Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

Original Research Article **DOI - 10.26479/2017.0205.06**

A HYPOTHESIS ON THE POSSIBLE EVOLUTIONARY ROLE OF HISTIDINE IN PROTEINS

George Suji*

Department of Microbiology* Virus Research and Diagnostic laboratory

Bangalore Medical College & Research Institute, KR Road, Fort, Bangalore Karnataka-560 002, India

ABSTRACT: Glycation, the nonenzymic interaction of sugars with proteins is known to play a key role in complications of many pathophysiological processes. Glycation by sugars is directed against the positively charged residues in the protein, namely arginine and lysine. It is an undesirable phenomenon and result in progressive loss of function of biological system. Glycated proteins are tolerated in the body compartment, as breakdown and replacing them amounts to spending extra energy. So the glycated protein to exist in body compartment should resort to compensatory mechanisms that resist the effects of glycation. The charge regulation is such a process, which resist the change in net charge due to glycation. Thus the increase in the net charge of the protein, results in increased negative potential on the surface of the protein which increases the local concentration of H+ ions around the histidine residues in proteins there by affecting their ionization. Since glycation is inevitable process, I hypothesise yet another possible evolutionary role of histidine in proteins to resist change in net charge arising due to deleterious modification.

KEYWORDS: Glycation, Charge regulation, Histidine, Proteins

***Corresponding Author: Dr. George Suji** Ph.D.

Department of Microbiology* Virus Research and Diagnostic laboratory

Bangalore Medical College & Research Institute, KR Road, Fort, Bangalore Karnataka-560 002, India

* Email Address: geosuji@gmail.com

1.INTRODUCTION

Reaction of virtually any protein with a reducing sugar can produce a stable glycated form of protein. These reactions are slow, are not catalysed by enzymes and, to certain extent take place in almost all forms of life. Glycation can occur within the cell or in the extracellular compartment of the body. Reducing sugars (glucose, mannose, fructose, galactose, ribose) and other carbohydrate metabolites are inherently reactive toward nucleophiles [1]. Thus carbonyl groups (electrophiles) of sugars interact with the amino groups (nucleophiles) of macromolecules (proteins, nucleic acids, and aminolipids) to form labile Schiff's base derivatives, which slowly isomerize to form ketoamine, adducts via the Amadori rearrangement. Subsequent to the Amadori rearrangement, the later less well understood stage, advanced glycation process begins resulting in the formation of Advanced Glycation end products (AGEs) [2]. Although to a certain extent glycation take place in all forms of life, in poorly controlled diabetic patients with hyperglycemia, elevated levels of glucose will cause increased glycation of proteins of plasma, other body fluids, intracellular proteins and extracellular matrix [3]. Hyperglycaemia and formation of AGEs contributes to the development of macrovascular disease, nephropathy, neuropathy and retinopathy [4]. AGEs once formed, can also bind to receptor for advanced glycation product there by resulting in the formation of reactive oxygen radical [5,6]. Some AGEs can form stable cross linking structures on protein there by leading to loss of function [7]. AGEs are also associated with increased coronary artery disease in diabetics [8]. Overall hyperglycemia and AGEs directly or indirectly is a major factor in the pathogenesis of diabetes, neurodegenerative diseases, as well as in the process of aging [9-11] Arginine and lysine being positively charged contributes significantly to the net charge of the protein. The net charge on the protein is the summation of charges carried by its electrostatic components including, charged amino acid residues, metal ions and charged cofactors [12]. However the effective charge of a protein is always less than the net charge and is determined by the extent of screening of the charge by the association of the protein by counter ions in the surrounding medium [12]. Charges on the protein can affect substrate binding and reactivity [13], protein-protein interactions [14] and secretion [15], localization of specific protein. Besides, osmotic pressure of a cell is affected by the sum of the charges of all of its protein [12]. In proteins arginine and lysine are the two residues predominantly glycated. Glycation of proteins and the subsequent formation of advanced glycation end products on arginine and lysine can convert the positively charged residues to a negative charge or a neutral charge depending on the chemical type of the AGEs [16, 17]. For instance the formation of carboxymethyllysine convert the positively charged lysine-ε-NH3+ to negatively charged carboxyl group. Although it might seem obvious that the effective charge of the protein on modification is

equal to the number of charge units blocked or modified. However this is a faulty assumption, since the change in protein charge upon modification is considerably more complex due to the phenomenon of charge regulation. Charge regulation in proteins take place in response to change in the charge of a protein by modification, the charge on the rest of the protein adjusts to the neutralization of charge on the ε-NH3+ in a way that reduces the total charge on the protein. Thus due to the charge regulation mechanism, the loss of positive charges on protein will considerably increase the net negative charge of the protein, which will result in increase in the local concentration of the hydrogen ion on the surface of the protein. This would considerably affect the acid base equilibrium of the charged amino acid there by resisting the change in the net charge [18]. Scheme 1 shows the physical basis of charge compensation mechanism. Here I take the example of carboxyl methyl lysine, an AGE formed by the reaction of lysine with glyoxal derivative [19] or autooxidation of early stage AGEs such as Amadori product [20]. At the physiological pH the carboxyl group of carboxyl methyl lysine would be negatively charged. Thus eliminating the positive charge from ε-NH+ group and replacing with negative charge would increase the net negative charge by 2 units. However this change would be less than 2 units due to charge regulation, due to the adjustment of protonation states of other ionizable residues predominantly histidine in a manner that reduces the total change in charge [18, 21].

2. MATERIALS AND METHODS

All the sequences for analysis were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov). Out of the 20 amino acids in protein, 5 of them have a pK values such that they should be partially charged at neutral pH. These includes acidic amino acids aspartic acid and glutamic acid, basic amino acids Arginine and lysine, and Histidine which somewhat behaves as a basic amino acid. At physiological pH, Asp and Glu with pK's far below 7.4 and lysine and arginine have pK's far above 7.4 are fully charged. Thus ASP and Glu contribute to –1 to the charge of the protein and Lys and Arg each contribte to +1. However His has pK of around 6-6.5 depending on the immediate environment in the protein. Thus the fraction of His charged at a given pH is given by rearranged Henderson-Hasselbalch equation [22] where pK is the pK value for the basic group. Taking a pK of 6.5 and at pH 7.4 the fraction of histidine charged is f =0.1. thus with the basic assumption for the charges on these 5 amino acids, the net charge Q on the protein is given by where nk ,nR , nH ,nD and nE are the total number of lysines, arginines, histidines, aspartic acids and glutamic acids respectively in the protein. Total number of lysine, arginine, histidine, aspartic acid and glutamic acid in aminoacid sequence of albumin and crystalline was counted using the program SAP (statistical analysis of protein sequences) available at http://www.isrec.isb-sib.ch/software/SAPS_form.html. The isoelectric point (pI) was calculated using the Expasy

isoelectric point calculator at http://www.expasy.ch/tools/pi_tool.html

HYPOTHESIS

The five charged amino acids, at physiological pH only histidine has a pKa close enough to be significantly influenced by the change in local hydrogen concentration. Glutamic acid and Aspartic acid has pKa far below and arginine and Lysine have far above the physiological pH. It is observed that effect of charge compensation is maximum at pH of 7.4, which is near to the pKa of histidine residues in proteins. Histidine contributes in many functional activities, including Cu2+, Fe2+, Zn2+ and heme coordination, is part of the classical catalytic triad of protease active sites and can adopt flexible roles in conformation [23]. Based on the experimental evidence of charge regulation in proteins and the important role played by histidine in the process. I here hypothesize a role for histidine to resist change in the net charge of the protein during glycation and other physiologically relevant nonenzymatic and deleterious post translational modification.

3. RESULTS AND DISCUSSION

It has been proposed that the properties of the protein are relatively unchanged in vivo, so that the glycated protein is tolerated within the body compartment [3]. If we assume that the (T1/2) is shortened as a result of Glycation, in that case a proportionate amount of energy must be expended by the body to catabolise the glycated protein and to replace the lost protein. Alternatively if we assume that the body tolerates glycated protein then there has to be a compensatory mechanism to nullify the effect of glycation. One of the major changes arising due to glycation is the loss of charge. As mentioned earlier, the charge of the protein plays a crucial role in many physiological processes [13, 14, 15]. Hence there exists a strong evolutionary pressure to maintain the net charge of the protein. Thus an increase in positive charge by mutation on the protein should be balanced by corresponding increase in the negative charge so that the net charge remains constant [24]. In fact it has been shown that there exists some evolutionary pressure to maintain the charge protein on carbonic anhydrase. Similarly we tested the hypothesis with albumin and crystallin proteins that are found to be glycated in vivo. Glycation of alpha crystalline by fructose, methylglyoxal and glucose form AGEs and contribute to cataract formation [25]. In human serum, albumin is the most readily glycated protein amounting for more than 80% of the content of the glycated molecules [26]. The plot of the normalized number of negative charged residue versus the normalized positive charged residues gave a straight-line graph with slope of 0.691, indicating each positive charge is balanced by only a 0.691 negative charges (Fig 1, table 1). Also it can be seen that species with high proportion of positively charged amino acid has a high proportion of negatively charged amino acid indicating that there exists some evolutionary pressure to maintain the charge on albumin. Similarly table 2 shows the net charge

in crystalline remains constant, which again shows that there is tendency to conserve the net charge.

DISCUSSION: If there is a strong evolutionary pressure to conserve the net charge on protein, then the changes in number of positive charges should be exactly balanced by corresponding changes in the negatively charged amino acids. Thus a mutation that increases or decreases the charge would be followed by a mutation that restores the charge to its original value. Unlike mutation where there can be increase or decrease of charge, post translational events such as glycation is most often followed by increase in net negative charge of the protein.In reality the situation is different in glycated protein, since AGEs shows chemical and physical diversity. For instance modification of BSA by methylglyoxal results in the formation argpyrimidine (7%), Nε -carboxyethyl-lysine (1%), methylglyoxal-derived lysine dimer and hydrimidazolone (91%) (27). The extent of glycation and the type of AGE formed in vivo is dependent upon the concentration of sugar, the type of sugar [1]. Further glycation is also influenced by presence of metal ions and oxidative conditions [28]. As a result more than one chemical type of AGEs may be formed on the protein. Thus in vivo the total charge on the glycated protein would be determined by the ionization behavior of various ionizable species including AGE structures formed and the charge compensation mechanism. However studies using specific modifying agent such as acetic anhydride where influence of complexity due to formation of more than one product can be avoided showed that charge regulation plays a crucial role. For instance the native charge on bovine carbonic anhydrase is –2.9, upon modification of all the 18 lysine the net charge is –19. The net change in the charge is approximately –16, which in the absence of charge regulation would be –18 [24]. Glucose has been selected as a principle currency of energy within the animal kingdom, since the divergence of the closed vascular system from the gastrointestinal tract in ancestral invertebrates. Also the selection of glucose over other metabolic sugars such as fructose, galactose has been proposed due to greater stability of its ring form compared to other aldose and ketose monosacharides and therefore it shows comparatively low reactivity with proteins [29]. In proteins, the extent of change in functional property will depend upon the quantity (number of arginines and lysines) as well as quality (which of the arginine and lysine is glycated). If the protein evolves to exclude variants expressing exposed and reactive lysyl and arginine residues, the variants that are relatively resistant to glycation will prevail in the population. However the reactivity of lysine and arginine is must for the overall functional integrity of the protein and also in disease state the increased sugar levels will glycate even those residues which are not inherently reactive. In such situation, glycation of protein is unavoidable and it is not feasible for the body to get rid of the glycated protein as it would expend extra amount of energy in replacing the lost protein [2]. The natural selection pressure within the host species, mutation of the gene's which yield

variants that are relatively resistant to glycation in our case buried histidine should be excluded and also proteins with more reactive lysine and arginine would have more number of histidine to nullify the effect of glycation. Analyses of the distribution of amino acids in proteins reveal that histidine is rather versatile as it is the only uniformly distributed amino acids in terms of the solvent accessibility levels in proteins [30]. The constant exposure of histidine would ensure that uniform charge compensation mechanism persists throughout the protein rather than localized effect.Besides glucose, trehalose is the circulating metabolic fuel of the animal kingdom insecta and is also the storage sugar of many species of fungi [31, 32]. Being non reducing, trehalose are non-reactive towards the lysine and arginine in proteins [33, 34]. This would allow insects to express proteins that are not under selective pressure on the evolution of lysine, arginine and histidine in proteins. Now it would be interesting to know that whether insect proteins, contain greater quantities of lysine and arginine, whether there exists freedom for the evolution of histdine in terms of its exposure and the correlation between positive charge residues and histidine.

CONCLUSION

The study presented in this manuscript shows that there exists a strong evolutionary pressure to maintain the net charge of albumin and crystalline. These proteins are target of glycation *in vivo*. A number of modifications by noxious chemical entities take place in an organism. For instance advanced lipid oxidation end products and free radicals modify proteins *in vivo*. Although the formations of these products are limited in normal individual, its level is increased in disease states. The hypothesis presented in this manuscript explains how modified protein can resist change in physical characteristics by the mechanism of charge regulation.

CONFLICT OF INTEREST

The authors have no conflict of interest.

Peer review under responsibility of Life Science Informatics Publications
2017 Jan- Feb RJLBPCS 2(5) Page No.69

REFERENCES

1.Sharma SD, Pandey BN, Mishra KP, Sivakami S. Amadori product and age formation during nonenzymatic glycation of bovine serum albumin in vitro. J Biochem Mol Biol Biophys. 2002;6(4):233-42.

2. Suji G and Sivakami S. Glucose, glycation and aging. Biogerontology. 2004;5(6): 365-73.

3. Hatton MW, Richardson M, and Winocour PD. On glucose transport and non-enzymic glycation of proteins in vivo. J Theor Biol. 1993; 161(4):481-90.

4. Huebschmann AG, Regensteiner JG, Vlassara H, Reusch JE. Diabetes and advanced glycation end products. Diabetes care 2006;29(6):1420-32.

5. Ramasamy R, Yan SF, Schmidt AM. The RAGE axis and endothelial dysfunction: maladaptive roles in the diabetic vasculature and beyond. Trends Cardiovasc Med. 2005; 15(7):237–43.

6. Wautier MP, Chappey O, Corda S, Stern DM, Schmidt AM, Wautier JL Activation of NADPH oxidase by AGE links oxidant stress to altered gene expression via RAGE. Am J Physiol Endocrinol Metab. 2001;280(5):E685–E694.

7. Eble AS, Thorpe SR, Baynes JW. Nonenzymatic glucosylation and glucose-dependent cross-linking of protein. J Biol Chem. 1983; 258(15): 9406–12.

8. Kilhovd BK, Berg TJ, Birkeland KI, Thorsby P, Hanssen KF. Serum levels of advanced glycation end products are increased in patients with type 2 diabetes and coronary heart disease. Diabetes Care. 1999; 22(9):1543–48.

9. Brownlee M. Advanced protein glycosylation in diabetes and aging. Annu. Rev. Med. 1995;46:223–34.

10. Colaco CA, and Harrington CR. Glycation: a pathological modification in neuropathies? A hypothesis Neuroreport 1994;5(8):859–61.

11. Thorpe SR and Baynes JW. Role of the Maillard reaction in diabetes mellitus and diseases of aging. Drugs Aging 1996;9(2):69–77.

12. Gao J, Gomez FA, Harter R, Whitesides GM. Determination of the effective charge of a protein in solution by capillary electrophoresis. Proc Natl Acad Sci USA 1994; 9(25)1:12027-30.

13.Garcia-Viloca M, Gao J, Karplus M, Truhlar DG, How enzymes work: analysis by modern rate theory and computer simulation. Science 2004;303(5655):186-195.

14. Davis SJ, Davies EA, Tucknott MG, Jones EY, van der Merwe, PA. The role of charged residues mediating low affinity protein –protein recognition at the cell surface by CD2. Proc. Natl. Acad. Sci. U.S.A. 1998;95(10):5490-94.
15. Stephenson K, Jensen CL, Jorgensen ST, Lakey JH, Harwood CR. The influence of secretory-protein charge on late stages of secretion from Gram-positive bacterium Bacillus subtilis. Biochem. J. 2000;350(1):31-39.
16. Luthra M, Balasubramanian D. Nonenzymatic Glycation Alters Protein Structure and Stability. A study of two eye lens crystallins. J Biol Chem 1993;268(24):18119-27.
17. Westwood ME, Thornalley PJ. Molecular characteristics of methylglyoxal-modified bovine and human serum albumins. Comparison with glucose-derived advanced glycation endproduct-modified serum albumins. J Protein Chem 1995;14(5):359-372.
18. Menon MK, Zydney AL. Determination of effective protein charge by capillary electrophoresis: effects of charge regulation in the analysis of charge ladders. Anal Chem 2000;72(22):5714-17.
19. Glomb MA, Monnier VM. Mechanism of protein modification by glyoxal and glycolaldehyde, reactive intermediates of the Maillard reaction. J Biol Chem 1995;2709(17):10017-26.
20. Frye EB, Degenhardt TP, Thorpe SR, Baynes JW. Role of the Maillard reaction in aging of tissue proteins. advanced glycation end product-dependent increase in imidazolium cross-links in human lens proteins. J Biol Chem 1998;273(30):18714-19.
21. Gitlin I, Mayer M, Whitesides GM. Significance of Charge Regulation in the Analysis of Protein Charge Ladders. J. Phys. Chem. B. 2003;107:1466-72.
22. Sear RP. The effects of added salt on the second virial coefficients of the complete proteome of E.coli. J Chem Phys 2003;118: 5157-61.
23. Creighton TE. Proteins, Structures and Molecular Properties (Freeman, New York). 1993.
24. Gudiksen KL, Gitlin I, Yang J, Urbach AR, Moustakas DT, Whitesides GM. Eliminating positively charged lysine □-NH3+ groups on the surface of carbonic anhydrase has no significant influence on its folding from sodium dodecyl sulfate. J Am Chem Soc. 2005; 127(13):4707-14.
25. Haik Jr GM, Lo TW, Thornalley PJ. Methylglyoxal concentration and the glyoxalase activities in the human lens. Exp. Eye Res 1994;59(4): 497–500.
26. Dolhofer R, Wieland OH. Increased glycosylation of serum albumin in diabetes mellitus. Diabetes

1980;29(6):417-22.

27. Ahmed N, Dobler D, Dean M, Thornalley PJ. Peptide mapping identifies hotspot site of modification in human serum albumin by methylglyoxal involved in ligand binding and esterase activity. J Biol Chem 2005;280(7):5724 - 32.

28. Baynes JW, Thorpe SR. The role of oxidative stress in diabetic complications: a new perspective on an old paradigm. Diabetes 1998;48(1): 1-9.

29. Bunn HF, Higgins PJ. Reaction of monosaccharides with proteins: possible evolutionary significance. Science 1981;213(4504):222-4.

30. Baud F, Karlin S. Measure of residue density in protein structures. Proc Natl Acad Sci USA 1999;96(22):12494-99.

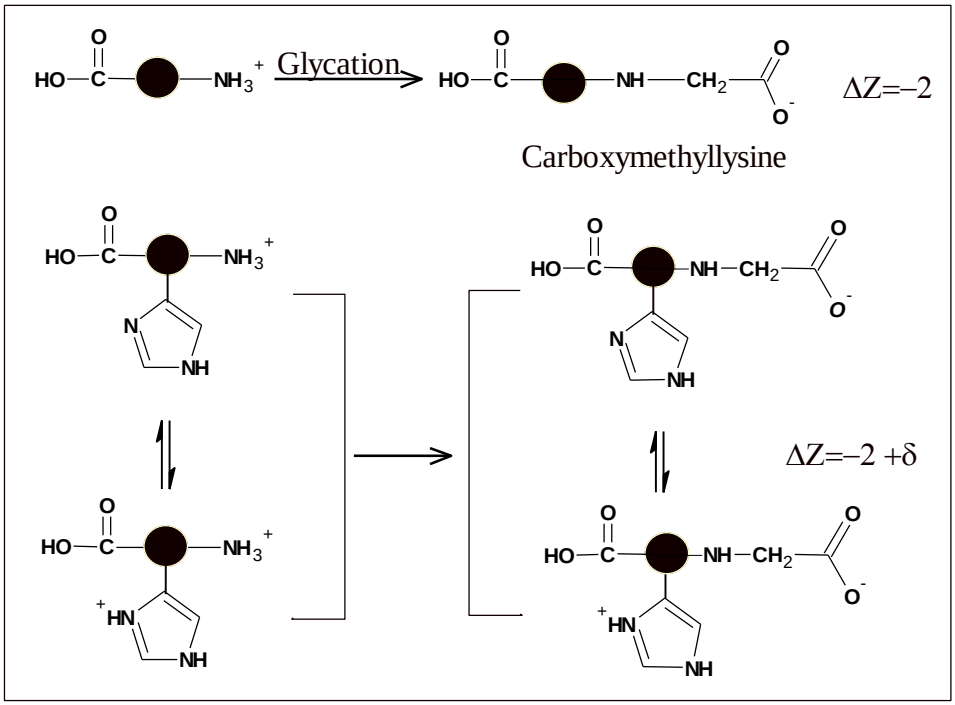
31. Matusuda H, Yamada T, Yoshida M, Nishimura T. Flies without trehalose. J Biol Chem 2015;290(2):1244-55.

32.Shukla E, Thorat LJ, Nath BB, Gaikwad SM. Insect trehalase: physiological significance and potential application. Glycobiology. 2015;25(4):357-67.

33. Motomiya y, Higashi T, Masuda M, Iwamoto H, Miura K, Yoshimura Y, Maruyama I. An in vitro evaluation of the glycation potential of a natural disaccharide, trehalose. Clin Exp Nephrol. 2003;7(3): 195-200.

34. Szwergold BS. Maillard reactions in hyperthermophilic archaea: implications for better understanding of nonenzymatic glycation in biology. Rejuvenation Res. 2013;16(4): 259-72.

Supplementary Files:



Scheme I. Physical basis of mechanism of charge compensation. Glycation results in formation of carboxymethyl lysine as a result change in the net charge of the protein is -2 (ΔZ=-2). Lower part of the scheme shows the charge compensation mechanism with histidine residue adjusting charge.

Label	Organism	pI[a]	Molecular Mass(Da)	Positive Residues[b]	Negative residues[c]	Total residues	Net Charge	Accession number
A	Human	5.67	66472	84.6	98	585	-13.4	P02768
B	Bovine	5.60	66432	83.7	99	583	-15.3	P02769
C	Dog	5.32	65710	80.2	98	584	-17.3	P49822
D	Cat	5.30	65845	76.8	100	584	-23.2	P49064
E	Rat	5.74	66915	75.1	84	587	-8.9	P36953
F	Horse	5.72	65752	84.9	99	583	-14.1	P35747
G	Pig	5.92	65966	80.8	91	584	-10.2	AAQ20088
H	Sheep	5.58	66327	83.8	100	583	-16.2	P14639
I	Mouse	5.44	66949	70.4	86	587	-15.6	O89020
J	Rabbit	5.65	66015	81.3	99	584	-17.7	P49065
K	Rhesus	5.67	65979	81.6	95	584	-13.4	Q28522
L	Orangutan	5.67	66472	84.6	98	585	-13.9	Q5NVH5
M	Salamander	5.40	65121	70.2	92	590	-21.8	P21848
N	Chicken	5.35	67189	74.3	91	592	-16.7	P19121
O	Elephant	5.62	65768	78.8	94	583	-15.2	AAT90502
P	Xenopus laevis	5.47	67501	75.5	106	583	-30.5	P14872
Q	Cobra	5.43	67946	73.0	94	598	-21	S59517

Table 1. Net charge and charged residues on albumin

[1] [a]The isoelectric point (pI) was calculated using the EXPASY isoelectric point calculator : http://expasy.org/tools/pi_tool.html. [b]The number of positively charged residues is the number of lysine + arginine + X × Histidine residues, where X is the fractional charge of the histidine residue at

pH of 7.4 (X=0.1 at pH 7.4). [c]The number of negatively charged residues is the number of aspartic acid + glutamic acid. [1]

Organism	pI[a]	Molecular Mass(Da)	Positive Residues[b]	Negative residues[c]	Total residues	Net charge	Accession number
Bovine	5.78	19790	20.7	25	173	-4.3	P02470
Rabbit	5.78	19837	20.7	25	173	-4.3	P02493
Gray Seal	5.78	19713	20.7	25	173	-4.3	P68289
Sea lion	5.78	19713	20.7	25	173	-4.3	P68288
Hippopotamus	5.78	19760	20.7	25	173	-4.3	P68285
Giraffe	5.78	19760	20.7	25	173	-4.3	P68284
Cat	5.78	19731	20.7	25	173	-4.3	P68282
Pig	5.78	19807	20.7	25	173	-4.3	P68281
Dog	5.78	19731	20.7	25	173	-4.3	P68280
Tree Shrew	5.78	19792	20.7	25	173	-4.3	P68406
Rhesus	5.77	19792	20.7	25	172	-4.3	P02488
Common tegu	5.78	19898	20.7	25	173	-4.3	P02506
Slider turtle	5.63	17214	17.6	22	149	-4.4	Q91517
Dog fish	5.39	20680	21.5	27	177	-5.5	P02509
Elephant	5.6	19837	19.7	25	173	-5.3	P02498
Opossum	5.88	19995	20.8	25	173	-4.2	P02503
Rhinoceros	5.78	19808	20.7	25	173	-4.3	P02479
Camel	5.78	19745	20.7	25	173	-4.3	P02472
Whale	5.78	19778	20.7	25	173	-4.3	P02474

Table 2. Net Charge and charged residues in alpha crystalline.

[a]The isoelectric point (pI) was calculated using the EXPASY isoelectric point calculator : http://expasy.org/tools/pi_tool.html. [b]The number of positively charged residues is the number of lysine + arginine + X × Histidine residues, where X is the fractional charge of the histidine residue at pH of 7.4 (X=0.1 at pH 7.4). [c]The number of negatively charged residues is the number of aspartic acid + glutamic acid.

Fig 1. Plot of negative amino acids divided by the total number of amino acids versus the number of positive amino acids divided by the total number amino acids for the albumin of different species listed in table 1 with label. The points D, P, G, E were excluded from linear regression analysis. The line of best fit has a slope of 0.691 and correlation coefficient of 0.912