Indian Journal of Fundamental and Applied Life Sciences ISSN: 2231– 6345 (Online)
An Open Access, Online International Journal Available at http://www.cibtech.org/jls.htm
2016 Vol. 6 (1) January-March, pp. 68-74/Abdullah et al.
Research Article

PRELIMINARY HOMOLOGY MODELING AND STRUCTURE ANALYSIS OF DUFFY ANTIGEN RECEPTOR 3D STRUCTURE

***Hasdianty Abdullah[1], Mohd Fadzli Ahmad[1], Motharasan Manogaran[1] and Nor Suhaila Yaacob[2]**
[1]Faculty of Science and Biotechnology, Universiti Selangor, Jalan Timur Tambahan, Bestari Jaya, Selangor
[2] Institute of Bio-IT Selangor, Universiti Selangor, Jalan Zirkon A7/A, Seksyen 7, 40000 Shah Alam, Selangor
**Author for Correspondence*

ABSTRACT
Plasmodium knowlesi uses its Duffy antigen receptor to attach on to Duffy antigen, located on the surface of erythrocytes cells. This provides the parasites the entry point to hijack red blood cells. This study is conducted its protein 3D structure and to analyze the structure. Throughout this study, the amino acid sequence of Duffy antigen obtained from Uniprot was used. The appropriate template for homology modeling was determined using BLASTP (Basic Local Alignment Search Tool Protein). The Swiss Model Server was used to build the protein model. The structure was validated using ProQ server where it produced LGScore of 2.904 which lies in the range of extremely good model and MaxSub score of 0.271 which lies in the range of very good model. The structure also satisfies the validation using Ramachandran plot which had resulted in 95.6% number of residues in the favored the region. Structure analysis done using ProFunc had found 2 motifs which indicated the similar function of this protein with other proteins.

Keywords: *Plasmodium Knowlesi, Duffy Antigen Receptor Protein, Bioinformatics, Protein Structure Prediction, Swiss Model*

INTRODUCTION
Malaria, a tropical disease caused by infection with single-celled parasites of the genus *Plasmodium,* is one of the most deadly parasitic diseases in the world (Goh *et al.*, 2013). This zoonotic malaria was considered to be extremely rare until a large case of *P. knowlesi* infections in the Kapit Division of Sarawak, Malaysian Borneo, was reported in 2004 (Singh, 2004). Since then, human cases have been reported in virtually all Southeast Asian countries, and *P. Knowlesi* is now considered the fifth species of Plasmodium causing malaria in humans (White, 2008). Invasion of a malaria parasite into its host erythrocyte depends on the interaction between the parasite's protein and the corresponding receptor on the surface of the erythrocyte (Fong *et al.*, 2014).
P. knowlesi, shares a close phylogenetic relationship with *Plasmodium vivax* (White, 2008) and its morphological features resemble those of either *Plasmodium falciparum* or *Plasmodium malariae* (Lee *et al.*, 2009). *P. knowlesi* have schizont and ring form trophozoite morphology. The asexual cycle *P. knowlesi* is unique compared to other sub species because it's the only primate malaria species characterized by a quotidian (24 hour) asexual blood stage development. Similar to *P. vivax, P. knowlesi* also uses the Duffy blood group antigen as a receptor to invade human erythrocytes (Singh, 2003). This simian malaria infects human erythrocytes through reorganization and binding at Duffy antigen (Gaur *et al.*, 2004). Duffy blood group negative human erythrocytes (FyFy) are resistant to infection by *Plasmodium knowlesi* (Miller *et al.*, 1975).
Protein modeling enables a structure to be predicted from its sequence with the accuracy that is comparable to the best results achieved experimentally. Homology models are used to get a rough idea where the alpha carbons of key residues sit the folded protein. Study on *P. Knowlesi*, Duffy antigen receptor protein is crucial in understanding the mechanism of its attachment.
Therefore, this study aims to predict the of the *P. Knowlesi* Duffy antigen receptor protein through homology modelling method and hence analyse the predicted structure using bioinformatics online tools.

Protein structure prediction is one of the most important goals pursued by bioinformatics and theoretical chemistry.

Protein structure prediction by using bioinformatics approach can involve sequence similarity searches, multiple sequence alignments, identification and characterization of domains, secondary structure prediction, solvent accessibility prediction, automatic protein fold recognition, 6 constructing three-dimensional models to atomic detail, and model validation (Edwards and Cottage, 2003).

MATERIALS AND METHODS

Retrieving and Choosing the Target Amino Acid Sequence from Uniprot Database

Throughout this study, the target sequence obtained from Uniprot was used. Choosing the best sequence had been a huge challenge since there were three protein entries with almost similar score which were the P22545, P50493 and P50494. When compared, only P22545 provided enough information on structural level and hence, it was chosen.

Preparation of Structure Template

The quality of the homology protein model depends on the quality of the sequence alignment and template structure (Nayeem *et al.*, 2006). Choosing a template which has the sequence identity more than 70% will produce a better result while template with less than 20% sequence identity can have very different structure.

To search for the best template, the target sequence was compared using BLASTP algorithm against the protein data bank (PDB). The sequence with similarity greater than 30% with known structure was considered as the template and retrieved in FASTA format. Another criteria being considered for the template selection was the structure validation results obtained from PDB website.

X-ray crystallography results from PDB of four possible templates were compared. 4NUU_A was chosen as the best template because it achieved the best validation result and its structure was resolved using x-ray crystallography with atomic resolution less than 2.0 Å.

Sequence Alignment and the Model Building

Clustal Omega online tool by EMBI-EBI was used to perform the alignment. The tool is developed by the European Bioinformatics Institute (EMBL-EBI) for multiple sequence alignment and is available at http://www.ebi.ac.uk/Tools/msa/clustalo/. The program uses guide tress and HMM profile-profile techniques to generate the alignments.

First, target sequence P22545 was uploaded followed by the template from BLASTP. Since P22545 consists of 1073 amino acids while the template, 4NUU_A consists 317 amino acids, the outcome result produced alignment at the centre of the target sequence. This has caused of no alignment produced at other sites of the target sequence. To reduce gaps errors and to produce better results, the target protein sequence had been shorten according to its neighboring alignment. Once the target chain was edited, both target and template sequences were uploaded and ran in Clustal Omega server again.

The 3D structure of the target protein was then built using the SWISS-MODEL server. SWISS-MODEL (http://swissmodel.expasy.org/) is an automated system for modeling the 3D structure of a protein from its amino acid sequence using homology modeling techniques.

Structure Validation

The 3D model generated by SWISS MODEL (pdb file) was later submitted to ProQ-Protein Quality Predictor and Ramachandran plot to check the quality of the model. According to Wallner B. & Elofsson (2003), ProQ is a neural-network-based method to predict the quality of a protein model that extracts structural features, such 16 as frequency of atom–atom contacts, and predicts the quality of a model, as measured either by LGscore or MaxSub.

Ramachandran Plot is a way to visualize dihedral angles ψ against φ of amino acid residues in protein structure (Hollingsworth and Karplus, 2011). For each conformation, the structure was examined for close contacts between atoms.

If the predicted structure satisfies the validation parameters of ProQ process and Ramachandran plot, then the structure is taken for further analysis.

Structure Analysis

The pdb file obtained from Swiss Model was loaded into the ProFunc online tool (https://www.ebi.ac.uk/thornton-srv/databases/profunc/index.html). The result was then sent by the server to the email provided. The sequence motifs, matching folds, nest analysis and a 3D functional template search result was obtained.

RESULTS AND DISCUSSION

Retrieving the Target Amino Acid Sequence from Uniprot Database

Upon submitting the search request, the server will identify the "request" in database to employ for the search. Once the search was done, the server listed 8,304 related records of Duffy protein receptor available in the database. Therefore, the result had to be filtered out to obtain a finer result.

The result was then filtered with the target bacteria, the *P. knowlesi*. This had successfully resulted in only 43 sequence records of Duffy Protein sequence in *P. knowlesi*. The challenge is now to select the best sequence as the target sequence to be used throughout this study. The best approach is to select the sequence based on sequence review by Swiss-Prot.

Swiss-prot is a curated protein sequence database which strives to provide a high level of annotations (such as the description of the function of a protein, structure of its domains, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration compared to TrEMBL (Apweiler and Bairoch, 1996). TrEMBL strives to comprise all protein sequences that are not yet represented in SWISS-PROT, by incorporating a increasing level of mostly automated annotation (Boeckmann *et al.*, 2003). By comparison, choosing entries from Swiss-prot will be priority since it have more accurate compared to TrEMBL entry. The 43 sequence records were then filtered based on the reviewed sequence by Swiss-Prot and top three sequences listed were taken into consideration.

Choosing the Right Sequence

The problem was choosing sequence in alpha, beta or gamma form, as this can influence the prediction of residues during template selection and alignment. To choose the right sequence further study was conducted.

One of the advantages of using Uniprot is, each sequence in Uniprot provides the protein function, names and taxonomy, structure, 3D database information if present, and cross-references. This allows to look for best criteria between this three entry; P22545, P50493, P50494. By comparing the three of them, entry P22545 was chosen because it provides enough information on the structural level.

Protein sequence P22545 was chosen as the target sequence because for the following reasons; (i) the sequence was obtained by experimental evidence at protein level (ii) shows preliminary result on secondary structure in its annotation, and most importantly, (iii) it was reviewed by Swiss-prot as ‘Experimental evidence at protein level’. According to Uniprot, the value ‘Experimental evidence at protein level’ indicates that there is a clear experimental evidence for the existence of the protein. The status 'Evidence at protein level' indicates that there is clear experimental evidence (such as a characterization paper, partial to complete Edman sequencing, clear identification by mass spectrometry (MSI), X-ray or NMR structure, detection by antibodies) for the existence of this protein (Agostino, 2013). In this study, only P22545 entry indicates its amino acid sequences was obtained through the mentioned method.

Selecting the Best Template

In comparative modeling method such as homology modeling, the three-dimensional structure of a given protein sequence (target) is predicted based primarily on its alignment to one or more proteins of known structure (template). Choosing the best template from among the candidates is a key step as it can affect the final accuracy of the structure significantly.

The search for the correct structure template was done using BLAST (Basic Local Alignment Search Tool). BLAST is a searching database which embeds the algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences with the most similar sequences in the databases.

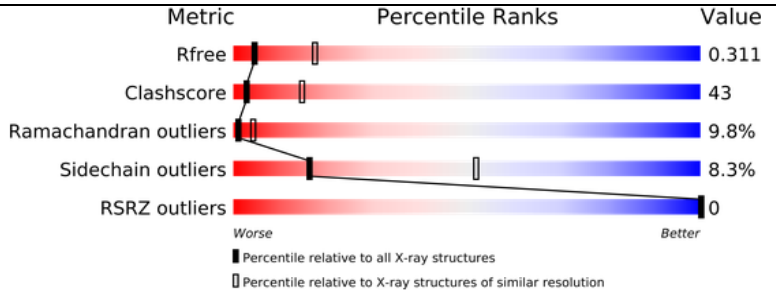
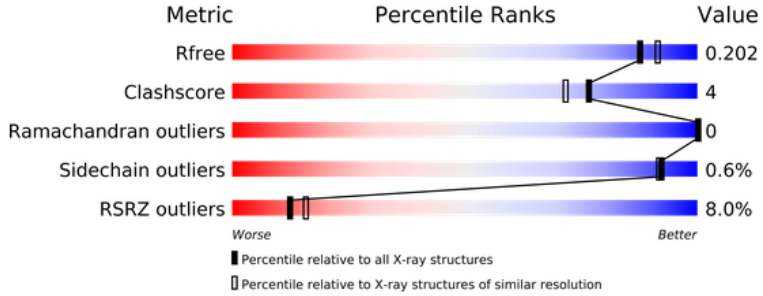
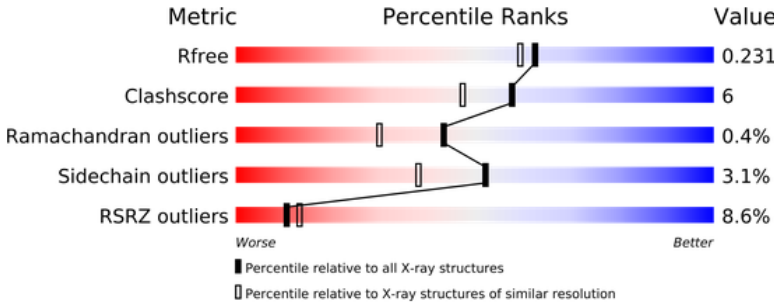
BLASTP analysis against the structural database which is the Protein Databank (PDB) had successfully resulted in three protein entries with identity value greater than 70%.

To avoid choosing the best template, certain factors had to be considered for choosing the right template such as resolution of x-ray crystallography structure, template structure validation results and the chosen structure best resolved one.

When proteins in the crystal are aligned in an identical way, forming almost perfect crystal, then all of the proteins will scatter X-rays the same way, and the diffraction pattern will show the fine details of crystal. On the other hand, if the proteins in the crystal are all slightly different, due to local flexibility or motion, the diffraction pattern will not contain as much fine information. The resolution results of the three possible template structures were obtained from Protein Databank (PDB) annotation and compared. Even though the structures had been determined using the same method, yet they possess different structure resolution.

When compared, only entry no. 4NUU_A and 3RRC_A were found possessing resolution less than 2.0 Å, producing better structural resolution than 2C6J_A.

Table 1: List of Possible Templates, Listed for Selection of Best Template

Acc. No	Structure Resolved Based on	Resolution [Å]	Structure Validation Result
2C6J_A	X-ray diffraction	3.00	
4NUU_A	X-ray diffraction	1.95	
3RRC_A	X-ray diffraction	1.95	

The structure validation result was then used to choose between the two structures as the best template. Table 1 listed out the structure resolution method, resolution value and the structure validation results. From the table, it is shown that 4NUU satisfied all the mentioned factors and hence it was selected as the template for this study.

Sequence Alignment

The subject sequence, the P22545 consists of 1073 amino acids while the template consists of 317 amino acids. When aligned, the alignment was only generated at the middle of the target sequence. This had caused the gaps errors. Therefore, the sequences that did not matched with the template sequence had to be deleted. Once deleted, both the target and template sequences were again aligned using Clustal Omega to obtain better alignment result and the edited template sequence was used to develop the predicted model using Swiss Model.

Prediction of the 3D Structure using Swiss Model

The Swiss Model Server had successfully predicted the 3D protein structure for the target protein and the result was visualized using RasMol viewer. Figure 1 show the predicted model generated.

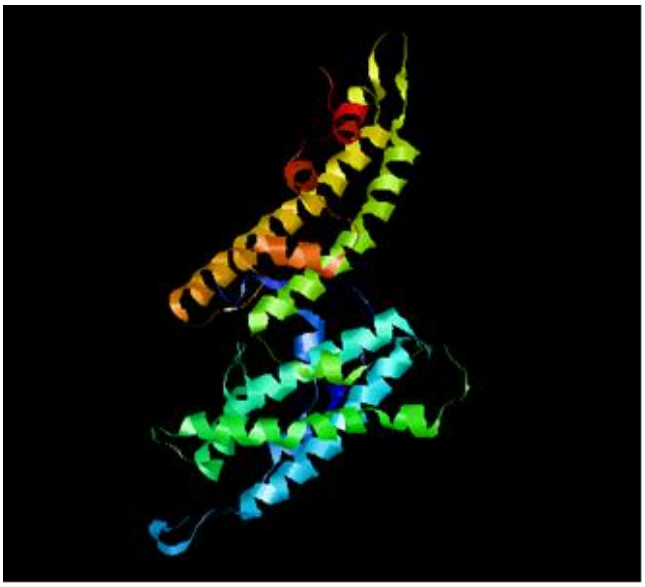


Figure 1: The 3D Protein Structure Predicted using Swiss Model Server Viewed using RasMol Software

Based on the generated 3D structure, one domain is identified, the Domain I. It consists of 229 amino acid residues which contains 8 alpha helices. It has 107 amino acids in hydrophobic residue and 119 amino acids in hydrophilic residue. The domain structure is conventionally divided into two sub domains; (i) sub domain 1 consists of residues LEU60 – PHE155 and, like most other DBL domains, lacks abundant secondary structure but for a single 5-residue helix, and (ii) sub domain 2 incorporates residues ILE181 - LEU296 and is composed of four structurally conserved helices (helices 6-9).

Proteins belonging to the erythrocyte binding-like (EBL) super family are known to play key roles in the complex series of interactions required for merozoite invasion of erythrocytes. According to the study done by Chitnis and Chaudhuri (1996), the interaction of *P. knowlesi* ligand with human and rhesus erythrocytes appears to be mediated by a peptide-peptide interaction. The finding shows that *P. knowlesi* requires the Duffy antigen to form a tight junction and to invade human erythrocytes suggested that this two sub domains forms outwards profile like structure which used to locate and binds to duffy antigen on erythrocyte cells, which aided by peptide-peptide interaction. Its amino acids hydrophobic residues are located at the inner side of the loop, while its hydrophilic residues point outwards suggesting the entire protein is projected outside of *P. knowlesi*.

Structure Validation

Before the predicted structure could be used for further analyses, it has to be validated to evaluate the quality of the structure. The validation was done using ProQ-Protein Quality Predictor and Ramachandran Plot available at http://mordred.bioc.cam.ac.uk/~rapper/rampage.php. This validation result from ProQ-Protein Quality Predictor is shown in Figure 2. The predicted LGscore for the Swiss Model is 2.904 and the predicted MaxSub is 0.271. The predicted LGscore lies in the range of an every good model of predicted 3D protein structure, while their predicted MaxSub are a highly acceptable score which lies in the range of a good model.

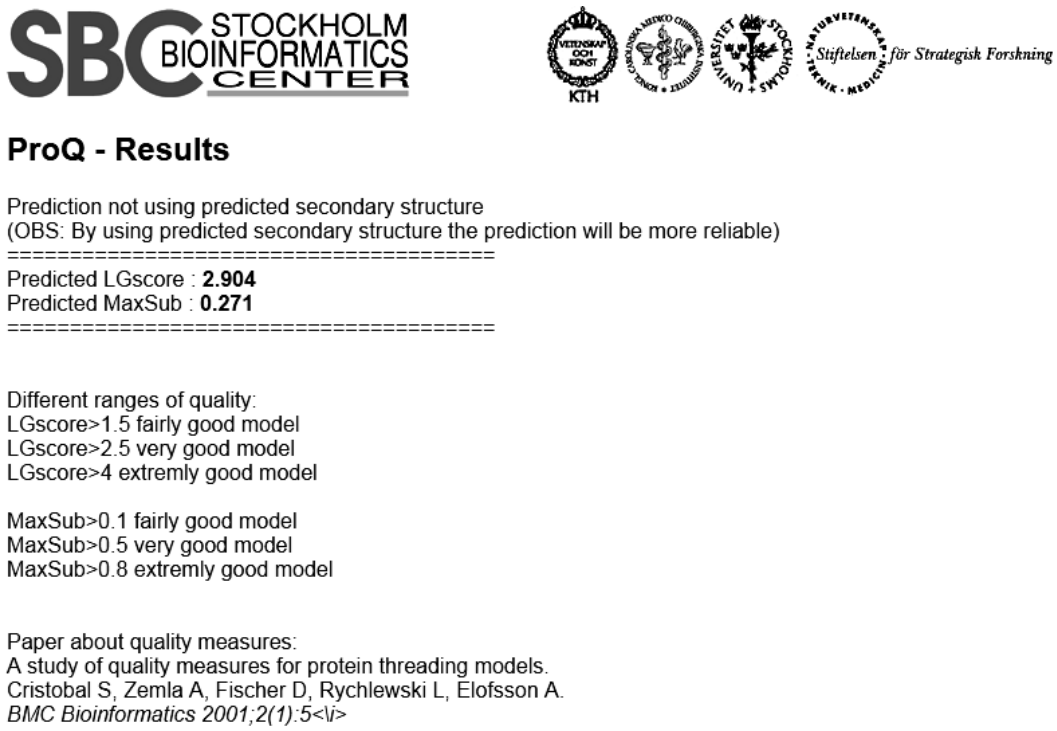


Figure 2: The Evaluation from ProQ-Protein Quality Predictor for Swiss Model Server Protein Structure, Showing Both LGscore and MaxSub Values Lies in Acceptable Range of Score

Validation result using Ramachandran plot is also promising as it produced good result with 95.6% residues in favored region and 1.3% of residue in outlier region. Results obtained from Ramachandran plot are also promising as most of the residues clustered tightly in the most favored regions with very few or none outliers.

Predicting Function from the Structure

The validated structure model is further analysed using Profunc. ProFunc is developed to identify the likely biochemical function of a protein from its 3D structure. It uses a series of methods, including fold matching, residue conservation, surface cleft analysis, and functional 3D templates, to identify both the protein's likely active site and possible homologues in the PDB. The result of the analysis shows the function of the predicted structure is similar to the function of Duffy antigen receptor in which ProFunc has successfully found two motifs. Motif in proteins is conjectured to have biological significance.

Conclusion

This study focused on modeling the 3D protein structure of *Plasmodium Knowlesi's* Duffy antigen receptor protein using homology modeling method and to determine its function using bioinformatics resources. The 3D protein structure built has resulted with LGscore of 2.904, yielding to an extremely

good model according to ProQ validation and Ramachandran plot criteria. Structure analysis using ProFunc found two motifs which indicate the similar function of this protein with other Duffy antigen proteins.

Therefore, this study has successfully produced a good basis for the structural study for the target protein. The generated model could later be used in further structural study such as in the determination of the active sites for molecular docking.

REFERENCES

Agostino M (2013). Bioinformatics Tools for the Laboratory. In: *Practical Bioinformatics*, (Garland Science, New York, USA) 172.

Apweiler R, Bairoch A, Wu H, Barker W and Boeckmann B (2004). UniProt: The Universal Protein knowledge base. *Nucleic Acids Research* **32** 115-119.

Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S and Schneider M (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research* **31**(1) 365–370.

Chitnis CE, Chaudhuri A, Horuk R, Pogo O and Miller LH (1996). The Domain on the Duffy Blood Group Antigen for Binding Plasmodium vivax and P. knowlesi Malarial Parasites to Erythrocytes. *The Journal of Experimental Medicine* **8** 1531-1536.

Edwards-Yvonne JK and Cottage A (2003). Bioinformatics methods to predict protein structure and function. *Molecular Biotechnology* **23**(2) 139-166.

Fong MY, Lau YL, Chang PY and Anthony CN (2014). Genetic diversity, haplotypes and allele groups of Duffy binding protein (PkDBPαII) of *Plasmodium knowlesi* clinical isolates from Peninsular Malaysia. *Parasites & Vectors* **7** 161.

Goh XT, Lim YAL, Vythilingam I, Chew CH, Lee, PC, Ngui R, Tan TC, Yap NJ, Nissapatorn V and Chua KH (2014). Increased detection of Plasmodium knowlesi in Sandakan division, Sabah as revealed by PlasmoNex™. *Malaria Journal* **12** 264.

Gaur D, Mayer DC and Miller LH (2004). Parasite ligand-host receptor interactions during invasion of erythrocytes by *Plasmodium merozoites*. *International Journal for Parasitology* **34** 1413–1429.

Hollingsworth S and Karplus P (2011). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomolecular Concepts* **1** 271-283.

Lee KS, Cox-Singh J and Singh B (2009). Morphological features and differential counts of Plasmodium knowlesi parasites in naturally acquired human infections. *Malaria Journal* **8** 73.

Miller L, Mason S, Dvorak J, Mcginniss M and Rothman I (1975). Erythrocyte receptors for (Plasmodium knowlesi) malaria: Duffy blood group determinants. *Science* **1** 561-563.

Nayeem A, Sitkoff D and Stanley Jr K (2006). A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. *Protein Science* **15**(4) 808-824.

Singh B, Sung L, Matusop A, Radhakrishnan A, Shamsul S, Cox-Singh J, Conway D (2004). A Large Focus of Naturally Acquired Plasmodium Knowlesi Infections in Human Beings. *The Lancet* **363** 1017-1024.

Wallner B and Elofsson A (2003). Can correct protein models be identified? *Protein Science* **12**(5) 1073-1086.

White NJ (2008). Plasmodium knowlesi: the fifth human malaria parasite. *Clinical Infectious Diseases* **46** 172–173.