MDPI

Article

Prediction of the Effects of Missense Mutations on Human Myeloperoxidase Protein Stability Using In Silico Saturation Mutagenesis

Adebiyi Sobitan [1], William Edwards [1], Md Shah Jalal [1], Ayanfeoluwa Kolawole [1], Hemayet Ullah [1], Atanu Duttaroy [1], Jiang Li [2] and Shaolei Teng [1,*]

[1] Department of Biology, Howard University, Washington, DC 20059, USA
[2] Department of Electrical Engineering and Computer Science, Howard University, Washington, DC 20059, USA
* Correspondence: shaolei.teng@howard.edu

Abstract: Myeloperoxidase (MPO) is a heme peroxidase with microbicidal properties. MPO plays a role in the host's innate immunity by producing reactive oxygen species inside the cell against foreign organisms. However, there is little functional evidence linking missense mutations to human diseases. We utilized in silico saturation mutagenesis to generate and analyze the effects of 10,811 potential missense mutations on MPO stability. Our results showed that ~71% of the potential missense mutations destabilize MPO, and ~8% stabilize the MPO protein. We showed that G402W, G402Y, G361W, G402F, and G655Y would have the highest destabilizing effect on MPO. Meanwhile, D264L, G501M, D264H, D264M, and G501L have the highest stabilization effect on the MPO protein. Our computational tool prediction showed the destabilizing effects in 13 out of 14 MPO missense mutations that cause diseases in humans. We also analyzed putative post-translational modification (PTM) sites on the MPO protein and mapped the PTM sites to disease-associated missense mutations for further analysis. Our analysis showed that R327H associated with frontotemporal dementia and R548W causing generalized pustular psoriasis are near these PTM sites. Our results will aid further research into MPO as a biomarker for human complex diseases and a candidate for drug target discovery.

Keywords: myeloperoxidase (MPO); in silico saturation mutagenesis; missense mutations; post-translational modification (PTM) sites; protein stability

1. Introduction

Myeloperoxidase (MPO) is a member of the human peroxidase family of proteins expressed mainly in immune cells such as neutrophils. This family of proteins catalyzes the breakdown of peroxides and is involved in killing foreign bodies [1]. Specifically, invading bacteria enhances their production of hydrogen peroxide. MPO uses hydrogen peroxide to produce chloramine and hypochlorite in the peroxidase or halogenation cycle. These chemicals are cytotoxic to invading bacteria. MPO also uses hydrogen peroxidase in the oxidation of chloride ions to hypochlorous acid [2]. Previous studies have shown that MPO has a polycationic nature; therefore, MPO has a high affinity for cell surfaces ranging from immune cells to lipoproteins and epithelial cells [3]. MPO binding to these surfaces can cause functional changes. For example, the binding of MPO to platelets causes the reorganization of the platelet cytoskeleton and thus alters the aggregation properties [3].

Many heritable missense mutations identified from past research are now known to result in human disease phenotypes. These missense mutations cause disease by disrupting the biosynthesis of MPO [4]. Myeloperoxidase deficiency (MPOD) is the most common neutrophil biochemical mutation that manifests in the phenotype as a lack of peroxidase activity [5]. MPOD was considered a rare disease, but recent studies have shown that

MPOD has a high incidence rate and prevalence. In the United States and Europe, MPOD affects 1 in 2000–4000 people [6]. Patients that exhibit MPO deficiency have limited ability to attack and kill pathogenic invaders. The R569W mutation leads to malformation of the MPO protein and the inability of neutrophil cells to complete the biosynthetic pathway of the MPO molecule by not inserting a heme into MPO precursors [7], thus stifling posttranslational processing and stopping the production of enzymatically active MPO species [7]. MPO missense mutations like R548W and R590C are associated with skin conditions, the most common being pustular skin disease and psoriasis [8,9]. Still, other identified missense mutations can cause cognitive diseases—for example, R327H in individuals with frontotemporal dementia and C285G and D403E having genetic associations with autism spectrum disorders [10–12].

Research in loss-of-function variants and linking these mutations to disease is central toward the future growth of genomic studies—especially those focused on computational analysis. Constructing quality databases using various tools and models of post-translational analysis, genome-wide association studies (GWAS), and targeted computational genomic analyses allow research to treat and possibly prevent future diseases. Unlike SNVs, loss-of-function variants are not well understood, and more research in computational biology is necessary to assist genetic data-driven approaches for application in human health [13].

Respiratory burst within neutrophils mediates the ejection of MPO, which is then catalyzed by hydrogen peroxide (H_2O_2). Upon the release of hydrogen peroxide, MPO oxidizes chlorine to hypochlorous acid (HOCl) [14]. Oxidative phosphorylation leads to the production of reactive oxygen species (ROS), which are products of MPO activities. However, phosphorylation of MPO has not been well-studied. Global analysis and mass spectrometry of MPO derived from neutrophils identified five different types of N-glycans or complex sugars at specific locations [15]. N-glycosylation of MPO at certain residues contributes to its functions. In addition to N-glycosylation sites, recent advancement in MS had led to the identification of the heterogeneity of the glycosylation sites on MPO [16].

The identification of post-translational modification (PTM) sites through experiments is a well-established process. In addition, most of the available PTM site data were extracted through experiments. However, the process of identifying PTM sites using experiments is labor-intensive and time-consuming. Experiments are limited by the availability of enzymatic reactions, which also require optimization. A promising alternative is a computational approach, which will reduce the labor and time involved. However, due to the inability to mimic in-vivo conditions, computational approaches are used as a preliminary approach. With the development of more rigorous algorithms, the prediction of PTM sites through a computational approach will be more reliable and accurate. In this study, we used Foldx, a computational tool, to analyze the effect of all possible missense mutations of MPO on its stability. Then, we investigated the functional impact of target mutations altering protein stability.

2. Materials and Methods

2.1. Sequence Selection and Structure Preparation

We selected the sequence isoform chosen as the canonical sequence, Isoform H17 (identifier: P05164-1) [17]. This entry has 745 residues, and no missing residues were reported in the UniProtKB (www.uniprot.org accessed on 20 July 2021) [18]. For the structure, we used the cryogenic crystal structure of human myeloperoxidase Isoform C (PDB ID: 1CXP) from the RCSB Protein Data Bank (PDB) [19]. This structure was crystallized in a previous study through an X-ray method with a high resolution of 1.8 A [19]. We further modify the PDB ID: 1CXP, which is a heterotetramer, into a homodimer.

2.2. Saturated Computational Mutagenesis and Structure-Based Energy Calculation

We applied a Perl programming language in-house script to substitute each residue on the MPO structure to 19 other residues. We used the list of mutations to perform structure-

based energy calculations on Foldx [20]. Foldx is a precise protein tool for predicting protein stability changes upon mutations, and its performance accuracy outperformed other prediction tools [21]. Foldx uses the 'RepairPdb' command to allow the protein structure to assume its native state. This optimizes the protein structure for further analysis. For stability analysis, Foldx uses the 'BuildModel' command to calculate the difference between the folding energy of the wild-type structure and the mutant structure. We then used the resulting folding energy change ($\Delta\Delta G$) to categorize the effect of each missense mutation. The mathematical equation for calculating folding energy change ($\Delta\Delta G$) is:

$$\Delta\Delta G_{\text{(stability)}} = \Delta G_{\text{(folding)}} \text{MUT} - \Delta G_{\text{(folding)}} \text{WT}$$

A positive $\Delta\Delta G$ means the missense mutation destabilizes the protein. A negative $\Delta\Delta G$ means the missense mutation stabilizes the protein. Comparatively, the DDGs derived from Foldx deviate from the experimental $\Delta\Delta G$s by 0.46 kcal/mol (~0.5 kcal/mol) [22]. In addition, 2.5 kcal/mol was identified as the best cutoff for the corrections between functional changes and $\Delta\Delta G$ [23]. Therefore, we placed the effects of missense mutations on protein stability into five categories, in 0.5 kcal/mol increments. These are highly stabilizing ($\Delta\Delta G < -2.5$ kcal/mol), moderately stabilizing ($-2.5 < \Delta\Delta G < -0.5$ kcal/mol), neutral ($0.5 < \Delta\Delta G < +0.5$ kcal/mol), moderately destabilizing ($+0.5 < \Delta\Delta G < 2.5$ kcal/mol), and highly destabilizing ($\Delta\Delta G > 2.5$ kcal/mol).

2.3. Sequence-Based Prediction of PTM Sites

We performed the prediction of PTM sites on the sequence of the MPO protein using novel algorithms developed by the CUCKOO Workgroup (www.biocuckoo.org accessed on 20 July 2021). These algorithms have been optimized and improved upon to enhance their PTM site prediction performances. Here, we predicted phosphorylation sites, sumoylation sites, and methylation sites using the default medium threshold. We used the NetOGlyc 4.0 [24] and NetNGlyc 1.0 [25] online servers to predict O-linked glycosylation and N-linked glycosylation sites, respectively. The threshold used for identifying positive sites is 0.5 for both N-linked and O-linked glycosylation site prediction. Finally, we used the MusiteDeep webserver to predict different PTM types using the default threshold (0.5) to minimize false positives [26].

2.4. Sequence-Based Pathogenicity Prediction

The tools we used for pathogenicity prediction of each computed missense mutation are the Polymorphism Phenotyping v2 (PolyPhen2) [27], Sorting Intolerant From Tolerant (SIFT) [28], and Screening for non-acceptable polymorphisms (SNAP) [29]. The webservers of PolyPhen2, SIFT, and SNAP accept FASTA sequence of the MPO protein as input. These are robust tools that accept and annotate large volume of data. We used the Batch query for all three webservers to analyze all possible missense mutations.

We used R programming language (https://www.r-project.org/ accessed on 20 July 2021) to perform analysis of variance (ANOVA) among the five $\Delta\Delta G$ categories and PolyPhen2, SIFT, and SNAP scores.

2.5. Detection of PTM-SNPs

We used a Perl script to curate the PTM and missense mutation data. We queried the dbNSFP for additional functional annotation of the missense Single Nucleotide Polymorphism (SNP) data. The dbNSFP is a comprehensive database for Non-synonymous SNPs and their Functional predictions in the human genome [30]. The Perl Express version 2.5 is available at http://www.perl.org (accessed on 20 July 2021).

We mapped the position of the missense SNP site and PTM site on the canonical FASTA sequence of the MPO protein to obtain the relative distance between the SNP site and the PTM site. We specified a window length of −7 aa to +7 aa. The + or − sign of the relative distance indicates if the SNP site is downstream or upstream of the PTM site, respectively.

3. Results

3.1. Analysis of Overall Computed Missense Mutations

Computational mutagenesis of the MPO protein generated 10,811 non-redundant missense mutations from 569 residues. In the experimental structure, C316 was omitted due to a modification [31]. FoldX uses an energy calculation approach to categorize the potential effect of each missense mutation. Of the potential non-redundant mutations, 4221 (~39%) mutations have a highly destabilizing effect ($\Delta\Delta G > 2.5$ kcal/mol), 3486 (~32%) mutations have a destabilizing effect ($0.5 < \Delta\Delta G \leq 2.5$ kcal/mol), 2248 (~21%) mutations have a neutral effect ($-0.5 < \Delta\Delta G \leq 0.5$ kcal/mol), 811 (~8%) mutations have a stabilizing effect ($-2.5\ \Delta\Delta G \leq -0.5$ kcal/mol), and 45 (~0.4%) mutations have a highly stabilizing effect ($\Delta\Delta G < -2.5$ kcal/mol) on the MPO protein (Figure 1). Furthermore, our analysis shows that missense mutations have more highly destabilizing effects within the Myeloperoxidase_like domain (41%) than outside the Myeloperoxidase_like domain (33%).
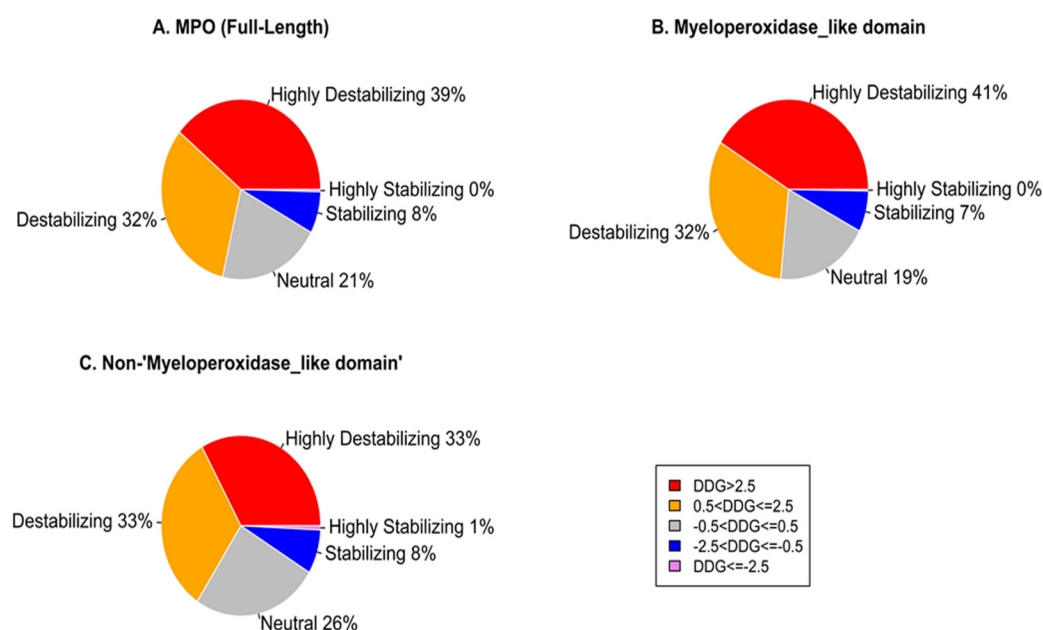

Figure 1. Distribution of the effects of missense mutations on MPO stability. (**A**) residues 167−744 of the whole sequence of 1CXP.pdb; (**B**) residues 317−728 of the Myeloperoxidase_like domain; (**C**) residues 167−315 & 729−744 outside the Myeloperoxidase_like domain.

3.2. Top Missense Mutations Affecting MPO Stability

The result of the computational mutagenesis shows the magnitude of the folding energy change caused by each missense mutation. The mean $\Delta\Delta G$ values of the folding energy change caused by all possible missense mutations on a residue, and $\Delta\Delta G$ values of alanine substitution were showed in Figure 2A. Figure 2B shows the top folding energy changes caused by individual missense mutations. The residues G402, G361, G655, G595, and G189 have the highest mean $\Delta\Delta G$ values among all MPO amino acids. Meanwhile, the residues D264, D260, G501, G615, and T428 have the lowest mean $\Delta\Delta G$ values in MPO residues. The top missense mutations stabilize and destabilize the MPO protein significantly. As shown in Figure 2B, the overall top destabilizing missense mutations G402W, G402Y, G361W, G402F, and G655Y caused a $\Delta\Delta G$ of 83.11 kcal/mol, 81.0 kcal/mol, 75.1 kcal/mol, 70.33 kcal/mol, and 62.39 kcal/mol, respectively. Interestingly, three of these top destabilizing mutations are on residue G402, and these mutations involve the substitution of the glycine residue. The presence of glycine in these positions must contribute to the stability of the MPO protein. Also shown in Figure 2B, the overall top stabilizing missense mutations D264L, G501M, D264H, D264M, and G501L caused a $\Delta\Delta G$ of −4.82 kcal/mol, −4.25 kcal/mol, −4.18 kcal/mol, −3.84 kcal/mol, and −3.79 kcal/mol, respectively.
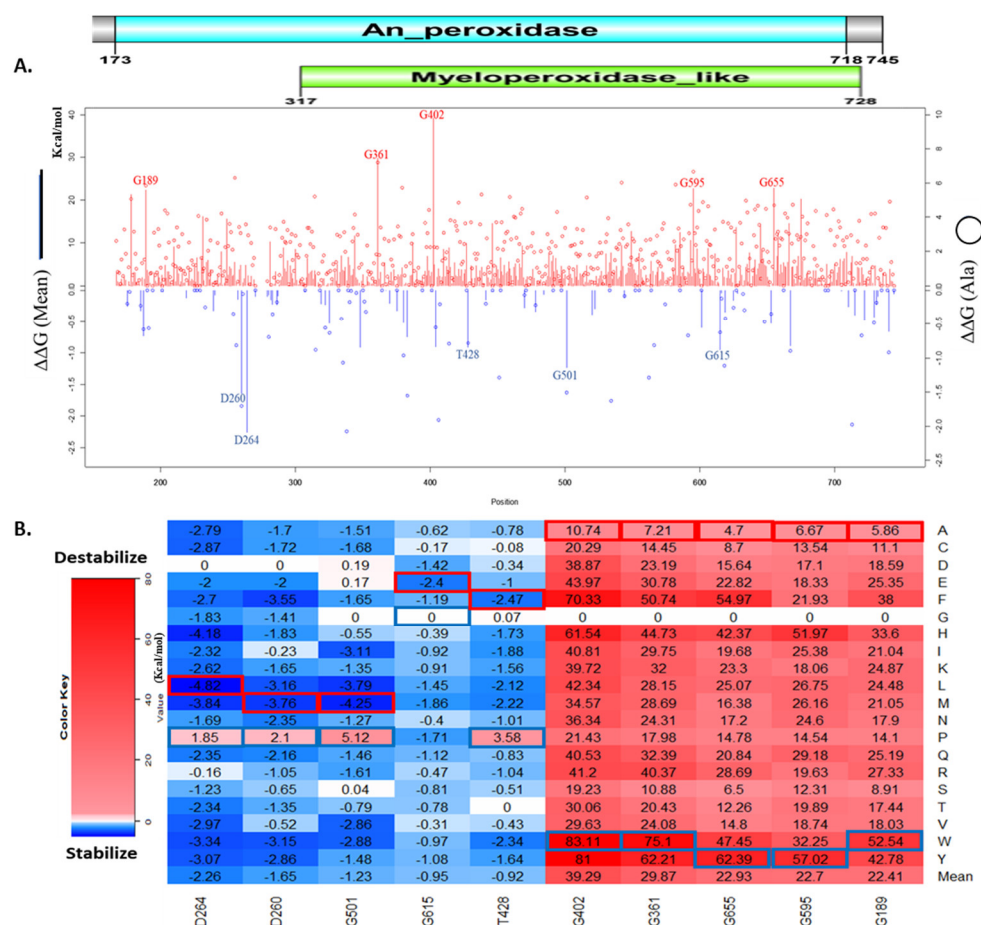
Figure 2. Effects of missense mutations on MPO stability. (**A**) line chart of residues 167−744 showing their mean folding energy changes (ΔΔG) in vertical bars and ΔΔG of alanine substitutions in circles. The An_peroxidase (Aqua bar) and the Myeloperoxidase (green bar) domains represent the superfamily and family domains, respectively; (**B**) heatmap of top 5 stabilizing and top 5 destabilizing mean ΔΔG. The blue rectangle represents maximum values and the red rectangle represents minimum values for each residue position. Key residues are highlighted on the line chart and heatmap.

We used PyMol to observe the spatial changes caused by key missense mutations on the MPO structure, as shown in Figure 3. In addition, there is an increase in Van der Waals (steric) strain and clashes in G361W, G402W, and G402Y. G402W had eight possible rotamers with steric strains ranging from 92.84 to 129.27. G402Y has five possible rotamers with steric strains ranging from 102.21 to 129.58. Our Foldx output also shows reduction in the sidechain hydrogen bond, backbone hydrogen bond, and Van der Waals attraction of the missense mutations shown in Figure 3 (Supplementary Table S1).

3.3. Sequence-Based Analysis of the Effects of Missense Mutations on MPO Stability

We analyzed the damaging and pathogenic impact of missense mutations on the MPO protein sequence. Figure 4 shows boxplots comparing sequence-based predictions against structure-based predictions. The five groups are statistically different from one another (p-value $< 2 \times 10^{-16}$). We observe that the highly destabilizing (ΔΔG > 2.5) category has the most pathogenic and damaging impact on the MPO protein, while the neutral ($-0.5 < \Delta\Delta G \leq 0.5$) category has the least pathogenic and damaging impact.
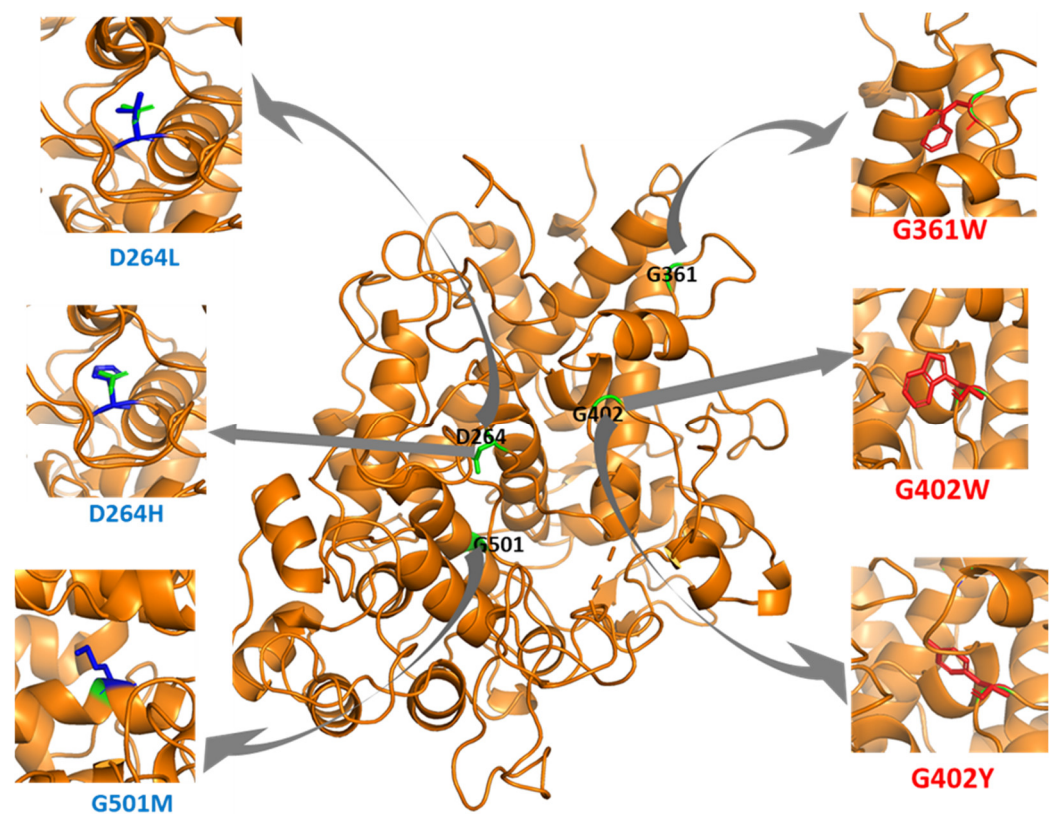
Figure 3. Structural representation of some key missense mutations. Missense mutations in red destabilize the MPO protein (**Right**); Missense mutations in blue stabilize the MPO protein (**Left**); Green represents the wild-type residues.

Table 1 shows the structure-based and sequence-based prediction effects of ten missense mutations on MPO stability. Polyphen2, SNAP, and SIFT predicted the top five highly destabilizing missense mutations, predicted by Foldx, to affect the MPO protein. The SNAP tool predicted that the top five stabilizing missense mutations would affect MPO protein function. On the other hand, SIFT tool predicted D264L and D264H as tolerant missense mutations, while Polyphen2 predicted D264H as benign.

3.4. Computational Predictions of Disease Phenotypes

We searched through the HGMD database and found 14 missense mutations associated with the MPO protein as shown in Table 2. Scientists have determined these mutations through experiments to cause disease phenotypes in humans. Foldx predicts all the disease-causing mutations to significantly destabilize the MPO structure except G501S, which increased the folding energy of MPO by only 0.04 kcal/mol. Of the 14 HGMD missense mutations, PolyPhen2 predicts A332V, C285G, and M251T as benign. SIFT predicts D371G and A332V as tolerant mutations. Lastly, SNAP predicts that only A332V does not affect MPO stability.

Noticeably, four distinct complex diseases were associated with missense mutations of the MPO protein. For example, R548W, R569W, R327H, and C285G were implicated in general pustular psoriasis (GPP), MPOD, Frontotemporal dementia (FTD) and autism spectrum disorder (ASD), respectively. Based on our computational predictions, R548W, R569W, R327H, and C285G increased the folding energy of MPO by >3 kcal/mol, thereby destabilizing the MPO structure.
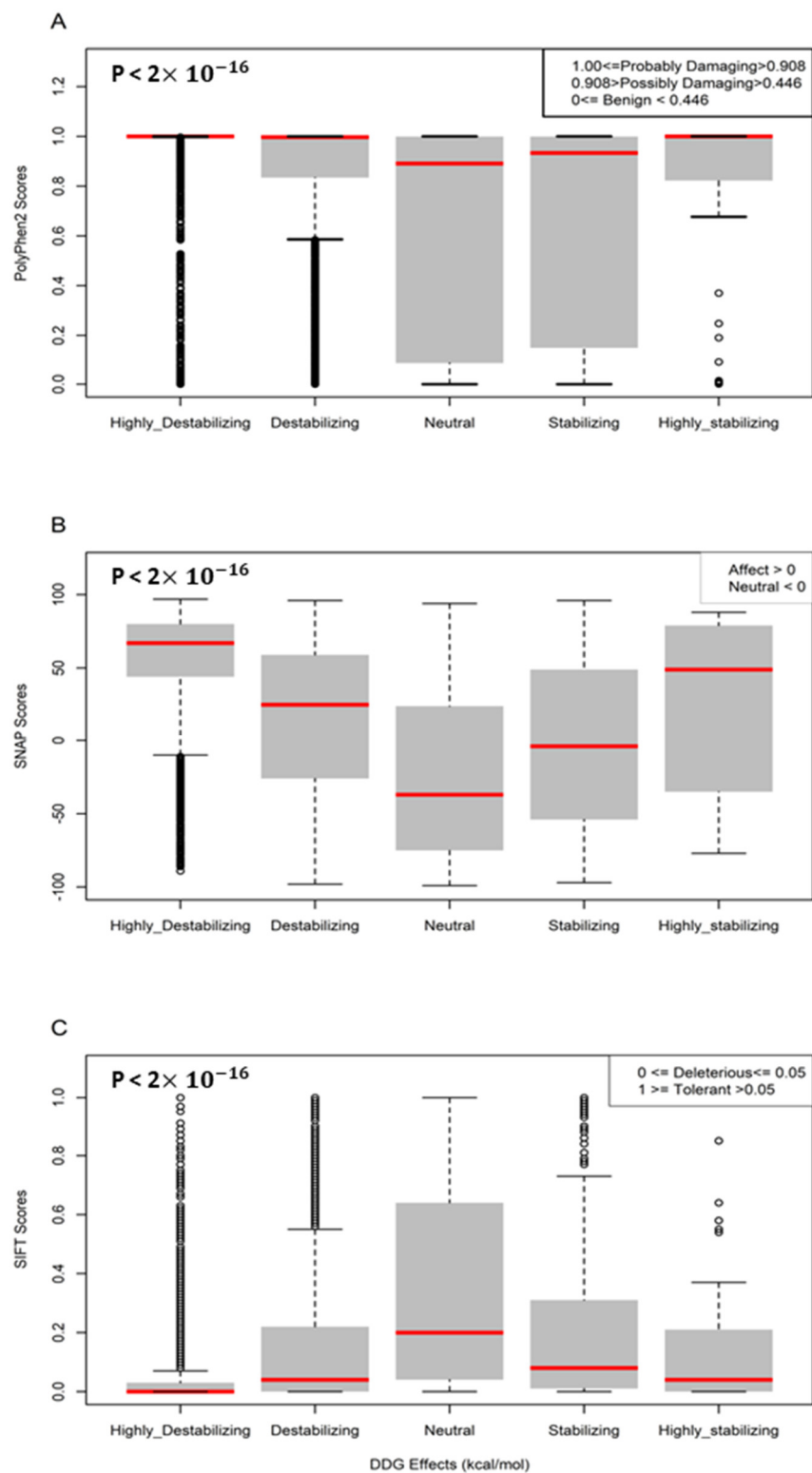
Figure 4. Boxplots of the prediction of mutation pathogenicity. (**A**) PolyPhen2 scores; (**B**) SNAP scores; and (**C**) SIFT scores against five categories of the effects of missense mutations on MPO protein stability.

Table 1. Functional effects of top ten mutations altering MPO protein stability.

| Missense Mutations | Structure-Based Prediction (Protein Stability) | | | Sequence-Based Prediction (Mutation Pathogenicity) | | | | | |
| | FoldX | | PolyPhen2 | | SIFT | | SNAP | | |
	ΔΔG (Kcal/mol)	Mutation Effect	Score	Effect	Score	Effect	Score	Effect	
G402W	83.11	H_Des	1	Pro_d	0	Del	80	Aff	
G402Y	81	H_Des	1	Pro_d	0	Del	75	Aff	
G361W	75.1	H_Des	1	Pro_d	0	Del	93	Aff	
G402F	70.33	H_Des	1	Pro_d	0	Del	77	Aff	
G655Y	62.39	H_Des	1	Pro_d	0.01	Del	84	Aff	
D264L	−4.82	H_Sta	0.972	Pro_d	0.11	Tol	20	Aff	
G501M	−4.25	H_Sta	1	Pro_d	0	Del	85	Aff	
D264H	−4.18	H_Sta	0.191	B	0.09	Tol	11	Aff	
D264M	−3.84	H_Sta	0.995	Pro_d	0.05	Del	44	Aff	
G501L	−3.79	H_Sta	1	Pro_d	0	Del	84	Aff	

Note: H_Des = Highly Destabilizing, H_Sta = Highly Stabilizing. Pro_d = Probably damaging, B = Benign, Del = Deleterious, Tol = Tolerated, Aff = Affect protein.

Table 2. Sequence-based and structure-based prediction of disease-associated missense mutations.

| Missense Mutations | HGMD | Structure-Based Prediction (Protein Stability) | | Sequence-Based Prediction (Mutation Pathogenicity) | | | | | |
| | | FoldX | | PolyPhen2 | | SIFT | | SNAP | |
	Phenotype	ΔΔG (Kcal/mol)	Mutation Effect	Score	Damaging Effect	Score	Effect	Score	Effect
R548W	GPP	11.22	H_Des	1	Pro_d	0	Del	84	Aff
R569W	MPOD	7.84	H_Des	1	Pro_d	0.04	Del	71	Aff
W643R	MPOD	5.92	H_Des	1	Pro_d	0	Del	93	Aff
M251T	MPOD	4.81	H_Des	0.032	B	0.01	Del	35	Aff
R327H	FTD	4.72	H_Des	0.986	Pro_d	0	Del	57	Aff
C285G	ASD	3.31	H_Des	0.239	B	0	Del	74	Aff
Y173C	MPOD	3.08	H_Des	0.995	Pro_d	0	Del	71	Aff
D371G	MPOD	2.25	Des	0.999	Pro_d	0.12	Tol	64	Aff
R499C	MPOD	1.62	Des	0.999	Pro_d	0	Del	62	Aff
R590C	GPP	1.32	Des	1	Pro_d	0	Del	39	Aff
D403E	ASD	1.27	Des	1	Pro_d	0	Del	13	Aff
A332V	MPOD	1.07	Des	0.059	B	0.1	Tol	−26	N
L572W	MPOD	0.51	Des	1	Pro_d	0	Del	73	Aff
G501S	MPOD	0.04	N	0.938	Pro_d	0	Del	23	Aff

Note: MPOD = Myeloperoxidase deficiency, FTD = Frontotemporal dementia, ASD = Autism spectrum disorder, GPP = generalized pustular Psoriasis. H_Des = Highly Destabilizing, N = Neutral, Des = Destabilizing. Pro_d = Probably damaging, B = Benign, Del= Deleterious, Tol = Tolerated, Aff = Affect protein.

Figure 5 highlights the top destabilizing missense mutations associated with the four complex diseases. Arginine (R) substitution with Tryptophan (W) in positions 548 and 569 alters the local environment by constricting neighboring residues. The change in orientation of the sidechains of C285 and R327 due to substitution with Glycine (G) and Histidine (H), respectively, alters their local environment leading to instability of the MPO protein.

3.5. Effect of Missense Mutations on Post-Translational Modification Sites on MPO

We used MusiteDeep predictions for our final analysis. MusiteDeep predicted 14 PTM sites within residues 167–745, five phosphorylation sites, two ubiquitination sites, one pyrrolidone carboxylic acid site, five N-Glycosylation sites, and one acetylation site. Figure 6 shows that S231 has the highest mean folding energy change (ΔΔG), 16.23 kcal/mol, while N729 has the least mean ΔΔG, −0.43 kcal/mol. All possible predicted missense mutations on S231 increase the folding energy of the wildtype MPO protein by at least 0.83 kcal/mol. The mutation S231Y increases ΔG by 49.81 kcal/mol, and N729P decreases ΔG by −1.94 kcal/mol. The ubiquitination site K556 shows the second-highest mean ΔΔG of 3.97 kcal/mol. Of 400 missense mutations predicted on 14 PTM sites on MPO, S231Y and N729P produce the highest and lowest ΔΔG, respectively.
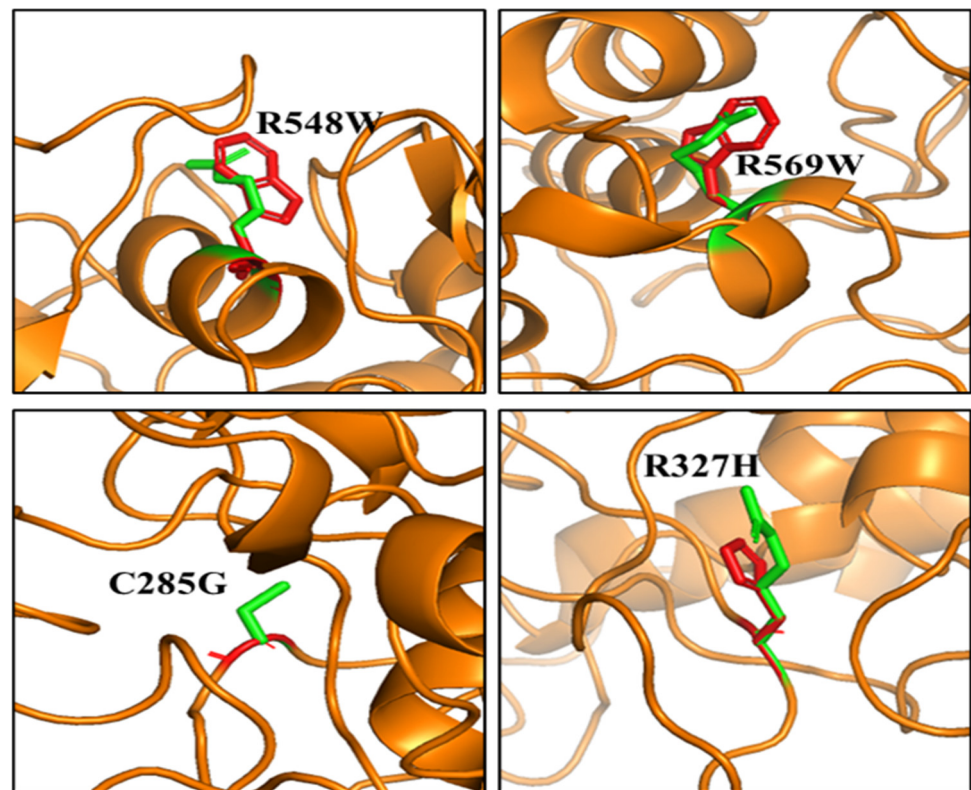
Figure 5. Structural representation of some disease-associated missense mutations of MPO; wildtype residues in Green; mutant residues in Red.
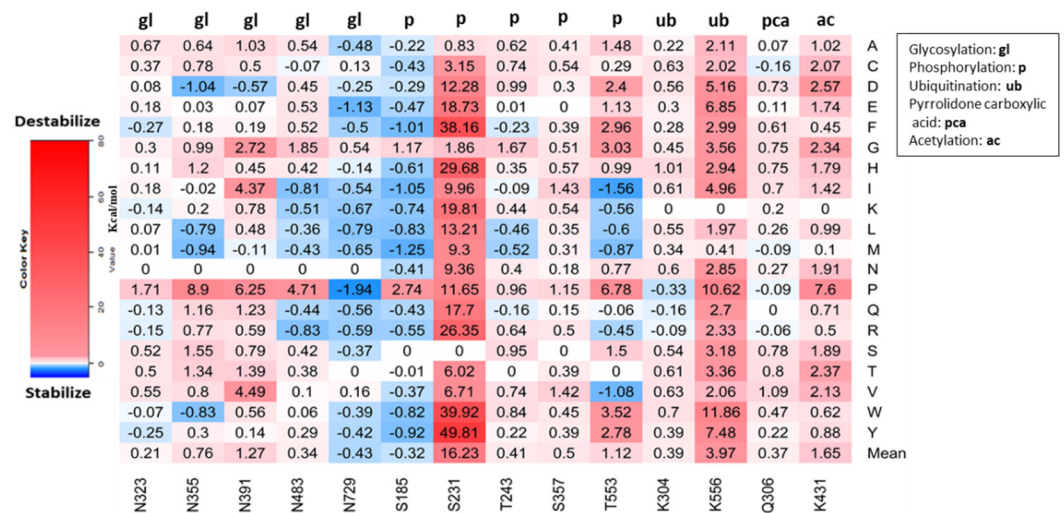
	gl	gl	gl	gl	gl	p	p	p	p	p	ub	ub	pca	ac	
	0.67	0.64	1.03	0.54	-0.48	-0.22	0.83	0.62	0.41	1.48	0.22	2.11	0.07	1.02	A
	0.37	0.78	0.5	-0.07	0.13	-0.43	3.15	0.74	0.54	0.29	0.63	2.02	-0.16	2.07	C
	0.08	-1.04	-0.57	0.45	-0.25	-0.29	12.28	0.99	0.3	2.4	0.56	5.16	0.73	2.57	D
	0.18	0.03	0.07	0.53	-1.13	-0.47	18.73	0.01	0	1.13	0.3	6.85	0.11	1.74	E
Destabilize	-0.27	0.18	0.19	0.52	-0.5	-1.01	38.16	-0.23	0.39	2.96	0.28	2.99	0.61	0.45	F
	0.3	0.99	2.72	1.85	0.54	1.17	1.86	1.67	0.51	3.03	0.45	3.56	0.75	2.34	G
	0.11	1.2	0.45	0.42	-0.14	-0.61	29.68	0.35	0.57	0.99	1.01	2.94	0.75	1.79	H
	0.18	-0.02	4.37	-0.81	-0.54	-1.05	9.96	-0.09	1.43	-1.56	0.61	4.96	0.7	1.42	I
	-0.14	0.2	0.78	-0.51	-0.67	-0.74	19.81	0.44	0.54	-0.56	0	0	0.2	0	K
	0.07	-0.79	0.48	-0.36	-0.79	-0.83	13.21	-0.46	0.35	-0.6	0.55	1.97	0.26	0.99	L
	0.01	-0.94	-0.11	-0.43	-0.65	-1.25	9.3	-0.52	0.31	-0.87	0.34	0.41	-0.09	0.1	M
	0	0	0	0	0	-0.41	9.36	0.4	0.18	0.77	0.6	2.85	0.27	1.91	N
	1.71	8.9	6.25	4.71	-1.94	2.74	11.65	0.96	1.15	6.78	-0.33	10.62	-0.09	7.6	P
	-0.13	1.16	1.23	-0.44	-0.56	-0.43	17.7	-0.16	0.15	-0.06	-0.16	2.7	0	0.71	Q
	-0.15	0.77	0.59	-0.83	-0.59	-0.55	26.35	0.64	0.5	-0.45	-0.09	2.33	-0.06	0.5	R
	0.52	1.55	0.79	0.42	-0.37	0	0	0.95	0	1.5	0.54	3.18	0.78	1.89	S
	0.5	1.34	0.39	0.38	0	-0.01	6.02	0	0.39	0	0.61	3.36	0.8	2.37	T
Stabilize	0.55	0.8	4.49	0.1	0.16	-0.37	6.71	0.74	1.42	-1.08	0.63	2.06	1.09	2.13	V
	-0.07	-0.83	0.56	0.06	-0.39	-0.82	39.92	0.84	0.45	3.52	0.7	11.86	0.47	0.62	W
	-0.25	0.3	0.14	0.29	-0.42	-0.92	49.81	0.22	0.39	2.78	0.39	7.48	0.22	0.88	Y
	0.21	0.76	1.27	0.34	-0.43	-0.32	16.23	0.41	0.5	1.12	0.39	3.97	0.37	1.65	Mean
	N323	N355	N391	N483	N729	S185	S231	T243	S357	T553	K304	K556	Q306	K431	

Legend inset:
Glycosylation: **gl**
Phosphorylation: **p**
Ubiquitination: **ub**
Pyrrolidone carboxylic acid: **pca**
Acetylation: **ac**

Figure 6. Post−translational modification sites on MPO showing calculated folding energy changes (ΔΔG) for all possible reference residues. Legends are inset.

Among the 14 missense mutations in MPO causing disease phenotypes, two are close to PTM sites (Table 3). R327H causes frontotemporal dementia (FTD) in humans, and it is located 4-position downstream (+4) to an N-lined glycosylation site, N323. R548W causes generalized pustular psoriasis (GPP) in humans, and it is located 5-position upstream (−5) to T553, a phosphorylation site.

Table 3. Disease-associated PTM-SNPs on the MPO protein.

rs_dbSNP151	Missense SNP	PTM SITE	PEPTIDE	DISTANCE APART (PTM ↔ SNP SITE)	PTM TYPE	Phenotype
rs760619802	R327H	N323	GSNITIRNQI	+4	N-LINKED GLYCOSYLATION	FTD
rs148802625	R548W	T553	ILRGLMATPA	−5	PHOSPHORYLATION	GPP

Note: SNP represents missense mutation at the protein level, FTD = Frontotemporal dementia, GPP = generalized pustular Psoriasis.

4. Discussion

MPO is present in unicellular to multicellular organisms. A well conserved myeloperoxidase_like domain in the human MPO highlights its importance, and the presence of a more conserved An_peroxidase domain in different organisms offers insight into its essential function. MPO plays a pivotal role in the host's innate and adaptive immunity [32]. MPO causes inflammation when it extravasates into the extracellular environment in high amounts [33]. However, MPO has become a therapeutic target. Some studies found MPOD as beneficial in preventing myocardial infarction [34]. It has been established through experiments that most missense mutations result in a less stable protein [35]. Our study explores the use of computational tools to predict the effect of missense mutations on MPO stability. Protein stability critically affects its function. Based on our predictions, ~71% of all possible mutations will destabilize the MPO protein compared to ~8% stabilizing the MPO protein. Missense mutations increasing protein stability could also alter its function, which could drive disease mechanisms or the evolution of new functions [36].

Experimental assays used in the characterization of disease-associated missense mutations exhibit selective sensitivity [37]. However, computational approaches provide unbiased insight into the functional characterization of disease-associated missense mutations [38,39]. Our structure-based analysis provided the basis for understanding how missense mutations could alter the structure of MPO and ultimately affect its function. In Figure 3, we showed that missense mutations that increase steric clashes would destabilize the structure of MPO. We highlighted the top ten mutations significantly affecting the MPO structure (Table 1). Furthermore, we observed that missense mutations that significantly destabilize or stabilize MPO protein could have damaging effects on its function. G402 and G361 contribute to the hydrophobic core of the MPO structure. Therefore, the substitution of glycine with tryptophan (G402W, G361W) or tyrosine (G402Y) in these positions caused steric clashes preventing the normal folding process of the active MPO protein. The top five most stabilizing missense mutations were in D264 and G501. Substituting aspartic acid with leucine (D264L), methionine (D264M), or histidine (D264H) must have a greater substantial hydrophobic effect on the MPO structure. Similarly, G501M and G501L significantly increase MPO stability. We also used sequence-based tools, Polyphen2, SIFT, and SNAP, for mutation pathogenicity prediction. We showed that the top destabilizing/stabilizing mutations have damaging effects on protein function, except for D264H, which is predicted to have neutral effects (Table 1). In our previous study [40], we discovered these tools are reliable for analyzing the effects of mutations on protein function.

A previous study utilized Foldx for mutagenesis to identify deleterious missense mutations [38]. In our research, Foldx predictions of disease-associated missense mutations of MPO showed destabilizing effects except for G501S. The whole-exome sequencing of families having a child with Autistic disorder (ASD) reveals the contributions of de novo (DN) missense mutations, C285G, to ASD diagnosis in affected probands [41]. Another study analyzed de novo mutations in Whole-exome sequencing (WES) data from 5947 families and found that D403E is a genetic risk of ASD [42]. A study discovered that Y173C contributed to the degradation of MPO in the cell. MPO degrades because Y173C prevents its proteolytic maturation and secretion, leading to MPOD [43]. The biochemical characterization of hereditary MPOD in a father and daughter shows alteration (14-base deletion) in their MPO mRNAs. The presence of M251T in some patients led to late degradation

of the MPO protein after an early granulocyte differentiation [44]. A study found R327H in the MPO protein after a comprehensive investigation in 38 patients with Dementia. In addition, patients with R327H exhibit behavioral changes associated with Alzheimer's disease (AD) [45]. Bone marrow analyses of subjects in Italy with MPOD reveal the structural changes of the MPO protein by A332V, D371G, L572W, and W643R [46]. In two independent studies among Japanese patients, scientists discovered two novel missense mutations, R499C and G501S, associated with MPOD. Although rare, R499C and G501S are proximal to H502, a functionally important residue for iron binding. Further analysis revealed that R499C and G501S disrupt the maturation of the MPO protein [47,48]. Nauseef and others discovered R569W in patients with complete MPOD. Restriction digestion of the gDNA of patients with complete MPOD shows a substitution at exon 10, translating to a substitution at codon 569 [7]. A study discovered R548W and R590C in the blood samples of patients with general pustular psoriasis (GPP). R548W and R590C also completely depleted MPO activity in these patients with GPP [9]. From Table 1, we found these missense mutations to destabilize the MPO protein structure. This structural destabilization might play a role in the disease phenotypes discovered in the different patients.

Polyphen2 predicted all disease-associated missense mutations to be probably damaging except A332V, C285G, and M251T. SIFT scores showed all to be deleterious except D371G and A332V. Lastly, the SNAP scores showed all to affect MPO protein but A332V. While Polyphen2, SIFT, and SNAP predicted A332V as not damaging or deleterious, Foldx predicted A332V as destabilizing. In a previous study, the expression of A332V presented an asymptomatic phenotype with partial MPO activity. SIFT, Polyphen2, and SNAP tools rely on the characterization of known protein sequences in databases and are limited in predicting the damaging effects of missense mutations. However, Foldx utilizes the structural information for predicting the protein stability changes upon mutations.

Whether MPO undergoes other post-translational modifications is not well known, However, MPO undergoes N-linked glycosylation [14]. With the deep learning framework, MusiteDeep, we predicted five different types of post-translational modifications on the MPO protein. We analyzed the effects of all possible missense mutations on the predicted PTM sites and found that mutation of the phosphorylation site, S231, has the most significant impact on destabilizing the MPO protein. T553, another phosphorylation site, could play an important role in modifying MPO. A study demonstrated that the catalytic activity of MPO is regulated through phosphorylation [49]. The substitution of asparagine with proline in all five N-glycosylated sites had significant impacts on the folding energy of MPO. Glycosylation increases the catalytic activity of MPO, and substitution with proline could lead to a decrease or absence of catalytic activity. The ubiquitination site, K556, showed a significant increase in folding energy when mutated. All possible missense mutations at K556 increased the folding energy of the wild-type MPO protein. The mutation of K556 might lead to abnormal or absence of MPO degradation. MusiteDeep also predicted that an acetylation site, K431. Our result showed that mutation of K431 would destabilize the MPO protein by changing its overall charge.

Missense mutations can disrupt PTM sites, thereby leading to diseases [50]. We provided additional analysis on how missense mutations can affect PTM. We assumed that the location of a missense mutation relative to the PTM site would influence the outcome of PTM. We mapped disease-associated missense mutations with our predicted PTMs, and we found two disease-associated PTM-SNPs, which indicates a correlation. The inhibition of T553 and N323 on MPO could influence the progression of generalized pustular psoriasis and frontotemporal dementia (FTD), respectively.

5. Conclusions

Based on our predictions, over 70% of all possible mutations will destabilize the MPO protein compared to 8% stabilizing missense mutations. The residues, G402 and G361, contribute to the hydrophobic core of the MPO structure, and their mutations destabilize MPO. Conversely, mutations of D264 and G501 significantly increase MPO stability.

Computational analysis of disease-associated missense mutations provides compelling evidence that Foldx is a reliable tool. Computational prediction of missense mutations on PTM sites on MPO revealed the vital role of the different PTM types on the MPO protein. Furthermore, missense mutations located close to PTM sites T553 and N323 on MPO could influence the progression of generalized pustular psoriasis and frontotemporal dementia (FTD), respectively. However, in-vitro analysis of these computational missense mutations is required.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes13081412/s1, Table S1: The energy changes introduced by the target variants.

References

1. Aratani, Y. Myeloperoxidase: Its role for host defense, inflammation, and neutrophil function. In *Archives of Biochemistry and Biophysics*; Academic Press Inc.: Cambridge, MA, USA, 2018; Volume 640, pp. 47–52.
2. Khan, A.; Alsahli, M.; Rahmani, A. Myeloperoxidase as an Active Disease Biomarker: Recent Biochemical and Pathological Perspectives. *Med. Sci.* **2018**, *6*, 33. [CrossRef] [PubMed]
3. Gorudko, I.V.; Sokolov, A.V.; Shamova, E.V.; Grigorieva, D.V.; Mironova, E.V.; Kudryavtsev, I.V.; Gusev, S.A.; Gusev, A.A.; Chekanov, A.V.; Vasilyev, V.B.; et al. Binding of human myeloperoxidase to red blood cells: Molecular targets and biophysical consequences at the plasma membrane level. *Arch. Biochem. Biophys.* **2016**, *591*, 87–97. [CrossRef] [PubMed]
4. Nauseef, W.M.; McCormick, S.; Goedken, M. Redox Report Communications in Free Radical Research Impact of missense mutations on biosynthesis of myeloperoxidase. *Redox Rep.* **2013**, *5*, 187–206. [CrossRef]
5. Marchetti, C.; Patriarca, P.; Solero, G.P.; Baralle, F.E.; Romano, M. Genetic studies on myeloperoxidase deficiency in Italy. *Jpn. J. Infect. Dis.* **2004**, *57*, S10–S12.
6. Parry, M.F.; Root, R.K.; Metcalf, J.A.; Delaney, K.K.; Kaplow, L.S.; Richar, W.J. Myeloperoxidase deficiency: Prevalence and clinical significance. *Ann. Intern. Med.* **1981**, *95*, 293–301. [CrossRef]
7. Nauseef, W.M.; Cogley, M.; McCormick, S. Effect of the R569W missense mutation on the biosynthesis of myeloperoxidase. *J. Biol. Chem.* **1996**, *271*, 9546–9549. [CrossRef]
8. Vergnano, M.; Grys, K.; Benzian-Olsson, N.; Mahil, S.K.; Chaloner, C.; Barbosa, I.A.; August, S.; Burden, A.D.; Choon, S.E.; Cooper, H.; et al. Loss-of-function myeloperoxidase mutations are associated with increased neutrophil counts and pustular skin disease. *Am. J. Hum. Genet.* **2020**, *107*, 539–543. [CrossRef]
9. Haskamp, S.; Bruns, H.; Hahn, M.; Hoffmann, M.; Gregor, A.; Löhr, S.; Hahn, J.; Schauer, C.; Ringer, M.; Flamann, C.; et al. Myeloperoxidase Modulates Inflammation in Generalized Pustular Psoriasis and Additional Rare Pustular Skin Diseases. *Am. J. Hum. Genet.* **2020**, *107*, 527–538. [CrossRef]
10. Volkman, R.; Ben-Zur, T.; Kahana, A.; Garty, B.Z.; Offen, D. Myeloperoxidase Deficiency Inhibits Cognitive Decline in the 5XFAD Mouse Model of Alzheimer's Disease. *Front. Neurosci.* **2019**, *13*, 990. [CrossRef]
11. Zappia, M.; Manna, I.; Serra, P.; Cittadella, R.; Andreoli, V.; Russa, A.L.; Annesi, F.; Spadafora, P.; Romeo, N.; Nicoletti, G.; et al. Increased Risk for Alzheimer Disease With the Interaction of MPO and A2M Polymorphisms. *Arch. Neurol.* **2004**, *61*, 341–344. [CrossRef]

12. Jiang, Y.H.; Yuen, R.K.C.; Jin, X.; Wang, M.; Chen, N.; Wu, X.; Ju, J.; Mei, J.; Shi, Y.; He, M.; et al. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am. J. Hum. Genet.* **2013**, *93*, 249–263. [CrossRef] [PubMed]

13. Pagel, K.A.; Pejaver, V.; Lin, G.N.; Nam, H.-J.; Mort, M.; Cooper, D.N.; Sebat, J.; Iakoucheva, L.M.; Mooney, S.D.; Radivojac, P. When loss-of-function is loss of function: Assessing mutational signatures and impact of loss-of-function genetic variants. *Bioinformatics* **2017**, *33*, i389–i398. [CrossRef]

14. Antwerpen, P.V.; Slomianny, M.-C.; Boudjeltia, K.Z.; Delporte, C.; Faid, V.; Calay, D.; Rousseau, A.; Moguilevsky, N.; Raes, M.; Vanhamme, L.; et al. Glycosylation Pattern of Mature Dimeric Leukocyte and Recombinant Monomeric Myeloperoxidase: Glycosylation is required for optimal enzymatic activity. *J. Biol. Chem.* **2010**, *285*, 16351–16359. [CrossRef] [PubMed]

15. Reiding, K.R.; Franc, V.; Huitema, M.G.; Brouwer, E.; Heeringa, P.; Heck, A.J.R. Neutrophil myeloperoxidase harbors distinct site-specific peculiarities in its glycosylation. *J. Biol. Chem.* **2019**, *294*, 20233–20245. [CrossRef] [PubMed]

16. Cieutat, A.M.; Lobel, P.; August, J.T.; Kjeldsen, L.; Sengeløv, H.; Borregaard, N.; Bainton, D.F. Azurophilic Granules of Human Neutrophilic Leukocytes Are Deficient in Lysosome-Associated Membrane Proteins but Retain the Mannose 6-Phosphate Recognition Marker. *Blood* **1998**, *91*, 1044–1058. [CrossRef]

17. UniProt, C. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [CrossRef]

18. Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef]

19. Fiedler, T.J.; Davey, C.A.; Fenna, R.E. X-ray crystal structure and characterization of halide-binding sites of human myeloperoxidase at 1.8 A resolution. *J. Biol. Chem.* **2000**, *275*, 11964–11971. [CrossRef]

20. Delgado, J.; Radusky, L.G.; Cianferoni, D.; Serrano, L. FoldX 5.0: Working with RNA, small molecules and a new graphical interface. *Bioinformatics* **2019**, *35*, 4168–4169. [CrossRef]

21. Buß, O.; Rudat, J.; Ochsenreither, K. FoldX as Protein Engineering Tool: Better Than Random Based Approaches? *Comput. Struct. Biotechnol. J.* **2018**, *16*, 25–33. [CrossRef]

22. Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX web server: An online force field. *Nucleic Acids Res.* **2005**, *33*, W382–W388. [CrossRef] [PubMed]

23. Bromberg, Y.; Rost, B. Correlating protein function and stability through the analysis of single amino acid substitutions. *BMC Bioinform.* **2009**, *10*, S8. [CrossRef]

24. Steentoft, C.; Vakhrushev, S.Y.; Joshi, H.J.; Kong, Y.; Vester-Christensen, M.B.; Schjoldager, K.T.; Lavrsen, K.; Dabelsteen, S.; Pedersen, N.B.; Marcos-Silva, L.; et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **2013**, *32*, 1478–1488. [CrossRef] [PubMed]

25. Gupta, R.; Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* **2002**, *7*, 310–322.

26. Wang, D.; Liu, D.; Yuchi, J.; He, F.; Jiang, Y.; Cai, S.; Li, J.; Xu, D. MusiteDeep: A deep-learning based webserver for protein post-translational modification site prediction and visualization. *Nucleic Acids Res.* **2020**, *48*, W140–W146. [CrossRef]

27. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. PolyPhen-2: Prediction of functional effects of human nsSNPs. *Nat. Methods* **2010**, *7*, 248–249. [CrossRef]

28. Sim, N.-L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P.C. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **2012**, *40*, W452–W457. [CrossRef]

29. Bromberg, Y.; Rost, B. SNAP: Predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.* **2007**, *35*, 3823–3835. [CrossRef]

30. Liu, X.; Li, C.; Mou, C.; Dong, Y.; Tu, Y. dbNSFP v4: A comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med.* **2020**, *12*, 103. [CrossRef]

31. Fenna, R.; Zeng, J.; Davey, C. Structure of the Green Heme in Myeloperoxidase. *Arch. Biochem. Biophys.* **1995**, *316*, 653–656. [CrossRef]

32. Grayson, P.C.; Kaplan, M.J. At the Bench: Neutrophil extracellular traps (NETs) highlight novel aspects of innate immune system involvement in autoimmune diseases. *J. Leukoc. Biol.* **2016**, *99*, 253–264. [CrossRef] [PubMed]

33. Odobasic, D.; Yang, Y.; Muljadi, R.C.M.; O'Sullivan, K.M.; Kao, W.; Smith, M.; Morand, E.F.; Holdsworth, S.R. Endogenous myeloperoxidase is a mediator of joint inflammation and damage in experimental arthritis. *Arthritis Rheumatol.* **2014**, *66*, 907–917. [CrossRef] [PubMed]

34. Ndrepepa, G. Myeloperoxidase—A bridge linking inflammation and oxidative stress with cardiovascular disease. *Clin. Chim. Acta Int. J. Clin. Chem.* **2019**, *493*, 36–51. [CrossRef] [PubMed]

35. Matreyek, K.A.; Starita, L.M.; Stephany, J.J.; Martin, B.; Chiasson, M.A.; Gray, V.E.; Kircher, M.; Khechaduri, A.; Dines, J.N.; Hause, R.J.; et al. Multiplex Assessment of Protein Variant Abundance by Massively Parallel Sequencing. *Nat. Genet.* **2018**, *50*, 874–882. [CrossRef]

36. Bloom, J.D.; Labthavikul, S.T.; Otey, C.R.; Arnold, F.H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5869–5874. [CrossRef]

37. Post, K.L.; Belmadani, M.; Ganguly, P.; Meili, F.; Dingwall, R.; McDiarmid, T.A.; Meyers, W.M.; Herrington, C.; Young, B.P.; Callaghan, D.B.; et al. Multi-model functionalization of disease-associated PTEN missense mutations identifies multiple molecular mechanisms underlying protein dysfunction. *Nat. Commun.* **2020**, *11*, 2073. [CrossRef]

38. Hasnain, M.J.U.; Amin, F.; Ghani, A.; Ahmad, S.; Rehman, Z.; Aslam, T.; Pervez, M.T. Structural and Functional Impact of Damaging Nonsynonymous Single Nucleotide Polymorphisms (nsSNPs) on Human VPS35 Protein Using Computational Approaches. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, 1. [CrossRef]

39. Pandey, S.; Dhusia, K.; Katara, P.; Singh, S.; Gautam, B. An in silico analysis of deleterious single nucleotide polymorphisms and molecular dynamics simulation of disease linked mutations in genes responsible for neurodegenerative disorder. *J. Biomol. Struct. Dyn.* **2020**, *38*, 4259–4272. [CrossRef]

40. Sobitan, A.; Mahase, V.; Rhoades, R.; Williams, D.; Liu, D.; Xie, Y.; Li, L.; Tang, Q.; Teng, S. Computational Saturation Mutagenesis of SARS-CoV-1 Spike Glycoprotein: Stability, Binding Affinity, and Comparison With SARS-CoV-2. *Front. Mol. Biosci.* **2021**, *8*, 784303. [CrossRef]

41. Iossifov, I.; O'Roak, B.J.; Sanders, S.J.; Ronemus, M.; Krumm, N.; Levy, D.; Stessman, H.A.; Witherspoon, K.T.; Vives, L.; Patterson, K.E.; et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **2014**, *515*, 216–221. [CrossRef]

42. Lim, E.T.; Uddin, M.; De Rubeis, S.; Chan, Y.; Kamumbu, A.S.; Zhang, X.; D'Gama, A.M.; Kim, S.N.; Hill, R.S.; Goldberg, A.P.; et al. Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. *Nat. Neurosci.* **2017**, *20*, 1217–1224. [CrossRef] [PubMed]

43. DeLeo, F.R.; Goedken, M.; McCormick, S.J.; Nauseef, W.M. A novel form of hereditary myeloperoxidase deficiency linked to endoplasmic reticulum/proteasome degradation. *J. Clin. Investig.* **1998**, *101*, 2900–2909. [CrossRef] [PubMed]

44. Romano, M.; Dri, P.; Dadalt, L.; Patriarca, P.; Baralle, F.E. Biochemical and Molecular Characterization of Hereditary Myeloperoxidase Deficiency. *Blood* **1997**, *90*, 4126–4134. [CrossRef] [PubMed]

45. Xu, Y.; Liu, X.; Shen, J.; Tian, W.; Fang, R.; Li, B.; Ma, J.; Cao, L.; Chen, S.; Li, G.; et al. The Whole Exome Sequencing Clarifies the Genotype- Phenotype Correlations in Patients with Early-Onset Dementia. *Aging Dis.* **2018**, *9*, 696–705. [CrossRef] [PubMed]

46. Marchetti, C.; Patriarca, P.; Solero, G.P.; Baralle, F.E.; Romano, M. Genetic Characterization of Myeloperoxidase Deficiency in Italy. *Hum. Mutat.* **2004**, *23*, 496–505. [CrossRef] [PubMed]

47. Ohashi, Y.Y.; Kameoka, Y.; Persad, A.S.; Koi, F.; Yamagoe, S.; Hashimoto, K.; Suzuki, K. Novel missense mutation found in a Japanese patient with myeloperoxidase deficiency. *Gene* **2004**, *327*, 195–200. [CrossRef]

48. Kameoka, Y.; Persad, A.S.; Suzuki, K. Genomic variations in myeloperoxidase gene in the Japanese population. *Jpn. J. Infect. Dis.* **2004**, *57*, S12–S13.

49. Mishra, O.P.; Popov, A.V.; Pietrofesa, R.A.; Hwang, W.T.; Andrake, M.; Nakamaru-Ogiso, E.; Christofidou-Solomidou, M. Radiation Activates Myeloperoxidase (MPO) to Generate Active Chlorine Species (ACS) via a Dephosphorylation Mechanism— Inhibitory Effect of LGM2605. *Biochim. Biophys. Acta. Gen. Subj.* **2020**, *1864*, 129548. [CrossRef]

50. Kim, Y.; Kang, C.; Min, B.; Yi, G.S. Detection and analysis of disease-associated single nucleotide polymorphism influencing post-translational modification. *BMC Med. Genom.* **2015**, *8*, S7. [CrossRef]