

Candidate genes controlling pulmonary function in mice: transcript profiling and predicted protein structure

Character Count: 108 characters

Koustav Ganguly^{1*}, Tobias Stoeger^{1*}, Scott C. Wesselkamper^{2*}, Claudia Reinhard¹, Maureen A. Sartor², Mario Medvedovic², Craig R. Tomlinson², Ines Bolle¹, John M. Mason², George D Leikauf^{2,3}, Holger Schulz¹

¹GSF-National Research Center for Environment and Health, Institute of Inhalation Biology, Neuherberg, Germany

²Center for Environmental Genetics, Department of Environmental Health, University of Cincinnati, OH, USA and ³Department of Environmental and Occupational Health, University of Pittsburgh, PA, USA.

* Contributed equally to this work.

Corresponding Author:

Prof. Dr. med. Holger Schulz
GSF-Institute of Inhalation Biology
Ingolstaedter Landstrasse. 1
D-85764 Neuherberg/Munich
Tel.: +49 89 3187 4119
Fax.: +49 89 3187 2400
e-mail: schulz@gsf.de
<http://www.gsf.de/ihb/>

Running Heading: Genetics of Lung Function in Mice

This article has a data supplement.

Research Funding Sources: NGFN2 (1GR0430); NIH (HL070542, ES010562, ES006096, HL065612, HL077763, ES015675).

Abstract:

Rationale: Impaired development and reduced lung capacity are risk factors of asthma and chronic obstructive pulmonary disease (COPD). Previously, our genome-wide linkage analysis of C3H/HeJ (C3H) and JF1/Msf (JF1) mouse strains identified quantitative trait loci (QTL) associated with the complex traits of dead space volume (V_D), total lung capacity (TLC), lung compliance (C_L), and diffusing capacity (D_{CO}). **Method/Findings:** We assessed positional candidate genes by comparing C3H with JF1 lung transcript levels using microarray and by comparing C3H, BALB/cByJ, C57BL/6J, A/J, PWD/PhJ, and JF1 using exon sequencing to predict protein structure. Microarray identified >900 transcripts differing in C3H and JF1 lungs related to lung development, function, and remodeling. Of these, 3 genes localized to QTLs associated with differences in lung function. C3H and JF1 strains differed in transcript and protein levels of superoxide dismutase 3, extracellular (SOD3; mCh5: V_D) and transcript of trefoil factor 2 (TFF2; mCh17: TLC& D_{CO}), and ectonucleotide pyrophosphatase/phosphodiesterase 2 (ENPP2; mCh15: TLC& C_L). Nucleotide sequencing of *Sod3*, *Tff2*, and previously identified relaxin 1 (*Rln1*; mCh19: C_L) uncovered polymorphisms that could lead to nonsynonymous amino acid changes and altered predicted protein structure. Gene-targeted *Sod3*^(-/-) mice had increased conducting airway volume (V_D /TLC) compared to strain-matched control *Sod3*^(+/+) mice, consistent with the QTL on mCh 5. **Conclusions:** Two novel genes (*Tff2* and *Enpp2*) have been identified and two suspected genes (*Sod3* and *Rln1*) have been supported as determinants of lung function in mice. Findings with gene-targeted mice suggest that SOD3 is a contributing factor defining the complex trait of conducting airway volume.

Word Count: 250 Words **Keywords:** lung function; genetics; complex trait; COPD; asthma, *Sod3*.

Abbreviations: **BALB:** BALB/cByJ, **BW:** Body weight, **C3H:** C3H/HeJ, **C57:** C57BL/6J, **C_L:** static lung compliance, **COPD:** Chronic obstructive pulmonary disease, **D_{co}:** diffusing capacity for carbon monoxide, **ENPP2:** ectonucleotide pyrophosphatase/phosphodiesterase 2, **FDR:** false discovery ratio, **FEV₁:** forced expiratory volume-1 second, **FRC:** Functional Reserve Capacity, **FVC:** Forced vital capacity, **GenMapp:** Gene Microarray Pathway Profiler, **hCh:** human chromosome, **JF1:** JF1/Msf, **KDa:** Kilodalton, **LOD:** logarithm of the odds, **LPA:** lysophosphatidic acid, **MappFinder:** Microarray Pathway Profile Finder, **mCh:** mouse chromosome, **NCBI:** National Center for Biological Information, **Pao:** airway opening pressure, **PDGFB:** platelet derived growth factor, B polypeptide, **PWD:** PWD/PhJ, **qRT-PCR:** quantitative real time-polymerase chain reaction, **QTL:** quantitative trait loci, **RLN1:** relaxin1, **RXRβ:** retinoid X receptor beta, **SNP:** single nucleotide polymorphism, **SOD3:** superoxide dismutase 3, extracellular, **TFF2:** trefoil factor 2, **TGFBR3:** transforming growth factor β receptor 3, **TLC:** total lung capacity, **5'-UTR:** 5'-untranslated region, **V_A:** alveolar volume, **V_D:** dead space volume.

Introduction

Increasing scientific interest has been placed on gene-environment interactions that determine human responses to cigarette smoke, metals, and ambient air pollution. Constant improvements in genetic analyses have enabled the acquisition of genome-wide information on numerous species and the rapid determination of phenotype-genotype linkages that continue to provide a deeper understanding of common human diseases (i.e., complex traits). Respiratory diseases like chronic obstructive pulmonary disease (COPD) are characterized by an accelerated decline in lung function and have been associated with numerous environmental risk factors, especially cigarette smoking. Although smoking is declining in some countries, COPD continues to rise worldwide and is now the fourth leading cause of death. A striking reality of smoking is that the mass of irritants and particulate matter deposited in a 2-pack per day smoker's lung over lifetime is staggering. Yet not everyone who smokes develops severe COPD, lung cancer, or other smoking-related diseases, which strongly supports a role for individual genetic susceptibility and resistance. For example, one study estimated that only 15% of the variation in lung function in smokers is conferred by the degree of smoking exposure (a primary environmental factor) [4] while another study estimated that over 40% of the variation is conferred by heredity (genetic factors) [9].

Impaired embryonic and neonatal lung development may contribute to adolescence and adult lung diseases including asthma [45] and COPD [41]. Shortened gestation is a proposed risk factor for adolescent asthma because it results in diminished airway diameter, and thus increased hyperreactivity. It has also been proposed that persons with a smaller initial lung volume are more likely to develop overt manifestations during the accelerated decline of lung function in COPD. Therefore the genetic factors that limit lung development can be viewed as risk factors for later appearance of clinically significant disease. Genome-wide association studies examining differences in forced expiratory volume-1 second (FEV₁) and forced vital capacity (FVC) and the ratio FEV₁/FVC have identified linkage to loci on 13

different human chromosomes (hCh) with replicated linkage reported on hCh 2, 5, and 6 [21, 42, 51-53].

In inbred mice, considerable inter-strain differences in lung function and lung morphometrics have been reported [43, 46]. Previously, we measured lung function in C3H/HeJ (C3H) BALB/cByJ (BALB), C57BL/6J (C57), A/J and JF1/Msf (JF1) mouse strains and found C3H and JF1 differed the greatest [34, 35]. For example, the total lung capacity (TLC, male and female) is 1640 ± 54 μ l in C3H mice, whereas TLC is 890 ± 15 μ l in JF1 mice. Accordingly, we completed a quantitative trait loci (QTL) analysis of multiple lung phenotypes in offspring derived from the C3H (*Mus musculus domesticus*) and JF1 (*Mus musculus molossinus*) strains [35]. Using female N₂ backcross and F₂ intercross progeny genome-wide linkage analysis was performed for TLC, dead space volume (V_D), lung compliance (C_L), and diffusing capacity for carbon monoxide (D_{CO}). Significant linkage was detected for TLC on mouse chromosome (mCh) 15 and 17 [logarithm of the odds (LOD) $\geq$ 5.0], V_D on mCh 5 and 15 (LOD $\geq$ 4.0), C_L on mCh 15 and 19 (LOD $\geq$ 5.0), and D_{CO} on mCh 15 and 17 (LOD $\geq$ 4.0)

After evaluating the possible physiological roles of genes harbored in these regions, we proposed possible candidate genes including superoxide dismutase 3, extracellular (*Sod3* a.k.a. extracellular SOD) on mCh 5. This region is syntenic to a region on hCh 4 with suggestive linkage to FEV₁/FVC ratio identified in the National Heart, Lung, and Blood Institute Family Heart Study (NHLBI FHS) [51]. We also proposed Relaxin 1 (*Rln1*) to be a candidate for the QTL on mCh 19. To further explore these linkages, we performed a microarray analysis contrasting C3H with JF1 strain lung transcript levels with confirmatory quantitative real-time polymerase chain reaction (qRT-PCR) and Western blot analyses. This was followed by exon sequencing of the candidate genes to predict protein structural modifications among 6 mouse strains, the 5 listed above and PWD/PhJ (PWD; *Mus musculus musculus*), a strain with small lung size (TLC 985 ± 19 μ l) similar to JF1. For functional analysis gene targeted *Sod3*^(-/-) mice were analyzed for lung function and were

found to have higher conducting airway volume (V_D/TLC), consistent with our QTL on mCh5 [35].

Methods

Mice: All mice were housed in animal care facility of the GSF-Institute of Inhalation Biology, Neuherberg/Munich, Germany under specific pathogen free conditions. Breeding pairs of PWD/PhJ mice were obtained from the Jackson Laboratory (Bar Harbour, ME, USA) and gene-targeted *Sod3*^(-/-) and wildtype control *Sod3*^(+/+) mice were kindly provided by Dr. Tim D Oury (University of Pittsburgh, PA, USA). Both wildtype and gene-targeted *Sod3* mice were derived from 129/Sv (sub-strain not identified) and C57BL/6 mouse strains [5]. The use of animals was in accordance with the German Law of Animal Protection and approved by the Bavarian Animal Research Authority and all procedures were also approved by the IACUC of the University of Cincinnati and University of Pittsburgh.

Lung function measurements of PWD and *Sod3*^(-/-) mice: Lung function measurements and statistics were performed as described previously [34, 35]. Briefly, anesthetized mice were intubated and ventilated using a computer-controlled piston-type servo ventilator and concentrations of the respiratory and the test gases (He and C¹⁸O) were measured by a magnetic sector field mass spectrometer. The volume, airway opening pressure (Pao), esophageal pressure, and gas concentration were continuously recorded and digitized for subsequent data analysis. Inspiratory reserve capacity was defined as that volume slowly inspired over 10 sec from a relaxed expiratory level to a Pao = 25 cm H₂O. Expiratory reserve volume was defined as that volume slowly expired over 10 sec from relaxed inspiratory level to Pao = -10 cm H₂O. To determine the functional reserve capacity (FRC) by the helium dilution technique, a rebreathing volume of 80% inspiratory reserve capacity containing 1% He: 21% O₂: 78% N₂ was set at a rate of 50 breaths/min for 15 cycles and the mixed He concentrations were determined by mass spectrometry. The quasi-static compliance of the pulmonary system and that of the lung was determined from the linear

portion of the respective pressure–volume curve obtained during a 6 sec lasting exhalation from TLC to residual volume and the C_L was derived from the transpulmonary pressure–volume curve. Series V_D was obtained using the Fowler method from single-breath wash-in measurements with 1% He in air administered for 3 sec from relaxed expiratory level to TLC to slightly below FRC over 7 sec without a breath-hold. D_{CO} was determined with single breath-holding method using a gas (0.3% $C^{18}O$: 1% He: 21% O_2 , 71.7% N_2) started from the relaxed expiratory level to TLC over 3 sec followed by a breath-holding time of 3 sec and an expiration slightly below FRC. The alveolar volume (V_A) was determined according to the Cotes method [8]. Values of lung function measurements were also normalized to body weight or lung size (e.g., TLC/BW, V_D /TLC). Further details of the technique including calibration and reproducibility are presented in Schulz et al., 2002 [39, 57]. Age (12 weeks)-, body weight- and sex-matched $Sod3^{+/+}$ were used as controls for comparisons of lung function with the $Sod3^{-/-}$ mice (n=6/group).

RNA isolation and microarray analysis: For this initial transcript analysis of the lung we have selected mice 4 weeks of age because this is a late adolescence, early adult age in mice. In humans, the lung develops relatively late in gestation (with premature infants often being at risk of lung failure due to incomplete surfactant synthesis), continues to grow throughout childhood, and peaks at early adulthood (typical at 20-30 years of age). With age lung function begins to decline, with an increasing rate of decline noted in chronic obstructive lung disease. Mice are born with incomplete lung morphology and continue to develop alveoli shortly after birth. The period of rapid alveolarization is nearly completed within the first two weeks after birth and followed by further septation and lung growth at a continuously declining pace until adolescence [43]. Therefore the age of 4 weeks is a late adolescence and early adulthood stage and we considered it to be the optimum time point for the first analysis. This stage can be considered as the transition between completion of lung development and beginning growth-maturation phase.

At exactly age 28 days, female C3H and JF1 mice (n = 8 mice/strain) were anesthetized (5 mg i.p. sodium pentobarbital, Abbott, Chicago, IL), and the lungs were removed and immediately frozen in liquid nitrogen. Total RNA was isolated from homogenized lung with TRIzol reagent (Invitrogen, Carlsbad, CA), and the concentration of total RNA was determined by spectrometric analysis. The quality of RNA was confirmed by running an aliquot of each sample on a denaturing formaldehyde/agarose/ethidium bromide gel and quantified by analysis with an Agilent Bioanalyzer (Quantum Analytics, Inc., Foster City, CA).

To examine differential gene expression of 31,775 70-mer oligonucleotides, microarrays were fabricated by the Genomic and Microarray Laboratory, Center for Environmental Genetics, University of Cincinnati, (<http://microarray.uc.edu>) using a commercial library (Qiagen-Operon, Alameda, CA) consisting of clones amplified by PCR and printed onto glass slides (Omnigrid Microarrayer, GeneMachines, San Carlos, CA). Eight microarrays per strain were compared using 20 µg total RNA per array. Each sample of mRNA was reverse transcribed and reciprocal tagged randomly with fluorescent Cy3 or Cy5 (*i.e.* the Cy3 and Cy5 dye switched, *e.g.*, Cy3 for C3H strain and Cy5 for JF1 strain). Cy3 and Cy5 samples were co-hybridized with the printed 70-mers. A sample from an individual mouse of one strain was randomly paired with a sample from a mouse of another strain and hybridized on a single array. Following hybridization, slides were washed and scanned at 635 (Cy5) and 532 (Cy3) nm (GenePix 4000B, Axon Instruments, Inc., Union City, CA). For inclusion in analysis, an oligo covered its grid location on a slide by a minimum of 50% and a maximum of 400% of the 100 µm feature. Area of the feature was set to one-third of the background area and had an intensity above background. Each array was visually inspected to exclude artifacts or fused or streaked features. Consistency of fluorescence measurements in different hybridizations was assessed by calculating pair-wise correlation coefficients between measurements for the same mouse in different hybridizations. All such pair-wise Pearson's correlation coefficients were 0.9 or above, indicating the reproducibility of measurements.

Microarray data normalization and pathway analysis: Data normalization was performed in three steps for each microarray as described previously [24]. Statistical analysis of microarray data was performed by fitting the following mixed effects linear model for each gene separately as previously described [50]. To further analyze the microarray dataset, we used the MAPPFinder (Microarray Pathway Profile Finder) program to dynamically link microarray data to the Gene Ontology (GO) hierarchy database as illustrated previously [11]. The results, calculated using Fisher's Exact test, are expressed as a "z score" for a given pathway, and values at least 2.0 were considered to be significant. GenMAPP (Gene Microarray Pathway Profiler) was used to view microarray data on biological pathways [10].

Confirmational qRT-PCR: For qRT-PCR, 100 ng lung RNA from C3H and JF1 mice (n = 5 mice/strain) was reverse transcribed into first strand cDNA using a High Capacity cDNA Archive Kit (Applied Biosystems, Foster City, CA) in a 100 µl reaction volume. cDNA (10 µl) was used in a subsequent PCR reaction using 25 µl TaqMan Universal PCR Master Mix (Applied Biosystems), 2.5 µl of each primer mixture, and 12.5 µl RNase-free water. Primer mixtures for *Tff2* (Cat. No. Mm00447491_m1), *Enpp2* (Cat. No. Mm00516572_m1), *Sod3* (Cat. No. Mm00448831_m1), and the actin, beta, cytoplasmic (*Actb*) (Cat. No. Mm00607939_s1) were purchased from Applied Biosystems, and analysis performed with an Applied Biosystems 7900HT System using the following conditions: 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 60 °C for 1 min. For relative quantization of expression of each gene between the C3H and JF1 strain, the comparative C_T method ($\Delta\Delta C_T$) was used. $\Delta C_T = C_T$ (gene of interest) - C_T (ACTB), and this value was calculated for each sample where C_T = cycle number threshold. The comparative $\Delta\Delta C_T$ calculation involved finding the difference between each sample's ΔC_T and the mean ΔC_T for the JF1 strain. These values were then transformed to absolute values using the formula where comparative expression level = $2^{-\Delta\Delta C_T}$.

Western blot analysis: Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed using Nu-PAGE® 4-12% Bis-Tris Gel (Cat. No. NP0322BOX, Invitrogen). Total lung protein (20µg) was loaded in each lane. Following SDS-PAGE, proteins were electrophoretically transferred to an Invitrolon™ PVDF membrane (Cat. No. LC2005, Invitrogen). The membrane was then blocked with 5% fat free milk in PBS-T (170 mM NaCl, 2.68 mM KCl, 3.21 mM Na₃PO₄, 1.47 mM KH₂PO₄, 0.1% Tween-20, pH 7.4) for 1 hour at room temperature. The membrane was then incubated with rabbit anti-mouse SOD3 antibody (kindly provided by Tim D. Oury) in 5% fat free milk in PBS-T, followed by incubation with Horseradish Peroxidase-conjugated goat anti rabbit IgG. The antibody was then detected using an ECL detection system (Amersham Life Science, Little Chalfont, Buckinghamshire, England). Two bands signifying the intact and cleaved SOD3 were observed at 32 and 31 KDa respectively were detected in all cases, and results were normalized to ACTB used as a protein loading control.

Genomic DNA, PCR amplification, and nucleotide sequencing: Genomic DNA from C3H, BALB, C57BL/6J (C57), A/J, PWD, and JF1 mouse strains (listed from larger to smaller lung volume) was purchased from The Jackson Laboratory (Bar Harbor, ME) (<http://www.jax.org/dnares/index.html>). PCR amplification of each candidate gene was performed in an exon-wise manner from the genomic DNA using standard procedure. PCR amplified products were then commercially sequenced both in forward and reverse directions to obtain safe sequences (Sequisherve, Vaterstetten, Germany). The sequences of candidate genes from the five strains (BALB, C57, A/J, JF1, and PWD) were compared with C3H to identify single nucleotide polymorphisms (SNPs) that could result in amino acid changes. List of oligonucleotides used for PCR amplification are provided in the online data supplement (**Table S1**).

Protein structure predictions and hydropathic profiles: SwissModel First Approach Mode [20] was used to generate the Protein Data Bank (.pdb) files. The secondary structure of proteins was developed using the DeepView Swiss-PDB viewer [20] and MDL Chimie programs, or PSIPRED Protein Structure Prediction Server [29]. The DeepView Swiss-PDB viewer was used to calculate the total molecular surface area and electrostatic potential of the proteins. PROSITE scanning was performed to detect the location of each polymorphism in relation to the known functional domains of a protein. Hydropathic profiles of each protein from each strain was calculated and compared using the Kyte and Doolittle [28] method (default window size: 9). World Wide Web addresses of all the programs applied and further details for protein prediction studies are provided in the data supplement.

Results

Lung function of PWD mice:

In our previous survey of lung function in various mouse strains [34, 35], we initially assessed lung function in PWD mice and found TLC in this strain to be similar to that of JF1 strain, which is much smaller than the TLC of C3H strain. In this study, we further examined additional lung functions in PWD mice to determine whether this strain shared other lung function measurements with JF1 mice (**Table 1**). In PWD mice, the TLC was 6% smaller in females (955 μ l) than males (1015 μ l) and was slightly larger than that of JF1 mice (females = 874 μ l; males = 904 μ l) but was significantly less than C3H mice (females = 1443 μ l, males = 1837 μ l). At full inflation V_D was not different between sexes and the mean V_D (200 μ l) in PWD mice was similar to the mean V_D (197 μ l) of JF1 mice, but significantly lower than that of C3H mice (234 μ l). The values of mean V_D /TLC (20.4%) in PWD mice was also similar to the mean V_D /TLC (22.2%) in JF1 mice and both strains had higher mean V_D /TLC than that in C3H mice (14.7%) (**Table 1**).

PWD mice also exhibited the least compliant lung [C_L or C_L /TLC] among all strains studied [34, 35]. In PWD mice, the D_{∞} was also less than that of C3H mice, but when

normalized to V_A , D_{co}/V_A in PWD was greater than that of C3H. Similar to the relationship between JF1 and C3H mice, these findings are consistent with higher gas exchange efficiency per unit of lung volume. Based on this analysis, PWD phenotypes were deemed to resemble those of the JF1 strain, and exon sequences of candidate genes of JF1 and PWD strains were compared to those of C3H strain.

Lung transcript levels differing between C3H and JF1 Strains.

For microarray analysis, we considered expression values significantly different between C3H and JF1 mice when the false discovery ratio (FDR) < 0.05 and an average intensity ≥ 150 . Of the 31,775 genes that were present on the microarray, 18,399 were detected in the lung RNA on each array, and of those, 933 transcripts significantly differed between strains (FDR<0.05) [Submitted to ArrayExpress accession: E-MEXP-855]. The assigned chromosomal location of the 933 genes was obtained from the Ensembl Mouse database using the NCBI m36 mouse assembly [C57 reference strain].

We focused the genome-wide search to transcripts located within a previously identified chromosomal region with significant linkage [35] and that contained a 5'-untranslated region (UTR) single nucleotide polymorphism (SNP) differing between C3H and JF1 as determined *in silico*. Using these criteria, we identified *Sod3* (C3H vs. JF1 = 2.0 ± 0.2 -fold, $p < 0.0000001$), previously identified as a candidate gene on mCh 5, and two novel candidate genes: *Tff2* (C3H vs. JF1 = 2.4 ± 0.5 -fold, $p < 0.004$) located on mCh 17, and *Enpp2* (C3H vs. JF1 = -1.6 ± 0.3 -fold, $p < 0.000002$) located on mCh 15. Confirmation of differences in lung expression of these genes between the C3H and JF1 strains was performed by qRT-PCR (**Figure 1a**). Of these three candidates, the strain difference in SOD3 transcript levels was the greatest and therefore a confirmational Western blot was performed (**1b, 1c**).

We also evaluated the transcript levels of other candidate genes previously identified based upon possible physiological significance and chromosomal location within QTL associated with lung function [35]. Transcript levels of *Rxrb* and *Tgfb β 3* were not significantly

different between the two strains and transcript for *Rln1* and *Pdgfb* were not detected in lung RNA from 4 week old JF1 or C3H mice.

Although transcript expression profiling has been supportive of QTL detection in the past [12, 23, 24], it is salutary to consider the limitation of this approach [15]. Variants at the level of protein structure may not influence gene expression, transcript stability, or steady-state transcript levels (yielding false negatives), or may illicit compensatory mechanisms (yielding false positives). In addition, expression differences may be limited to distal tissues (e.g., hormonal or neural signaling) or to developmental stages not examined. Lastly, the physical proximity of a gene that is expressed in a linkage region does not prove causality. Thus, it is important to sequence suspected candidate genes and to explore the strain differences in structure.

Pathway analysis:

Using MAPPFinder, we further evaluated the functional significance of the lung expression differences between C3H and JF1 mice as assessed by microarray. Comparing classification terms in the Gene Ontology (GO) Consortium categories of cellular components, biological processes, and molecular function, several molecular pathways were identified from the 933 significant transcript expression differences that contained an over-represented number of altered transcripts (i.e., z score ≥ 2.0 with ≥ 7 transcripts altered). Several of effected pathways were likely to play a role in lung development, function, and remodeling and included extracellular matrix ($z = 3.0$), metalloendopeptidase activity ($z = 3.0$), growth factor activity ($z = 2.8$), and vasculature development ($z = 2.4$).

Polymorphisms and predicted protein structure of candidate genes:

Based on the microarray analysis and previous linkage analysis [35], *Sod3*, *Tff2*, *Rln1*, *Tgfb β 3*, *Pdgfb*, and *Rxrb* were selected for further exon sequence analysis to assess amino acid polymorphisms among the mouse strains (including: JF1, PWD, C3H, BALB, C57, and A/J) (**Table 2**). Several polymorphisms were detected in the nucleotide sequences of five

genes (*Sod3*, *Tff2*, *Rln1*, *Tgfbr3*, *Pdgfb*). Of these, no obvious differences in predicted protein resulted from 4 polymorphisms detected in *Tgfbr3* (**Figure S6**) or 1 polymorphism detected in *Pdgfb* (**Figure S7**). Thus, the 3 remaining candidates *Sod3*, *Tff2*, and *Rln1* were further analyzed.

Sod3: Four missense polymorphisms were detected in *Sod3* that included: (1.) A→G in *Sod3*^{JF1, PWD}, Asn-21-Asp, uncharged polar→acidic, (2.) G→C in *Sod3*^{PWD}, Glu-57-Gln, acidic→uncharged polar, (3.) C→T in *Sod3*^{JF1}, Pro-152-Leu, non-polar→ non-polar, and 4) G→T in *Sod3*^{BALB, JF1,PWD}, Ala-186-Ser, non-polar→uncharged polar (**Figure S1**). The predicted amino acid sequence of SOD3 is identical among C3H, C57 and A/J strains. Under identical coordinates of rotational angle by either Swissprot or MDL Chimie analysis, the predicted protein structure of the conformation of the protein backbone for JF1 mice (SOD3^{JF1}) was markedly different from that of other strains including BALB and PWD. Presumably this is due to the combined or singular effect of the polymorphisms at position 152 and 186 (**Figure 2A-B**). In addition to altered protein structure, the predicted SOD3^{JF1} protein had a different electrostatic potential (**Figure 2C**), molecular surface area (**Figure 2D**) and hydropathic profile (**Figure 2E**) in the amino acid region 140-200 (which contains two predicted amino acid conversions). The PSIPRED prediction of SOD3 protein structure also varied in JF1 mice when compared to that of other strains (**Figure S2**). The Asn-21-Asp is within the propeptide sequence of SOD3^{JF1} and could lead to difference in peptide processing. The amino acids in positions 57 and 152 are conserved among mouse, rat, and human.

Tff2: Three missense polymorphisms were detected in *Tff2* that included: (1.) G→A *Tff2*^{PWD}, Arg-4-Gln, basic→non-polar, (2.) A→C *Tff2*^{PWD}, Glu-77-Ala, acidic→non-polar, and (3.) A→C *Tff2*^{PWD}, Lys-89-Arg basic→basic (**Figure S3**). The influence of the nucleotide changes on amino acids at positions 77 and 89 is illustrated in the predicted protein structures (**Figure 3A**) and electrostatic potential (**Figure 3B**). The observed *Tff2*^{PWD} SNPs could also alter the

hydropathic profile (**Figure 3C**) between amino acids 70-90 (containing 2 polymorphisms). No significant alteration was detected in the predicted total molecular surface area of TFF2^{PWD}. The PSIPRED prediction of TFF2^{PWD} protein structure also varied as compared to other mouse strains (**Figure S4**). The predicted Lys-89-Arg conversion is located within the P-type trefoil domain signature and is proximal to a protein kinase C phosphorylation site (PROSITE scanning). Moreover the Arg-4-Gln conversion is within the signal sequence of the protein. The amino acid position 4 in TFF2 is conserved among mouse, rat and human whereas position 77 is conserved among only mouse and human.

Rln1: As compared to the C3H mouse strain, JF1 and PWD mice had 3 missense polymorphisms in *Rln1* that included: (1.) T→C conversion, Leu-98-Pro, non-polar→non-polar, (2.) C→T conversion, Leu-109-Phe, non-polar→aromatic, and (3.) G→A conversion, Val-173-Ile, non-polar→non-polar (**Figure S5**). The PSIPRED RLN1^{JF1, PWD} protein prediction revealed an altered helical structure (amino acid position 80-120) and coiling of the protein (amino acid position 161-175) (**Figure 4A**). The 90-120 amino acid region of predicted RLN1^{JF1, PWD} also had an altered hydropathic profile as compared to other strains (**Figure 4B**). The polymorphism Leu-109-Phe is located within a casein kinase II phosphorylation site (PROSITE Scanning). This non-polar to aromatic amino acid shift could have severe functional consequences in that the biochemical properties of the related protein region would be significantly affected. However, none of the amino acids associated with these polymorphisms are conserved among mouse, human, and rat.

Lung Function of gene-targeted *Sod*^(-/-) mice:

To assess the loss of SOD3 on respiratory function, *Sod3*^(-/-) mice (Age: 12 weeks; n=6) were compared to strain-matched *Sod3*^(+/+) control mice. Baseline lung function values of the *Sod3*^(+/+) mice were not significantly different than those values previously determined of C57BL/6J mice [34], the strain from which these mice were derived [5]. The *Sod3*^(+/+) vs. C57BL/6J lung function values (Mean ± SE) were as follows: TLC = 1203 ± 25 vs. 1174 ±

34 μ l for female and 1337 ± 35 vs. 1380 ± 33 μ l for male; $V_D/TLC = 18.8 \pm 0.4$ vs. $19.3 \pm 0.3\%$ for female and 17.2 ± 0.6 vs. $16.6 \pm 0.3\%$; $C_L = 56.7 \pm 2.4$ vs. 54.9 ± 1.9 μ l/cmH₂O for female and 61.2 ± 3.5 vs. 63.0 ± 2.0 μ l/cmH₂O for male; $D_{co} = 11.4 \pm 0.7$ vs. 11.2 ± 0.2 μ mol/min/hPa for female 13.1 ± 0.4 vs. 13.8 ± 0.3 μ mol/min/hPa for male.

The lung function values (Mean $\pm$ SE) of TLC (1082 ± 33 μ l), C_L (60.6 ± 3.0 μ l/cmH₂O), and D_{co} (10.4 ± 0.5 μ mol/min/hPa) in female *Sod3*^(-/-) mice were not significantly different from those of female *Sod3*^(+/+) mice. Similarly C_L (60.9 ± 3.6 μ l/cmH₂O) in male *Sod3*^(-/-) mice was not significantly different from that of male *Sod3*^(+/+) mice. A slight but significant ($p < 0.05$) decrease in TLC (1167 ± 36 μ l) and D_{co} (12.2 ± 0.6 μ mol/min/hPa) was noted only in male *Sod3*^(-/-) mice as compared to these values in male *Sod3*^(+/+) mice.

In contrast, V_D/TLC was consistently increased in both female ($23.5 \pm 0.4\%$) and male ($22.1 \pm 0.4\%$) *Sod3*^(-/-) mice as compared to the V_D/TLC in sex-matched *Sod3*^(+/+) mice ($18.8 \pm 0.4\%$ in female and $17.2 \pm 0.6\%$ in male) ($p < 0.001$). The V_D/TLC in both female and male *Sod3*^(-/-) mice were similar to those observed in PWD/PhJ and JF1/Msf strains, which are also significantly greater than C3H values (Table 1).

Discussion

Transcript profiling:

To begin to uncover the genetic determinants of lung function and possible links to pulmonary disease, we previously performed a genome-wide linkage analysis study with inbred mice and identified multiple QTLs on chromosomes 5, 15, 17, and 19 [35]. While supporting the contention that lung function is a complex trait, the task of identifying candidate genes remains a considerable challenge, as noted with numerous other QTLs found in mice [19]. Therefore, we employed a systematic approach utilizing microarray analysis to identify difference in transcript levels of genes harbored within the identified chromosomal regions followed by exon screening for SNPs and subsequent *in silico* analysis

of the predicted structural alterations in candidate genes. We also performed functional analysis by measuring lung function in gene-targeted mice that lack one candidate gene.

The transcript profiling was limited to a single time during development and thus should be viewed with caution. Considering lung development, growth and maturation in rodents [43], the age of 4 weeks can be considered as late adolescence and early adulthood in mice. We considered this to be an optimum time for an initial microarray analysis as it can be considered as the transition between completion of the lung development and beginning of growth-maturation phase. Using stringent inclusion criteria ($FDR < 0.05$ and an average intensity ≥ 150), we detected 933 transcripts (of 18,399 detected in the lung) that differed between 4 week old C3H and JF1 mice. These results suggest that a relatively large portion (about 5%) of the transcripts expressed in the lung is altered between these 2 strains. Functional assessment of these differences by MAPPFinder indicated that several of affected pathways were likely to play a role in lung development, function, and remodeling and included pathways controlling the extracellular matrix ($z = 3.0$), metalloendopeptidase activity ($z = 3.0$), growth factor activity ($z = 2.8$), and vasculature development ($z = 2.4$).

Candidate genes identified from our microarray analysis that reside within the previously identified QTLs included *Sod3* on mCh 5 (previously identified as a candidate [35]) and two novel candidates: *Tff2* on mCh 17, and *Enpp2* on mCh 15. The roles of these candidate genes are consistent with the biological pathways identified by MAPPFinder in eliciting downstream effects on key processes of lung development, such as extracellular matrix production and reduction, and cell growth and differentiation, and vasculogenesis. In addition, preliminary histological studies comparing C3H and JF1 support the involvement of these pathways. JF1 mice have significantly lower mean alveolar airspace chord length and higher elastin/collagen ratio in the interstitial compartment of alveolar septal walls compared to C3H mice (Personal communication: Heinz Fehrenbach). These morphometric differences in the alveolus and pulmonary interstitial tissue are consistent with the observed differences in respiratory mechanics, particularly C_L , inasmuch compliance was determined in the linear region of the pressure-volume curve that is controlled by elastic fiber recoil [34,

35]. The altered elastin/collagen ratio in JF1 mice supports the global changes in transcript expression that may be influenced by the candidate genes identified in this study in maintaining the elastin/collagen homeostasis.

Exon sequencing of candidate genes:

Based on the microarray results, we selected two differentially expressed candidate genes, *Sod3* and *Tff2*, for sequencing. In addition, we sequenced 4 other previously identified candidates (*Rln1*, *Pdgfb*, *Tgfrb*, and *Rxrb*) and predicted the consequence to protein properties. Exon sequencing identified polymorphisms in *Sod3*, *Tff2*, and *Rln1* nucleotide sequences that could influence predicted protein structure, property, and function. Effects of the polymorphisms detected in *Tgfrb3* and *Pdgfb* were not immediately obvious. Additionally, no polymorphisms were detected within *Rxrb* that might result in an amino acid exchange. We have yet to sequence *Enpp2* because of its complexity (i.e. 25 exons). The current mouse sequencing database (NCBI m36 mouse assembly) lists 4 SNPs in the 5'-untranslated region (5'-UTR) of *Enpp2* that differed between the C3H, BALB, C57, and A/J strains and 3 SNPs in the 5'-UTR, and one non-synonymous SNP (T→C, Asp-743-Asn) that differed between the C3H and PWD strains.

SOD3: The combined microarray, RT-PCR and Western blot results demonstrated a consistent 2 to 3- fold difference in the level of SOD3 expression between strains. Although the fold difference of transcript and protein may be slightly different between strains, the two isoforms of the protein were not different within each strain. This suggests that the difference in expression between strains is probably due to differences at the transcript level. Altering the level of SOD3 protein is likely to have physiological significance for several reasons.

The protein encoded by *Sod3* is highly expressed in the lung as an antioxidant enzyme associated with the extracellular matrix. Interestingly, *Sod3* has been suggested as a candidate gene for susceptibility to lung function decline in COPD [56]. In the mouse lung, *Sod3* expression is highly localized to pulmonary epithelial cells [17], and augmentation of

Sod3 expression protects against lung injury induced by hyperoxia [5; 16] or bleomycin [2]. Similarly, gene-targeted *Sod3*^(-/-) mice are sensitive to bleomycin or asbestos [14]. Individuals vary greatly (as much as 10-fold) in serum SOD3 protein levels, a phenotype associated with Arg-213-Gly heterozygosity [38, 55]. Moreover, this polymorphism is associated with development of tobacco-smoke induced COPD [22, 56], and is known to alter proteolytic SOD3 processing [33]. In addition, JF1 mice had decreased lung SOD3 protein levels and missense exon polymorphisms that could lead to amino acid substitutions within conserved regions.

The predicted SOD3^{JF1} protein with its altered structure, electrostatic potential, molecular surface area, and hydropathy, together with a lowering of steady-state expression levels, suggests susceptibility of JF1 mice to oxidative stress. Interestingly, C3H mice with the normal wildtype allele are resistant to bleomycin, and we would predict this phenotype based on the predicted protein structure. Significantly lowered expression of SOD3 protein in JF1 lung clearly indicates proportionally lowered enzymatic activity. The polymorphism Asn-21-Asp in the signal sequence of SOD3 is predicted to affect the cotranslational signal peptidase cleavage site of the protein. Therefore this polymorphism might affect the cleavage and secretion of SOD3 preprotein. Moreover, the two polymorphisms (Pro-152-Leu, Ala-186-Ser) in SOD3^{JF1} are located within the catalytic domain of the protein and this region is conserved across species [17]. However, it remains to be investigated as to whether novel polymorphisms uncovered in this study would lead to susceptibility of JF1 and PWD strains to oxidative stress.

Supportive evidence for a role of SOD3 in determining lung function was also obtained with gene targeted *Sod3*^(-/-) mice. When compared to strain-matched *Sod3*^(+/+) mice, gene targeted *Sod3*^(-/-) had increased V_D/TLC , a measure of the percent of total lung capacity that is occupied by the dead space of the conducting airways. Thus, with increasing V_D/TLC ventilation efficiency is reduced. When compared to C3H mice, increased V_D/TLC was previously noted in JF1 mice and we also found this to occur in PWD mice. The difference in V_D/TLC between *Sod3*^(-/-) and *Sod3*^(+/+) mice accounted for about 67% of the phenotypic

difference in V_D /TLC noted between JF1 and C3H mouse strains (Table 1). Previously, the QTL linked to difference in dead space volume between JF1 and C3H mice was localized to mCh 5 ($LOD \geq 4.0$) and was mapped to 26cM (50.6Mbp) to 32cM (54.5Mbp), (delimited by D5Mit14 and D5Mit133, respectively). This led us to suggest *Sod3* which is at 31cM (52.6Mbp), to be a candidate gene for this trait. Because heritability of respiratory function is likely to be complex and not determined by a single gene, the observed difference suggests that other genes or gene-gene interactions contribute to this phenotype.

TFF2: Initially, increased TFF2 protein was localized to mucosal epithelial restitution and wound healing in the gastrointestinal tract. However, additional studies have demonstrated a role for TFF2 in the lung. Transcript levels are increased following allergen challenge of sensitized mice (via a mechanism involving Th2 cytokines and STAT6 signaling) [32], and in airway epithelial cells in subjects with asthma [27]. The altered amino acid sequence in the TFF2^{PWD} protein predicted from the DNA sequence polymorphisms would slightly alter secondary structure. However, the predicted alteration in hydropathic profile and electrostatic potential of TFF2^{PWD} could be significant enough to influence its function. The polymorphism (Arg-4-Gln) in the signal peptide of the protein, and location of another polymorphism (Lys-89-Arg) in the P-type trefoil domain that is close to a protein kinase C phosphorylation site (PROSITE scanning), makes it an interesting candidate for functional analysis. The decreased transcripts in JF1 lung suggest interference in injury, repair, and remodeling processes in the lung.

RLN1: Although RLN1 transcript levels did not differ between strains, RLN1 is the major circulating and stored form of relaxin in rodents and the detected sequence differences were noteworthy. Originally thought mainly to be a hormone related to pregnancy, recent studies support a role for RLN1 in the lung, heart, and brain [40] as an extracellular matrix remodeling hormone that reduces lung fibrosis and promotes wound healing [37]. Progressive, age-related increases in lung weight, collagen content and collagen

concentration have been reported in gene-targeted *Rln1*^(-/-) mice, resulting in the distortion of alveolar structure and altered lung function [36]. Moreover, treatment of *Rln1*^(-/-) mice with human relaxin 2 in early and developed stages of fibrosis resulted in the reversal of lung collagen deposition, confirming its ability to inhibit fibrosis by diminishing collagen biosynthesis and promoting collagen breakdown [37]. Relaxins may also play critical roles in the regulation of collagen deposition in the airway/lung in animal models of allergic airway disease [31].

The identified *Rln1* polymorphisms predicted to lead to Leu-98-Pro and Leu-109-Phe are contained within the connecting peptide domain and the Val-173-Ile is within the A chain of the protein. Position 173 is thought to be determinative in the specificity of ligand binding to its receptor [3] and position 109 could influence the rate of phosphorylation of the protein. In addition, the predicted changes in hydropathic profile within the connecting peptide (amino acid 58-156) and in protein folding might significantly influence the functionality of the RLN1^{JF1, PWD} protein. The predicted diminution of RLN1 function could contribute to altered extracellular matrix (e.g., elastin/collagen ratio) and lung tissue architecture noted in JF1 mice.

ENPP2: Another possible candidate gene implicated by microarray was *Enpp2*. Initially found to increase cell motility *in vitro* [44], ENPP2 is released into the extracellular space [26] where it has shared function with purified extracellular lysophospholipase D [47, 49]. The biological effects of ENPP2 are mediated by degradation of the abundant lung sphingosylphosphorylcholine and lysophosphatidylcholine into and sphingosine 1-phosphate and lysophosphatidic acid (LPA), respectively [7, 18, 54]. LPA can exert diverse biological effects pivotal to lung development and remodeling including mitogenesis [13] and collagen gel reduction by lung fibroblasts [30], LPA also mediates fibronectin release and epithelial cell spreading and migration [1], as well as smooth muscle cell proliferation [6] and contractility [48]. Thus, differences in lung transcript levels may be a genetic determinant of functions regulated through LPA-mediated signaling in multiple target lung cells. Cautionary

consideration of *Enpp2* as a suggestive candidate gene is limited due to lack of additional exon sequence.

Summary: Global differential expression identified > 900 transcripts that differed in C3H and JF1 lungs, many contributing to an enrichment in pathways related to lung development, function, and remodeling (e.g., extracellular matrix). This group was further narrowed down to 3 candidate genes, *Sod3*, *Tff2*, and *Enpp2*, based on localization to chromosome regions previously associated with differences in lung function. Lung function measurements of *Sod3*^(-/-) mice suggest that a larger portion of total lung capacity is occupied by dead space of the conducting airways, which is consistent with our previously identified QTL on mCh 5. In addition to differences in transcript levels, the predicted protein changes associated with polymorphisms in *Sod3* and *Tff2* could, in part, contribute to variation of lung function among commonly inbred laboratory strains of mice. Exon sequencing of another candidate gene *Rln1* identified several alterations of predicted protein structure. Because gene-targeted *Rln1*^(-/-) mice have altered lung structure [36] and *Tff2* is altered in allergic airway sensitivity [32], it would be valuable to examine the lung function *Rln1*^(-/-) and *Tff2*^(-/-) mice to further validate these candidates. These candidate genes also need to be functionally tested to influence the phenotype in a segregant cohort or haplotype across several other mouse strains that have been genotyped. Lastly, the human variation in these genes clearly needs further study both in lung development and in the development of lung disease.

Acknowledgements:

The oligonucleotides were synthesized and supplied by Utz Linzner (GSF-Institute of Experimental Genetics). *Sod3*^(-/-) mouse and anti-mouse SOD3 antibody was kindly provided by Dr. Tim D Oury, University of Pittsburgh, PA, USA. The authors sincerely acknowledge the cooperation of Dr. med. vet. Shinji Takenaka and excellent technical assistance of Martina Schreiber, Maria Neuner, Claudia Zeller and Gunter Eder (GSF-Institute of Inhalation Biology).

Author Contributions: KG, SCW, CR, TS, GDL, and HS conceived and designed the experiments. KG, SCW, CR, IB, MAS, MM, CRT, and JMM performed the experiments. MAS, MM, GDL, and SCW performed microarray data analysis. KG, TS, HS and GDL analyzed the lung function and DNA sequence data. GDL, KG, SCW, and HS wrote the paper.

Funding: This research was supported by the National Genome Research Network 2 (NGFN2: 1GR0430), National Institutes of Health, US Department of Health and Human Services (NIH: HL070542, ES010562, ES006096, HL065612, HL077763, ES085655).

Competing interests: The authors have declared that no competing interests exist.

References:

- [1] Beckmann JD, Romberger DJ, Rennard SI, Spurzem JR. (1995) Induction of bovine bronchial epithelial cell filopodia by tetradecanoyl phorbol acetate, calcium ionophore, and lysophosphatidic acid. *J Cell Physiol* 164:123-131.

- [2] Bowler RP, Nicks M, Warnick K, Crapo JD. Role of extracellular superoxide dismutase in bleomycin-induced pulmonary fibrosis (2002) *Am J Physiol Lung Cell Mol Physiol*; 282(4):L719-26.

- [3] Bullesbach EE, Schwabe C. (1994) Functional importance of the A chain loop in relaxin and insulin. *J Biol Chem*. 269(18):13124-8.

- [4] Burrows B, Knudson RJ, Cline MG, Lebowitz MD. (1977) Quantitative relationship between cigarette smoking and ventilatory function. *Am Rev Respir Dis* 115:195–205.

- [5] Carlsson LM, Jonsson J, Edlund T, Marklund SL. (1995) Mice lacking extracellular superoxide dismutase are more sensitive to hyperoxia.; *Proc Natl Acad Sci U S A*; 92(14):6264-8.

- [6] Cerutis DR, Nogami M, Anderson JL, Churchill JD, Romberger DJ, Rennard SI, Toews ML. (1997) Lysophosphatidic acid and EGF stimulate mitogenesis in human airway smooth muscle cells. *Am J Physiol Lung Cell Mol Physiol* 273:L10-L15.

- [7] Clair T, Aoki J, Koh E, Bandle RW, Nam SW, Ptaszynska MM, Mills GB, Schiffmann E, Liotta LA, Stracke ML. (2003) Autotaxin hydrolyzes sphingosylphosphorylcholine to produce the regulator of migration, sphingosine-1-phosphate. *Cancer Res* 63:5446-5453.

[8] Cotes JE, Chinn DJ, Quanjer PH, Roca J, Yernault JC. (1993) Standardization of the measurement of transfer factor (diffusing capacity). Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J Suppl.* 16:41-52.

[9] Coultas DB, Hanis CL, Howard CA, Skipper BJ, Samet JM (1991) Heritability of ventilatory function in smoking and non-smoking New Mexico Hispanics *Am Rev Respir Dis* 144:770–5.

[10] Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet* 31:19-20.

[11] Doniger SW, Salomonis N, Dahlquist KD, Vranizan K, Lawlor SC, Conklin BR. (2003) MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol* 4:R7.

[12] Doss S, Schadt EE, Drake TA, Lusk AJ (2005) Cis-acting expression quantitative trait loci in mice. *Genome Res* 15(5):681-91.

[13] Ediger TL, Toews ML. (2000) Synergistic stimulation of airway smooth muscle cell mitogenesis. *J Pharmacol Exp Ther* 294:1076-1082.

[14] Fattman CL, Tan RJ, Tobolewski JM, Oury TD. (2006) Increased sensitivity to asbestos-induced lung injury in mice lacking extracellular superoxide dismutase *Free Radic Biol Med.*; 40(4):601-7.

- [15] Flint J, Valdar W, Shifman S, Mott R (2005) Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet* 6(4):271-86.
- [16] Folz RJ, Abushamaa AM, Suliman HB (1999) Extracellular superoxide dismutase in the airways of transgenic mice reduces inflammation and attenuates lung toxicity following hyperoxia. *J Clin Invest*. 103(7):1055-66.
- [17] Folz RJ, Guan J, Seldin MF, Oury TD, Enghild JJ, Crapo JD. (1997) Mouse extracellular superoxide dismutase: primary structure, tissue-specific gene expression, chromosomal localization, and lung in situ hybridization. *Am J Respir Cell Mol Biol* 17(4):393-403.
- [18] Gijssbers R, Aoki J, Arai H, Bollen M. (2003) The hydrolysis of lysophospholipids and nucleotides by autotaxin (NPP2) involves a single catalytic site. *FEBS Lett* 538:60-64.
- [19] Glazier AM, Nadeau JH, Altman TJ (2002) Finding genes that underlie complex traits. *Science* 298: 2345-2349.
- [20] Guex N, Peitsch MC. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 18(15):2714-23.
- [21] Joost O, Wilk JB, Cupples LA, Harmon M, Shearman AM, Baldwin CT, O'Connor GT, Myers RH, Gottlieb DJ (2002) Genetic loci influencing lung function: a genome-wide scan in the Framingham Study. *Am J Respir Crit Care Med* 165(6):795-9.
- [22] Juul K, Tybjaerg-Hansen A, Marklund S, Lange P, Nordestgaard BG. (2006) Genetically increased antioxidative protection and decreased chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 173(8):858-64.

- [23] Karp CL, Grupe A, Schadt E, Ewart SL, Keane-Moore M, Cuomo PJ, Kohl J, Wahl L, Kuperman D, Germer S, Aud D, Peltz G, Wills-Karp M (2000) Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nat Immunol.* 1(3):221-6
- [24] Karyala S, Guo J, Sartor M, Medvedovic M, Kann S, Puga A, Ryan P, Tomlinson CR. (2004) Different global gene expression profiles in benzo[a]pyrene- and dioxin-treated vascular smooth muscle cells of AHR-knockout and wild-type mice. *Cardiovasc Toxicol* 4:47-73.
- [25] Klein RF, Allard J, Avnur Z, Nikolcheva T, Rotstein D, Carlos AS, Shea M, Waters RV, Belknap JK, Peltz G, Orwoll ES (2004) Regulation of bone mass in mice by the lipoxygenase gene *Alox15*. *Science* 303(5655):229-32.
- [26] Koike S, Keino-Masu K, Ohto T, Masu M. (2006) The N-terminal hydrophobic sequence of autotaxin (ENPP2) functions as a signal peptide. *Genes Cells* 11:133-142.
- [27] Kuperman DA, Lewis CC, Woodruff PG, Rodriguez MW, Yang YH, Dolganov GM, Fahy JV, Erle DJ. (2005) Dissecting asthma using focused transgenic modeling and functional genomics. *J Allergy Clin Immunol.* 116(2):305-11.
- [28] Kyte J, Doolittle RF. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 157(1):105-32.
- [29] McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics.* 16(4):404-5.
- [30] Mio T, Liu X, Toews ML, Rennard SI. (2002) Lysophosphatidic acid augments fibroblast-mediated contraction of released collagen gels. *J Lab Clin Med* 139:20-27.

- [31] Mookerjee I, Solly NR, Royce SG, Tregear GW, Samuel CS, Tang ML. (2006) Endogenous relaxin regulates collagen deposition in an animal model of allergic airway disease. *Endocrinology*. 147(2):754-61.
- [32] Nikolaidis NM, Zimmermann N, King NE, Mishra A, Pope SM, Finkelman FD, Rothenberg ME. (2003) Trefoil factor-2 is an allergen-induced gene regulated by Th2 cytokines and STAT6 in the lung. *Am J Respir Cell Mol Biol*. 29(4):458-64.
- [33] Olsen DA, Petersen SV, Oury TD, Valnickova Z, Thogersen IB, Kristensen T, Bowler RP, Crapo JD, Enghild JJ. (2004) The intracellular processing of extracellular dismutase (EC-SOD) is a two step event. *J Biol Chem* 279:22152–7.
- [34] Reinhard C, Eder G, Fuchs H, Ziesenis A, Heyder J, Schulz H (2002) Inbred strain variation in lung function. *Mamm Genome*. 13(8):429-37.
- [35] Reinhard C, Meyer B, Fuchs H, Stoeger T, Eder G, Ruschendorf F, Heyder J, Nurnberg P, de Angelis MH, Schulz H (2005) Genomewide linkage analysis identifies novel genetic Loci for lung function in mice. *Am J Respir Crit Care Med*. 171(8):880-8.
- [36] Samuel CS, Zhao C, Bathgate RA, DU XJ, Summers RJ, Amento EP, Walker LL, McBurnie M, Zhao L, Tregear GW. (2005) The relaxin gene-knockout mouse: a model of progressive fibrosis. *Ann N Y Acad Sci*. 1041:173-81.
- [37] Samuel CS. (2005) Relaxin: antifibrotic properties and effects in models of disease. *Clin Med Res*. 3(4):241-9.

- [38] Sandstrom, J.; Nilsson, P.; Karlsson, K.; Marklund, S. L (1994) 10-fold increase in human plasma extracellular superoxide dismutase content caused by a mutation in heparin-binding domain. *J. Biol. Chem.* 269: 19163-19166,
- [39] Schulz H, Johner C, Eder G, Ziesenis A, Reitmeier P, Heyder J, Balling R. (2002) Respiratory mechanics in mice: strain and sex specific differences.; *Acta Physiol Scand.* 174(4):367-75.
- [40] Scott DJ, Layfield S, Riesewijk A, Morita H, Tregear GW, Bathgate RA. (2004) Identification and characterization of the mouse and rat relaxin receptors as the novel orthologues of human leucine-rich repeat-containing G-protein-coupled receptor 7. *Clin Exp Pharmacol Physiol.* 31(11):828-32.
- [41] Shaheen S. (1997) The beginnings of chronic airflow obstruction. *Br Med Bull.*; 53(1):58-70.
- [42] Silverman EK, Palmer LJ, Mosley JD, Barth M, Senter JM, Brown A, Drazen JM, Kwiatkowski DJ, Chapman HA, Campbell EJ, Province MA, Rao DC, Reilly JJ, Ginns LC, Speizer FE, Weiss ST. Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. *Am J Hum Genet.* 2002 May;70(5):1229-39.
- [43] Soutiere SE, Mitzner W. (2006) Comparison of postnatal lung growth and development between C3H/HeJ and C57BL/6J mice. *J Appl Physiol.* 100(5):1577-83.
- [44] Stracke ML, Krutzsch HC, Unsworth EJ, Arestad A, Cioce V, Schiffmann E, Liotta LA. (1992) Identification, purification, and partial sequence analysis of autotaxin, a novel motility-stimulating protein. *J Biol Chem* 267:2524-2529.

[45] Tager IB, Weiss ST, Munoz A, Rosner B, Speizer FE. (1983) Longitudinal study of the effects of maternal smoking on pulmonary function in children. *N Engl J Med.*; 309(12):699-703.

[46] Tankersley CG, Fitzgerald RS, Kleeberger SR. (1994) Differential control of ventilation among inbred strains of mice. *Am J Physiol*; 267:R1371–R1377.

[47] Tokumura A, Majima E, Kariya Y, Tominaga K, Kogure K, Yasuda K, Fukuzawa K. (2002) Identification of human plasma lysophospholipase D, a lysophosphatidic acid-producing enzyme, as autotaxin, a multifunctional phosphodiesterase. *J Biol Chem* 277:39436-39442.

[48] Towes ML, Ustinova EE, Schultz HD. (1997) Lysophosphatidic acid enhances contractility of isolated airway smooth muscle. *J Appl Physiol* 83:1216-1222.

[49] Umezu-Goto M, Kishi Y, Taira A, Hama K, Dohmae N, Takio K, Yamori T, Mills GB, Inoue K, Aoki J, Arai H. (2002) Autotaxin has lysophospholipase D activity leading to tumor cell growth and motility by lysophosphatidic acid production. *J Cell Biol* 158:227-233.

[50] Wesselkamper SC, Case LM, Henning LN, Borchers MT, Tichelaar JW, Mason JM, Dragin N, Medvedovic M, Sartor MA, Tomlinson CR, Leikauf GD. (2005) Gene expression changes during the development of acute lung injury: role of transforming growth factor-beta. *Am J Respir Crit Care Med* 172:1399-1411.

[51] Wilk JB, DeStefano AL, Arnett DK, Rich SS, Djousse L, Crapo RO, Leppert MF, Province MA, Cupples LA, Gottlieb DJ, Myers RH. (2003) A genome-wide scan of pulmonary

function measures in the National Heart, Lung, and Blood Institute Family Heart Study. *Am J Respir Crit Care Med.*; 167:1528-33.

[52] Wilk JB, DeStefano AL, Joost O, Myers RH, Cupples LA, Slater K, Atwood LD, Heard-Costa NL, Herbert A, O'Connor GT, Gottlieb DJ (2003) Linkage and association with pulmonary function measures on chromosome 6q27 in the Framingham Heart Study. *Hum Mol Genet.* 12(21):2745-51.

[53] Wilk JB, D'Jousse L, Arnett DK, Rich SS, Province MA, Hunt SC, Crapo RO, Higgins M, Myers RH (2000) Evidence for major genes influencing pulmonary function in the NHLBI family heart study. *Genet Epidemiol* 19(1):81-94.

[54] Xie Y, Meier KE. Lysophospholipase D and its role in LPA production. (2004) *Cell Signal* 16:975-981.

[55] Yamada, H.; Yamada, Y.; Adachi, T.; Goto, H.; Ogasawara, N.; Futenma, A.; Kitano, M.; Hirano, K.; Kato, K. (1995) Molecular analysis of extracellular-superoxide dismutase gene associated with high level in serum. *Jpn. J. Hum. Genet.* 40: 177-184.

[56] Young RP, Hopkins R, Black PN, Eddy C, Wu L, Gamble GD, Mills GD, Garrett JE, Eaton TE, Rees MI. (2006) Functional variants of antioxidant genes in smokers with COPD and in those with normal lung function. *Thorax.* 61(5):394-9.

[57] <http://aretha.jax.org/pub-cgi/phenome/mpdcgi?rtn=projects/details&sym=Schulz1>.

Figure Legends

Figure 1: (A.) Lung transcript levels of superoxide dismutase 3, extracellular (SOD3), trefoil factor 2 (TFF2), and ectonucleotide pyrophosphatase/phosphodiesterase 2 (ENPP2), in 4 week old C3H/HeJ (C3H) and JF1/Msf (JF1) mice. Real-time qRT-PCR was performed, and the comparative C_T method ($\Delta\Delta C_T$) was used [where $\Delta C_T = C_T (\text{gene}) - C_T (\text{ACTB})$] for quantitation of the data. The $\Delta\Delta C_T$ calculation involved finding the difference between each sample's ΔC_T and the mean ΔC_T for the JF1 strain. Data are presented as relative JF1 expression (means $\pm$ SEM, $n = 5$ mice/strain). Significant differences in expression levels were found between the C3H and JF1 strains for all transcripts (t-test, $p \leq 0.05$). **(B.)** Western blot analysis of lung homogenate of 4 week old C3H and JF1 mice ($n = 3$ mice/strain) showing significantly low immunoreactive SOD3 protein level (both intact and cleaved forms) in the lung. Total protein (20 μ g) was loaded in each lane and normalized β -Actin (ACTB) as the protein loading control. **(C.)** Lung SOD3 protein levels in JF1 strains is 3.0 ± 0.9 -fold less than that of C3H strains. The data is measured from the Western Blot (Figure 1B) with $n = 3$ mice/strain and normalized to β -Actin.

Figure 2. Predicted protein structure of superoxide dismutase 3, extracellular (SOD3) comparing C3H/HeJ (C3H) and JF1/Msf (JF1) mouse strains. (A.) Under identical coordinates of rotational angle, the predicted protein backbone of SOD3^{JF1} structure differs from the predicted structures of the other strains. The amino acids at positions 152 and 186 which participate in the protein backbone formation are highlighted. This 3D model have been developed using the Swissprot PDB viewer. **(B.)** The 3D modelling of the SOD3 protein from C3H and JF1 strains using MDL Chimie also reflect the same observation as in Figure 2A. **(C.) The total**

molecular surface area of SOD3^{JF1} is larger than that of SOD3^{C3H}. Such alterations in molecular surface area can influence on its activity and functionality in general. **(D.) The electrostatic potential of SOD3^{JF1} differs from that of SOD3^{C3H}.** Such alterations in electrostatic potential influence on the biochemical property of the protein thereby can alter the functionality in general. **(E.) Amino acids region between positions 140-200 of the SOD3^{JF1} protein shows significantly altered hydropathic profile compared to SOD3^{C3H}.** Magnified image of the area with major change is provided in inset.

Figure 3: Predicted protein structure of trefoil factor 2 (TFF2) comparing PWD/PhJ (PWD) and C3H/HeJ (C3H) mouse strains. (A.) Amino acids at positions 77 and 89 which participate in the protein backbone formation (highlighted in the figure) shows altered charge distribution in TFF2^{PWD}. Under identical coordinates of rotational angle, the predicted TFF2^{PWD} structure only differs slightly in the predicted conformation of the protein backbone from those of other strains. This 3D model have been developed using the Swissprot PDB viewer. **(B.) The electrostatic potential of TFF2^{PWD} predicted protein differs from that of TFF2^{C3H}.** Such alterations in electrostatic potential influence on the biochemical property of the protein thereby can interfere with the functionality in general. **(C.) Altered hydropathic profile (magnified image in inset) between amino acids 70-90 (containing 2 polymorphisms) is observed in TFF2^{PWD} when compared with TFF2^{C3H}.**

Figure 4: Predicted protein structure of relaxin 1 (RLN1) comparing C3H/HeJ (C3H), PWD/PhJ (PWD), and JF1/Msf (JF1) mouse strains. (A.) The PSIPRED protein prediction shows altered helical structure of RLN1^{JF1, PWD} between the amino acid positions 80-120 and altered coiling of the protein between amino acid positions 161-175 compared to that of other strains. **(B.)** The amino acids region between positions 90-120 of the RLN1^{JF1, PWD} protein has altered hydrophobic profile. Magnified image of the areas with major change is provided in inset.

Strain	PWD/PhJ female	PWD/PhJ male	JF1/Msf female	JF1/Msf male	C3H/HeJ female	C3H/HeJ male
Parameters						
BW [g]	15.8 ± 0.2	19.1 ± 0.4*	16.5 ± 0.4	22.7 ± 0.7	22.4 ± 0.6	29.0 ± 0.7
Lung volumes						
TLC [μl]	955 ± 17 ^{##}	1015 ± 20 ^{##*}	874 ± 17 ⁺	904 ± 11 ⁺	1443 ± 29	1837 ± 79
TLC / BW [μl/g]	60.7 ± 1.6 [#]	53.2 ± 0.9 ^{##*}	53.6 ± 1.7 ⁺	40.2 ± 1.5 ⁺	64.9 ± 2.1	63.5 ± 2.8
V _D [μl]	202 ± 5 ⁺	198 ± 4 ⁺	193 ± 3 ⁺	201 ± 4 ⁺	227 ± 2	245 ± 4
V _D / TLC (%)	21.3 ± 0.7 ⁺	19.5 ± 0.6 ⁺	22.1 ± 0.5 ⁺	22.3 ± 0.3 ⁺	15.8 ± 0.3	13.6 ± 0.5
Compliance of the lung						
C _L [μl/cm H ₂ O]	31.9 ± 1.3 ⁺	34.0 ± 1.3 ⁺	35.5 ± 2.2 ⁺	37.8 ± 1.1 ⁺	84.8 ± 3.6	100.3 ± 8.7
C _L / TLC [μl/cm H ₂ O/ml TLC]	33.4 ± 1.1 ^{##}	33.4 ± 0.8 ^{##}	40.9 ± 3.0 ⁺	41.8 ± 1.3 ⁺	58.5 ± 1.7	54.9 ± 2.0
Pulmonary diffusing capacity						
D _{CO} [μmol/min/hPa]	9.3 ± 0.4 ⁺	10.5 ± 0.6 ⁺	8.3 ± 0.4 ⁺	9.9 ± 0.4 ⁺	12.4 ± 0.3	14.1 ± 0.5
D _{CO} /V _A [μmol/min/hPa/ml V _A]	11.4 ± 0.3 ^{##}	11.9 ± 0.3 ⁺	9.9 ± 0.4	11.8 ± 0.3 ⁺	9.3 ± 0.2	8.7 ± 0.4

For comparison, data from C3H/HeJ and JF1/Msf mice are provided [34, 35]. BW: Body weight, TLC: total lung capacity, V_D: dead space volume, C_L: static lung compliance, D_{CO}: diffusing capacity for carbon monoxide, V_A: alveolar volume. # denotes significant differences to JF1/MsF (p<0.01), + those to C3H/HeJ (p<0.05), * denotes significant difference between sex in PWD/PhJ (p<0.05). Lung function data of other inbred mouse strains included in the ‘Mouse Phenome Project’ are available at <http://aretha.jax.org/pub-cgi/phenome/mpdcgi?rtn=projects/details&sym=Schulz1>.

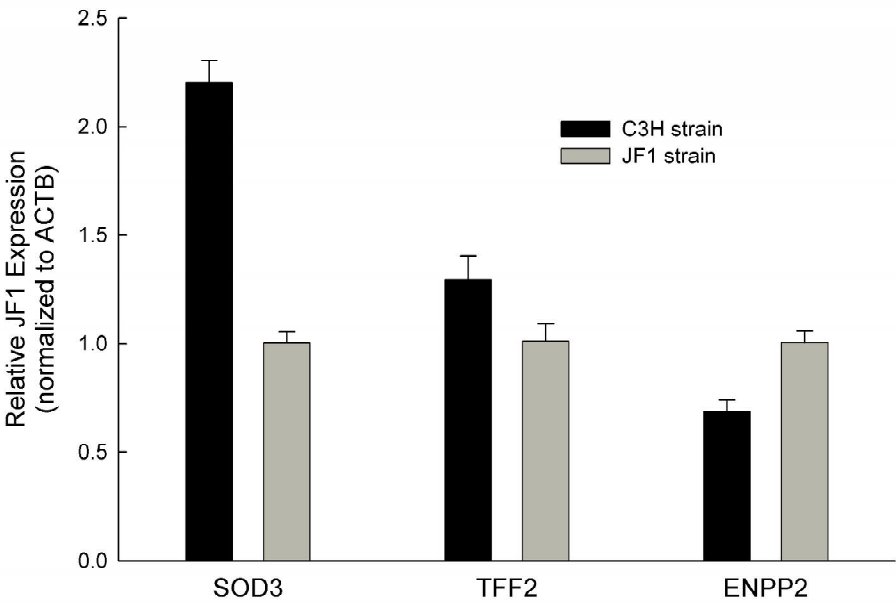
Table 2: Polymorphisms in candidate genes linked to lung function in mice and their predicted effect on the respective protein.

Gene	Associated lung function	Polymorphisms detected and predicted amino acid change	Secondary protein structure	Total molecular surface area	Electrostatic potential	Hydropathic profile	Functional domain
Sod3 mCh 5	C _L and V _D	Total: 4: JF1 (3), PWD (3),BALB (1) Asn-21-Asp: JF1, PWD Glu-57-Gln: PWD Pro-152-Leu: JF1 Ala-186-Ser: JF1, PWD, BALB	++ (JF1)	++ (JF1)	++ (JF1)	++ (JF1) (between amino acid positions 140-200) + (PWD) (around amino acid position 57)	Amino acid 21: Predicted cleavage site of propeptide Amino acid 152, 186: Located within the catalytic domain.
Tff2 mCh. 17	TLC and D _{co}	Total: 3 (PWD) Arg-4-Gln Glu-77-Ala Lys-89-Arg	+ (PWD)	-	++ (JF1)	+ (PWD) (between amino acid positions 70-90)	Amino acid 4: Located within the signal sequence. Amino acid 89: Located within P-type trefoil domain signature and proximal to a protein kinase C phosphorylation site
Rln1 mCh. 19	C _L	Total: 3 (JF1, PWD) Leu-98-Pro Leu-109-Phe Val-173-Ile	++ (JF1, PWD)	NA	NA	++ (JF1, PWD) (between amino acid positions 90-120)	Amino acid 89: Located within Casein kinase C phosphorylation site
Pdgfb mCh 15	TLC and D _{co}	Total:1 (C57, JF1, PWD) Ile-183-Val	-	-	-	-	-
Tgfb3 mCh 5	C _L and V _D	Total: 4 (JF1, PWD) His-96-Tyr Ser-391-Pro Met-765-Pro Pro-772-Ala	+ (JF1, PWD)	NA	NA	+ (JF1, PWD)	-
Rxrb mCh 17	TLC and D _{co}	None	-	-	-	-	-

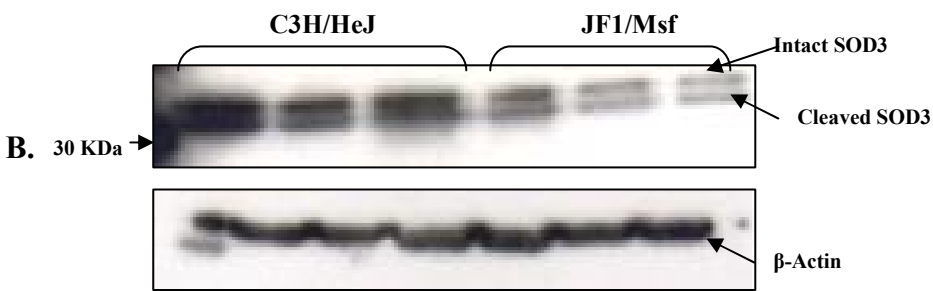
All polymorphisms among the mouse strains (including: JF1, PWD, BALB, C57, and A/J) are determined in comparison to C3H/HeJ. Strains analyzed but with no polymorphism listed are identical to C3H/HeJ for the respective protein. **Symbols:** mCh.= Mouse chromosome, **Sod3** = superoxide dismutase 3, extracellular, **Tff2** = trefoil factor 2, **Rln1** = relaxin 1, **Pdgfb** = platelet-derived growth factor, beta polypeptide, **Tgfb3** = transforming growth factor, beta receptor III, **Rxrb** = retinoid X receptor beta, C_L = lung compliance, V_D dead space volume, TLC = total lung capacity, Strains sequenced: **C3H**= C3H/HeJ; **JF1**= JF1/Msf, **PWD** = PWD/PhJ, **BALB**= BALB/cByJ, **C57**= C57BL/6J; **A/J**; (+) alteration observed; (++) disruptive alteration observed, (NA) not applicable as .pdb file could not be generated; (-) no change observed. Lung functions are associated to the respective chromosome region [35].

Figure 1

A.



B.



C.

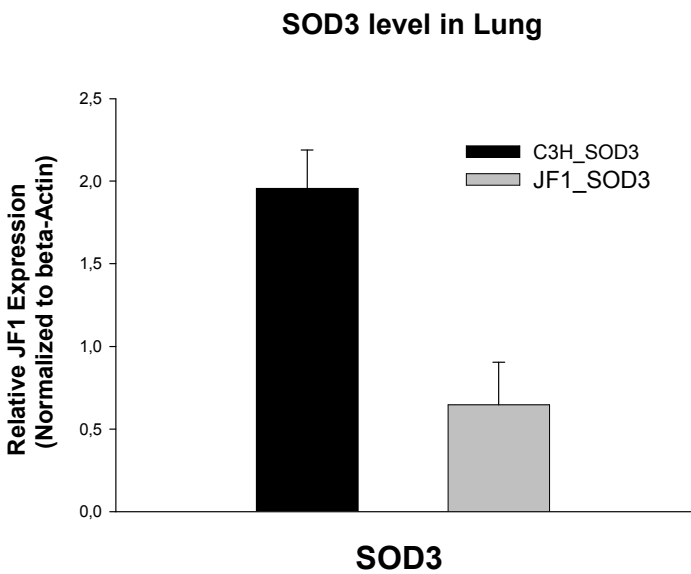


Figure 2:

A.

B.

C.

D.

E.

**3D Structure
(Swissprot)**

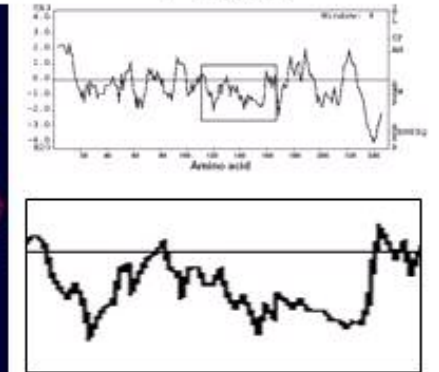
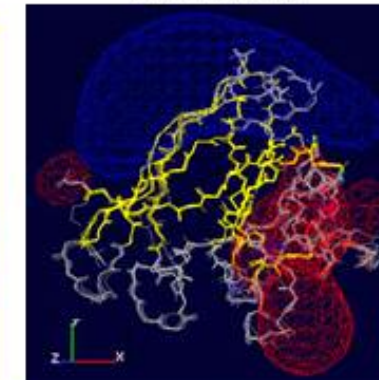
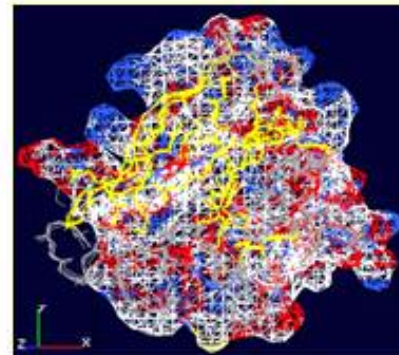
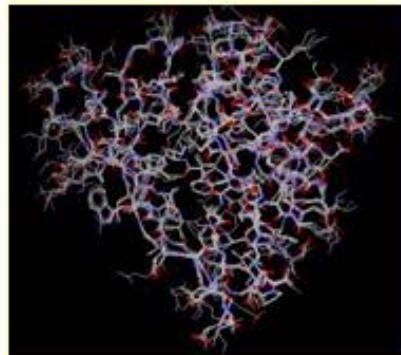
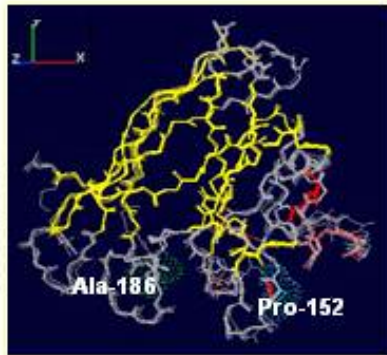
**3D Structure
(MDL Chimie)**

**Molecular
Surface Area**

**Electrostatic
Potential**

**Hydropathic
Profile**

SOD3^{C3H}



SOD3^{JF1}

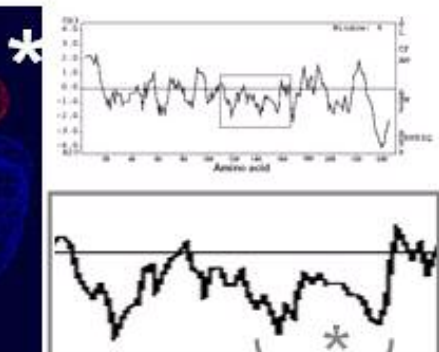
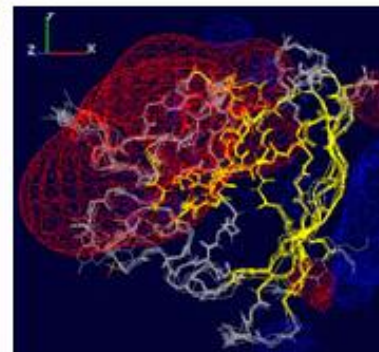
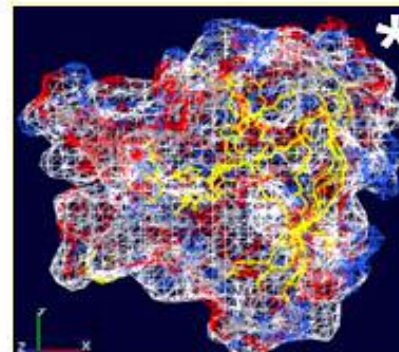
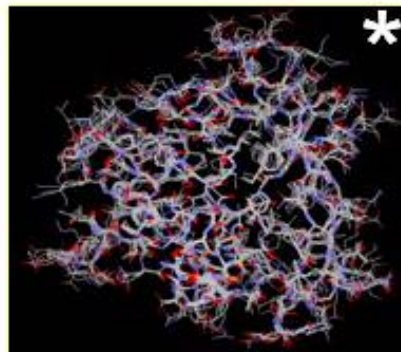
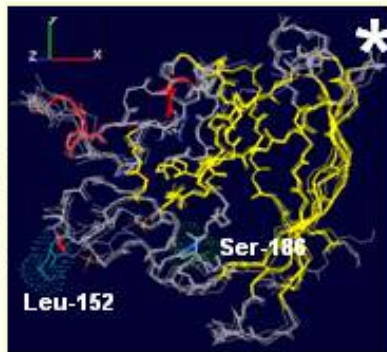
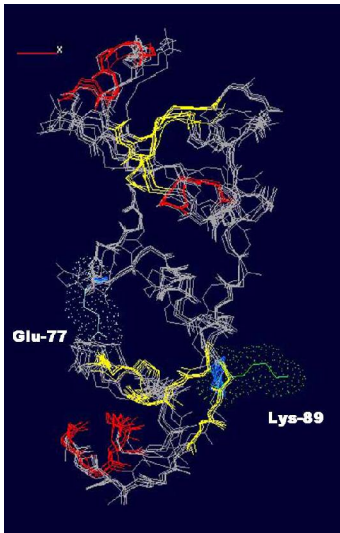
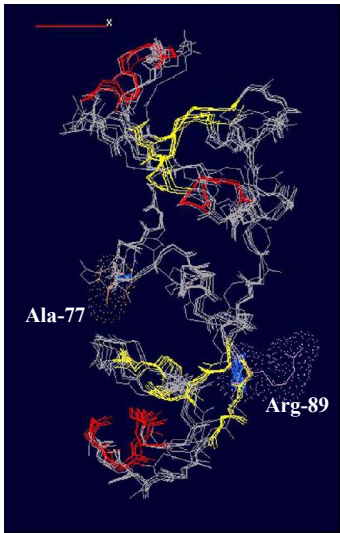


Figure 3 **Structure
(Swissprot)**

TFF2^{C3H}

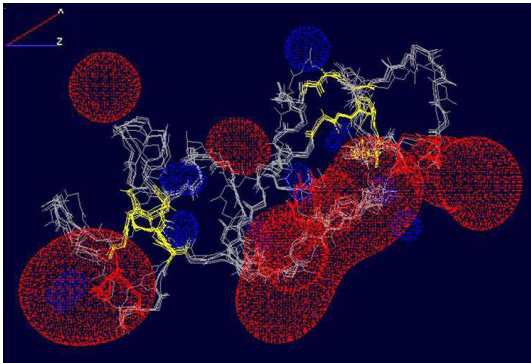
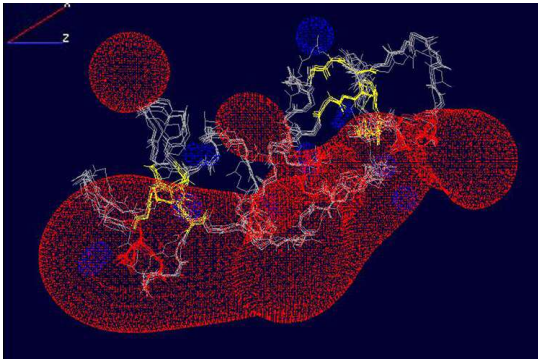


TFF2^{PWD}



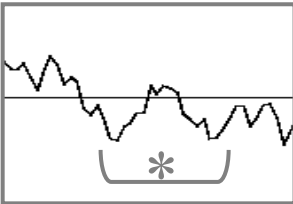
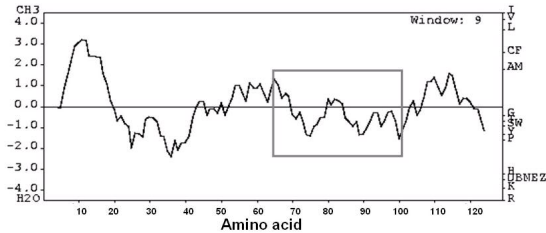
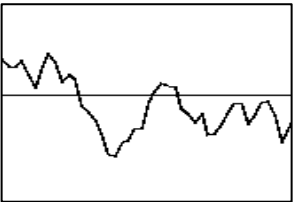
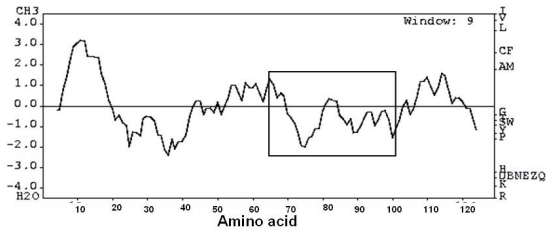
A.

**Electrostatic
Potential**



B.

**Hydropathic
Profile**



C.

