Side chains of Lys, Arg, Glu, and Asp residues were assigned single positive or negative charges as appropriate, and the surface coloration varies smoothly from blue in areas of high positive charge through white to red in negatively charged regions. The yellow areas of the surface are colored according to carbon atoms of the hydrophobic (or partly so) side chains of several semi-conserved residues in the inner vestibule (Thr^{75}, Ile^{100}, Phe^{103}, Thr^{107}, Ala^{108}, Ala^{111}, Val^{115}). The green CPK spheres represent K^+ ion positions in the conduction pathway. (**B**) Stereoview of the entire internal pore. Within a stick model of the channel structure is a three-dimensional representation of the minimum radial distance from the center of the channel pore to the nearest van der Waals protein contact. The display was created with the program HOLE (*34*).

Biophysical Journal Volume 83 October 2002 1867–1876

Open-State Models of a Potassium Channel

Philip C. Biggin and Mark S. P. Sansom
Laboratory of Molecular Biophysics, Department of Biochemistry, The University of Oxford, Oxford OX1 3QU, United Kingdom

ABSTRACT The structure of the bacterial potassium channel, KcsA, corresponds to the channel in a closed state. Two lines of evidence suggest that the channel must widen its intracellular mouth when in an open state: 1) internal block by a series of tetraalkylammonium ions and 2) spin labeling experiments. Thus it is known that the protein moves in this region, but it is unclear by how much and the mechanisms that are involved. To address this issue we have applied a novel approach to generate plausible open-state models of KcsA. The approach can be thought of as placing a balloon inside the channel and gradually inflating it. Only the protein sees the balloon, and so water is free to move in and out of the channel. The balloon is a van der Waals sphere whose parameters change by a small amount at each time step, an approach similar to methods used in free energy perturbation calculations. We show that positioning of this balloon at various positions along the pore axis generates similar open-state models, thus indicating that there may be a preferred pathway to an open state. We also show that the resulting structures from this process are conformationally unstable and need to undergo a relaxation process for up to 4 ns. We show that the channel can relax into a new state that has a larger pore radius at the region of the intracellular mouth. The resulting models may be useful in exploring models of the channel in the context of ion permeation and blocking agents.

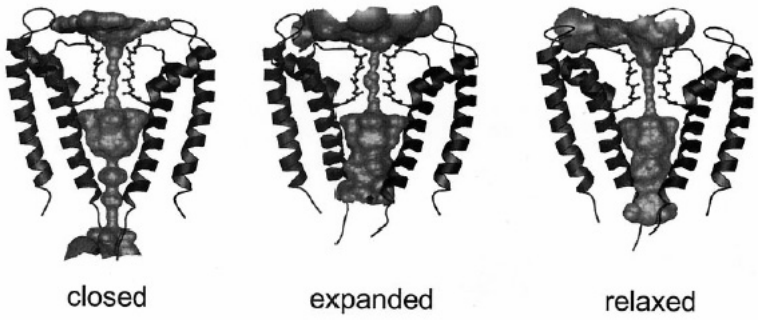


FIGURE 7 The inner surface of the pore for the three model states: closed (the final structure after 1 ns of simulation based upon the crystal structure), expanded (the final structure after simulation Ex3), and relaxed (Rx1 after 11 ns of simulation). Inner surfaces were generated with HOLE (Smart et al., 1997) and rendered within Molscript (Kraulis, 1991) and POV-Ray.

Biophysical Journal 83(4) 1867–1876

Brownian Dynamics Simulations of Interaction Between Scorpion Toxin Lq2 and Potassium Ion Channel

Meng Cui,* Jianhua Shen,* James M. Briggs,† Xiaomin Luo,* Xiaojian Tan,* Hualiang Jiang,* Kaixian Chen,* and Ruyun Ji*
*Center for Drug Discovery and Design, State Key Laboratory of New Drug Reseearch, Shanghai Institute of Meteria Medica, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, Peoples Republic of China; and †Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5513 USA

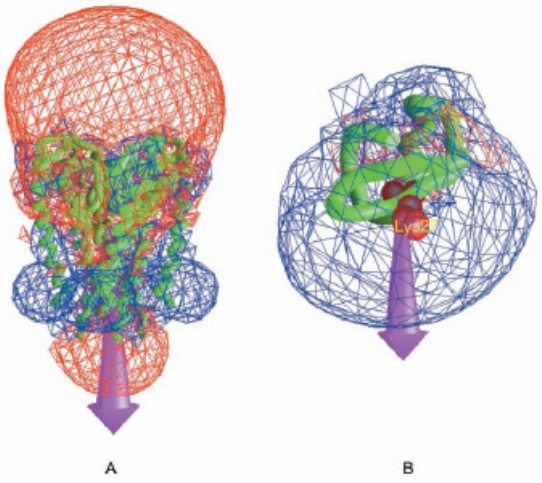


FIGURE 1 The electrostatic potential contour maps for the K^+ channel and the scorpion toxin Lq2. (*A*) Electrostatic potential for the K^+ channel. The red contours represent isopotential surfaces where charge 1e possesses electrostatic free energy equal to −1.0 kT/e; the blue isopotential surfaces are for energy +1.0 *kT/e*. (*B*) Electrostatic potential for the scorpion toxin Lq2. The red contours represent isopotential surfaces where charge 1e possesses electrostatic free energy equal to −2.0 kT/e; the blue isopotential surfaces are for energy +2.0 kT/e. Arrows indicate the directions of the dipoles in the proteins. The figure was generated with the program GRASP (Nicholls et al., 1991).

3524 Biophysical Journal Volume 92 May 2007 3524–3540

Probing the Outer Mouth Structure of the hERG Channel with Peptide Toxin Footprinting and Molecular Modeling

Gea-Ny Tseng,* Kailas D. Sonawane,† Yuliya V. Korolkova,‡ Mei Zhang,* Jie Liu,* Eugene V. Grishin,‡ and H. Robert Guy†

*Department of Physiology, Virginia Commonwealth University, Richmond, Virginia; †Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland; and ‡Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

ABSTRACT Previous studies have shown that the unusually long S5-P linker lining *human ether a-go-go related gene*'s (hERG's) outer vestibule is critical for its channel function: point mutations at high-impact positions here can interfere with the inactivation process and, in many cases, also reduce the pore's K^+ selectivity. Because no data are available on the equivalent region in the available K channel crystal structures to allow for homology modeling, we used alternative approaches to model its three-dimensional structure. The first part of this article describes mutant cycle analysis used to identify residues on hERG's outer vestibule that interact with specific residues on the interaction surface of BeKm-1, a peptide toxin with known NMR structure and a high binding affinity to hERG. The second part describes molecular modeling of hERG's pore domain. The transmembrane region was modeled after the crystal structure of KvAP pore domain. The S5-P linker was docked to the transmembrane region based on data from previous NMR and mutagenesis experiments, as well as a set of modeling criteria. The models were further restrained by contact points between hERG's outer vestibule and the bound BeKm-1 toxin molecule deduced from the mutant cycle analysis. Based on these analyses, we propose a working model for the open conformation of the outer vestibule of the hERG channel, in which the S5-P linkers interact with the pore loops to influence ion flux through the pore.

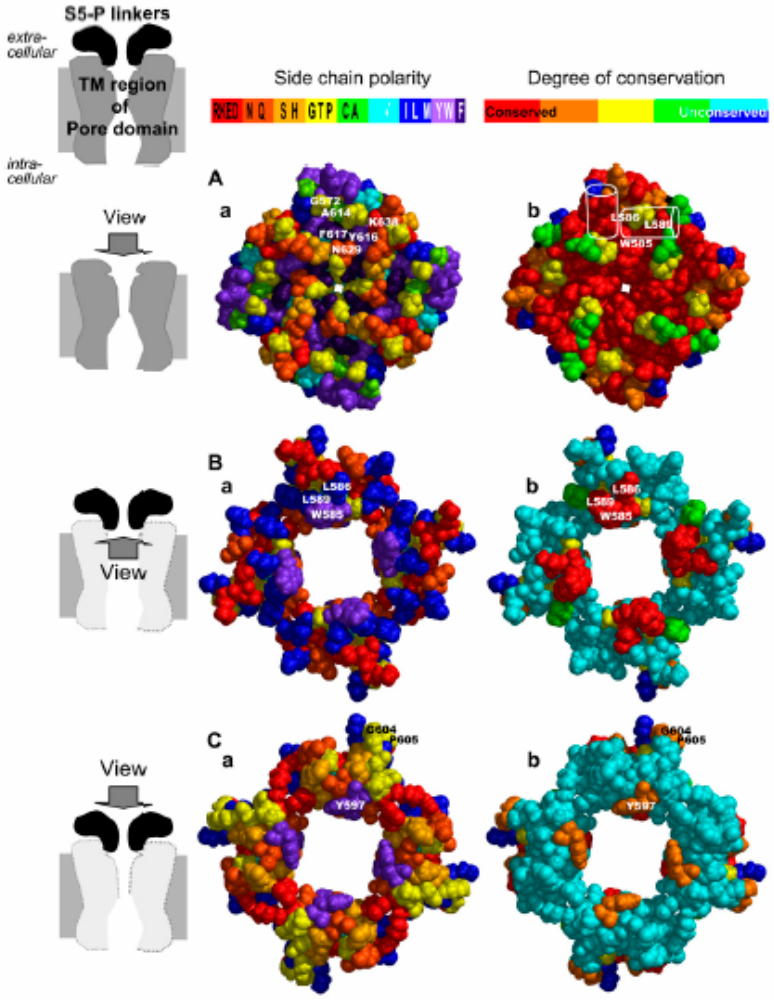


FIGURE 7 Putative interactions between the S5-P linkers and the outer surface of the transmembrane (*TM*) region of hERG's pore domain in our model. (*Right column*) Diagrams of channel structures (TM region of pore domain and extracellular S5-P linkers) and directions of view in rows *A*, *B*, and *C*. (*Central column*) Space-filled models of channel structures viewed as indicated on the left with amino acid side chains color coded according to polarity (color scheme shown on top with one-letter amino acid codes), red = charged, orange = hydrophilic, yellow/light green = amphivalent, cyan/blue = hydrophobic, and purple = aromatic. (*Right column*) Same views as in central column with amino acid side chains color coded according to the degree of conservation within the EAG family. The color scheme is as illustrated in Fig. 4. Panels *Aa* and *Ab* illustrate outer surface view of the TM region of hERG pore domain. Labeled in the top subunit of *Aa* are residues with which highly conserved S5-P linker residues (W585, L586, and L589) may interact. Labeled in the top subunit of *Ab* are the approximate backbone locations of S5-P1 and S5-P2 helices (*white outline* of *cylinders*) and approximate locations of W585, L586, and L589 side chains. Panels *Ba* and *Bb* illustrate inner surface view of the S5-P linkers. Note that most of the hydrophobic residues of the S5-P linkers reside on this surface (*Ba*), and that some residues form a highly conserved cluster (*red* area in *Bb*). W585, L586, and L589 residues interact with highly conserved residues of the S5/P/ S6 segments shown in *Aa* and *Ab* (see details in text). *Ca* and *Cb* illustrate outer surface view of S5-P linkers. Most of these residues are hydrophilic (*Ca*) and poorly conserved (*Cb*), consistent with the notion that they are exposed to the extracellular aqueous phase.

El de la serpiente

doi:10.1006/jmbi.2000.3522 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 296, 1283–1294

JMB

Energetic and Structural Interactions between δ-Dendrotoxin and a Voltage-gated Potassium Channel

John P. Imredy and Roderick MacKinnon*

Howard Hughes Medical Institute, Laboratory of Molecular Neurobiology and Biophysics, Rockefeller University, New York NY 10021, USA

**Corresponding author*

Dendrotoxin proteins isolated from Mamba snake venom block potassium channels with a high degree of specificity and selectivity. Using site-directed mutagenesis we have identified residues that constitute the functional interaction surfaces of δ-dendrotoxin and its voltage-gated potassium channel receptor. δ-Dendrotoxin uses a triangular patch formed by seven side-chains (Lys3, Tyr4, Lys6, Leu7, Pro8, Arg10, Lys26) to block K^+ currents carried by a Shaker potassium channel variant. The inhibitory surface of the toxin interacts with channel residues at Shaker positions 423, 425, 427, 431, and 449 near the pore. Amino acid mutations that interact across the toxin-channel interface were identified by mutant cycle analysis. These results constrain the possible orientation of dendrotoxin with respect to the K^+ channel structure. We propose that dendrotoxin binds near the pore entryway but does not act as a physical plug.

Keywords: Dendroaspis angusticeps; snake dendrotoxin; *Shaker;* Kv1.1; KcsA

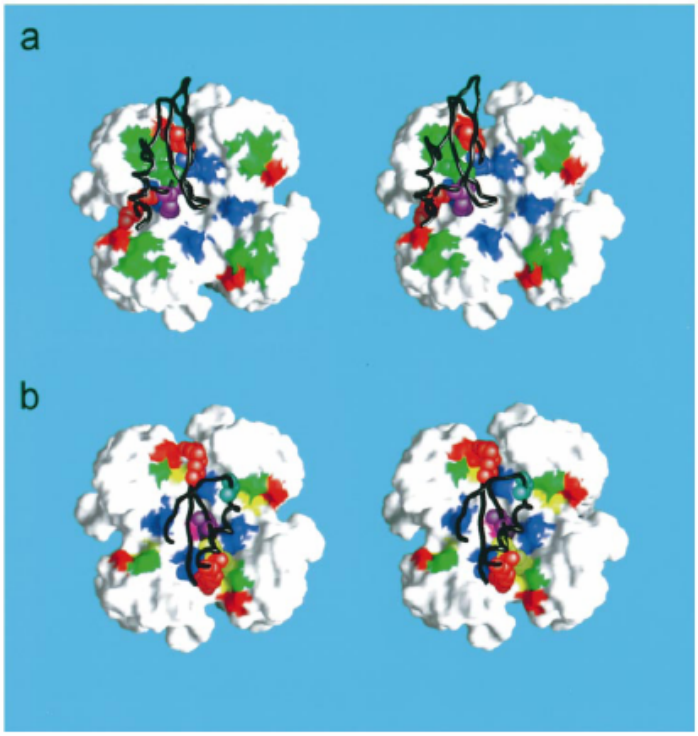


Figure 5. Hypothetical binding orientation for dendrotoxin. (a) Stereo diagram of δ-dendrotoxin relative to the Shaker K^+ channel. The channel is shown in the same orientation as in Figure 3(c), looking directly down the pore from the extracellular side of the channel. The functional interaction residues of the channel have been mapped to the surface with the following color scheme: 423/425/427 (green), 431(red), and 449 (blue). The toxin side-chains shown are from residues that strongly interact (through mutant cycles) with their color matched interaction partner(s) on the channel (Figure 4(a); Lys3 and Arg10, red; Leu7, green; Pro8, blue; Tyr4, omitted for clarity). Lys6 of δ-dendrotoxin, which interacts to roughly the same extent with both 431 (red) and 449 (blue), is purple. (b) Binding orientation of scorpion toxin on the K^+ channel. This orientation is based on mutant cycle analysis of Agitoxin2 interaction with the Shaker channel (Ranganathan *et al.*, 1996; MacKinnon *et al.*, 1998). The channel residues are colored as in (a), except that positions 423 and 427 have been omitted. Additional Shaker residues that interact with Agitoxin2 are M448 (yellow) and Y445 (purple). The blue-green residue on Agitoxin2 interacts with both position 449 (blue) and 425 (green). The purple residue on Agitoxin2 is Lys27 which interacts with the external K^+ binding site of the channel and thus acts as a "plug" in the pore (Park & Miller, 1992a,b; Ranganathan *et al.*, 1996). The orientation of the toxins relative to the channel was arrived at by visually minimizing the distances between the spacefilled models of the interacting toxin residue atoms (omitting hydrogen atoms) and the molecular surface of the coupled channel residues within the