Scientific Research Publishing

Testing the Mutation Accumulation Theory of Aging Using Bioinformatic Tools

Abdullah Salah Elamoudi

Department of Internal Medicine, King Abdulaziz Medical City, Jeddah
Email: Abdullah.s.elamoudi@gmail.com

Abstract

Objective: An interplay of many variant mechanisms is thought to underlie aging or senescence. The Mutation Accumulation Theory proposes the accumulation of mutations in proteins to engender their aging phenotype. Testing whether random mutations lead to the aging phenotype was never done and is deemed infeasible. Bioinformatic algorithms provide an *a-priori* approach that allows testing; they employ prior knowledge of well-studied proteins to predict the changes brought on by mutations. Here, the Mutation Accumulation Theory of aging is tested using such bioinformatic tools. Methods: This is a simulation study, conducted 2017, September, using algorithms with Web accessibility. Three well-studied proteins implicated in aging were chosen: Collagen, Beta-amyloid Precursor Protein (β-APP) and Low-density-lipoprotein-receptor (LDL-receptor). Random mutations were introduced to their native coding sequences. Then, the mutated sequences were tested using three different prediction algorithms: SPpred for solubility, I-mutant for stability (delta-free energy), SNP and GO for pathogenicity. The new mutated phenotype was then correlated to the aging phenotype of the protein; decrease in solubility for Collagen and β-APP; and accelerated atherosclerosis for LDL-receptor. Results: 15 mutated variants for each protein (45 in total). For collagen and β-APP, the SPpred algorithm did not predict changes in solubility of the naked protein, but the I-mutant and SNP and GO definitely predicted changes that fit the aging phenotype. However, for LDL-receptors, none of the mutated variants when studied could account for the aging phenotype. Conclusion: for Collagen and B-APP, it is shown here that random mutations and their accumulation could explain the aging phenotype of both proteins; backing the Mutation Accumulation Theory for aging.

Keywords

Aging, Senescence, Mutation Accumulation Theory

1. Introduction

Not long after the settlement of early humans in small communities, forgoing the hunter-gatherer lifestyle, did the question of aging and mortality engross many aspects of our lives. Our awareness of our mortality and decay, a distinguishing feature of our species among living organisms, dawned our collective mind with angst and became the drive of our mores, mythology and metaphysical expression; it even became the subject of our art.

Historically, aging was likened to wear and tear accumulating with time. This view however came to change when the thermodynamic law of entropy took its formal shape. *i.e.*, entropy is unidirectional; disorder is always increasing. However, biological systems are not closed systems and living beings take in free energy then release its entropy as waste. Living organisms reproduce and repair and there is no thermodynamic necessity for senescence (aging).

The turn of the twentieth century saw prominent biologists addressing the issue of aging. Ilya Ilyich Mechnikov, a prominent Russian biologist and a Nobel laureate for the discovery of phagocytes (macrophages) in 1908, was one of the first biologists to address the issue of aging and he is the one to coin the terms *Geriatrics* and *Gerontology* from the Greek word *Geras* for old age. Since then, many theories for aging were put forward. Some backed by more evidence than others. The free radical theory and DNA damage theory are two of the older ones. In the free radical theory, accumulative damage that underlies senescence is brought on by free radicals with time, hence, calorie restriction and the increase in consumption of anti-oxidant for an individual can be said to increase longevity [1]. The DNA damage theory suggests breaks in the DNA and other mechanical damage to the DNA and what they ensue on the cellular machinery to be the main culprit of aging. Much evidence was put forward for other theories to take shape; programmed longevity theory suggests aging to be a similar phenomenon to development where the expression of genes alters with time changing with it the phenotypic traits of the individual; the endocrine theory suggests hormonal changes to be a contributing factor to aging [2].

One of the oldest theories to address aging is the Mutation Accumulation Theory: first proposed by Sir Peter Brian Medawar, a prominent biologist and a Noble Laureate (1960), it proposes mutations accumulating with age in the DNA to be the main culprit for aging. For this to happen, mutations have to accumulate in stem cells leading to the expression of defected proteins which underlie the phenotypic traits of aging [3]. There has already been description of genetic mutations that are associated with aging [4]. Although the rationale behind this theory remains solid, linking these mutations to the aging phenotype had not been done yet, and testing for this would mean testing for mutations in stem cells; mutations that are neither consistent nor repetitive; from here, the mutation accumulation theory remains hypothesized and not proven. However, evidence abounds proving that changes in protein expression do take place with aging, especially recently with the introduction of high-throughput proteomics;

referenced here are three papers that did that [5] [6] [7].

It is easily understood why testing this theory was not done. It would be cumbersome and unfathomable to test this theory in a lab; the number of possible mutations, the variance of the single mutation itself, and the addition of more mutations, etc. Not to mention the ensuing study of the expressed protein in different milieus, different conditions and different systems each with a great number of variables to account for. Bioinformatic tools are *a-priori* tools that use existent information on proteins to predict the outcomes of a change introduced.

For this project, the mutation accumulation theory will be tested; random mutations will be introduced into protein sequences. The effect of these mutations will be assayed using bioinformatics tools. This effect will then be correlated to the known aging phenotype of the chosen proteins.

2. Experimental Procedure

This is a simulation study; conducted September 2017, using accessible algorithms available online.

1) Choice of proteins: Firstly, the proteins have to be implicated in aging phenotypic traits. Secondly, the proteins chosen have to be well-studied since one of the algorithms will relate the mutated sequence to diseased phenotypes brought on by hereditary mutations. Also, the physiology where the protein is involved will help correlate the effects of mutations predicted by the algorithms to the aging phenotype that is known. Therefore, for this project, three proteins were chosen: Collagen, beta-Amyloid Precursor Protein (β-APP) and Low-density-lipoprotein-receptor (LDL-receptor).

Collagen is the most abundant protein in the body, forming the connective matrix for almost all tissues. Studies have been done to implicate collagen's drop in solubility to aging phenomena; in the skin, lens, blood vessels, bones and cartilage [7]. B-APP is known to precipitate in Alzeimer's and in aged individuals causing senility [8]. LDL-receptors mediate the deposition of atherosclerotic plaques that are known to accelerate with aging [9].

2) Choice of mutations: in principle, mutations accumulating with age should be random; to do that, the sequences were input in a simple algorithm that introduced these mutations randomly. Here, the ***Sequence Manipulation Suite-Mutate Protein*** tool provided by Bioinformatics.org was used. It introduces mutations randomly with more predilections to parts of the sequence that are more prone to mutate, mimicking what goes on in actual DNA machinery. The single native sequence was input five times independently to produce five mutated sequences. Then, each mutated sequence was input in the same algorithm two more times sequentially to accumulate two independent mutations. Therefore, for each native sequence, fifteen mutated variants were produced; forty five for all three proteins. All these mutations were single-point mutations leading to amino acid substitution.

To accumulate multiple mutations in a single sequence, might seem a scarce phenomenon; since mutations occur with a frequency of 1/10,000,000 mutation per replication cycle. But, taking into account the number of replications, the number of different cells producing the protein and the life-span of the average human, accumulating two more mutations to the first mutated sequence merits the study. Every mutated sequence was then input in three different algorithms:

-**SPpred** (**Soluble Protein Prediction**): to predict the solubility of the protein. Support Vector Machine values (SVM) were output; a positive sign for soluble and a negative sign for insoluble; the absolute value is scalar for how soluble or insoluble the protein was predicted to be; this solubility algorithm was chosen for simplicity and not including expression systems as a required input.

-**I-mutant**: To predict the stability (delta-free energy) brought on by the mutation; a positive sign for an increase in the stability of the protein, a negative sign for a drop in stability. This algorithm was chosen after the recommendation by a review article that tested different stability algorithms and concluded two to be the best ones; I-mutant is one those two [10].

For the added mutations, delta-free energies were arithmetically added; example, delta-G of second mutant is the sum of delta-G of the first mutant + delta-G of the second mutant as a mutation in the first mutated sequence.

-**SNP and GO**: for predicting pathogenicity. It correlates the mutation introduced to the sequence to the known hereditary mutations and outputs a probability to predict the diseased phenotype. This algorithm was also recommended by a review article that tested several algorithms for pathogenicity prediction [11].

The results were then correlated to the aging phenotype; this is done in the discussion part of this paper.

As a simulation study, no Ethical approval was required.

3. Results

All 45 mutants were run in the three algorithms, none was excluded.

1) Collagen:

For collagen, the *SPpred* algorithm for solubility reported values that are comparable to the wild type (see Table 1). The *I-mutant* algorithm predicted a significant effect on stability; 8 out of 15 mutated sequences had a delta-free energy value above 0.5. Three of these 8 had a value above 1; this effect on stability was gradient for 100% of mutated sequences; increasing added mutations. The *SNP and GO* algorithm predicted 6 out of the 15 mutated sequences would have a diseased phenotype (see Table 1). Some mutations when added reversed the pathology predicted by the preceding mutation.

Here, Collagen sequences with a single mutation and those with multiple (added one or two) are predicted to have the aging phenotype.

2) Beta-Amyloid Precursor Protein:

For β-APP, the SPpred algorithm for solubility reported values that are

Table 1. Output of the three algorithms for **Collagen**; SVM (SPpred) for support vector machine, DDG (I-mutant) for delta free energy, Disease/probability (SNP and GO) N for neutral and D for disease/the probability of disease. WT; wild type (unmutated sequence). Mere numbers in the rows are for the different initial mutations, +1 for the added mutation to the first mutated sequence, +2 for the second added mutation to the sequence with two mutations.

	Mutation	SVM	DDG	Disease/Probability
WT		−0.59203		
1	D116E	−0.58325	0.2	N/0.047
1 + 1	G902M	−0.5779	−0.67	D/0.983
1 + 2	P544T	−0.57267	−0.91	N/0.061
2	P849E	−0.58967	−0.04	N/0.192
2 + 1	P336V	−0.5866	0.39	N/0.15
2 + 2	P841Q	−0.5932	−0.17	N/0.144
3	G1106V	−0.58419	0.2	D/0.901
3 + 1	G494D	−0.58414	−0.17	D/0.888
3 + 2	G614L	−0.58747	−1.16	D/0.883
4	G221A	−0.5844	0.93	D/0.801
4 + 1	A1134Q	−0.58508	0.08	N/0.122
4 + 2	G834M	−0.57972	−0.8	D/0.841
5	G1187L	−0.5852	−0.8	N/0.201
5 + 1	K1236F	−0.58934	−2.25	N/0.108
5 + 2	P504A	−0.58316	−2.75	N/0.125

comparable to the wild type (see Table 2). The I-mutant algorithm predicts significant effect on stability; 14 out of 15 mutated sequences had a delta-free energy value above 0.5. 9 of these 14 had a value above 1. This effect on stability was gradient for 100% of mutated sequences; increasing instability with adding mutations. The SNP and GO algorithm predicts that 6 out of the 15 mutated sequence will have a diseased phenotype. However, 3 out of the 15 mutated sequences could not be output with the SNP and GO algorithm and were output as "error" (see Table 2). Here, mutated sequences with single or multiple mutations led to the aging phenotype.

3) LDL-receptor:

For LDL-receptor, the SPpred algorithm for solubility reported values that are comparable to the wild type (see Table 3). The I-mutant algorithm predicts significant effect on stability; 9 out of 15 mutated sequences had a delta-free energy value above 0.5. 8 out of these 9 had a value above 1. This effect was NOT gradient for 50% of the mutated sequences; did not necessarily increase with adding mutations. The SNP and GO algorithm predicts that only 2 out of the 15 mutated sequences will have a diseased phenotype (see Table 3). Here, almost all mutated sequences were not fazed these introduced mutations.

Table 2. Output of the three algorithms for **B-APP**; SVM (SPpred) for support vector machine, DDG (I-mutant) for delta free energy, Disease/probability (SNP and GO) N for neutral and D for disease/the probability of disease. WT; wild type (unmutated sequence) Mere numbers in the rows are for the different initial mutations, +1 for the added mutation to the first mutated sequence, +2 for the second added mutation to the sequence with two mutations.

	Mutation	SVM	DDG	Disease/Probability
WT		0.93634		
1	V551K	0.92703	−1.87	D/0.681
1 + 1	A472R	0.90571	−1.87	D/0.138
1 + 2	M752V	0.91678	−3.11	N/0.138
2	P130S	0.94349	−0.98	D/0.723
2 + 1	L61K	0.95451	−0.8	N/0.204
2 + 2	P424T	0.95451	−1.22	Error
3	R470K	0.93454	0.5	D/0.747
3 + 1	L350K	0.9396	−1.65	N/0.054
3 + 2	K726T	0.93428	−2.71	N/0.189
4	D739L	0.94609	−0.63	Error
4 + 1	A294Y	0.93888	−0.66	D/0.806
4 + 2	G649Y	0.93307	−1.21	N/0.064
5	L118P	0.90773	−0.22	Error
5 + 1	Q481G	0.90386	−1.43	D/0.851
5 + 2	A213C	0.89283	−2.57	N/0.133

4. Discussion

In this paper, three proteins were chosen that have a prominent role in aging; Collagen, β-Amyloid Precursor Protein and LDL-receptor. The aging phenotype of these proteins is well-known. Each native sequence has 15 mutated variants. For each, five had one mutation, five had two (one added to the first one), and five had accumulated three mutations. To accumulate multiple mutations in a single sequence, might seem a scarce phenomenon at first; since mutations occur with a frequency of 1/10,000,000 per replication cycle. However, taking into account the number of replication cycles and the life-span of the average human being, accumulating two more mutations to the first mutated sequence merits the study. The Mutation accumulation theory suggests that mutations accumulating in these proteins would lead to the aging phenotype. For each of the forty-five mutations in total, three bioinformatics tests were run to each.

For collagen, insolubility is the main aging phenotype all over the body; in skin where turgor is increased and the connective tissue doesn't hold as much water producing a ubiquitous aging phenomenon in all mammals; in lenses where losing solubility will cause collagen to precipitate and to haze the lenses "cataract"; in joints where the cartilage flexibility is diminished exposing the

Table 3. Output of the three algorithms for LDL-receptor; SVM (SPpred) for support vector machine, DDG (I-mutant) for delta free energy, Disease/probability (SNP and GO) N for neutral and D for disease/ the probability of disease. WT; wild type (unmutated sequence) Mere numbers in the rows are for the different initial mutations, +1 for the added mutation to the first mutated sequence, +2 for the second added mutation to the sequence with two mutations.

	Mutation	SVM	DDG	Disease/Probability
WT		−0.90019		
1	S174C	−0.88236	−1.31	N/0.089
1 + 1	L191N	−0.90011	−2.22	D/0.752
1 + 2	W34R	−0.82773	−2.47	N/0.102
2	D4C	−0.90995	−0.42	N/0.96
2 + 1	Q247E	−0.90247	−1.32	N/0.129
2 + 2	S235V	−0.84264	−1.78	N/0.104
3	K25T	−0.90866	−1.13	N/0.061
3 + 1	S12N	−0.87993	−2.47	N/0.058
3 + 2	L6G	−0.96008	−2.39	N/0.104
4	S61L	−0.9292	−0.14	N/0.074
4 + 1	E153L	−0.98389	0.18	D/0.667
4 + 2	R209Q	−0.93933	−0.1	N/0.091
5	G48S	−0.87999	−0.94	N/0.215
5 + 1	T237V	−0.83627	−0.43	N/0.054
5 + 2	E101K	−0.85581	−0.47	N/0.025

surface to further wear and tear "osteoarthritis"; in blood vessels where rigidity ensues and higher blood pressures become a common phenomenon among the aging population [7]. For this rather pervasive phenotypic trait, the mutation accumulation theory of aging suggests mutations in the coding sequence of collagen to be the main culprit in reducing solubility. After running five random mutations and accumulating two more mutations to them sequentially, the bioinformatic tool used here concluded minor changes brought on by the changing side chains on the solubility of the protein (*see* **Table 1**). Although a change is reported, it is never significant enough to account for the major drop in solubility in aging collagen [7]. However, these mutations can still impact solubility through other means and to a much higher degree; these mutations can interfere with post-translational modifications and glycosylation events that can enhance solubility tremendously; for example hydroxylation of lysine and proline residues in collagen are major modifications [12]. This hydroxylation depends on vitamin C as a cofactor and therefore Vitamin C deficiency weakens Connective tissue support of blood vessels and other components and leads to Scurvy; this underscores further the role of such modifications in collagen solubility. Other modifications are done to collagen. Cross-linking events are important ones;

cross-linking events are known to solidify collagen reducing its solubility [13]. These post-translational phenomenae alter Collagen's solubility to a great degree; mutations can hinder these modifications or enhance cross-linking and therefore lower solubility through these indirect means that cannot be assayed here. The role played by these modifications is underscored by the fact that our wild-type collagen sequence had low solubility score that was comparable to those sequences housing one, two or even three mutations.

Another way for these mutations to alter collagen solubility is by the significant drop in stability predicted by the bioinformatics tool used here "I-mutant" (*see* Table 1). Here, we note significant changes in the free energy of the protein brought on by a single mutation at times. If this effect of these mutations were not masked by post-translational modifications, then it is safe to assume unstable protein conformations that favor precipitation and interfere with collagen as a substrate for enzymes modifying it. In some instances, adding a mutation or two, caused even a more significant drop in stability "all groups of mutations except the third group"; this is also in check with the mutation accumulation theory of aging where accumulating mutations are believed to play a role as well.

Finally, the SNP and GO algorithm was run to all collagen sequences and for six out of 15 sequences, the algorithm predicted the introduced mutation or mutations to cause a diseased phenotype; as predicted, mutations that substituted glycine, an amino acid with no side chains, were almost consistently leading to the diseased phenotype; in collagen, these glycine residues are important for compacting the quaternary structure of Collagen [14]. Collagen hereditary diseases all have a feature in common; collagen losing its supportive role in tissues, for example, Alport Syndrome, Epidermolysis Bullosa Dystrophica, Ehlers-Danlos Syndrome and Osteogenesis Imperfecta. Hence, it is safe to assume that these random mutations "six mutated sequences" can be chosen as predicted mutations that take place with aging; this should back up the Mutation Accumulation Theory as an important phenomenon that underlies the aging phenotype. Here however, accumulating mutations didn't further the disease phenotype (*see* Table 1).

In conclusion: *Results here conclude that collagen as a naked protein has low solubility, post-translational modifications and cross-linking events modify the solubility of collagen to a great degree as is reported in the literature. Therefore, introducing mutations to the naked sequence should not alter solubility that much as concluded in this study. However, these mutations caused somewhat significant changes in the stability of collagen which can interfere with the post-translational handling of collagen by different enzymes. This is backed even further when our random mutations were compared to hereditary collagen opathies using the algorithm SNP and GO proving that some of the random mutations, sometimes even solitary ones, can produce a diseased phenotype; by running the experiment in this last algorithm, post-translational modifications had to be included in the scope of our argument. This should prove that random mutations standing alone can produce the aging phenotype.*

For beta-Amyloid Precursor Protein, the same number of mutations was produced randomly and similar bioinformatic runs were done to all mutated sequences. B-APP is notorious for its role in senility and Alzehimer's; this protein will precipitate in all individuals indiscriminately, however certain genetic variants and mutations will favor the early precipitation and early manifestation of Alzehimer's symptoms starting with memory loss [15]. Therefore, it's a reasonable first-step to test the solubility of randomly mutated sequences; here however, no significant differences were observed and the wild-type sequence had similar low solubility values (see Table 2). This is easily accounted for by the transmembrane domain of β-APP. β-APP is a surface protein. Plus, it is the β-APP cleavage product that precipitates and not the entire β-APP; we know that once β-APP is cleaved by beta-secretase, a product that precipitates is produced and allelic variants that favor this particular cleavage associate with a higher prevalence and earlier onset of Alzheimer's [8]. Therefore, these solubility values can be ignored at this point. As for the stabilities concluded from mutations, mutations lowered the stability of β-APP; all groups of mutations had a significant negative effect on stability except the second group; these can be proposed as mutations that could underlie the aging phenotype; lowering the stability should interfere with enzymatic handling of the protein and therefore encourage its degradation; a higher turn-over of β-APP should mean greater cleavage product "cleavage is a degradation mechanism of the β-APP" [8].

Finally, the SNP and GO algorithm was run to all sequences; this algorithm should relate the mutations to known mutations and allelic variants favoring the diseased phenotype; in the case of β-APP it's almost exclusively associated with senility and Alzheimer's. The algorithm's output predicted the diseased phenotypes for six mutated sequences out of twelve retrieved outcomes. Also the diseased phenotype was favored with high probabilities at times; 80.5% and 85.1% for the fourth and fifth groups respectively. This without question should underscore the role of these random mutations (50% of the time, 6 out 12) in producing senility in the aging individual; an important hallmark of aging.

In conclusion: for β-APP, solubility algorithms didn't show a significant difference but this could be accounted for by the cellular location of the protein and the fact that only part of β-APP is concerned with precipitation and disease. The stability algorithms predicted significant drop in stability for most mutations which could raise the turn-over of the protein and therefore increase the degradation products; one of which is what precipitates and causes the aging/disease phenotype. Finally, the SNP and GO algorithm concluded half of these randomly mutated sequences to produce a phenotypic change similar to that of Alzehimer's. Therefore, these random mutations could just as well underlie the aging phenotype validating the Mutation Accumulation Theory of Aging.

LDL-receptors play an important role in Atherosclerosis [9]. Atherosclerosis is known to accelerate with age and is therefore a hallmark of aging [16]. Because of ischemia brought on by atherosclerosis "acute or chronic", heart disease

is the leading cause of mortality among men and women of old age (CDC, Heart Disease Facts). Here, we ran the same number of mutations in the same three aforementioned algorithms.

As predicted, no significant changes in solubility were observed since the protein is a receptor with a significant part spanning the phospholipid bilayer (transmembrane). The stability algorithms concluded a drop in stability in the first three groups of mutations, and a stabilizing effect of the last two groups of mutations on LDL-receptors. For the SNP and GO algorithms, all mutations were neutral except two; and for these two the probability of disease was not that significant; 75% and 66%. Plus, both mutations were the second one added to the mutated sequence and were neutralized by the addition of the third random mutation. All these factors lead us to exclude any significant effect brought on by these random mutations to produce an aging phenotype (accelerated atherosclerosis). Therefore, besides the destabilizing effect of some of the mutations, no significant effect of random mutations was observed on the protein. These destabilizing mutations can however increase the turn-over of the receptor, but even here no pathology can be projected; the receptor acts to internalize LDL continuously and a high turn-over is a known trait of the protein [9]. Another reason why these algorithms might be poor in making predictions for LDL-receptor, is the fact that in atherosclerosis the LDL-receptor pathway is overly activated, therefore processes that up-regulate the pathway or suppress its down-regulation, are ones that favor atherosclerosis; in Atherosclerosis, macrophages using their LDL-receptors uptake LDL from the blood and fail to be suppressed from doing so; with this over-uptake of LDL, the macrophages eventually die and deposit below the endothelium [17]. Therefore, looking for mutations that destabilize the protein go against the accelerated atherosclerosis assumption. Plus, in SNP and GO mutations are correlated to mutations in LDL-receptor that cause familial hypercholesterolemias where LDL-receptor loses function. But, in aging, we expect the LDL-receptor to over-activate leading to accelerated atherolsclerosis, a known phenotypic trait of aging.

In conclusion; for LDL-receptors, it was not possible to predict the aging phenotype of accelerated atherosclerosis using these bioinformatics tools. An important reason for that, is changes that accelerate atherosclerosis are ones that activate the pathway where LDL-receptors are involved. Plus, the SNP and GO algorithm which proved very useful with Collagen and β-APP count on hereditary diseases in LDL receptors to make predictions; in LDL-receptors the only hereditary diseases involving them are Familial hypercholesterolemias, where LDL-receptor-mediated uptake of LDL from the blood is down-regulated which is the opposite of what occurs in macrophages to produce atherosclerosis [18].

Some significant limitations are noted in this study. One is limiting the study to mutations that led to amino acid substitutions. No other mutations were studied; insertions/deletions, missense, frame-shift, to name a few. This was not done as no algorithms could simulate those mutations that were tested and

recommended by other authors. However, the alteration in protein expression with aging is well described and did not report significant changes in the proteins in terms of size, conformation, and no reference to truncation either [5]. This weighs more to hint at substitutions being the predominant mutations in these proteins as we age. No mutations were studied that involved the non-coding regulatory sequences either of for the same two reasons.

As for the number of accumulated mutations being small, considering the rate of mutagenesis in DNA being around 1/10,000,000, plus the proof-reading machinery in our DNA, accumulating one or two mutation can be said to be generous. Another argument is that for Collagen and β-APP, the specific substitution and the site of mutation did not seem to factor much in the predictions. The bottom line being that the bulk of mutations would lead more or less to the aging phenotype. This would be a more significant limitation had that not been the case.

5. Conclusion

The mutation Accumulation Theory of Aging proposes that random mutations with aging underlie, to some extent, the aging phenotype; here three proteins were mutated randomly and the effect of mutations predicted using bioinformatics tools; these effects were then correlated to the aging phenotype. For Collagen and β-APP, randomly produced mutations had predicted effects that can easily be correlated to the aging phenotype and therefore, back up the mutation accumulation theory for aging. However, for LDL-receptors, although a good candidate protein to study for aging, it could not be correlated to the aging phenotype; this can be attributed to the biological role LDL-receptors play in aging which is not analogous to established pathologies that the pathogenicity algorithm SNP and GO use for predictions. In conclusion, in this project, using bioinformatics tools, random mutations that incur with aging can produce the aging phenotype, validating the Mutation Accumulation Theory. This was the case for Collagen and β-APP.

References

[1] Harman, D. (1992) Free Radical Theory of Aging. *Mutation Research/DNAging*, **275**, 257-266. https://doi.org/10.1016/0921-8734(92)90030-S

[2] Jin, K.L. (2010) Modern Biological Theories for Aging. *Aging and Disease*, 72-74.

[3] Kennedy, S.R., Loeb, L.A. and Herr, A.J. (2012) Somatic Mutations in Aging, Cancer and Neurodegeneration. *Mechanisms of Ageing and Development*, **133**, 118-126. https://doi.org/10.1016/j.mad.2011.10.009

[4] Sharov, V.S. and Schöneich, C. (2007) Proteomic Approach to Aging Research. *Journal Expert Review of Proteomics*, **4**, 309-321. https://doi.org/10.1586/14789450.4.2.309

[5] Waldera-Lupa, D.M., Kalfalah, F., Florea, A.-M., Sass, S., Kruse, F., Rieder, V., Tigges, J., *et al.* (2014) Proteome-Wide Analysis Reveals an Age-Associated Cellular Phenotype of *in Situ* Aged Human Fibroblasts. *Aging* (*Albany NY*), **6**, 856-872.

https://doi.org/10.18632/aging.100698

[6] Grant, J.E., Bradshaw, A.D., Schwacke, J.H., Baicu, C.F., Zile, M.R. and Schey, K.L. (2009) Quantification of Protein Expression Changes in the Aging Left Ventricle of *Rattus norvegicus. Journal of Proteome Research*, **8**, 4252-4263. https://doi.org/10.1021/pr900297f

[7] Schnider, S.L. and Kohn, R.R. (1981) Effects of Age and Diabetes Mellitus on the Solubility and Nonenzymatic Glucosylation of Human Skin Collagen. *Journal of Clinical Investigation*, **67**, 1630. https://doi.org/10.1172/JCI110198

[8] Selkoe, D.J. (1994) Normal and Abnormal Biology of the Beta-Amyloid Precursor Protein. *Annual Review of Neuroscience*, **17**, 489-517. https://doi.org/10.1146/annurev.ne.17.030194.002421

[9] Brown, M.S. and Goldstein, J.L. (1984) How LDL Receptors Influence Cholesterol and Atherosclerosis. *Scientific American*, **251**, 58-66. https://doi.org/10.1038/scientificamerican1184-58

[10] Potapov, V., Cohen, M. and Schreiber, G. (2009) Assessing Computational Methods for Predicting Protein Stability upon Mutation: Good on Average but Not in the Details. *Protein Engineering Design and Selection*, **22**, 553-560. https://doi.org/10.1093/protein/gzp030

[11] Thusberg, J., Olatubosun, A. and Vihinen, M. (2011) Performance of Mutation Pathogenicity Prediction Methods on Missense Variants. *Human Mutation*, **32**, 358-368. https://doi.org/10.1002/humu.21445

[12] Kivirikko, K.I. and Prockop, D.J. (1967) Enzymatic Hydroxylation of Proline and Lysine in Protocollagen. *Proceedings of the National Academy of Sciences of the United States of America*, **57**, 782. https://doi.org/10.1073/pnas.57.3.782

[13] Deshmukh, K. and Nimni, M.E. (1969) Chemical Changes Associated with Aging of Collagen *in Vivo* and *in Vitro. Biochemmical Journal*, **112**, 397-405. https://doi.org/10.1042/bj1120397

[14] Fratzl, P., Misof, K., Zizak, I., Rapp, G., Amenitsch, H. and Bernstorff, S. (1998) Fibrillar Structure and Mechanical Properties of Collagen. *Journal of Structural Biology*, **122**, 119-122. https://doi.org/10.1006/jsbi.1998.3966

[15] Holcomb, L., Gordon, M.N., McGowan, E., Yu, X., Benkovic, S., Jantzen, P., Wright, K., *et al.* (1998) Accelerated Alzheimer-Type Phenotype in Transgenic Mice Carrying Both Mutant Amyloid Precursor Protein and Presenilin 1 Transgenes. *Nature Medicine*, **4**, 97-100. https://doi.org/10.1038/nm0198-097

[16] Bierman, E.L. (1978) Atherosclerosis and Aging. *Federation Proceedings*, **37**, 2832-2836.

[17] Moore, K.J. and Tabas, I. (2011) Macrophages in the Pathogenesis of Atherosclerosis. *Cell*, **145**, 341-355. https://doi.org/10.1016/j.cell.2011.04.005

[18] Hobbs, H.H., Brown, M.S. and Goldstein, J.L. (1992) Molecular Genetics of the LDL Receptor Gene in Familial Hypercholesterolemia. *Human Mutation*, **1**, 445-466. https://doi.org/10.1002/humu.1380010602