CHAPTER 3

Nucleosides and Nitrogenous Bases Inhibit Protein Amyloidosis: A Biophysical and Computational Analysis Using Hen Egg White Lysozyme as Model System*

Abstract

The available literature indicates that the imbalance in metabolites like nucleosides and nitrogenous bases is correlated with amyloid-associated neurodegenerative diseases. The current study has investigated the potential anti-amyloidogenic property of selected intracellular nitrogenous bases like Cytosine, Guanine, Thymine, Uracil, and a few nucleosides. Adenosine and Guanosine. We have employed various spectroscopic, microscopic, and computational techniques considering Hen Egg-White Lysozyme (HEWL) as the model protein. The data suggest the anti-amyloidogenic nature of these metabolites. Our results indicate that the metabolites can drive the hen egg-white lysozyme towards forming aggregates over 120 hours in near-physiological and pH 12.2. Further, we have also shown a significant decrease in the aggregation index of the protein solutions incubated with various selected metabolites. Moreover, the molecular simulation studies also indicate that these metabolites have stable binding at amyloidogenic sites of HEWL and possibly prevent amyloid by inhibiting intermolecular interactions. During simulation, favourable secondary structural changes in the presence of these metabolites also indicate that the protein tends to move away from protein aggregation-mediated pathways. The results provide critical insights into the role of selected metabolites in modulating protein aggregation pathways and the pathogenicity of diseases caused due to their metabolic imbalance.

---

* Part of the work is published is submitted for publication

3.1. Introduction

Protein misfolding diseases cause neuropathic and non-neuropathic pathological conditions in humans, mainly due to protein misfolding and subsequent formation of amyloid or aggregates. Amyloid plaque, characteristic of various neurodegenerative conditions, is an ordered structure, mostly fibrous (Kundu *et al*., 2020). The acceleration in the rate of amyloid deposition is supposedly a primary event in the pathogenesis of Alzheimer's disease (AD). A wide array of endogenous factors could affect the rate of amyloid formation *in vitro* and play a role in the pathogenesis of AD (Evan *et al*., 1995). Cellular systems have their quality control mechanism, maintaining a balance between protein production and degradation. Deregulating these systems could prove vital for being a causative factor in various diseases, primarily neurological disorders and systemic amyloidosis (Bukau *et al*., 2006, Chaari *et al*., 2015; Linder and Dermarez *et al*., 2009). Nucleotide balance is an essential factor in both dividing and quiescent cells. The emphasis is more on neural tissues as there is a high rate of metabolism and hence a high requirement of ATP (Kundu *et al*., Federico *et al*., Fasullo and Endres, 2015) for proper functioning. One exciting aspect of protein misfolding diseases that has come into being is the existence of intermediates during the formation of amyloid fibrils or complete protein unfolding. These oligomeric intermediates are potentially more cytotoxic and have also increased the scope of further research (Ghosh *et al*., 2014).

Previous studies have reported that most proteins and short peptides can form *in vitro* aggregates and amyloids, which shows that the formation of amyloids is majorly an intrinsic property. Hen egg-white lysozyme (HEWL) is a suitable candidate as a model protein to investigate various mechanisms of studying amyloid fibrillogenesis (Imoto *et al*., 1972, Das *et al*., 2018). Small globular protein, six Trp residues and high homology with human lysozyme make it suitable for studying the effect of various metabolites in HEWL amyloidosis formation and degradation (Das *et al*., 2018, Strynadka and James, 1991). The hen egg-white lysozyme amyloids generated *in vitro* are also toxic to cell cultures (Gharibayan *et al*., 2007, Chaari *et al*., 2015). In alkaline pH hen, egg-white lysozyme forms molten globule states and amyloid formation reported in previous studies

(Ghosh *et al*., 2014, Dhulesia *et al*., 2010, Kumar *et al*., 2008, Kumar *et al*., 2009). The amyloid fibrils formed by human lysozyme in diseased conditions are similar in shape and morphology to the fibrils formed by hen egg-white lysozyme under *in vitro* conditions (Ghosh *et al*., 2014, Arnaudov and Vries, 2005). *In vitro* studies in hen egg-white lysozyme is also appreciated because of the well-acclaimed functions of HEWL fibrils in membranes and their capability to induce apoptosis in neuroblastoma cells (Gharibayan *et al*., 2007, Ghosh *et al.*, 2014).

In the current study, we have tried to elucidate the effect of selected nitrogenous bases viz—cytosine, guanine, thymine and uracil, and nucleosides viz. adenosine and Guanosine on amyloid formation using HEWL amyloids as a model. We have also tried to investigate the possible formation of intermediate states via ANS assay and have used atomic force microscopy (AFM) to monitor changes in morphology and shape of the possible oligomers and amyloid fibrils. Further, we have monitored the aggregation index and turbidity both with and without metabolites. We have employed a dynamic light scattering technique to gain an insight into the size and heterogeneity of the protein aggregates in solution, both with and without selected metabolites. Lastly, we have utilised molecular simulation studies based on our previous molecular docking studies to gain more insights into the possible mechanism of the effect of metabolites on the model protein.

3.2. Materials and Methods

3.2.1 Materials: We procured Hen Egg White Lysozyme (HEWL) (EC 3.2.1.17) lyophilised powder (L6876) from Sigma Aldrich (5g), storage at -20ºC. Metabolites Adenosine, Cytosine, Guanine, Guanosine, Uracil and Thymine, were procured from Himedia (5g) Animal Cell Culture tested and stored at respective temperatures. We procured thioflavin T (ThT) (RM10365) and ANS (A5144) from Sigma Aldrich and Tris Buffer (GRM262) from Himedia. We performed all the experiments in triplicates and employed statistical analysis to obtain the significance of the study.

3.2.2 Methods

3.2.2.1 Buffer, Protein and metabolites solutions: We carried out the studies employing 50mM Tris HCl, pH 7.4, and Tris NaOH buffer, pH 12.2. Studies at pH 7.4 were done in buffer Tris-HCl to understand the pH-dependent behaviour of protein and metabolites. The buffers were freshly

prepared and filtered through a 0.22 µm Sartorius filter of 45mm diameter. We measured the absorbance of the buffers at 280 nm, 350 and 405nm. The absorbance of the buffer at all the wavelengths was below 0.1. We stored the buffers at 4ºC for no more than fifteen days.

We prepared a stock solution of HEWL at 500µM by dissolving the purified protein of the required amount in the filtered buffer and storing it at 4ºC until used. The working concentration of the protein was 50µM, prepared right before the preparation of samples. We prepared working solutions of HEWL with and without the metabolites in a ratio of 1:10 concentration.

We prepared stock solutions of all the chosen metabolites at a concentration of 1mM, and we optimised the working concentration to be 500 µM. Essentially the ratio between protein and the ligand in solution was 1:10 in a reaction mixture of 1ml, and the addition of the same buffer made up the remaining volume. The buffer for each set of experiments at different pHs was the one for that experiment. All the solutions were stored at 4ºC until used for no more than 15 days.

3.2.2.2 Hen Egg White Lysozyme Amyloid formation*:* The final cocktail for HEWL amyloids consisted of 50µM of the protein solution with and without metabolites in a ratio of 1:10, 10 mM NaCl in both pH 12.2 and 7.4, respectively, incubated at 55ºC, at 180 rpm (Ghosh *et al*., 2014, Choudhury and Kumar, 2021). Initially, the objective was to see the formation of amyloids until the initial oligomeric state, which we achieved by keeping the protein solution under these conditions for 72 hours. We have utilised three-time points in the current study, 24, 72, and 120 hours.

3.2.2.3 Thioflavin T assay: We utilised Cary Eclipse Varian Fluorescence Spectrophotometer, Agilent Technologies, to monitor the intrinsic fluorescence property of the protein, Thioflavin T and ANS assays. The slit widths for excitation and emission were kept at 5nm. 1mM stock solution of ThT was weighed and dissolved in 50mM Tris buffer, pH 7.4, and 5% sodium azide was added to the final solution. The entire solution was filtered using a 0.2 µM and stored at 4ºC wrapped in an aluminium foil. We freshly prepared all the solutions at room temperature. The ratio of the protein and ThT used for this assay was 1:2. 100µM of Thioflavin T was added to the solution approximately 30 minutes before the fluorescence was measured at room temperature. We scanned

the ThT assay from 460 nm-610 nm with an excitation at 445 nm and λmax 483-488 nm (Kumar *et al*., 2008).

3.2.2.3 ANS Assay*:* 1-anilinonaphthalene–8-sulphonate dye interacts non-covalently with the protein's hydrophobic and charged groups. Based on this fact, the relative hydrophobicity of the surface of a protein is often quantified (Klotz and Hunston, 1971, Hawe *et al*., 2008, Brudar and Lee-Hribar, 2019). We prepared a 1mM stock ANS solution following a protocol similar to the stock ThT solution. Samples prepared were kept in the dark prior to preparation for the measurement. The ratio of the protein and ANS used for this assay was 1:2. 100µM of ANS was added to the solution approximately 30 minutes before the fluorescence was measured at room temperature. We scanned the ANS assay from 400–600 nm., with excitation at 380 nm. We used a 1 cm quartz cuvette for fluorescence measurements with excitation and emission slit width at 5 nm.

3.2.2.4 Aggregation Index measurements: We used Cary Vis UV Spectrophotometer, Agilent Technologies, to measure protein solutions' absorbance with and without metabolites and buffers. We measured the aggregation index of the HEWL amyloids with and without the metabolites for each period. We measured the absorbance of the protein solution at 280 nm, followed by absorbance at 350 nm. Further, we calculated the aggregation index using $OD_{350}/ (OD_{280}-OD_{350})$.

3.2.2.5 Dynamic Light Scattering: To measure the heterogeneity of the samples prepared, coupled, and estimated aggregate size, we employed Dynamic Light Scattering (DLS) technique. We measured the size and heterogeneity of the control and the treated samples near-physiological and pH 12.2 after 120 hours of incubation. We used the Malvern Zetasizer Pro, Blue Label Instrument (4mW He-Ne 632.8 nm laser). The samples were pre-incubated at 25ºC for 120 seconds prior to the run. The results obtained were of five batch scans, each replicated five times. We performed the analysis using ZS Xplorer software.

3.2.2.6 Atomic Force Microscopy (AFM): We employed Atomic Force Microscopy imaging of HEWL amyloids with and without metabolites. We carried out the imaging using NT-MDT Nova Technologies. We used 10µl of prepared HEWL sample oligomers, both untreated and treated after

five days of incubation. We cleaned the substrate slides using leboline, methanol and 2-propanol for 10 minutes each in a professional ultrasonic cleaner prior to using. The samples were applied onto freshly washed silicon dioxide slides with 2µl of 10mM Magnesium Chloride for better sample adherence. After a few minutes, we rinsed the samples with milli Q water and removed extra water using filter paper. We kept the samples for drying overnight in a dust-free environment (Sarkar *et al*., 2009). We collected the AFM images in semi-contact (tapping) mode. The cantilever force constant was 3.5N/m±20%, with a resonate frequency of 140 kHz and a scan rate of 0.5 Hz, 256 lines per second over an area of 10µm * 10 µm. We acquired and analysed the AFM images using NT-MDT software Nova Px (Molecular Imaging and tools for Nanotechnology) v.3.2.5.

3.2.2.7 Aggregation Prone Zones and pH-dependent disorder prediction: We used Aggregator 3D 2.0 to predict the amyloid propensity regions of the protein (Kuriata *et al*., 2019). Further, we also evaluated the pH-dependent behaviour of the protein in terms of hydrophobicity and estimated disorder scores using the DispHScan web server (Pintado-Grima *et al*., 2019, Santos *et al*., 2020).

3.2.2.8 Molecular Dynamics and Simulation: Based on our previous molecular docking studies (Kundu *et al.*, 2020), we have employed molecular dynamics using GROMACS v2018.8 of the chosen protein-ligand complexes. We have proceeded with the biophysical investigations with those metabolites based on their binding energy from molecular docking. We have performed 100ns molecular simulations of the chosen complexes utilising the GROMOS 54a7 forcefield. As described previously in our works, we generated topology files for the metabolites from the PRODRG server (Van Alten *et al*., 1996, Pande *et al*., 2021, Kundu and Dubey, 2021). We set up the simulation using the methodology described in our work (Pande *et al*., 2021, Kundu and Dubey, 2021). As for the binding energy analysis post-simulation, we utilised the widely accepted MM/PBSA method. We also utilised the g_mmpbsa tool of GROMACS, calculating the per residue-free energy analysis (Pande *et al*., 2021, Kundu and Dubey, 2021, Hess *et al*., 1997, Kawata and Nagashima, 2001).

3.2.2.9 Statistical analyses using OriginPro (Learning Edition)

We have utilised the OriginPro Learning Edition to perform the statistical analyses and generate the graphs. We performed the statistical analyses, calculated standard deviations and p- values for all experiments performed in triplicates (n=3), and measured the significance level for p values <0.05.

3.3 Results

3.3.1 HEWL aggregates have a higher growth rate in near physiological pH than pH 12.2

We have characterized the formation of an oligomeric load of hen egg-white lysozyme at both near physiological pH and pH 12.2. Recent studies have also shown the formation of non-fibrillar aggregates of HEWL at near physiological pH (Al-Adem *et al*., 2020). Our results show that the growth rate of amyloid fibrils is higher in near physiological pH than pH 12.2 for the same concentration of protein used, and all other conditions replicated. In other words, in extreme pH 12.2, the transition rate is slowed between the lag and exponential phase of amyloid fibrillation kinetics. To validate this, we have shown higher turbidity, aggregation index, and ThT intensity. We report no significant differences in the oligomeric load between 0 and 24 hours but a significant difference at 72 and 120 hours. We observed the same in both the pH (Fig 3.1 A-F).

We monitored the amyloid fibrillation kinetics for HEWL at 20μM and 100μM to validate the statement mentioned above at near-physiological and pH 12.2. We measured the lag time, growth rate constant, and t_{50} time for all the conditions and tried to understand the ThT kinetic profile (Table 1). We normalised the data using the Boltzmann Equation (Moorthy *et al*., 2015). Further, we utilised Amylofit- The Aggregation Fitter webserver, to fit the normalised data for obtaining the primary nucleation rate (k_n) (Miesl *et al.*, 2016). We employed the nucleation elongation model for the same. The kn helped us get more insights into understanding the role of primary nucleation as a function of pH and concentration.

Table 3.1- ThioflavinT Kinetics profiling of HEWL at 20 and 100µM concentration at pH 7.4 and pH 12.2

pH	Conc (µM)	Lag phase (hrs)	t_{50} (hrs)	Growth rate (h^{-1})	Primary nucleation rate (M^{-1} hr^{-1})	Kinetics Mode
7.4	20	127.9	140.10	0.16	4.43	Sigmoidal
	100	52.56	121.00	0.029	0.001	Sigmoidal
12.2	20	182.99	206.53	0.084	1.99	Sigmoidal
	100	87.85	113.37	0.078	0.36	Sigmoidal

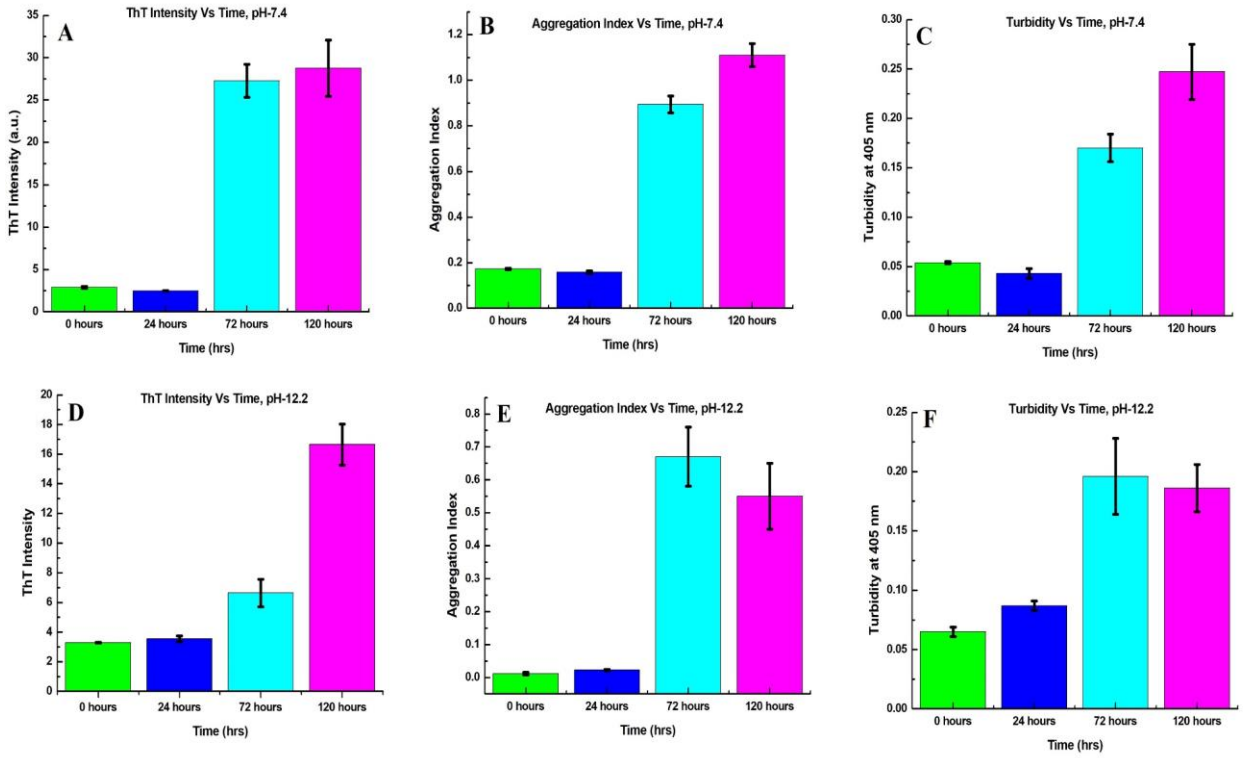

Fig 3.1 Hen egg-white lysozyme aggregation trends at different time points: (A) Thioflavin T intensity **(B)** Aggregation index **(C)** Turbidity at pH 7.4 and **(D)** Thioflavin T intensity **(E)** Aggregation index **(F)** Turbidity at pH 12.2.

3.3.2 ThT and ANS assay at near physiological pH- Thioflavin T is a commonly used dye to detect the formation of amyloids *in vitro*. The fact that the emission of the dye is directly proportional to the amyloid content in the solution makes it a gold standard test for the detection of amyloids (Sarkar and Dubey, 2011). When ThT is exposed to β-sheet regions of the amyloid in solution, it interacts with them, resulting in a change in the λ max emission shifting from 445 nm to 482 nm respectively, when excited at 450 nm (Naiki *et al*., 1989, Srivastava *et al*., 2010, Xue *et al*., 2017).

We have used metabolites from two diverse classes in the current study: nitrogenous bases and nucleosides. In the first 24 hours, the control sample (hen egg-white lysozyme only) shows no significant formation of oligomeric aggregates. Simultaneously when the same sample is incubated with nucleosides, adenosine and Guanosine, we observe no significant change in ThT intensity. After 72 hours, among the nitrogenous bases, guanine, thymine, and uracil reduce ThT intensity (Fig 2A-B). In contrast, cytosine increases ThT intensity. In the ANS assay, which detects the extent of hydrophobic patches on protein surfaces and oligomerisation state, we observe that ANS intensity lowers in the presence of three bases, viz. cytosine, thymine and guanine. Only uracil-treated lysozyme samples showed an enhanced ANS intensity than the control sample (Fig 3.2 C-D).

Among the two nucleosides, adenosine shows no change in ThT intensity, although there is a reduction in the ANS intensity. Guanosine reduces both ThT and ANS intensity after 72 hours. After 120 hours, we observe a reduction in ThT and ANS intensity in samples where protein and metabolites are simultaneously incubated (Fig 3.4).

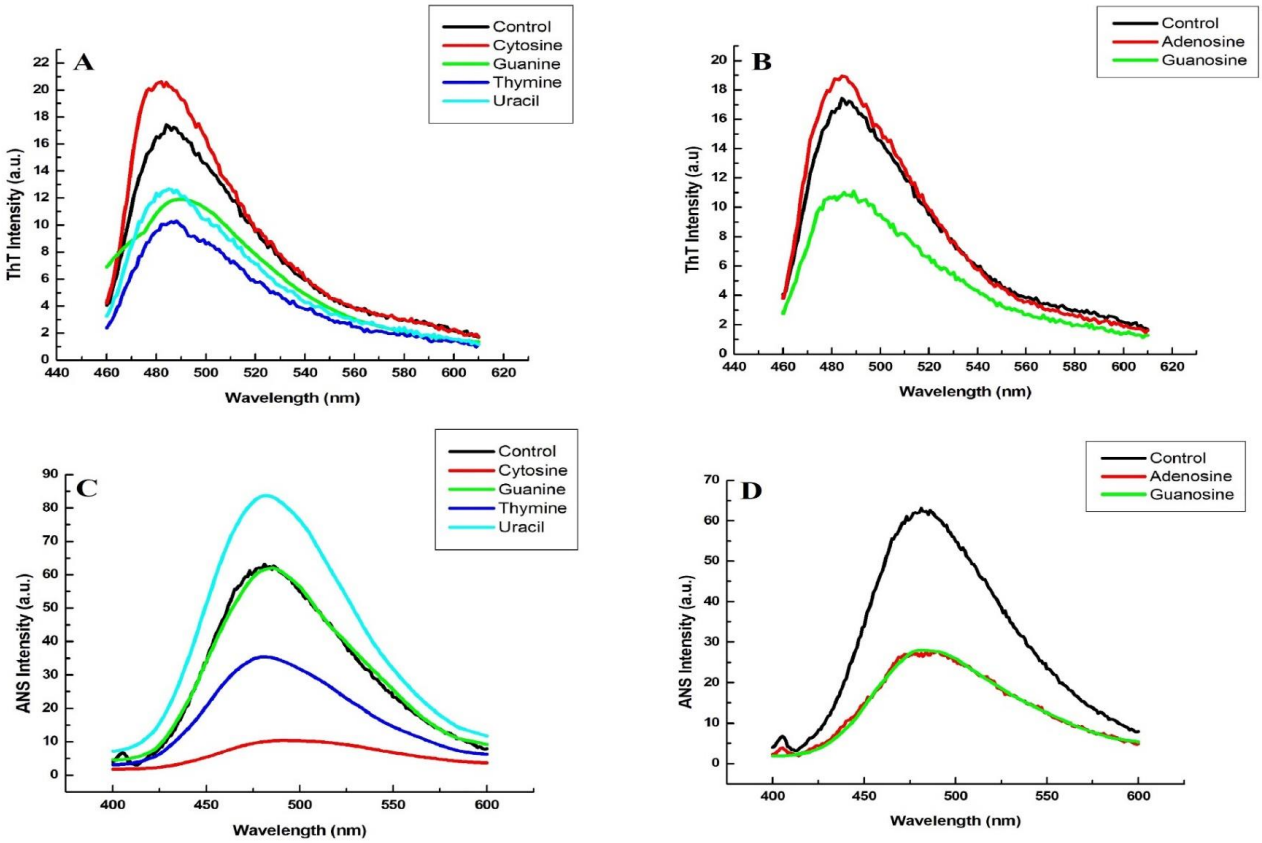

Fig 3.2 Thioflavin T and ANS intensity of 50 µM HEWL samples incubated with selected nitrogenous bases and nucleosides at pH 7.4: (A) Thioflavin T intensity with nitrogenous bases **(B)** Thioflavin T intensity with nucleosides **(C)** ANS intensity with nitrogenous bases and **(D)** ANS intensity with nucleosides after 72 hours of incubation.

3.3.3 Aggregation index at near physiological pH: Further establishing the effect of the metabolites in lowering the amyloid oligomeric aggregates, we show a significant reduction in the aggregation index (AI) of the protein solution in the presence of adenosine, Guanosine and guanine. In the presence of cytosine and thymine, we observe no significant changes in AI. Comparing our results of ThT and ANS, we observe a significant reduction in AI time-dependent in the presence of metabolites irrespective of the class of metabolites (Fig 3.5A & 5B).

3.3.4 ThT and ANS assay at pH 12.2: We conducted simultaneous studies replicating the same protocols at pH 12.2. Among the nitrogenous bases, cytosine, uracil, and thymine enhance ThT intensity after 72 hours compared to guanine. Slightly higher ANS intensity in the presence of the metabolites after 72 hours indicates the oligomeric state formation in the presence of nitrogenous bases. Among the bases, only thymine reduces ANS intensity. Between the two nucleosides, both

adenosine and Guanosine reduce ThT. Adenosine further reduces ANS intensity, whereas Guanosine increases ANS intensity (Fig 3.4A-4D).

We observe a significant change in ThT intensity after 120 hours of incubation. We also see the metabolites being more effective at this point in lowering ThT intensity. In the ANS assay, all the metabolites except guanine had lowered ANS intensity than the control. Lower ThT intensity and higher ANS intensity for protein samples in the presence of guanine probably indicate a distinct form of an intermediate state between oligomers and full-length amyloid fibrils (Fig 3.4 A- 4D).

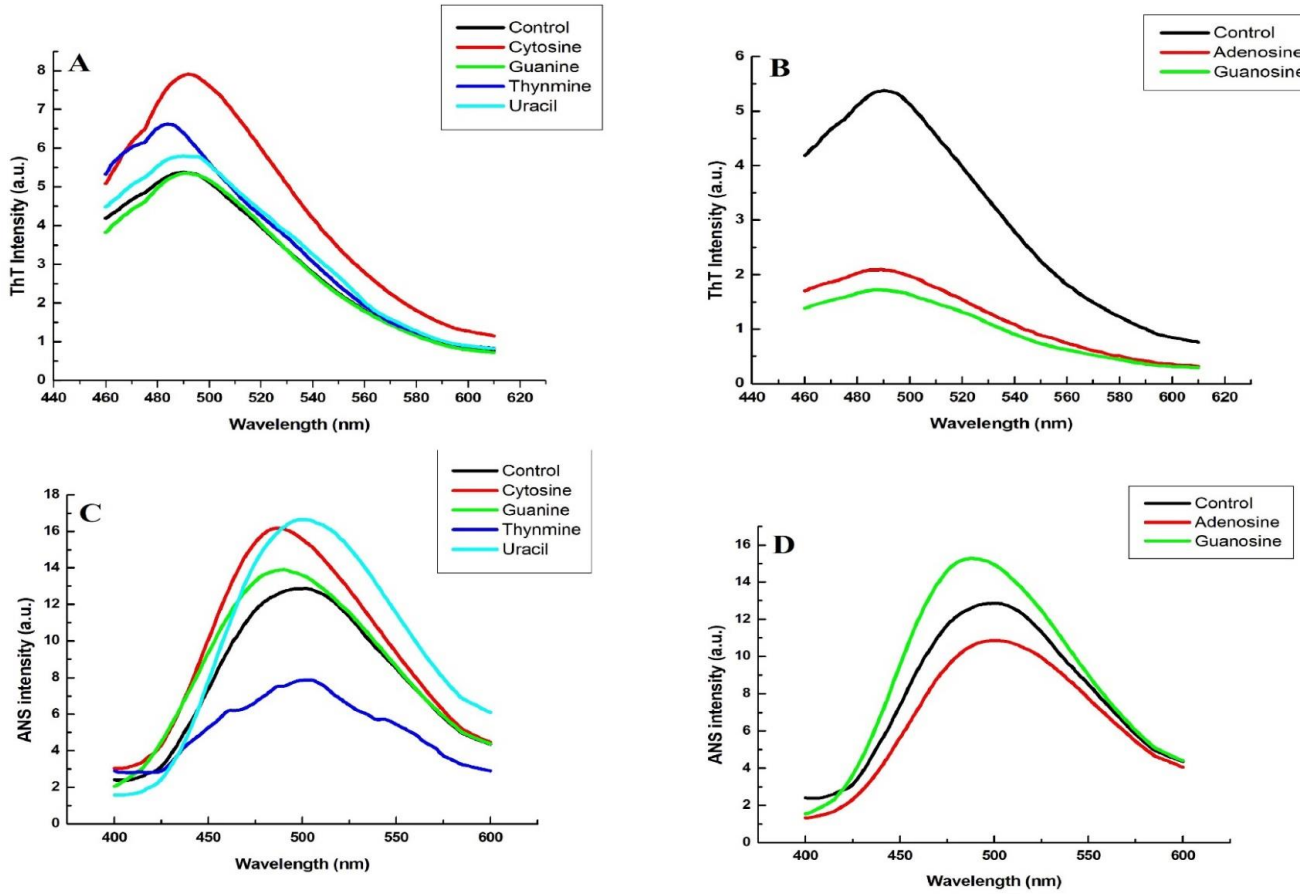

Fig 3.3 Thioflavin T and ANS intensity of 50 µM HEWL samples incubated with selected nitrogenous bases and nucleosides at pH 12.2: (A) Thioflavin T intensity with nitrogenous bases **(B)** Thioflavin T intensity with nucleosides **(C)** ANS intensity with nitrogenous bases and **(D)** ANS intensity with nucleosides after 72 hours of incubation.

3.3.5 Aggregation index at pH 12.2: We monitored the aggregation index and protein solution turbidity at pH 12.2, both with and without metabolites. Throughout the study, there were significant changes in the AI in all the lysozyme solutions incubated with metabolites at different time points compared to the control. Comparing our results of ThT and ANS, we observe a significant reduction in AI time-dependent in the presence of metabolites irrespective of the class of metabolites (Fig 3.5C-3.5D).

3.3.6 Dynamic Light Scattering Study (DLS): We carried out our DLS studies after incubating the samples for five days with and without metabolites at both pHs. At near physiological pH, the peak intensity is higher in samples treated with metabolites. The Z-average hydrodynamic radii of control aggregates are 229.8 nm. Among nucleosides, adenosine lowers aggregate size with a hydrodynamic radius of 7.57 nm and Guanosine around 197.6 nm. Among the nitrogenous bases, uracil shows a vigorous-intensity indicating a higher homogenous population in the solution with a radius of ~ 125 nm. In the presence of Guanine, Thymine and Cytosine, aggregates of lower hydrodynamic radii compared to control samples ranging between 145-200 nm. (Fig 3.6A). The polydispersity index of all the samples is within 0.7.

At pH 12.2, there is a stark difference in the heterogeneous nature of the aggregates compared to near physiological pH. The higher intensity in samples in pH 12.2 indicates homogeneity of the sample in shape and size. The control sample showed a maximum population of aggregates with a Z-average of 1433 nm and a tiny population of around ~100 nm. In ligand-treated samples, aggregates are the smallest in the presence of Guanosine (778 nm). All the other lysozyme aggregates in the presence of metabolites show a similar population size and intensity compared to the control sample (Fig 3.6 B), with hydrodynamic radii of more than 1000 nm. The detailed results are represented in Table 3.2.

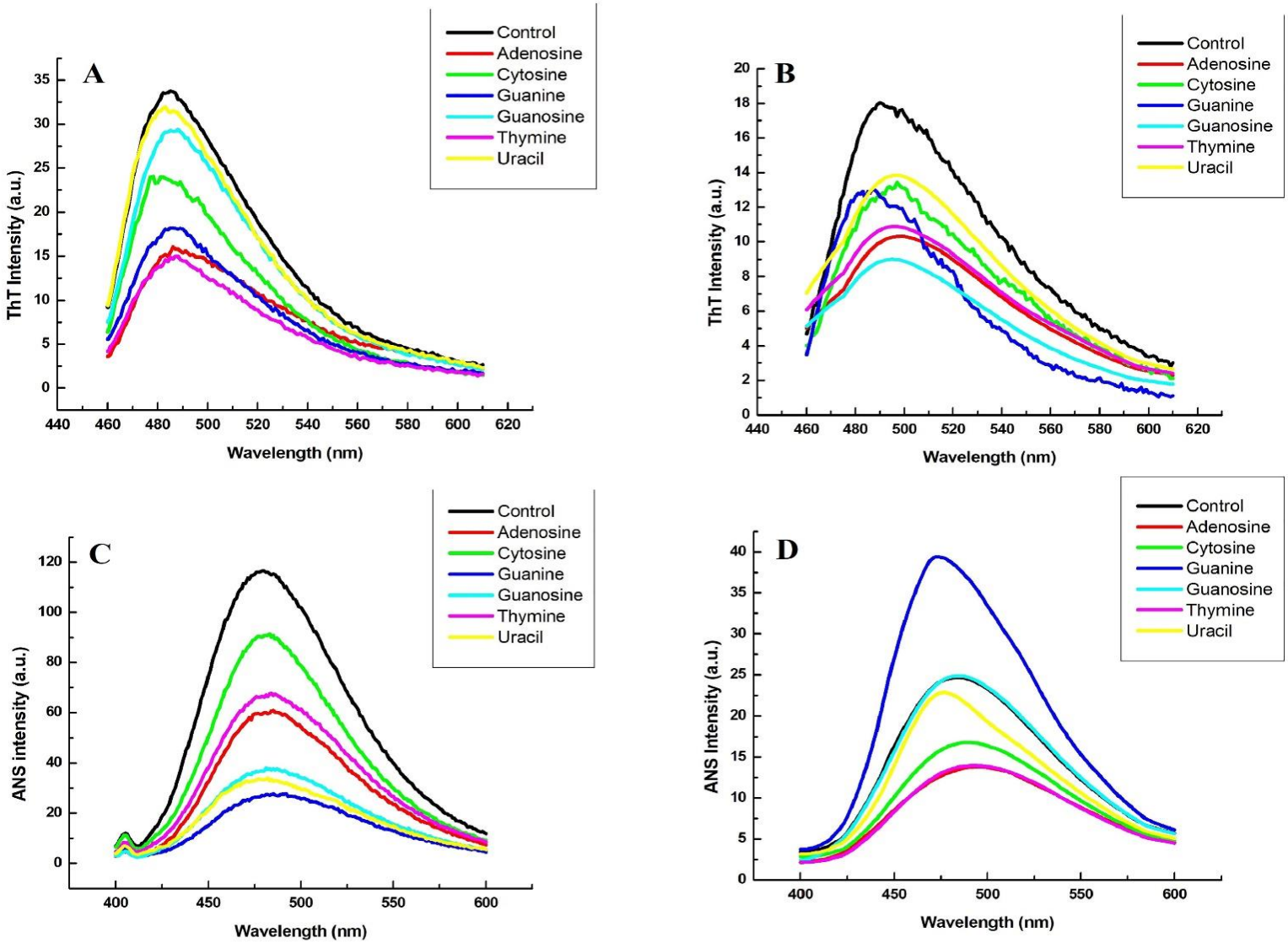

Fig 3.4 Thioflavin T of HEWL samples incubated with various ligands at 1:10 concentration after 120 hours: (A) pH 7.4 **(B)** pH 12.2, and ANS intensity at **(C)** pH 7.4 and **(D)** pH 12.2.

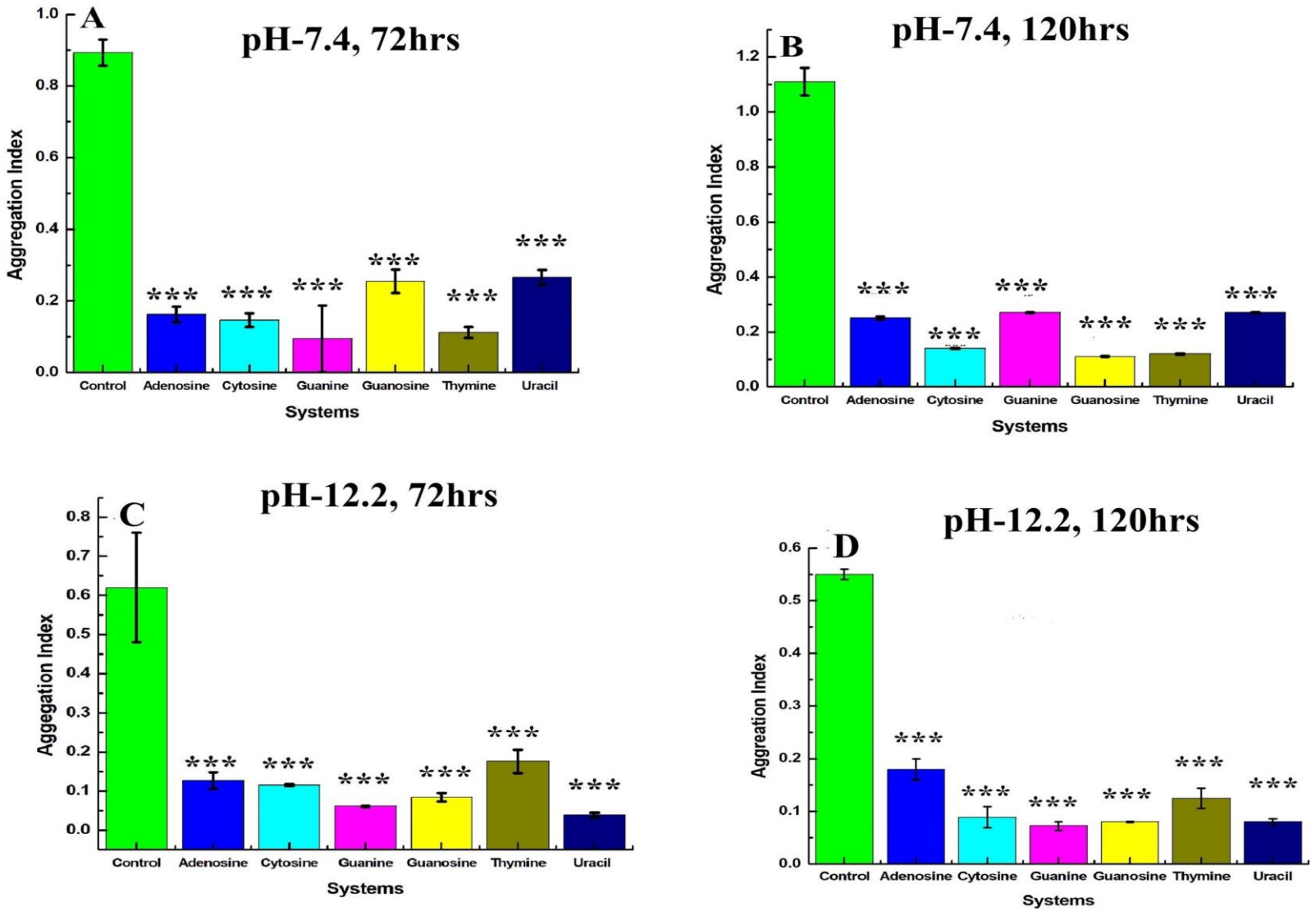

Fig 3.5 Aggregation Index of HEWL incubated with bases and nucleosides at pH 7.4 and 12.2: (A) 72 hours **(B)** 120 hours **(C)** 72 hours, and **(D)** 120 hours, respectively (all p values have been calculated compared to control with p-value <0.05).

Table 3.2- Dynamic Light Scattering analysis of physiological and alkaline pH HEWL aggregates with and without ligands

pH	System	Polydispersity Index (PI)	Z-average (nm)
7.4	Control	0.4498	229.8
	Adenosine	0.4768	7.57
	Cytosine	0.5605	197.6
	Guanine	0.6428	146.1
	Guanosine	0.2926	141.5
	Thymine	0.4854	197.6
	Uracil	1.2	125.6
12.2	Control	0.799	1433
	Adenosine	0.6496	1312
	Cytosine	0.6175	778
	Guanine	0.616	1232
	Guanosine	0.5849	1015
	Thymine	0.6271	1239
	Uracil	0.6492	1368

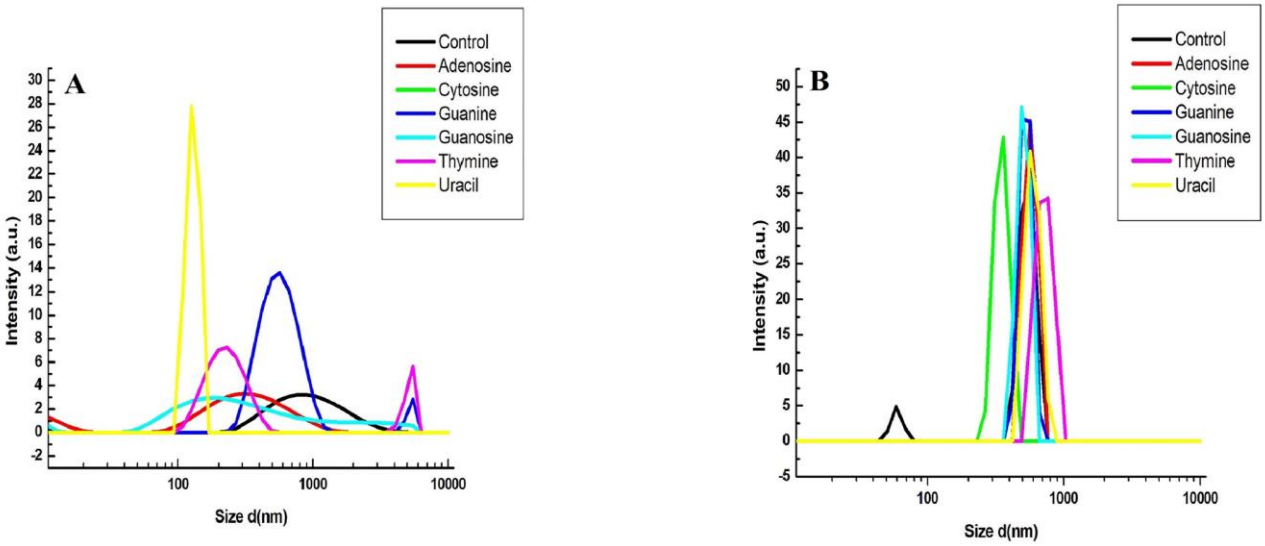

Fig 3.6 Dynamic Light Scattering of Hen Egg White Lysozyme aggregates: (A) pH 7.4 and **(B)** pH 12.2 with and without ligands after 120 hours of incubation at 37°C.

3.3.7 AFM imaging of HEWL oligomers /amyloids: We employed atomic force microscopy to monitor the structural morphology of the changes in the HEWL oligomers with the treatment of metabolites. We have discussed the height of the respective samples here for both pHs. At the day 5 stage in pH 12.2, the samples treated with metabolites produced aggregates of slightly higher heights compared to the control sample. The average length varied mainly between ~0.3μm to ~1.0μm in treated samples. The height of the control sample measured up to 0.419μm.

Further, at pH-7.4, we obtained amyloid fibrils up to a length of 0.543μm in the control samples. The HEWL fibrils treated with various metabolites at this pH produced heights ranging from 0.33 to ~0.45μm. The results here show that the metabolites affect the overall length and morphology of the HEWL amyloid fibrils (Fig 3.7). The histogram plots are represented in Fig 8A-B.

3.3.8 Aggregation prone regions and pH-dependent disorder: The Aggrescan 3D 2.0 software helped us predict the amyloidogenic regions in the entire protein. More than 50% of the primary structure sequence is prone to form aggregates. Since such a high percentage of the sequence is prone to form aggregates, the protein can quickly form soluble/insoluble aggregates in unfavourable environmental conditions. We further evaluated the possible effect of the pH changes in the protein in terms of disordered content using the DispHScan server. We observe that at pH 12, the protein has the highest disordered score and a minimum score at neutral pH (Fig 3.9-A). It is also interesting to note that, in neutral pH, there are more aggregation-prone zones than pH 12.2. Our experimental methods observe that the protein forms more soluble aggregates at neutral pH, making the solutions more turbid. The protein forms insoluble aggregates at pH 12.2 visible precipitates in the solution. We have also predicted the pH-dependent disordered regions in HEWL (Fig 3.9-B).

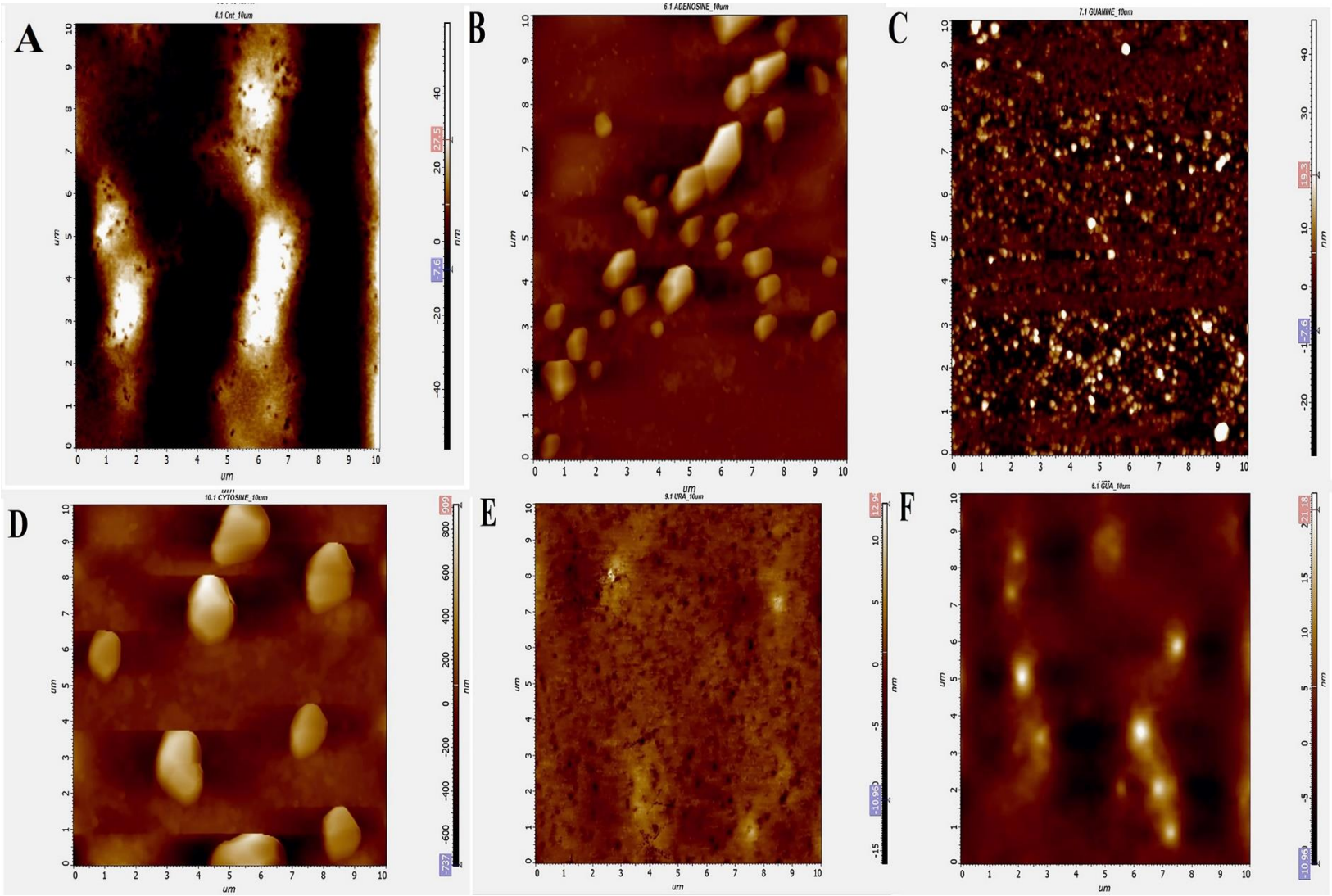

Fig 3.7 AFM images of HEWL amyloid aggregates/fibrils in the absence and presence of ligands after 72 hours of incubation at pH 12.2 (A) Control **(B)** treated with Adenosine **(C)** treated with Guanine **(D)** treated with Cytosine **(E)** treated with uracil and **(F)** treated with Guanosine.

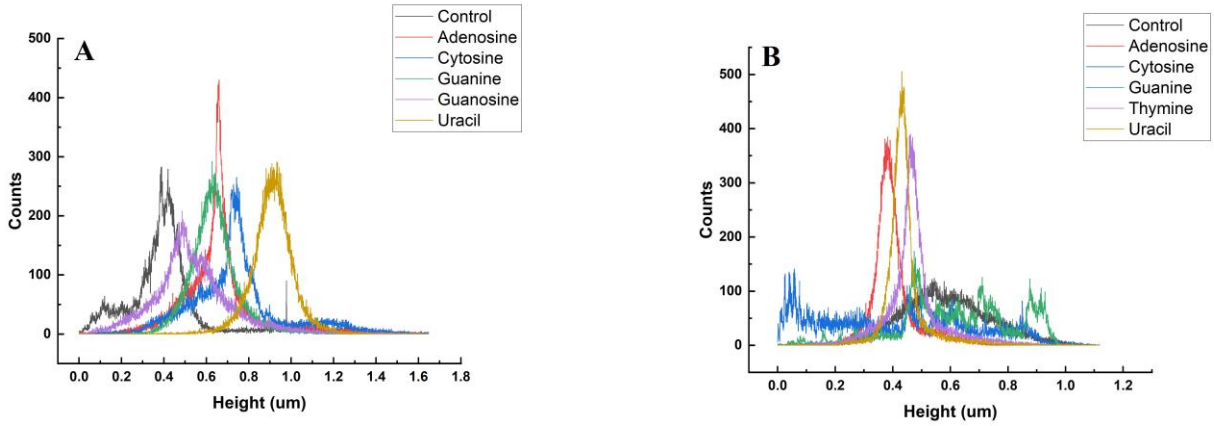

Fig 3.8 Histogram analysis of AFM images: Showing Heights Vs Counts of various systems in **(A)** pH 12.2 and **(B)** 7.4

3.3.9 Molecular Dynamics and Simulation

3.3.9.1 RMSD, RMSF and Radius of Gyration analyses: We performed molecular dynamics and simulation for 100 ns to get more insights into overall protein stability, compactness, and structural conformations in the presence of metabolites. When evaluated for all the systems, the root mean squared deviation analyses showed that all the systems gained stability within a 10-15 ns period, beyond which the systems were broadly stable with minor fluctuations. The overall stability of the systems was satisfactory. The RMSD of HEWL reached a maximum value of 0.25 nm. In the presence of the metabolites, uracil and guanine showed slightly lower RMSD with overlapping trajectories stabilising around a maximum value of 0.2 nm. Adenosine and thymine did not significantly change the system's overall stability compared to the HEWL protein. In the presence of cytosine, the system had a slightly higher RMSD indicating a possible change in structural conformations during the simulation period. The RMSD rose around 0.30 nm in the presence of cytosine. The radius of gyration analyses yielded the overall compactness of the HEWL and the HEWL complexes with metabolites. Guanosine among nucleosides showed slightly higher fluctuations. Validating our RMSD analysis, the radius of gyration also indicates no significant destabilisation effect on HEWL due to the presence of the metabolites. Cytosine and adenosine produced some deviations in their trajectories. The overall RMSD trajectories spanned between 1.36 nm to 1.44 nm, indicating maintenance of both stability and compactness during the

simulation. In line with RMSD and radius of gyration, the overall RMSF trajectories also show that the side chain residues are not much exposed to the solvent and have minimal fluctuations and solvent exposure. Although when we take a deeper insight into specific aggregation-prone zones (53-62), we see that the presence of metabolites increases the RMSF of these residues. Metabolites such as adenosine, cytosine, uracil and guanine are more impactful in making these residues more solvent prone. The results indicate that interaction of these metabolites in the aggregation-prone zone and increasing their solubility can prevent these sites from aggregation (Fig 10A-D).

3.3.9.2 SASA, DSSP and hydrogen bond analyses*:* We used the g_sasa tool and dssp module to compute the general area of all the systems and the secondary structural changes, respectively. The solvent-accessible surface area (SASA) analyses of the apoprotein and the holoprotein complexes showed that there was not any significant difference between the SASA of the apoprotein and the protein-ligand complexes. This also shows that the metabolites bind and interact with the protein in a stable manner, which could prevent the protein from intermolecular interactions under stressful conditions. The potential interactions with the amino acids in aggregation-prone zones could prevent the protein from taking aggregation-mediated pathways. In the supplementary section, we report the computed SASA of the systems in Table 3.3.

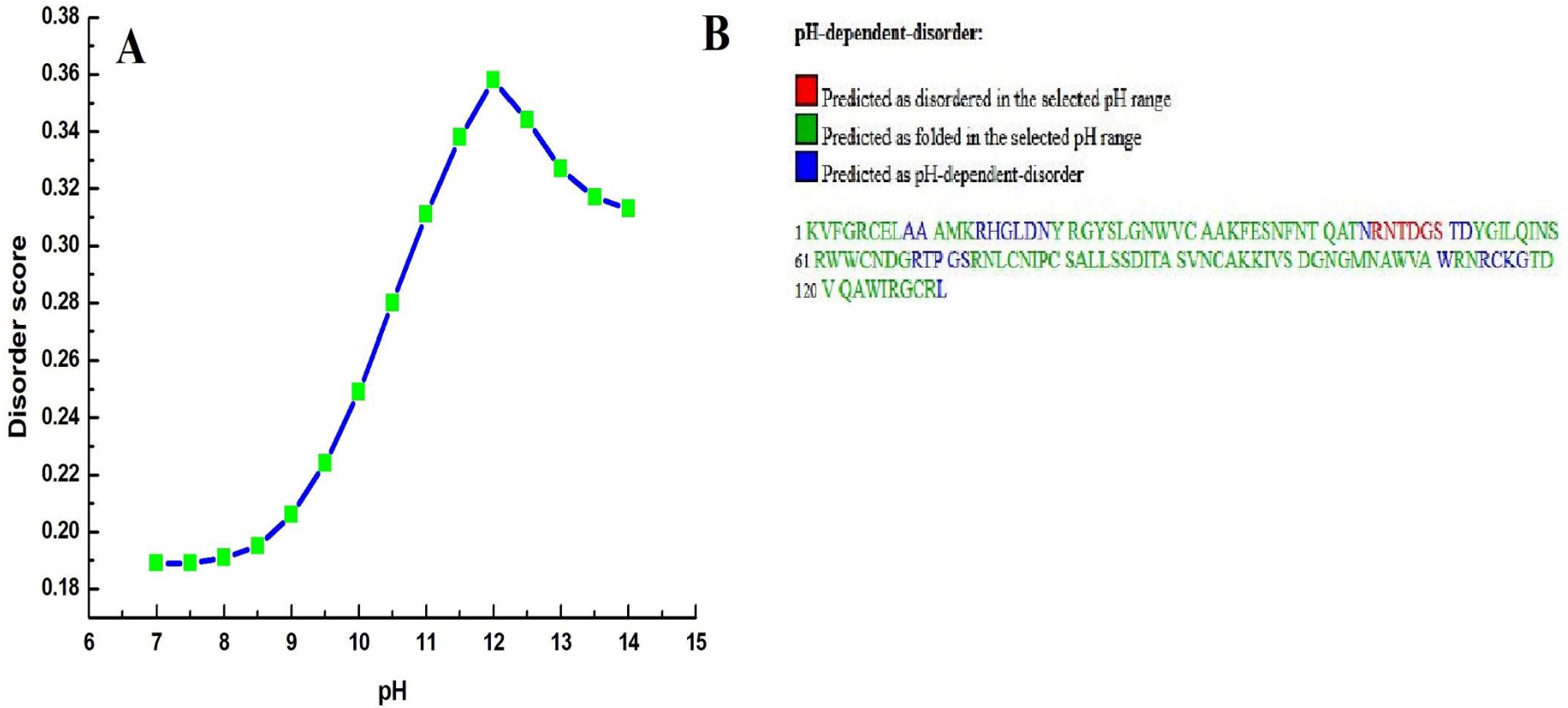

Fig 3.9 Disorder prediction using DispHscan Web server: (A) Disorder prediction score of HEWL protein in a pH-dependent manner ranging from pH 7 to pH 14 and **(B)** Disordered sequences in the HEWL protein in the selected pH range.

Detailed analysis of the secondary structure transitions during simulation across the systems yielded encouraging results. We represented the secondary structure transitions in Table 3.4. We report no change in overall β sheet content between HEWL and the HEWL-metabolite complexes. A marked decrease in β turns with a concomitant increase in alpha-helical content in the overall structure indicates a possible transition of turns to alpha-helices, highly prevalent in globular proteins. The possible transition from turns to helices, which occur majorly on the protein's surface, shows that the protein tends to be packed with more ordered secondary structures in the presence of the metabolites having more stability. The metabolites can deviate the propensity of the protein to aggregate by this transition as alpha helices are majorly stable secondary structures which, even if they tend to unfold, would not be more prone to form aggregates without many changes in β structures content.

Adenosine and cytosine form, on average, two hydrogen bonds during the entire simulation, whereas uracil forms approximately one hydrogen bond. Thymine does not form any stable hydrogen bond interactions with the protein. All the hydrogen bonds formed have a bond length of less than 0.35 nm. All the hydrogen bonds of the metabolites are mainly surrounded by aggregation-prone zones, which is an important observation of our current hypothesis.

3.3.9.3 MM/PBSA energy analyses:

We employed the MM/PBSA tool to compute the metabolites' overall binding energy and mode of interaction with hen egg-white lysozyme. We used the g_mmpbsa tool and pbsa script in the GROMACS environment to perform previously described analyses (Kundu and Dubey, 2021, Pande *et al.*, 2021, Umesh *et al.*, 2020). We performed the analyses for the last 40 ns of the trajectory, with a time step of 100 ps. The solute dielectric constant was set at 2 and solvent at 80, respectively. Table 3.5 contains detailed energy analyses. All the metabolites emphasised the importance of Van Der Waal's interaction energy and electrostatic interactions in the overall contribution of the binding free energy. These contributions also show the importance of protein folding in the presence of metabolites. Among the five metabolites, adenosine (-52.839±58.74

kJ/mol), guanine (-52.01±14.50 kJ/mol), and thymine (-52.14±24.65 kJ/mol) show similar binding energy, guanosine (-84.316 +/-50.635 kJ/mol) which also compliments our biophysical experiments at both at near physiological pH and at pH 12.2.

Table 3.3- Computed solvent-accessible surface areas of the HEWL and the HEWL-metabolites complexes

System	**SASA ($Å^2$)**
HEWL	72.86
HEWL+ Adenosine	74.29
HEWL+ Cytosine	74.61
HEWL+ Guanine	73.24
HEWL+ Guanosine	75.61
HEWL+ Thymine	73.06
HEWL+ Uracil	72.84

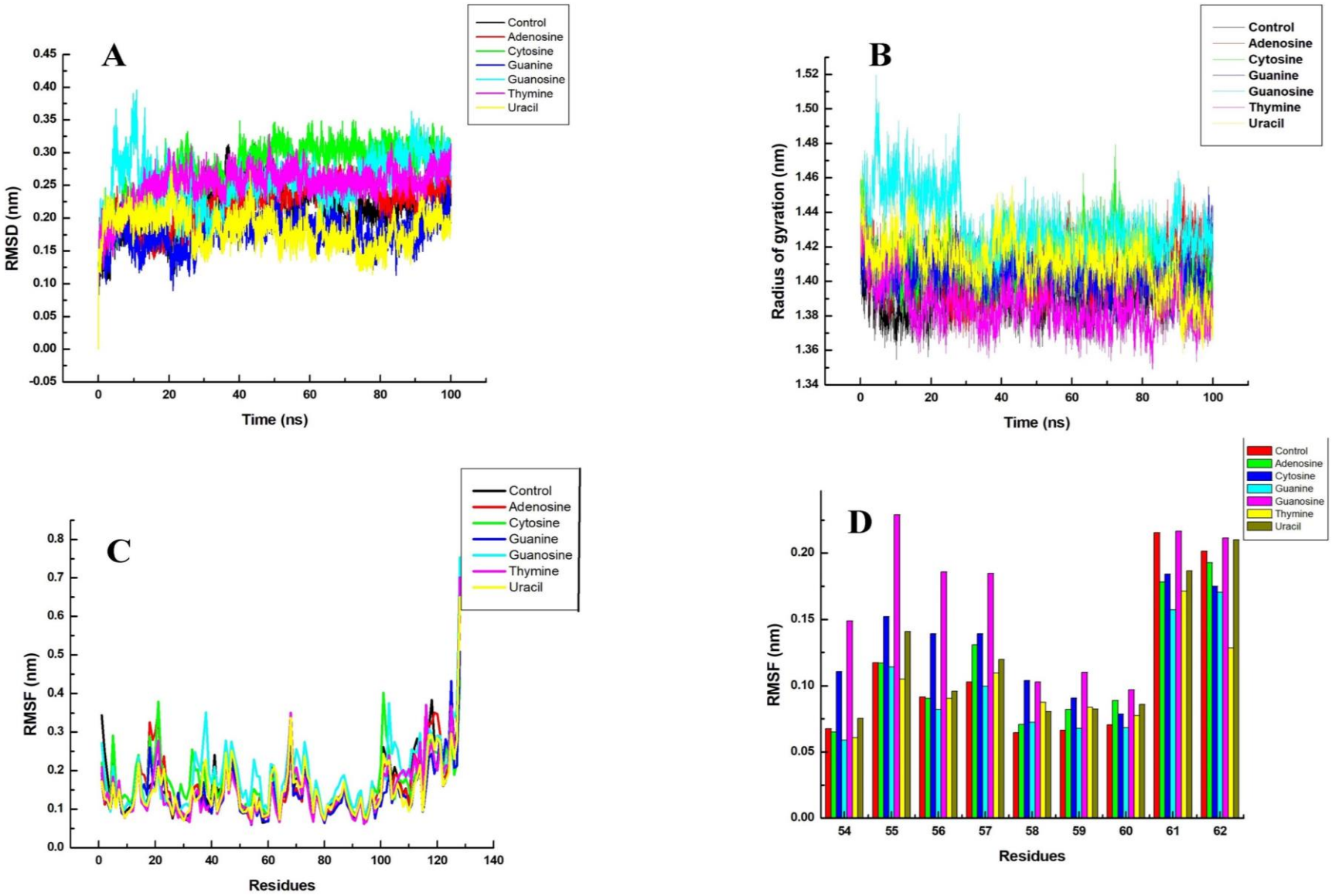

Fig 3.10 Molecular Dynamics and simulation analyses of HEWL and HEWL-metabolites complexes for 100 ns run: **(A)** RMSD **(B)** Radius of gyration **(C)** RMSF, and **(D)** RMSF of the residues in the most aggregation-prone zone of HEWL

Table 3.4- Details of Secondary structure of apoprotein and holoprotein complexes computed using DSSP tool

Protein Model	Coiled-Coil	β-sheet	Bridge	Bend	Turn	Alpha-helix
HEWL	18%	6%	3%	13%	25%	34%
HEWL+Ade	17%	6%	3%	13%	21%	39%
HEWL+Cyt	22%	6%	4%	14%	18%	35%
HEWL+Gua	16%	6%	4%	14%	21%	39%
HEWL+Guan	18%	7%	4%	14%	20%	34%
HEWL+Thy	19%	6%	4%	15%	21%	36%
HEWL+Ura	16%	6%	3%	15%	20%	38%

Table 3.5- MM/PBSA binding free energy analyses of the various ligands with hen egg-white lysozyme

Ligand	VDW (kJ/mol)	Electrostatic Interaction (kJ/mol)	Polar Solvation (kJ/mol)	SASA (kJ/mol)	Binding energy (kJ/mol)
Adenosine	-74.98±73.03	-4.16±5.44	33.36±32.5	-7.058±6.895	-52.839±58.74
Cytosine	-41.39±35.913	-13.82±14.306	33.779±32.65	-4.656±4.080	-26.09±25.667
Guanine	-80.69±19.68	-13.11±6.2	50.46±13.89	-8.66±1.69	-52.01±14.50
Guanosine	-113.051 +/- 72.407	-5.206 +/- 11.066	44.757 +/- 38.656	-10.816 +/- 6.890	-84.316 +/- 50.635
Thymine	-73.19±26.44	-10.07±6.81	38.16±16.79	-7.05±2.49	-52.14±24.65
Uracil	-4.61±10.82	-1.29±5.10	6.59±46.10	-0.91±2.35	-0.233± 43.9

3.4. Discussion

Amyloid fibrils are stable, insoluble, and notorious protein aggregates formed by various proteins under specific conditions (Chiti *et al*., 1999); it is essential (Sarkar and Dubey, 2011). Further, amyloids show extreme stability and resistance to proteases (Sarkar and Dubey, 2011, Chatani and Goto, 2005). It is essential to understand various structural phenomena a protein undergoes for better insights into protein folding, misfolding and aggregation (Kundu and Dubey, 2021). Since Hen egg-white lysozyme is highly homologous to a human counterpart, it is readily available, economical, and highly stable; it has become an automatic choice for many studies with

the aim of investigations into protein aggregation and misfolding studies. Previous literature suggests various small molecules and other derived compounds as potential anti-aggregation agents, viz natural polyphenolic compounds like curcumin and kaempferol (Borana *et al.*, 2014), oxidised ECGC (An *et al.*, 2017), also quinone and quinone derivatives (Feng *et al.*, 2012). We have utilised the underused selected endogenous molecules, nitrogenous bases and nucleosides in the current study to investigate their anti-aggregation properties.

We have earlier shown that these metabolites can bind to the amyloidogenic region of HEWL and thereby possibly inhibiting intermolecular interaction and amyloid formation (Kundu *et al.*, 2020). The current work has experimentally validated the finding using various experimental techniques. Using critical factors such as pH, protein concentration and agitation, we have generated soluble and insoluble protein aggregates using HEWL solution. We examined our hypothesis using experimental and further corroborated molecular dynamics and simulations. We also report that HEWL aggregation has a higher growth rate at near physiological pH than pH 12.2, irrespective of the protein concentration. The observation indicates that near physiological pH tends to produce more matured fibrils than pH 12.2, as indicated by our ThT intensities and kinetics than pH 12.2. Secondly, we also show that the overall formation of protein aggregates at pH 12.2 is much slower than near physiological pH. The lag phase, primary nucleation rate and growth constant of amyloid fibrils depend on the solution's pH and are more pronounced in near physiological pH. An increase in ThT intensity with increased protein concentration indicates that ThT binds more readily with matured fibrils than oligomers. Also, since the rate of fibrillar growth at pH 12.2 is slower than at near physiological pH, ThT binding intensities will be higher at near physiological pH than at pH 12.2. Although HEWL is more disordered at pH around its isoelectric point, HEWL tends to form insoluble visible aggregates and less soluble aggregates, which further impairs ThT and ANS binding and subsequent fluorescence intensities, as shown in our data.

The energy difference between the native state of the protein and the unfolded state is approximately 5-20 kcal/mole. In contrast, the energy barrier between the unfolded and the amyloid

fibrillar state is much higher. This is why amyloid fibrillation is theoretically an irreversible process. We compared the effect of the metabolites at two different time points, 72 hours and 120 hours incubation. We noted that simultaneous incubation of the protein in the presence of metabolites modulates HEWL oligomerisation with time. Once a protein starts misfolding, the protein readily forms amorphous aggregates followed by oligomers and the more stable amyloid fibrils. After 72 hours of incubation of HEWL with metabolites at pH 7.4, most of the metabolites drive the HEWL aggregation towards an amorphous aggregates state. In adenosine, the HEWL attains an intermediate state between oligomeric and amyloid fibrils, whereas HEWL aggregates tend to be towards an amyloid fibrillar in the presence of cytosine state. After 120 hours of incubation, all the metabolites tend to drive the HEWL aggregation away from amyloid fibrillation towards a more pronounced amorphous aggregate state, as indicated by our ThT and ANS assays. A decrease in ThT intensity in the HEWL solutions in the presence of the metabolites compared to the control indicates a lowering of β-sheet content in these oligomeric solutions.

Along with a decrease in ThT intensity, the presence of the metabolites also reduces ANS intensity, indicating a decrease in exposed hydrophobic patches on the surface of the protein. The possible mechanism of the metabolites includes a reduction in β sheets coupled with a reduction in exposed hydrophobic patches on the protein's surface. Some studies reported a similar mechanism (Sarkar and Dubey, 2011, Konar *et al*., 2017). The DLS results indicate that the metabolites impact the size of the aggregates more at near physiological pH than at pH 12.2. The solution at near physiological pH consists primarily of heterogeneous species with variable sizes well below 1000 nm. At pH 12.2, the aggregates are of much homogenous nature. The metabolites do not impact the overall size of the aggregates forming larger than near physiological pH. Our AFM results show that at the day 5 stage after incubation in pH 12.2, the samples treated with metabolites do not have much-reduced height. In near physiological pH, there is a consistent decrease in the height of the samples in the presence of various aggregates. The decrease in the height of the samples also supports our previous results, implying the impact of the metabolites on preventing the formation of

matured fibrils. Generally, the average length of matured fibrils varies from 100nm-1000nm. Hence, we have also shown that HEWL forms matured fibrils at both pHs with a starting monomer concentration used in this study.

Our molecular dynamics and simulation results encourage solvent-accessible surface area and secondary structure analyses. SASA analyses showed that HEWL SASA does not have much difference from the ones in which metabolites are interacting. The slight tendency of the protein to increase SASA in the presence of the majority of metabolites indicates a possible tendency for the protein to have minor changes in compactness in the presence of the metabolites. Uracil did not affect the SASA of the HEWL. Delving into the secondary structure analyses, we observe no impact on the overall β sheet content, although the HEWL-metabolite complexes showed a reduction in β turns and an increase in α-helices. The role of turns in protein folding is well explained in the published literature (Marcelino and Geirasch, 2008). β turns are essential in protein folding as they can initiate the formation of more ordered secondary structures without much loss in chain entropy. In our simulation studies, we see a transition of β turns to alpha-helix, both of which occur on the surface of the proteins and are composed of hydrophilic amino acids. The transition into alpha-helical content could explain the slight increase in overall SASA. The tendency to increase SASA slightly in the presence of metabolites is encouraging because proteins with rich alpha-helical content and limited β she*et al*so imply that proteins will probably not tend to aggregate when they unfold, which highlights this central theme manuscript.

Hydrogen bonding between metabolites and HEWL with critical amino acid residues during simulation is also important in gauging the binding capabilities of the metabolites with the protein. In HEWL, Trp62, Trp63 and Trp108 are essential residues of the hydrophobic cleft and binding site for metabolites, ultimately governing the folding-unfolding mechanism of the native protein (Ghosh *et al*., 2008). We analysed the hydrogen bonding patterns of our metabolites, most of which formed stable interactions with Trp62, Trp63 and Trp108. They also interacted with other residues of the amyloid-prone zone in HEWL (Formoso and Foster, 1975). Since compounds that can stabilise

these specific amino acids during the unfolding process can inhibit the progress of fibrillation stages, we show that molecules like adenosine, cytosine, and uracil can also have a similar impact on HEWL-associated aggregation.

3.5 Concluding remarks

The current study marks the importance of using complementary spectroscopic analysis better to understand amyloidosis as a matter of physiological importance. Various methods like UV-Vis, Fluorescence and CD spectroscopy can be used to monitor changes in the fibrillation processes of amyloids grown *in vitro*. Atomic Force Microscopy remains one of the most sought-out techniques to gain a direct insight into the structural morphology of the aggregates or fibrils formed. Experimentally studying HEWL amyloidosis *in vitro* is a well-established method to study drug development for any amyloid-associated diseased condition because it is easily reproducible under laboratory conditions, and the oligomeric states show cytotoxic effects.

The study focuses on the impact of selected intracellular metabolites on the primary nucleation step of protein aggregation. Fibril-dependent processes are absent in reactions starting with protein monomers under aggregating conditions. Since fibril-dependent processes are absent, primary nucleation is the driving force for protein aggregation under these conditions. Our results show that primary nucleation processes are suppressed in the presence of metabolites as there is a simultaneous loss of ThT and ANS intensity. Interaction of small molecules directly with aggregation-prone disordered zones within the full-length protein is also an exciting strategy for finding small molecules as protein aggregation modifiers. We have employed combined computational and biophysical approaches in the current study.

This study involving intracellular molecules with extensive biophysical investigations has not been reported before, further opening avenues for research into this field. We feel that this work raises more questions than it answers. This study has pointed out areas where investigations can be directed involving protein folding and intermediate states in the pathway of amyloidosis. This study exhibits various metabolites which can be potential agents and explored further with many other

molecules. Greater understanding is also required precisely as the amyloid formation pathway is highly dependent on many factors, especially pH and protein concentration and other environmental factors such as temperature and agitation, not only on amino acid composition. This is evident from this study as some unexpected results need further study. Lastly, the processes undertaken in this study are still highly appreciable as a standard procedure for studying amyloid formation by many proteins that are of more physiological importance for direct drug development processes.