Statistical Analysis of Thermodynamic Quantities in the Binding of Ligands to Bovine and Human Serum Albumin

A Thesis submitted
to the Graduate School Valdosta State University

in partial fulfillment of requirements

for the degree of

MASTER OF SCIENCE

in Biology

in the Department of Biology of the College of Science and Mathematics

July 2022

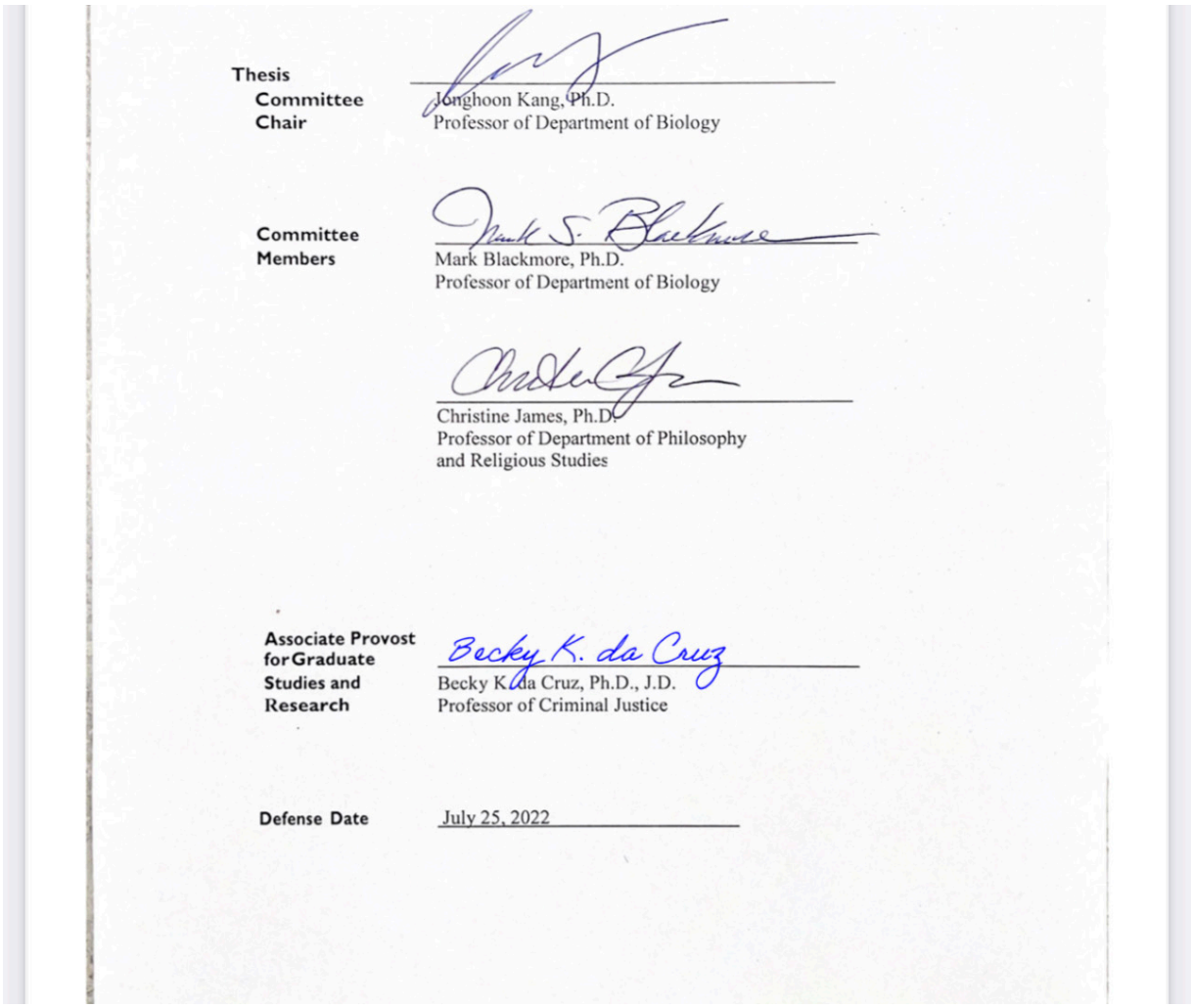
Willie Darrell Hill Jr.

B.S., Valdosta State University, 2015

© Copyright 2022 Willie Darrell Hill Jr.

All Rights Reserved

This thesis, “Statistical Analysis of Thermodynamic Quantities in the Binding of Ligands to Bovine and Human Serum Albumin” by Willie Darrell Hill Jr. is approved by:



Thesis Committee Chair

Jonghoon Kang, Ph.D.
Professor of Department of Biology

Committee Members

Mark Blackmore, Ph.D.
Professor of Department of Biology

Christine James, Ph.D.
Professor of Department of Philosophy and Religious Studies

Associate Provost for Graduate Studies and Research

Becky K. da Cruz, Ph.D., J.D.
Professor of Criminal Justice

Defense Date July 25, 2022

FAIR USE

This Thesis is protected by the Copyright Laws of the United States (Public Law 94-553, revised in 1976). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgement. Use of the material for financial gain without the author's expressed written permission is not allowed.

DUPLICATION

I authorize the Head of Interlibrary Loan or the Head of Archives at the Odum Library at Valdosta State University to arrange for duplication of this Thesis for educational or scholarly purposes when so requested by a library user. The duplication shall be at the user's expense.

Signature____*Willie Darrell Hill Jr.* _____

I refuse permission for this Thesis to be duplicated in whole or in part.

Signature

_____

ABSTRACT

Albumin is one of the most studied proteins because of its diverse functions (Peters, 1995). Albumin varies between species, but past research on Human Serum Albumin (HSA) and Bovine Serum Albumin (BSA) has shown structural similarities between the two. Binding properties of HSA and BSA can be compared using the thermodynamic parameters of enthalpy, entropy, and Gibbs free energy (ΔH, ΔS, and ΔG). Previous studies have shown a linear relationship between enthalpy and entropy when temperature is held constant (enthalpy-entropy compensation). Prior data used thermodynamics to compare the two proteins on a micro-level (1-10 ligands). This thesis is the first to statistically analyze thermodynamic quantities in the binding of ligands to HSA and BSA on a macro-level (200+ ligands). A bioinformatic approach was used to obtain thermodynamic data from 197 primary literature sources available in PubMed (Sayer, et al., 2022). Linear regression was applied and showed a significant positive correlation between enthalpy and entropy in both proteins examined. The findings indicate the existence of a significant enthalpy-entropy compensation. This thesis quantitatively described the variation of ΔH°, ΔS°, ΔG° in the binding of ligands to BSA and HSA by using Cumulative Distribution Function (CDF). The obtained CDFs were used to derived Normalized Probability Density Function (NPDF) for each thermodynamic quality. The Shannon entropy was calculated to assess the variability on ΔH°, ΔS°, ΔG°. The resulting values of Shannon entropy shows that both ΔH° and TΔS° have larger values than that of ΔG° for both proteins. ΔG° contained lower levels of information (uncertainty) than ΔH° and TΔS° for both proteins.

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank the graduate school for giving me an opportunity pursue my master's in biology. It has been a very tough journey for me. I started this program in the Spring of 2016 studying mosquitoes with Dr. Blackmore. Due to my own negligence, I lost all of my research data. After nearly a two-year hiatus, I returned to start new research with Dr. Kang in the Fall of 2020. I appreciate my past and present faculty advisors for always working with me to complete my goal. I also appreciate Dr. James for taking the time to be a part of my committee.

I will also like to thank Dr. Myung Soog Lee for her efforts in verifying my data collection.

I will also like to thank my mother, Carolyn Chambliss Hill for always encouraging me to finish my master's degree in biology. I love you for all of the sacrifices that you made for me.

Lastly, I would like to thank myself, Willie Darrell Hill Jr. for never giving up and finishing what I started.

I acknowledge Dr. Jonghoon Kang, Dr. Mark Blackmore, Dr. Christine James, and Dr. Myung Soog Lee. I acknowledge my mother, Carolyn Chambliss Hill, and myself, Willie Darrell Hill Jr.!

Chapter I

INTRODUCTION

Albumin Family

The albumin family in animals consists of serum albumin, a-fetoprotein, vitamin D binding protein and afamin (Haefliger, et al., 1989). Attention is often directed towards serum albumin because it is the most abundant blood plasma protein.

Albumin is widely distributed in vertebrate species with a few exceptions. Some plants contain albumins, but they were not phylogenetically related to animal albumin. Flatworms and roundworms contained no albumin. Insects don't show any signs albumin except in Culex quinquefasciatus (southern house mosquito). The most mosquito species require a blood meal from mammals or birds to develop their eggs, which could explain the presence of albumin. Albumin has a large presence in mammals (Li, et al., 2017).

ECM1

In a study by led by Shugang Li, a new addition to the albumin family was discovered named extracellular matrix protein (ECM1) was discovered. The scientists stated that ECM1 has a close phylogenetic relationship with serum albumin. EMC1 is involved in endochondral bone formation and angiogenesis (development of new blood vessel). ECM1 contributed to skin integrity, homeostasis, and wound healing (Li, et al., 2017).

Serum albumin is the most abundant plasma protein in mammals. It is a globular protein consisting of a single polypeptide chain with about 580 residues. Globular proteins are recognized by their spherical shape and water solubility. Albumin is subject to heat denaturation and is moderately soluble in salt solutions (Puoci, 2015).

Biological Function of Serum Albumin

Serum albumin is produced in the liver and has the main function of controlling the osmotic pressure in the blood. HSA has several other functions in body: regulate pH levels in blood, lipid metabolism, release toxins, and resist oxidative stress (Peters, 1985).

Serum albumin acts as a vehicle to transport biological important ligands (Tayyab, & Feroz, 2021). There are other transport proteins, but serum albumin is the only one that can bind to a wide diversity of ligands (Kragh-Hansen, 1990). The binding of ligand is extremely important for transporting nutrients in the blood.

Human / Bovine: (The two most common)

Human serum albumin (HSA) occurs in humans and bovine serum albumin (BSA) is found in cattle (*Bos taurus*). HSA and BSA are structurally very similar but also contain some differences. They bind to ligands and transport them throughout the body. Serum albumin acts as a vehicle to transport biologically important ligands (Peters, 1970). Based on past observations, it was determined that the two proteins share similar properties (Kragh-Hansen, 1990). BSA and HSA are very similar in structure and environmental binding properties, but BSA has a more rigid structure than HSA (Ketrat, et al., 2020), and HSA is generally more stable than BSA.

Delta H (ΔH) and Delta (ΔS)

Delta H represents Enthalpy. Enthalpy is the sum of the internal energy and the product of the pressure and volume of a thermodynamic system. Delta S is represented in entropy. Entropy is a measure of how much energy is not available to do work (Dragan, et al., 2017). Using Thermodynamics to measure protein interaction

Thermodynamics is the branch of physics that focuses on the relationships between heat and other forms of energy. Thermodynamics describes how thermal energy is converted to and from other forms of energy and how it affects matter. Binding between two interacting partners has both enthalpic (ΔH) and entropic (ΔS) components, which means the recognition event is associated with changes of both the structure and dynamics of each counterpart (Chodera & Mobley, 2013). As pointed out in (Kang & Kang, 2021), it is important to include both ΔH and ΔS when analyzing thermodynamic compensation. The thermodynamic parameters of enthalpy, entropy, and Gibbs free energy (ΔH°, ΔS°, ΔG°) are extremely important in analyzing the binding properties of serum albumins.

Gibbs Free Energy (ΔG) also plays a major role in the binding. Thermodynamic compensation has been used to discover more information about proteins. Past studies from Valdosta State University used thermodynamics compensation to provide key facts about the protein keratin (Cowen & Kang, 2019).

Thermodynamic Parameters in framing EEC

The relationship between the four thermodynamic components are expressed in the following equation: $\Delta G = \Delta H + T\Delta S$. There are different methods to determine each component.

Thermodynamics consist of four major components:

1. ΔG – Gibbs Free Energy (the amount of energy left over after a chemical reaction has taken place)
2. ΔH – Enthalpy (sum of the internal energy and the product of the pressure and volume of a thermodynamic system)
3. ΔS – Entropy (a measure of how much energy is not available to do work)
4. T – Temperature (constant)

Enthalpy-entropy compensation

For over a century, studies have showed a linear compensation between enthalpy and entropy when there was a constant temperature. In most cases enthalpy increases with entropy (Moulik, et al., 2019). The enthalpy and entropy interactions produce a positive slope. Scientists described this phenomenon as "Enthalpy-Entropy Compensation" (EEC) Liu & Guo, 2001). Studies have also shown that the enthalpy-entropy slope would be different when comparing constant temperature and experimental temperature (Moulik, et al., 2019). In most cases enthalpy increases with entropy. The enthalpy and entropy interactions produce a positive slope. Scientists described this phenomenon as "Enthalpy-Entropy Compensation" (EEC).

Serum Albumin and Medicine

The study of serum albumin has been very useful in medicine. During the World War Two, medics used purified human serum albumin to aid wounded soldiers (Peters, 1995). In the mid-1980s, serum albumin concentrations were used to prognosis myeloma (Whicher & Spence, 1987). The low concentrations of serum albumin indicated chances of the patient developing myeloma. The concentration of serum albumin plays a major role in an individual's quality of life. (Kobayashi, et al., 1991). Scientists are now using serum albumin data to research COVID-19. The researchers hypothesized that patients with low serum albumin were at higher risk of hospitalization within three weeks of discharge from the Emergency Room (Acharya, et al., 2021). Studies on serum albumin can provide was better ways of designing drugs and providing medical treatment (Maier, 2021).

Chapter II

MATERIALS AND METHODS

The database of PubMed itself is composed of biomedical publications archived in MEDLINE containing fields of interest like those centered around immunology in the biomedical society (Sayer, et al., 2022).

Data collection

Thermodynamic data, $\Delta H°$ and $\Delta S°$, in the binding of ligands to BSA or HSA were obtained from primary literature available in PubMed. The units of $\Delta H°$ and $\Delta S°$ are variable among the papers. Therefore, we converted them to the units of the SI system, namely kJ/mol and kJ/mol/K (Kelvin) for $\Delta H°$ and $\Delta S°$, respectively.

Linear regression analysis of $\Delta H°$ and $\Delta S°$

Numerical relationship between $\Delta H°$ and $\Delta S°$ was obtained by fitting Eq. (1) to the data:

$$\Delta H° = T_C \Delta S° + \beta \quad (1)$$

where T_C is the compensation temperature in Kelvin (Griessen & Dam, 2021) and β is the y-intercept, which is $\Delta H°$ when $\Delta S° = 0$.

Cumulative distributions of $\Delta H°$, $T\Delta S°$, and $\Delta G°$

Thermodynamic quantities were arranged in an increasing order, and the rank of each quantity was scaled to 1 to 100 using Eq. (2):

$$Rank = \frac{99x}{N_{Total} - 1} + \frac{N_{Total} - 100}{N_{Total} - 1} \quad (2)$$

where x is the order of the quantity and N_{Total} is the total number of data, which is 185 for BSA and 97 for HSA.

Cumulative distribution function (*CDF*) of each thermodynamic quantity was obtained by fitting a sigmoid equation to the data of the rank of each thermodynamic quantity obtained using Eq. (2). For $\Delta H°$ and $T\Delta S°$, Eq. (3) was used in fitting for $\Delta H°$ and $T\Delta S°$, and Eq. (4) was used for $\Delta G°$:

$$Rank(x) = \frac{a}{1 + \exp[\frac{x_0 - x}{b}]} + \frac{c}{1 + \exp[\frac{x_1 - x}{d}]} \quad (3)$$

$$Rank(x) = \frac{a}{1 + \exp[\frac{x_0 - x}{b}]} \quad (4)$$

where a, b, c, d, x_0, and x_1 are fitting parameters.

Normalized Probability distribution function of $\Delta H°$, $T\Delta S°$, and $\Delta G°$

Probability density function (*PDF*) was obtained from the *CDF* using Eq. (5):

$$PDF(x) = \frac{d}{dx} Rank(x) \quad (5)$$

Normalized PDF (*NPDF*) was obtained using Eq. (6):

$$NPDF(x) = \frac{1}{N} PDF(x) \quad (6)$$

where N is the normalizing constant, which was calculated using Eq. (7):

$$N = \int_{-\infty}^{\infty} PDF(x)\, dx \quad (7)$$

Venn diagrams for $\Delta H°$, $T\Delta S°$, and $\Delta G°$ of BSA and HSA

Similarity of *NPDF* of each thermodynamic quantity for BSA and HSA was calculated using Eq. (8):

$$\eta(BSA, HSA) = 1 - \frac{1}{2} \int_{-\infty}^{\infty} |BSA(x) - HSA(x)|\, dx \quad (8)$$

where $\eta(BSA, HSA)$ is the overlapping index between two *NPDF*s of BSA and HSA, x is the corresponding thermodynamic quantity, and $|.|$ represents the absolute value

operator (Pastore & Calcagnì, 2019). $\eta(BSA, HSA)$ represents the intersection of NDPFs of BSA and HSA.

Shannon entropy for NDPF of $\Delta H°$, $T\Delta S°$, and $\Delta G°$

Shannon entropy (H_{Shannon}) for the probability distribution of each thermodynamic quantity was calculated using Eq. (9) (Ben-Naim, 2012; Pierce, 1980; Stone, 2015):

$$H_{Shannon}(NDPF) = -\int_{-\infty}^{\infty} p_x \times \log_2 p_x \quad (9)$$

where p_x is the *NDPF* of $\Delta H°$, $T\Delta S°$, or $\Delta G°$.

Mathematical software

Numerical and statistical analyses were conducted using Mathematica (version 12.1, ©Wolfram Research, Inc., Champaign, IL) and © SigmaPlot (version 11, Systat Software Inc., San Jose, CA).

Chapter III

RESULTS

Enthalpy–entropy compensation

We obtained 185 and 97 ligand binding thermodynamic data sets ($\Delta H°$ and $\Delta S°$) for BSA or HSA (Fig. 1A). The plot shows a wide variation of $\Delta H°$ and $\Delta S°$. The relationship between $\Delta H°$ and $\Delta S°$ in the binding of ligands to BSA or HSA was examined using linear regression with Eq. (1). Linear regression indicated that a significant correlation between $\Delta H°$ and $\Delta S°$ existed in both proteins as the R^2 values are 0.9823 for BSA and 0.9445 for HSA (Table 1). Furthermore, no systematic relation between the standard residuals for $\Delta H°$ and $\Delta S°$ was detected, validating the linear regression (Fig. 1B). Standard deviation of each thermodynamic quantity was calculated, and while both ΔH° and ΔS° have large values of standard deviation, about seven times larger than that of ΔG° (Fig. 1C).

The correlation between $\Delta H°$ and $\Delta S°$ observed in Fig. 1A is known as enthalpy–entropy compensation (Fox et al., 2018). The values of T_C in Eq. (1) and its standard errors for BSA and HSA are 293.57 ± 2.91 K and 316.32 ± 7.87 K, respectively. A *t*-test indicated that the difference of T_C between BSA and HSA was statistically significant (t = -3.269, df = 278, P = 0.001). In the *t*-test, the value of degree of freedom (*df*) (Fowler, et al., 2008) was calculated as df = (n1 – 2) + (n2 – 2), where n1 and n2 are the number of data points of BSA and HSA, respectively: n1 = 185 and n2 = 97. It is known that T_C can be a quantitative measure of the degree of compensation between $\Delta H°$ and $\Delta S°$. The statistical difference of T_C between BSA and HSA strongly suggests that they have

distinct ligand binding properties despite a high degree of homology in their protein sequences (Supplementary).

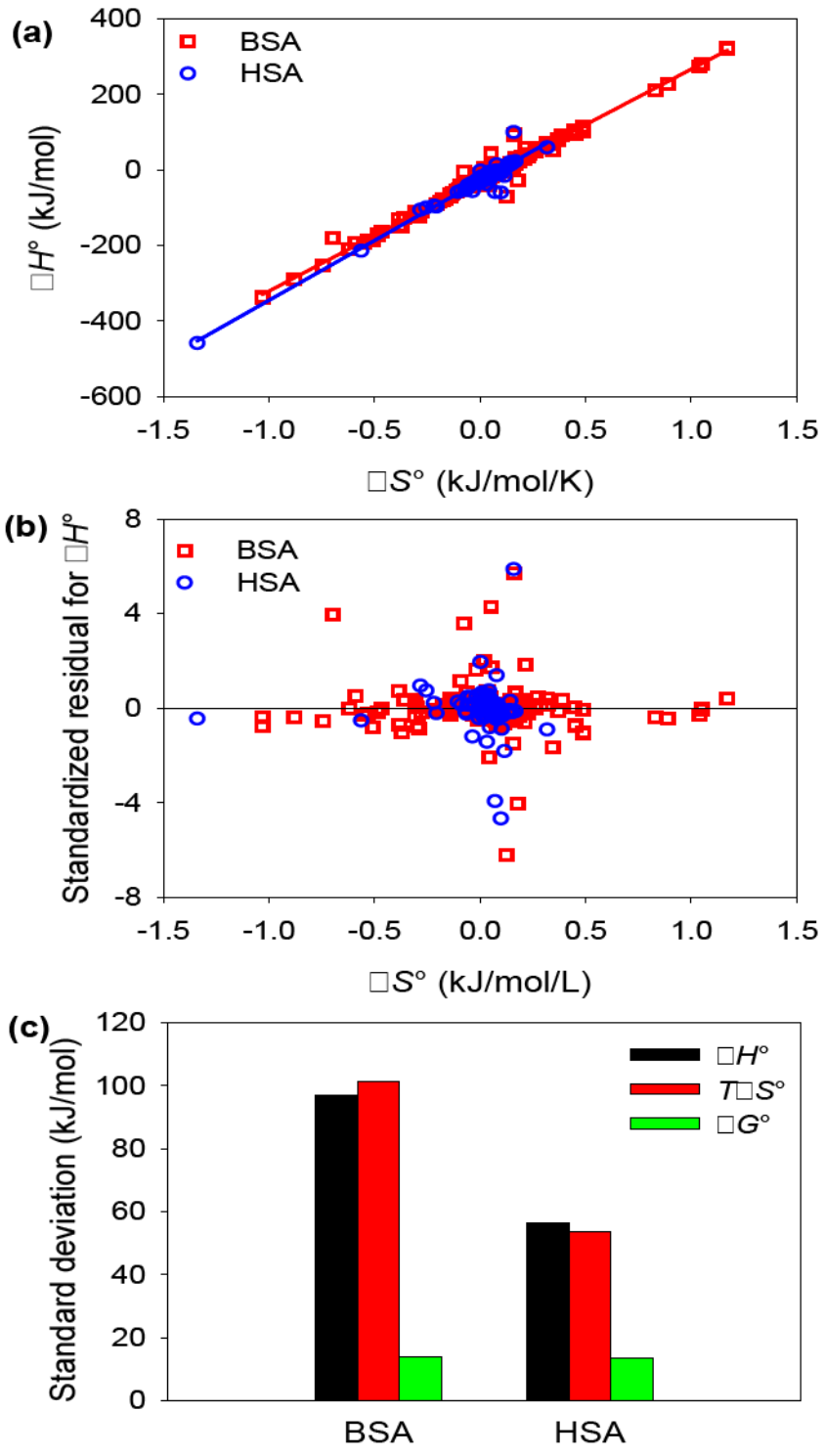


Figure 1. Ligand binding thermodynamic data sets (ΔH° and ΔS°) for BSA or HSA (Fig. 1A and 1B). Standard deviation of each thermodynamic quantity for BSA and HSA (Fig. 1C). The database of PubMed was accessed between January, 2021to December 2021 All graphs were performed on SigmaPlot (version 11, Systat Software Inc, San Jose, CA, USA).

Table 1 Fitting parameters and standard errors (SE) in linear regression between $\Delta H°$ and $\Delta S°$

Protein	$T_C \pm$ SE	$\beta \pm$ SE	R^2
BSA	293.6 ± 2.9	-27.5 ± 0.9	0.9823
HSA	316.3 ± 7.9	-29.1 ± 1.4	0.9445

Cumulative distribution of thermodynamic quantities

In order to quantitatively describe the variations of $\Delta H°$ and $\Delta S°$ in the binding of ligands to BSA or HSA, we performed statistical analysis aiming to get the *NPDF* of each thermodynamic quantity. First, we tested the normality of $\Delta H°$, $T\Delta S°$, and $\Delta G°$ for BSA and HSA, and found none of them were normally distributed (Table 2), indicating normal distribution is not appropriate for the description of the data. In order to find appropriate distribution functions for the thermodynamic quantities, we constructed a *CDF* of each quantity. In the construction of the *CDF*, the order indices of the binding ligands were rescaled to 1 to 100 in an increasing order of each thermodynamic quantity (Fig. 2). Various standard nonlinear equations were tested for the regression analysis of the $\Delta H°$, and none of them produced satisfying fits. Therefore, we applied a user-defined function, Eq. (3), and found a satisfying fit for the $\Delta H°$, as the R^2 value is larger than 0.99 in both BSA and HSA (Fig. 2A and Table 3). $T\Delta S°$, entropic energy at T = 310.15 K, showed the same features as $\Delta H°$, and the same form of equation, Eq. (3) was applied to get the best fit (Fig. 2B and Table 4). In the case of $\Delta G°$ at 310.15 K, a simpler form of sigmoid function, Eq. (4) was able to effectively describe the data (Fig. 2C and Table 5)

Normalized probability density distributions of thermodynamic quantities

The CDFs obtained for each thermodynamic quantity allowed us to derive their NPDFs following the procedure shown in the section of Materials and methods. One

critical step in obtaining an NPDF is identification of the normalizing constant. We obtained the normalizing constant for each thermodynamic parameter using Eq. (7). One should note that the normalizing constants are close to 100 as expected, but are not exactly 100 (Table 6). All NPDFs obtained are smooth functions, as so are their CDFs (Fig 3A-C).

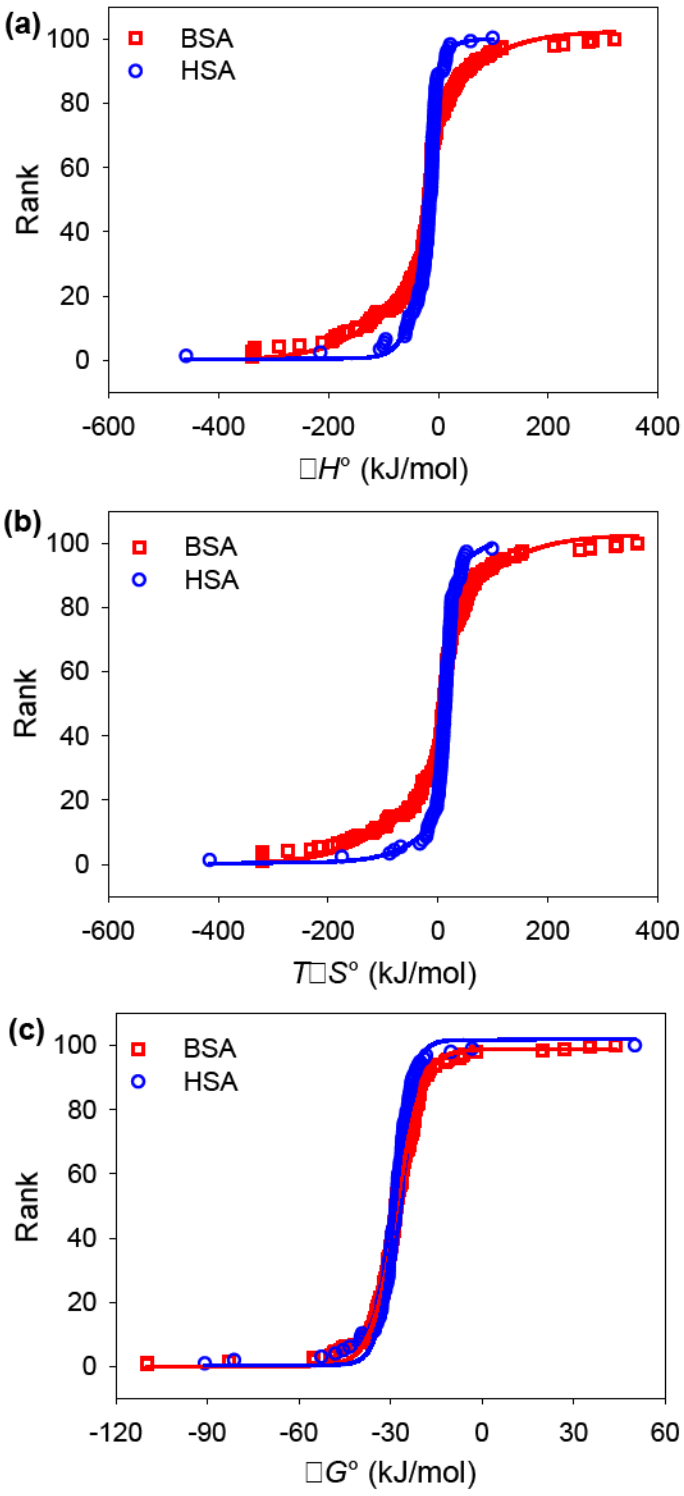


Figure 2. Cumulative distributions function of each thermodynamic quantity for BSA and HSA (Fig. 2A-C) The database of PubMed was accessed between January 2021 to December 2021. All graphs were performed on SigmaPlot (version 11, Systat Software Inc, San Jose, CA, USA).

Table 2 Shapiro-Wilk W-statistic of each thermodynamic parameter at 310.15 K

Protein	$\Delta H°$	$T\Delta S°$	$\Delta G°$
BSA	0.844	0.846	0.749
HSA	0.529	0.492	0.637

Table 3 Fitting parameters and SE for *CDF* of $\Delta H°$

Protein	$a \pm$ SE	$b \pm$ SE	$x_0 \pm$ SE	$c \pm$ SE	$d \pm$ SE	$x_1 \pm$ SE	R^2
BSA	43.7 ± 4.1	6.0 ± 0.9	-15.6 ± 0.2	58.4 ± 1.2	67.4 ± 2.2	-28.9 ± 2.1	0.9972
HSA	51.8 ± 1.0	4.1 ± 0.3	-11.6 ± 0.3	47.5 ± 5.1	19.5 ± 1.2	-30.6 ± 3.6	0.9957

Table 4 Fitting parameters and SE for *CDF* of $T\Delta S°$

Protein	$a \pm$ SE	$b \pm$ SE	$x_0 \pm$ SE	$c \pm$ SE	$d \pm$ SE	$x_1 \pm$ SE	R^2
BSA	58.7 ± 1.3	70.8 ± 2.5	0.2 ± 2.3	43.6 ± 1.1	6.8 ± 0.3	12.4 ± 0.2	0.9970
HSA	28.3 ± 4.5	45.4 ± 9.1	10.3 ± 12.2	74.1 ± 2.8	6.7 ± 0.2	14.0 ± 0.2	0.9974

Table 5 Fitting parameters and SE for *CDF* of $\Delta G°$

Protein	$a \pm$ SE	$b \pm$ SE	$x_0 \pm$ SE	R^2
BSA	98.5 ± 0.4	4.15 ± 0.04	-27.94 ± 0.05	0.9969
HSA	101.2 ± 0.9	2.87 ± 0.06	-28.29 ± 0.07	0.9937

Table 6 Normalizing constants for each thermodynamic parameter at 310.15 °K

Protein	$\Delta H°$	$T\Delta S°$	$\Delta G°$
BSA	102.161	102.345	98.5053
HSA	99.3365	102.446	101.215

The NPDFs of each thermodynamic parameter are similar for BSA and HSA, but are not identical (Fig 3A-C). In order to quantitatively assess the statistical difference of each thermodynamic quantity between BSA and HSA, the NPDF of HSA was subtracted from that of BSA as shown in Eq. (8). In the results of subtraction (Fig 3D-F), the positive-valued area under the curve represents normalized probability density exclusive to BSA, and the negative-valued area above the curve represents normalized probability density specific to HSA. Those two areas must have the same value, fulfilling the basic principle of probability. In the case of ΔH°, the area exclusive to each protein is 0.31, and therefore, the size of the intersection is 0.69 (Fig 3G). The same analysis was applied to TΔS° and ΔG° (Fig 3 H-I).

Shannon entropy of ΔH°, TΔS°, and ΔG°

The standard deviation is the most widely applied measure of variability within data (Fowler, et al., 2008), and it was used in assessing the variability in ΔH°, TΔS°, and ΔG°. However, strictly speaking, the use of standard deviation requires normality of the data (Sainani, 2012), therefore, it may not be appropriate for the description of variability in ΔH°, TΔS°, and ΔG°, since they are not normally distributed (Table 2). Even though, the normality requirement is decreasing in the case of a large dataset that has a sample size greater than 80 (Sainani, 2012), it would be desirable to measure the variability of

data with an alternative method. We applied Shannon entropy to assess the variability of $\Delta H°$, $T\Delta S°$, and $\Delta G°$ using Eq. (9). The resulting values of Shannon entropy are shown in Figure 4 and Table 7, where it is clear that both $\Delta H°$ and $T\Delta S°$ have larger values than that of $\Delta G°$ for both proteins.

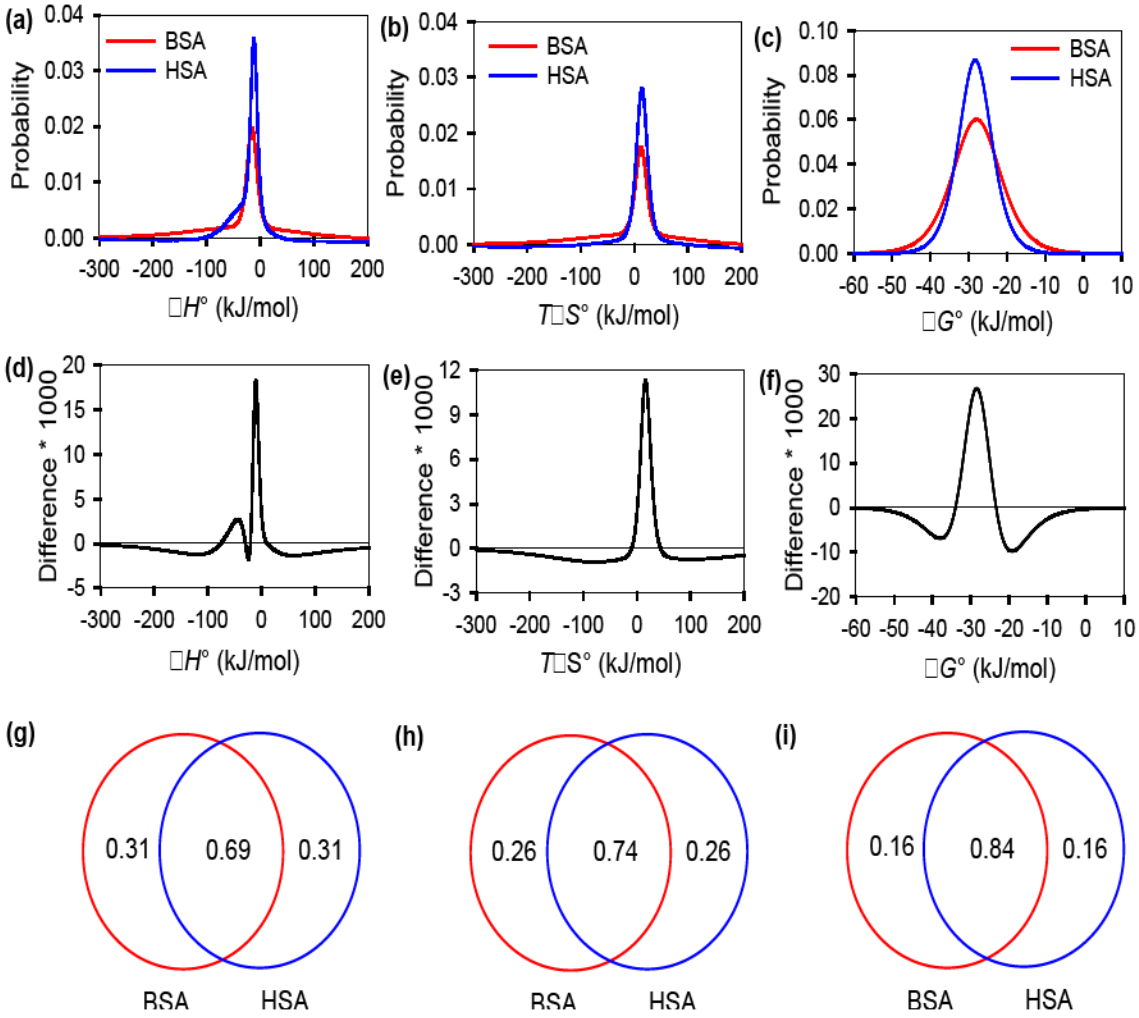
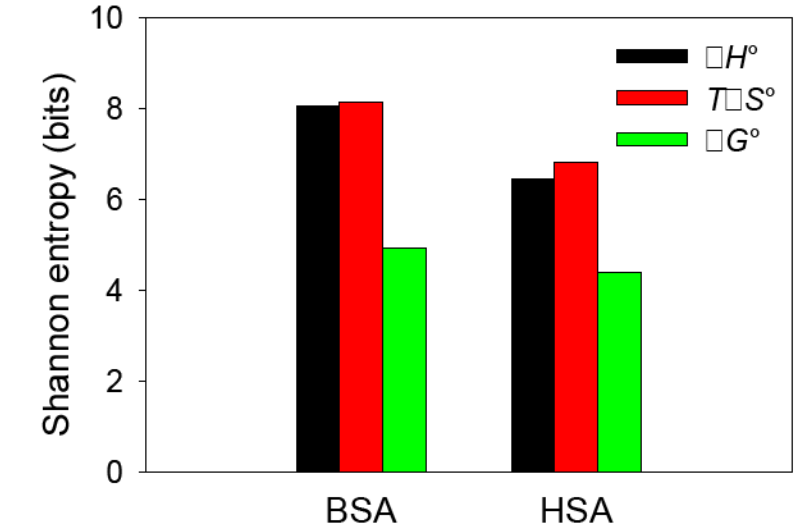


Figure 3. Normalized probability density distributions of thermodynamic quantities for BSA and HSA.(Fig. 3A-C). NPDF of HSA was subtracted from that of BSA (Fig. 3D-F). The basic principle of probability each thermodynamic quantity for BSA and HSA (Fig. 3G-I). The database of PubMed was accessed between January 2021 to December 2021. All graphs were performed on SigmaPlot (version 11, Systat Software Inc, San Jose, CA, USA).



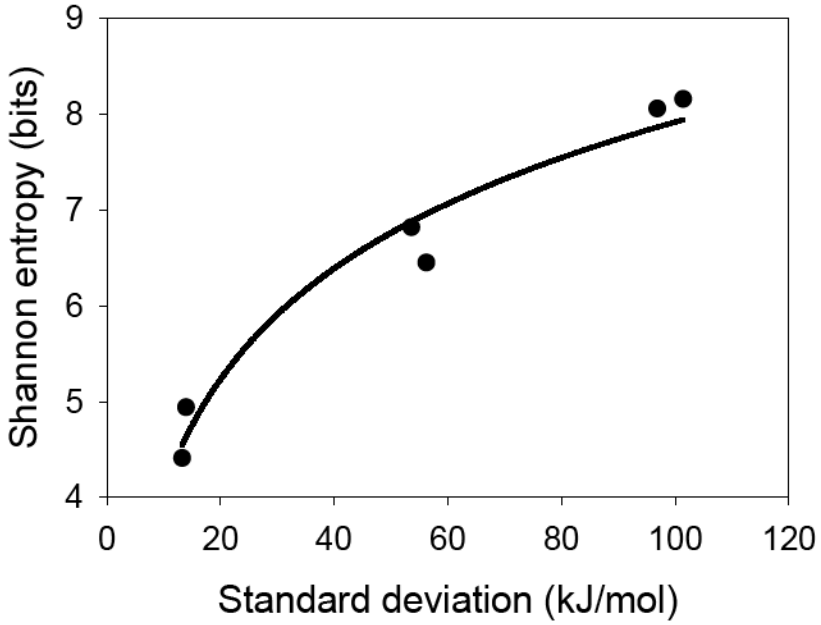


Figure 4. Shannon entropy to assess the variability of ΔH°, TΔS°, and ΔG° (Fig 4.A). The database of PubMed was accessed between January 2021 to December 2021. All graphs were performed on SigmaPlot (version 11, Systat Software Inc, San Jose, CA, USA).

Table 7 Shannon entropy (bits) of thermodynamic parameters at 310.15 K

Protein	$\Delta H°$	$T\Delta S°$	$\Delta G°$
BSA	8.06	8.16	4.94
HSA	6.45	6.82	4.41

Chapter IV:
DISCUSSION

Serum albumin has been studied for decades. This thesis provided a statistical analysis of human and bovine albumin. It consisted of analyzing 185 and 97 ligand binding thermodynamic data sets for BSA and HSA. We were able to view the finding of these interaction through PubMed. PubMed provided a large population to form a sample size. I was able to determine the exact percentage of similarities and differences of the two proteins by using thermodynamics.

I used linear regression and determined that a significant correlation between enthalpy and entropy in both proteins existed. It was determined no systematic relation between the standard residuals for enthalpy and entropy which validated the linear regression. I determined the standard deviation on all three thermodynamic components. Our finding indicated the existence of a significant enthalpy-entropy compensation.

This thesis quantitatively described the variation of ΔH° and ΔS° in the binding of ligands to BSA and HSA using Cumulative Distribution Function (CDF). The obtained CDFs allowed us to derive Normalized Probability Density Function (NPDF) for each thermodynamic quality. I discovered that BSA and HSA have similar NPDFs but they are not identical. BSA and HSA had the most similarities when comparing Gibbs free energy. BSA and HSA had the least amount of similarities when comparing Enthalpy. The rigid structure of BSA could be used to describe the difference in our statistical data when comparing it to HSA. HSA is generally more stable than BSA (Michnik, et al., 2006).

I concluded my research by calculating the Shannon entropy on ΔH°, ΔS°, ΔG°. Shannon Entropy measures the amount of information in the value of uncertainty. The entropy increases as the amount of uncertainty increases. The Shannon Entropy is a scientific expression that is used to show variability within a distribution (Lesne, 2014). It takes a mathematical approach to show how different elements of distribution are from each other. Enthalpy and entropy had higher Shannon entropy values than Gibbs free energy for both proteins. The enthalpy and entropy data provided the experiment with more information (bits) than Gibbs free energy.

Future research should analyze the thermodynamic parameter of ΔH°, ΔS°, ΔG° of BSA and HSA binding to specific types of ligands. This thesis analyzed a variety of different ligands such as vitamins, drugs, herbicides, etc. A broader analysis would provide important facts about the protein binding interaction of serum albumins to certain ligands. The statistical findings should be compared to the findings in this experiment.

This type of research is very helpful for pharmaceutical companies and medical professionals. Understanding the protein binding interaction of serum albumin to ligands can assist in the development of more effective drugs. The findings in this experiment can aid in the research of discovering more ways of using serum albumin in medicine.

REFERENCES

Acharya, R., Poudel, D., Patel, A., Schultz, E., Bourgeois, M., Paswan, R., & Sarwar, H. (2021). Low serum albumin and the risk of hospitalization in COVID-19 infection: A retrospective case-control study. *PloS One*, *16*(4), e0250906.

Ben-Naim, A. (2012). Entropy and the second law: Interpretation and mis-interpretations. *World Scientific*, Singapore

Chodera, J. D., & Mobley, D. L. (2013). Entropy-enthalpy compensation: Role and ramifications in biomolecular ligand recognition and design. *Annual Review of Biophysics*, *42*, 121.

Cowan, M. L., & Kang, J. (2019). Enthalpy–entropy compensation in the melting of human hair. *International Journal of Trichology, 11*, 141-142. DOI: 10.4103/ijt.ijt_21_19

Dragan, A. I., Read, C. M., & Crane-Robinson, C. (2017). Enthalpy-entropy compensation: the role of solvation. *European Biophysics Journal: EBJ*, *46*(4), 301–308. https://doi.org/10.1007/s00249-016-1182-6

Fowler, J., Cohen, L., & Jarvis, P., (2008). *Practical statistics for field biology*. John Wiley & Sons, Chichester, England.

Fox, J. M., Zhao, M., Fink, M. J., Kang, K., & Whitesides, G. M. (2018). The molecular origin of enthalpy/entropy compensation in biomolecular recognition. *Annual Review of Biophysics*, *47*, 223–250. https://doi.org/10.1146/annurev-biophys-070816-033743

Griessen, R., & Dam, B. (2021). Simple accurate verification of enthalpy-entropy compensation and isoequilibrium relationship. *Chemphyschem : A European*

Journal of Chemical Physics and Physical Chemistry, *22*(17), 1774–1784. https://doi.org/10.1002/cphc.202100431

Haefliger, D. N., Moskaitis, J. E., Schoenberg, D. R., & Wahli, W. (1989). Amphibian albumins as members of the albumin, alpha-fetoprotein, vitamin D-binding protein multigene family. *Journal of Molecular Evolution*, *29*(4), 344-354.

Kang, J., Kang, A.M. (2021). Statistical and thermodynamic analysis of the binding of trans-activation response-binding proteins to HIV-1 TAR RNA. *Journal of Biological Chemistry 296*:100067. https://doi.org/10.1016/j.jbc.2020.100067

Ketrat, S., Japrung, D., & Pongprayoon, P. (2020). Exploring how structural and dynamic properties of bovine and canine serum albumins differ from human serum albumin. *Journal of Molecular Graphics and Modelling, 98,* 107601.

Kobayashi, K., Ogasahara, N., & Matsuoka, A. (1991). Glycation of human serum albumin in long-term incubation with low and high concentrations of glucose. *Chemical & Pharmaceutical Bulletin*, *39*(4), 1080–1081. https://doi.org/10.1248/cpb.39.1080

Kragh-Hansen, U. (1990). Structure and ligand binding properties of human serum albumin. *Danish Medical Bulletin*, *37*(1), 57-84.

Lesne, A. (2014). Shannon entropy: A rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics. *Mathematical Structures in Computer Science*, *24*(3), 25-31.

Li, S., Cao, Y., & Geng, F. (2017). Genome-wide identification and comparative analysis of albumin family in vertebrates. *Evolutionary Bioinformatics*, *13*, 1176934317716089.

Liu, L., & Guo, Q. X. (2001). Isokinetic relationship, isoequilibrium relationship, and enthalpy− entropy compensation. *Chemical Reviews*, *101*(3), 673-696.

Maier, R., Fries, M. R., Buchholz, C., Zhang, F., & Schreiber, F. (2021). Human versus bovine serum albumin: A subtle difference in hydrophobicity leads to large differences in bulk and interface behavior. *Crystal Growth & Design*, *21*(9), 5451-5459.

Michnik, A., Michalik, K., Kluczewska, A., & Drzazga, Z. (2006). Comparative DSC study of human and bovine serum albumin. *Journal of Thermal Analysis and Calorimetry*, *84*(1), 113-117.

Moulik, S. P., Naskar, B., & Rakshit, A. K. (2019). Current status of enthalpy-entropy compensation phenomenon. *Current Science*, *117*(8), 1286-1291.

Pastore, M., Calcagnì, A., (2019). Measuring distribution similarities between samples: a distribution-free overlapping index. *Frontiers Psychology, 10,* 1089-1091. https://doi.org/10.3389/fpsyg.2019.01089

Peters Jr, T. (1970). Serum albumin. *Advances in clinical chemistry, 13*, 37-111.

Peters Jr, T. (1985). Serum albumin. *Advances in protein chemistry, 37*, 161-245.

Peters Jr, T. (1995). All about albumin: Biochemistry, genetics, and medical applications. Academic press.

Pierce, J.R. (1980). An introduction to information theory: Symbols, signals and noise. Dover Publications, New York

Puoci, F. (Ed.). (2015). *Advanced Polymers in Medicine*. Berlin: Springer.

Sainani, K. L. (2012). Dealing with non-normal data. *PM&R , 4*(12), 1001–1005. https://doi.org/10.1016/j.pmrj.2012.10.013

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., Tse, T., Sherry, S. T. (2022). Database resources of the national center for biotechnology information. https://doi.org/10.1093/nar/gkab1112

Stone, J.V. (2015). Information theory: A tutorial introduction. Sebtel Press, Sheffield, England

Tayyab, S., & Feroz, S. R. (2021). Serum albumin: Clinical significance of drug binding and development as drug delivery vehicle. *Advances in Protein Chemistry and Structural Biology, 123,* 193–218. https://doi.org/10.1016/bs.apcsb.2020.08.003

Whicher, J., & Spence, C. (1987). When is serum albumin worth measuring?. *Annals of - Clinical Biochemistry*, *24*(6), 572-580.

APPENDIX A:

Data table: Values of ΔH° and ΔS° for BSA and HSA binding to different ligands

Bovine serum albumin data

PMID	Ligand	Entropy (kJ/mol/K)	Enthalpy (kJ/mol)	Reference
33125021	Rifampicin D12DAB	0.234	41.9338	1
	Rifampicin D14DAB	0.203	31.3861	
	Rifampicin D16DAB	0.264	49.9916	
	Rifampicin D18DAB	0.128	9.6599	
	Dapsone D12DAB	0.195	30.5976	
	Dapsone D14DAB	0.287	58.5974	
	Dapsone D16DAB	0.167	19.4294	
	Dapsone D18DAB	0.179	24.6738	
32968899	PVC MPs	0.04317	-41.77	2
32115798	Octyl gallate	0.4475	104.4	3
31212764	Surfactin	-0.0892	-48.42	4
	Ca2+-surfactin	-0.0911	-49.07	
	Mg2+-surfactin	-0.09567	-50.42	
	Cu2+-surfactin	-0.09823	-51.3	
	Zn2+-surfactin	-0.10002	-51.59	
31190435	HAs-1	-0.38537	-131.38	5
	HAs-2	-0.53276	-187.34	
30976176	Compound 5e	-0.485	-172	6
30446713	Bromo-Noscapine	-0.35949	-128.47	7
30430934	Curcumin	0.212	27.184	8
30176766	Fisetholz	0.03541	-17.2	9
30065238	Tilmicosin	0.0471	-11.06	10
29927538	Copper ferrite nanoparticles	-0.58888	-193.85	11
29920814	4-methylmephedrone	0.0643	-6.23	12
29726682	PEtOx (0.4 chains/nm2)	-0.62	-210	13
	PEG 3.3 nm	-0.88	-291	
	PEG 6.7 nm	-1.03	-335	
	PEG 8.0 nm	-1.03	-335	
	PEtOx (0.7 chains/nm2)	-1.03	-340	
	PiPOx-co-PEtOx	-1.03	-340	
	PiPOx	-1.03	-340	
	PNiPAm	-1.03	-340	
29721735	Anthocyanins	0.14972	18.45	14
29659061	Fosinopril	-0.14013	-63.3	15
29600590	Thymol	0.37053	79.55	16
29564326	Neratinib	-0.164	-76.9	17
29476861	Thimerosal	0.01446	-16.1	18
29435431	Ascorbyl palmitate	0.214622	59.206	19
29403963	AHDMAPPC	0.16536	15.151	20
29314579	TTH	0.052	-16.11	21
29282888	PEG-InP/ZnS QDs	0.151	9.5	22

29064419	HL2	-0.30271	-116.21	23
	HL1	-0.30434	-112.3	
	Compound 2	-0.30816	-122.24	
	Compound 1	-0.50835	-187.51	
28976065	Cefalexin	0.03968	-6.9	24
	Cefixime	0.03051	-17.96	
28792461	Compound 1	0.31668	67.1	25
	Compound 2	-0.19114	-84.47	
	Compound 3	-0.01155	-29.55	
	Compound 4	-0.20818	-89.64	
	Compound 5	-0.08871	-54.79	
	Compound 6	-0.0381	-38	
	Compound 7	0.38743	90.73	
28752073	Neomycin	-0.697	-181	26
28749443	Anti-inflammatory agent	-0.744	-253	27
28575123	Divanillin	0.0408	-19.42	28
28494586	CNCs	1.05	280	29
	CNCs	1.04	274	
	CNCs	0.89	228	
	CNCs	0.83	211	
28480368	Triptolide	-0.382	-148.6	30
28419132	Linifanib	-0.11174	-55.91	31
28253382	Rifampicin	0.052058	43.1	32
27830402	Complex 2	1.17015	321.31	33
	Complex 1	0.4856	114.26	
27808463	CdTe 520 QD	0.22603	35.8	34
	CdTe 540 QD	0.18672	22.96	
27640606	Auramine O	0.144	19.2	35
27508228	AG50 (NaCl 0 M, pH 4.8)	0.05558	-12.11	36
	AG50 (NaCl 0.09 M, pH 4.8)	0.05542	-11.52	
	AR1 (NaCl 0 M, pH 4.8)	0.05334	-18.59	
	AR1 (NaCl 0.04 M, pH 4.8)	0.05221	-17.81	
	AG50 (NaCl 0.04 M, pH 4.8)	0.05155	-12.94	
	AG50 (NaCl 0.15 M, pH 4.8)	0.04559	-14.14	
	AG50 (NaCl 0 M, pH 7.4)	0.03406	-13.63	
	AG50 (NaCl 0.04 M, pH 7.4)	0.02447	-16.29	
	AG50 (NaCl 0.15 M, pH 7.4)	0.01933	-17.53	
	AR1 (NaCl 0.15 M, pH 4.8)	0.01756	-25.87	
	AR1 (NaCl 0.04 M, pH 7.4)	0.01711	-18.95	
	AG50 (NaCl 0.09 M, pH 7.4)	0.01351	-19.38	
	AR1 (NaCl 0.09 M, pH 4.8)	0.01229	-28.48	
	AR1 (NaCl 0 M, pH 7.4)	0.00871	-21.8	
	AR1 (NaCl 0.09 M, pH 7.4)	0.00708	-21.75	
	AR1 (NaCl 0.15 M, pH 7.4)	0.00103	-23.12	
27478738	GABA derivative	0.22267	37.08	37

27377878	Ethambutol	-0.1342	-65.34	38
27118686	CdSeS/ZnS quantum dot	0.34411	51.671	39
26751077	Afatinib	0.04534	-20.69	40
26668672	Aceclofenac	0.144913	14.051	41
26139573	N-acetyl cysteine	-0.06139	-36.7	42
25985127	Sorafenib	-0.1404	-72.2	43
25828537	16-E2-16	0.06449	-8.408	44
25652544	Curcumin (pH 4.0)	0.27254	58.16	45
	Curcumin (pH 7.4)	0.0911	0.498	
	Curcumin (7.4 with 5 M urea)	-0.27772	-111.62	
25382435	Limonene	-0.02732152	-30.1248	46
24852109	Megestrol acetate	-0.2956	-124	47
24398555	Lysionotin	-0.03593	-40.81	48
24123897	Cyanidin-3-glucoside	-0.29082984	-124.01376	49
24030872	Clenbuterol hydrochloride	-0.01313	-30.44	50
23831983	Tetrandrine	0.02277	-27.61	51
23527100	Naringin	0.04723	-13.64	52
	Naringin palmitate	-0.07659	-4.11	
23308237	BBE	0.04087	-15.13	53
23261625	Prednisolone	-0.3707	-149.6	54
23237845	Isoflavone genistein	0.04275	-14.64	55
23136717	Isoprenaline hydrochloride	0.12	10.02	56
22916242	Epicatechin gallate	0.09535336	-2.5719048	57
	Epicatechin	0.08196456	-4.0530408	
22746363	ZnO NP	0.025104	-17.9912	58
	ZnO-PEI	0.003138	-26.7776	
22733490	Ergosterol	0.01912	-16.241	59
22730061	Chlortetracycline	-0.0976	-51.2	60
22615615	Erlotinib hydrochloride	0.217278	38.91	61
22402577	Anastrozole	0.15918	92.99	62
22173815	GdW2	0.07485	-6.84	63
	NdW2	0.0647	-9.59	
	YbW2	0.05561	-11.58	
	DyW2	0.05414	-11.59	
	PrW2	0.04835	-13.99	
	TbW2	0.0359	-17.56	
	SmW2	0.02297	-22.33	
21945127	Pegylated puerarin	0.02001	4.09	64
21901752	Tetrabromobisphenol A	0.1345	7.18	65
21831703	Farrerol	0.00506	-29.92	66
21804222	Anhydrotetracycline	0.01684	-25.5224	67
21718002	Luminol	0.057997	-13.529	68
21670962	MCPA-Na	0.07269	-8.254	69
21561769	BON	0.0104	-23.27	70
21088910	Benzophenone	-0.05305	-43.73	71

20936333	DNP	0.02351	-21.12	72
20593228	CdSe/CdS quantum dots	-0.02436	-33.39	73
20427226	Alpinetin	0.16604	22.1	74
20397730	PFOA	-0.00332	-27.61	75
	PFDA	-0.01196	-37.14	
20129752	Ractopamine	0.07839	-13.47	76
19727573	Co(phen)3	0.08227	-2.73	77
19583232	MNPs	0.15738	-0.9	78
19480002	Myricetin	0.034695	-17.091	79
19437138	Benzidine	-0.02589	-34.11	80
19420779	Hoechst 33258	0.49018	102.785	81
19367911	Flindersine	-0.55	-193	82
19185535	Berbamine	-0.1734	-76.5	83
19093579	Cefuroxime axetil	0.0819	-13.43	84
18844172	Levofloxacin	0.1721	20.94	85
18800727	Vincristine	-0.12938	-62.07	86
18786760	Malachite green	-0.01123	-27.25	87
18719740	Efonidipine	0.31942	68.04	88
18478191	Cu(II)L	-0.17548	-80.79	89
18452134	La(III)L(2)	-0.03261	-41.03	90
18351302	Trans-resveratrol	-0.0192	-12.58	91
18310949	Chlorogenic acid	0.08331	-3.56	92
	3,4-diCQM	0.07968	-13.76	
	3,5-diCQM	0.07526	-12.43	
	3,5-diCQA	0.07349	-11.94	
	3,4-diCQA	0.07314	-12.99	
18051539	Phenazopyridine hydrochloride	0.1248	-71.2	93
17960331	Cu phen2+3	0.05435	10.74	94
17703351	Florasulam	-0.1136	-57.89	95
17623271	3-nitroaniline	0.03346	-11.51	96
	4-nitroaniline	0.03327	-11.78	
	2-nitroaniline	0.03038	-12.53	
17610308	TPP	0.31835	70.96	97
	TMeOPP	0.16797	30.02	
	TClPP	-0.13389	-64.56	
17541184	Cinnamic acid	0.038028	-16.457	98
17303235	Vitamin K(3)	-0.46508	-164.09	99
17227676	Naproxen	-0.305432	-114.47424	100
17030362	LMF	0.047438	-7.97	101
	LMF-Cu	0.033542	-12.469	
16946527	Indirubin	0.112756	-2.744	102
16733318	Carbamazepine	0.19485	33.9	103
16730793	Artemisinin	0.107419	-3.625	104
15684430	Alhydrogel	0.45	95	105
15324481	Schiff base selenide	0.1776	-27.82	106

15137995	Daphnetin	0.0193	-24.21	107
15134145	Nigerloxin	0.11506	0	108
14991661	Isofraxidin	0.05138	-17.63	109
12830359	Magnolol	0.01588	-28.53	110
12736459	Barbaloin	0.0236	-23.7	111
11951988	Cadmium	-0.0103	-28.4	112
9101704	Halothane	-0.092	-40	113
9083685	Chloroform	0.0281	-10.37	114
2923909	3,3',5'-triiodo-L-thyronine	0.0426768	-19.20456	115

Human serum albumin data

PMID	Ligand	Entropy (kJ/mol/K)	Enthalpy (kJ/mol)	Reference
34867382	Delicaflavone	-0.006831	-33.69	116
34610201	Remimazolam	-0.207	-97.6	117
33870854	Pazopanib	0.09837	-60.31	118
33455539	Aspartame	-0.05819	-41.2	119
33390094	Cafaminol	-0.28234	-105.88	120
32685852	9-Hydroxyphenanthrene	0.134	11.08	121
32450395	PVC MPs	0.07076	-59.27	122
30909827	EPPC	0.168	22.154	123
30238853	Folic acid	0.016	-20	124
30200461	2'R-ochratoxin A	0.10262	-8.45	125
30176766	Fisetholz	0.02402	-18.28	9
30084622	Carboxyl porphyrin	0.027905	-16.132	126
29403926	Sulfamethoxazole	-0.0413	-36	127
29251407	Canthaxanthin	-0.059697	-46.083	128
29113080	Acyclovir	0.0794	-1.79	129
	Penciclovir	0.06995	-4.47	
29068381	Aflatoxin B1	0.05509	-10.48	130
	Citrinin	0.0209	-24.15	
	Zearalenone	-0.0035	-30.09	
28987797	Busulfan	0.0423852	-5.8576	131
28911681	S-allyl cysteine	-0.255	-100	132
28541121	PPL	0.078	13.99	133
	BXL	-0.563	-214.3	
28504004	Diazinon	0.116	-16.695	134
28374530	Fluphenazine	0.06842	-4.637	135
27794621	PND	-0.06838	-52.86	136
27768950	3,4,5-triCQA	0.012429	-28.802	137
27478738	GABA derivative	0.0695	-10.63	37
27424099	Axitinib	0.06821	-8.38	138
26142245	Tranilast	0.0469	-25.2	139
26055127	6-Shogaol	0.05252	-11.76	140
25761396	A8HQ (pH 9.0)	0.05961	-12.42	141
	A8HQ (pH 7.4)	0.0457	-15.57	
	A8HQ (pH 6.0)	0.0371	-17.94	
25007630	Levofloxacin	-0.10538	-59	142
24878559	Chloramphenicol	-0.02581	-34.77	143
24481880	Aspartame	-1.339	-458.67	144
24463244	Sulfamethoxazole	0.03233	-16.4	145
24039032	Vincamine	0.07626	-4.57	146
24030872	Clenbuterol hydrochloride	0.05017	-10.6	50
23831983	Tetrandrine	0.31758	59.15	51

23339149	Strictosamide	0.07775	-3.01	147
23322489	Sulfadiazine sodium	0.03856	-9.5	148
23261611	Demeclocycline	-0.06513	-53.01	149
23233363	Methacycline	-0.21813	-95.29	150
23140723	Vanillin	0.0058	-20	151
22828882	Triclosan	0.0326	-37.9	152
22810650	3-(2-cyanoethyl) cytosine	0.0777	-6.47	153
22733490	Ergosterol	0.143245	20.146	59
22658498	Amodiaquine	-0.05003	-43.27	154
22624666	Flavokawain B	0.0695	-6.87	155
22449330	TNN	0.07275	-7.31	156
22402577	Anastrozole	0.15919	99.43	62
22307229	Diclofenac	0.0655	-6.6	157
	Ketoprofen	0.0436	-10.5	
	Clofibric acid	0.0392	-11.4	
22292768	Dihydromyricetin	0.01821	-28.76	158
22169567	Rhaponticin	0.00158	-2.75	159
22077361	[Pt(3)LCl(3)](ClO(4))(3)	0.1455	14.6	160
21845372	ODNR	0.03652	-17.97	161
21681428	Chlorogenic acid	0.012	-22	162
20931763	Vindoline	0.07998	-10.3	163
20827515	FBTZ	-0.0049	-36.55	164
20671814	[(cymene)Ru(ATSC)Cl]PF6	0.152	16.5	165
20171066	Thymoquinone	0.045	-10.24	166
19616992	BAFP	0.06733	-7.75	167
19588173	5-Methyluridine	0.033	-12.62	168
19345112	CDNR	0.02623	-19.89	169
19215910	2'-deoxyuridine	0.024	-18.87	170
18838170	Ni(OAc)(2)L(2).2H(2)O	0.046396	-11.026	171
	Cu(OAc)(2)L(2).2H(2)O	0.046339	-11.533	
18585272	Gallic acid	0.04942	-8.1	172
18456467	CNPB	0.00957	-17.35	173
18330291	Colchicine	0.051507	-11.66	174
18297375	Bergenin	0.03076	-14.45	175
18261869	Vitamin B12	0.06673	-13.38	176
18178358	Morin	0.12915	8.97	177
18058205	Puerarin	-0.00563	-28.01	178
18036655	Daunorubicin	0.02786	-16.13	179
18035550	Glyphosate	0.00638	-21.78	180
17966133	HEA	0.03023	-24.05	181
17575357	EPNT	0.0782702	-6.292	182
16705651	(-)-Epigallocatechin-3-gallate	0.01623	-22.59	183
16569428	All-trans retinoic acid	0.10614	0.10617	184
16420470	Isoflavones	-0.087864	-55.06144	185

16219344	5,7,4'-trihydroxy-6,3',5'-trimethoxyflavone	0.03662	-18.7	186
15942949	Cardamonin	0.00704	-25.312	187
15715992	Jatrorrhizine	0.056267	-10.891	188
15698801	Alpinetin	0.05397	-10.2	189
15664334	Carbamazepine	-0.01489504	-26.9868	190
15556236	Genistein	0.0196	-22.24	191
15468244	Daphnetin	0.05248	-12.45	192
14723965	Isofraxidin	0.07357	-10.08	193
7132952	Warfarin	0.087864	-5.0208	194
5771186	Digitoxin	0.1414192	14.35112	195
	Digitoxin	0.1351432	11.88256	
844942	Bilirubin	-0.035564	-56.484	196

ΔH° and ΔS° data table: BSA and HSA binding to different ligands. PMID is the PubMed Identification number of each reference article. The database of PubMed was accessed between January 2021 to December 2021.