

Bioinformatic Analysis of Toll-Like Receptor Sequences and Structures

Tom P. Monie, Nicholas J. Gay, and Monique Gangloff

Abstract

Continual advancements in computing power and sophistication, coupled with rapid increases in protein sequence and structural information, have made bioinformatic tools an invaluable resource for the molecular and structural biologist. With the degree of sequence information continuing to expand at an almost exponential rate, it is essential that scientists today have a basic understanding of how to utilise, manipulate and analyse this information for the benefit of their own experiments. In the context of Toll-Interleukin I Receptor domain containing proteins, we describe here a series of the more common and user-friendly bioinformatic tools available as Internet-based resources. These will enable the identification and alignment of protein sequences; the identification of functional motifs; the characterisation of protein secondary structure; the identification of protein structural folds and distantly homologous proteins; and the validation of the structural geometry of modelled protein structures.

Key words Toll-like receptor, TLR, Toll-Interleukin-1 Receptor Domain (TIR), Bioinformatics, Sequence alignment, Sequence comparison, Homology, Structure validation, FUGUE

1 Introduction

Toll-like receptors (TLRs) are type I transmembrane receptors. They are constituted of a leucine-rich repeat ligand-binding domain, a single membrane spanning helix and a signalling Toll-Interleukin-1 Receptor (TIR) domain [1, 2]. TLRs recognise a diverse range of microbial ligands. Following ligand binding, the TLRs undergo conformational change enabling the initiation of signal transduction [3]. The TIR domains possess a conserved $\alpha\beta$ structural organisation essential for signal transduction [4]. Indeed, paralogues of individual TLR TIRs show particularly high levels of amino acid conservation.

In this chapter, we describe the use of the classic bioinformatic tools BLAST [5, 6] and Clustal Ω [7], for the identification and alignment of TLR TIR paralogues. We also address the identification of structurally homologous proteins and the annotation of a

protein's three-dimensional environment through the use of the programs FUGUE [8] and JOY [9]. Moreover, we describe the use of available resources for the identification of functional motifs within proteins and the validation of the stereochemistry of protein structures. These techniques are highlighted with examples from TIR containing proteins.

These tools provide an important set of resources that, when used either individually, or in conjunction with one another, can greatly assist with multiple aspects of the study of TLRs. For example, they enable important functional and structural observations to be made about specific proteins. Additionally, they can aid the design of expression constructs for structural and biochemical studies and assist in the design of rational mutagenesis for functional work.

2 Materials

- | | |
|---|---|
| 2.1 TLR Orthologues | <ol style="list-style-type: none"> 1. Human TLR4 amino acid sequence (Accession Number O00206). 2. Multiple TLR4 orthologue sequences (<i>see Note 1</i>). |
| 2.2 Sequence–Structure Homology | <ol style="list-style-type: none"> 1. Human TLR4 amino acid sequence (Accession Number O00206). Select the region from residue 674 to 839. 2. Key to formatted Joy alignments (<i>see Table 1</i>). |
| 2.3 Three-Dimensional Structure Comparison | <ol style="list-style-type: none"> 1. The TLR1, TLR2 and TLR10 TIR crystal structure PDB (Protein Data Bank) codes. These are 1fyv, 1fyw and 2j67 respectively (<i>see Note 2</i>). |

Table 1
Key to formatted Joy alignments

Structural features	Labelling	Residue format
Alpha helix	Red	x
Beta strand	Blue	x
3 ₁₀ helix	Maroon	x
Solvent accessible	Lower case	x
Solvent inaccessible	Upper case	X
Hydrogen bond to main-chain amide	Bold	x
Hydrogen bond to main-chain carbonyl	Underline	<u>x</u>
Disulfide bond	Cedilla	ç
Positive phi torsion angle	Italic	<i>x</i>

X: any amino acid; ç: a half-cystine residue

2.4 Structural Validity

1. PDB file for model to be validated.

2.5 Post-translational Modifications

1. Human TRIF-related adaptor molecule, TRAM, also known as TICAM-2, amino acid sequence (Accession number NP_067681).

3 Methods

3.1 TLR Orthologues

Structurally and functionally important regions of homologous proteins often have high levels of amino acid conservation. Alignment and comparison of the amino acid sequence of homologous proteins from different species (i.e. protein orthologues) can be extremely helpful experimentally through the identification of key functional residues and protein domain boundaries. Here, we describe how to identify and align orthologues of TLR4.

3.1.1 BLAST Search

1. BLAST (Basic Local Alignment Search Tool) identifies regions of local similarity between the query and database sequences.
2. Paste the human TLR4 amino acid sequence into the query window of the NCBI-BLAST2—Protein Database page (<http://www.ebi.ac.uk/Tools/sss/ncbiblast/>).
3. Check that the program selected is blastp and the database is protein and UniProtKB/Swiss-Prot. Run Blast (*see Note 3*).
4. A table of results will be generated showing information about homologous sequences such as: protein description and source, length, identity, score and E value (*see Note 4*). From these results, it is possible to select TLR4 orthologues identified and download the sequences in a FASTA format (*see Note 5*).

3.1.2 ClustalΩ Multiple Sequence Alignment

1. Copy the FASTA formatted orthologues downloaded from the BLAST search (Subheading 3.1.1) into the input query field on the EMBL-EBI ClustalΩ server web-page (www.ebi.ac.uk/Tools/msa/clustalo/).
2. The default parameters can normally be retained. Run ClustalΩ (*see Note 3*).
3. A series of alignment and similarity results will be generated. These include pairwise scores for each sequence aligned, phylogram and cladogram trees, and a multiple sequence alignment (Fig. 1).
4. The multiple sequence alignment is especially useful for identifying regions of high and/or low conservation, domain boundaries and potential substitutions for mutagenic studies.

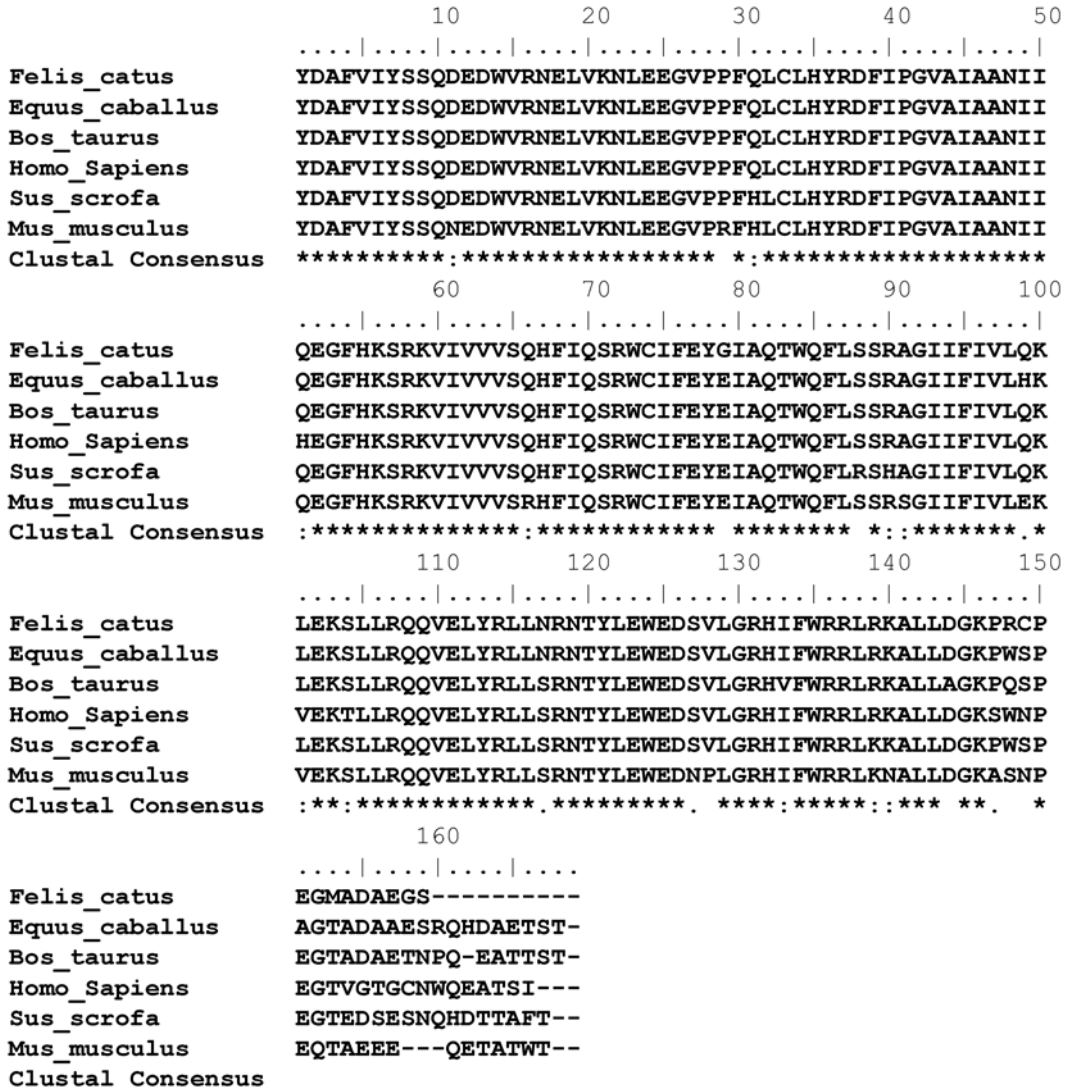


Fig. 1 Example ClustalΩ multiple sequence alignment. The TIR signalling domains of six of the TLR4 orthologues (host species as labelled in *figure panel*) identified by a BLAST search (Subheading 3.1.1) were submitted for ClustalΩ multiple sequence alignment (Subheading 3.1.2). The Clustal consensus sequence identifies fully conserved residues (*), strongly similar substitutions (:), weakly similar substitutions (.), and a lack of consensus (). The consensus sequence highlights the high degree of conservation in the TLR4 TIR, in contrast the very C-terminus of the protein shows significant variation

3.2 Sequence-Structure Homology

Sequence and structural information can be simultaneously used to improve the homology recognition power and the accuracy of sequence alignments (*see Note 6*). Identifying structural homology between a protein sequence of unknown three-dimensional structure and one with known structure provides useful information for understanding protein function. It also provides another

layer of information and reflects the high evolutionary pressure for structurally and functionally important residues in a given protein family. In other words, such alignments help identify divergently evolved (homologous) proteins with structural and functional relationships. Furthermore, it allows prediction of the three-dimensional structure through comparative modelling, a technique which is beyond the scope of this chapter (*see* **Note 7**). Here, we demonstrate how to use the program FUGUE to identify structural homologues for the TLR4 TIR domain. Unlike the TIR domains of TLR1, 2 and 10, the structure of the TLR4 TIR domain has not yet been solved experimentally.

Annotation of protein sequence alignments with three-dimensional structural features is a useful tool for identifying key structural and functional residues. This can be achieved with a program such as Joy, which provides a modified version of the one-letter amino-acid code in order to convey structural information (*see* Table 1).

3.2.1 Sequence– Structure Homology with Fugue

1. Open the Fugue web-page (<http://tardis.nibio.go.jp/fugue/prfsearch.html>) and enter your e-mail address and the amino acid sequence of the human TLR4 TIR domain (residues 673–839).
2. Keep the default parameters and click on search. The output is sent via e-mail and results can be accessed at <http://tardis.nibio.go.jp/result/fugue/1146/fugue.html>.
3. The Fugue result for human TLR4 TIR domain reveals that the HOMSTRAD [10] profile hs1fyxa (*see* **Notes 8 and 9**) has the highest *Z*-score. With over 99 % confidence, the suggested homology is certain (Fig. 2a).
4. The HOMSTRAD family called TIR (*see* **Note 9**), which was built on the crystal structures of human TLR1 and TLR2 TIR domains, is the second best hit. The low *Z*-score of other hits makes them less reliable.
5. Focus only on the two alignments with the highest *Z*-scores by clicking on ‘alignment’ in the results.

3.2.2 Sequence– Structure Alignment with Joy

1. The alignments mentioned in **step 5**, Subheading 3.2.1 (Fig. 2b) are represented using the Joy annotation described in Table 1. In addition to providing a secondary structure prediction for the query sequence they can also be used to highlight differences and/or problem areas within the sequence-structure alignments. These could be, for example: insertions or deletions in regions of helical structure; proline residues in regions of predicted helix; the presence, or substitution, of charged residues (e.g. lysine, arginine) for hydrophobic (e.g. phenylalanine, leucine, isoleucine, tyrosine) ones, and vice versa.

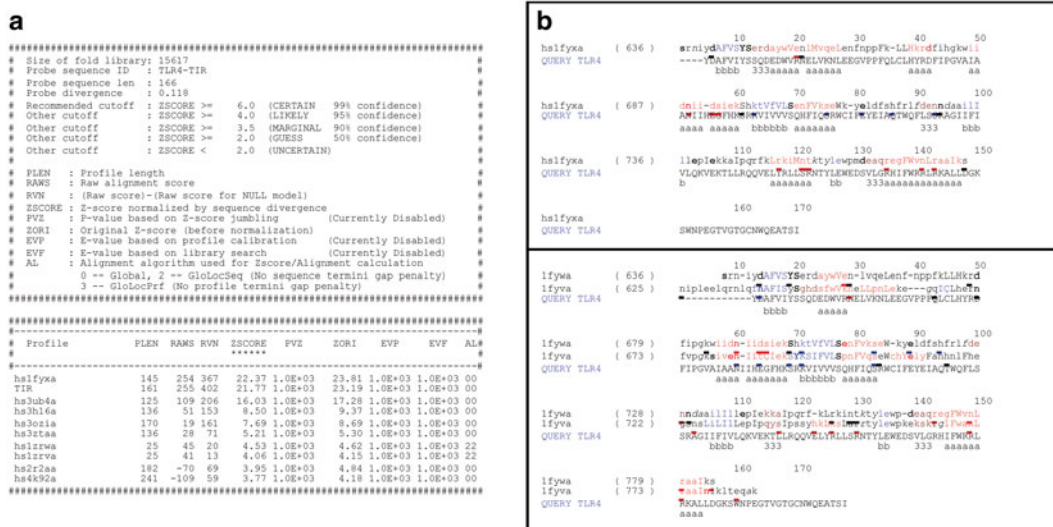


Fig. 2 Example analysis of a TLR4-TIR domain homology search using Fugue. Extract of the output for Fugue sent via e-mail (**a**) and the sequence–structure alignments of the two top hits based on the Joy annotation (**b**)

2. Analysis of the structural alignments reveals that the core of the TIR domain is well conserved between TLR1 (1fyv), TLR2 (1fyw) and TLR4 (Query). There are however apparent differences. For instance, an extra histidine residue at position 724 in human TLR4 interrupts an alpha-helix and is likely to cause some structural distortion. In addition, compared to the structural templates, there are extra residues at the C-terminus of the TLR4 sequence. These are not part of the TIR domain but constitute a tail of unpredictable structure.

3.3 Three-Dimensional Structure Comparison

It can be very helpful to evaluate the degree of three-dimensional structural similarity between either two or more experimentally determined or computer-modelled structures. This can help provide an estimation of structural similarity and/or model/structure reliability. The following method uses the Secondary Structure Matching (SSM) program PDBFold, available at <http://www.ebi.ac.uk/msd-srv/ssm/>, to determine the similarity between experimentally determined TLR TIR domains.

3.3.1 Pairwise Structural Comparison

1. Choose the pairwise 3D alignment submission option and select ‘PDB entry’ for both the query and target sequences.
2. Insert the PDB codes for TLR1 (1fyv) and TLR2 (1fyw) TIR domains in the query and target fields respectively.
3. Retain the default parameters and submit query (*see* **Note 10**).
4. An output table detailing the 3D structural similarity will be generated. In general, the higher the number of aligned residues and the lower the rmsd (Root mean square deviation of

C α atoms) the greater the degree of structural similarity (*see Note 11*). The values for the TLR1 and TLR2 structural comparison suggest a high degree of structural similarity.

3.3.2 Multiple Structural Comparison

1. Choose the multiple 3D alignment submission option and select 'PDB entry' as the source.
2. Input the TLR1 TIR PDB code (1fyv) and press the 'Actualize' button, followed by the new entry button. Repeat for TLR2 (1fyw).
3. Input the TLR10 (2j67) TIR structure files and press 'Find Chains'; delete B, Y and Z from the text box then press 'Actualize'. This removes unnecessary information as the TLR10 structure was a dimer. Submit query.
4. The results page will contain information relating to the similarities of the 3D superposition of the structure. This will include rmsd and Q scores, alignment of secondary structure elements and a structural alignment of input files. The aligned files can be viewed individually, as a superposition, or downloaded.

3.4 Structural Validity

There are many computer packages that will produce structural models with little more user input than an amino acid sequence. However, the models produced may contain regions of either poor, or disallowed, stereochemistry. It is always advisable to validate the geometry of any models generated. Two good ways to do this: use the programs Verify3D and Rampage. Verify3D assesses the sequence position and structural environment of the model and compares them to databases of known high-quality structures. Rampage provides a Ramachandran plot analysis to assess the stereochemical environment of the backbone torsion angles in the modelled structure.

1. Upload, and submit for analysis, the co-ordinate PDB file of the modelled structure to the servers for Verify3D (http://services.mbi.ucla.edu/Verify_3D/) and RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>).
2. Sample results for a TLR4 TIR homodimer model are in Fig. 3.
3. Verify3D scores each residue on a scale of -1 to +1 and a score of >0.2 suggests that the residue is in a structurally favourable environment. Regions with scores below this suggest that those parts of the model must be viewed as less reliable. The region of output shown in Fig. 3 indicates that the TLR4 model submitted has all Verify3D scores over 0.2 and therefore possesses high-quality stereochemistry, with the individual residues being found in structurally favoured environments.

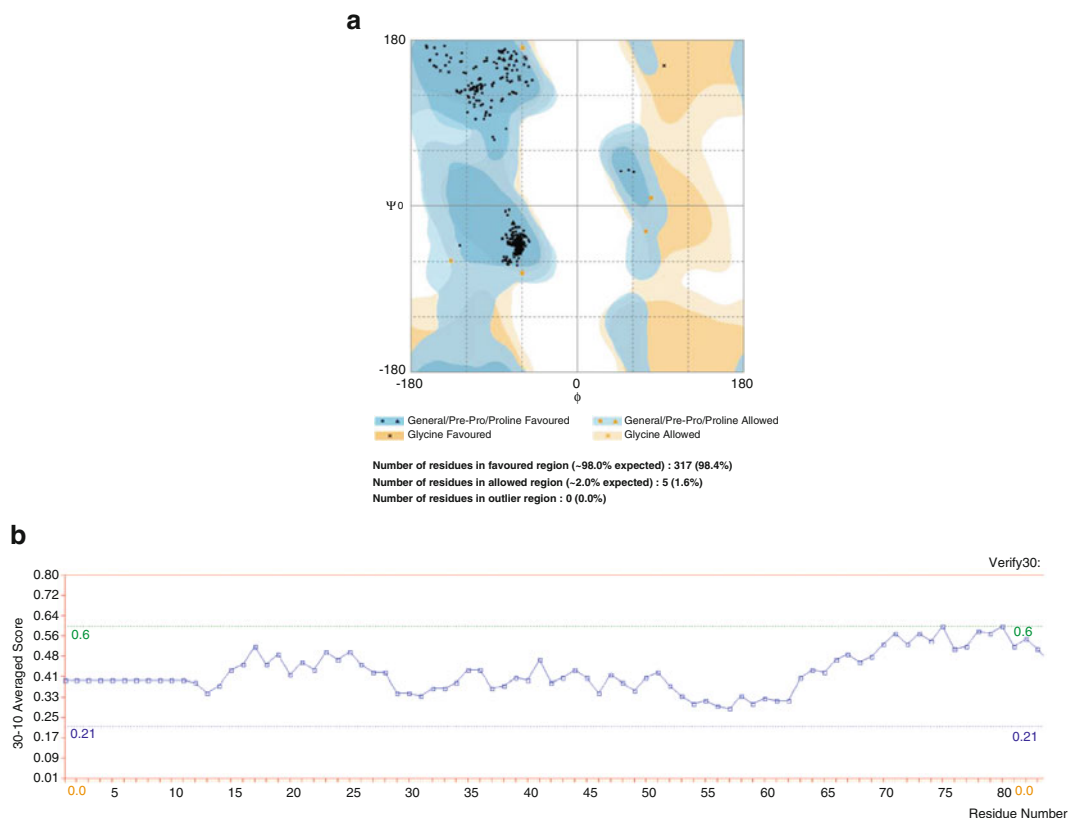


Fig. 3 Example analysis of a modelled TLR4 homodimer using RAMPAGE (a) and Verify3D (b)

4. RAMPAGE produces a clear graphical output of the Ramachandran plot that identifies the proportion of residues in favoured, allowed and disallowed regions. This provides a clear indication of the stereochemical quality of the model. For the TLR4 model submitted (Fig. 3) over 98 % of the residues have torsion angles in the favoured regions, less than 2 % in allowed regions, and there are no outliers. This helps confirm the high-quality stereochemistry of the model.

3.5 Post-translational Modifications

Assessing the presence of post-translational modifications in the Toll receptor pathway proteins is critical for understanding the biology of Toll signalling. Many tools exist for this purpose (*see Note 12*) and here we use one to identify a protein myristoylation site on the TIR containing adaptor protein TRAM. Myristoylation anchors the adaptor protein to the plasma membrane, where it fulfils its biological role in transferring the signal of activated Toll receptors. The linkage occurs on a consensus sequence consisting of Gly-X-X-X-Ser/Thr-Lys/Arg, where X stands for any amino acid. The 14 carbon fatty acid, myristic acid, is covalently attached

by amide linkage to the N-terminal glycine of a protein by an N-terminal myristoyltransferase.

1. Copy the FASTA formatted TRAM protein sequence into the query field on the NMT server web-page (<http://mendel.imp.ac.at/myristate/SUPLpredictor.htm>).
2. Keep the default parameter of 'Eukaryota' as it fits the taxonomy of the sequence.
3. Run the prediction.
4. A reliable myristoylation site is predicted at residue G2 within the sequence GIGKSKINSCPLSLSWG, with an overall score of 0.85 and a probability of false-positive prediction of 1.98×10^{-3} .
5. A logical progression would be to confirm the presence and biological relevance of this modification. Indeed site-directed mutagenesis of the predicted myristylation residue (Gly2Ala) and confocal microscopy experiments have determined that wild-type TRAM is myristylated and localises to the plasma membrane. In contrast, a G2A mutant TRAM has a cytoplasmic distribution and is unresponsive to lipopolysaccharide stimulation [11].

4 Notes

1. These can be obtained from a BLAST search, *see* Subheading 3.1.1.
2. The full PDB files can be downloaded from the Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>).
3. The default search parameters should be fine for these applications. If the user wants further information regarding parameter attributes and variation it is recommended that they read the related program documentation available through the EMBL-EBI web-site (<http://www.ebi.ac.uk>). The UniProtKB/Swiss-Prot database is the smaller portion of the UniProt database and contains fully annotated sequence information. Using this stops multiple redundant hits being identified. If it was unknown whether orthologues existed then use of the UniProtKB/TrEMBL or UniProt Clusters databases would be more appropriate.
4. The score takes into account the number of gaps and substitutions in the alignment. The greater this number, the better the quality of the alignment. The E value is a measure of the likelihood of the alignment occurring by chance. The smaller this number the less likely the alignment is a result of chance.

5. The first line of a FASTA formatted protein sequence starts with a > followed by descriptive text about the sequence. The second, and subsequent, lines contain the protein sequence in single letter code with no spaces or numbering.
6. A good overview of structural homology modelling can be found in the following reference [12].
7. To find out about the homology modelling approach, go to the Swiss-model (<http://swissmodel.expasy.org>) and the Modeller (<http://www.salilab.org/modeller>) web-pages.
8. HOMSTRAD (HOMologous STRucture Alignment Database) is a curated database of structure-based alignments for homologous protein families. Its web-site can be found at <http://tardis.nibio.go.jp/homstrad/>.
9. FUGUE results are given as a list of potentially matching HOMSTRAD profiles. The code hs1fyxa corresponds to the crystal structure of the TLR2 mutant P681H. 1fyxa relates to the PDB identifier (1fyx; chain A) in the HOMSTRAD 'hs' database. The code TIR refers to the HOMSTRAD family containing the TLR1 and TLR2 crystal structures (PDB 1fyv and 1fyw). Clicking of the listed HOMSTRAD profile in the FUGUE results will open the HOMSTRAD entry and show details of its composition.
10. If using a different query sequence and the whole PDB archive as the target then it may be necessary to lower the percentage similarity cut-off for the lowest acceptable target match in order to get any positive hits.
11. Full details of the interpretation of results and scores can be found at http://www.ebi.ac.uk/msd/EMBO/ssm-tutorial/ssm_tutorial.html. The higher the Z and Q scores the better.
12. A list of programs for the prediction of post-translation modifications can be found on the Expasy tools web-site at <http://www.expasy.org/tools>.

References

1. Akira S, Takeda K (2004) Toll-like receptor signalling. *Nat Rev Immunol* 4:499–511
2. Gay NJ, Gangloff M (2007) Structure and function of toll receptors and their ligands. *Annu Rev Biochem* 76:141–65
3. Gay NJ, Gangloff M, Weber AN (2006) Toll-like receptors as molecular switches. *Nat Rev Immunol* 6:693–8
4. Xu Y, Tao X, Shen B, Horng T, Medzhitov R, Manley JL, Tong L (2000) Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature* 408:111–5
5. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–10
6. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–402
7. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497–500

8. Shi J, Blundell TL, Mizuguchi K (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol* 310:243–57
9. Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP (1998) JOY: protein sequence-structure representation and analysis. *Bioinformatics* 14:617–23
10. Mizuguchi K, Deane CM, Blundell TL, Overington JP (1998) HOMSTRAD: a database of protein structure alignments for homologous families. *Protein Sci* 7:2469–71
11. Rowe DC, McGettrick AF, Latz E, Monks BG, Gay NJ, Yamamoto M, Akira S, O'Neill LA, Fitzgerald KA, Golenbock DT (2006) The myristoylation of TRIF-related adaptor molecule is essential for Toll-like receptor 4 signal transduction. *Proc Natl Acad Sci U S A* 103:6299–304
12. Nunez Miguel R, Shi J, Mizuguchi K (2001) Protein fold recognition and comparative modeling using HOMSTRAD, JOY, and FUGUE. In: Tsigelny IF (ed) *Protein structure prediction: bioinformatic approach*. International University Line, La Jolla, CA

Toll-Like Receptors

Practice and Methods

McCoy, C.E. (Ed.)

2016, XIV, 455 p. 58 illus., 29 illus. in color., Hardcover

ISBN: 978-1-4939-3333-4

A product of Humana Press