

---

## Docking Studies of Serine Protease Inhibitor with Various SPI Activators to Treat COPD

---

<sup>1</sup>Chlumalla Srikanth, <sup>2</sup>Rahul Bharadwaj

<sup>1,2</sup>A.V Post Graduate Centre, Hyderabad

Email: [shreekanthchelumalla@gmail.com](mailto:shreekanthchelumalla@gmail.com)

Received: 10<sup>th</sup> April 2017, Accepted: 19<sup>th</sup> April 2017 Published: 1<sup>st</sup> May 2017

### Abstract

The SERPINA1 gene provides instructions for making a protein called alpha-1 antitrypsin, which is a type of serine protease inhibitor (serpin). Alpha-1 antitrypsin protects the lungs from neutrophil elastase, which can damage lung tissue if not properly controlled. The current work involves the complete annotation of SERPINA1 followed by analysing its role in COPD disease. The work can be extended in designing a potential ligand (drug) that can exhibit a good docking energy and can be used in the therapy of the same. By using different Bioinformatics Tools Structural & functional analysis of alpha-1 antitrypsin is done. Serpins (serine proteinase inhibitors) belong to MEROPS inhibitor family I4, clan ID. Several softwares and Databases have been used in the work. NCBI, Gene Cards, PDB, Pubchem were the databases used for collecting the protein Sequence, Functional Information, Structure and chemical compounds respectively. Certain tools like Dundee ProDrg Server, Protparam, CASTP etc are used for data analysis. The docking was performed using Hex software. Based on the energy values exhibited, the best ligand selected was Gabexate mesylate with its docking energy of -264.27.

### Keywords

SERPINA1, Antitrypsin, Elastase, COPD, MEROPS inhibitor family, Protparam, CASTP

### Introduction

COPD, or chronic obstructive pulmonary (PULL-mun-ary) disease, is a progressive disease that makes it hard to breathe. "Progressive" means the disease gets worse over time. COPD can cause coughing that produces large amounts of mucus (a slimy substance), wheezing, shortness of breath, chest tightness, and other symptoms. Cigarette smoking is the leading cause of COPD. Most people who have COPD smoke or used to smoke. Long-term exposure to other lung irritants—such as air pollution, chemical fumes, or dust—also may contribute to COPD. The airways and air sacs are elastic (stretchy). When you breathe in, each air sac fills up with air like a small balloon. When you breathe out, the air sacs deflate and the air goes out. In COPD, less air flows in and out of the airways because of one or more of the following: The airways

and air sacs lose their elastic quality. The walls between many of the air sacs are destroyed. The walls of the airways become thick and inflamed. The airways make more mucus than usual, which can clog them.

In the United States, the term "COPD" includes two main conditions—emphysema (em-fih-SE-ma) and chronic bronchitis (bron-KI-tis). (Note: The Health Topics article about bronchitis discusses both acute and chronic bronchitis.)

In emphysema, the walls between many of the air sacs are damaged. As a result, the air sacs lose their shape and become floppy. This damage also can destroy the walls of the air sacs, leading to fewer and larger air sacs instead of many tiny ones. If this happens, the amount of gas exchange in the lungs is reduced.

The SERPINA1 gene provides instructions for making a protein called alpha-1 antitrypsin, which is a type of serine protease inhibitor (serpin). Serpins help control several types of chemical reactions by blocking (inhibiting) the activity of certain enzymes. Alpha-1 antitrypsin prevents the digestive enzyme trypsin from breaking down proteins until trypsin reaches the intestines. Alpha-1 antitrypsin also inhibits other enzymes, including a powerful enzyme called neutrophil elastase that is released from white blood cells to fight infection. Alpha-1 antitrypsin protects the lungs from neutrophil elastase, which can damage lung tissue if not properly controlled. Alpha-1 antitrypsin is produced in the liver and then transported to the lungs via the blood.

More than 120 mutations in the SERPINA1 gene have been identified. Some of these mutations do not affect the production of alpha-1 antitrypsin, while others cause a shortage (deficiency) of the protein. Without enough functional alpha-1 antitrypsin, neutrophil elastase destroys the small air sacs in the lungs (alveoli) and causes lung disease. Excessive damage to the alveoli leads to emphysema, an irreversible lung disease that causes extreme shortness of breath.

## **Materials and Methods**

### **Identification of protein cause in the disease:**

#### **GeneCards®: The Human Gene Database**

GeneCards is a searchable, integrative database that provides comprehensive, user-friendly information on all annotated and predicted human genes. It contains genomic, transcriptomic, proteomic, genetic, clinical and functional information. The gene related to a disease can be easily identified by its data search engine. The tool provides a list of genes involved in the onset of a disease along with the other information like location, function, mutations structure etc.

#### **Protein Sequence Retrieval from NCBI:**

NCBI is the database which provides the Molecular and Biotechnology information. NCBI (National Center for Biotechnology Information) is an archive of information related to proteins and genes with respect to its medical applications. It gives complete information of gene, sequences and protein structures. By using the NCBI data base we can retrieve the protein and gene sequences. It also contains information about various journals publications, books, SNP etc.

#### **FUNCTIONAL ANALYSIS:**

The functional Analysis of the protein is done by using following tools, they are.

**SMART** (a Simple Modular Architecture Research Tool) allows the identification and annotation of genetically mobile domains and the analysis of domain architectures. Domain indicates a restricted, independent functional unit within the protein sequence. Many domains are 'genetically mobile', meaning that they can be found associated with different domain combinations in different proteins. The information provided by SMART includes mainly the number and the name of the domains present for a specific protein and the function of each domain are mentioned.

#### **Protein Interaction Study:**

##### **String EMBL:**

STRING is a database which provides information about the protein-protein interactions. The interactions are two types' direct and indirect associations. Here a compact network chart of the study protein is provided showing the total number of direct and indirectly interacting proteins with their function and importance. This tool is mainly used to identify the important proteins linked to the study protein and their functions.

#### **STRUCTURAL ANALYSIS:**

The structure Analysis of the can be performed using the following tool: **Protparam** from Primary characterization of protein, **SOPMA** for secondary structure confirmation and **Phyre** for tertiary structure prediction.

##### **Protparam:**

A simple tool running on a theoretical calculations and algorithms providing maximum information about the protein with only the sequence input. Basic physico-chemical characterization of the protein is performed by the tool providing ample information which can be used for designing the experiments for the lab based analysis. Information like protein sequence length, Molecular Weight, stability, Polar nature, pH, Isoelectric Point etc can be analyzed by the tool.

##### **SOPMA for Secondary Structure Prediction**

SOPMA is a secondary structure prediction tool used to characterize the secondary structural confirmations of the protein that include helix, coil, sheet, turn, extended strand etc. This tool would be useful to analyse the total helical regions present in the protein structure and the graphical representation of each kind of structure. The tool can be obtained at

[https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_sopma.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)

##### **PHYRE2**

This is a tool linked to the PDB database of the protein structures. It would provide the pdb id for the protein whose 3D structure can be retrieved from the RCSB Protein Databank. Once the sequence is entered into the tool it would perform an internal BLAST with the PDB to provide all the similar PDB Ids for each hit performed. The best structure is the one which possesses the closest sequence to the query. Phyre is available at the link <http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>

##### **Rasmol Visualization**

Rasmol is a visualization software that is used to visualize the 3D structure of the protein and identify the various structural elements in the protein like the secondary confirmation, embedded water molecules, presence of ligands or co factors, polar and Polar Regions etc. However the software is used to study the existing structure but manipulation of the structure is not available. Rasmol can be freely downloaded from the online server.

##### **Binding site prediction using CASTP**

Pocket or binding sites are the important regions on the protein used for docking with the external ligands.

These sites are detected using the CASTP server. The tool would provide the total no of pockets along with the residues of amino acids involved in making of the pocket. The pockets are also sorted as per the volume and area. The tool can be obtained at <http://sts.bioe.uic.edu/castp/>

### Drug Designing

The process of developing or identifying a suitable drug candidate for a particular disease or receptor requires a tedious analysis and a step by step protocol making it a rational drug designing protocol. Some of the major steps include Chemical library creation, Screening for Drug likeliness, Ambiguity checking and finally docking. Chemical library can be created using suitable chemicals as per their function required from the pubchem database, for whom the Lipinski parameters of Molecular weight (300 to 600), H bond donor (1-5), H bond receptors (>5) Log p value (1-5) are all recorded in the chemical library. These parameters would make a chemical either drug or a non-drug. Once all the chemicals possessing drug likeliness are selected they are subjected for a screening called ambiguity check performed using Dundee Prodrgr server. This is available online and would detect the standard nature of each bond, bong angle, length etc., making the chemical non ambiguous or standard. All the selected chemicals are downloaded and are used for docking with the identified receptor of the disease.

The docking can be performed using any of the softwares like MolDock, Sybyl, Argus lab etc., here Hex docking has been performed.

### Results and Discussion

#### Protein Sequence Retrieval:

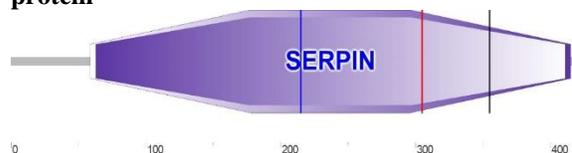
The Protein sequence of alpha 1 Anti-Trypsin Inhibitor is collected from the NCBI, Accession No: NP\_001121179, It is Linear structure containing 418 AA, it is present on the chromosome 14 (map="14q32.13"), the signal peptide sequence start from 1 – 24 AA, and molecular weight is 2430. Mat peptide sequence start from 25 – 418 AA, product is alpha-1-antitrypsin and calculated molecular weight is 44325. Region 55 – 412 this family contains a variety of different members of clade A of the serpin superfamily. They include the classical serine proteinase inhibitors, alpha-1-antitrypsin and alpha-1-antichymotrypsin, protein C inhibitor. The protein sequence of alpha 1 anti-trypsin inhibitor. Represented id FASTA format is shown below which is used for the complete analysis

```
>NP_001121179.1 alpha-1-antitrypsin precursor
[Homo sapiens]
```

```
MPSSVSWGILLLAGLCLVPVSLAEDPQGDA
KQTDTSHHDDQDHPFNKITPNLAEFASLYR
QLAHQSNSTNIFFSPVSIATAFAMLSLGT
KADTHDEILEGLNFNLTEIPEAQIHEGF
QELLRTLNPQDSQLQLTTGNGLFLSEGL
KLVDFKLEDVKKLYHSEAFVNFVGDTEE
AKKQINDYVEKGTQGKIVDLVKELDRDT
VFALVNYIFFKGGKWERPFVEVKDTE
EEDFHVDQVTTVKVPMMKRLGMFNIQHCK
KLLSSWVLLMKYLGNATAIFFLPDEGK
LQHLLENELTHDIITKFLNEDRRSASL
HLPKLSITGTYYDLKSVLGQLGITKVF
SNGADLSGVTEEAPLKLKSKAVHKA
VLTIDEKGTAAAGAMFLEAIPMSIPPE
VKFNKPFVFLMIEQNTKSPLFMGKVVN
PTQK
```

### SMART analysis

Fig 1 SMART showing the SERPIN domain in the protein

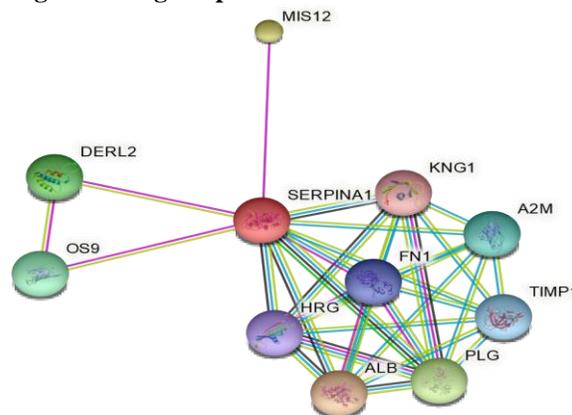


The above result of SMART shows that there is only one domain in the protein ie., SERPIN. This belong to MEROPS inhibitor family I4. This domain extends from 59 to 415 on the sequence.

### STRING for protein network study

STRING has generated the following network chart for the protein showing several other interacting proteins.

Fig 2 showing the protein network

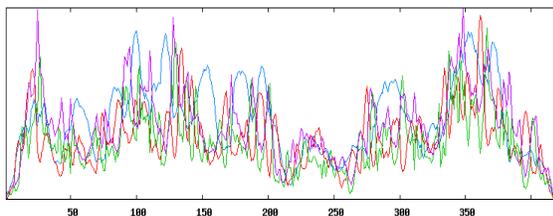


From the above result of STRING it can be inferred that the there are several proteins which would interact with the study protein at various molecular levels. Some of the interacting proteins or genes include KNG1, A2M, TIM1 PLG, ALB etc.

**Protparam for physicochemical characterization of protein:**

The result of PROTPARAM depicts that the sequence has the length of 418AA, Its PI value is 5.37, it is an acidic protein with its aliphatic index 91.20, the protein is a polar molecule with the hydrophaticity index to be -0.183. The protein is acidic with more number of negatively charged residues.

**Fig 3 SOPMA result showing the secondary structural confirmations**



The picture indicates the percentage of each type of secondary structure present in the sequence. From the above result of SOPMA it can be inferred that the secondary structure of the sequence contain a majority of alpha helical confirmation constituting 42% of the protein structure followed by random coil 32%, 15% extended strand and 10% beta turn.

**Phyre Analysis for the PDB id of the protein**

To retrieve the PDB id of the protein so as to download the 3D structure from the RCSB Protein Data Bank Phyre analysis is performed. As per the result of Phyre best pdb id selected as the structure representative for the above protein is 1QLP.

**Rasmol Visualization**

The 3D structure of the protein is visualized using RASMOL software. The command line programme is used to control the structure and instruct the software.

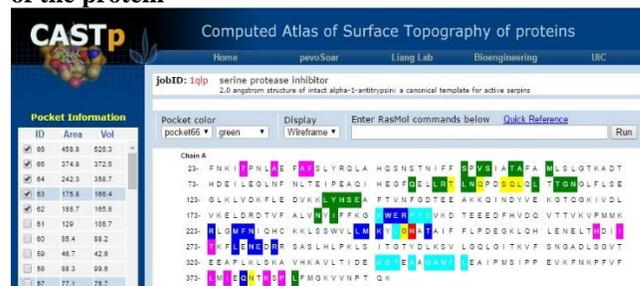
**Fig 4 showing the 3D structure of the protein as shown by RASMOL**



**Binding site Analysis using CASTP**

The binding sites or pockets present within the 3D structure are identified using CASTP. A total number of 66 pockets have been identified which can be targeted for docking. The pockets are shown in the figure 5.

**Fig 5: Showing the binding sites on the 3D structure of the protein**



From the above result of CASTP it can be inferred that that a total of 66 pockets have been identified in the selected protein structure. The pocket details are further furnished in the above results.

**Chemical Analysis:**

**Compound library generation from Pubchem**

In the Pubchem database a total of 106 chemicals were detected possessing the required function. However only 68 chemicals were found to be qualifying the Lipinsky rule of Druglikeness. Thus only these 68 were further analyzed. Further these chemicals are subjected for secondary screening for ambiguity.

**Ambiguity Check using Dundee prodrgr server:**

Dundee prodrgr server detects the defects in the molecular structures with respect to Bond angle, Partial Charges, Dihedrals etc. The compound in its standard state would have least ambiguity and the one with more defects is known to have high ambiguity.

**Fig 6: Ambiguity details for the compound Gabexate mesylate**

```

PRODRG> Starting up PRODRG version AA100323.0717
PRODRG> Parameter set 'pd/gromos96' (fftype=2).
PRODRG> MOL mode detected.
PRODRG> WARNING: deleted hydrogen(s) from your input.
PRODRG> WARNING: No stereo information found in input file.
PRODRG> WARNING: more than one fragment.
PRODRG> Molecule complexity index: 2.00.
PRODRG> 24 hydrogen(s) added.
PRODRG> Using charge groups.
PRODRG> Net charge on molecule: 0.000
PRODRG> 31 partial charges 3 ambiguous
PRODRG> 47 bonds 7 ambiguous
PRODRG> 80 bond angles 25 ambiguous
PRODRG> 25 improper dihedrals 2 ambiguous
PRODRG> 15 dihedrals 0 ambiguous
PRODRG> Writing GROMACS topology.
PRODRG> GROMACS topology quality on 0-10 scale: 7.3
GENDRG> Best structure was iteration 51 with 9.90484142
PRODRG> Spawning GROMACS version 3.3.3...
PRODRG> RMSD from GROMOS bond ideality (Angstrom) : 0.009
PRODRG> RMSD from GROMOS angle ideality (degrees) : 2.905
PRODRG> RMSD from GROMOS plane ideality (degrees) : 1.998
PRODRG> Number of improper improper dihedrals : 0
PRODRG> Writing: SCRTHOWMMPG
PRODRG> Normal program end.
    
```

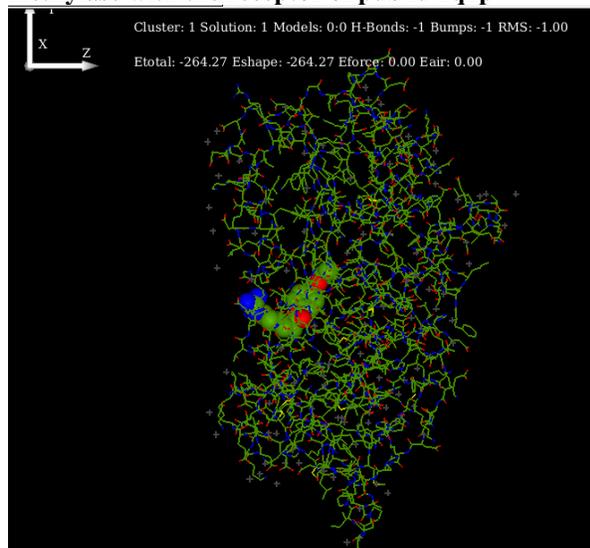
The above figure shows the ambiguity details for the compound **Gabexate mesylate**. The similar analysis has been performed for several other chemicals and the one showing the lowest ambiguity is further selected for docking. Followed by the ambiguity check using Dundee prodrug server 4 chemicals have been selected as lead molecules from a total list of 68.

### Docking Analysis

Docking of the above receptor with 4 final ligands is performed using HEX 8.0 Software and the results are depicted below.

The Hex results indicate that the docking value of Gabexate mesylate is  $E_{total} = -264.27$ , Nco 650; OR348951 is  $E_{total} = -298.30$ , trans-4-Guanidinomethylcyclohexanecarboxylic acid p-tert-butylphenyl ester hydrochloride is  $E_{total} = -278.66$ , Batebulast is  $E_{total} = -272.62$  out of which the best ligand selected was Gabexate mesylate due to its lowest docking energy.

**Fig 7: Showing the docking result of Gabexate methylase with the receptor of pdb id 1q1p**



The above picture shows the docking between the selected receptor 1q1p with Gabexate methylase showing a docking energy of  $-264.27$  making a stable complex.

### Conclusion

In the above analysis the protein sequence of SERPINA1 was collected from NCBI and the functional study was undertaken in genecards. The protein sequence was used for the Sequence based analysis which includes Protparam for physicochemical characterization, SMART for Domain identification, String for network analysis etc. The protein structural analysis was performed using

SOPMA and Phyre followed by its 3D visualization using Rasmol. The total binding sites present in the protein were identified using CASTP tool. The study was directed for Drug development using the regular CADD approach. Activators of serine protease inhibitor (serpin) are collected from the PubChem database, out of 108 Activators obtained only 68 activators are selected for analysis which followed lipinsky rule as they possess the drug likeness. These chemicals were further screened for Ambiguity at the secondary screening level. For this Dundee Prodrug server has been used. The structures were downloaded from the same data base. A docking study was performed between the selected list of ligands and the one receptor SERPINA1 using HEX 8.0 Software. The final ligands with lowest binding energy was found to be Gabexate methylate which was selected as the best drug to target the above receptor. It can be used as a therapeutic drug to treat COPD.

### References

1. Lovasi GS, et al. (2010). Association of environmental tobacco smoke exposure in childhood with early emphysema in adulthood among nonsmokers. *American Journal of Epidemiology*, 171(1): 54-62.
2. U.S. Preventive Services Task Force (2008). Screening for chronic obstructive pulmonary disease using spirometry: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine*, 148(7): 529-534.
3. Qaseem A, et al. (2011). Diagnosis and management of stable chronic obstructive pulmonary disease: A clinical practice guideline update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. *Annals of Internal Medicine*, 155(3): 179-191.
4. Criner GJ, Sternberg AL (2008). A clinician's guide to the use of lung volume reduction surgery. *Proceedings of the American Thoracic Society*, 5(4): 461-467.
5. Global Initiative for Chronic Obstructive Lung Disease (2014). Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. <http://www.goldcopd.org/guidelines-global-strategy-for-diagnosis-management.html>. Accessed May 21, 2014.
6. King DA, et al. (2008). Nutritional aspects of chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society*, 5(4): 519-523.

7. Qaseem A, et al. (2011). Diagnosis and management of stable chronic obstructive pulmonary disease: A clinical practice guideline update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. *Annals of Internal Medicine*, 155(3): 179-191.
8. Soriano, JB, Maier, WC, Egger, P et al. Recent trends in physician diagnosed COPD in women and men in the UK. *Thorax*. 2000; 55: 789-794.
9. Lung and Asthma Information Agency. Trends in emergency hospital admissions for lung disease. Fact sheet 2001/4. London: St George's Hospital Medical School.
10. Jones, PW, Quirk, FH, Baveystock, CM, and Littlejohns, P. A self-complete measure of health status for chronic airflow limitation: the St George's respiratory questionnaire. *Am Rev Respir Dis*. 1992; 145: 1321-1327.
11. McSweeney, AJ, Grant, I, Heaton, RK, Adams, KM, and Timms, RM. Life quality of patients with chronic obstructive pulmonary disease. *Arch Intern Med*. 1982; 142: 473-478.
12. Ketelaars, CA, Schlosser, MA, Mostert, R et al. Determinants of health-related quality of life in patients with chronic obstructive pulmonary disease. *Thorax*. 1996; 51: 39-43.
13. Hajiro, T, Nishimura, K, Tsukino, M, Ikeda, A, Koyama, H, and Izumi, T. Comparison of discriminative properties among disease-specific questionnaires for measuring health-related quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 1998; 157: 785-790
14. Morgan, MDL, Threlfall, E, Sanders, D, Walden, IM, and Britton, M. Living with COPD: results of the British Lung Foundation survey. European Respiratory Society Congress, Berlin; Sept 22-26, 2001(abstr 3266).
15. Rennard, S, Decramer, M, Calverley, PMA et al. The impact of COPD in North America and Europe in 2000: the patients perspective of the Confronting COPD International Survey. *Eur Respir J*. 2002;20: 1-7.
16. American Thoracic Society. Standards for the diagnosis of patients with chronic obstructive pulmonary disease (COPD) and asthma. *AmRev Respir Dis*. 1987; 136: 225-244.
17. Canadian Thoracic Society Working Group. Guidelines for the assessment and management of chronic obstructive pulmonary disease. *Can Med Assoc J*. 1992; 147: 420-428.
18. European Respiratory Society Task Force. Optimal assessment and management of chronic obstructive pulmonary disease. *Eur Respir J*. 1995; 8: 1398-1420.
19. European Respiratory Society Task Force. Optimal assessment and management of chronic obstructive pulmonary disease. *Eur Respir J*. 1995; 8: 1398-1420.
20. Pauwels, RA, Buist, AS, Calverley, PMA, Jenkins, CR, and Hurd, SS. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop Summary. *Am J Respir Crit Care Med*. 2001; 163: 1256-1276.
21. Fletcher, C and Peto, R. The natural history of chronic airflow obstruction. *BMJ*. 1977; 1: 1645-1648.
22. Peto, R, Speizer, FE, Cochrane, AL et al. The relevance in adults of air-flow obstruction, but not of mucus hypersecretion, to mortality from chronic lung disease: results from 20 years of prospective observation. *Am Rev Respir Dis*. 1983; 128: 491-500.
23. Pandey, MR, Regmi, HN, Neupane, RP, Gautam, A, and Bhandari, DP. Domestic smoke pollution and respiratory function in rural Nepal. *Tokai J Exp Clin Med*. 1985; 10: 471-481.
24. Pandey, MR. Prevalence of chronic bronchitis in a rural community of the hill region of Nepal. *Thorax*. 1984; 39: 331-336.
25. Perez-Padilla, R, Regalado, J, Vedal, S, Pare, P, Sansores, R, and Seman, M. Exposure to biomass smoke and chronic airway disease in Mexican women: a case-control study. *Am J Respir Crit Care Med*. 1996; 154: 701-706.
26. Cozio, M, Ghezzi, H, Hogg, JC et al. The relations between structural changes in small airways and pulmonary function tests. *N Engl J Med*. 1978; 298: 1277-1281.
27. Thurlbeck, WM and Wright, JL. Chronic airflow obstruction. 2nd edn. BC Dekker, Hamilton; 1999.