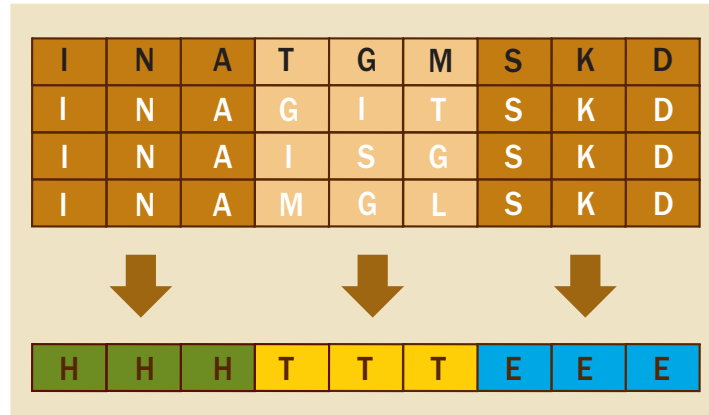


Predicción de la estructura Secundaria a partir de la secuencia

8
Capítulo




La existencia de una asociación entre la secuencia de aminoácidos y las estructuras secundarias adoptadas en la proteína es la hipótesis fundamental sobre la cual están basados todos los métodos de predicción de estructura secundaria.

Los métodos ampliamente utilizados en la predicción de estructura secundaria incluyen los de Chou-Fasman y GOR (Métodos estadísticos), modelos de redes neuronales y los basados en vecinos cercanos (nearest-neighbor). A continuación serán descritos cada uno de ellos.

MÉTODOS CHOU-FASMAN Y GOR

En 1974, Chou y Fasman desarrollaron métodos estadísticos de predicción de estructura secundaria basados en la probabilidad de los aminoácidos en adoptar orientaciones espaciales específicas a partir de la observación de la localización de cada aminoácido en 15 estructuras de proteínas determinadas por difracción de





Estos métodos utilizan las propiedades estereoquímicas y fisicoquímicas de los diferentes aminoácidos. A través de los años, dichos métodos han sido perfeccionados al involucrar un mayor número de proteínas.

A diferencia de los métodos de Chou-Fasman, los cuales asumen que cada aminoácido por individual influye en la estructura secundaria, los métodos GOR utilizan la estadística para relacionar la propensión del residuo central, de segmentos típicamente de 11-21 aminoácidos, con la adopción de una estructura en particular. Lo anterior aumentó la fiabilidad de los métodos de predicción puesto que la adopción de estructuras secundarias también está gobernada por interacciones de largo alcance (Garnier, 1996).

MÉTODOS QUE EMPLEAN REDES NEURONALES

Los métodos de predicción de estructura secundaria han mejorando sustancialmente en los últimos lustros por el empleo de información evolutiva, obtenida a partir de la divergencia de familias de proteínas resueltas estructuralmente. De hecho, la precisión de la predicción ha aumentado por más del 4% en particular por la utilización de redes neuronales (Rost *et al.*, 1996).

Los programas computacionales son entrenados para reconocer patrones de estructuras secundarias conocidas empleando sistemas de entrenamiento de estructuras no homólogas, y son evaluadas con proteínas de estructura conocida. El servidor de predicción de estructura secundaria más representativo que utiliza este tipo de métodos es PHD (Rost *et al.*, 1993a; Rost *et al.*, 1996), el cual permite alcanzar un nivel de exactitud total del 73% para cada residuo en los tres tipos de estructura secundaria (Hélices, hojas y giros). Otros programas de predicción que emplean redes neuronales son NNpredict (<http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html>) (Kneller *et al.*, 1990) y Psipred (<http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>) (Jones *et al.*, 1999).

La razón de una mayor exactitud en la predicción de estructuras secundarias a partir de redes neuronales es atribuida a la capacidad de alinear la secuencia de la pro-



teína objeto de estudio con otras de la misma familia, encontrando integrantes de la misma con estructura conocida, lo cual facilita la asignación de la estructura secundaria.

Las redes neuronales tienen grandes dificultades en ocupar motivos de longitud variables puesto que la capa de entrada de la red es típicamente una estructura rígida que acepta secuencias de un tipo de longitud definida. Así, mientras que las redes neuronales pueden ser herramientas de gran alcance en la predicción de la función, generalmente no nos dicen información relevante sobre el proceso de re-

MÉTODOS DE LOS VECINOS CERCANOS

La idea básica de los métodos vecinos cercanos (nearest-neighbor) es la predicción de los estados de estructura secundaria tomando información de aminoácidos centrales que hacen parte de segmentos o ventanas de prueba a partir de datos estructurales de segmentos homólogos de proteínas con estructuras conocidas depositadas en bases de datos de proteínas.

La variabilidad que ofrecen los métodos vecinos cercanos es presentada en la selección de subsecuencias adyacentes a una ventana alrededor del aminoácido al cual se está asignando la estructura secundaria. Cada programa utiliza diversos parámetros, tales como definición del nivel de semejanza y el tamaño de la ventana de la secuencia que debe ser examinado.

Varios métodos emplean la estrategia descrita con anterioridad, tales como SOPM (Geourjon y Deleage, 1994), SOPMA (Geourjon y Deleage, 1995), NNSSP (Salamos *et al.*, 1995) y PREDATOR (Argos *et al.*, 1996), entre otros.

Cualquier método de predicción de estructuras de proteínas a partir de la secuencia tiene dos fuentes de errores: (1) errores sistemáticos, tales como efectos no locales directos, y (2) errores causados por instancia de sucesiones de ejemplos durante el entrenamiento de redes neuronales. Teóricamente, combinar cualquier número de métodos de predicción mejora la exactitud mientras que los errores de los métodos individuales son mutuamente independientes y no son solamente sistemáticos.



El desarrollo de aproximaciones consenso basados en la combinación de varios métodos de predicción de estructuras secundarias ofrece mejores resultados que los singulares (Selbig *et al.*, 1999; Albrecht *et al.*, 2003). A continuación serán descritos métodos que emplean resultados consenso para predecir estructuras secundarias de proteínas a partir de la secuencia.



NPS@ (Red de Análisis de Secuencias de Proteínas)

URL: http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html

ENTIDAD ADMINISTRADORA

Instituto de Biología y Química de Proteínas, Laboratorio de Conformación de Proteínas, Lyon, Francia.

DESCRIPCIÓN

NPS@ Combina métodos usando diferentes aproximaciones para predecir la estructura secundaria, siendo útil para validar la predicción. En el primer paso es posible seleccionar los métodos de predicción de estructura secundaria. Posteriormente, el programa de predicción genera un consenso de estructura secundaria para cada aminoácido presente en la proteína, además, muestra el porcentaje de estructura secundaria en la proteína obtenido con cada uno de los servidores de predicción (Combet *et al.*, 2000).

Los servidores presentes son SOPM (Geourjon y Deleage, 1994), SOPMA (Geourjon y Deleage, 1995), HNN (Guermeur, 1997), MLRC en GOR4, SIMPA96 y SOPMA (Guermeur *et al.*, 1998) DPM (Deleage y Roux, 1987), DSC (King y Stenberg, 1996), GOR I (Garnier *et al.*, 1978), GOR III (Gibrat *et al.*, 1987), GOR IV (Garnier *et al.*, 1996), PHD (Rost *et al.*, 1993), PREDATOR (Argos *et al.*, 1996) y SIMPA96 (Levin *et al.*,



JPRED (Método Consenso para la Predicción de Estructuras Secundarias)

URL: <http://www.compbio.dundee.ac.uk/~www-jpred/>

UNIVERSIDAD DE CARTAGENA



ENTIDAD ADMINISTRADORA

Universidad de Dune, Grupo de Barton, Paris, Francia.

DESCRIPCIÓN

El servidor Jpred es un servidor interactivo de predicción de estructura secundaria. Jpred acepta como archivos de entrada secuencias singulares y alineamientos múltiples de secuencia. El reporte de salida resume la predicción generada a partir de 6 algoritmos diferentes que aprovechan información evolutiva tomada de múltiples secuencias: DSC (King y Sternberg, 1996), PHD (Rost y Sander, 1996), NNSSP (Salamos y Solovyev, 1995), PREDATOR (Frishman y Argos 1997), ZPRED (Zbelevil *et al.*, 1987) y MULPRED (Barton, 1988).

Cada uno de estos algoritmos emplea diferentes métodos heurísticos, tales como redes neuronales (NNSSP), redes neuronales basadas en decisión (PHD), discriminación lineal (DSC), Métodos combinados de secuencias singulares (MULPRED), propiedades de enlace de hidrógeno (PREDATOR), y predicción de número de pesos conservados (ZPRED). El archivo de salida además, contiene información de propiedades fisicoquímicas, accesibilidad al solvente, entre otros, para cada uno de los aminoácidos evaluados. El servidor Jpred posee la opción de descargar los datos originales de cada método de predicción y simplifica el uso de los algoritmos de predicción permitiendo identificar importantes estructuras

Centre for Integrative Bioinformatics VU
vrije Universiteit amsterdam



SYMPRED (Optimización simétrica de predicción estructural de proteínas)
URL:<http://zeus.cs.vu.nl/programs/sympredwww/>

ENTIDAD ADMINISTRADORA

Centro Integral de Bioinformática, Vrije Universidad de Amsterdam, Holanda.



UNIVERSIDAD DE CARTAGENA

DESCRIPCIÓN

SYMPRED es un consenso de programas de predicción de estructura secundaria que emplean algoritmos BLAST para producir un archivo de entrada por método individual que sea empleado.

El consenso utiliza algoritmos de programación dinámica en la que todas las predicciones son óptimamente segmentadas a un alto nivel de precisión (Simossis *et al.*, 2004). SYMPRED puede ser ejecutado empleando combinaciones de PHDpsi (Przybylski y Rost, 2002), PROFsec (Rost y Eyrich, 2001) SSPro2.01 (Pollastri *et al.*, 2002), YASPIN (<http://ibivu.cs.vu.nl/programs/yaspin> www/) (Lin *et al.*, 2005), JNet (Cuff y Barton, 2000) y PSIPred (Jones, 1999). Todos estos programas emplean varios tipos de redes neuronales (NNs) que realizan predicciones a partir de alineamientos de múltiples secuencias (MSAs) homólogas.

Ejemplo: La Kiwelin es una proteína de 28 kD rica en cisteínas que reconoce IgE de pacientes alérgicos al Kiwi (*Actinidia chinensis*) (Dalal *et al.*, 2005). La secuencia en formato fasta (Co. Acc. Swiss-Prot P84527) de esta proteína fue utilizada como información de entrada por el servidor SYMPRED para predecir la estructura secundaria. El reporte de salida en primera instancia indica que los resultados de la predicción fueron enviados al correo electrónico suministrado por el usuario.

Es posible tener acceso a los resultados generados en un máximo de 15 días a partir del momento en que fue introducida la secuencia al servidor. De igual manera es descrito el servidor consenso y aparece una tabla con vínculos donde es posible acceder a los resultados en diversos formatos (Figura 8.1A). La sección siguiente presenta una representación gráfica que indica el asignamiento de estructuras secundarias para cada aminoácido en la secuencia (H = hélices, E = hojas beta). Los puntos indican que el servidor no asignó estructura secundaria para dichos aminoácidos en la secuencia.

En la Figura 8.1B observamos que la proteína kiwelin tiene una mayor probabilidad de tener hojas beta. Existe una única región conformada por una alfa hélice del residuo 166 al 172.



SYMPRED Consensus Secondary Structure Prediction results for job Untitled

[Back to SYMPRED main page](#)

An email has been sent to jeffrey2004@yahoo.es with the results.

The SYMPRED consensus secondary structure prediction and the predictions of your chosen individual methods for your query sequence are directly under their corresponding amino acid.

The methods that were unable to complete the prediction successfully are denoted by a continuous dotted line.

In the event that a DSSP entry has been found for your query sequence, the errors made by each prediction method and SYMPRED will be visible under each prediction as a Length (l), Overprediction (o), Underprediction (u) or Wrong (w) according to the scheme used by [McGuffin and Jones 2003](#).

SYMPRED files	Raw output Human readable Machine Readable
PSI-BLAST files (nr)	Alignment file PSSM file
Prediction files	SSPRO YASPIN JNet PSIPred

You have run SYMPRED using the Dynamic Programming strategy with no weighting.

FILE: uploaded.seq
 # SEQUENCE NAME: sp|P84527|KIWEL_ACTCHKiwelin(AllergenAct?)-Actinidiachinensis(Kiw)(Yangtao).
 # SEQUENCE LENGTH: 189

8.1A

Figura 8.1.

Resultados de salida del servidor Sympred para la secuencia de la Kiwelina.

- A) Breve descripción del servidor y archivos de resultados.
- B) Asignamiento de la estructura secundaria a cada una de los aminoácidos de la secuencia blanco.

Ruler11.....21.....31.....41.....51.....61.....
Sequence	ISSCNGPCRDLMDCDGLICIRKRCNDDPQVGVTHICRGTTTSHQPGGCKP8GLTCRGSYPTDCCSPV					
PHD					
PROF					
SSPRO	EEEE		EE			
PREDATOR					
YASPIN	EEEE		EE		EEEE	
JNET	EEEE		E		EE	
PSIPRED	EEEE		EE			
SYMPRED	EEEE		EE			

Ruler81.....91.....101.....111.....121.....131.....
Sequence	TSSTPAKLTNDDFSEGGDDGCPSECDSEYHNNRIVALSTCWYNGSPCGMIRITASNGKSVSAKVVD					
PHD					
PROF					
SSPRO			EEEE		EEEE EEEEE	
PREDATOR					
YASPIN	EEEE		EEEE		EEEE EEEEE	
JNET	EE		EEEE		EEEE EEEEE	
PSIPRED	HHH		EEEE		EEEE EEEEE	
SYMPRED			EEEE		EEEE EEEEE	

Ruler151.....161.....171.....181.....
Sequence	ECDSRHCDREHACQPPCRNNIVDGSNAVWSALGLDKNVGVVDITWSMA			
PHD			
PROF			
SSPRO	HHHHHH		EEEE	
PREDATOR			
YASPIN	HHHHHH		EEEE	
JNET	HHHHHH		EEEE	
PSIPRED	E	HHHHHH		EEEE

8.1B



OTROS SERVIDORES DE PREDICCIÓN DE ESTRUCTURA SECUNDARIA

HYPROSPII - A knowledge-based hybrid method for protein secondary structure prediction based on local prediction confidence



HYPROSP II

URL: <http://bio-cluster.iis.sinica.edu.tw/~bioapp/hyprosp2/>
(Lin *et al.*, 2005).



PORTER

URL: <http://distill.ucd.ie/porter/>
(Pollastri *et al.*, 2005).

Un número extenso de métodos han sido desarrollados para predecir estructuras secundarias regulares tales como giros α (Kaurt *et al.*, 2004a), giros β (Kaurt *et al.*, 2002a; Kaurt *et al.*, 2002b; Kaurt *et al.*, 2003a; Kaurt *et al.*, 2003b; Kaurt *et al.*, 2004b; Fuchs *et al.*, 2005) y giros γ (Kaurt *et al.*, 2003c). La predicción de giros β y giros γ son útiles para el diseño de nuevas moléculas tales como fármacos, pesticidas y antígenos, entre otros. Por tal razón, se han creado métodos de predicción basados en máquinas con vectores de soporte (SVM) llamados BTSVM y GTSVM, los cuales predicen giros β y giros γ , respectivamente. Estos procedimientos están dentro de los métodos de aprendizaje de máquinas, la cual son modelos de entrenamiento supervisado utilizados en problemas de clasificación binaria, fácilmente extensibles a modelos de clasificación múltiple.

UNIVERSIDAD DE CARTAGENA



OJO

Los resultados de estos análisis son mas comprensivos y fáciles de usar que los resultados previos utilizados en el diseño de giros en proteínas (Pham *et al.*, 2005).

SERVIDORES DE INTERNET QUE PREDICEN GIROS EN ESTRUCTURAS SECUNDARIAS.

BetaTPred2: Prediction of β -turns in proteins using neural networks and multiple alignment

BetaTPred2

URL:

<http://www.imtech.res.in/raghava/betatpred2/>
(Kaurt *et al.*, 2003a).

BTEVAL: EVALUATION OF BETA TURN PREDICTION METHODS

BTEVAL

URL: <http://mail1.imtech.res.in/raghava/bteval/>
(Kaurt *et al.*, 2003b).

A server for β -turn types prediction
Bioinformatics Centre, Institute of Microbial Technology
Chandigarh, INDIA

BetaTurns

URL:

<http://www.imtech.res.in/raghava/beteturns/>
(kaurt *et al.*, 2004b).

ALPHAPRED: A server for prediction of alpha turns in proteins

ALPHAPRED

URL: <http://www.imtech.res.in/raghava/alphapred/submit.html>
(Kaurt *et al.*, 2004).

Las hojas β y algunos motivos estructurales suprasecundarios tales como las horquillas β juegan un papel fundamental en la estructura, función, evolución y bioingeniería de proteínas. El servidor BetaPro es un ejemplo de predicción de hojas β en proteínas (Chen *et al.*, 2005).



UNIVERSIDAD DE CARTAGENA

SERVIDORES DE INTERNET QUE PREDICEN ESTRUCTURAS HOJAS B



BETAPro

URL: <http://www.ics.uci.edu/~baldig/betasheet.html>
(Chen *et al.*, 2005).



BhairPred

URL: <http://www.imtech.res.in/raghava/bhairpred/>
(Kumar *et al.*, 2005).

Los ovillos (enrollamientos) constituyen motivos que aparecen con alta frecuencia en proteínas, presentando importancia para la estructura y función de diversas clases de proteínas fibrosas, factores de transcripción y fusión de membranas de proteínas. La predicción de regiones coil en proteínas puede ser usada para identificar dominios putativos de oligomerización y postular mecanismos funcionales. Además, tales predicciones son necesarias como primer paso en el entendimiento de interacciones entre regiones de enrollamiento y el estudio de procesos complejos de mantenimiento de la infraestructura subcelular.

De otra parte, estos métodos para predecir con exactitud regiones de enrollamiento son importantes para coleccionar información producto de los proyectos de secuenciación de genomas. Los métodos basados en la predicción de regiones de enrollamiento, tales como COILS (http://www.ch.embnet.org/software/COILS_form.html) (Lupas *et al.*, 1991), Multicoil (Wolf *et al.*, 1997) (<http://multicoil.lcs.mit.edu/cgi-bin/multicoil>) y Marcoil (<http://bioinf.wehi.edu.au/folders/mauro/marcoil/index.html>) (Delorenzi y Speed, 2002) han sido absolutamente acertados.

SERVIDORES Y RECURSOS DE INTERNET QUE PREDICEN REGIONES DE ENROLLAMIENTO (COIL) Y CON ESTRUCTURA ALEATORIA.



Coiled-Coil prediction

URL: http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_lupas.htm





PairCoil2

URL: <http://groups.csail.mit.edu/cb/paircoil2/>
(McDonnell *et al.*, 2006)



DISPROT

URL : <http://www.disprot.org/>
(Vucetic *et al.*, 2005).



IUPRED

URL: <http://iupred.enzim.hu/index.html>
(Dosztanyi *et al.*, 2005).



SCRATCH (Servidor de predicción de futuros estructurales Y estructuras de proteínas).

URL: <http://www.ics.uci.edu/~baldig/scratch/>
(Cheng *et al.*, 2005).



UNIVERSIDAD DE CARTAGENA

GLOBPLOT 2



GLOBPLOT 2

URL: <http://globplot.embl.de/>
(Linding *et al.*, 2003).



Disopred

URL: <http://dis.embl.de/>



RONN (Predicción de regiones en desorden de proteínas) URL :<http://www.strubi.ox.ac.uk/RONN>
(Shen *et al.*, 2005).

REFERENCAS

- Albrecht, M., Tosatto, S., Lengauer, T., Valle, G. (2003). Simple consensus procedures are effective and sufficient in secondary structure prediction. *Protein Eng.* 16(7):459-462.
- Bryson, K., McGuffin, J., Marsden, R. L., Ward, J., Sodhi, J., Jones, D. (2005). Protein structure prediction servers at University College London. *Nucleic Acids Res.* 33(Web Server issue):36-38.
- Cheng, J., Baldi, P. (2005). Three-Stage Prediction of Protein Beta-Sheets by Neural Networks, Alignments, and Graph Algorithms. *Bioinformatics.* 21(Suppl 1):75-84.
- Combet C., Blanchet C., Geourjon C., Deléage G. (2000). NPS@: Network Protein Sequence Analysis. *TIBS* 2000. 25(3):147-150.
- Cuff, J., Barton, G. (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins.* 40(3):502-511.
- Cuff, A., Clamp, E., Siddiqui, S., Finlay, M., Barton, J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics.* 14(10):892-893.
- De la Cruz, X., Hutchinson, G., Shepherd, A., Thornton, M. (2002). Toward Predicting protein topology: an approach to identifying B hairpins. *Proc. Natl Acad. Sci. USA.* 99(17):11157-11162.
- Deleage, G., Roux, B. (1987). An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng.* 1(4):289-294.
- Delorenzi, M., Speed, T. (2002). An HMM model for coiledcoil domains and a comparison with PSSM-based predictions. *Bioinformatics.* 18(4):617-625.
- Dosztanyi, Z., Csizmok, V., Tompa, P., Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.* 21(16):3433-3434.
- Frishman, D., Argos, P. (1996). Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng.* 9(2):133-142.
- Frishman, D., Argos, P. (1997). Seventy-five percent accuracy in protein secondary structure prediction. *Proteins.* 27(3):329-335.
- Fuchs, P.F., Alix, A.J. (2005). High accuracy prediction of beta-turns and their types using propensities and multiple alignments. *Proteins.* 59(4):828-39.



- Garnier, J., Gibrat, J-F., Robson, B. (1996). GOR secondary structure prediction method version IV. *Meth. in Enzymol.* 266:540-553.
- Geourjon, C., Deleage, G. (1994). SOPM: a self-optimized method for protein secondary structure prediction. *Protein Eng.* 7(2):157-164.
- Gibrat, F., Garnier, J., Robson, B. (1987). Further developments of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs. *J. Mol. Biol.* 198 (3):425-443.
- Geourjon, C., Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* 1(6):681-684.
- Guermeur, Y. (1997). Combinaison de classifieurs statistiques. Application a la prediction de structure secondaire des proteines. Ph.D. Thesis, Ecole Normale Superieure de Lyon, France.
- Guermeur, Y., Geourjon, C., Gallinari, P., Deleage, G. (1998). Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics.* 15(5):413-421.
- Jones, T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292(2):195-202.
- Kaur, H., Raghava, S. (2004a). Prediction of alpha-turns in proteins using PSI-BLAST profiles and secondary structure information. *Proteins.* 55(1):83-90
- Kaur, H., Raghava, S. (2004b). A neural network method for prediction of β -turn types in proteins using evolutionary information. *Bioinformatics.* 20(16):2751-2758.
- Kaur, H., Raghava, S. (2002a). BetaTPred: Prediction of β -turns in a protein using statistical algorithms. *Bioinformatics.* 18:498-499.
- Kaur, H., Raghava, P. (2002b). An evaluation of β -turn prediction methods. *Bioinformatics.* 18:1508-1514.
- Kaur, H., Raghava, P. (2003a). Prediction of β -turns in proteins from multiple alignment using neural network. *Protein Sci.* 12:627-634.
- Kaur, H., Raghava, P. (2003b). BTEVAL: a server for evaluation of β -turn prediction methods. *J. Bioinform. Comput. Biol.* 1:495-504.



multiple sequence alignment. *Protein Sci.* 12:923-929.

• King, D., Sternberg, J. (1996). Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci.* 5(11):2298-2310.

• Kneller, D., Cohen, F., Langridge, R. (1990). Improvements in protein secondary structure prediction by an enhanced neural network. *J. Mol. Biol.* 214(1):171-182.

• Kuhn, M., Meiler, J., Baker, D. (2004). Strand-loop-strand motifs: prediction of hairpins and diverging turns in proteins. *Proteins.* 5:282-288.

• Kumar, M., Bhasin, M., Natt, K., Raghava, P. (2005). BhairPred: prediction of beta-hairpins in a protein from multiple alignment information using ANN and SVM techniques. *Nucleic Acids Res.* 33(Web Server issue):154-159.

• Levin, M., Robson, B., Garnier, J. (1986). An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS Lett.* 205(2):303-308.

• Linding, R., Russell, B., Neduva, V., Gibson, J. (2003). GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 31(13):3701-3708.

• Lin, K., Simossis, A., Taylor, R., Heringa, J. (2005). A simple and fast secondary structure prediction algorithm using hidden neural networks. *Bioinformatics.* 21(2):152-159.

• Lin, N., Chang, M., Wu, P., Sung, Y. (2005). Hsu WL. HYPROSP II—a knowledge-based hybrid method for protein secondary structure prediction based on local prediction confidence. *Bioinformatics.* 21(15):3227-3233.

• Lupas, A., Van Dyke, M., Stock, J. (1991). Predicting coiled coils from protein sequences. *Science.* 252(5010):1162-1164.

• Qin, S., He, Y., Pan, M. (2005). Predicting protein secondary structure and solvent accessibility with an improved multiple linear regression method. *Proteins.* 61(3):473-480.

• McDonnell, V., Jiang, T., Keating, E., Berger, B. (2006). Paircoil2: Improved prediction of coiled coils from sequence. *Bioinformatics.* 22(3):356-358.

• Pham, H., Satou, K., Ho, B. (2005). Support vector machines for prediction and analysis of beta and gamma-turns in proteins. *J. Bioinform. Comput. Biol.* 3(2):343-358.

• Pollastri, G., Przybylski, D., Rost, B., Baldi, P. (2002). Improving the prediction of protein secondary



structure in three and eight classes using recurrent neural networks and profiles. *Proteins*. 47(2):228-235.

• Przybylski, D., Rost, B. (2002). Alignments grow, secondary structure prediction improves. *Proteins*. 46(2):197-205.

• Rost, B., Sander, C. (1993). Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Natl. Acad. Sci. USA*. 90(16):7558-7762.

• Rost, B., Eyrich, V. (2002). Predictions to Automatic Servers EVA: Large-scale analysis of secondary structure prediction. *Proteins*. 45:192-199.

• Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods Enzymol*. 266:525-539.

• Pollastri, G., McLysaght, A. (2005). Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics*. 21(8):1719-20.

• Salamov, A., Solovyev, V. (1995). Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments. *J. Mol. Biol*. 247(1):11-15.

• Selbig, J., Argos, P. (1998). Relationships between protein sequence and structure patterns based on residue contacts. *Proteins*. 31(2):172-85.

• Simossis, A., Heringa, J. (2004). Optimally segmented consensus secondary structure prediction. *Bioinformatics*. (Submitted).

• Sun, Z., Rao, X., Peng, L., Xu, D. (1997). Prediction of protein supersecondary structures based on artificial neural network method. *Protein Eng*. 10:763-769.

• Vucetic, S., Obradovic, Z., Vacic, V., Radivojac, P., Peng, K., Lakoucheva, M., Cortese, S., Lawson, D., Brown, J., Sikes, G., Newton, D., Dunker, K. (2004). DisProt: a database of protein disorder. *Bioinformatics*. 21(1):137-140.

• Wolf, E., Kim, S., Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci*. 6:1179-1189.

• Yang, R., Thomson, R., McNeil, P., Esnouf, M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*. 21(16):3369-3376.



• Zvelebil, J., Barton, J., Taylor, R., Sternberg, J. (1987). Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J. Mol. Biol.* 195(4):957-961.

I	N	A	T	G	M
I	N	A	G	I	T
I	N	A	I	S	G
I	N	A	M	G	L



H	H	H	T	T	T
---	---	---	---	---	---

