

STRUCTURAL HOTSPOT MAPPING OF MISSENSE VARIANTS IN SERPINA1: AN IN-SILICO FRAMEWORK LINKING SEQUENCE VARIATION, PROTEIN STABILITY AND SERPIN DYSFUNCTION

Eman Ejaz¹, Ghulam Murtaza^{*2}, Rabia Ijaz³, Muhammad Ahsan Raza⁴, Shakeela Mahvish⁵, Azalfa Younas⁶, Amina Shahbaz⁷, Kainat Nawaz⁸, Rida Fatima⁹, Shahid Bashir¹⁰

^{1,*2,5,6,7,8,9}Department of Zoology, University of Gujrat, Gujrat, Punjab, Pakistan

³Department of Biochemistry and Biotechnology, University of Gujrat, Gujrat Punjab, Pakistan

⁴Government Graduate College (Boys), Gujranwala, Punjab 50250, Pakistan

¹⁰Neuroscience Center, King Fahad Specialist Hospital, Dammam, 31444, Saudi Arabia.

¹emanejaz100@gmail.com, ²drghulam.murtaza@uog.edu.pk, ³rabiaijaz37@gmail.com, ⁴ahsanraza1810@gmail.com, ⁵shakeela.mahvish15@gmail.com, ⁶azalfayounas@gmail.com, ⁷shahbazmatta60@gmail.com, ⁸nawazkainat671@gmail.com, ⁹rf4555482@gmail.com, ¹⁰shahidbpk13@gmail.com

DOI: <https://doi.org/10.5281/zenodo.23334866>

Keywords

AATD, missense SNP, COPD, in-silico analysis, molecular docking, protein stability, SERPINA1, AAT

Article History

Received: 25 June 2026

Accepted: 16 September 2026

Published: 13 October 2026

Copyright @Author

Corresponding Author: *
Ghulam Murtaza

Abstract

Chronic obstructive pulmonary disease (COPD) is a major cause of morbidity and mortality globally, the major inherited cause is alpha-1 antitrypsin deficiency (AATD) which is caused by pathogenic variants of a protein called alpha-1 antitrypsin that is the main inhibitor of neutrophil elastase in the lung. NGS has revealed a large number of rare variants of SERPINA1 that are still not understood and need to be characterized, especially missense variants that may affect AAT protein folding, stability, secretion, polymerization and protease inhibitory function, and thus impair the ability of AAT to protect lung tissue from elastase-mediated damage. This review addresses various computational approaches in the characterization of rare missense variants of SERPINA1, including sequence-based pathogenicity prediction, protein stability estimation, 3D structural modeling, computation of molecular dynamics, and molecular docking with neutrophil elastase. The complementarity of these methods and their respective limitations are evaluated, and the integration of these methods into a systematic framework for variant prioritization is discussed. Substitutions in structurally or functionally important areas may lead to changes in protein stability and conformation, as shown by the computational analysis, and molecular dynamics modeling can show dynamic effects that may not be apparent in static structural models. But interpretation is still limited due to the lack of experimental validation, population biases in variant databases, and methodological heterogeneity of prediction tools. Standardized, integrative computational workflows and experimental validation are thus required to enhance functional interpretation of rare SERPINA1 variants and to select those that may be

relevant to AATD-associated COPD susceptibility.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a leading cause of mortality around the world. It is a chronic progressive disease of airflow limitation due to chronic respiratory inflammation of the airways (Mei et al., 2022). Tobacco smoking is the most important environmental risk factor for COPD, but only a small proportion of heavy smokers develop clinically significant COPD, suggesting there may be genetic susceptibility (Tiensuu et al.). AAT is synthesized from a hepatocyte as a 394 residue glycoprotein and secreted into the circulation, where it enters the lung interstitium (Yang & Kim, 2025), where it combines with neutrophil elastase released during an inflammatory response (Crystal, 1990). Failure to maintain this balance will cause unopposed elastase to destroy elastin in the alveolar wall, causing tissue destruction that causes emphysema and early-onset COPD. The mutations in SERPINA1 disrupt this balance in two very general ways: firstly, mutations that cause misfolding of the mutant protein, polymerization of the protein, and retention in the endoplasmic reticulum of the hepatocyte rather than its secretion into the lung; and secondly, mutations that impair the elastase-inhibitory capacity of AAT that does make it to the lung. The Z allele (Glu342Lys) is the most thoroughly studied example of the first mechanism and the only major contributor to severe AATD, which results in hepatic polymer accumulation and substantial decreases in serum AAT. Over 100 different variants of SERPINA1 have since been identified, and the use of next-generation sequencing and large screening programs is steadily increasing the list of variants (Blanco et al., 2017). *In-silico* characterization is likely to serve as a helpful tool to fill the void in functional annotation, especially in the context to the most common cause of chronic disease and mortality in the world: Obstructive Pulmonary Disease (OPD) characterized by

chronic airway inflammation and progressive destruction of the lung parenchyma, leading to persistent airway obstruction. Even though smoking is the most important environmental risk factor for COPD, only some of the heavy smokers develop clinically significant COPD, indicating that there is a genetic susceptibility. We are now able to identify over one hundred different variants of SERPINA1, and will continue to add to this list with the use of ever-advancing next-generation sequencing and through large screening programs. In fact, over 500,000 people were screened at one U.S. referral laboratory in 2023 to identify 22 previously unreported alleles of SERPINA1, and case series have continued to report on novel rare variants with unknown pathogenicity (Havvarimana, 2023). In most of these rare alleles, there is no functional or structural data, and so clinicians and genetic counselors have no reliable basis to interpret incidental or family-screening results. This is an ever-expanding gap between the amount of sequences being produced and the amount of functional annotation, which is where *in-silico* characterization steps in (Kadota et al., 2022). Computational biology provides a scalable alternative to thorough experimental characterization. Several sequence-based predictors make use of evolutionary conservation and physicochemical substitution patterns and identify amino acid changes that are likely to affect protein function, several structure-based predictors are based on the modeling of the three-dimensional effects of a substitution and the estimation of the effect on the folding stability; and several molecular docking approaches simulate the effect of the change in the surface geometry or loop geometry of the modified protein on its interaction with a physiological partner. As AATD is essentially a protein misfolding, secretion failure, and protein-protein interaction disease, SERPINA1 is a good

candidate to be part of this class of integrated computational analysis (Javed et al., 2026).

This review presents current advances in *in-silico* methods used to assess the pathogenicity of missense mutations in SERPINA1 broken down into four stages: sequence-based pathogenicity prediction, protein stability prediction, protein structural modeling, in particular of the serpin fold, and molecular docking between AAT and neutrophil elastase, encompassing the most recent developments using the AlphaFold software and molecular dynamics simulations of the serpin fold (Sandhaus & Turino, 2013). We don't treat these approaches as standalone methods, but focus on the rationale, evidentiary strength, and weaknesses of using them together in tiered and consensus pipelines and note the methodological gaps which remain to make confident computational prioritization of rare SERPINA1 variants possible for COPD risk assessment (Wang et al., 2021).

Molecular Biology of Alpha-1 Antitrypsin and Inhibitory Mechanism of Serpins

The inhibitory serpins belong to a family of over 1,500 identified proteins called the serpin (serine protease inhibitor) superfamily and are distinguished by a metastable native (stressed, or S) conformation that is not a consequence of the protein's instability, but is its structural feature that gives rise to the inhibitory mechanism. The mechanism of serpin inhibition is unique and resembles a 'mousetrap' mechanism. The scissile bond is found within an exposed, highly flexible region, called the reactive center loop (RCL), which is located at the target protease, neutrophil elastase, recognition site in AAT and is first broken (Gangemi et al., 2025). The cleavage induces an extraordinarily large conformational change, involving the insertion of the cleaved RCL as a new fourth strand into beta-sheet A and the translation of the covalently linked protease by more than 70 Å to the opposite end of the serpin, to the point that the catalytic site is distorted so much that it is locked in an inactive, kinetically irreversible complex. This stressed-to-

relaxed (S-to-R) transition is associated with a significant increase in thermodynamic stability, which is just what makes the trapped complex "trapped" and keeps the protease from escaping (Marijanovic et al., 2019).

The conformational movement and flexibility that allows this inhibiting function is the same that makes serpins highly susceptible to misfolding. In the presence of the RCL which does not fit into the partially opened beta-sheet A of another molecule, the serpin assumes an inactive latent conformation; if the RCL of one molecule fits into the partially opened beta-sheet A of another molecule, ordered polymers are formed. The classical AATD-causing Z variant (Glu342Lys), which destabilizes the region next to the shutter domain of beta-sheet A, favors premature and self-propagating loop-sheet polymerization in the endoplasmic reticulum (ER) of hepatocytes, rather than normal secretion (O'Brien et al., 2022). Thus, the polymerized form of AAT leads to loss-of-function because polymerized AAT cannot reach the circulation and serve as a protective agent for the lung, and gain of toxic-function because the accumulated polymer triggers stress responses in cells leading to liver injury (Stoller & Aboussouan, 2012).

The mechanics of destabilizing mutations are most likely to be those that affect the shutter region, the RCL or the beta-sheet A-strands and least likely to be those in solvent-exposed loops, far from those functional regions. The reasons behind the choice of using multiple predictors based on sequence conservation, as well as explicit structure and dynamic modeling: a variant might be evolutionarily rare but not structurally destabilizing, and vice versa (Tsutsui, 2008).

SERPINA1 Variants and Clinical Correlations

There are two deficiency alleles of SERPINA1 that dominate the clinical genetics of the gene. The severity of AATD is seen with the Z allele, which has a strong correlation with low AAT levels in serum, hepatic polymer deposition and increased risk of early onset emphysema,

especially in smokers. The S allele (Glu264Val) results in a lesser decrease in serum AAT activity and typically carries a lower but a clinically significant risk. The S and Z alleles are responsible for more than 95% of the clinically recognized AATD and are combined together to make up the majority of the diagnostic testing for AATD (Hatipoğlu & Stoller, 2016).

In addition to these two well-described alleles, over one hundred other SERPINA1 alleles have been reported, ranging from completely null (frameshift and PTC mutations) to decreased secretion (deficient alleles), to secreted at near-normal levels but with defective elastase-inhibitory activity (dysfunctional alleles), and finally rare alleles with unknown function (Reeves et al., 2019). Thus, this catalogue continues to grow with more recent screening and case-series studies, including a University of Florida screening initiative that identified 22 new alleles, mostly null or deficient, through screening of over half a million individuals and additional case series that identified other rare alleles, such as missense variants like p.Ala349Pro, which were confidently classified as pathogenic only after *in-silico* prediction, cellular expression assays and structural reasoning about the distance to the RCL (Matamala et al., 2026).

Newly discovered variants of SERPINA1 are now placed in context thanks to the use of population and clinical databases. The NCBI dbSNP resource, which has been updated daily for 30 years, is the database of choice for cataloguing human sequence variation and is the basis for most pipelines for retrieval of SERPINA1 variants, while the Ensembl Variant Effect Predictor enables rapid annotations of variants and prioritization based on the context of the transcript, protein and regulatory sequence (Mróz et al., 2024). Because gnomAD resources and other population allele-frequency databases carry millions of population-level missense changes many of which are likely to be merely evolutionarily neutral, rare polymorphisms it is important to perform a filtering step by cross-referencing candidate missense variants with

them, yet few other databases are available to do this. An additional complexity of the observations is that SERPINA1 variant frequency and phenotype varies across populations. Both the Z and S alleles have distinct geographical gradients throughout the Europe and rare null and deficient alleles have been reported representing different variants in different ethnicities with implications for the design of ethnic-specific screening panels and the use of the frequency of a variant as a measure of pathogenicity. This population heterogeneity supports population oriented, systematic *in-silico* characterization instead of using functional data from mainly European populations (Claussnitzer et al., 2020).

Sequence-Based Computational Prediction of Missense SNP Pathogenicity

Most missense-SNP characterization pipelines provide the first analytical layer in the form of sequence based pathogenicity prediction, which relies on evolutionary conservation and on the likelihood of substitution instead of making an explicit structural prediction (Adzhubei et al., 2013). SIFT (Sorting Intolerant From Tolerant) scores the amino acid conservation at a particular site in a sequence alignment of homologous proteins and assigns a score to a substitution if it is below a normalized threshold (usually 0.05). The PolyPhen-2 (Polymorphism Phenotyping v2) tool, however, considers all of these sequence-based features, as well as structural ones like solvent accessibility, functional domain membership, and the proximity to known ligand binding or interface residues, and outputs a probabilistic score; values above roughly 0.9 are generally classified as "probably damaging," while those in the range of 0.4 to 0.9 are generally classified as "possibly damaging. The delta-alignment score, which is widely implemented and popular, is used in PROVEAN (Protein Variation Effect Analyzer) to account for the effect of a substitution across a cluster of homologous sequences and this is an additional,

independent validation line in variant triage pipelines (Sandell & Sharp, 2022).

In each of the independent studies in different disease genes, none of the predictors has been both highly sensitive and highly specific; moreover, there was not complete agreement between the accuracy of the predictors in the different studies. Using confirmed pathogenic and benign PRRT2 variants, PROVEAN and MutationTaster performed best and produced the highest overall accuracy while SIFT and PolyPhen-2 performed best in terms of sensitivity, though at the expense of accuracy. A similar study of missense variants in BRCA1 and BRCA2 yielded similar results that PolyPhen-2 had the lowest sensitivity, specificity, and accuracy among the tools and that the combination of (Miller, 2018). Likewise, a comparative study of GJB2, GJB3 and GJB6 deafness associated variants with nine different tools, showed that SIFT and PROVEAN gave the best accuracy of the tools tested, while other classifiers predicted too many false positive pathogenicity calls and were deemed inappropriate for this gene family.

These findings have direct relevance to the characterization of SERPINA1 for two reasons: These findings have direct relevance to characterization of SERPINA1 for two reasons: First, they argue that the near universal requirement to predict a missense SNP as "high confidence" for downstream structural and docking analyses, with at least two (and preferably three) independent, sequence based tools, should be viewed as consensus voting such that the false-positive and false-negative misclassification rates are significantly lower than if the single tools used them individually. Second, they emphasize that a sequence-based call of 'deleterious' or 'probably damaging' should not be taken as sole evidence to prioritize a variant as a pathogenicity disrupting AAT folding, secretion, or elastase-inhibitory activity without structural and preferably dynamic evidence (Ben Khelifa, 2023).

One other SERPINA1 warning is about the numbering systems. In the past, the numbering

system employed in both AAT variants has been the mature-protein numbering system which excludes the 24-residue signal peptide and for this reason it is common to see the references 'Glu342Lys' and 'Gly367Lys' for the same variant; in contemporary reporting, the numbering system is HGVS-compliant with the UniProt P01009 reference sequence, and gives 'Glu366Lys' and 'Gly367Lys' for the same variant. Setting the conventions when submitting a variant to SIFT, PolyPhen-2, or PROVEAN is not something commonly done and indeed, several case reports on newly identified rare mutations have highlighted this as a re-occurring diagnostic challenge (Mróz et al., 2024).

Protein Stability Prediction and Thermodynamic Assessment

In most cases, AATD is not a catalytic loss-of-function disease, as such, but rather a disease in which protein misfolding and impaired secretion affect the protein. In this regard, thermodynamic stability prediction plays a central role in the characterization of SERPINA1 variants. I-Mutant 2.0 is still one of the most popular methods, which is based on support vector machines trained on experimentally determined $\Delta\Delta G$, to predict the change in the Gibbs free energy for unfolding ($\Delta\Delta G$) upon a specific mutation, along with a qualitative call of increased or decreased stability (Masso & Vaisman, 2008). The negative predicted $\Delta\Delta G$ implies that the mutant protein should be thermodynamically less stable than the WT, which in the context of a metastable serpin like AAT (Giri Rao & Gosavi, 2018).

The pathogenic logic of stability loss leading to AATD phenotype is clearly defined for the prototype Z variant which has a Glu342Lys mutation that destabilizes the area adjacent to the shutter domain of beta-sheet A in such a way as to allow an alternative folding pathway to the normal native structure to become the favored pathway and lead to the formation of polymer. This logic has been continued by structural and biophysical studies of rarer variants, as the predicted or measured effect of a variant's

position on folding stability increasingly becomes a way to interpret newly described substitutions; variants near the shutter, the breach and the base of the RCL are increasingly characterized by a higher frequency of reduced secretion and polymer accumulation in cellular expression assays (Wanner & Sandhaus, 2015).

The prediction of stability, however, has its own restrictions, which are quite prominent for a protein like AAT which is liable to polymerization. Most $\Delta\Delta G$ predictors, including I-Mutant, are based on datasets of globular, generally non-metastable proteins, and predict a difference in stability compared to a single reference native structure; they have been developed to estimate the resilience of a protein to simple unfolding, not to the alternative, sheet-opening conformational pathway that is characteristic of serpin polymerization. In principle a mutation can be predicted by an unspecified $\Delta\Delta G$ estimation to be a mild destabilizer, and yet create an exaggerated effect on whether the protein is correctly folded or is in the polymer form. One of the main reasons for treating stability predictions as one part of an integrated pipeline instead of a prediction of pathogenicity alone is that they are best used alongside other predictions, in particular those that are now being increasingly used in parallel, namely explicit structural modeling and molecular dynamics simulation, which are discussed next (Garrido-Rodríguez et al., 2024).

Alphafold and Molecular Dynamics

Unlike the sequence and stability-based predictors that give a single score of the likely effect of a mutation, the structural modeling tries to visualize and quantify the effect of the substitution on the architecture of AAT in 3D structure. Despite its approximate nature, homology modelling (e.g. using SWISS-MODEL) is still a widely used method in which a solved or predicted structure of the wild-type AAT is selected and used to build a mutant, after which the two structures can be compared both structurally and in terms of loop conformation,

packing around the altered residue and the secondary structure content. The models produced in this way are usually subjected to further validation for stereochemical plausibility analysis including Ramachandran plots, ERRAT and other quality metrics before they are used for downstream interpretations, because a poorly validated model could lead to structural artefacts that could be incorrectly interpreted as real distortion due to mutation (Supriya & Sunkar, 2026).

AlphaFold has significantly enriched the set of structural tools available for the analysis of serpin variants, while at the same time providing valuable insights in the limitations of the method for analysis of metastable proteins. In a 2024 study, AlphaFold was shown to accurately predict the wild-type SERPINA1 native structure, while Rosetta-based stability calculations correctly picked out the most severe mutants when tested against a set of pathogenic missense mutations in the closely related wild-type serpin, SERPINA1, the fold predicted by AlphaFold was essentially the native fold for all tested mutants, irrespective of the actual biochemical severity or known clinical consequences of the mutation. Put another way, AlphaFold did not discriminate between mutations that were found by orthogonal approaches to be strongly destabilizing versus only mildly destabilizing mutations, since the model was not trained to predict the alternative, non-native conformational states (latent and/or polymeric) that are where serpin pathologies lie. The finding has only a direct cautionary implication for the studies of SERPINA1: They should rely on explicit thermodynamic stability calculation in addition to AlphaFold-based structure comparisons, and ideally, molecular dynamics simulation (Garrido-Rodríguez et al., 2024).

While static structural comparison can only provide information on the primary structure, molecular dynamics (MD) simulation provides a temporal dimension that can be used to observe the behavior of a mutant AAT structure over the course of a simulated trajectory, instead of just a

single minimized structure. Pathogenic missense variants, analyzed across different protein systems using MD, have all been found to cause measurable changes in the fluctuation of the structure, often measured as root mean square fluctuation (RMSF), in terms of reduced compactness (radius of gyration) and changes in hydrogen bonding networks as compared to the wild-type trajectory. In the serpin literature, time-resolved simulation has been used to support a preference of destabilizing mutations to affect the shutter region and the strands of beta-sheet A that are expected to open to allow insertion of RCL, which is consistent with a mechanism that predicts that these regions are functional hotspots for polymerization-prone variants (Pastorella et al., 2026). For SERPINA1, this implies that the static homology or AlphaFold models should be used as a starting point, not as an endpoint, for a thorough structural characterization pipeline where only variants with a destabilizing signature that is supported by dynamic simulation should

be called confidently structural pathogenicity (Sala et al., 2023).

Molecular Docking and AAT-Neutrophil Elastase Interaction

The last analytical layer is functional consequence that is clinically relevant, in terms of whether or not a specific variant of AAT can bind and irreversibly inhibit neutrophil elastase. Molecular docking mimics this protein-protein interaction in a computational fashion; this is usually accomplished by generating the docked pose of the mutant AAT reactive center loop with the active site of neutrophil elastase (ELANE), and comparing the binding free energy, hydrogen-bonding pattern and hydrophobic contact surface to the corresponding wild-type complex. Molecular Operating Environment (MOE) is a popular application used for docking purposes, as it supports both rigid and flexible docking protocols and offers a variety of scoring functions which can be compared in a standardized manner in a panel of variants (Moroy et al., 2012).

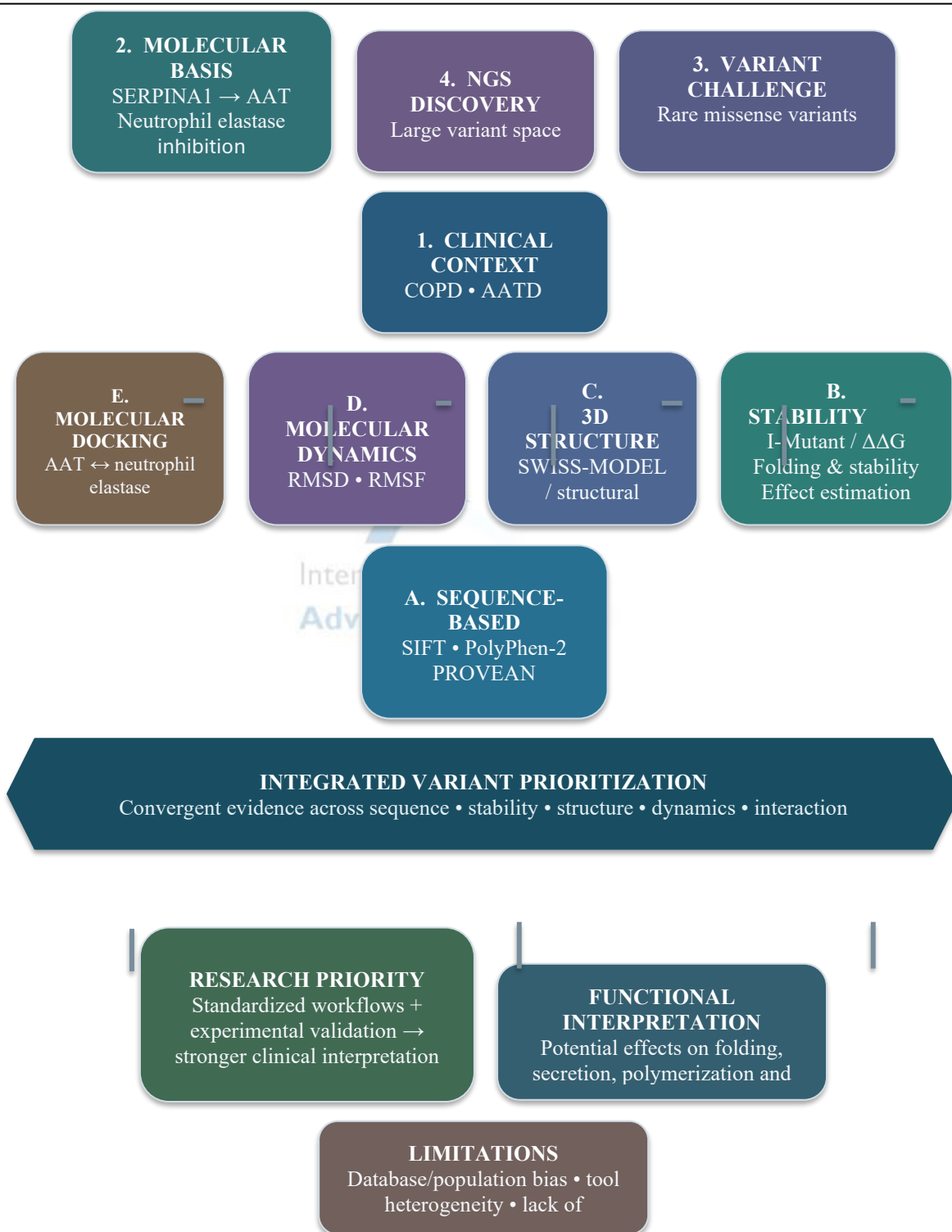


Figure 1. Integrated Computational workflow for Characterization of Rare Missense SERPINA1 Variants in COPD

As the mousetrap inhibitory mechanism relies on insertion of the mousetrap RCL into beta-sheet A, and the same is true for the site of the scissile bond that makes the peptide substrate susceptible to neutrophil elastase, missense substitutions in the RCL have a higher risk of disrupting both, based on the fact that they are scattered throughout the peptide and within beta-sheet A. A docking approach to the study of serpin-protease interactions has highlighted that even conservative mutations at RCL positions can substantially reduce predicted binding affinity and/or influence the geometry of the catalytic triad engagement, which agrees with clinical experience that RCL-proximal AAT variants are more likely to have a dysfunctional rather than simply absent phenotype (Marijanovic et al., 2019).

A recent computational study of the common SERPINA1 rs17580 variant is a prime example of how docking evidence is incorporated into a genetics and computation framework for the identification of susceptibility biomarker information for progression of COPD that integrates variant retrieval, pathogenicity prediction and structural reasoning to support the relevance of the variant to disease risk. Such studies demonstrate the practical workflow that this review advocates: docking conclusions are best used as a last resort, after concordant sequence-based predictions have flagged a variant, and upon its ability to cause a change in the elastase-binding behavior it must be structurally destabilizing so that the docking-derived change in elastase binding can be interpreted in the context of a coherent mechanism rather than as a stand-alone numerical score (Qadir et al., 2026).

Towards Integrated *In-Silico* Pipelines: Rationale, Strengths and Limitations

Sequenced based pathogenicity prediction, thermodynamic stability estimation, structural and dynamic modeling, and molecular docking are all complimentary and complementary in nature and each will tackle a different part of the question of whether or not a missense variant of

SERPINA1 will be clinically relevant. Sequence-based tools are quick and cheap to use on large lists of variants, and suitable for a first-pass screen, but are not disease-specific in terms of their underlying structural biology; and, as benchmarking studies across different disease genes have repeatedly demonstrated, have an intrinsic tendency to miss the correct classification of individual variants. Conventional $\Delta\Delta G$ estimators were not developed for the serpin-specific polymerization pathways; stability prediction is an additional (mechanistically grounded) filter applicable to a misfolding disease. Structural modeling, especially when supplemented with molecular dynamics simulations, provides direct visualization of the structural distortion caused by mutations, but is more costly and informative if the structural model is interpreted in the light of a known structural "hotspot map" for the RCL, shutter and beta-sheet A (Tanoli et al., 2025).

The recent studies focusing on the Serpina1 gene and genes adjacent to Serpina1 typically follow this pipeline: variant retrieval and filtering (from resources like dbSNP, Ensembl, gnomAD, and ClinVar); consensus sequence-based prediction; stability estimation; and, if computational resources are available, dynamic simulation of the structure; and finally, docking-based functional interpretation, where concordant evidence from each tier is used to nominate a shortlist of high-confidence pathogenic candidates that are subject to a final data-integration step. The multi-modal experimental approach, involving pre-existing serum AAT concentration, phenotyping, sequencing, and, in growing number of cases, cellular expression assays, is reflected in this layered approach, where the computational approaches mirror the steps taken by diagnostic laboratories for characterizing novel alleles of SERPINA1 and determining pathogenicity (Giacopuzzi et al., 2018).

Nevertheless, it is crucial to make it clear what an in-silico pipeline of such type can and cannot prove. However, computational prioritization does not prove pathogenicity per se, as would be

required by variant classification schemes like those of the American College of Medical Genetics and Genomics. The most appropriate way to evaluate *in-silico* evidence is to see it as a way of efficiently prioritizing limited experimental resources in order to target the variants that are most likely to be informative and to warrant continued investigation, and

experimental evidence, measured by either cellular expression assays measuring AAT secretion and intracellular accumulation, biochemical assays of elastase-inhibitory activity, or, most significantly, by correlation with clinical evidence and serum AAT levels and pulmonary phenotype, is the gold standard (Snyder et al., 2006).

Table 1: Characterization of SERPINA1 variants

Computational Characterization of Rare SERPINA1 Missense Variants in AATD Associated COPD					
Stage	Approach	Key Question	Output	Biological Meaning	Limitations
Variant Identification	NGS	Which rare SERPINA1 variants need characterization?	Curated rare missense variants	Candidate variants for analysis	Database and population bias (Kueppers et al., 2019).
Sequence Prediction	SIFT/ PolyPhen-2	Could the substitution impair protein function?	Deleterious predictions	Initial pathogenicity prioritization	Predictors may disagree (Castellana & Mazza, 2013).
Stability Analysis	I-Mutant/ VVG predictors	Does substitution alter stability?	Predicted stability change	Possible folding/stability defects	Model-dependent predictions (Mahajan et al., 2026).
3D Structural Modeling	AlphaFold	Where is residue?	Mutant structural model	Structural and functional disruption	Static model don't capture dynamics (Szczepski & Jaremko, 2025).
Molecular Dynamics	All-atom MD simulation	Does mutation alter flexibility?	Radius gyration of H-bond	Dynamic structural effects	Force-field/ simulation line dependence (Li et al., 2025).
Molecular Docking	AAT- Neutrophil elastase docking	Could the variant alter protease recognition?	Docking poses, scores, interface contacts	Elastase inhibition	Docking is not experimental affinity (Landers, 2016).
Integrated Prioritization	Multilayer computational integration	Which variants show convergent evidence?	Variant prioritization profile	Sequence, Stability, structure, and dynamic	No single method proves pathogenicity (Giacopuzzi et

				interaction	al., 2018).
Experimental Validation	Biochemical Validation	Does computational prediction reflect biological function?	Functional experimental evidence	Computational interpretation	Maybe unavailable for rare variants (Spielmann & Kircher, 2022).

Challenges and Gaps

The confidence in the use of purely computational SERPINA1 variant characterization is occurred by a few methodological shortcomings. First, the lack of experimentally validated benchmark data sets specific to serpin genes, such as SERPINA1, from which sequences, stabilities and structures can be calibrated by the predictors, introduces uncertainty when they are used to extrapolate their degree of accuracy at serpin genes to AAT.

Second, population representation in the variant databases used to fuel filtering and frequency-based triage is still skewed towards European-ancestry cohorts, which may lead to under-cataloging of all truly rare pathogenic variants in under-represented populations as well as to misclassifying polymorphisms from a population as suspicious just because they are not present in the reference panels. However, *in-silico* characterization efforts, such as those done mainly using public databases, is a real constraint, as documented population heterogeneity in the frequency of the SERPINA1 alleles and their phenotypes exists.

Third, as shown in the AlphaFold versus molecular dynamics comparison of SERPINA1, existing structure prediction methods are not inherently suitable to model conformational states that are latent and polymeric as seen in the pathology of serpins and defined by mutations, and hence single-structure predictions may not adequately reflect the actual structural liability of a mutation unless combined with stability calculation and simulation of dynamics. Fourthly, molecular docking of the AAT-neutrophil elastase interaction is generally carried out with rigid or slightly flexible RCL structures and scores are interpreted as an initial snap fit or initial binding

capacity and not a complete simulation of the inhibitory mousetrap.

Lastly, there is no consistent approach to weighting and combining the sequence-based, stability-based, structural, and docking-based evidence for nominating a SERPINA1 variant as a high priority candidate, making it difficult to compare across studies and leaving a lot of room for methodologically subjective interpretation of individual research groups. Community agreement on minimum reporting standards would help here, much like the minimum reporting standards for the classification of clinical variants generally have improved the reproducibility of these.

Future Directions

Some concrete developments would benefit future work in this area. Further validation of SERPINA1 variants in a coordinated study of the novel alleles discovered in large-scale screening programs through cellular expression of the serpin would yield the serpin-specific benchmark data needed to be able to tune the accuracy of the predictor. In the context of AlphaFold-based serpin studies, where the application of molecular dynamics simulation in addition to sequence and stability tools would be helpful to close the gap identified, near the shutter and RCL, conformational dynamics seems to be the more informative signal. Increased population diversity in variant discovery and database curation, including systematic screening in currently underrepresented populations, such as those of South Asia, would enhance sensitivity detected for rare variants and would enhance accuracy of population-frequency-based filtering. Last but not least, tighter coupling of computational pipelines and convenient experimental validation, such as

the prioritization of a small set of the best predicted variants for cellular secretion and elastase-inhibition assays instead of using *in-silico* prediction as an end, would render the results of future SERPINA1 characterization studies more translatable and help develop serpin-specific ground-truth datasets that are currently lacking in the field.

Conclusion

SERPINA1 is in an important class of disease genes that can be fully characterized *in silico*: the structural mechanism of disease is well understood, the mechanics of the conformational transitions involving a metastable serpin-fold and the relationships between the transitions and loss of secretion, polymerization, and impaired elastase-inhibition are all directly linked; however, the great majority of missense variants reported in the literature are not functionally characterized. All of these predictors or estimators of stability, structural or dynamic models and molecular docking provide unique and complementary evidence that could help address this gap, and the growing literature increasingly suggests the adoption of integrated pipelines rather than the use of any single tool. These pipelines, with their continued improvement by leveraging greater experimental benchmarks, population-based variant discovery, and explicit modeling of serpin-specific conformational states, show promise in boosting the computational prioritization of clinically relevant SERPINA1 variants and, as a result, in enhancing genetic risk assessment and early diagnosis of COPD due to alpha-1 antitrypsin deficiency.

REFERENCES

- Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, 76(1), 7.20. 21-27.20. 41.
- Ben Khelifa, E. (2023). Effects of mutations in the reactive center loop of alpha-1-antitrypsin on inhibitory activity against neutrophil elastase.
- Blanco, I., Bueno, P., Diego, I., Pérez-Holanda, S., Casas-Maldonado, F., Esquinas, C., & Miravittles, M. (2017). Alpha-1 antitrypsin Pi* Z gene frequency and Pi* ZZ genotype numbers worldwide: an update. *International journal of chronic obstructive pulmonary disease*, 561-569.
- Castellana, S., & Mazza, T. (2013). Congruency in the prediction of pathogenic missense mutations: state-of-the-art web-based tools. *Briefings in bioinformatics*, 14(4), 448-459.
- Claussnitzer, M., Cho, J. H., Collins, R., Cox, N. J., Dermitzakis, E. T., Hurles, M. E., Kathiresan, S., Kenny, E. E., Lindgren, C. M., & MacArthur, D. G. (2020). A brief history of human disease genetics. *Nature*, 577(7789), 179-189.
- Crystal, R. G. (1990). Alpha 1-antitrypsin deficiency, emphysema, and liver disease. Genetic basis and strategies for therapy. *The Journal of clinical investigation*, 85(5), 1343-1352.
- Gangemi, R., Bignotti, M., Denardo, A., Pearce, C. N., Ronzoni, R., Lomas, D. A., Irving, J. A., Fra, A., & Gangemi, F. (2025). Identification of an exosite at the neutrophil elastase/alpha-1-antitrypsin interface. *The FEBS Journal*, 292(8), 1887-1903.
- Garrido-Rodríguez, P., Carmena-Bargueño, M., de la Morena-Barrio, M. E., Bravo-Pérez, C., de la Morena-Barrio, B., Cifuentes-Riquelme, R., Lozano, M. L., Pérez-Sánchez, H., & Corral, J. (2024). Analysis of AlphaFold and molecular dynamics structure predictions of mutations in serpins. *Plos one*, 19(7), e0304451.

- Giacopuzzi, E., Laffranchi, M., Berardelli, R., Ravasio, V., Ferrarotti, I., Gooptu, B., Borsani, G., & Fra, A. (2018). Real-world clinical applicability of pathogenicity predictors assessed on SERPINA1 mutations in alpha-1-antitrypsin deficiency. *Human Mutation*, 39(9), 1203-1213.
- Giri Rao, V. H., & Gosavi, S. (2018). On the folding of a structurally complex protein to its metastable active state. *Proceedings of the National Academy of Sciences*, 115(9), 1998-2003.
- Hatipoğlu, U., & Stoller, J. K. (2016). α 1-Antitrypsin deficiency. *Clinics in Chest Medicine*, 37(3), 487-504.
- Havyarimana, E. (2023). *Understanding environmental risk factors associated with vasculitis in United Kingdom* University of Glasgow].
- Javed, A., Soomro, A. S., Mughal, S. A., Shahzad, S., Yousuf, A., Khushk, I., & Qureshi, A. S. (2026). SERPINA1 and Alpha-1 Antitrypsin: A Review of Molecular Structure, Clinical Variants and Computational Study of Diseases Involving the Protein. *BioMed Target Journal*, 4(1), 1-1.
- Kadota, K., Matsumoto, K., Uchiyama, H., Tobita, S., Maeda, M., Maki, D., Kinehara, Y., Tachibana, I., Sosnowski, T. R., & Tozuka, Y. (2022). In silico evaluation of particle transport and deposition in the airways of individual patients with chronic obstructive pulmonary disease. *European Journal of Pharmaceutics and Biopharmaceutics*, 174, 10-19.
- Kueppers, F., Andrade, M. D., Xu, Q., Dunbrack Jr, R. L., Kim, J., & Sanders, C. L. (2019). Protein modeling to assess the pathogenicity of rare variants of SERPINA1 in patients suspected of having Alpha 1 Antitrypsin Deficiency. *BMC Medical Genetics*, 20(1), 125.
- Landers, S. A. (2016). *An Investigation of Dysregulated Neutrophil Activity in Alpha-1 Antitrypsin Deficiency* Royal College of Surgeons in Ireland].
- Li, W., Yuan, M., Li, J., Tu, Z., Cui, K., Liao, Y., Liao, H., Liu, J., & Gong, X. (2025). Exploring the mechanism of guben kechuan granules in the treatment of COPD via network pharmacology, molecular docking and molecular dynamics simulations. *Scientific Reports*, 15(1), 41391.
- Mahajan, N., Singh, S., Chopra, V., Garg, K., Gund, V., & Sharma, S. (2026). Molecular modelling and simulation-based profiling of nsSNPs in hedgehog pathway genes linked to COPD susceptibility. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 15(1), 107.
- Marijanovic, E. M., Fodor, J., Riley, B. T., Porebski, B. T., Costa, M. G., Kass, I., Hoke, D. E., McGowan, S., & Buckle, A. M. (2019). Reactive centre loop dynamics and serpin specificity. *Scientific Reports*, 9(1), 3870.
- Masso, M., & Vaisman, I. I. (2008). Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis. *Bioinformatics*, 24(18), 2002-2009.
- Matamala, N., Osaba, L., Drobic, E., Gil-Martin, S., Martin-Galiano, A. J., Almadana-Pacheco, V., Cazorla, M. J., Costa, F., Souza-Sokoloski, C., & Çörtük, M. (2026). Functional and clinical significance of novel SERPINA1 variants on alpha-1 antitrypsin deficiency. *Respiratory Research*, 27(1), 95.
- Mei, F., Dalmartello, M., Bonifazi, M., Bertuccio, P., Levi, F., Boffetta, P., Negri, E., La Vecchia, C., & Malvezzi, M. (2022). Chronic obstructive pulmonary disease (COPD) mortality trends worldwide: An update to 2019. *Respirology*, 27(11), 941-950.

- Miller, C. M. (2018). *Functional profiles of health: elucidating genetics and microbiome of human disease* Technische Universität München].
- Moroy, G., JP Alix, A., Sapi, J., Hornebeck, W., & Bourguet, E. (2012). Neutrophil elastase as a target in lung cancer. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*, 12(6), 565-579.
- Mróz, J., Pelc, M., Mitusińska, K., Chorostowska-Wynimko, J., & Jezela-Stanek, A. (2024). Computational tools to assist in analyzing effects of the SERPINA1 gene variation on alpha-1 antitrypsin (AAT). *Genes*, 15(3), 340.
- O'Brien, M. E., Murray, G., Gogoi, D., Yusuf, A., McCarthy, C., Wormald, M. R., Casey, M., Gabillard-Lefort, C., McElvaney, N. G., & Reeves, E. P. (2022). A review of alpha-1 antitrypsin binding partners for immune regulation and potential therapeutic application. *International Journal of Molecular Sciences*, 23(5), 2441.
- Pastoressa, L., Pivetti, V., Valente, M., Beghè, B., Clini, E., Tonelli, R., & Cerri, S. (2026). Alpha-1 Antitrypsin Deficiency Beyond COPD and Emphysema: A Narrative Review. *Medical Sciences*, 14(1), 106.
- Qadir, M. F., Jamal, A., Ikhtiar, F., Zafar, I., & Mehdi, S. M. S. (2026). Genetic and computational insights into SERPINA1 rs17580 as a susceptibility biomarker for COPD progression. *risk*, 9, 10.
- Reeves, E. P., Dunlea, D. M., McQuillan, K., O'Dwyer, C. A., Carroll, T. P., Saldova, R., Akepati, P. R., Wormald, M. R., McElvaney, O. J., & Shutchaidat, V. (2019). Circulating truncated alpha-1 antitrypsin glycoprotein in patient plasma retains anti-inflammatory capacity. *The Journal of Immunology*, 202(8), 2240-2253.
- Sala, D., Engelberger, F., Mchaourab, H. S., & Meiler, J. (2023). Modeling conformational states of proteins with AlphaFold. *Current Opinion in Structural Biology*, 81, 102645.
- Sandell, L., & Sharp, N. P. (2022). Fitness effects of mutations: an assessment of PROVEAN predictions using mutation accumulation data. *Genome biology and evolution*, 14(1), evac004.
- Sandhaus, R. A., & Turino, G. (2013). Neutrophil elastase-mediated lung disease. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 10(sup1), 60-63.
- Snyder, M. R., Katzmman, J. A., Butz, M. L., Yang, P., Dawson, D. B., Halling, K. C., Highsmith, W. E., & Thibodeau, S. N. (2006). Diagnosis of α -1-antitrypsin deficiency: an algorithm of quantification, genotyping, and phenotyping. *Clinical chemistry*, 52(12), 2236-2242.
- Spielmann, M., & Kircher, M. (2022). Computational and experimental methods for classifying variants of unknown clinical significance. *Molecular Case Studies*, 8(3), a006196.
- Stoller, J. K., & Aboussouan, L. S. (2012). A review of α 1-antitrypsin deficiency. *American journal of respiratory and critical care medicine*, 185(3), 246-259.
- Supriya, S., & Sunkar, S. (2026). Molecular Modeling and Dynamics-Based Evaluation of siRNA-Ago2 Interaction Targeting SERPINA1 in Alpha-1 Antitrypsin Deficiency. *In Silico Research in Biomedicine*, 100479.
- Szczepski, K., & Jaremko, Ł. (2025). AlphaFold and what is next: bridging functional, systems and structural biology. *Expert Review of Proteomics*, 22(2), 45-58.
- Tanoli, Z., Fernández-Torras, A., Özcan, U. O., Kushnir, A., Nader, K. M., Gadiya, Y., Fiorenza, L., Ianevski, A., Vähä-Koskela, M., & Miihkinen, M. (2025). Computational drug repurposing: approaches, evaluation of in silico resources and case studies. *Nature Reviews Drug Discovery*, 24(7), 521-542.

Tiensuu, H., Haapalainen, A. M., Tissarinen, P., Pasanen, A., Määttä, T. A., Huusko, J. M., Ohlmeier, S., Bergmann, U., Ojaniemi, M., & Muglia, L. J. (2022). Human placental proteomics and exon variant studies link AAT/SERPINA1 with spontaneous preterm birth. *BMC medicine*, 20(1), 141.

Tsutsui, Y. (2008). *Exploring Functional and Folding Energy Landscapes by Hydrogen-deuterium Exchange Mass Spectrometry* Case Western Reserve University].

Wang, J. M., Han, M. K., & Labaki, W. W. (2021). Chronic obstructive pulmonary disease risk assessment tools: is one better than the others? *Current opinion in pulmonary medicine*, 28(2), 99.

Wanner, A., & Sandhaus, R. A. (2015). *Alpha-1 antitrypsin: role in health and disease*. Humana Press.

Yang, S.-R., & Kim, H.-R. (2025). Next-Generation Regenerative Therapies for Alpha-1 Antitrypsin Deficiency: Molecular Pathogenesis to Clinical Translation. *International Journal of Molecular Sciences*, 26(17), 8504.