



## 저작자표시 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#) 

이학석사 학위논문

**Understanding Antibody Viscosity  
Based on Protein-Protein Docking**

단백질-단백질 도킹을 이용한 항체의 점성 분석

2022 년 8 월

서울대학교 대학원

화학부

Katsuhito Inui

# Understanding Antibody Viscosity Based on Protein-Protein Docking

지도교수 석 차 옥

이 논문을 이학석사 학위논문으로 제출함

2022 년 8 월

서울대학교 대학원

화학부 전공

Katsuhito Inui

Katsuhito Inui 의 이학석사 학위논문을 인준함

2022 년 8 월

위원장 \_\_\_\_\_ 정연준 \_\_\_\_\_ (인)

부위원장 \_\_\_\_\_ 석차옥 \_\_\_\_\_ (인)

위 원 \_\_\_\_\_ 신석민 \_\_\_\_\_ (인)

## ABSTRACT

# Understanding Antibody Viscosity Based on Protein-Protein Docking

Katsuhito Inui

Department of Chemistry

The Graduate School

Seoul National University

There are many antibody drugs that have been approved and practically used. Antibodies have been considered for therapeutic purposes because of their high affinity and specificity for their targets and their ability to elicit immune responses. However, antibodies must meet some biophysical properties to be used as drugs. One of them is viscosity. Antibody viscosity usually gets higher with antibody concentration. High viscosity makes it hard to be injected into human bodies in high concentrations. However, experimental methods to improve antibody viscosity are time-consuming and expensive. Therefore, computational methods to understand and to predict antibody viscosity is desired. In this thesis, we introduce a new physics-based method to explain the sequence-dependence of antibody viscosity by using antibody structure prediction and protein-protein docking.

**keywords:** antibody, viscosity, protein-protein docking, protein-protein interaction

***Student Number:*** 2020-27795

# TABLE OF CONTENTS

ABSTRACT .....	i
TABLE OF CONTENTS .....	ii
LIST OF FIGURES .....	iv
LIST OF TABLES .....	v
1. INTRODUCTION .....	1
1.1. Review of previous studies on viscosity prediction .....	2
1.1.1. Empirical method: Tomar <i>et al.</i> .....	2
1.1.2. Empirical method: Li <i>et al.</i> .....	4
1.1.3. Empirical method: Sharma <i>et al.</i> .....	5
1.1.4. Physics-based method: Chaudhri <i>et al.</i> .....	6
1.1.5. Physics-based method: Bülow <i>et al.</i> .....	6
2. METHODS .....	8
2.1. Overview of the new method based on protein-protein docking .....	8
2.2. Antibody viscosity data sets .....	8
2.3. 3D structure prediction of antibodies in the data set .....	12
2.4. Protein-protein docking .....	12
2.5. Greedy clustering .....	14
2.6. Solvation energy analysis .....	15
3. RESULTS .....	16
3.1. Fv-Fv docking results .....	16
3.2. Origin of the spurious negative correlation of viscosity with the electrostatic	

score.....	22
4. CONCLUSION.....	26
SUPPLEMENTARY INFORMATION.....	27
BIBLIOGRAPHY .....	30
국문초록.....	33

# LIST OF FIGURES

<b>Figure 1.1.</b> Amino acid sequences of hinge regions of different antibody isotypes.	3
<b>Figure 2.1.</b> Blocked region in protein-protein docking.....	13
<b>Figure 3.1.</b> Viscosity versus average GalaxyTongDock score for Top 500 conformations for 40 antibodies.....	17
<b>Figure 3.2.</b> Viscosity versus the average of score components for Top 500 conformations.....	21
<b>Figure 3.3.</b> Three-dimensional dimer structure of a mutant with low viscosity that shows high electrostatic score.....	23
<b>Figure 3.4.</b> Correlation between the viscosity and the solvation energy gap.....	25

# LIST OF TABLES

<b>Table 2.1.</b> The sequence of an anti-PDGF-BB antibody AB-001.....	10
<b>Table 2.2.</b> Viscosity of an anti-PDGF-BB Antibody and its mutants.....	10
<b>Table 3.1.</b> Correlation coefficients $R^2$ of viscosity with different energy components of GalaxyTongDock.....	18
<b>Table 3.2.</b> Top 500 average score components including weight parameters for 40 antibodies.....	18
<b>Table 3.3.</b> The comparison between R1-011 (the highest viscosity in round 1) and R1-016 (the lowest viscosity in round 1).....	24

# 1. INTRODUCTION

Antibody drugs have become one of the essential parts of drug discovery fields for their target specificity and their roles in immune responses (Marsden *et al.*, 2014). In addition to their target specificity, antibody drugs must satisfy various other conditions such as stability, immunogenicity, solubility, viscosity, and toxicity (Kuroda *et al.*, 2020, Brady *et al.*, 2014, Brennan *et al.*, 2010). Among these properties, we focus on antibody viscosity in this thesis.

Antibody viscosity is usually explained as a fluid's resistance to flow. It has been known that antibody viscosity gets higher when it is concentrated. For self-administration by the patient, subcutaneous administration is preferred in many cases. (Jezek *et al.*, 2011) For self-administration at home, the volume of the antibody should be small enough, which could make the viscosity higher. (Tomar *et al.*, 2016) However, high viscosity sometimes makes it hard for antibody drugs to be injected into human bodies. For administration to patients, an ideal viscosity value of drug antibodies is under 30 cP when the concentration is 150 mg/mL (Shire *et al.*, 2004).

If antibody viscosity can be predicted effectively by a computational method at an early stage of development, such a method can be used to save time and cost for therapeutic antibody development.

In this thesis, attempts to explain observed viscosity based on empirical parameters or on molecular interaction energy calculations for different sets of antibodies are reviewed. Then, a new physics-based method that employs protein-protein docking is introduced. A key idea behind this method is that viscosity originates from intermolecular interactions among antibody molecules in solution, and that strengths of intermolecular interactions may be estimated by structure-

based, protein-protein docking method. A new insight from the interaction analysis is that non-polar interactions between antibody molecules and the polar solvation effect is critical in explaining sequence-dependent antibody viscosity.

Reviews of previous studies on antibody viscosity prediction are presented in **Section 1.1.** and the methods tested in this thesis are described in **Section 2.** Results and conclusions are presented in **Sections 3.** and **4.**, respectively.

## 1.1. Review of previous studies on viscosity prediction

### 1.1.1. Empirical method: Tomar *et al.*

Tomar *et al.* (2017) performed 105 concentration-dependent measurements on 16 different antibodies and tried to find out relationships between antibody viscosity and computationally obtained parameters such as charges on each antibody region, aggregation propensity predicted by existing methods, dipole moment of the antibody, and zeta potential, the electrical potential at the slipping plane, of the antibody. The relation between viscosity and concentration was described as

$$\frac{\eta}{\eta_0} = A \exp(Bc), \quad (1)$$

where  $\eta$  is the viscosity of the antibody solution with concentration  $c$ , and  $\eta_0$  is the viscosity of the platform buffer. A and B are the coefficients that can take different values based on the antibody type or the experimental condition. Taking the logarithm of the two sides of the equation leads to

$$\ln \frac{\eta}{\eta_0} = \ln A + Bc. \quad (2)$$

By plotting  $\frac{\eta}{\eta_0}$  against  $c$ , the linear function of the slope  $B$  and the intercept  $\ln A$  can be calculated. By obtaining the values of the coefficients  $A$  and  $B$  from available experimental data, viscosity values at a different concentration can be calculated using **Eq (2)**.

The authors employed empirical descriptors to predict the parameter  $B$ . Square of correlation coefficient ( $R^2$ ) between the experimentally derived value of parameter  $B$  and each computationally estimated descriptor was at most 0.32, meaning no descriptors alone can fully explain concentration-dependent viscosity behavior. By combining these descriptors, they obtained a best performing model for the prediction of parameter  $B$  as follows:

$$B = 1 + x_3 * x_7 + x_3 * x_{13} + x_5 * x_7 + x_7 * x_{13} \quad (3)$$

where  $x_i * x_j = x_i + x_j + x_i x_j$ , and each of  $x_3$ ,  $x_5$ ,  $x_7$ , and  $x_{13}$  denotes the charge on  $V_H$  region, the net charge on the hinge region (sequence shown below in **Figure 1.1**), the net charge on  $V_L$  region, and the solvent-accessible hydrophobic surface area of the full-length antibody calculated from homology model structures.

**Figure 1.1. Amino acid sequences of hinge regions of different antibody isotypes.** (Tomar, D. S. *et al.* (2017) *MAbs*. 9. 3. 476-489.)

Hinge Region Sequence													Isotype									
E	P	K	S	C	D	K	T	H	T	C	P	P	C	P	A	P	E	L	L	G	G	IgG <sub>1</sub>
E	R	K	C	C	V	E	-	-	-	C	P	P	C	P	A	P	P	A	A	A	A	IgG <sub>2</sub>
E	S	K	Y	G	P	P	-	-	-	C	P	P	C	P	A	P	E	F	L	G	G	IgG <sub>4</sub>

One merit of this method is one can predict viscosity of the certain antibody for the arbitrary concentration. However, this scheme has some limitations. In this method, only 16 antibody molecules were used, so it is not certain how the descriptors employed in the study would be generalizable to other sets of antibody viscosity data. Second, only a linear model was used. Third, they used homology-based molecular models for calculating descriptors for **Eq (3)** but what is missed here is that antibodies are dynamic macromolecules, which means their structures can change over time. Using the only 1 modeled structure for these descriptors, therefore, could lead to less accuracy. They claimed that using the average values for these descriptors derived from conformational ensembles may help the performance improved. The fourth limitation is that **Eq (3)** is not transferable to different formulations, buffer and pH. The last limitation is that even though mathematical analysis implies the correlation of the viscosity and each descriptor, it does not explain the causal relationships. They claim that molecular descriptors mentioned in their method may correlate with more than one of antibody phenomenon including viscosity, solubility, and aggregation because they all arise from protein-protein interactions.

### **1.1.2. Empirical method: Li *et al.***

Li *et al.* (2014) generated concentration-dependent viscosity data on 11 different therapeutic antibodies. This experiment was conducted under the condition of pH 5.8 and with no added salt in the buffer to facilitate direct comparisons among the mAbs. The particular pH was chosen because the optimum pH range for high concentration mAb formulations is 5.5-6.3 since colloidal interactions involving only the constant domains (Fc) of human antibodies are expected to be predominantly repulsive in the pH range. The correlation between the logarithm of

relative viscosity and concentration was high with  $R^2 > 0.96$  for all 11 antibodies.

The authors found both two descriptors, the net charge of F<sub>V</sub> portion,  $Z_{FV}$ , and the zeta potential of F<sub>V</sub> portion,  $\xi_{FV}$ , have good correlations with antibody viscosity. Zeta potential is an indicator of the stability of colloidal dispersions. Suspensions of particles showing a high absolute value of zeta potential are reported to be more stable than those exhibiting lower absolute zeta potential values (Hanaor *et al.*, 2012) For the 11 antibodies, the zeta potential was calculated based on structure models. When both  $Z_{FV}$  and  $\xi_{FV}$  are positive, the net charge and the zeta potential of the whole antibody become larger than those of the constant region because the net charge of F<sub>C</sub> is positive, then the whole antibody molecule becomes more repulsive, which leads to low viscosity. When the net charge and the zeta potential of the whole antibody become smaller than those of the constant region, the electric polarization is enhanced, which allows more self-associations, which in turn leads to high viscosity.

This work has similar limitations to that of Tomar *et al.* (2017). Among them, it is a critical limitation that data for only 11 antibodies were interpreted and the generalizability of the model is not proven yet.

### **1.1.3. Empirical method: Sharma *et al.***

Sharma *et al.* (2014) calculated the correlation between the experimentally measured viscosity of 14 antibodies and three sequence-dependent properties: net charge of F<sub>V</sub> at pH 5.5, F<sub>V</sub> charge symmetry parameter (F<sub>V</sub>CSP) between the V<sub>H</sub> and V<sub>L</sub> domain at pH 5.5, and the hydrophobicity index (HI) of the F<sub>V</sub>. The net charge potentially contributes to repulsive interactions, whereas F<sub>V</sub>CSP and HI can potentially contribute to attractive interactions. A linear model based on the three

features showed a strong correlation with the viscosity for the 14 antibodies. This model also has similar problems as the other empirical methods described above.

#### **1.1.4. Physics-based method: Chaudhri *et al.***

Chaudhri *et al.* (2020) performed a coarse-grained (CG) simulation for large systems of 1,000 antibody molecules to explain the sequence dependence of antibody viscosity. They developed two reduced CG models (12-site and 26-site) of two antibodies where the 12-site model has one bead for each of four domains for each of two Fv's and an Fc, and the 26-site model is the same model with an additional bead for each of three H-CDRs and three L-CDRs and in each of two hinge regions. The CG molecular dynamics simulations were performed for antibodies with swapped charges, leading to very different behavior even for small changes in the amino acid sequence from heterogeneous sense cluster formations to homogeneous molecular distributions. Although this CG model provides interesting insights into sequence-dependent antibody viscosity, it is hard to be applied directly to real antibodies.

#### **1.1.5. Physics-based method: Bülow *et al.***

Bülow *et al.* (2019) performed large-scale all-atom, explicit-water molecular dynamics simulations of dense protein solutions in order to understand viscosity at high concentration. Although the simulated proteins are not antibodies, both concentration dependency and absolute values of viscosity were explained very

accurately for multiple proteins. Atomic details about the intermolecular interactions responsible for high viscosity at high concentrations were also revealed. However, such large-scale molecular simulations are not affordable on a routine basis because of the requirement for large computation resources, and they are also hard to be applied to large antibody molecules.

## 2. METHODS

### 2.1. Overview of the new method based on protein-protein docking

In this thesis, a new physics-based method for understanding sequence-dependent viscosity change using protein-protein docking is presented. First, single-molecule antibody structures are predicted from the amino acid sequences. Next, protein-protein docking is performed to generate antibody-antibody complex structures that may form transiently at high concentrations. Finally, score components of the docking algorithm and generated structures are analyzed to understand correlations with the observed viscosity values.

### 2.2. Antibody viscosity data sets

At present only a limited amount of data is publicly available on experimentally measured viscosity values for different monoclonal antibodies. In this thesis, the data set reported by Apgar *et al.* (2020) was used to test the physics-based viscosity prediction model based on molecular docking. Apgar *et al.* performed two rounds of mutation design of an anti-PDGF-BB (platelet-derived growth factor B homodimer) antibody to reduce viscosity in which 39 mutants were generated in addition to the initial antibody. The sequence of the initial antibody is provided in **Table 2.1**. They measured viscosity for the original antibody (AB-001 in **Table 2.2**) and 39 mutants (**Table 2.2**). The mutants R1-002 to R1-018 were generated in the first round and the mutants R2-001 to R2-022 were generated in the second round based on the lowest-viscosity mutant in the first round, R1-016. Amino acid mutations introduced to the mutants are specified in **Supplementary Table S1 (a)**

for the heavy chain and (b) for the light chain. According to **Table 2.2**, the viscosity values range from high ( $> 150$  cP), medium ( $30 \sim 150$  cP), and low ( $< 30$  cP) values. Round 1 mutants have high viscosity except for two mutants with medium viscosity, R1-010 and R1-016, and Round 2 mutants have medium (12 mutants), and low (8 mutants) viscosity with two mutants with unknown viscosity. The viscosity of the mutant R2-002 and R2-003 could not be measured because both had formed precipitates during concentrating.

In the typical buffer pH (4.5  $\sim$  7.5), antibodies have a net positive charge and the majority of the constant region also has a positive net charge. (Li *et al.*, 2014, Boswell *et al.*, 2010, Bumbaca *et al.*, 2012) So negative patches on complementarity-determining regions (CDR) or variable regions could lead to self-association, which in turn leads to high viscosity. (Apgar *et al.*, 2020) In the original paper by Apgar *et al.* (2020), where they tried to make mutants whose viscosity is less viscous, they made 39 mutants by basically adding positively charged residues or removing negatively charged residues.

**Table 2.1. The sequence of an anti-PDGF-BB antibody AB-001.** The heavy chain shown in the upper panel (a) is composed of 120 amino acids, and the light chain shown in the lower panel (b) is composed of 106 amino acids.

(a) Heavy chain

<b>FW1</b>	<b>CDR1</b>	<b>FW2</b>	<b>CDR2</b>
EVQLLESGGGLVQPGGSLRLSCAAS	GFTFSSYAMS	WVRQAPGKGLEWVS	YISDDGSLKYYADSVKG
<b>FW3</b>	<b>CDR3</b>	<b>FW4</b>	
RFTISRDN SKNTLYLQMN SLRAEDTAVYYCAK	HPYWYGGQLDL	WGQGLVTVSS	

(b) Light chain

<b>FW1</b>	<b>CDR1</b>	<b>FW2</b>	<b>CDR2</b>
SYELTQPPSVSVSPGQTASITC	SGDSLGSYFVH	WYQQKPGQSPVLVIY	DDSNRPS
<b>FW3</b>	<b>CDR3</b>	<b>FW4</b>	
GIPERFSGNSNGNTATLTISGTQAMDEADYYC	SAFTHNSDV	FGGGTKLTVL	

**Table 2.2. Viscosity of an anti-PDGF-BB Antibody and its mutants.**

MUTANTS	Viscosity [cP] @ 150 mg/mL	class
AB-001	439.7	HIGH
R1-002	288.3	HIGH
R1-003	522.9	HIGH
R1-004	310.4	HIGH
R1-005	190.2	HIGH
R1-006	313.7	HIGH
R1-007	233.2	HIGH
R1-008	567.2	HIGH
R1-009	430.4	HIGH
R1-010	98.6	MEDIUM
R1-011	519.4	HIGH
R1-012	471	HIGH
R1-013	414.3	HIGH
R1-014	414.7	HIGH
R1-015	452.1	HIGH

R1-016	73.1	MEDIUM
R1-017	1534	HIGH
R1-018	415.8	HIGH
R2-001	37.3	MEDIUM
R2-002	-	-
R2-003	-	-
R2-004	54.4	MEDIUM
R2-005	36.7	MEDIUM
R2-006	12.6	LOW
R2-007	21.1	LOW
R2-008	23	LOW
R2-009	19.3	LOW
R2-010	35.1	MEDIUM
R2-011	25.8	LOW
R2-012	38.9	MEDIUM
R2-013	26.4	LOW
R2-014	50.8	MEDIUM
R2-015	83.2	MEDIUM
R2-016	66.5	MEDIUM
R2-017	83.5	MEDIUM
R2-018	19.9	LOW
R2-019	59.9	MEDIUM
R2-020	10	LOW
R2-021	118.9	MEDIUM
R2-022	134.7	MEDIUM

### 2.3. 3D structure prediction of antibodies in the data set

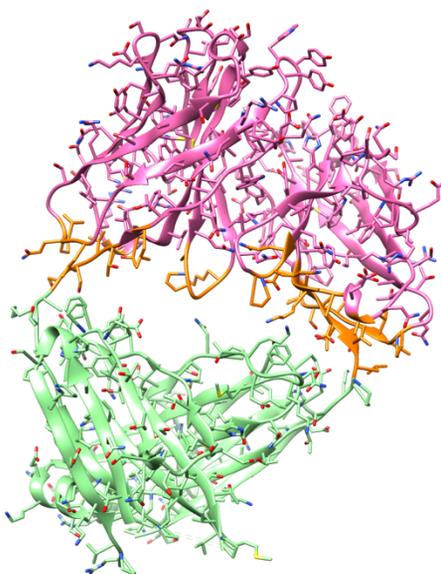
Three-dimensional antibody structures are built with AlphaFold2 (Jumper *et al.*, 2021) for the initial anti-PDGF-BB antibody and its 39 mutants before protein-protein docking. Only Fv regions were modeled in this study for efficient protein-protein docking. Because a related 3D structure of the PDGF-B blocking antibody is available in the protein data bank (PDB ID: 4QCI, Kuai *et al.*, 2015), the predicted structures are expected to be very accurate. The protein sequence of AB-001 is different from that of 4QCI only by 7 residues. The root-mean-square deviation (RMSD) between all pairs of antibody structures was at most 0.54 Å.

### 2.4. Protein-protein docking

To generate complex conformations for two antibody molecules that may form transiently at high concentrations, GalaxyTongDock (Park *et al.*, 2019) was used. GalaxyTongDock is an *ab initio* protein-protein docking program that performs rigid-body docking by a Fast Fourier transformation. GalaxyTongDock was run for each of Fv sequences with default options with the block options which prohibit designated residues from interfaces. Since only the F<sub>v</sub> region of the antibody structure (shown in pink in **Figure 2.1**) was used as an input, direct use of GalaxyTongDock may result in cases where one antibody interacts with the other in the area which is connected to the deleted part of the antibody (shown in green in **Figure 2.1**). To prevent this, those atoms in close proximity (within 8 Å) with the deleted area (shown in orange in **Figure 2.1**) were assigned as blocked residues. The generated conformations were clustered with RMSD cut-off of  $\sqrt[3]{n}$ , where  $n$

is the number of residues, and the 3D structures of the highest-score representatives were observed in detail.

**Figure 2.1. Blocked region in protein-protein docking.** The region colored in pink and orange is the Fv region which was actually used as an input for docking simulation. The region colored in green was removed in docking for computational efficiency, so the region colored in orange was prevented to interact with the other antibody during docking.



Since the score components of GalaxyTongDock were examined, the score components are explained here. The GalaxyTongDock energy is composed of six components as follows:

$$E_{GalaxyTongDock} = E_{SC_{rep}} + w_1 E_{SC_{att}} + w_2 E_{Elec} + w_3 E_{ACE} + w_4 E_{IFACE} + w_5 E_{consv} \quad (4)$$

where the first five terms are from ZDOCK (Mintseris *et al.*, 2007, Chen *et al.*, 2003) (repulsive and attractive parts of the shape complementarity score (Mintseris *et al.*, 2007)  $E_{SC_{rep}}$  and  $E_{SC_{att}}$ , Coulomb energy with distance-dependent dielectric constant  $E_{Elec}$  (Gabb *et al.*, 1997), atomic contact energy  $E_{ACE}$  (Zhang *et al.*, 1997) of ZDOCK2.3.2 (Chen *et al.*, 2003), and interface atomic contact energy (Mintseris *et al.*, 2007)  $E_{IFACE}$  of ZDOCK3.0.2 (Mintseris *et al.*, 2007)). The last term  $E_{consv}$  is a conservation score which accounts for amino acid sequence conservation among evolutionary related proteins (Liang *et al.*, 2009). The weight factors  $w_1 \sim w_5$  were optimized for the best docking performance for benchmark sets (Park *et al.*, 2019).

## 2.5. Greedy clustering

Since docking structures are dynamic system, a proper clustering method to select representative docking structures is needed. In this thesis, we adopted a greedy method. (REF) Greedy clustering starts from sorting the docking structures based on GalaxyTongDock scores. A clustering radius of  $\sqrt[3]{N}$ , where  $N$  is the number of residues, is used, and RMSD of C $\alpha$  of each residue was considered to measure the distance. In greedy clustering, clustering of the docking structure which has the best docking score is first considered, then the next step of clustering is done after removing all the structures included in the first step. Basically a greedy clustering

method has the trait that the clustering result is affected by scoring function. So it is important to carefully consider which function should be used as a scoring function.

## **2.6. Solvation energy analysis**

For the theoretical explanation for the correlation between the viscosity and the electrostatic term of GalaxyTongDock score, the solvation energy was also calculated. The solvation energy was calculated with the higher resolution program GalaxyRefineComplex (Heo *et al.*, 2016) by giving bound structures and unbound structures as input.

## 3. RESULTS

### 3.1. Fv-Fv docking results

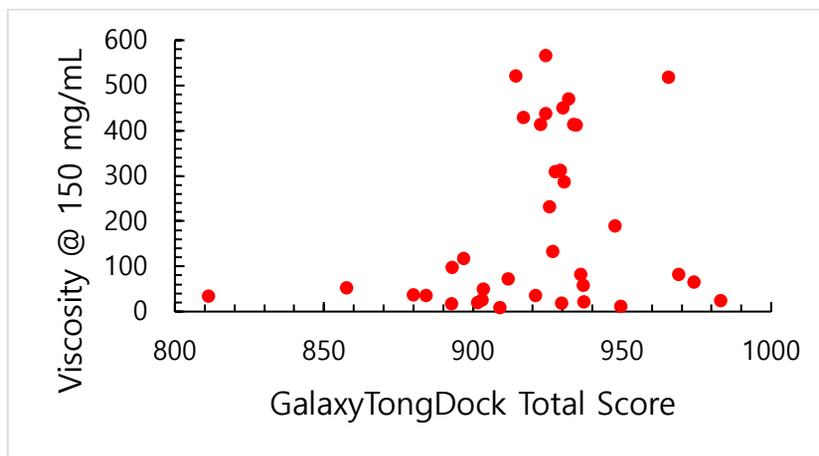
It is expected that a high-concentration antibody solution involves a diverse set of transient dimer and higher-order conformations in its microscopic states. Although a protein-protein docking method was developed to identify the lowest-free energy conformation effectively, we assume here that additional conformations generated during docking also represent some plausible conformations that may exist at high temperatures or high concentrations. This reasoning comes from the fact that the GalaxyTongDock score consists of calculating physics-based energy terms such as van der Waals interactions ( $E_{SC\_rep}$ ,  $E_{SC\_att}$ ) and Coulomb interactions ( $E_{Elec}$ ). Since it is not possible to simulate a rigorous thermodynamic ensemble with the docked conformations, we simply observed the average docking score of Top N poses. We examined  $N = 100, 300, 500,$  and  $1000$  and the discussion and conclusions below do not change depending on the choice of  $N$ . Throughout this thesis, we use  $N = 500$  for convenience when results for a single  $N$  value are discussed.

The average of the total score of GalaxyTongDock does not show any correlation with the viscosity value at  $150 \text{ mg/ml}$  with  $R^2 = 0.03$ , as shown in **Figure 3.1**. The score is designed to be negative of energy, meaning that a positive score corresponds to higher stability. Because GalaxyTongDock consists of various score components statistical potential ( $E_{ACE}$ ,  $E_{IFACE}$ ) and conservation energy ( $E_{CONSV}$ ) in addition to physics-based terms, the score was decomposed into its components ( $E_{SC\_rep}$ ,  $E_{SC\_att}$ ,  $E_{Elec}$ ,  $E_{ACE}$ ,  $E_{IFACE}$ , and  $E_{CONSV}$ ) for further analysis.

We found that the  $R^2$  values for the two components, the attractive part of

the shape complementarity score,  $E_{SC\_att}$ , and the Coulomb electrostatic potential energy score,  $E_{Elec}$ , are significantly higher than all other components, as shown in **Table 3.1**. The  $R^2$  values for  $E_{SC\_att}$ , and  $E_{Elec}$  are 0.36 and 0.40, respectively, for Top 500 average. The individual score values (Top 500 average) calculated with the weight factors for each mutant are also presented in **Table 3.2**.

**Figure 3.1. Viscosity versus average GalaxyTongDock score for Top 500 conformations for 40 antibodies.**



**Table 3.1. Correlation coefficients  $R^2$  of viscosity with different energy components of GalaxyTongDock.**

	<b>Top 100</b>	<b>Top 300</b>	<b>Top 500</b>	<b>Top 1000</b>
$E_{GalaxyTongDock}$	0.0003	0.0158	0.0330	0.0596
$E_{SC\_rep}$	0.0644	0.1068	0.1124	0.1220
$E_{SC\_att}$	0.3519	0.3445	0.3637	0.3575
$E_{Elec}$	0.4065	0.3939	0.3990	0.3783
$E_{ACE}$	0.2744	0.2206	0.2113	0.1979
$E_{IFACE}$	0.0716	0.0802	0.0781	0.0788
$E_{consv}$	0.1776	0.1883	0.1638	0.1128

**Table 3.2. Top 500 average score components including weight parameters for 40 antibodies. Viscosity values are at 150 mg/mL.**

<b>Name</b>	<b>viscosity</b>	$E_{GalaxyTongDock}$	$E_{SC\_rep}$	$E_{SC\_att}$	$E_{Elec}$	$E_{ACE}$	$E_{IFACE}$	$E_{consv}$
<b>AB-001</b>	439.7	924.23	-344.68	2262.39	98.82	256.52	642.66	1.33
<b>R1-002</b>	288.3	930.34	-345.59	2331.24	90.02	253.43	612.50	1.33
<b>R1-003</b>	522.9	914.14	-348.38	2250.12	98.61	249.58	637.47	1.34
<b>R1-004</b>	310.4	927.32	-374.49	2316.68	87.52	285.37	674.58	1.30
<b>R1-005</b>	190.2	947.40	-399.54	2405.06	68.36	306.24	713.80	1.29
<b>R1-006</b>	313.7	929.18	-339.16	2249.82	128.40	240.88	630.30	1.35
<b>R1-007</b>	233.2	925.51	-371.59	2327.89	73.67	289.38	667.36	1.30
<b>R1-008</b>	567.2	924.15	-339.61	2231.71	78.44	278.04	664.20	1.33
<b>R1-009</b>	430.4	916.72	-340.63	2188.62	110.52	269.49	661.82	1.33
<b>R1-010</b>	98.6	892.79	-324.90	2140.97	206.83	205.06	561.53	1.31
<b>R1-011</b>	519.4	965.35	-401.37	2369.00	52.86	321.08	788.02	1.31
<b>R1-012</b>	471	931.87	-334.90	2231.16	137.81	246.93	634.23	1.34
<b>R1-013</b>	414.3	934.36	-355.72	2326.39	103.29	267.77	635.18	1.31
<b>R1-014</b>	414.7	933.68	-348.45	2284.12	123.01	254.36	636.70	1.33
<b>R1-015</b>	452.1	930.00	-345.82	2258.43	120.46	256.08	640.53	1.35

R1-016	73.1	911.55	-358.30	2147.08	195.91	243.32	663.48	1.31
R1-017	1534	927.33	-371.71	2348.09	93.99	267.47	645.70	1.30
R1-018	415.8	922.55	-363.00	2319.13	82.85	269.06	646.76	1.30
R2-001	37.3	920.75	-367.17	2162.51	207.57	246.13	677.89	1.33
R2-002	-	941.09	-339.44	2192.15	206.17	249.91	631.49	1.40
R2-003	-	864.68	-305.94	1979.16	245.79	211.06	550.03	1.33
R2-004	54.4	857.39	-285.40	2046.99	242.31	189.48	445.22	1.37
R2-005	36.7	884.11	-333.42	1999.35	411.04	173.87	520.73	1.32
R2-006	12.6	949.41	-322.40	2121.90	281.94	215.87	620.42	1.38
R2-007	21.1	901.43	-308.23	2095.53	236.87	204.93	550.48	1.35
R2-008	23	936.95	-358.53	2190.08	233.92	241.61	656.00	1.32
R2-009	19.3	892.65	-331.61	2141.47	219.14	207.26	555.78	1.37
R2-010	35.1	811.10	-261.66	1997.82	237.35	175.06	343.22	1.37
R2-011	25.8	982.98	-341.29	2210.09	287.51	230.25	663.84	1.32
R2-012	38.9	879.75	-339.26	2201.26	219.40	188.82	504.89	1.37
R2-013	26.4	902.88	-358.80	2148.57	156.55	281.49	662.86	1.33
R2-014	50.8	903.37	-319.66	2140.70	174.70	214.00	588.71	1.36
R2-015	83.2	935.91	-326.02	2177.44	208.46	230.88	609.19	1.38
R2-016	66.5	973.85	-353.36	2200.63	231.05	267.42	707.95	1.33
R2-017	83.5	968.82	-377.56	2220.26	253.91	249.66	719.69	1.34
R2-018	19.9	929.58	-350.03	2141.54	209.32	246.79	673.76	1.34
R2-019	59.9	936.75	-340.35	2136.82	231.93	260.63	643.89	1.40
R2-020	10	908.84	-316.55	2136.73	203.43	216.96	574.00	1.35
R2-021	118.9	896.58	-356.46	2177.65	222.68	229.79	578.93	1.39
R2-022	134.7	926.45	-354.04	2206.74	221.71	243.26	613.82	1.37

Viscosity and the average score components are shown for 40 mutants in **Figure 3.2**. The attractive shape score and the electrostatic score show the highest correlations, but the electrostatic score shows negative correlation with viscosity. It is understandable that attractive shape score has a positive correlation with viscosity because stronger pair-wise attractive interactions between antibodies would cause higher viscosity. The small correlation of viscosity with the repulsive shape score is presumably due to the highly sensitive nature of the repulsive score

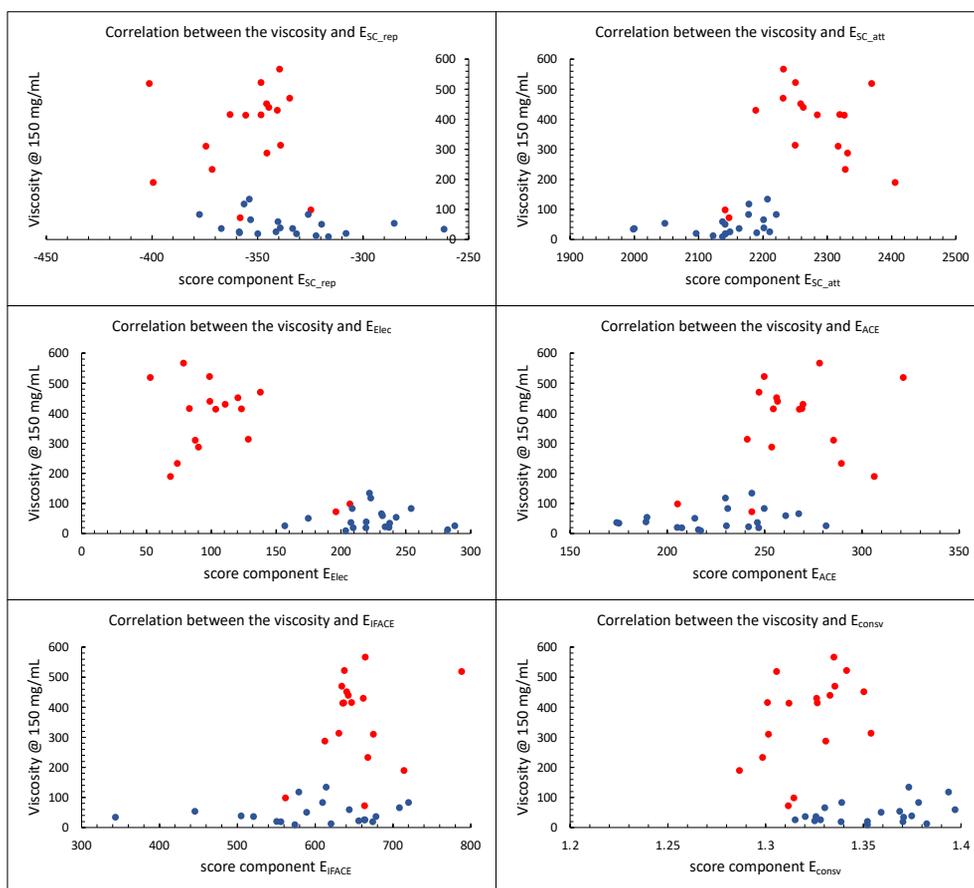
to the level of local structural refinement. The positive correlations of viscosity with the statistical scores ACE and IFACE are also reasonable because these scores were developed to represent attractive energy between atoms. The small correlation with the conservation score also makes sense because the variable regions of the antibody are not well-conserved.

The negative correlation of viscosity with the electrostatic score was puzzling at first, so we investigate further regarding this point in the next subsection.

In terms of the classification, **Figure 3.2.** shows the attractive shape score and the electrostatic score divided the set into 2 groups. While the attractive shape score has roughly divided the set around  $E_{SC\_att} = 2200$ , the electrostatic score seems to divide the set into low and medium viscosity group and high viscosity group more clearly. From **Table 3.2.** and **Figure 3.2.**, all of the mutants whose viscosity at 150 mg/mL is higher than 190.2 cP (R1-005) have  $E_{Elec}$  value smaller than 150, and all of the mutants whose viscosity at 150 mg/mL is less than 134.7 cP (R2-022) have  $E_{Elec}$  value larger than 150. Here, we label the group whose viscosity value is larger than 190.2 cP (at 150 mg/mL) Group 1, and another group whose viscosity value is smaller than 134.7 cP (at 150 mg/mL) Group 2. Group 2 includes all of the Round 2 mutation antibodies and 2 antibodies from Round 1 mutation, which are R1-010 and R1-016. Group 1 includes most of the Round 1 mutation antibodies but does not include R1-010 and R1-016. The interesting point here is these score components can classify the data set into 2 groups regardless of antibody's sequence identity. As seen in mutants' sequences (see Supplementary Information **Table S1.**), sequences of the Round 2 mutation antibodies are different from the sequence of R2-001 by only 1 residue. However, sequences of R1-010 and R1-016, which also belong to Group 2, are different from R2-001 by 7 or 4

residues respectively. This indicates the classification between 2 groups was not caused only by the sequence similarity because 2 mutants which have the lowest viscosity in the Round 1 mutations were classified to low viscosity group.

**Figure 3.2. Viscosity versus the average of score components for Top 500 conformations.** Round 1 mutants are plotted in red color while Round 2 mutants are plotted in blue color.

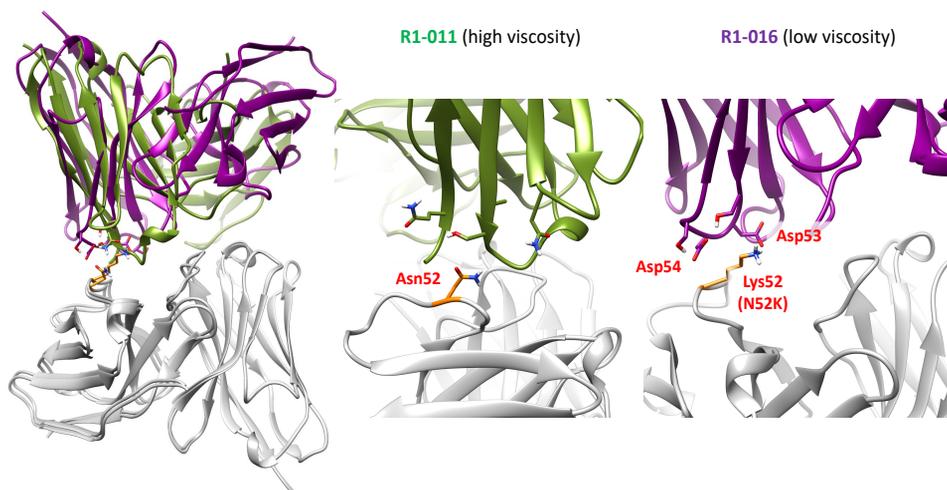


### **3.2. Origin of the spurious negative correlation of viscosity with the electrostatic score**

We could understand the spurious negative correlation of viscosity with the electrostatic score after examining the complex structures generated by docking for the low-viscosity mutants with the high electrostatic score. Those mutants have mutation from ASN to LYS at residue position 52. The three-dimensional structure in **Figure 3.3** shows that the positively charged, mutated LYS residue (colored in cyan) of an antibody molecule (colored in orange) interacts closely with the negative residues, Asp 53 and ASP 54 of another antibody (purple) in a symmetric dimer conformation. Because of these charge-charge interactions, those low-viscosity mutants with LYS at 52 show high electrostatic scores. However, this conformation leaves more spatial voids at the interface, resulting in a lower shape complementarity score.

The weight factors for different score components of GalaxyTongDock were trained to obtain the best possible results for protein-protein complex structure prediction given that the two proteins interact with each other. Therefore, the weight factors are not expected to be optimal for describing physical energy that determines how strongly two given proteins interact with each other. One critical point is that GalaxyTongDock score does not include an explicit solvation free energy term. The solvation effect is considered only implicitly in the statistical score terms ACE and IFACE. The solvation free energy tends to behave in the opposite direction from the Coulomb interaction energy because conformations with favorable intermolecular electrostatic energy have less favorable interactions with solvent water molecules and vice versa.

**Figure 3.3. Three-dimensional dimer structure of a mutant with low viscosity that shows high electrostatic score.**



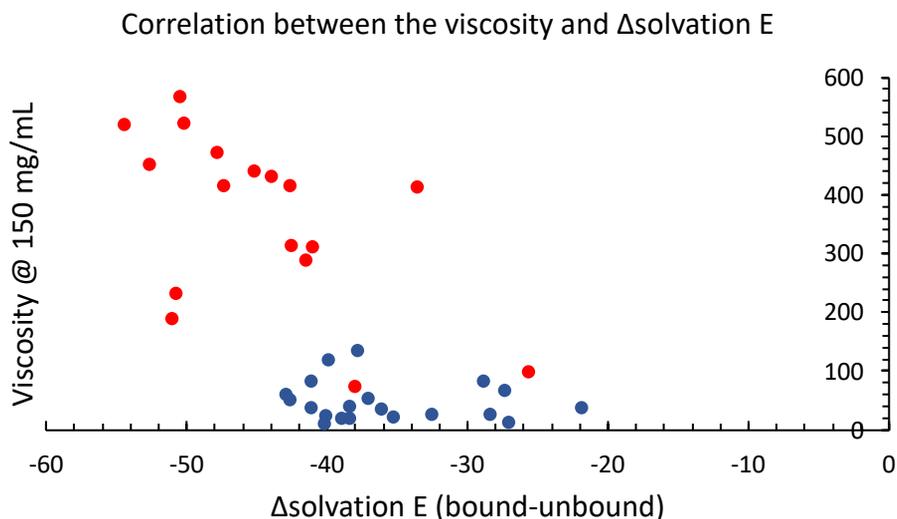
In order to confirm this hypothesis, solvation energy was calculated using 2 structures, the bound state and the unbound state. **Table 3.3.** shows the viscosity, Top1  $E_{Elec}$ , Top 1  $E_{SC\_att}$ ,  $\Delta$ solvation, Top 500  $E_{Elec}$ , Top 500  $E_{SC\_att}$  of R1-011 and R1-016. Notice  $\Delta$ solvation is the gap of the solvation energy while others are score components, which get higher when the complex is energetically stable. The values of  $\Delta$ solvation of R1-011 and R1-016 are both negative, which means that for both mutants, solvation is energetically favorable, However, the absolute value of  $\Delta$ solvation of R1-016 is smaller than one of R1-011. This implies that R1-011 becomes relatively more stable by being solvated than R1-016.

**Table 3.3. The comparison between R1-011 (the highest viscosity in round 1) and R1-016 (the lowest viscosity in round 1).** Notice that  $\Delta$ solvation is the energy gap while other values are score

	Viscosity	Top1 $E_{Elec}$	Top1 $E_{SC\_att}$	$\Delta$ solvation	Top500 $E_{Elec}$	Top500 $E_{SC\_att}$
R1-011	519.4	152.9	996.5	-40.4	19.0	852.8
R1-016	73.1	247.9	519.8	-28.6	70.5	773.0

Next,  $\Delta$ solvation was calculated for other mutants besides R1-011 and R1-016. **Figure 3.4.** shows the correlation between the viscosity values and solvation term calculated by local optimize. The correlation coefficient  $R^2$  is 0.410, even better than the one for  $E_{SC\_att}$  (0.364) and  $E_{Elec}$  (0.399), which were shown in **Table 3.1.** From the value of the correlation coefficient, it is expected that by considering the solvation effect, a better method for viscosity prediction might be achieved.

**Figure 3.4. Correlation between the viscosity and the solvation energy gap.** Red points are mutants from the Round 1 while blue points are mutants from the Round 2.



In summary, the attractive shape score of GalaxyTongDock, which corresponds to van der Waals interaction, could account for stronger interaction between antibody molecules with high viscosity. The electrostatic score component showed a negative correlation with the viscosity because the high electrostatic score originated from conformations with smaller contacts between antibody molecules in the current data set. The negative correlation is also due to the fact that GalaxyTongDock score underestimates the solvation effect that can cancel the favorable electrostatic score.

## 4. CONCLUSION

A new model to explain antibody viscosity in terms of molecular associations at a high concentration was tested in this thesis, inspired by a recent molecular dynamics simulation study. In order to explore possibilities of treating molecular associations much more efficiently than molecular dynamics simulations, a protein-protein docking method was employed. Although the docking score itself did not show a correlation with the antibody viscosity at a high concentration, its score components show correlations. In particular, a score that corresponds to van der Waals interaction shows a strong correlation with viscosity. This implies that protein-protein docking can be used to explain antibody viscosity if a proper score function is developed. Analysis of the docking results on the current data set reveals a problem in the docking energy which underestimates the solvation effect. For the development of a predictive model for antibody viscosity based on protein-protein docking, several requirements arise. First, individual antibody structures have to be predicted accurately for accurate docking results. Protein-protein docking has to be also accurate with accurate score function and effective conformational sampling. Finally, a large set of antibody viscosity data is necessary for training and validation of the predictive model. The current study provides a basis for future study of a predictive viscosity prediction model that effectively accounts for underlying principles. Such a model could be a physics-based model based on docking or a machine learning model inspired by physics-based studies or purely data-driven.

# SUPPLEMENTARY INFORMATION

**Table S1.** (a) List of heavy chain variable regions and (b) of light chain variable regions for mutated sequences. Sequences are compared to the parental AB-001 sequence. [Taken from Apgar et al. (2020) PLoS ONE. 15. 5. e0232713.]

(a)

Name	FW1	CDR1	FW2	CDR2	FW3	CDR3	FW4
AB-001	EVQLLESGGGLVQPGGSLRLSCAAS	GFTFSSYAMS	WVRQAPGKGLEWVS	YISDDGSLKYADSVKG	RFTISRDNKNTLYLQMNSLRAEDTAVYYCAK	HPYWYGGQLDL	WGQGTLVTVSS
4QCI	---V-----	-----	-----	-----	-----R	-----	-----
R1-002	---Q-----	-----	-----	-----	-----R	-----	-----
R1-003	-----K-----	-----	-----	-----	-----R	-----	-----
R1-004	-----R--	-----	-----	-----	-----R	-----	-----
R1-005	-----	-----	-----	--N-----	-----R	-----	-----
R1-006	-----	-----	-----	-----	-----R	-----	--R-----
R1-007	---Q---K---	-----	-----	-----	-----R	-----	-----
R1-008	-----	-----	-----	-----	-----	-----	-----
R1-009	-----	-----	-----	-----	-----	-----	-----
R1-010	-----	-----	-----	-----	-----	-----	-----
R1-011	-----	-----	-----	-----	-----	-----	-----
R1-012	---Q---K---	-----	-----	-----	-----R	-----	-----
R1-013	---Q-----	-----	-----	-----	-----R	-----	-----
R1-014	---Q---K---	-----	-----	-----	-----R	-----	-----
R1-015	---Q---K---	-----	-----	-----	-----R	-----	-----
R1-016	---Q---K---	-----	-----	-----	-----R	-----	-----
R1-017	-----K-----	-----	-----	-----	-----R	-----	-----
R1-018	-----	-----	-----	-----	-----R	-----	--R-----
R2-001	---Q---K---	-----	-----	-----	-----R	-----	-----
R2-002	---Q---K---	-----	-----	-----	-----R	-----	-----
R2-003	---Q---K---	-----	-----	-----	-----R	-----	-----
R2-004	---Q---K---	-----	-----	-----	-----R	-----	-----
R2-005	---Q---K---	-----	-----	-----	-----R	-----	-----
R2-006	---Q---K---	-----	-----	-----	-----R	-----	-----

R2-007	---Q---K-----	-----	-----	-----	-----	-----R	-----	-----
R2-008	---Q---K-----	-----	-----	-----	-----	-----R	-----	-----
R2-009	---Q---K-----	-----	-----	-----	-----	-----R	-----	-----
R2-010	---Q---K-----	-----	-----	-----	-----	-----R	---H---	-----
R2-011	---Q---K-----	-----	-----	-----	-----	-----R	---K---	-----
R2-012	---Q---K-----	-----	-----	---K---	-----	-----R	-----	-----
R2-013	---Q---K-----	-----	-----	---N---	-----	-----R	-----	-----
R2-014	---Q---K-----	-----	-----	---Q---	-----	-----R	-----	-----
R2-015	---Q---K-----	-----	-----	---N---	-----	-----R	-----	-----
R2-016	---Q---K-----	-----	-----	-----	-----	-----R	---N---	-----
R2-017	---Q---K-----	-----	-----	-----	-----	-----R	---Y---	-----
R2-018	---Q---K-----	-----	-----	-----	-----	-----R	-----	-----
R2-019	---Q---K-----	-----	-----	-----	-----	-----R	-----	-----
R2-020	---Q---K-----	-----	-----	-----	-----	-----R	-----	-----
R2-021	---Q---K-----	-----	-----	-----	-----	-----R	-----	-----
R2-022	---Q---K-----	-----	-----	-----	-----	-----R	-----	-----

(b)

Name	FW1	CDR1	FW2	CDR2	FW3	CDR3	FW4
AB-001	SYELTQPPSVSVSPGQTASITC	SGDSLGSYFVH	WYQQKPGQSPVLVIY	DDSNRPS	GIPERFSGSNGNTATLTISGTQAMDEADYYC	SAFTHNSDV	FGGTKLTVL
4QCI	-----A-----R-S-	-----	-----A-----	-----	-----E-----	-----	-----
R1-002	-----	-----	-----	-----	-----	-----	-----
R1-003	-----	-----	-----	-----	-----	-----	-----
R1-004	-----	-----	-----	-----	-----	-----	-----
R1-005	-----	-----	-----	-----	-----	-----	-----
R1-006	-----	-----	-----	-----	-----	-----	-----
R1-007	-----	-----	-----	-----	-----	-----	-----
R1-008	--V-----	-----	-----	-----	-----	-----	-----
R1-009	-----R-----	-----	-----	-----	-----	-----	-----
R1-010	-----	-----	-----	---K---	-----	-----	-----
R1-011	-----	-----	-----	-----	-----	---N---	-----
R1-012	--V-----	-----	-----	-----	-----	-----	-----
R1-013	-----R-----	-----	-----	-----	-----	-----	-----
R1-014	-----R-----	-----	-----	-----	-----	-----	-----
R1-015	-----	-----	-----	-----	-----	---N---	-----
R1-016	-----	-----	-----	---K---	-----	-----	-----
R1-017	-----R-----	-----	-----	-----	-----	-----	-----
R1-018	-----R-----	-----	-----	-----	-----	-----	-----
R2-001	--V-----A--K--R--	-----	-----	---K---	-----	-----	-----
R2-002	--V-----A--K--R--	---K---	-----	---K---	-----	-----	-----
R2-003	--V-----A--K--R--	-----K	-----	---K---	-----	-----	-----
R2-004	--V-----A--K--R--	-----	-----H	---K---	-----	-----	-----

R2-005	--V-----A--K--R--	-----	-----R	--K--	-----	-----	-----
R2-006	--V-----A--K--R--	-----	-----	--KK--	-----	-----	-----
R2-007	--V-----A--K--R--	-----	-----	--K--	-----K-----	-----	-----
R2-008	--V-----A--K--R--	-----	-----	--K--	-----K-----	-----	-----
R2-009	--V-----A--K--R--	-----	-----	--K--	-----K-----	-----	-----
R2-010	--V-----A--K--R--	-----	-----	--K--	-----	-----	-----
R2-011	--V-----A--K--R--	-----	-----	--K--	-----	-----	-----
R2-012	--V-----A--K--R--	-----	-----	--K--	-----	-----	-----
R2-013	--V-----A--K--R--	-----	-----	--K--	-----	-----	-----
R2-014	--V-----A--K--R--	-----	-----	--K--	-----	-----	-----
R2-015	--V-----A--K--R--	-----	-----	--K--	-----	-----	-----
R2-016	--V-----A--K--R--	-----	-----	--K--	-----	-----	-----
R2-017	--V-----A--K--R--	-----	-----	--K--	-----	-----	-----
R2-018	--V-----A--K--R--	--N-----	-----	--K--	-----	-----	-----
R2-019	--V-----A--K--R--	-----	-----	I--K--	-----	-----	-----
R2-020	--V-----A--K--R--	-----	-----	--N--K--	-----	-----	-----
R2-021	--V-----A--K--R--	-----	-----	--K--	-----	-----K--	-----
R2-022	--V-----A--K--R--	-----	-----	--K--	-----	-----N--	-----

## BIBLIOGRAPHY

- APGAR, J. R., TAM, A. S. P., SORM, R., MOESTA, S., KING, A. C., YANG, H., KELLEHER, K., MURPHY, D., D'ANTONA, A. M., YANG, GUOYING., ZHONG, X., RODRIGUEZ, L., MA, W., FERGUSON, D. E., CARVEN, G. J., BENNETT, E. M. & LIN, L. 2020. Modeling and mitigation of high-concentration antibody viscosity through structure-based computer-aided protein design. *PLoS ONE*, 5, 15, e0232713.
- BOSWELL, C. A., TESAR, D. B., MUKHYALA, K., THEIL, F.-P., FIELDER, P. J. & KHAWLI, L. A. Effects of charge on antibody tissue distribution and pharmacokinetics. *Bioconjug Chem*, 2010. 21. 12. 2153-2163.
- BRADY, J. L., HARRISON, L. C., GOODMAN, D. J., COWAN, P. J., HAWTHORNE, W. J., O'CONNELL, P. J., SUTHERLAND, R. M. & LEW, A. M. 2014. Preclinical screening for acute toxicity of therapeutic monoclonal antibodies in a hu-SCID model. *Clin Trans Immunol*, 3. e29.
- BRENNAN, F. R., MORTON, L. D., SPINDELREHER, S., KIESSLING, A., ALLENSPACH, R., HEY, A., MULLER, P. Y., FRINGS, W. & SIMS, J. Safety and immunotoxicity assessment of immunomodulatory monoclonal antibodies. 2010. *MAbs*, 2. 233-255.
- BÜLOW, S. von., SIGGEL, M., LINKE, M. & HUMMER, G. 2019. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *PNAS*, 116. 20. 9843-9852.
- BUMBACA, D., BOSWELL, C. A., FIELDER, P. J. & KHAWLI, L. A. 2012. Physicochemical and biochemical factors influencing the pharmacokinetics of antibody therapeutics. *AAPS J*, 14. 3. 554-558.
- CHAUDHRI, A., ZARRAGA, I. E., YADAV, S., PATAPOFF, T. W., SHIRE, S. J. & VOTH, G. A. 2020. The role of amino acid sequence in the self-association of therapeutic monoclonal antibodies: insights from coarse-grained modeling. *J Phys Chem B*, 117, 1269-1279.
- CHEN, R., LI, L. & WENG, Z. 2003. ZDOCK: an initial-stage protein-docking algorithm. *Proteins*, 52. 1. 80-87.
- CLOUTIER, T., SUDRIK, C., MODY, N., SATHISH, H. A. & TROUT, B. L. 2019. Molecular Computations of preferential interaction coefficients of IgG1 monoclonal antibodies with sorbitol, sucrose, and trehalose and the impact of these excipients on aggregation and viscosity. *Mol Pharmaceutics*, 16, 3657-3664.
- GABB, H. A., JACKSON, R. M. & STERNBERG, M. J. 1997. Modelling protein docking using shape complementary, electrostatics and biochemical information. *J Mol Biol*, 272. 1. 106-120.
- HANAOR, D., MICHELAZZI, M., LEONELLI, C. & SORRELL, C. C. 2012. The effect of carboxylic acids on the aqueous dispersion and electrophoretic deposition of ZrO<sub>2</sub>. *Journal of the European Ceramic Society*, 32. 1. 235-244.
- HEO, L., LEE, H. & SEOK, C. 2016. GalaxyRefineComplex: refinement of protein-protein complex model structures driven by interface repacking. *Sci Rep*, 18. 6. 32153.

- JEZEK, J., RIDES, M., DERHAM, B., MOORE, J., CERASOLI, E., SIMLER, R. & PEREZ-RAMIREZ, B. 2011. Viscosity of concentrated therapeutic protein compositions. *Adv Drug Deliv Rev*, 63. 13. 1107-1117.
- JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ŽÍDEK, A., POTAPENKO, A., BRIDGLAND, A., MEYER, C., KOHL, S. A. A., BALLARD, A. J., COWIE, A., ROMERA-PAREDES, B., NIKOLOV, S., JAIN, R., ADLER, J., BACK, T., PETERSEN, S., REIMAN, D., CLANCY, E., ZIELINSKI, M., STEINEGGER, M., PACHOLSKA, M., BERGHAMMER T., BODENSTEIN, S., SILVER, D., VINYALS, O., SENIOR, A. W., KAVUKCUOGLU, K., KOHLI, P. & HASSABIS, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596. 583-589.
- KUAI, J., MOSYAK, L., BROOKS, J., CAIN, M., CARVEN, G. J., OGAWA, S., ISHINO, T., TAM, M., LAVALLIE, E. R., YANG, Z., PONSEL, D., RAUCHENBERGER, R., ARCH, R. & PULLEN, N. 2015. *Biochemistry*, 54. 10. 1918-1929.
- KURODA, D. & TSUMOTO, K. 2020. Engineering stability, viscosity, and immunogenicity of antibodies by computational design. *Journal of Pharmaceutical Sciences*, 109, 1631-1651.
- LAI, P. K., SWAN, J. W. & TROUT, B. L. 2021. Calculation of therapeutic antibody viscosity with coarse-grained models, hydrodynamic calculations and machine learning-based parameters. *mAbs*, 13, 1, 1907882.
- LEE, J., CHONG, S. H. & HAM, S. 2020. Local environment effects on charged mutations for developing aggregation-resistant monoclonal antibodies. *Scientific Reports*, 10, 21191.
- LI, L., KUMAR, S., BUCK, P. M., BURNS, C., LAVOIE, J., SINGH, S. K., WAME, N. W., NICHOLS, P., LUKSHA, N. & BOARDMAN, D. 2014. Concentration dependent viscosity of monoclonal antibody solutions: explaining experimental behavior in terms of molecular properties. *Pharm Res*, 31, 3161-3178.
- LIANG, S., MEROUEH, S. O., WANG, G., QIU, C. & ZHOU, Y. 2009. Consensus scoring for enriching near-native structures from protein-protein docking decoys. *Proteins*, 75. 2. 397-403.
- MARSDEN, C. J., ECKERSLEY, S., HEBDITCH, M., KVIST, A. J., MILNER, R., MITCHELL, D., WARWICKER, J. & MARLEY A. E. 2014. The use of antibodies in small-molecule drug discovery. *Journal of Biomolecular Screening*, 19, 6, 829-838.
- MINTSERIS, J., PIERCE, B., WIEHE, K., ANDERSON, R., CHEN, R. & WENG, Z. 2007. Integrating statistical pair potentials into protein complex prediction. *Proteins*, 69. 3. 511-520.
- PARK, T., BAEK, M., LEE, H. & SEOK, C. 2019. GalaxyTongDock: symmetric and asymmetric protein-protein docking web server with improved energy parameters. *J Comput Chem*, 40. 27. 2413-2417.
- SHIRE, S. J., SHAHROKH, Z. & LIU, J. 2004. Challenges in the development of high protein concentration formulations. *Journal of Pharmaceutical*

*Sciences*, 93. 6. 1390-1402.

- SHARMA, V. K., PATAOFF, T. W., KABAKOFF, B., PAI, S., HILARIO, E., ZHANG, B., LI, C., BORISOV, O., KELLEY, R. F., CHORNY, I., ZHOU, J. Z., DILL, K. A., SWARTZ, T. E. 2014. In silico selection of therapeutic antibodies for development: Viscosity, clearance, and chemical stability. *PNAS*, 111, 18601-18606.
- TOMAR, D. S., LI, L., BROULIDAKIS, M. P., LUKSHA, N. G., BURNS, C. T., SINGH, S. K. & KUMAR, S. 2017. In-silico prediction of concentration-dependent viscosity curves for monoclonal antibody solutions. *mAbs*, 9, 3, 476-489.
- TOMAR, D. S., KUMAR, S., SINGH, S. K., GOSWAMI, S. & LI, L. 2016. Molecular basis of high viscosity in concentrated antibody solutions: Strategies for high concentration drug product development. *mAbs*, 8. 2. 216-228.
- ZHANG, C., VASMATZIS, G., CORNETTE, J. L. & DELISI, C. 1997. Determination of atomic desolvation energies from the structures of crystallized proteins. *J Mol Biol*, 267. 3. 707-726.

## 국문초록

의약품 용도로 쓰이는 항체가 주목을 받고 있다. 하지만 항체를 실용적으로 사용하기에는 물성에 관한 여러가지 조건을 만족할 필요가 있다. 그 중 하나가 점성 (viscosity) 다. 일반적으로 점성은 농도가 높을 수록 커지는 걸로 알려져 있다. 보통 의약품을 주사로 놓을 때 높은 농도로 조절할 필요가 있지만, 약품의 점성이 높으면 주사를 놓기에 큰 힘이 필요하거나 그것으로 인해 환자에게 큰 아픔을 줄 수가 있다. 하지만 점성을 실험적으로 예측할 방법에는 많은 돈과 시간이 들기 때문에 실험 없이 계산적인 방법을 이용한 점성 예측의 필요성이 높아지고 있다. 이 논문에서는 단백질-단백질 도킹을 이용해서 점성과 항체의 서열의 관계성을 설명한다.

**주요어:** 항체, 점성, 단백질-단백질 도킹

**학 번:** 2020-27795