

Motivos Lineales en Proteínas

Lucía Chemes y Juliana Glavina

Recursos a utilizar:

- Regex101 <https://regex101.com>
- UniProt <http://www.uniprot.org/>
- ELM <http://elm.eu.org>
- SlimSearch <http://slim.ucd.ie/slimsearch/>
- ProViz <http://proviz.ucd.ie>

Ejercicio 1. Familiarizándonos con las Expresiones Regulares

Objetivos:

- Familiarizarse con la simbología utilizada en expresiones regulares
- Utilizar la simbología para poder realizar búsquedas basadas en texto

La simbología comúnmente utilizada en expresiones regulares es:

Símbolo	Definición
.	Cualquier aminoácido es permitido
[XY]	Solo los aminoácidos X e Y son permitidos
[^XY]	Los aminoácidos X e Y están prohibidos
{min,max}	Número mínimo y máximo de veces que se puede repetir una posición
^X	El aminoácido X se encuentra en el extremo N-terminal
X\$	El aminoácido X se encuentra en el extremo C-terminal
(AB) (CD)	Se encuentran, o bien, los aminoácidos AB, o bien, los aminoácidos CD

Estos símbolos nos permiten definir patrones que son observados en proteínas naturales para luego identificarlos en otras proteínas y ser puestos a prueba experimentalmente.

Los receptores nucleares interactúan con diversas proteínas mediante un motivo lineal llamado NRBox (*Nuclear Receptor Box*) (Heery, 1997). Existen numerosas estructuras de péptidos unidos a diferentes receptores nucleares (PDBs: 3CS8, 2GPO, 1GWQ, 1RJK, 1M2Z) que permitieron estudiar y entender algunas características de la interacción.

La evidencia experimental recolectada de la literatura indica que:

- I. El motivo NRBox forma una hélice alfa
- II. Existen tres leucinas que se **encuentran en una misma cara de la hélice** que interactúan con un bolsillo hidrofóbico en la superficie del receptor nuclear.

Los siguientes fragmentos de secuencia corresponden a regiones de distintas proteínas que interactúan con diversos receptores nucleares y cuya interacción se verificó de manera experimental por distintos métodos.

```
>sp|Q15648|MED1_HUMAN|644-650
SMAGNTKNHPMLMNLLKDNPAQDFSTL
>sp|O43593|HAIR_HUMAN|565-571
AKHLLSGLGDRLCRLRREREALAWAQ
>sp|Q16881-4|TRXR1_HUMAN|46-52
GPTLKAYQEGRLQKLLKMNGPEDLPKS
>sp|P48552|NRIP1_HUMAN|500-506
DVHQDSIVLTYLEGLLMHQAAGGSGTA
>sp|Q9UQ80|PA2G4_HUMAN|353-359
YKSEMEVQDAELKALLQSSASRKTQKK
>sp|Q90ZL7|Q90ZL7_DANRE|69-75
VQHADGEKSNVLRKLLKRANSYEDAVM
>sp|Q9UBK2|PRGC1_HUMAN|143-149
PPPQEAEEPSLLKKLLLAPANTQLSYN
>sp|Q9JL19|NCOA6_MOUSE|1494-1500
MSPAMREAPTSLSQLLDNSGAPNVTIK
>sp|Q15596|NCOA2_HUMAN|689-695
HGTSLEKEKHKILHRLQLDSSSPVDLAK
>sp|Q92793|CBP_HUMAN|69-75
LVPDAASKHKQLSELLRGGSGSSINPG
```

- a. Copia y pega las secuencias en el recuadro de *Test String* en regex101 y prueba encontrar una expresión regular que permita identificar el motivo que media la

interacción de estas proteínas con los receptores nucleares y que cumpla con la evidencia experimental observada.

- b. Considerando que el motivo se encuentra en una hélice, ¿modificaría la expresión regular que obtuvo?

Ejercicio 2. Base de datos de motivos lineales en Eucariotas (ELMdb)

La base de datos ELM (*Eukaryotic Linear Motifs*) es una base de datos que se enfoca principalmente en la anotación y detección de motivos lineales (MLs). Para ello cuenta con un repositorio de motivos manualmente anotados, por lo cual está altamente curada y una herramienta de predicción de motivos. Esta predicción de motivos se realiza mediante una búsqueda de patrones de secuencia basada en texto utilizando expresiones regulares.

Objetivos:

- Familiarizarse con la herramienta de predicción de motivos de ELM.
- Aplicar la herramienta de predicción de motivos de ELM a una proteína viral.

La familia viral *Adenoviridae* (adenovirus) son virus ADNdc desnudos. Los adenovirus que infectan a humanos son responsables de muchas enfermedades respiratorias y de numerosos casos de gastroenteritis en niños. El único género de adenovirus que posee la proteína E1A es el género *Mastadenovirus* que infecta a **mamíferos**. Hasta la fecha, no existe ningún homólogo reportado en los restantes géneros de esta familia viral. La proteína E1A posee un rol importante en la replicación del genoma viral ya que desregula el ciclo celular induciendo la división celular. Esta estimulación de la progresión de la fase G1 a la fase S, permite que el virus use la maquinaria celular de replicación del ADN para replicar su propio genoma. Una vez expresada la proteína E1A su localización en la célula infectada es **nuclear** y minoritariamente **citoplasmática**.

1. Busca en ELM (<http://elm.eu.org>) en la pestaña *Prediction* la proteína E1A del virus Human adenovirus 5 (E1A_ADE05).

A. Utiliza los siguientes parámetros:

Cell Compartment: **Not specified**

Motif Probability Cutoff: **100**

Taxonomic context: **(leave blank)**

¿Cuántas clases y cuántas instancias de motivos encuentras?

- B. En base a los conocimientos que poseemos de E1A_ADE05 modifica los parámetros **Cell Compartment** (se puede seleccionar más de un compartimento celular utilizando la tecla **ctrl**) y **taxonomic context**. ¿Cómo cambia el número de motivos encontrados?
- C. ¿Qué otros filtros observas que está utilizando ELM? ¿Por qué se te ocurre que se eligen automáticamente esos filtros?
- D. ¿Qué se puede decir de la estructura de la proteína E1A? ¿Se observa algún dominio? ¿Se observan regiones desordenadas?
- E. En cada una de las clases de motivos encontrados, se indica con distintos símbolos (descritos en la parte superior de la página) si la instancia del motivo es predicha o fue identificada experimentalmente (instancias anotadas o “True Positives”). ¿Cuántas instancias anotadas existen?
- F. E1A tiene dos motivos de interacción con la proteína Retinoblastoma, un regulador del ciclo celular (motivo AB_groove y motivo LxCxE). Explora la clase LIG_Rb_LxCxE_1, para esto haz click sobre el nombre de la misma, en la lista de la izquierda. Se abrirá la página correspondiente a esa clase donde se listan todas las instancias reportadas en la literatura que están anotadas en ELM.
- I- ¿En qué tipos de proteínas se encuentra el motivo LxCxE?
- II- Existen dos tipos de instancias.
- True Positives (TP): Son instancias identificadas por la expresión regular y que la evidencia experimental muestra que es funcional.
 - False Positives (FP): Son instancias identificadas por la expresión regular, que la evidencia experimental sugieren que son funcionales, pero cuando fue evaluada se cree que no es realmente funcional.
- ¿Se te ocurre algún ejemplo donde esto pueda ocurrir?

2. Busca en ELM E1A_ADECR.

- A. ¿Cuál es el contexto taxonómico?
- B. ¿Cuántas instancias anotadas hay? ¿Se encuentran los motivos anotados de E1A_ADE05? ¿A qué puede deberse?

3. La proteína Retinoblastoma (Rb) controla la transición en el ciclo celular de la fase G1 a la fase S mediante la interacción con factores de transcripción de la familia E2F.

- a. Vaya a ProViz (<http://proviz.ucd.ie/>) y busque la proteína E2F1_HUMAN (Q01094).
- b. ¿Puede identificar el motivo de interacción con Rb? (Pista: Hay una línea a la izquierda que se llama ELM).
- c. El motivo ¿Está en un contexto estructural desordenado? ¿Se encuentra conservado? ¿Es el mismo motivo usado por la proteína E1A para interactuar

con Rb?

- d. ¿Qué otros motivos identifica? ¿Algunos de estos motivos están involucrados en el ciclo celular?

EJERCICIOS ADICIONALES

Ejercicio adicional 1. Un poquito más de expresiones regulares

La reparación del ADN durante la replicación ocurre por un proceso llamado Translesion synthesis (TLS). En este proceso, una polimerasa TLS, inserta un nucleótido en la lesión del ADN y luego, una polimerasa de la familia B extiende el templado. La acción coordinada de estas polimerasas, se logra por la interacción de proteínas scaffold como PCNA (*Proliferating Cell Nuclear Antigen*) y la polimerasa TLS Rev1.

Existen estructuras cristalográficas de distintos péptidos unidos a Rev1 (PDBs: 2N1G, 2LSK, 2LSJ, 4FJO, 2LSI y 4GK5) que permiten entender algunas características de la interacción.

La evidencia experimental recolectada de la literatura indica:

- I. La interacción está mediada principalmente por dos residuos consecutivos de fenilalanina (Ohashi, 2009).
- II. Las fenilalaninas interactúan con un bolsillo hidrofóbico en la superficie de Rev1 (Pozhidaeva, 2012; Zhao, 2017).
- III. Las fenilalaninas se encuentran en el primer giro de una hélice α .
- IV. Se requieren al menos 4 residuos posteriores a las fenilalaninas que formen parte de una hélice (Ohashi, 2009)
- V. El resto de la región de interacción se pliega formando hélices α de longitud variable (Pustovalova, 2016)
- VI. En general se observan residuos cargados positivamente en la 2da y/o 3ra posición luego de las fenilalaninas que median interacciones electrostáticas con una superficie acídica de Rev1. Aunque la posición de estos residuos puede variar.

Los siguientes fragmentos de secuencia corresponden a regiones de distintas proteínas que participan en la reparación del ADN que se unen la proteína Rev1 y cuya interacción se verificó de manera experimental por distintos métodos.

```

>sp|Q03834|MSH6_YEAST|31-38
SQKKMKQSSLLSFFSKQVPSGTPSKKVQ
>sp|Q04049|POLH_YEAST|625-632
KKQVTSSKNILSFFTRKK
>sp|Q60596|XRCC1_MOUSE|191-200
DDSANSLKPGALFFSRINKTSSASTSDPAG
>sp|Q9H040|SPRTN_HUMAN|418-428
RPRLEDKTVFDNFFIKKEQIKSSGNDPKYST
>sp|Q15054|DPOD3_HUMAN|236-245
NKAPGKGNMMSNFFGKAAMNKFKVNLDSEQ
>sp|Q9UNA4|POLI_HUMAN|569-579
SCPLHASRGVLSFFSKKQMQDIPINPRDHLS
>sp|Q9Y253|POLH_HUMAN|481-490
TATKKATTSLLESFFQKAAERQK VK EA SL SS
>sp|Q9Y253|POLH_HUMAN|529-539
PFQTSQSTGTEPFFKQKSLLLKQKQLNNSV
>sp|Q9QUG2|POLK_MOUSE|564-575
LAKPLEMSHKKSFFDKKRSEKISNCQDTSRCK
>sp|Q9UBT6|POLK_HUMAN|565-576
FVKPLEMSHKKSFFDKKRSEKWSHQDTFKCE

```

- a. Copia y pega las secuencias en el recuadro de *Test String* en regex101 y prueba encontrar una expresión regular que permita identificar el motivo que media la interacción de estas proteínas con Rev1 y que cumpla con la evidencia experimental observada.
- b. Busca en ELM alguna de las proteínas. ¿Tu expresión regular difiere mucho de la propuesta por ELM?
- c. SlimSearch es una herramienta que utilizando expresiones regulares permite buscar la presencia de motivos en las proteína almacenadas en Uniprot. Ve a la web de SlimSearch (<http://slim.ucd.ie/slimsearch/>) e ingresa la expresión regular del motivo como figura en ELM.
 - i. ¿Cuántas proteínas obtuviste?
 - ii. ¿Cuál es la localización celular de Rev1 (Q9UBZ9)? Explora la lista de proteínas. ¿Hay alguna que no tenga la misma localización?
 - iii. Encuentra en la lista la proteína Kinesin-like protein KIF11 (P52732). ¿Cuál es su localización?
- d. ProViz es una herramienta que colecta y muestra información desde distintas fuentes facilitando la detección de motivos lineales. Ingresa en el servidor de Proviz y busca la proteína Kinesin-like protein KIF11

(KIF11) (P52732). Ubica la región donde se encuentra el posible motivo sugerido por SlimSearch. ¿Está conservado? ¿A qué se debe esa conservación? ¿Te parece que es un posible motivo?

Ejercicio Adicional 2. Familiarizándose con la base de datos ELM.

1. Pega y copia la siguiente secuencia en ELM y utiliza los parámetros que se indican a continuación.

> **P12931**

```
MGSNKS KPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPASADGHRGPSAAFAPAAAE
PKLFGGFNSSDTVTSPQRAGPLAGGVTTFVALYDYESRTETDLSFKKGERLQIVNNTGEGD
WWLAHSLSTGQTGYIPSNYVAPSDSIQAEWYFGKITRRESERLLLNAENPRGTFLVRES
ETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFYITSRTQFNSLQQLVAYYSKHADGL
CHRLTTVCPTSKPQTQGLAKDAWEIPRESLRLEVKLGQGC FGEVWMGTWNGTTRVAIKTL
KPGTMSPEAFLQEAQVMKKLRHEKLVQLYAVVSEEP IYIVTEYMSKGSLLDFLKGETGKY
LRLPQLVDMAAQIASGMAYVERMNYVHRDLRAANILVGENLVCKVADFG LARLIEDNEYT
ARQGAKFPIKWTAPEAALYGRFTIKSDVWSFGILLTELTKGRVPYPGMVNREVL DQVER
GYRMPCPPECPESLHDLMCQCWRKEPEERPTFEY LQAFLEDYFTSTEPQYQPGENL
```

Cell Compartment: **Not specified**

Motif Probability Cutoff: **100**

Context information: **(leave blank)**

- ¿Cuántas instancias del motivo se encuentran?
- ¿Que se puede decir sobre la estructura de la proteína? ¿Se observa algún dominio? ¿Se observan regiones desordenadas?

2. Realiza la misma búsqueda (P12931) utilizando los siguientes parámetros:

- Cell Compartment: **cytosol**
- Motif Probability Cutoff: **0.01**
- Context information: **Homo sapiens**

- ¿Cuántas instancias aproximadamente se encuentran ahora?
- ¿Cuántas instancias están anotadas?
- ¿Los predictores estructurales y filtros (SMART, GlobPlot, IUPRED, Secondary Structure) coinciden sobre qué regiones son estructuradas/desordenadas?
- Compara la ubicación de las instancias anotadas con la información estructural proveniente de IUPRED

3. Realiza la búsqueda de la secuencia de la proteína Paxillina (P49023) en ELM, utilizando los parámetros por defecto. Compara los resultados con una búsqueda de la misma secuencia pero modificando el parámetro **cellular compartment** 'plasma membrane'.
4. Busca la proteína SRC_MOUSE (P05480) en ELM.
 - a. ¿Existen instancias anotadas?
 - b. Si no, cuál es la instancia anotada más cercana que se puede encontrar. Investiga de dónde proviene esta información.
5. Busca la proteína CDN1A_HUMAN (P38936) en ELM.
 - a. ¿Cuántas instancias de la clase DOC_PP1_RVXF_1 se encontraron?
 - b. ¿Cuál es la diferencia entre estas instancias? ¿Que tiene de especial la instancia en la posición 155? ¿Porqué?
6. Busca la proteína 'P53_HUMAN'
 - a. ¿Qué compartimentos celulares se le asignaron? ¿Tienen sentido?
 - b. ¿Cuántos degrons hay en p53?
 - c. ¿Existe algún sitio CDK en p53? ¿Hay un Cyclin Box en p53?
7. Busca en ELM la proteína 'MDM4_HUMAN' y encuentra el motivo de unión a USP (DOC_USP7_MATH_1). ¿Cuántas instancias del motivo se encuentran en esta secuencia?
8. Busca en ELM la proteína 'AMPH_HUMAN' y encuentra la clase 'LIG_Clathr_ClatBox_1'
 - a. ¿Cuál es la relevancia biológica de cada una de estas instancias?
 - b. ¿La anotación de la relevancia biológica coincide con la estructura globular?
9. Busca todas las instancias anotadas para "*Homo sapiens*" que contienen el término "cilium" (pista: Usa url http://elm.eu.org/elms/browse_instances.html).
 - a. ¿Cuántas instancias hay?
 - b. ¿Qué evidencia experimental está anotada y cuán confiable es esta evidencia?
 - c. Descarga estas instancias en un archivo ".tsv" (tab-separated values)
10. Busca todas las instancias anotadas que contienen el término "retinoblastoma" (Pista: usa url http://elm.eu.org/elms/browse_instances.html)
 - a. Compara el número de instancias humanas con el número de instancias virales.
 - b. ¿Porqué hay tantas proteínas virales que interactúan con retinoblastoma? (Pista: La respuesta está en el abstract de la clase del motivo [LIG_Rb_LxCxE_1](#))

Ejercicio Adicional 3. Proteína CagA de *Helicobacter pylori*.

La infección por *H. pylori* puede causar gastritis, úlcera péptica o cáncer de estómago. Hay una mayor probabilidad de desarrollar cáncer estomacal si la infección es producida por una cepa del Este asiático (como F32) en comparación a una cepa del Oeste (como NCTC 11637). Estas cepas difieren en el número y contexto de secuencia de los motivos EPIYA (Higashi, H., et al., 2002; Jones, K.R., et al., 2009).

- a. Copia y pega en ELM las secuencias la proteína CagA de una cepa del Oeste y una cepa del Este asiático, especificando Cytosol como compartimento celular, *Homo sapiens* como especie y un umbral de corte de la probabilidad del motivo de 0.001.

> NCTC11637_CagA

```
MTNETIDQQPQTEAAFPNQQFINNLQVAF LKVDNAVASYDPDQKPIVDKNDNRQAFDGISQL
REEYSNKAIAKNPTKKNQYFSD FINKSNDLINKDNLIDIGSSIKSFQKFGTQRYRIFTSWVSHQN
DPSKINTRSIRNFMENIIQPPIPD DKEKAEFLKSAKQSFAGIIIGNQIRTDQKFMGVFDEFLKE
RQEAEKNGEPTGGDWLDIFLSFV FNKEQSSDVKEAINQEPVPHVQPDIAATTTTHIQGLPPESRD
LLDERGNFSKFTLGDMEMLDVEGV ADIDPNYKFNQLLIHNNALSSVLMGSHNGIEPEKVSLLYA
GNGGFGAKHDWNATVGYKNQQGD NVATLINVHMKNNGSGLVIAGGEKGINNPSFCLYKEDQLTGS
QRALSQEEIRNKIDFMEFLAQNN AKLDNLSEKEKEKEKFQNEIEDFQKDSKAYLDALGNDRIA FVS
KKDPKHSALITEFGKGDLSYTLK DYGKKADRALDREKNVTLQGNLKHDSVMFVNYSNFKYTNAS
KSPDKGVGV TNGVSHLDAGFSKVAVFNLPDLNNLAITSFVRRNLENKLVTEGLSLQEANKLIKD
FLSSNKELVGKALNFNKA VADAKNTGNYDEVKKAQKDLEKSLRKREHLEKEVEKKLESKSGNKN
KMEAKAQANSQKDKIFALINKEAN RDARAIAYSQNLKGIKRELSDKLEKINKDLKDFS KSFDEF
KNGKNKDFSKAEETLKALKGSVKDLGINPEWISKVENLNAALNEFKNGKNKDFSKVTQAKSDLE
NSVKDVIVNQKITDKVDNLNQAVS MAKATGDFSRVEQALADLKNFSKEQLAQQTQKNESFNVGK
KSEIYQSVKNGVNGTLVGNGLSG IEATALAKNFSDIKKELNEKFKNFNNNNNNNGLENEPIYAKV
NKKKTGQVASPEEPIYAQVAKKVNA KIDRLNQAASGLGGVGQAGFPLKRHDKVDDLSKVGRSVS
PEPIYATIDDLGGPFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDLSKVG
RSVSPEPIYATIDDLGGPFPLKRHDKVDDLSKVGLSRNQELAQKIDNLSQAVSEAKAGFFSNLE
QTIDKLKDSTKYNSVNLWVESAKKVPASLSAKLDNYATNSHTRINSNIQNGAINEKATGMLTQK
NPEWLKLVNDKIVAHNVGSVPLSEYDKIGFNQKNMKDYSDSFKFSTKLNNAVKDVKSSFTQFLA
NAFSTGYYS LARENAEHGIKNVNTKGGFQKS
```

> F32_CagA

```
MTNETIDQTTTPDQTGFVPQRFINNLQVAFIKVDNAVASFDPDQKPIVDKNDKDNRQAYEKISQ
```

LREEYANKAIKNPAKKNQYFSDFINXSNDLINKDNLIAVDSSVESFRKFGDQRYQIFTSWVSLQ
 KDPSKINTQQIRNFMENVIKPPISSDDKEKAEFLRSAKQSFAGIIIGNQIRSDEKFMGVFDESLE
 ARQEAENAEPAGGDWLDIFLSFVFNNKQSSDLKETLNQEPRPDFEQNLATTTTDDIQGLPPEAR
 DLLDERGNFFKFTLGDVEMLDVEGVADKDPNYKFNQLLIHNNALSSMLMGSHSNIEPEKVSLLY
 GDNGGPEARHDWNATVGYKNQQGNNVATLINAHNNNGSGLIIAGNEDGIKNPSFYLYKEDQLTG
 LKQALSQEEIQNKVDFMEFLAQNNAKLDNLSEKEKEKFQTEIENFQKDRKAYLDALGNDHIAFV
 SKKDPKHLALVTEFGNGELSYTLKDYGKKQDKALDGETKTTLQGSLKYDGVFMFVNYSNFKYTNA
 SKSPNKGLGTTNGVSHLEANFSKVAVFNLPLNNLAITNYIRRDLEDKLWAKGLSPQEANKLIK
 DFLNSNKEMVGKVSNNFNKAVAEAKNTGNYDEVKKAQKDLEKSLRKREHLEKEVAKKLESNDNK
 NRMEAKAQANSQKDKIFALISQEASKEARVATFDPYLKGVRSELSKLENINKNLKDFGKSFDE
 LKSGKNNDFSKAEETLKALKDSVKDLGINPEWISKIENLNALNDFKNGKNKDFSKVTQAKSDL
 ENSIKDVIINQKITDKVDNLNQAVSEIKLTGDFSKVEQALAEKKNLSLDLGKNSDLQKSVKNGV
 NGTLVSNGLSKTEATTLTKNFSDIRKELNEKLFGNSSNNNNGLKNNTEPIYAQVNKKKTGQATS
 PEEPIYAQVAKKVSADIDQLNEATSAINRKIDRINKIASAGKGVGGFSGAGRSASPEPIYATID
 FDEANQAGFPLRRSAAVNDLSKVGLSREQELTRRIGDLSQAVSEAKTGHFGNLEQKIDELKDST
 KKNALKLWVESAKQVPTSLQAKLDNYATNSHTRINSNVQSGTINEKATGMLTQKNPEWLKLVND
 KIVAHNVGSAPLSAYDKIGFNQKNMKDYSDFSFKSTKLNNAVKDIKSSFVQFLTNTFSTGSYSL
 MKANVEHGVKNTNTKGGFQKS

- b. ¿Cuáles son las diferencias en las predicciones del motivo EPIYA? ¿Existen diferencias en la asignación por homología?

Bibliografía Sugerida:

- Alexander et al. Sci. Sig 2011 "Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling" [\[URL\]](#)
- Davey NE, Travé G and Gibson TJ (2011), "*How viruses hijack cell regulation*", Trends Biochem Sci., Mar, 2011. Vol. 36, pp. 159-169. [\[DOI\]](#) [\[URL\]](#)
- Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H and Gibson TJ (2012), "*Attributes of short linear motifs*", Mol Biosyst., Jan, 2012. Vol. 8, pp. 268-281. [\[DOI\]](#) [\[URL\]](#)
- Dinkel H, Van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, Speck T, Krüger D, Grebnev G, Kuban M, Strumillo M, Uyar B, Budd A, Altenberg B, Seiler M, Chemes LB, Glavina J, Sánchez IE, Diella F, Gibson TJ. (2015), "*The eukaryotic linear motif resource ELM: 10 years and counting.*" Nucleic Acids Res., Nov, 2013. [\[DOI\]](#) [\[URL\]](#)
- Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ and Diella F (2011), "*Phospho.ELM: a database of phosphorylation sites--update 2011.*", Nucleic Acids Res., Jan, 2011. Vol. 39 (Database issue), pp. D261-D267. [\[DOI\]](#) [\[URL\]](#)

- Dyson HJ and Wright PE (2005), *"Intrinsically unstructured proteins and their functions"*, Nat Rev Mol Cell Biol., Mar, 2005. Vol. 6, pp. 197-208. [\[DOI\]](#) [\[URL\]](#)
- Van Roey K, Orchard S, Kerrien S, Dumousseau M, Ricard-Blum S, Hermjakob H and Gibson TJ (2013), *"Capturing cooperative interactions with the PSI-MI format"*, Database (Oxford). Vol. 2013, pp. bat066. [\[DOI\]](#) [\[URL\]](#)
- Van Roey K, Dinkel H, Weatheritt RJ, Gibson TJ, Davey NE. (2013) *"The switches.ELM resource: a compendium of conditional regulatory interaction interfaces."* Sci Signal. 2013 Apr 2;6(269):rs7. [\[DOI\]](#) [\[URL\]](#)
- Gibson TJ, Dinkel H, Van Roey K, Diella F (2015) "Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad", Cell Communication & Signalling [\[URL\]](#)