In-silico discovery and experimental verification of excipients for biologics

Lok Hin Lui

Thesis submitted to University College London for

the degree of Doctor of Philosophy

School of Pharmacy

University College London

September 2018

Declarations

I, Lok Hin Lui confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

First and foremost I would like to acknowledge my three supervisors Prof Ajoy Velayudhan, Dr Christopher van der Walle and Prof Steve Brocchini for introducing me to the worlds of computational modelling and formulation science. I appreciate their ongoing support and commitment throughout the course of this research.

I would like to thank other fellow students and the staff members at both University College London and the University of Nottingham of the Centre for Doctoral Training in Advanced Therapeutics and Nanomedicines for giving me this unique opportunity to study in this unique programme as well as the funding and support from the Engineering and Physical Sciences Research Council (EPSRC), the EPSRC Centre for Doctoral Training in Advanced Therapeutics and Nanomedicines, University College London, the University of Nottingham and MedImmune (EP/L01646X). Without their financial support, the achievements detailed in this report would not have been possible.

I would like to extend my gratitude to all staff members and fellow PhD students at UCL School of Pharmacy for their help and inspiring conversations over the years.

I owe special gratitude to UCL High Performance Computing Facilities (Grace@UCL and Legion@UCL) and associated support services. Without their input, most of the work described in this thesis would not have been carried out.

Lastly, and definitely not least, I would like to thank my parents, Jane and Edmond for their continuous encouragement and invaluable support through all the ups and downs of throughout this journey.

Table of Contents

Abstract

Protein-based pharmaceuticals such as monoclonal antibodies are the fastest growing class of therapeutic agent. As with all protein therapeutics, antibody aggregation must be avoided during production, storage and use. With recent advances in computing power, it is becoming feasible to simulate protein-protein interactions *in-silico*. Combining computational and experimental studies may offer a platform solution to design specific in-process stabilisers and excipients to accelerate the development of aggregation-resistant formulations.

An antibody Fv fragment was first evaluated to understand the early stages of aggregate formation by identifying aggregation-prone regions. Three-dimensional structural information and protein-protein docking were used to identify exposed hydrophobic patches. Virtual screening was used to identify compounds that bind to the exposed hydrophobic patches as a means to prevent Fv-Fv interactions that could result in aggregation. An excipient with the highest calculated binding affinity was found to prevent Fv-Fv interactions as determined with the diffusion interaction parameter (k_D) using DLS.

Excipient performance was then evaluated using coarse-grained molecular dynamics (MD) simulations with MARTINI force field to provide a more in-depth view on Fv fragment dimer complex formation. Simulation results were further evaluated with free-energy calculations but these free-energy calculations were found to produce highly variable and therefore unreliable results. This coarse-grained MD approach was also used to virtually screen a library of dipeptides to identify peptide excipients. The results revealed a positive correlation between the calculated mean interaction energies and the diffusion interaction parameter measured with DLS. Use of the MD approach was further extended to accommodate challenging an antibody without published structural data through homology modelling and to suggest possible excipients to prevent high-affinity antibody-antibody interactions. Therefore this MD approach could potentially be used as a first step for the selection of excipients for antibodies.

Impact Statement

Aggregation is considered as one of the most common issues that significantly affects the process of manufacturing, quality and efficacy of therapeutic antibodies. This EPSRC funded CDT PhD project hoped to gain understanding of the early stages of aggregation, by identifying aggregation-prone sequences and using that information to design in-process stabilisers to control formation of aggregates without sacrificing the biological function of the therapeutic protein at atomic level. In particular to evaluate the possibility of using a combination of *in-silico* techniques to identify aggregation-prone regions on therapeutic protein and use such information for screening of potential excipients those prevent aggregation. This project also evaluated the possibility of using molecular dynamics simulation for screening therapeutic protein formulations and recommends next-generation excipients with huge potentials.

The results and methods described in this thesis is particularly useful for the rational design of future bioprocesses and the formulations of protein-based products. Ultimately, these novel in-process stabilisers could potentially improve future products by reducing immunogenicity, improving shelf-life of the product, improving patient access to novel therapies, reducing time to market, improving the batch-to-batch reproducibility, reducing batch rejection rates, and lowering the cost of manufacturing, storage and transportation.

Presentations and Awards

2016 CDT colloquium, Pfizer, Sandwich, United Kingdom (Oral Presentation)

2016 PhD Research Day, UCL School of Pharmacy, United Kingdom (Poster Presentation)

2016 St. Margaret's Co-educational English Secondary and Primary School, Hong Kong (Oral Presentation)

2017 CDT colloquium, UCL, London, United Kingdom (Oral Presentation)

2017 ULLA Summer School, KU Leuven, Belgium (Poster Presentation)

2018 PhD Research Day, UCL School of Pharmacy, United Kingdom (Oral Presentation)
Winner: Originality of Presentation Style

2018 CDT colloquium, UCL, London, United Kingdom (Oral Presentation)

List of Abbreviations

ADC	Antibody Drug Conjugates
bsAb	Bispecific Antibody
BAR	Bennett Acceptance Ratio
CG	Coarse-grained
CD	Cluster of Differentiation
CDR	Complementarity Determining Region
CHARMM	Chemistry at Harvard Molecular Mechanics
CHO	Chinese Hamster Ovary
DAFT	Docking Assay For Transmembrane Components
DHRF	Dihydrofolate Reductase
DLS	Dynamic Light Scattering
Fab	Fragment Antigen-Binding
Fc	Fragment Crystallisable Region
FcR	Fc Receptor
FcRn	Neonatal Fc Receptor
Fv	Variable Fragment
FDA	Food and Drug Administration
G	Gibbs Free Energy
GROMACS	Groningen Machine for Chemical Simulations
H	Enthalpy
Ig	Immunoglobulin
IM	Intramuscular
INN	International Non-proprietary Name
IV	Intravenous
k_D	Diffusion Interaction Parameter
mAb	Monoclonal Antibody
MD	Molecular Dynamics
MM-GBSA	Molecular Mechanics Generalised Born Surface Area
MM-PBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PLGA	Poly(lactic-co-glycolic) Acid
RMSD	Root Mean Square Deviation

RESID	Residue ID
RMSF	Root Mean Square Fluctuation
S	Entropy
SAP	Spatial Aggregation Propensity
SASA	Solvent Accessible Surface Area
SC	Subcutaneous
scFv	Single-chain Variable Fragment
sdAb	Single-domain Antibody
SEC	Size-exclusion Chromatography
SEM	Standard Error of the Mean
T	Temperature
T_g	Glass Transition Temperature

Amino acids

One-letter code	Three-letter code	Amino acid
A	**ALA**	Alanine
C	**CYS**	Cysteine
D	**ASP**	Aspartic acid
E	**GLU**	Glutamic acid
F	**PHE**	Phenylalanine
G	**GLY**	Glycine
H	**HIS (HSP)**	Histidine (Protonated Histidine)
I	**ILE**	Isoleucine
K	**LYS**	Lysine
L	**LEU**	Leucine
M	**MET**	Methionine
N	**ASN**	Asparagine
P	**PRO**	Proline
Q	**GLN**	Glutamine
R	**ARG**	Arginine
S	**SER**	Serine
T	**THR**	Threonine
V	**VAL**	Valine
W	**TRP**	Tryptophan
Y	**TYR**	Tyrosine

List of Figures

List of Tables

1. Introduction

1.1. Monoclonal Antibodies

Biological therapeutics include proteins, peptides, nucleic acids and vaccines represent a large and growing class of therapeutic agents with eleven new therapeutic biologics under Biologics License Applications out of forty-six novel new drugs approved by the United States Food and Drug Administration (FDA) in 2017 with ten of those approved biologics are monoclonal antibodies-based products [1]. Monoclonal antibodies (mAb) and their related biologics are currently used in clinic to treat a wide-range of diseases including cancers, autoimmune diseases and cardiovascular disorders [2].

The field of therapeutic monoclonal antibodies has devoted decades of effort to develop both the understandings of the mechanism of actions of antibodies and reliable processes to manufacture monoclonal antibodies for therapeutic use. Antibodies, also referred to as immunoglobulins, are soluble glycoproteins naturally produced in response to immunogenic foreign molecules and these are termed as antigens. The epitope of an antibody can be defined as the minimal structural determinant that it binds to the target antigen.

The studies of antibodies began in 1890 when Emil von Behring and Shibasaburo Kitasato published their findings of serum component of animals immune to diphtheria or tetanus could transfer immunity [3]. In 1939, A Tiselius and EA Kabat revealed that there is a substantial increase in the gamma fraction of electrophoresed immunised rabbits serum, suggested globular proteins in the gamma fraction was responsible for immunity. The field of therapeutic monoclonal antibodies research was further extended by two key discoveries in 1970s. In 1973, Cohen and co-workers have reported that recombinant

DNA technology can be achieved by isolating bacterial plasmid, slicing it with restriction endonuclease EcoRI, inserting the desired DNA and then reforming the plasmid with DNA ligase so more variations of proteins can be made in large quantities [5]. In 1975, Köhler and Milstein published their revolutionary work on successful production of monoclonal antibodies with defined specificity by fusing of myeloma cells and B-lymphocytes derived from the spleen of immunised mice against a specific molecular target [6].

Since then, monoclonal antibodies have become a valuable tool in biomedical research. In 1986, the first mAb, Muromonab-CD3 (Brand name: Orthoclone OKT3), was approved to use in humans indicated for transplantation rejection [7]. The first generation mAbs were primarily produced by murine hybridomas, and were fully murine sequence antibodies. As a result, they generate human anti-mouse antibodies [8], reducing half-life and effectiveness of the product or even eliciting an anaphylactic hypersensitivity response [9]. This has led to the development of chimeric and humanised mAbs which aimed to reduce immunogenicity, improve half-life and efficacy, allowing for the application in different clinical conditions. Antibody technology continue to advanced and resulted in the explosive development of therapeutic mAbs over the last 20 years which has changed the way we manage a number of clinical conditions. There are a number of processing and formulation challenges for mAbs, such as aggregation, denaturation and chemical modifications, excipients are used to improve stability of therapeutic mAb preparations (Table 1).

Year of FDA Approval	INN	Trade Name	Species	Format	Target	Excipients
1992	muromonab-CD3	ORTHOCLONE OKT3	Mouse	IgG2a	CD3	Monobasic sodium phosphate, Dibasic sodium phosphate, Sodium chloride, Polysorbate 80
1993	abciximab	REOPRO	Chimeric	Fab	Glycoprotein IIb/IIIa	Disodium phosphate, Sodium dihydrogen phosphate, Sodium chloride, Polysorbate 80
1997	daclizumab	ZENAPAX	Humanised	IgG1	IL-2RA	Disodium phosphate, Sodium dihydrogen phosphate, Sodium chloride, Hydrochloric acid, Sodium hydroxide, Polysorbate 80
1997	rituximab	RITUXAN	Chimeric	IgG1	CD20	Sodium chloride, Sodium citrate, Citric acid, Polysorbate 80
1998	basiliximab	SIMULECT	Chimeric	IgG1	IL-2RA	Monobasic potassium phosphate, Disodium hydrogen phosphate, Sodium chloride, Sucrose, Mannitol, Glycine
1998	etanercept	ENBREL	Human	Fc fusion	TNF-alpha	Mannitol, Sucrose, Tromethamine
1998	infliximab	REMICADE	Chimeric	IgG1	TNF-alpha	Monobasic sodium phosphate, Dibasic sodium phosphate, Sucrose, Polysorbate 80
1998	palivizumab	SYNAGIS	Humanised	IgG1	Respiratory syncytial virus	Histidine, Glycine
1998	trastuzumab	HERCEPTIN	Humanised	IgG1	HER-2	L-histidineHCl, L-histidine, α,α-trehalose,Polysorbate 20
2000	gemtuzumab	MYLOTARG	Humanised	IgG4	CD33	Monobasic sodium phosphate, Dibasic sodium phosphate, Dextran 40, Sucrose, Sodium chloride
2001	alemtuzumab	CAMPATH	Humanised	IgG1	CD52	Sodium chloride, Dibasic sodium phosphate, Potassium chloride, Monobasic potassium phosphate, Disodium edetate, Polysorbate 80
2002	adalimumab	HUMIRA	Human	IgG1	TNF-alpha	Citric acid monohydrate, Sodium citrate, Monobasic sodium phosphate, Dibasic sodium phosphate, Mannitol, Monobasic sodium phosphate, Sodium chloride, Sodium hydroxide, Polysorbate 80
2002	ibritumomab tiuxetan	ZEVALIN	Mouse	IgG2	CD20	Sodium chloride
2003	alefacept	AMEVIVE	Human	Fc fusion	CD2	Sucrose, Glycine, Sodium citrate dihydrate, Citric acid
2003	efalizumab	RAPTIVA	Humanised	IgG1	CD11a	Sucrose, L-histidine hydrochloride, L-histidine, Polysorbate 20
2003	omalizumab	XOLAIR	Humanised	IgG1	IgE	Sucrose, L-histidine hydrochloride, L-histidine, Polysorbate 20
2003	tositumomab	BEXXAR	Mouse	IgG2a	CD20	Maltose, Sodium chloride, 1 mg/mL phosphate, Potassium hydroxide

Year of FDA Approval	INN	Trade Name	Species	Format	Target	Excipients
2004	bevacizumab	AVASTIN	Humanised	IgG1	VEGF-A	Monobasic sodium phosphate, Dibasic sodium phosphate, α,α-trehalose, Polysorbate 20
2004	cetuximab	ERBITUX	Chimeric	IgG1	EGFR	Monobasic sodium phosphate, Dibasic sodium phosphate, Sodium chloride
2004	natalizumab	TYSABRI	Humanised	IgG4	CD49d	Monobasic sodium phosphate, Dibasic sodium phosphate, Sodium chloride, Polysorbate 80
2005	abatacept	ORENCI	Human	Fc fusion	CD80/CD86	Monobasic sodium phosphate , Sodium chloride, Maltose
2008	tocilizumab	ACTEMRA	Humanised	IgG1	IL-6	Monobasic sodium phosphate, Dibasic sodium phosphate, Sucrose, Polysorbate 80
2006	panitumumab	VECTIBIX	Human	IgG2	EGFR	Sodium chloride, Sodium acetate
2006	ranibizumab	LUCENTIS	Humanised	Fab	VEGF-A	L-histidine, α,α-trehalose, Polysorbate 20
2007	eculizumab	SOLIRIS	Humanised	IgG2/4	Complement protein C5	Monobasic sodium phosphate, Dibasic sodium phosphate, Sodium chloride, Polysorbate 80
2008	certolizumab pegol	CIMZIA	Humanised	Fab	TNF-alpha	PEG, Sodium acetate, Sodium chloride
2008	rilonacept	ARCALYST	Human	Fc fusion	IL-1B	Histidine, Arginine, Polyethylene glycol 3350, Sucrose, Glycine
2008	romiplostim	NPLATE	Human	Fc fusion	CD110	L-histidine, Hydrochloric acid, Mannitol, Sucrose, Polysorbate 20
2009	canakinumab	ILARIS	Human	IgG1	IL-1B	L-histidine hydrochloride, L-histidine, Sucrose, Polysorbate 80
2009	golimumab	SIMPONI	Human	IgG1	TNF-alpha	L-histidine hydrochloride, L-histidine, Sorbitol, Polysorbate 80
2009	ofatumumab	ARZERRA	Human	IgG1	CD20	Hydrochloric acid, Edetate disodium, Arginine, Sodium acetate, Sodium chloride, Polysorbate 80
2009	ustekinumab	STELARA	Human	IgG1	IL-12B	L-histidine, Sucrose, Polysorbate 80
2010	denosumab	PROLIA	Human	IgG2	RANKL	Sorbitol, Acetate, Sodium hydroxide, Polysorbate 20
2011	aflibercept	EYLEA	Human	Fc fusion	VEGF-A	Sodium phospha, Sodium chloride, Sucrose, Polysorbate 20
2011	belatacept	NULOJIX	Human	Fc fusion	CD80/CD86	Monobasic sodium phosphate , Sodium chloride, Sucrose
2011	belimumab	BENLYSTA	Human	IgG1	BAFF	Citric acid, Sodium citrate, Sucrose, Polysorbate 80
2011	brentuximab vedotin	ADCETRIS	Chimeric	IgG1	CD30	Citric acid, Sodium citrate, Trehalose, Polysorbate 80

Year of FDA Approval	INN	Trade Name	Species	Format	Target	Excipients
2011	ipilimumab	YERVOY	Human	IgG1	CTLA-4	Diethylene triamine pentaacetic acid (DTPA), Sodium chloride, Tris hydrochloride, Mannitol, Polysorbate 80
2012	pertuzumab	PERJETA	Humanised	IgG1	HER-2	L-histidine acetate, Sucrose, Polysorbate 20
2012	raxibacumab	ABTHRAX	Human	IgG1	Anthrax protective antigen	Sodium citrate, Glycine, Sucrose, Polysorbate 80
2013	obinutuzumab	GAZYVA	Humanised	IgG1	CD20	L-histidine hydrochloride, L-histidine, Trehalose, Poloxamer 188
2013	trastuzumab emtansine	KADCYLA	Humanised	IgG1-ADC	HER-2	Sodium succinate, Sucrose, Polysorbate 20
2014	blinatumomab	BLINCYTO	Mouse	Bispecific	CD3/CD19	Citric acid, lysine hydrochloride, Trehalose, Sodium hydroxide, Polysorbate 80
2014	dulaglutide	TRULICITY	Human	Fc fusion	GLP1R	citric acid, trisodium citrate, mannitol, Polysorbate 80
2014	efmoroctocog alfa	ELOCTATE	Human	Fc fusion	Replacement of factor VIII	L-histidine, Sodium chloride, Calcium chloride, Sucrose, Polysorbate 20
2014	eftrenonacog alfa	ALPROLIX	Human	Fc fusion	Replacement of factor IX	L-histidine, Sodium chloride, Mannitol, Sucrose, Polysorbate 20
2014	nivolumab	OPDIVO	Human	IgG4	PD-1	Sodium chloride, Sodium citrate, Pentetic acid, Sodium Hydroxide, Hydrochloric acid, Mannitol, Polysorbate 80
2014	pembrolizumab	KEYTRUDA	Humanised	IgG4	PD-1	L-histidine, Sucrose, Polysorbate 80
2014	ramucirumab	CYRAMZA	Human	IgG1	VEGFR2	Histidine, Histidine hydrochloride, Sodium chloride, Glycine, Polysorbate 80
2014	vedolizumab	ENTYVIO	Humanised	IgG1	Integrin α4β7	L-histidine, L-histidine hydrochloride, L-arginine hydrochloride, Sucrose, Polysorbate 80
2015	alirocumab	PRALUENT	Human	IgG1	PCSK9	Histidine, Sucrose, Polysorbate 20
2015	asfotase alfa	STRENSIQ	Human	Fc fusion	Replacement of TNSAPL	Monobasic sodium phosphate, Dibasic sodium phosphate, Sodium chloride
2015	daratumumab	DARZALEX	Human	IgG1	CD38	Acetic acid, Sodium acetate, Sodium chloride, Mannitol, Polysorbate 20
2015	dinutuximab	UNITUXIN	Chimeric	IgG1	GD2	Histidine, Sodium chloride, Hydrochloric acid, Polysorbate 20
2015	elotuzumab	EMPLICITI	Humanised	IgG1	SLAMF7	Citric acid, Sodium citrate, Sucrose, Polysorbate 80
2015	evolocumab	REPATHA	Human	IgG2	PCSK9	Proline, Acetate, Polysorbate 80

Year of FDA Approval	INN	Trade Name	Species	Format	Target	Excipients
2015	mepolizumab	NUCALA	Human	IgG1	IL-5	Dibasic sodium phosphate, Sucrose, Polysorbate 80
2015	necitumumab	PORTRAZZA	Human	IgG1	EGFR	Glycine, Citric acid, Sodium citrate, Sodium chloride, Mannitol, Polysorbate 80
2015	secukinumab	COSENTYX	Human	IgG1	IL-17A	L-histidine, L-histidine hydrochloride, Sucrose, Polysorbate 80
2016	atezolizumab	TECENTRIQ	Humanised	IgG1	PD-L1	Acetic acid, L- histidine, Sucrose, Polysorbate 20
2016	bezlotoxumab	ZINPLAVA	Human	IgG1	C. difficile toxin B	Citric acid, Sodium citrate, Diethylenetriaminepentaacetic acid, Sodium chloride, Polysorbate 80
2016	ixekizumab	TALTZ	Humanised	IgG4	IL-17A	Citric acid, Sodium citrate, Sodium chloride, Polysorbate 80
2016	olaratumab	LARTRUVO	Human	IgG1	PDGFR-α	Glycine, L-histidine, L-histidine hydrochloride, Sodium chloride, Mannitol, Polysorbate 20
2016	reslizumab	CINQAIR	Human	IgG1	IL-5	Acetic acid, Sodium acetate, Sucrose
2017	avelumab	BAVENCIO	Human	IgG1	PD-L1	D-mannitol, Acetic acid, Polysorbate 20, Sodium hydroxide
2017	benralizumab	FASENRA	Humanised	IgG1	IL-5RA	L-histidine, Lhistidine hydrochloride monohydrate, Polysorbate 20, α,α-trehalose dihydrate
2017	brodalumab	SILIQ	Human	IgG2	IL-17RA	Glutamate, Polysorbate 20, Proline
2017	dupilumab	DUPIXENT	Human	IgG4	IL-4R	L-arginine hydrochloride, Lhistidine , Polysorbate 80, Sodium acetate, Sucrose
2017	durvalumab	IMFINZI	Human	IgG1	PD-L1	L-histidine, L-histidine hydrochloride monohydrate, α,α-trehalose dihydrate, Polysorbate 80
2017	emicizumab	HEMLIBRA	Humanised	Bispecific - IgG4	Coagulant factors IXa- and X-directed	L-arginine, L-histidine, Poloxamer 188, L-Aspartic acid
2017	guselkumab	TREMFYA	Human	IgG1	IL-23A	Lhistidine, L-histidine monohydrochloride monohydrate, Polysorbate 80, Sucrose
2017	ocrelizumab	OCREVUS	Humanised	IgG1	MS4A1	Acetic acid, Polysorbate 20, Sodium acetate trihydrate, Trehalose dihydrate
2017	sarilumab	KEVZARA	Human	IgG1	IL-6R	Arginine, Histidine, Polysorbate 20, Sucrose

Table 1: List of FDA-approved therapeutic antibody and related biologics by year with their INN name, trade name, antibody species, antibody format, therapeutic target and the list of excipients.

1.1.1. Antibody Structure

Immunoglobulin is a major constituent of the human immune system which recognises and binds to antigens via the epitope. The quaternary structure of protein is held by both covalent bonds and non-covalent bonds such as hydrogen bonds, hydrophobic interactions, electrostatic and van der Waals forces. Immunoglobulins are soluble glycoproteins that are mainly composed of Y-shaped monomer containing variable regions at the N-terminus and constant region at the C-terminus (Figure 1). There are two identical heavy chains each containing four domains (approximately 450 residues) and two identical light chains containing two domains (approximately 220 residues) connected via multiple disulphide bonds [10].

Immunoglobulin are classified into five isotypes of immunoglobulins based on their constant regions: IgG, IgM, IgA, IgD and IgE, designated as γ, μ, α, δ and ε respectively. Most immunoglobulins are present as monomers with a monocular weight of approximately 150 kDa, except IgA and IgE which present as dimers and pentamers respectively. The IgG subclass is the most abundant of the immunoglobulins in human serum comprising 75% of total serum immunoglobulins [11]. IgGs are further divided into four subclasses: IgG1, IgG2, IgG3, and IgG4 [12]. The IgG1 subclass is the most abundant immunoglobulins found in human serum and being used in the clinic. With the exception of subclass IgG3 which has an elimination half-life of approximately 7 days, the elimination half-lives in all other IgG subclasses are much longer, around 20 – 21 days [13]. The heavy chains are connected in the hinge region by variable number of disulphide bonds. The structure are different among IgG subtypes with different heavy chains and the number and the location of disulphide bonds (IgG1: 2, IgG2: 4, IgG3: 11, IgG4: 2). Immunoglobulins are glycosylated proteins with various number and location of glycosylation site between isotypes.

A human IgG molecule contains two N-linked biantennary oligosaccharide chains each attached to ASN residue in the CH2 domain of heavy chain. These oligosaccharide chains are responsible for biological functions including antibody-dependent cellular cytotoxicity [14-15], pharmacokinetic properties, aggregation propensity and physicochemical stability [16-18]. Light chains are classified into two subtypes: lambda (λ) and kappa (κ). Light chains are connected to the heavy chain via a disulphide bond between the last cysteine residue in the light chain at the C terminus and either the fifth cysteine residue of the heavy chain of IgG1 or the third cysteine residue of the heavy chain of IgG2, IgG3 and IgG4.

In IgG1, approximately the first 110 residues in both heavy and light chains are considered as the variable region (Fv) forming the antigen-binding fragment (Fab) and the remaining sequence is considered as the constant regions. The variable regions on the heavy chain and light chain contain three complementarity determining regions (CDRs). The fragment crystallisable (Fc) region corresponds to the paired CH2 and CH3 domains in the heavy chains and is the part of the antibody molecule that interacts with Fc receptors (FcRs) which are expressed in a number of immune cells regulating various immunological responses [19]. Generally, the CH3 domain is the most stable domain and the CH2 domain is the least stable within an antibody [20].

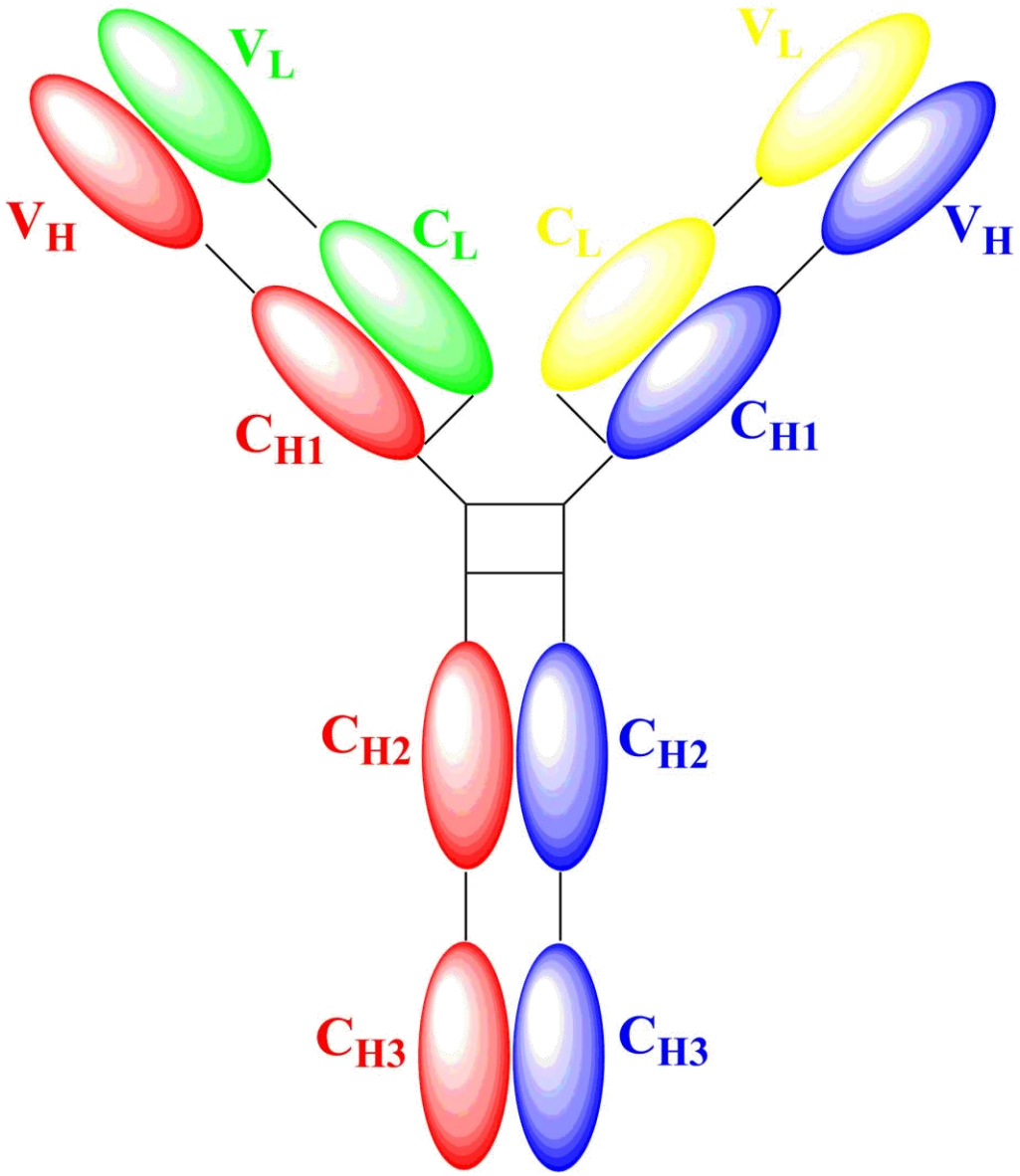


Figure 1: The basic structure of an antibody IgG1. Heavy chains are coloured in red and blue while light chains are coloured in green and yellow. IgG1 can be divided into the Fab fragment (VL and CL light chain and the VH and CH1 heavy chain domains) and the Fc portion (CH2 and CH3 heavy chain domains). Two heavy chains are connected by two disulfide bonds in the hinge region.

1.1.2. Antibody Technologies

Despite the success of monoclonal antibodies treating various disease conditions, the use of therapeutic IgGs are limited by a range of issues such as pharmacokinetic properties, limited ability to target tissues and slow production rates. Newly developed antibody-engineering technologies provided the opportunities to alter antibody characteristics to overcome these challenges to meet different clinical needs. The CDR of an antibody

determines the specificity and the affinity for antigens while the rest of the molecule do not interact with antigens directly. It is generally believed that a protein with a molecular weight of 150 kDa would be too large to penetrate cellular membranes [21].

IgG fragments such as Fab and smaller fragments such as single chain variable fragments (scFv) and single-domain antibody (sdAb) can reduce the size of the molecule without losing specificity and the affinity to the antigens. The elimination of the Fc region results in a large reduction in sizes. scFv and sdAb further reduce the size by elimination of the constant region of the Fab fragment. These fragments have a higher diffusion rates, better cell penetrations, quicker clearance from the systemic circulation and reduced immunogenicity [22].

The long serum half-life of IgGs mainly owing to the large molecular weight above the renal threshold and the pH dependent binding of the neonatal Fc receptor (FcRn). FcRn was originally discovered as the receptor that transports maternal IgG in mother's milk across the neonatal gut. FcRns are expressed primarily in the endothelial cells and form a specific recycling system which regulates IgG concentration in serum [23]. Fc fusion proteins have been a successful approach to prolong half-lives of many biological therapeutics where the Fc fragment have been modified and linked to protein that binds to its target.

Typically, each antibody molecule has at least two identical antigen-binding sites binding to the same target so they are considered as monospecific. Bispecific antibodies (bsAbs) are capable of binding two antigens simultaneously and it was first generated by Nisonoff and Rivers by fusing two different Fab fragments derived from two different antibody preparations by chemical oxidation [24]. bsAbs can be engineered with or without the Fc

fragment with bsAbs bearing a functional Fc fragment also termed as trifunctional antibody. With different antibody-engineering technologies, there are currently over 60 alternative bsAb formats have been described [25]. As the immunologic effector functions of antibodies have been enhanced mediated through interaction with multiple receptors simultaneously, bsAbs have shown cytotoxic effects those have not seen in the corresponding antibody [26].

Cytotoxic agents are broadly used to treat malignancies and solid tumours. These agents target rapidly proliferating cells by inhibiting interfering cellular processes including nucleic acids synthesis, DNA repair, microtubule function and cellular metabolism. However, cytotoxic agents often display little selectivity to target cells and substantial toxicity to normal tissues. As a result, these drugs are often associated with serious side-effects and toxicities such as bone marrow depression. By attaching cytotoxic agents to mAbs, antibody drug conjugates (ADC) delivers the attached cytotoxic agents to antigen-expressing tumour cells more selectively resulting in a large improved progression survival and a more favourable safety profile [27].

1.2. Administration Routes

The vast majority of antibody therapies are administered via injectable routes such as intravenous (IV), subcutaneous (SC) and intramuscular (IM) route. These administration routes offer clear advantage of better accessibility to the target tissues avoiding the need of diffusion across of biological barriers where the large molecular size, charge and hydrophilic nature of the antibody prohibit them diffusing across biological barriers. The earlier generation of antibody products are mainly deliver IV route but this administration route is unfavourable among healthcare professionals. IV administration require dedicated infusion facilities, a specific trained personnel giving the treatment, patient-

adjusted dosage calculation, aseptic preparation of infusion volumes and monitoring during and after the infusing leading to hospitalisation and additional costs to the healthcare system [28]. In addition, a single-use fixed-dosed injection device allowing self-administration is more favourable by patients because it reduces the need of hospital visits permitting flexibility in making appointments [29]. In a view of offering a treatment that is less invasive and easier administration would be more convenient to both healthcare professionals and patient, newer antibody therapies also utilise SC and IM routes for drug delivery.

The SC route is of growing interest for the administration of mAbs. Following SC administration a drug must diffuse out of the interstitial matrix and reaching traverse capillaries or lymphatic vessels depending on the molecular weight of the drug [30]. Molecules with molecular weights less than or equal to 16 kDa can reach the systemic circulation through blood capillaries. For molecules with molecular weights larger than the threshold such as mAbs are primarily absorbed by the lymphatic system [31]. The extracellular matrix of the subcutaneous tissue consists of structural macromolecules such as collagen and elastin consisting of hyaluronan and other glycosaminoglycans and polysaccharides forming a large structural network and control the diffusion of molecules. This network results in small interstitial spaces for fluids limiting the injection volumes, usually 1-2 mL with SC administration [30]. Hyaluronan has an estimated half-life of 15–20 hours and it is broken down and replenished as part of normal tissue homeostasis. Therefore, it is considered an ideal target for transiently disruption of extracellular matrix. Recombinant human hyaluronidase (rHuPH20) breaks down hyaluronic acid reversibly in the interstitial tissue which can be used to temporally increase the interstitial spaces for dispersion of a larger injected fluid. Similarly, IM injections deliver molecules below the subcutaneous space, molecules are being absorbed under similar principles and typically

injection volumes for IM injections are limit to 5 mL [32]. Nevertheless, for most mAb therapeutics, a large dose (i.e 8 mg/kg) often required for the treatment to reach therapeutic effects, successful SC/IM formulations have to be in high concentration due to the relatively low limit on the injection volume.

Oral is the most widely use delivery route for many therapeutic agents. As discussed earlier, FcRn facilitate the transfer of IgG from the mother to the fetus across the placenta and the proximal small intestine so IgG from ingested breast milk can be absorbed into the systemic circulations. As the gastrointestinal tract becoming more mature, the hostile proteolytic environments and extreme pH conditions adversely affect the stability of antibodies in the gastrointestinal tract. In addition, systematic uptake of antibodies via oral administration through FcRn binding alone is not sufficient for achieving therapeutic levels [33]. The large molecular size of antibodies and poor permeation also prohibits them being absorbed passively. Formulation strategies such as enzyme inhibitors, antibody engineering, polymeric coatings and nano-particulate systems have been explored but there is not yet a successful oral formulation for antibodies [33-36].

Lungs are useful targets for pulmonary route of administration for both local delivery treating respiratory disorders and systemic delivery. Pulmonary delivery of therapeutic agents can be achieved by aerosol or dry powder formulations. Particle size is of particular important in pulmonary administration as it determines the location within the lungs where it deposits in lungs. For optimal deposition into the alveoli, particle sizes between 1 and 5 μm are considered ideal for deposition. Local delivery via the pulmonary route is advantageous for respiratory conditions as the therapeutic agent can be delivered to the site directly allowing faster onset of action, minimising the risk of systemic side-effects. Large surface area with rapid absorption due to the high vascularisation and avoidance of

hepatic first pass effects are particularly useful for systemic delivery of therapeutic agents. FcRn is also expressed on the pulmonary epithelial cells in lungs and it is believed that IgG can be absorbed to systemic circulation via FcRn-mediated pathway [37]. Despite the advantages that the pulmonary route offers, the absorption of antibodies is potentially limited by the respiratory mucus, mucociliary clearance, pulmonary enzymes and macrophages [38]. The only approved inhaled protein product is an insulin dried-power formulation [39]. Nevertheless, with better design of the inhalation device and performance enhancements through formulation strategies, pulmonary delivery of antibodies remains a promising non-invasive route of delivery.

1.3. Antibody Processing and Formulation Challenges

1.3.1. Antibody Processing

Monoclonal antibodies were originally produced in murine hybridoma cells. Over the years, various production technologies have demonstrated their potential for producing antibodies in microbial [40] or plant cells [41] but mammalian cells are still the most dominant production system for the majority of full-length antibodies [42]. A number of factors have to be taken into account for the selection of the host cells, including suitability for the protein intended to be produced, glycosylation pattern, safety, cell turnover rate, protein yield, downstream purifications, safety, cost of production and maintenance. Recombinant Chinese Hamster Ovary (CHO) cells and Mouse Myeloma (NSO) cells are the most common and well-studied mammalian cell lines for antibody production [42]. A variety of transfection methods have been developed to introduce target genes into mammalian cells, such as calcium phosphate, electroporation, cationic lipid-based lipofection, and polymer or dendrimer-based methods [43]. Gene amplification is necessary because generally with greater number of insertion per cell will increase the host cell's efficiency in producing the protein of interest. DHFR-deficient

CHO cell lines are often used in combination with methotrexate. Methotrexate is a DHFR inhibitor and inhibits the conversion of dihydrofolate to tetrahydrofolate which is necessary for the de novo biosynthesis of purines, pyrimidines and glycine. Only those DHFR-deficient CHO cells that have been transfected with the DHFR gene can develop resistance to methotrexate through amplification of the DHFR gene in a growth medium lacking glycine, hypoxanthine and thymidine. In this way, the gene for the protein of interest are also co-amplified [44]. Transfected cells are then cultivated in cell culture medium and screened for robust cell growth and high productivity. As the insertion of genes are randomly integrated into the host cell, leading to variability in cell growth, productivity and stability. Isolation and selection of rare high yielding and stable clone with high growth rate is important for large scale production. Traditionally, this can be done by dispensing diluting transfected cells onto microplates at a concentration of less than one cell per well. Cell growth and productivity are then analysed to identify potential clones.

Cell culture media are essential to support the biological processes for the production of antibodies. Media usually contain amino acids, carbohydrates, vitamins, trace elements, inorganic ion, salts, lipids and growth factors [45] in ratios that are optimised for antibody titre and stability. It has been shown that the addition of some short lysine-peptides into the medium can lead to increase in the production of monoclonal antibody production [46]. Bovine serum and animal-derived raw materials are generally avoided due to risks associated with transmissible spongiform encephalopathy (TSE) and other contaminants [43].

The use of mammalian cells can be potentially contaminated by both endogenous retroviruses and adventitious viral contaminants introduced during manufacturing. In

addition, downstream processing is usually not considered to be a sterile operation and filtration is commonly employed with 0.2 μm filters in many processes to reduce bioburden and to remove particulates [47]. The use of filters provide a size-based clearance mechanism so viruses above the size of the pore are not be able to pass. After fermentation, antibodies are harvested by centrifugation or microfiltration, resulting in a harvested cell culture that contain various proteins and other cellular debris in addition to antibodies. Therefore the harvested cell culture has to undergo a series of filtration and column chromatography steps to recover the antibodies in cell culture and to meet the purity, safety and regulatory requirements of the product [48].

Various column chromatographic methods are available to purify antibodies based on their biochemical properties such as charge, molecular size, solubility, hydrophobicity and binding affinity, and the FDA demands two different steps to be included in the process. Among all these column chromatographic methods, affinity chromatography is the most specific and widely employed as the first process in antibodies purification processes. The separation is based on specific binding between the antibody and a ligand which is covalently coupled to the chromatography matrix allowing selective adsorption of the target antibody while other contents are removed [49].

Bacterially derived proteins, such as Staphylococcal protein A and Streptococcal protein G are commonly used as ligands in affinity chromatography for antibody isolation. Both Staphylococcal protein A and Streptococcal protein G bind to the Fc fragment of the IgG. Staphylococcal protein A binds to most human IgG subclasses except human IgG3 and mouse IgG1 while Streptococcal protein G binds to all four subclasses of human IgG [50]. The retained antibody is then eluted from the column at a very low pH value that inactivates viruses, reaching a purity of about 90% [51]. Small amounts of ligands co-

eluted with the antibody from the resin and other impurities may be present in the eluate which must be removed to reach the purity required in the final product. Most antibody production processes contain at least one ion exchange chromatography step for further purification. Ion-exchange chromatography separates molecules based on the differences between the overall charge of the proteins, allowing the reduction of charge variants, high molecular weight aggregates, leached ligands, residual nucleic acids, residual host cell protein and viral particles.

1.3.2. Antibody Formulation

Antibodies are proteins and hence subject to a number of physical and chemical degradation pathways which affects the delicate three-dimensional structure and may result in biological inactivation. The primary goal of antibody formulation development is to provide a pharmaceutical composition that will support the stability of an antibody during final stages of its production, storage, transportation and deliver consistently the intended performance in a dosage form that ensures safety and efficacy.

1.3.2.1. Liquid Formulation

All antibody therapeutics are formulated in either liquid dosage form or lyophilised form. The liquid dosage form is usually more favourable as there is no requirement for reconstitution before administration. As discussed earlier, for the ease of self-administration, a single monthly SC injection is more favourable among patients. However, for some therapeutic antibodies, the dosage requirement may be as high as 1400 mg/dose [52]. Due to the volume restriction for some administration routes, the concentration of the therapeutic antibody in the final formulation has to be as high as 120 mg/mL. In particular, for some less soluble mAbs, achieving such high concentration may exhibit significant stability and formulation issues, including solubility, viscosity and

aggregation [53]. Cleland and co-workers have demonstrated that the increase of the concentration of trastuzumab from 25 mg/mL to 50 mg/mL resulted in significant increase in aggregates formation after 1 month storage at 40 ºC [54]. Therefore a careful selection of pH, storage temperature, excipients and their concentration is essential for a successful formulation.

1.3.2.2. Dry Formulation

Not all mAbs are stable in a liquid preparation and very often the degradation rate is quicker in solution, reducing its potential shelf-life. For all product development projects, the time to market is crucial, there may not be a sufficient time for the development of a stable liquid formulation. Therefore, lyophilisation or spray drying can be used as an alternative technique to maintain a both chemically and physically stable formulation for improved stability. By using a smaller volume of diluent for reconstitution, lyophilisation can also achieve high antibody concentration for clinical applications. A traditional lyophilisation process consists of three steps: freezing, primary drying, and secondary drying. During the freezing step, the liquid formulation is cooled until pure crystalline ice forms from the liquid and ultimately the liquid formulation solidifies. The ice formed during freezing is removed by sublimation under vacuum in an ambient temperature below the characteristic collapse temperature for amorphous solutes or eutectic melting temperature for crystalline solutes [55]. Upon completion of primary drying, there are small amounts of unfrozen water remains in the matrix which is removed by desorption in elevated temperature and vacuum during the secondary drying step, achieving a dry moisture cake with the desired residual moisture [56].The aim for lyophilisation is to achieve a stable formulation but the lyophilisation process itself may degrade the mAb by various stresses including initial solute concentration, low temperature, ice formation and dehydration, resulting in denaturation and loss of biological activity [57].

There is currently an increasing interest in the use of spray drying to formulate proteins and mAbs, especially for pulmonary administration in a dried powder form. Spray drying has an advantage of producing free flowing powders with narrow particle size distribution [58]. In a typical spray drying process, the formulated liquid is first separated into small droplets by a nozzle, creating a large surface area which is exposed to a hot drying gas. This facilitates the rapid evaporation of solvent and the dried product is collected. The size of the nozzle is an important parameters as it determines the size of the droplets and ultimately the characteristics of the dried particles. For spray dryers currently used for pharmaceutical applications, typical droplet diameters can range from less than 10 μm for pulmonary applications to 100 μm, which generates typical dry particles with diameter between 0.5 μm and 50 μm [59]. Similarly, spray drying also cauce various stresses during the process including high temperature, mechanical stress and adsorption at the air–water interface during atomisation [60].

1.3.3. Chemical and Physical Instabilities

Despite the therapeutic advantages that mAbs and their related biologics offer in various disease conditions, the stability of protein-based biologics poses a major challenge to the biopharmaceutical industry. Antibodies, like any other protein therapeutics, are subject to a variety of degradation pathways and can be classified into two general classes: chemical degradation and physical degradation. These degradation pathways may be interrelated and happen simultaneously. The mechanism and impact of these degradation pathways will be discussed in this session. The instabilities of antibodies cause significant challenges to the biopharmaceutical industry as the degradation can happen during processing and storage of the product.

1.3.3.1. Chemical Degradation

1.3.3.1.1. Deamidation

Deamidation is one of the major chemical degradation mechanisms in therapeutic mAb products, affecting the quality of the product. OKT-3, the first marketed monoclonal antibody product, undergoes significant deamidation upon storage [61]. Non-enzymatic hydrolysis of the amide side-chain of amino acids ASN results in conversion of the amide group into a carboxylic acid group [62-63]. Under neutral to basic conditions, deamidation of ASN involves nucleophilic attack by the adjacent amide nitrogen onto the side chain carbonyl carbon atom forming a cyclic imide and ammonia (NH_3) [63]. Ammonia formed in solution is lost as a gas, hence this reaction is usually irreversible. The formation of the cyclic imide is the rate-limiting step of the reaction. The cyclic imide will subsequently be hydrolysed to form the ASP and isoASP products in a ratio of 1:3 [64]. In general, ASN residues are more prone to deamidation with the sequence ASN–GLY particularly prone to deamidation [65]. The 3D structure of protein can influence the deamidation rate and there are multiple sites on antibody those are prone to deamidation located on both the Fab and Fc regions. It has been identified that both chains in the Fab region of trastuzumab are susceptible to deamidation, ASN55 on the heavy chain and ASN30 on the light chain [66]. ASN55 is a highly conserved residue in IgG1 and in case of trastuzumab, theoretically labile to deamidation due to the sequence ASN55-GLY56 [66]. Interestingly, ASN30 on the light chain has a local sequence of ASN30-THR31 and it only undergo deamidation to ASP during cell culture production, suggesting the deamidation to ASP does not proceed via cyclic imide intermediate. Alternative mechanism such as nucleophilic attack may be involved [66]. Fc region contains multiple ASN residues and they are also subjected to deamidation. In case of OKT-3, ASN386-GLY387 has been demonstrated that more than 90% of this site is deamidated after 2 months storage at 37 ºC [61].

1.3.3.1.2. Isomerisation

ASN deamidation and ASP isomerisation are closely related as both reactions go through the same cyclic imide intermediate to form isoASP and it is difficult to control. Similarly, the rate-limiting step in ASP isomerisation is the formation of the cyclic imide intermediate. The first step in isomerisation involves nucleophilic attack by the adjacent amide nitrogen onto the side chain carbonyl carbon atom forming a cyclic imide and subsequently converted hydrolysed into ASP or isoASP in a ratio of 1:3. Local sequence is an important factor determining the rate of isomerisation with ASP-GLY exhibit a faster rate of isomerisation [67]. Structural factors also influence the rate of ASP isomerisation with highly conformational flexible and solvent-accessible ASP residue more prone to isomerisation [68]. In contrast to deamidation, ASP isomerisation favours slightly acidic conditions with pH around 5.0 to 6.0 [69]. There are several reports in the literature that ASP isomerization in the CDR regions can lead to significant reduction of biological activity. Cacia and co-workers have demonstrated that the isomerisation of an ASP32 on the CDR1 of one of the light chain of omalizumab will reduce activity by 58% and 85% of the activity if both light chains are affected [70]. ASP isomerisation has also been reported at ASP102 located on the CDR3 of the heavy chain on trastuzumab and it is estimated that only 12–30% relative antiproliferative activity remains [66].

1.3.3.1.3. Oxidation

Oxidation is commonly detected in the solvent-accessible amino acid residues CYS, HIS, MET, PHE, TRP and TYR which is due to the susceptibility of aromatic and sulphur-containing moieties towards various reactive oxygen species [71]. Methionine is particularly vulnerable to oxidation and the reaction product, methionine sulfoxide, will affect the protein conformation and stability by changing the size, shape and flexibility of methionine. In addition, methionine sulfoxide is more hydrophilic than methionine and

increased hydrogen bonding capability [61, 72]. Oxidation can occur during protein production and storage [73]. Oxidation can also arise from incompatibilities with impurities in the formulation such as PEG caused by low levels of peroxides and aldehydes found in PEG as impurities [74]. In human IgG1, there are two highly conserved MET residues in the heavy chain of the Fc fragment at MET252 and MET428, located at the end of CH2 domain and MET428 in the CH3 domain respectively [75]. MET252 and MET428 are located at the interface of the CH2-CH3 domains, where FcRn binds. Hence, oxidation of MET252 and MET428 will reduce the binding affinities of the Fc to the FcRn leading to reduction of serum half-lives [73]. Wang and co-workers have demonstrated in transgenic mice with human FcRn that increasing oxidation levels of trastuzumab will significantly reduce the serum half-live with complete oxidation of MET252 and MET428 decrease to only 17% of the serum half-live compared to the control [73]. Antioxidants such as methionine, histidine and ethylenediaminetetraacetic acid (EDTA), are commonly added to the formulation to prevent such oxidative reactions.

1.3.3.1.4. Disulphide Exchange

Generally, an IgG molecule contains at least 32 CYS residues forming at least 4 interchain and 12 intrachain disulphide bonds. These disulphide bonds are important in maintenance of the structural integrity and biological activities of antibodies. As discussed earlier, free solvent exposed CYS residues can be oxidised easily. An ionised CYS thiolate can undergo nucleophilic attack on a disulphide bond on another disulphide linkage, breaking the original disulphide bond and form a new disulphide bond with a free thiolate ion, resulting in structural distortion. The free thiolate ion generated can further propagation of this process and causing aggregation [62]. In general, the rate of disulphide exchange depends on the level of ionisation of the nucleophilic thiol group [76]. Disulphide exchange can arise from storage of antibodies and disulphide linked covalent aggregates

have been detected after one year storage of omalizumab in a spray-dried formulation [77].

1.3.3.2. Physical Degradation

1.3.3.2.1. Denaturation

Protein denaturation refers to partial or complete unfolding of the tertiary and secondary structure of a protein from its native state and results in deformation of the delicate three-dimensional conformation, leading to loss of binding affinity and its biological activity. Local unfolding of an aggregation-prone region allow it to be solvent exposed and available to associate with another molecule, and subsequently accelerating the aggregation process [78]. Proteins are only marginally stable in their folded state with multiple local stable intermediates between folded states [79]. A variety of stress conditions such as changes in pH, temperature, solvent environment, salt concentrations and shear force, can facilitate denaturation. Heat is generally the most prominent factor for denaturation but mAbs are relatively stable at high temperatures with melting temperatures around 70 ºC [80]. Low pH environment and high salt concentrations can lead to destabilisation of the antibody [81]. Proteins often are only stable in water, the presence of organic solvent, even at low concentrations, can cause denaturation [82]. A variety of pumps are used in the production and purification processes of antibodies and these introduce different physical stresses on the antibody as the solution is being pumped, such as shear, agitation and cavitation which may cause denaturation.

1.3.3.2.2. Adsorption

Proteins can be adsorbed to different surfaces so some free proteins may be lost to the surfaces where positively charged proteins tend to adsorb to a negatively charged glass surface and neutral proteins tend to adsorb to hydrophobic surfaces. Adsorption generally

does not cause much problems, except for products to be given in low concentration and require patient-specific dose titration, the additional preparation processes may significantly reduce the amount of proteins in the final dosage form. In addition, it has been reported that adsorption of antibody at the solid-liquid interface can mediate the aggregation process, posing a potential risk factor for long term storage [83].

1.3.3.2.3. Fragmentation

Fragmentation is a general term to describe cleavage of covalent bonds in the protein backbone that results in fragmentation of the protein structure which can happen at any residue in the protein structure. Peptide bonds exhibits exceptional chemical stability with non-enzymatic hydrolysis of peptide bonds unlikely to occur naturally unless extreme pH and high temperatures are used to accelerate the process. Most fragmentation in mAbs is commonly observed in one of the following residues: ASP, GLY, SER, THR, CYS or ASN [84]. ASP fragmentation is a common degradation pathways of mAbs under acidic conditions with hydrolysis occurs more readily for ASP-GLY and ASP-PRO bonds [84]. In addition to peptide sequence, flexibility of the local protein structure also play an important role in fragmentation. Fragmentation is commonly observed at the hinge region in mAbs, largely due to the hinge region is both very flexible and solvent exposed. Fragmentation has been detected during antibody production and purification processes and it can also occur during storage [84].

1.3.3.2.4. Aggregation

Aggregation is considered to be a dominant pathway of degradation, resulting in biological inactivation [85-86]. Aggregates can also induce immunogenic responses and result in self-immunity to the therapeutic protein causing loss of therapeutic effects [87].

The delicate structure of protein is held by both covalent bonds and non-bonded interactions such as hydrogen bonds, hydrophobic interactions, electrostatic and van der Waals forces. In most cases, the same forces also aid protein monomers to assemble into stable complexes [88-90]. Although many proteins require association for their biological functions, aggregation in protein-based products only refer to the undesired multimeric forms of the therapeutic protein [90].

A protein with a given sequence can be either in its native folded state or in the unfolded states. A protein, regardless of whether it is folded or not can form non-covalent contacts with another protein or proteins forming dimers or oligomers. Early precursors of aggregation, such as soluble dimers and oligomers, do not always affect the folding of the therapeutic protein, and are often reversible through dilutions which could be beneficial to product stability [88-90]. Nucleation is a process where aggregates grow into a stable irreversible species and the proteins within the 'nuclei' have undergone structural deformation to from a stable complex. Nuclei can continue to grow into higher order aggregates such as soluble filaments, soluble agglomerated aggregates and macroscopic aggregates [88-89].

It is difficult to predict the rates of aggregation that can occur for a protein, as a wide range of factors are known to accelerate aggregate formation such as protein concentration, pressure, lipophilicity, pH, ionic strength and temperature. Commercial biologic products must ensure safety and efficacy to meet regulatory requirements [88-90]. In industrial practice, aggregation is often controlled by a combination of precise control of the process, using anti-aggregation agents, buffers and surfactants, and optimising storage conditions. Traditional protein formulations use various excipients, such as amino acids, glucose, trehalose, dextran, sorbitol and polysorbate, to modify the

environment around the protein [89-90]. However, each mAb is different in nature with different aggregation-prone regions which these generic excipients may not able to offer sufficient protection at these regions.

Aggregation is not a problem limited to the formulated product. Instead, it can occur throughout the manufacturing process, from cell culturing to final filling of the product [90]. Aggregate levels in the manufacturing process are monitored and must stay within their product specification [90]. The purified products are highly pure but irreversible aggregates will eventually form as the product is transported, stored, dispensed and administered to patients.

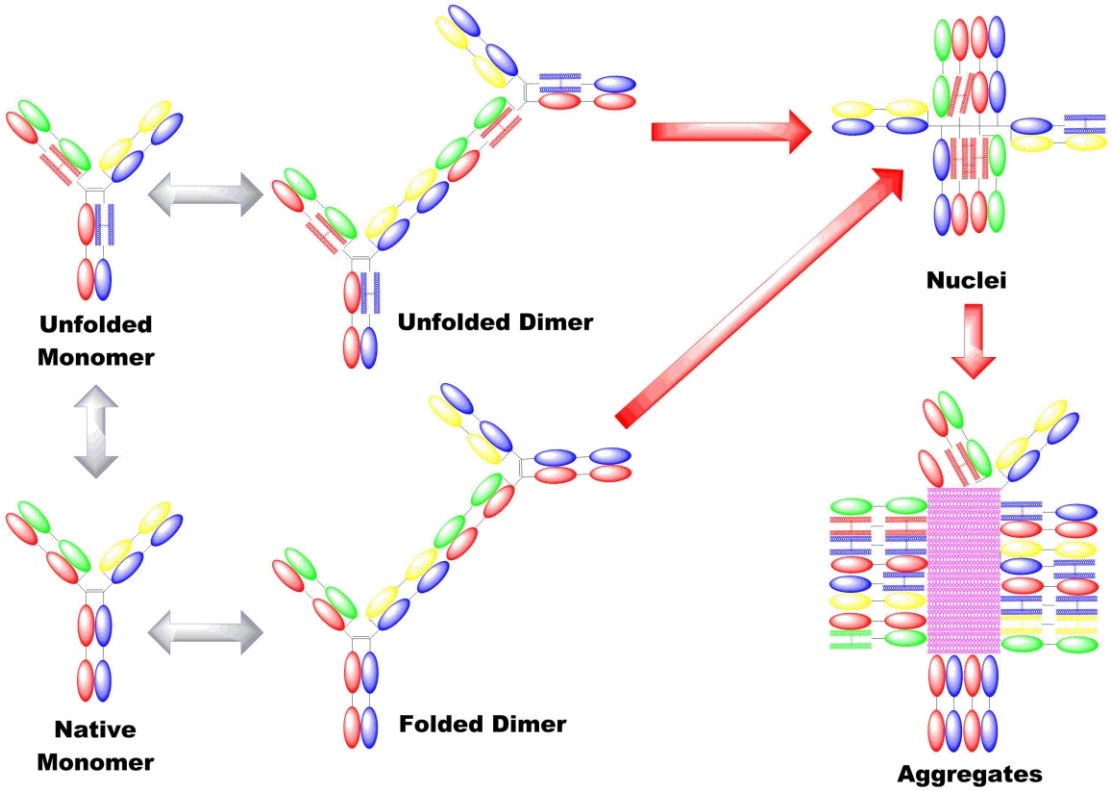


Figure 2: Schematic diagram of aggregates formation. Antibody is flexible in nature so it can partially unfolded and expose aggregation-prone regions. Whether antibody is misfolded or not, antibody can interact with adjacent antibody molecule to form dimers and these processes are usually reversible. However it can also progress into a nuclei, the net irreversible species, by association with more antibody molecules and eventually grow into an aggregate.

1.3.4. Excipients and Stabilisers

Excipients and stabilisers are commonly used in pharmaceutical formulations and during bioprocessing to stabilise mAbs and to prevent them from degradation. The choice and composition of these excipients and stabilisers are crucial for the performance of the process or the quality of the product. Each antibody has unique chemical and physical characteristics so the selection and the composition of these excipients and stabilisers have to be considered individually.

1.3.4.1. Buffer

Buffering agents have been routinely used in protein pharmaceuticals during bioprocessing and in formulation for pH control as pH affects the structural integrity of the protein and may play important roles in various degradation pathways. As each degradation pathway exhibits different degradation rates at different pH values, the optimal pH is based on the overall effects of these pathways. An acidic environment usually promotes fragmentation, hydrolysis and isomerisation whereas a basic environment can promote aggregation, deamidation and disulphide exchange [91]. Hence, the pH requirement for best antibody stability is different for each antibody. In general, most antibody therapeutics are formulated in slightly acidic conditions. Buffering agents have to be compatible with the antibody and other excipients so they do not accelerate the degradation process which may adversely affect the quality of the product. Salinas and co-workers have demonstrated that fragmentation of antibody occurs more readily in phosphate buffer compared to histidine buffer [92]. Choice of buffering agent and its concentration generally depends on their buffer capacity. Common buffering agent include acetate, citrate, histidine, phosphate and succinate. Histidine has a pKa of approximately 6, which is commonly used as a buffer for a slightly acidic environment (i.e. pH 5.5 to 6.5) to stay away from the isoelectric point of antibody (i.e. pI 8.5). It has

been shown to enhance the stability of an IgG2 by reducing viscosity of the formulation and inhibits aggregates formation [93]. However, in another case, formulation in histidine buffer resulted in more aggregates formation at pH 5 after 8 weeks storage at 40 ºC compared with acetate buffer [94]. Therefore, therefore each antibody should be considered individually as the buffer that works on one antibody may not work on another.

1.3.4.2. Carbohydrates

Carbohydrates as cryoprotectants and lyoprotectants are widely used to stabilise antibodies in lyophilised formulations. The major chemical interaction associated with sugar is the Maillard reaction where the carbonyl group of a reducing sugar can react with nucleophilic amine groups in a protein to form a glycosylated protein, especially at high temperatures. It is frequently observed in lysine side chains but also observed in other basic residues such as arginine, asparagine and glutamine [95-96]. Reducing sugars such as glucose, lactose, fructose and maltose are typically avoided by formulation scientists due to this effect. In addition, some non-reducing sugars such as sucrose can be hydrolysed into its reducing constituent monosaccharides glucose and fructose and this has been observed in IgG1 formulation with sucrose at elevated temperatures [97]. Trehalose is a non-reducing disaccharide with two glucose units which is commonly used in antibody formulations. Trehalose has three possible anomers but only α,α-trehalose have been isolated from living organisms [98]. The glycosidic bond in α,α-trehalose is less likely to be hydrolysed than sucrose [99]. It has been reported that the rates of glycation of sucrose can be 2000-fold greater than trehalose [99]. The molar ratio of carbohydrates to antibody is an important consideration in formulation design and it has been shown that trehalose exhibits a concentration dependent anti-aggregation behaviour in a lyophilised trastuzumab formulation after long term storage. Cleland and co-workers

suggested that a minimum carbohydrates to antibody molar ratio of 360:1 in a lyophilised trastuzumab formulation was required to inhibit any significant aggregation and deamidation [54].

1.3.4.3. Bulking Agents

In lyophilised products, bulking agent is usually crystallised in formulation to increase product mass and provide mechanical rigidity for cake formation. In addition to prevent cake collapse, it may also be used to adjust tonicity for injection. Common bulking agents include glycine, alanine and mannitol. The presence of bulking agents, choice of bulking agents and their concentrations may affect the stability of the antibody [100]. Meyer and co-workers have demonstrated that a greater storage stability in murine IgG formulations with amorphous glycine formulated with sucrose than sucrose alone [100].

1.3.4.4. Surfactant

Therapeutic antibody formulations encounter various interfaces (air–liquid, solid–liquid, and liquid–liquid) during processing and storage. As discussed earlier, antibodies are generally surface active and can be adsorbed to many interfaces which can lead to loss of antibody and aggregation. A common approach to prevent interfacial damage in the pharmaceutical industry is to use non-ionic surfactants such as polysorbate. Addition of polysorbate have been shown to protect antibodies against aggregation [101-102]. It has been proposed that polysorbate acts by binding to hydrophobic areas and decreases intermolecular contacts [103].

1.3.4.5. Amino Acids

Apart from using amino acids for bulking, buffer and as antioxidants, small amino acids are widely used as stabilisers in protein formulations. The use of amino acids in

biopharmaceuticals are largely because of being classified as endogenous compounds and generally considered as safe to use in pharmaceutical applications. Histidine have been found to reduce solution viscosity and reduce aggregates formation in both lyophilised and liquid formulations of human IgG2 [93, 104]. Arginine has been known for preventing aggregation, increasing solubility, reducing surface tension and reducing viscosity [105-106]. Interestingly arginine also improves the recovery of IgGs with a study shown the recovery rates significant better than citrate and the recovery rates almost doubled with 2 M arginine at pH 4.4 compared with 0.5 M arginine at pH 4.3 [107]. In another study, amino acids arginine, aspartic acid, histidine and glycine have been studied their effects in stabilising human antibodies. Among these amino acids, arginine and histidine were found to be the most effective excipient in preventing secondary structural changes in antibodies whereas glycine was the least effective to prevent secondary structure changes upon lyophilisation [108].

1.3.4.6. Controlled Release Formulations

Controlled release formulations offer many potential advantages over conventional dosage forms. They provide a more predictable drug release profile over a desired time period, reducing fluctuations of drug levels, reducing dosing intervals, reducing occurrences of adverse effects, improving patient compliance and reducing stress associated with the treatment. One promising approach to develop sustained release formulations is to biocompatible and biodegradable polymers with antibodies encapsulated within the polymer.

1.3.4.6.1. Pegylation

In general, the binding of the Fc portion of the antibody molecule to the FcRn exhibit generally long circulation half-lives that is suitable for once monthly administration but

it is not the case for mAb fragments such as Fab which is cleared more rapidly so prolonging the circulation time of an antibody fragment can become necessary to reduce the frequency of administration [109]. Abuchowsky and co-workers have demonstrated in 1977 the non-specific covalent conjugation of a protein with a poly(ethylene glycol) (PEG) [110] results in a PEG-protein conjugate with an increased molecular mass and hydrodynamic radius that increases its serum half-life while also reducing the immunogenicity of the protein [111]. This technique is also known as PEGylation. The non-specific conjugation is usually through LYS residues which can lead to substantial loss of binding activity [112]. To achieve site-specific conjugation, an additional free CYS residue is attached in the engineered hinge region and conjugated with PEG through the thiol group of the CYS residue. Site-specific conjugation have been shown to maintain biological activities and increase half-live to around 2 week [113].

1.3.4.6.2. Polymeric Microspheres

Poly(lactic-co-glycolic acid) (PLGA) is a one of the most successfully used biocompatible and biodegradable polymer composed of lactic acid and glycolic acid with wide therapeutic applications. Mordenti and co-workers have demonstrated the encapsulation of trastuzumab in PLGA microspheres but it failed to show any pharmacokinetic advantages over the liquid formulation with only 32.4% of the encapsulated trastuzumab detected *in-vivo* [114]. Choice of fabrication method can have a great influence on antibody stabilities during the fabrication process, product storage and *in-vivo* release. Wang and co-workers have demonstrated using Fourier-transformation Infrared (FTIR) that the fabrication of IgG-loaded PLGA microspheres can adversely affected the structural integrity of IgG and substantial aggregation is observed with only 20% activity detected in microspheres fabricated with water in oil in water (W/O/W) double emulsion method and 37% activity detected in microspheres

fabricated with solid in oil in water (S/O/W) method spray using freeze-dried IgG [115]. There are a number of factors that can affect protein stability during fabrication, including oil-water interfaces, water-air interfaces, acidic microenvironment resulted from hydrolysis of PLGA, PLGA–protein interactions, non-specific adsorption to the PLGA matrix and protein acylation [116].

1.3.4.7. Accelerated Stability Studies

During formulation development, it is often required to compare a large number of different formulation variables such as buffer, pH and excipient compositions needed to be screened to determine shelf-life of the protein and the best conditions for protein stability. Due to time constraints for formulation development, it is often not practical to compare each test formulation under real-time storage conditions. One of the most common methods to accelerate the degradation of protein in a formulation is through storage at elevated temperatures where the rates of degradation are increased and used to predict the degradation rates at typical storage conditions according to the Arrhenius equation.

$$k = Ae^{-\frac{E_a}{RT}} \qquad (1.1)$$

where k is reaction rate constant, A is the pre-exponential factor, E_a is the activation energy, R is the ideal gas constant and T is temperature.

For freeze-dried formulations stored in the glassy state, the accelerated stability studies should be conducted in the glassy state. Breen and co-workers have demonstrated that most chemical and physical degradation pathways followed Arrhenius behaviour during storage in the glassy state except ASP isomerization which only follows Arrhenius kinetics above the T_g of the glass of an humanised antibody [117]. Duddu and co-workers

have demonstrated that the data generated from accelerated stability studies in the glassy state (40 ºC) probably a better indicator of relative stability than the data generated at a higher temperature (60 ºC) where the antibody formulation in question may be near or above T_g mAb [118].

1.4. Computation Methods

In 1965 Moore described his observation that the number of transistors in a dense integrated circuit doubles approximately every two years [119] which is usually translated into computational power. With ever increasing computational power, *in-silico* studies have become more widely possible in scientific research and data analysis. Today, *in-silico* studies complement experimental approaches in both academic and pharmaceutical research, especially non-specific protein-protein interaction is often too small to characterise experimentally using existing biophysical methods. The molecular systems studied ranged from low molecular weight molecules to large molecular weight biologically derived molecules where computational resources have evolved to predict their dynamic behaviours over time. More recently, the use of *in-silico* modelling to study mAb aggregation behaviour in different compositional environment provided by varying formulation conditions has attracted a lot of attention and various computational methods have been developed [120-125].

Molecular docking strategies have been used to identify potential excipients and to study their interactions with antibodies. Veurink and co-workers described an *in-silico* approach to model aggregation-prone regions of bevacizumab and discovered that dexamethasone can prevent formation of bevacizumab dimers at LYS445 located on the Fc region which interacts with the Fab on the second bevacizumab [120]. *In-silico* prediction was further supported by *in-vitro* studies and dexamethasone showed positive

effects in reducing the formation of bevacizumab dimers, trimers and higher order aggregates after 28 days at 40°C compared with bevacizumab alone. Barata and co-workers used docking methods to determine hotspots for Fab A33 protein-protein interactions and interactions with commercially available excipients such as arginine, glycine, mannitol, sorbitol, sucrose, trehalose and Tween 20 [121]. Binding energies of these excipients as predicted by docking were in good agreement to the T_m values from a stepped thermal study.

MD simulations are useful in simulating behaviours of molecules in small liquid systems over timescales of nanoseconds. The dynamic conformations of a system are calculated by integrating Newton's equations of motion and captured in a trajectory that contains the positions and velocities of all atoms at a set simulation time point. The results from molecular dynamics simulations often guide experimental design to try to reduce the number of experiments required. Chennamsetty and co-workers performed short MD simulations on two therapeutic antibodies and their mutants using all-atom force field CHARMM to predict aggregation prone regions based on surface features of the molecule including side-chain hydrophobicity and solvent accessible surface area [122]. The wild type and mutants of these two antibodies were then experimentally evaluated for their aggregation behaviour. The authors found by replacing hydrophobic residues within an aggregation prone region with less hydrophobic residues can improve stability compared with the wild type antibody, suggesting this tool can be useful for engineering antibodies for stability by target mutations.

Bucks and co-workers used an AMBER99 all-atom model to study under thermal stress an IgG1 b12 antibody Fab that recognises the CD4-binding site of the human immunodeficiency virus-1 (HIV-1) gp120 protein [123]. The authors have found the

interfacial beta-strands within the Fab structure are more stable under thermal stress compared to other strands or loops, therefore aggregation prone regions located in interfacial beta-strands are effectively protected from self-associations. Changes were observed in the structures of the CDR regions, edge beta-strands and adjoining framework beta-strands causing increased solvent-accessible surface area and exposing aggregation prone regions in these regions. From this study, the authors identified two potential aggregation regions in the VH domain in the Fab located between PHE32 and VAL37 and between THR107 and SER112.

Therapeutic IgGs are relatively large in size and no more than one molecule was simulated using an all-atom force field which had been described in literature. Many interesting phenomena such as antibody-antibody assembly occur at much longer time scales that fall beyond the capabilities of current computation. Coarse-grained models have been developed to facilitate simulations of very large biomolecular systems involving multiple mAb molecules. Chaudhri and co-workers have developed two coarse-grained models to represent a full mAb molecule; a 12-site lower resolution model and a 26-site higher resolution model [124]. One thousand mAb molecules were evaluated in each simulation and self-associated dimers and trimers were calculated from equilibrated structures. The authors found positive correlations between the number of dimers formed in simulation and the observed viscosity differences between different simulated mAb systems. This study further confirmed that by reducing attractive charges in the Fab region will lead to a decrease in self-association of mAb molecules and reduce viscosity at high concentrations, especially at 70 and 100 mg/mL. Cluster formations were also analysed and the authors also suggested smaller effective excluded volumes could possibly explain lower viscosities at high concentrations where packing effects are important. Buck and co-workers have used this 12-site model to study four additional

mAbs and evaluate diffusion coefficients computed from simulations. *In-silico* results were in good agreement with experimentally determined viscosities for these four mAbs [125].

1.4.1. Aggregation-scoring Algorithms

One computational approach in determining the aggregation-prone region in a protein is through the use of aggregation-scoring algorithms based on the amino acid sequence of the protein, such as AGGRESCAN [126], Hot Spot is Found [127], PASTA [128], TANGO [129] and Zyggregator [130]. These algorithms use the sequence of the target protein to identify sequence motifs that can lead to aggregate formation based on physico-chemical properties such as β-sheet propensity, charge and hydrophobicity. These sequence-based scoring algorithms are mostly derived from amyloidogenic proteins and their sequences (Table 2) which are highly susceptible to form fibrous protein aggregates. Aggregation-scoring algorithms have been used to identify aggregation-prone region in antibodies and these aggregation-prone region are mostly located on the CDRs of the VH domain [131].

Algorithm	**Method**
AGGRESCAN	Relative aggregation propensity of Aβ42 mutants
Hot Spot is Found	Relative amyloidigenic and β-aggregation propensities based on experimental data of small peptides.
PASTA	The calculation is based on statistical mechanics of pair-wise interaction of a pair of residues those are likely to form fibrils cross-β structure.
TANGO	The calculation is based on statistical mechanics to predict aggregation by using physico-chemical principles of β-sheet and the assumption that the core regions of an aggregate are fully buried.

Zyggregator	Relative propensity calculation based on hydrophobicity, α-helix propensity, β-sheet propensity, hydrophobic pattern charge and local stability.

Table 2: Comparison of aggregation-scoring algorithms.

1.4.2. Crystal Structure

Protein structure can be determined by experimental methods such as X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy, cryo-electron microscopy or small angle X-ray scattering (SAXS). X-ray crystallography is the most common method in determining protein structures with high resolutions. Extraction, purification and crystallisation of the protein is required for X-ray crystallography. Since most crystallisation conditions are different to the physiological or cellular conditions for the protein of interest (i.e. trans-membrane proteins) and this may lead to structural artefacts. In addition, crystal structure is static so it may only represent a transient protein structure. In contrast, NMR spectroscopy can detect the dynamic behaviour of protein in solution but it is typically for proteins with a molecular weight below 25 kDa [132].

1.4.3. Molecular Docking

Molecular docking is a powerful and efficient method to predict the atomic interactions of two molecules using three dimensional structural information. Protein–protein interactions play crucial roles in the execution of biological functions in living organisms, including cellular proliferation, metabolism, gene expression and signal transduction. There are over 125,000 structures available on the Protein Data Bank (PDB), many of these proteins are able to form complexes. However, only a small proportion of the structures are available as complexes. Protein-protein docking is therefore an important tool to predict the structures of protein complexes.

Docking of two molecules requires the translational and rotational positioning of each molecule in relation to each another. Information regarding binding sites may not be complete so initial sampling using a geometry-based algorithm for protein-protein docking to cover all possible orientations resulting in a large number of possible protein complexes. If additional biological information regarding the location of the interface is available, many of these algorithms allow this information to be included to simplify and improve accuracy of the search.

In most cases, the refinement of the initial docking results are based on a scoring function taking into account shape complementarity, charge, desolvation and residue potentials [133]. Similarly, the same principles can also be applied to small molecule ligands docking to proteins. In protein-ligand docking, the binding of a ligand to a target protein is simulated and the binding free energy is calculated. Binding of a ligand to the receptor may result in conformational changes and some algorithms take into account the conformational flexibility for improved accuracy [134]. Protein-ligand docking has been widely used as virtual screening method to identify potential small molecule drug candidates where a large library of computational ligands are screened against a defined protein target for lead identifications.

1.4.4. Molecular Dynamics

The Nobel Prize in Chemistry 2013 was awarded to Martin Karplus, Michael Levitt and Arieh Warshel for their efforts in the development of computer simulations of biological systems. Molecular simulations (MD) at atomic level can give powerful insights into how protein interacts with other molecules and their surrounding environment under different physical and thermal conditions which is very difficult to study experimentally. Various

molecular dynamics programs such as AMBER [135], CHARMM [136], DL POLY [137], LAMMPS [138], NAMD [139] and GROMACS [140] are now routinely used for a wide range of applications.

MD simulations have been used extensively to identify protein conformations and interactions. Classical MD simulations are typically run within a box of specified size treat atoms as spheres and the bonds connecting them as springs. These simulations are designed to solve classical equations those describe the forces between pairs of atoms using predefined potentials such as distance, angle, dihedral, van der Waals and electrostatic forces [141-143]. During a MD simulation, the force of every atom in the system is calculated based on Newton's equations of motion at very small time steps. In order to eliminate the effect of the boundary, periodic boundary conditions are used where the box is replicated in all dimensions, atoms reaching a boundary are translocated to the opposing boundary. A large number of bonded and non-bonded interactions between each atom pair in the system needed to be calculated. As a result, enormous computation power is require for exploration of large biomolecular systems.

1.4.4.1. Classical Force Fields

Each atom in the simulation system is considered as a charged sphere mass of the corresponding atom and bonds are regarded as springs. The motion in the system is determined by the forces exerted upon it by all the other atoms in the system and the changes in the forces between all of the atom pairs are calculated throughout the simulation. In molecular dynamics, the quality and accuracy of biomolecular dynamic simulations relies critically on the force field which is a set of mathematical terms and associated constants to describe the time evolution of bonded and non-bonded interactions between particles in the system.

The parameters used in the force fields can be derived from a combinations of experimental data and quantum mechanics calculations. After continuous refinements of commonly used force fields, such as AMBER [141], CHARMM [142] and GROMOS [143], over the years, these force fields achieved satisfactory agreements with experimental data [141-143]. Each force field is derived from different methods in practice, the main differences include functional formula, particle type and definition as well as the quantum mechanics calculation methods and experimental data used in parameterising molecules.

Molecular dynamics primarily solves the Newtonian equation of motion for each atom of the system to determine new positions of atoms at a future time point.

$$F_i = m_i a_i \tag{1.2}$$

where *F* is force, *m* is mass and *a* is acceleration.

Therefore the change of variables with respect to time

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \tag{1.3}$$

where *F* is force, *m* is mass acting on it.

There are a different integrator algorithms available. One of the common algorithms is the Verlet algorithm which use the positions and accelerations at time t $(t + \Delta t)$, and the positions from previous step $(t - \Delta t)$ to calculate new positions at time t.

$$r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + a_i(t)\Delta t^2 \tag{1.4}$$

where *a* is the acceleration of particle at time *t*.

So velocity can be calculated by working out the difference between the positions and accelerations at time t (t+Δt), and the positions from previous step (t-Δt).

$$v(t) = \frac{r_i(t + \Delta t) - r_i(t - \Delta t)}{2\Delta t} \quad (1.5)$$

where v is the acceleration of particle at time *t*.

At each time step, the forces on each atom are determined and the coordinates of each atom are calculated. This process is carried out repeatedly for a desired number of steps to generate a trajectory that describes how the positions and velocities of the atoms in the system vary with time. In order to balance simulation timescale and to ensure integration errors are small enough to be ignored, a typical time step of 1 fs is used as this is the smallest timescale of motions within a molecule and is correlated to bond vibrations. Constraint algorithms, such as SHAKE [144], RATTLE [145] and LINCS [146], are frequently used to constrain bond lengths involving hydrogen atoms so a longer time step can be employed to reduce computational costs.

For a given atom *i*, the interactions between atoms are described as is a sum of bonded and non-bonded interactions:

$$U_i(\vec{R}) = \sum U_{i,bonded}(\vec{R}) + \sum U_{i,non-bonded}(\vec{R}) \quad (1.6)$$

where *U* denotes the total potential energy based on the positions of the atoms and *R* indicates the positions of atoms.

1.4.4.2. Bonded and Non-Bonded Terms

The bonded terms include bond stretching, angle bending, bond rotation (torsion) arising from the quantum mechanical interactions (Figure 3).

$$\sum U_{i,bonded}(\vec{R}) = U_{Bond} + U_{Angle} + U_{Dihedral} \tag{1.7}$$

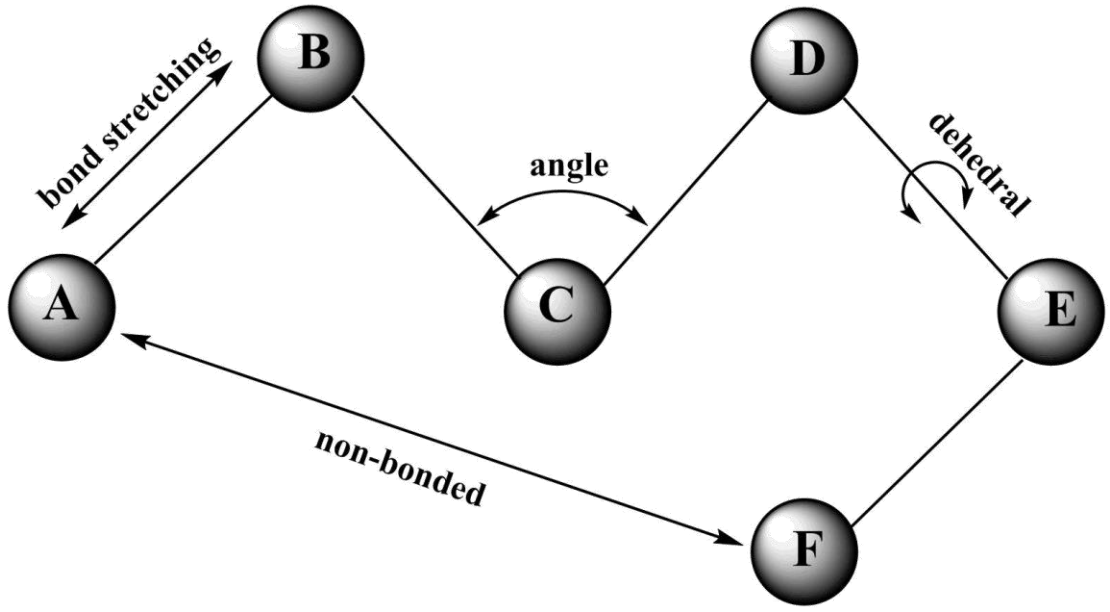


Figure 3: Graphical representation of components of a typical mechanical force field. Bonded interactions include the bond stretching, angle and torsional terms. Non-bonded terms consist of Van der Waals and electrostatic interactions.

The bond stretching term is a harmonic potential that describes the stretching energy of atomic pairs separated by a single bond (1-2 interaction), derived from a Taylor expansion.

$$U_{Bond} = \sum_{Bonds} k_i^{bond}(r_i - r_0)^2 \tag{1.8}$$

where *r* is the distance between two atoms and *k* is the bond stretching force constant.

These terms can be derived from experimental or quantum mechanics calculation. Whilst computationally efficient, this prohibits the possibility of rearrangement and reformation of covalent bonds, preventing chemical reactions to occur in the simulation.

The angle term is the energy for angle bending between three atoms which are bonded together (1-3 interaction). Similar to the bond stretching term, the angle term is also a harmonic potential and expressed as a Taylor expansion.

$$U_{Angle} = \sum_{Angles} k_i^{angle}(\theta_i - \theta_0)^2 \qquad (1.9)$$

where θ is the angle between two bonds and k is the force constant for the angle between two bonds. Again, these terms can be derived from experimental or quantum mechanics calculation.

The dihedral bond angle, or torsion angle, is the angle between the planes formed by two atom pairs. This is a term to describe the energy for rotating the central bond within a connected four atoms sequence (1-4 interaction). The dihedral bond can be rotated 360° and the energy for rotation is often low. The rotational barrier not only determined by the dihedral term, but also be influenced by non-bonded terms. Thus, the dihedral bond angle is expressed as a Fourier series

$$U_{Dihedral} = \sum_{Dihedrals} k_i^{dihedral}[1 + \cos(n_i\theta_i + \delta_i)] \qquad (1.10)$$

where θ is the dihedral angle, n is the periodicity, δ is the phase difference and k is the force constant for the dihedral angle. In this case, quantum mechanics data is more

preferable to experimental data as calculation of the entire energy surface is necessary for optimisation of the dihedral term with respect to the energy surfaces.

The non–bonded terms comprise the van der Waals forces and electrostatic interactions.

$$\sum U_{i,non-bonded}(\vec{R}) = U_{vdw} + U_{elec} \tag{1.11}$$

Van der Waals forces are the attractive or repulsive forces between molecular entities other than electrostatic energy arising from atomic charges. Van der Waals forces are very positive at short distance and after it reaches its minimum, the attractive forces reduces with increasing distance between molecules. This is caused by the electrons clouds of adjacent atoms or groups. At short distances, the atomic electron clouds overlap causing a strong repulsion. As the distance increases, fluctuations in the electron clouds lead to the formation of instantaneous dipoles which can induce the formation of instantaneous dipoles in adjacent atoms or groups. The attractive force is strongest at this distance and gradually reduces as the distance increases.

Van der Waals forces is widely expressed as the Lennard-Jones (LJ) potential:

$$U_{vdw} = \sum_{vdw} \varepsilon_{ij} \left[\left(\frac{R_{\min ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min ij}}{r_{ij}} \right)^{6} \right] \tag{1.12}$$

where ε is the well depth, R_{min} is the distance at which the interaction energy reached minimum and r is the distance between molecules. The extent of the force depends on the distance of the molecule pair. The strong repulsive r^{12} term dominates at short range and the r^6 term denotes attractive van der Waals forces at long range.

As previously discussed, van der Waals forces are caused by instantaneous dipoles formation with constantly shifting electron distributions, so electrostatic interactions are result from partial or whole charges on atoms of a molecule. This is modelled through the assignment of partial charges to each atom of a molecule. The electrostatic potential energy results from Coulomb forces between different charged atoms is according to their point charge.

$$U_{elec} = \sum_{elec} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{1.13}$$

where ε_0 is vacuum permittivity, q is relative point charges and r is the relative distance. The magnitude of the electrostatic energy varies inversely with the distance between two charged points.

As mentioned earlier, the total potential energy of the system is calculated as the sum of all bonded and non-bonded interactions. Since calculating all non-bonded interactions between all pairs of atoms is not practical for large models, a cut-off can be used to exclude atoms outside of the specified range from computation for Lennard-Jones potential as the magnitude of the force becomes too small to make any significant difference. Unlike Lennard-Jones, assigning a cut-off for Coulomb forces can potentially cause accumulation at the cut-off boundary. Ewald summation (PME) [147] and Particle-Particle Particle Mesh (P3M) [148] are methods that calculates the electrostatic energy of a system on a lattice to facilitate faster computation.

To enable simulations to be conducted under controlled conditions and to conserve energy over the course of the simulation, ensembles are used. In a canonical (NVT) ensemble, the number of atoms (N), the volume (V), and the temperature (T) of the system remains constant. NVT ensembles frequently employ a thermostat to maintain the temperature of

the system by coupling the kinetics of the system to an external heat bath at desired temperature, where the deviations in temperature are slowly adjusted by rescaling the velocities of atoms. The Berendsen thermostat rescales atomic velocities and the rate of temperature change is kept proportional to the temperature difference so the temperature change does not happen rigorously to generate the correct statistical ensemble. In an isothermic-isobaric (NPT) ensemble, the system is maintained under constant temperature and pressure, where the number of atoms (N), the pressure (P), and the temperature (T) of the system remains constant. NPT ensembles also employ a thermostat to maintain the temperature of the system along with a barostat to maintain the pressure of the system in a similar way using a pressure bath. Berendsen barostat works by dynamically scaling the coordinates and box size in the simulation to match the pressure bath. For most applications, NPT ensembles are the most appropriate, as temperature and pressure controls are generally required to reproduce real-life events.

1.4.4.3. Periodic Boundary Conditions

For biological molecules, the presence of water and ions have important roles in their biological behaviour. Therefore solvent environment must be carefully considered. In a real system, atoms are not confined to the size of the box and should be treated as in infinitely, therefore enormous computational power would require to compute the solvation environment. Periodic boundary conditions (PBC) are often adapted to overcome this problem and to eliminate surface effects. Under PBC, the solvated protein is simulated in a solvated box in the centre (Figure 4) with the box of molecules being surrounded by images of itself. When an atom is passing through the boundary of the simulated box, the atom will enter the opposite side as if there is no boundary. Each molecule in the system will therefore not just interact with molecules inside of the simulated box but also its images. In this way, the bulk solvation environment can be

represented by simulating a reasonable number of molecules, preventing evaporation and reduce the number of molecules required in the simulated system.

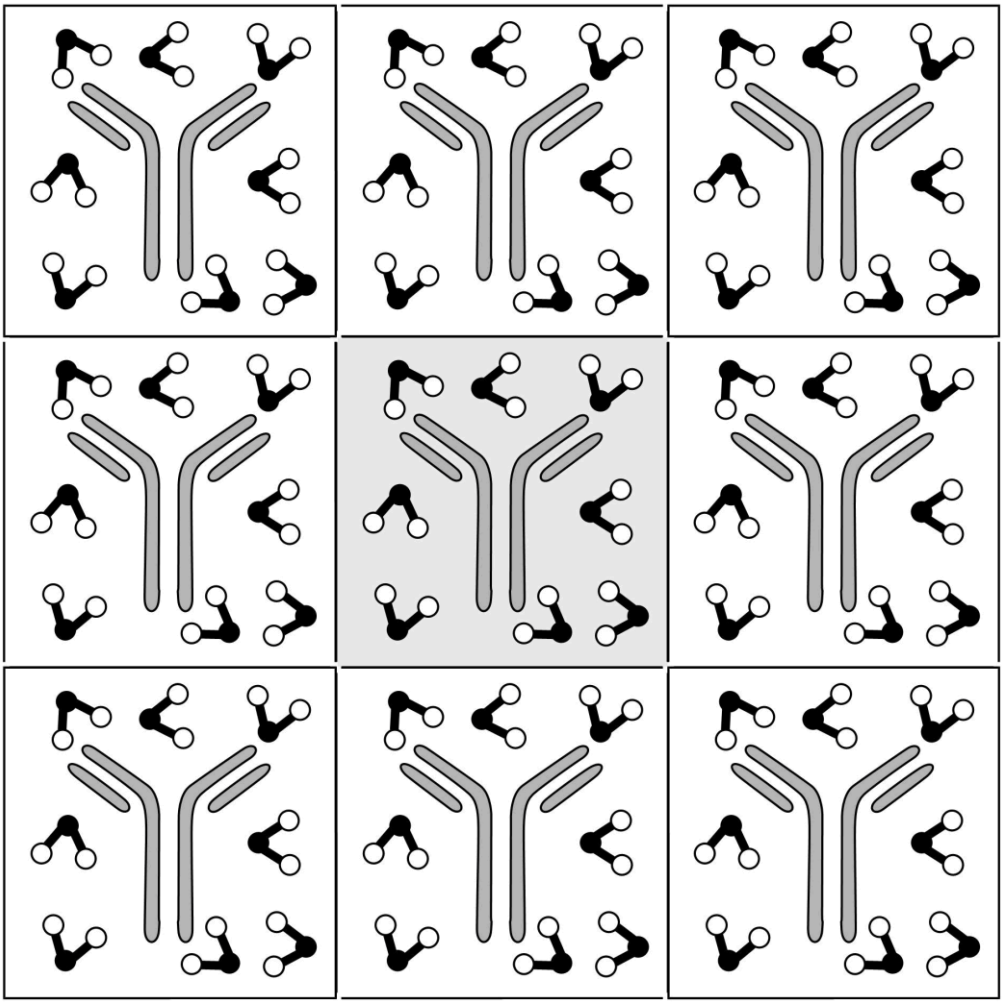


Figure 4: Graphical representation of periodic boundary conditions. The simulated system is located in the centre and the periodic images of it repeated in all directions around the system. If any atoms of the system leaves the simulated box, they will enter the box from the opposite face.

1.4.4.4. Coarse-Grained Force Fields

Coarse-graining is a useful technique in MD simulations to reduce the degrees of freedom by reducing the amount of particles to be computed in a simulation. This is achieved by grouping atoms within a molecule into larger beads representing corresponding parts of the molecule and the properties of the beads are derived from their constituent atoms.

This allows longer simulated times and larger systems to be simulated in exchange for a reduction of resolutions.

1.4.5. High Performance Computing

The increase of high computational power has made molecular dynamics simulation a more common tool to study proteins such as antibodies, complementing scientific theories and experiments. A model for a full-length mAb involves about 20,000 atoms and a typical model for a single mAb surrounded by a box of explicit water involves at least 500,000 atoms. It would be impossible to run molecular dynamics simulations involving multiple mAbs for sufficient sampling of the energy landscape using a standard PC. With the advanced computational power, a large system containing multiple molecules of mAb can potentially be studied for a longer period of time. Most of the computational work described in this thesis were conducted using UCL High Performance Computing Facilities (Grace@UCL and Legion@UCL).

1.5. Aims and Thesis Overview

Apart from protein aggregate-induced immunogenicity caused by therapeutic biologics, protein aggregation has been associated with a number of human diseases such as Alzheimers and Huntington diseases, and systemic amyloidosis [149]. The pathogenesis of these disorders involves abnormal accumulation and aggregation of specific proteins [149]. Although the morphologies of the protein in these diseases may differ the formation of protein aggregates depends on the propensity to form β-sheets [150]. To maintain cellular homeostasis, molecular chaperones such as Heat Shock Proteins, facilitate folding or refolding of mis-folded proteins to prevent protein aggregation [151]. Unlike cells, there are no aggregation detection and correction mechanisms in formulations of monoclonal antibodies. Excipients described in this chapter can act as a

protective mediator against protein aggregation. This raises the interesting question of whether it is possible to use computational strategies to screen for excipients that prevent aggregation in the final dosage form, reducing the potential failure in development and preserve product quality.

The aim of this thesis was to develop a method to screen for potential excipients for mAbs. MEDI1912 is an antibody that faces various challenging due to aggregation [152] but its structure has not been resolved by X-ray crystallography. Initial work began with studying the parent molecule of MEDI1912 which was generated by in-vitro affinity maturation of MEDI-578.

The crystallographic data of MEDI-578's Fv fragment was used to calculate the most solvent accessible hydrophobic area and commercially available small molecules were screened against this area with molecular docking tool AutoDock Vina in a view to disrupt protein-protein interaction in Chapter 2. The molecule with the highest binding affinity was further examined using short all-atom molecular dynamics simulations. Formulations of MEDI-578 with the top ranked molecule were tested in Chapter 3 and this molecule did not destabilise MEDI-578.

Consistent, rapid and reproducible models are always preferred. Most studies reported in the literature have either used a single molecule of an antibody or a highly coarse-grained representations of antibody to simulate antibody self-associations. In Chapter 4, MARTINI, which is a coarse-grained molecular dynamics model was used to study the free formation of MEDI-578 Fv fragment-Fv fragment complexes. The MARTINI method enabled a more in-depth understanding of the molecular events of 578 Fv

fragment-Fv fragment complex formations, observing the interaction pattern changes in present of different excipients.

The effect of intermolecular protein-protein and protein-solvent interactions are important in defining dimer formation and often lead to improvement in the predictions of binding energies from a MD simulation. The central issue discussed in Chapter 5 is the calculation of free energies from the MD simulation data generated in Chapter 4. A detailed set of simulations was performed on various aspect of these free energy calculations. Accurate estimates of binding affinity are important to distinguish between stable protein complexes. Evaluating energy changes in protein-protein interaction requires the calculations of classical electrostatic and van der Waals enthalpy terms between protein molecules as well as evaluating conformational flexibility and entropic effects associated with desolvation of the buried surfaces. This additional piece of work aims to calculate these entropic contributions and evaluate their relative importance by comparing them to the simulation results in Chapter 4.

Chapter 6 further utilise the methods described in Chapter 4 to examine the potentials of hydrophilic dipeptides as excipients for MEDI-578. The changes of interaction energies were in good agreement with the DLS data. A homology model of MEDI-1912 Fv fragment and a full Fab model of Motavizumab were also used and compared. The overall results were similar among these models with dipeptides with charged amino acid residues resulted in reduced interactions between antibody molecules. The work presented in this thesis offers a novel computational method that can be used to screen for potential excipients reliably. Such method would be expected to have utility in developing mAb formulations and also has the potential to be applied to different types of therapeutic proteins.

2. *In-silico* Screening for Potential Excipients

2.1. Introduction

There are a wide selection of *in-silico* modelling methods available for aggregation prediction ranging from less-computationally expensive sequence-based methods to more advanced methods such as docking and molecular dynamics simulations. Predicting the relative aggregation propensity from the protein sequence is the most convenient method to study aggregation-prone regions of a protein. These aggregation-prone predicting algorithms were mostly validated with small known amyloidogenic peptide sequences which are known for having β-sheet structures with a strong tendency to aggregate [126 -130]. The overall structure of antibody is highly conserved and in most cases only differ in few key residues on the CDR, these sequences are highly likely to appear on other antibodies as well.

The use of aggregation prediction algorithms may offer good predictions of aggregation propensity of amyloidogenic proteins. However, the lack of three-dimensional structural information in these tools, specific local aggregation-prone surface features such as charge, solvent accessible surface areas, side-chain hydrophobicity and conformational flexibility are not known from these studies. These tools only offer an oversimplified approach to protein stabilities and perhaps relevant in amyloidogenic proteins. Therefore, using aggregation-scoring algorithms are not conclusive to determine the global stability of antibodies and unable to predict key formulation parameters such as pH environment, concentration and effect of excipient in the formulation. It is important to understand the aggregation-prone regions of a protein in relation to its 3D structure and the dynamics under different formulation environments. Hence, Molecular Dynamics (MD) simulations would be a more suitable technique in determining the stability of proteins.

MD is a well-established method to simulate the structure and dynamics of biomolecular systems at atomistic level. Proteins are dynamic in nature so the binding of a ligand to a protein results in complex intramolecular and intermolecular interactions those lead to changes in conformation and exposure of aggregation-prone regions. In this context, MD simulations can dissect protein-ligand binding with high accuracy and can probe these complex events with atomistic details and revealing binding profiles. The hypothesis of this work is that low molecular weight excipients can be identified that bind to solvent accessible hydrophobic regions of an antibody to disrupt antibody-antibody interactions to improve mAb conformational stability to prevent or delay mAb aggregation processes.

The use of three-dimensional structural information with Spatial Aggregation Propensity (SAP) calculations can provide a more detailed analysis of the surface features of the molecule including side-chain hydrophobicity and solvent accessible surface area [122]. As a protein is dynamic in nature, the SAP calculations can take advantage of MD simulations and SAP can be determined over the course of the simulation. These type of simulations are computationally costly but it allows for the dynamic exposure of the solvent-accessible hydrophobic region and the effect of relevant parameters such as solvation to be simulated.

The self-association of two monomers to form a dimer is the first step towards nucleation. Since this process is usually reversible, a better understanding of the shape, structure and protein-protein interacting sequences of early precursors of aggregates such as dimers would be valuable in addition to special aggregation propensity calculations. Molecular docking was therefore selected to study the interaction between two proteins. ZDOCK is a protein docking algorithm based on Fast Fourier Transform (FFT) rigid-body search

along with scoring functions composed of desolvation energy, electrostatics, and grid-based shape complementarity [153].

Conformational complementary of the molecule is important for a ligand that is binding to its target. Single docking experiments are useful for how individual compound interact with the target in an attempt to find out the best binding position between a receptor and a ligand. Virtual screening uses a large library of compounds to dock on the same target and subsequently ranked the compounds to identify potential lead candidates.

The virtual screening in this study was powered by AutoDock Vina which is one of the most popular and widely used free and open source software that is widely used for protein-ligand docking. AutoDock Vina was developed by Oleg Trott in the Molecular Graphics Lab at The Scripps Research Institute [154]. It supports parallel computation offering improved performance and accuracy compared to the previous versions of AutoDock [154]. AutoDock Vina has an advantage of treating both the ligands and receptors flexibility, so the dynamic nature of the protein can be accounted for.

AutoDock Vina uses a file format of pdbqt for the input files for the ligand and receptor. Kollman charges are assigned to the protein receptor and Gasteiger charges are assigned to the ligands based electronegativity equilibration [154]. AutoDock Vina uses a united atom model and only polar hydrogens are considered. The search area has to be specified by the size and the three-dimensional coordinates. AutoDock Vina calculates the free energy of the binding to the receptor, negative calculated affinity indicates the binding is thermodynamically favourable. The output contains a pdbqt file with the best binding modes, calculated affinity to the receptor and RMSD with respect to each binding modes.

To ensure a wide range of commercially-available compounds were evaluated in this study, ZINC 15 database was used for ligand selections. The ZINC database [155] contains over 230 million compounds, and can be split into different subsets based on chemical properties such as log P, molecular weight, pH and charge. Ligand files are formatted in mol2, sdf and pdbqt supporting various docking software packages. In addition, all compounds within the ZINC database provide information on ligand purchases and their suppliers so the identified lead compound could be purchased and tested.

MEDI1912 is an antibody that faces various challenging due to aggregation [152]. MEDI-578 is the parent molecule of MEDI1912 so MEDI-578 was chosen as the model molecule for this set of studies. Ultimately, it is hoped the knowledge gained in this study can help designing formulations of MEDI1912. This chapter began with identifying aggregation prone region using Spatial Aggregation Propensity (SAP), a structure-based aggregation prediction algorithm, and ZDOCK, a protein-protein docking algorithm. The structural information of the aggregation prone region was used for virtual screening for potential ligands with suitable chemical properties preventing the dynamic exposure of the aggregation prone region. The ligand with the best affinity to the aggregation prone region is further examined with MD simulations.

2.2. Materials and Methods

2.2.1. MEDI-578

MEDI-578 has been used as the model molecule in this set of studies and its structure has been deposited in the protein data bank (PDB ID: 5JZ7). The published structure of MEDI-578 was resolved by X-ray crystallography with a maximum resolution of 3.4 Å.

The published structure contains four MEDI-578 Fv fragments and four β-NGF molecule, with two MEDI-578 Fv fragments bound to a β-NGF dimer (Figure 5) [152].

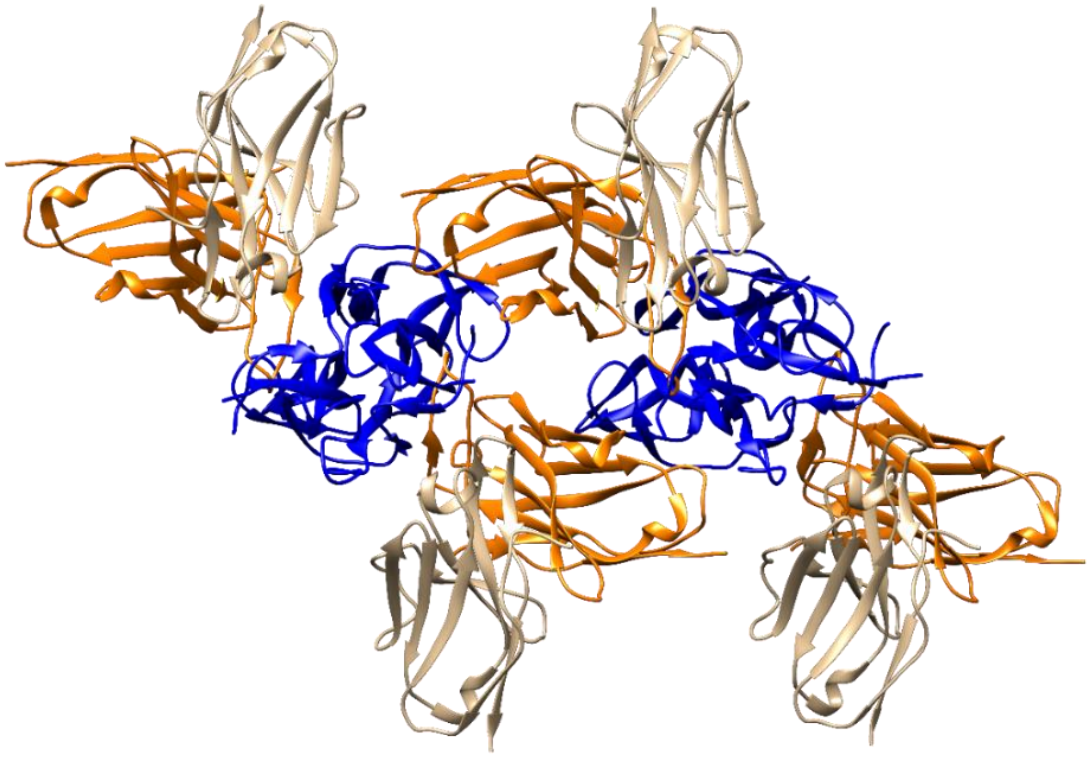


Figure 5: The resolved structure of MEDI-578/β-NGF complex. Light chain, Heavy chain and β-NGF were coloured in orange, beige and blue respectively.

Light Chain: QSVLTQPPSVSAAPGQKVTISCSGSSSNIGNNYVSWYQQLPGTAPKLLIYDNN KRPSGIPDRFSGSKSGTSATLGITGLQTGDEADYYCGTWDSSLSAWVFGGGT KLTVL
Heavy Chain: QVQLVQSGAEVKKPGSSVKVSCKASGGTFSTYGISWVRQAPGQGLEWMGGI IPIFDTGNSAQSFQGRVTITADESTSTAYMELSSLRSEDTAVYYCARSSRIYDL NPSLTAYYDMDVWGQGTMVTVSS
β-NGF: FHRGEFSVCDSVSVWVGDKTTATDIKGKEVMVLGEVNINNSVFKQYFFETK CRDPNPVDSGCRGIDSKHWNSYCTTTHTFVKALTMDGKQAAWRFIRIDTAC VCVLSRKA

Table 3: Sequence of MEDI-578 Fv fragment and β-NGF as in the published structure.

2.2.2. Spatial Aggregation Propensity

Hydrophobicity is a major factor in protein aggregation as aggregation-prone regions are usually characterised by an enrichment in hydrophobic, both aliphatic (e.g. VAL, LEU, ILE) and aromatic (e.g. PHE, TRP, TYR) residues. Spatial aggregation propensity (SAP) is an *in-silico* tool to measure the hydrophobicity of these exposed surface patches of a three dimensional protein structure. The SAP value is calculated based on the following equation [122].

$$SAP_{atom_i} = \sum_{\substack{Simulation \\ average}} \left\{ \sum_{\substack{Residue\ with \\ at\ least\ one \\ atom\ within \\ R\ of\ atom\ i}} \left[\frac{\begin{array}{c} SAA\ of\ the\ side\ chain\ atoms \\ within\ radus\ R \\ \times\ Residue\ hydrophobicity \end{array}}{\begin{array}{c} SAA\ of\ side\ chain\ atoms \\ of\ fully\ exposed\ residue \end{array}} \right] \right\} \quad (2.1)$$

where

(1) SAA is the "solvent accessible area" of side chain atoms contained within radius R from the central atom.

(2) SAA of side chain of fully exposed residue (for amino acid "X") is obtained by calculating the SAA of the side chain of the middle residue in the fully extended conformation of tripeptide "Alanine–X–Alanine."

(3) Residue hydrophobicity is obtained from the hydrophobicity scale of Black and Mould. The scale is normalized such that glycine has a hydrophobicity of zero, the most hydrophobic residue (PHE) has a value of 0.5, and the least hydrophobic residue (ARG) has a value of −0.5. Hydrophobic residues have residue hydrophobicity values greater than 0. Residue hydrophobicity values less than zero are less hydrophobic than glycine.

2.2.2.1. SAP Calculations

The SAP scores of the MEDI-578 Fv fragment was calculated at R=10 Å using the X-ray crystallography structure of MEDI-578 Fv fragments-β-NGF complex [152].

2.2.2.2. Dynamic SAP Calculations

A short MD simulation was prepared for SAP calculations using the X-ray crystallography structure of MEDI-578 Fv fragments-β-NGF complex as discussed above. The MD simulation systems were prepared and ran using GROMACS 2016.3 [140] and the CHARMM36 topologies [142]. The system was solvated with TIP3 waters and 28 NA ion molecule and 28 CL ion molecules were added to the system which is equivalent to 100 mM NaCl. This forms a rectangular box of 7.77 × 7.77 × 7.77 nm^3. Overall the simulation system has a total 46569 atoms. The system was first minimised and equilibrated for 1 ns. Subsequently, the MD simulation was ran at 2 fs/step with the use of periodic boundary conditions, constant temperature of 300 K and pressure at 1 bar using V-rescale thermostat and Parrinello-Rahman barostat respectively. Electrostatic and van der Waals interactions were calculated using a Verlet cutoff-scheme with Potential-shift-verlet as modifier with a cut-off of 1.0 nm. Snapshots were taken every 0.01 ns which is used for SAP calculations. SAP values of MEDI-578 Fv fragment were calculated at R=10 Å over a 60 ns MD simulation.

2.2.3. ZDOCK

Coordinates for the MEDI-578 Fv fragment were taken from the X-ray structure of the MEDI-578-βNGF complex (PDB ID: 5JZ7) and two monomers of MEDI-578 Fv fragments were docked together using ZDOCK 3.0.2 [153] without any residue blocking with a total of 2000 dimer structures were generated. The top 2000 predictions were

analysed and the distance between any pair of atoms from the respective of MEDI-578 Fv fragments were calculated and the atom pair with minimum distance is analysed.

2.2.4. Virtual Screening

In this study, the lead compound is considered to be having similar properties as disaccharides such as maltose, sucrose and trehalose. Ideally, the lead compound has to be soluble in water and have a small molecular weight. Therefore, the following selection criteria have been applied.

1. Currently in stock with a supplier
2. Log P below 1 to ensure solubility
3. A molecular weight below 400 g/mol
4. Chemically unreactive
5. Overall net charge of 0 at pH 7 was selected to ensure the effect of the ligand as a potential excipient in formulation is due to binding to the hydrophobic CDR

This yield a subset of 83845 compounds and input structures in pdbqt format were downloaded from ZINC database. The structures of sucrose, maltose and α,α-trehalose were also downloaded from the ZINC database for comparison. The MEDI-578 Fv fragment protein structure was obtained from the X-ray crystallography data of the MEDI-578-βNGF complex (PDB ID: 5JZ7) as described above. Kollman charges and polar hydrogen atoms were added to the structures of the MEDI-578 Fv fragment and the file was saved as PDBQT format using AutoDockTools [156]. The search grid focused on the most hydrophobic part of the MEDI-578 Fv fragment from SAP calculation within a cubic box size set at 25 grid points for all three dimensions extended from the Cα atom that has the highest SAP score. The exhaustiveness of the global search was set at 125 to emphasise docking accuracy and number of generated binding modes set to 9. Results

were pooled together and ranked. The ligand with best free energy of binding was selected for analysing the interactions between MEDI-578 Fv Fragment and all these compounds.

2.2.5. Molecular Dynamics

Atomistic simulations of MEDI-578 Fv fragment in solution is based on CHARMM36 topologies [142]. For MD simulation systems with trehalose and the best scored ligand (Compound X) from virtual screening, UCSF Chimera [157] was used to generate the atomic coordinates of the complex. Trehalose has been parameterised with the CHARMM carbohydrate force-field parameters [158] and Compound X was parameterised with CHARMM General Force Field [159]. All systems were solvated with TIP3 waters and NA ion molecules and CL ion molecules were added to the system to achieve 100 mM NaCl. The system was first minimised and equilibrated for 1 ns. The MD simulation systems were prepared and ran using GROMACS 2016.3 [140]. Subsequently, the MD simulation was ran at 2 fs/step with the use of periodic boundary conditions, constant temperature of 300 K and pressure at 1 bar using V-rescale thermostat and Parrinello-Rahman barostat respectively. Electrostatic and van der Waals interactions were calculated using a Verlet cutoff-scheme with Potential-shift-verlet as modifier with a cut-off of 1.0 nm. All MD simulation systems were ran in triplicates and snapshots were taken every 0.01 ns which is used for Root mean square fluctuations (RMSF) calculations and hydrogen bonds calculations using GROMACS.

RMSF measures of the deviation between the position of an atom and its reference position. RMSF is averaged over time and gives a value for each particle *i*.

$$RMSF_i = \left[\frac{1}{T}\sum_{t_j=1}^{T}\left(r_i(t_j) - r_i^{ref}\right)^2\right]^{½} \quad (2.2)$$

where T is the time over which one wants to average and r_i^{ref} is the reference position of particle *i*. Typically this reference position will be the time-averaged position of the same particle in the MD simulation.

2.3. Results and Discussions

2.3.1. Aggregation Prone Region Identifications

The major contributors of exposed hydrophobicity in MEDI-578 Fv fragment is the CDR3 loop (ILE102-LEU109) on the heavy chain with LEU105 has the highest overall SAP score (Figure 6). Other exposed hydrophobic patches include CDR1 loop (THR31), CDR2 loop (ILE52-PHE55) and the last β-sheet strand within the Fv fragment (MET123). In general, the light chain is more hydrophilic where most of its SAP values are negative. Exposed hydrophobic patches (ALA13, THR43) are located at the loops between β-sheet strands close to junction between Fv Fragment and CL domain. Solvent-accessible hydrophobic loops, including CDRs and other loops within the Fv fragment are flexible and hydrophobic residues interact with other molecules, therefore having a more positive score.

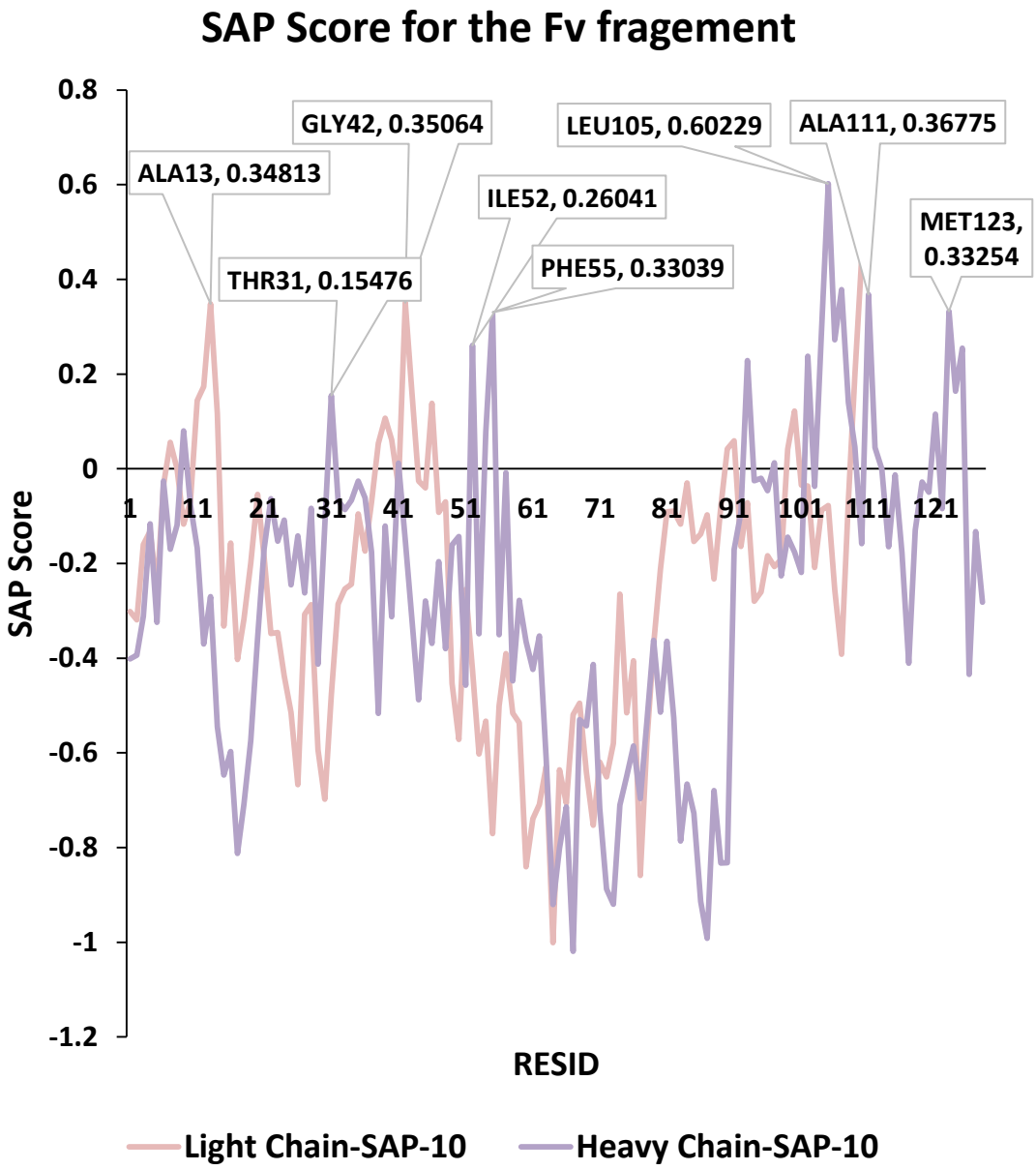


Figure 6: SAP calculation results at R=10 Å for MEDI-578 Fv fragment. Peaks with SAP scores greater than 0.15 were highlighted. LEU105 on the heavy chain has the highest calculated SAP score within the Fv fragment.

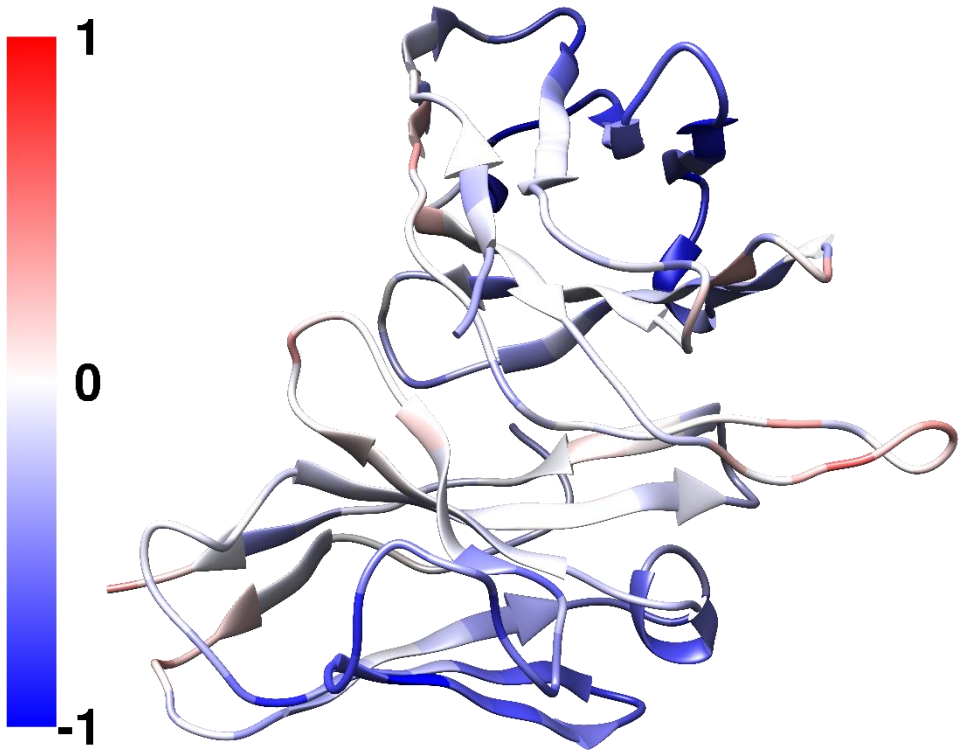


Figure 7: Visual representation of the SAP calculation results for MEDI-578 Fv fragment. The molecule is oriented with the heavy chain shown at the top of the model and the light chain at the bottom with CDRs at the right side.

During the 60 ns MD simulation, the aggregation propensities of each residues of MEDI-578 Fv fragment is similar to the static molecule with LEU105 within the CDR3 of the heavy chain show the highest aggregation propensities (Figure 8). For the heavy chain, the signal from CDR1 loop (THR31) become more positive after 25ns into the simulation with no significant changes in SAP score in other residues observed. In general, the light chain of MEDI-578 is more hydrophilic than the heavy chain with the signal from the hydrophobic regions surrounding THR43 and ALA13 become less intense after 40 ns compared to the static structure. Nevertheless, the CDR3 of the heavy chain has the strongest aggregation propensities within the whole molecule and it is the region of interest for designing specific in-process stabilisers.

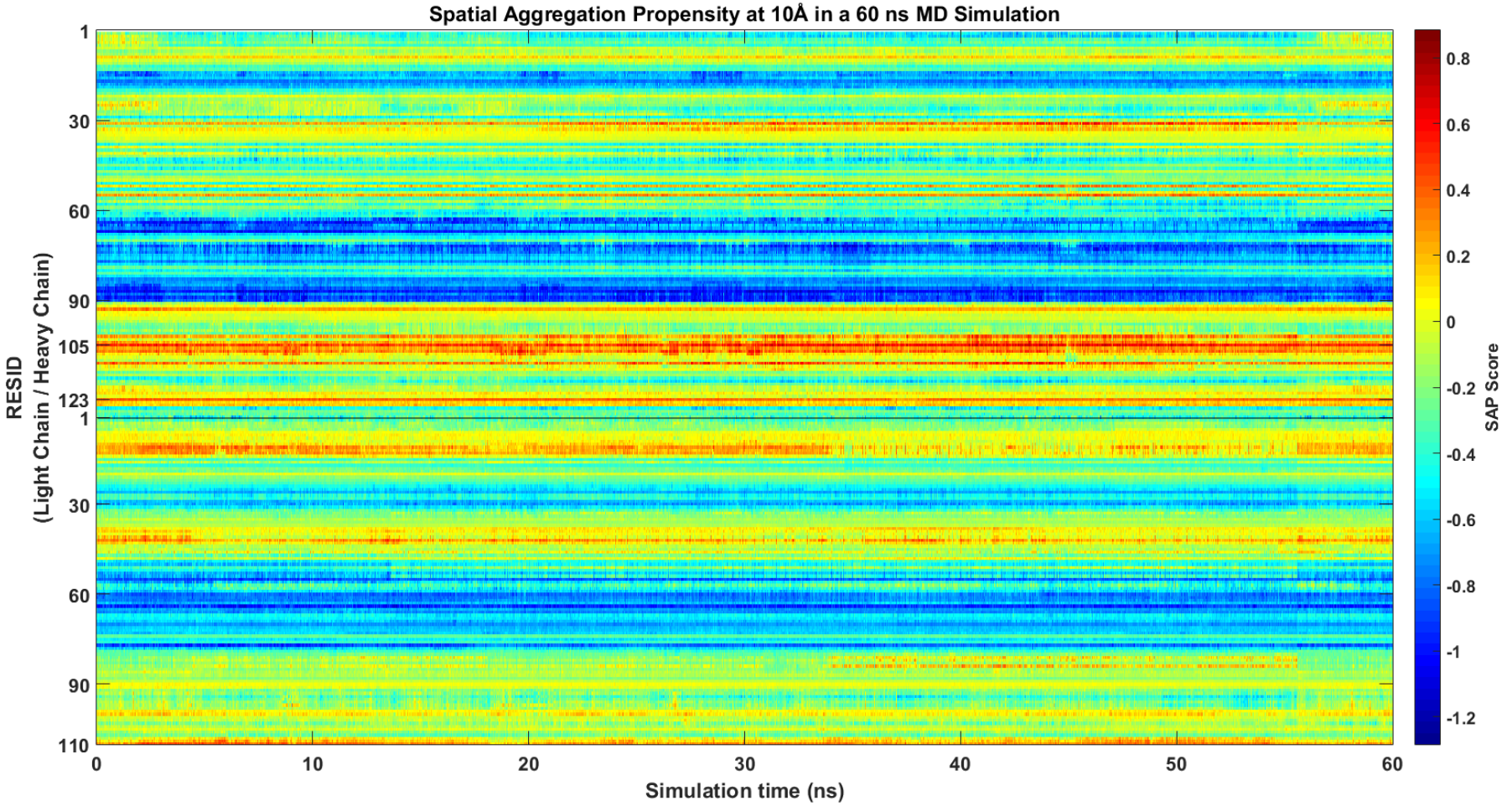


Figure 8: SAP values calculated at R=10 Å over a 60 ns MD simulation. The SAP score for each residue remained stable in a 60 ns all-atom MD simulation with LEU105 on the heavy chain identified as the most hydrophobic solvent-accessible patch on the Fv fragment.

2.3.2. ZDOCK

Figure 9 shows the top prediction in this study, two MEDI-578 Fv fragments were docked together successfully with the CDR3 of one molecule closely in contact with the loop at the side of the MEDI-578 Fv fragment heavy chain (ALA40 to LEU45). Figure 10 shows the representative results from these MEDI-578 Fv fragment complexes. Many of the top 2000 dimer structures show extensive contacts through the CDR regions of the Fv Fragment. Self-interaction through CDR3 (ILE102 to TYR113) and interaction between CDR3 and the β-sheet structure at the side of the Fv fragment of the heavy chain (GLU46 to ILE52) account for most of the interactions among these 2000 predictions (Figure 10). This is in agreement with the observation seen in the SAP studies. Typically, CDRs are solvent-accessible on the VH domain of an antibody structure to facilitate the binding of antibodies to its biological target. In absence of the biological target, these flexible solvent-accessible CDR loops may promote interaction between the molecules or even self-assembly.

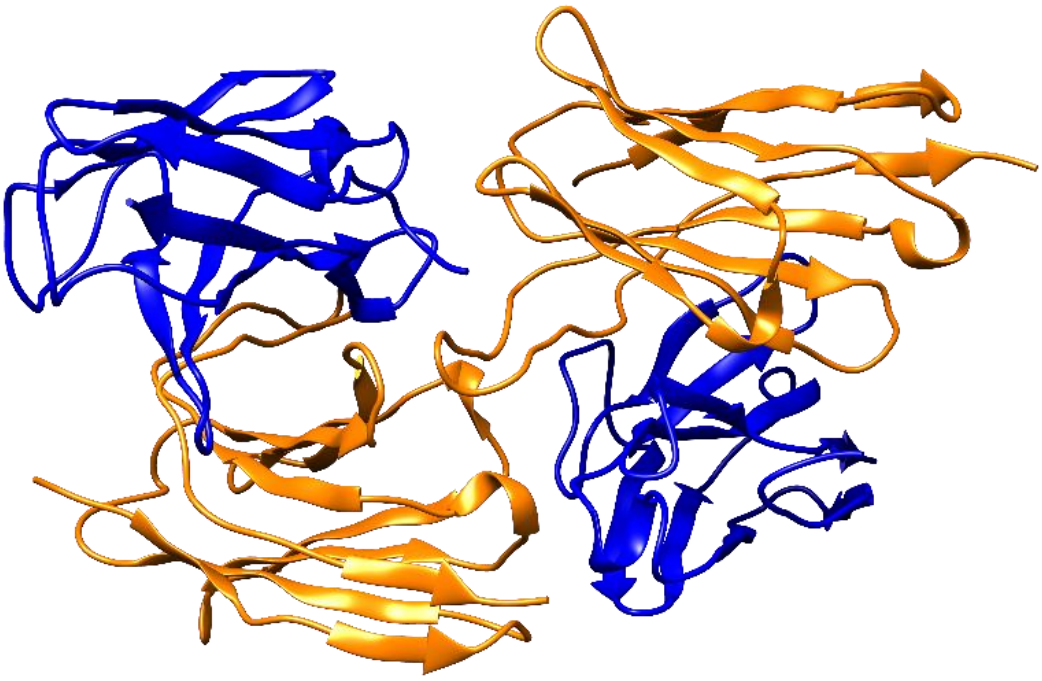


Figure 9: Best docked pose of MEDI-578 Fv fragment dimer generated by ZDOCK 3.0.2. The light chains are coloured in blue and the heavy chains are coloured in orange.

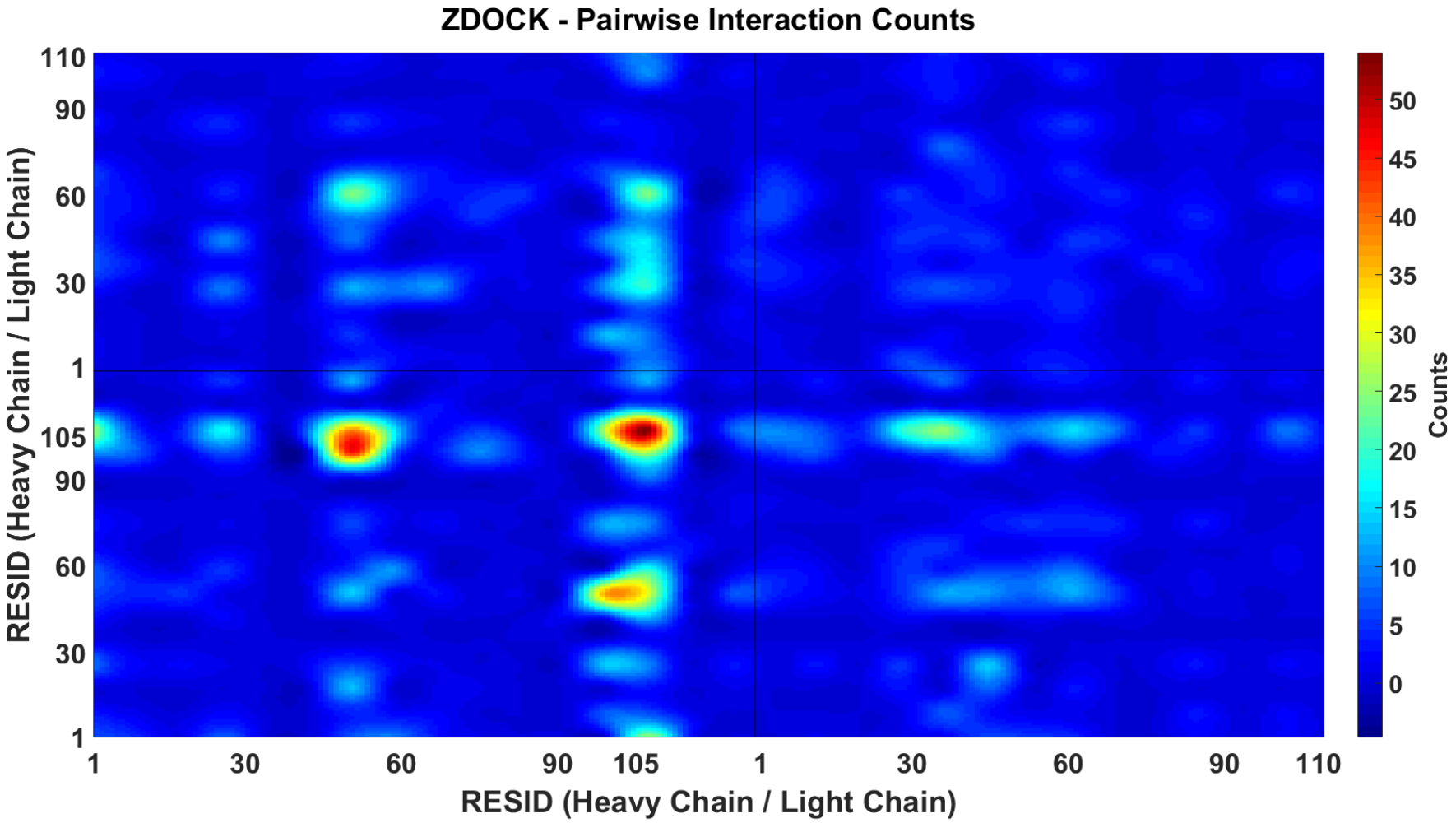


Figure 10: Top 2000 docking predictions of MEDI-578 Fv fragment dimer generated by ZDOCK 3.0.2. The area coloured in red denotes the binding pairs with most occurrence. Most of these dimer complexes shown interactions through LEU105 on the heavy chain.

2.3.3. Virtual Screening

Both SAP and protein-protein docking studies indicated that the CDR3 loop is the most aggregation-prone region on MEDI-578 Fv fragment but it is not an ideal candidate for mutagenesis because it is responsible for the binding of the flat β-sheet region of β-NGF so mutation of residues in this region may affect affinity to β-NGF.

Current small common excipients in protein formulations such as α,α-trehalose, sucrose and maltose have shown weak affinities towards the CDR3 of the heavy chain with α,α-trehalose achieving a greater affinity towards the CDR3 region compared with other disaccharides tested. Disaccharides are hydrophilic in nature and their calculated logP smaller than -3 [160], hence they do not exhibit high affinity towards the hydrophobic site via van der Waals interactions. For this reason, it is unlikely that disaccharides can achieve high specificity for the most aggregation-prone region on MEDI-578 Fv fragment. In addition, it has been reported that there are approximately 550 water-binding sites on antibody [161], and this could explain why a relative high molar ratio typically required to stabilise antibody formulations. In protein-protein docking study, it has shown to form a dimer through the CDR3 on the heavy chain.

A typical Gaussian distribution of predicted binding affinities was observed in the virtual screening with predicted binding affinities between -8.1 and -1.8 kcal/mol (Figure 11). N-{2-[3-(furan-2-yl)-6-oxo-1,6-dihydropyridazin-1-yl]ethyl}-2-[(6-oxo-1,6-dihydropyridin-3-yl)formamido]acetamide (Compound X, CPX) were found to bind to the CDR3 in a pocket between the heavy chain and the light chain with the highest affinity within the search. Compound X contain aromatic hydrophobic rings furan, pyridazine and pyridine. The docking results pointed out that residues ARG102, TYR103, ASP104, LEU109 and TYR112 on heavy chain and residues ASN32, TYR33, SER35, ASP51,

SER94 and TRP99 are involved in hydrogen bond and hydrophobic interactions with Compound X (Figure 12). The binding affinities calculated in this study is weak compared to typical antibody binding to its target [152], therefore it is reasonable to suggest that binding of these small peptides to the CDR3 on the heavy chain will not affect the binding of MEDI-578 to β-NGF as this is more energetically more favourable.

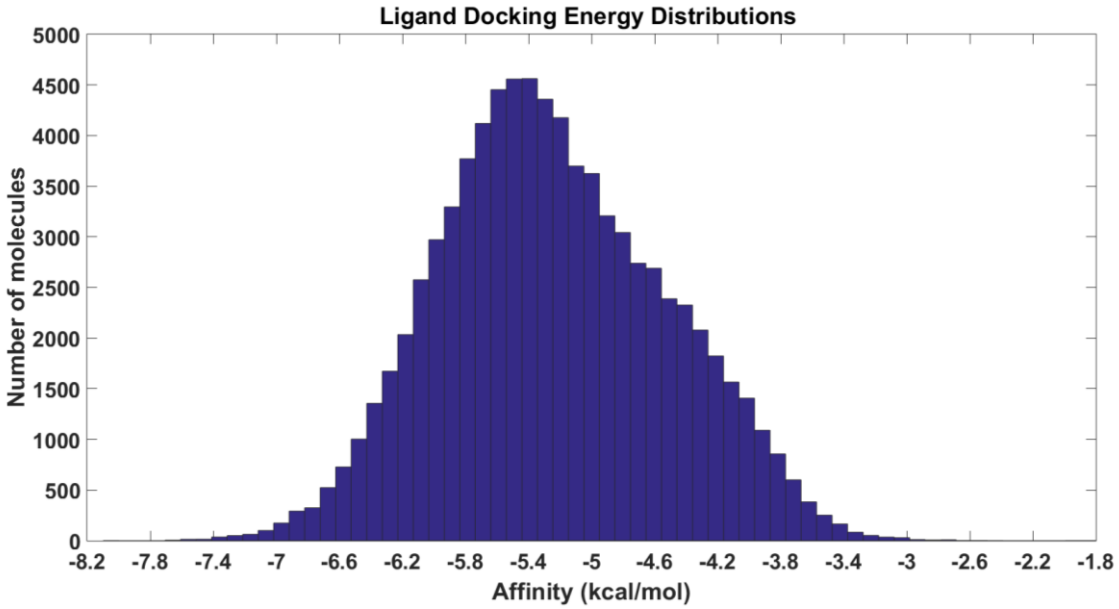


Figure 11: A typical Gaussian distribution of predicted binding affinities was observed in the virtual screening with predicted binding affinities between -8.1 and -1.8 kcal/mol.

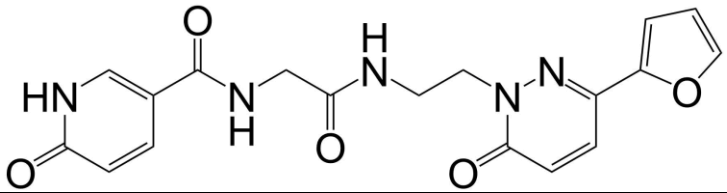
ZINC ID	ZINC000042384850
Molecular Formula	$C_{18}H_{17}N_5O_5$
Name	N-[2-[2-[3-(furan-2-yl)-6-oxopyridazin-1-yl]ethylamino]-2-oxoethyl]-6-oxo-1H-pyridine-3-carboxamide
Chemical Structure	
xlogP	-0.19
Molecular Weight	383.364 g/mol
H-bond donors	3
H-bond acceptors	10

Table 4: Summary of chemical properties of Compound X.

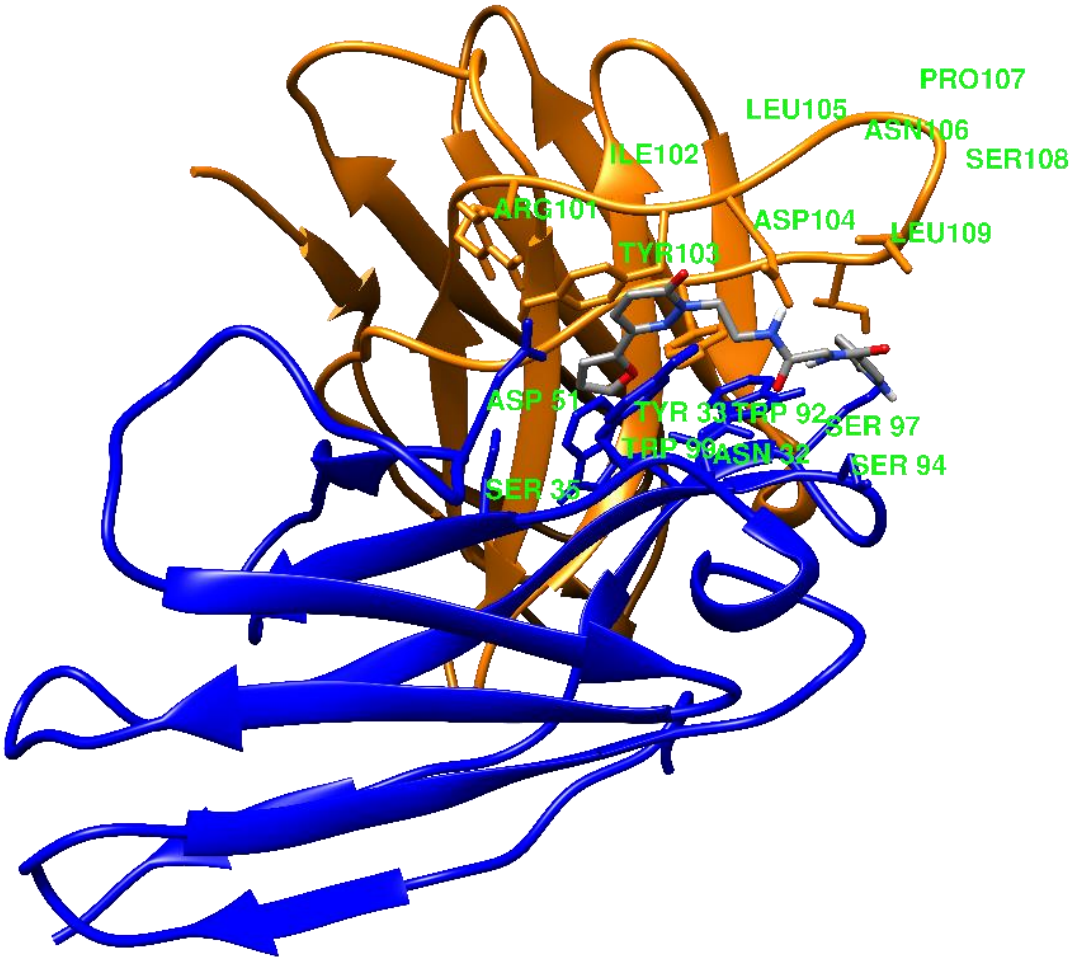


Figure 12: Best docking position of Compound X on MEDI-578 heavy chain CDR3. Compound X binds to a pocket located away from the most solvent-accessible hydrophobic residue LEU105 on heavy chain.

Compound	Predicted best binding affinity (kcal/mol)
α,α-Trehalose	-5.6
Sucrose	-4.8
Maltose	-4.8
CPX	-8.1

Table 5: Docking energies of other small molecules and mirroring peptide against the CDR3 of the MEDI-578 heavy chain. Binding of Compound X to MEDI-578 is more energetically favourable compared to disaccharides.

2.3.4. Molecular Dynamics

MD simulations were conducted to determine the structural stability within a short time scale for MEDI-578 Fv fragment, MEDI-578 Fv fragment-α,α-trehalose complex and MEDI-578 Fv fragment-CPX complex. RMSF is an important parameter determining structural flexibility of Cα atoms of each residue within a MD simulation.

The distance between the most hydrophobic residue, LEU105 on heavy chain and the excipient in the simulations were calculated (Figure 13). Across all 3 replicates of MEDI-578 Fv fragment-α,α-trehalose simulations, α,α-trehalose disassociates instantly and do not bind to any region of the MEDI-578 Fv fragment to form a stable complex with only a few transient interactions with the docked region was recorded whereas all 3 replicates shown CPX binds to the same pocket throughout the simulation.

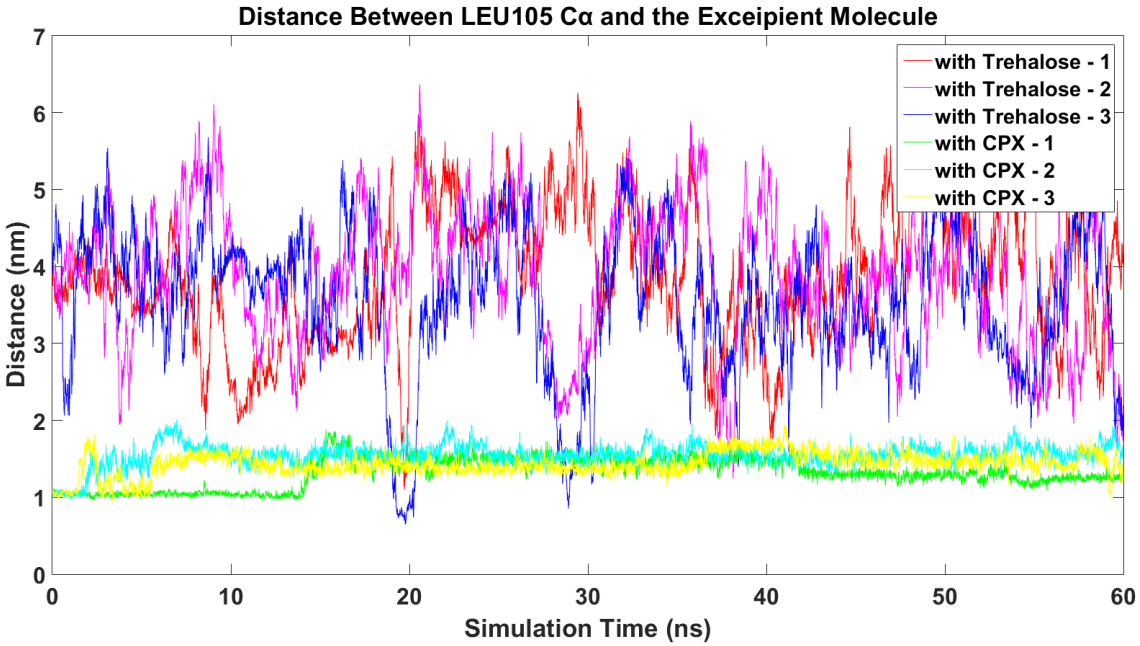
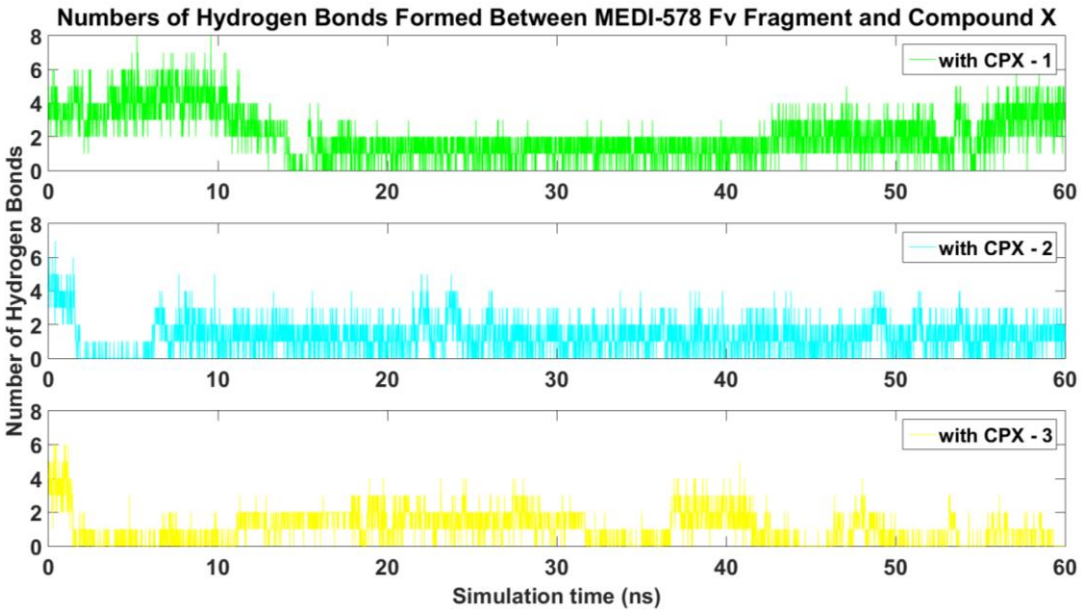


Figure 13: Time evolution changes in the distance between LEU105 and the excipient. Compound X stays in the same binding site throughout the 60 ns simulations while trehalose did not stay in close proximity to the most solvent-accessible hydrophobic residue LEU105 on heavy chain.

Hydrogen bonds are important in stabilising the ligand-protein complexes. A higher number of intermolecular hydrogen-bond interactions results in greater stability of ligand-protein complex. In all 3 replicates of MEDI-578 Fv fragment-CPX simulations, Compound X formed a maximum of eight hydrogen bonds with the MEDI-578 Fv fragment with at least two remained for most of the time (Figure 14a). In the cases of α,α-trehalose, hydrogen bonds were absent for most of the simulation time and only few intermittent hydrogen bonds formed (Figure 14b).

As previously discussed, aggregation is determined by the dynamic exposure of the aggregation prone region. As shown in Figure 15a, all MD simulations showed a similar RMSF profile with main fluctuations occur in the loops of the MEDI-578 Fv fragment as they connect β-sheets of the antibody structure. The results also suggested that the RMSF of the CDR3 on heavy chain is the most flexible part of the structure. The RMSF of MEDI-578 Fv fragment-CPX complex is lower than that of the MEDI-578 Fv fragment alone and MEDI-578 Fv fragment-α,α-trehalose complex. Hence, the lower RMSF values of the MEDI-578 Fv fragment-CPX complex suggested reduced conformational changes. The RMSF values of MEDI-578 Fv fragment-α,α-trehalose complex is higher than MEDI-578 Fv fragment-CPX complex because α,α-trehalose do not form a stable complex with MEDI-578 Fv fragment on the CDR3 loop on the heavy chain so stabilising effect is not observed. The short 60 ns MD simulations used in this study did not show much similarities across different replicates. Therefore studying the binding of excipients to antibody molecule may require significantly longer simulation time to achieve a good convergence. Nevertheless, the data presented here are sufficiently robust to provide an overview of how these excipients interact with a solvent accessible hydrophobic area on a mAb.

a)



b)

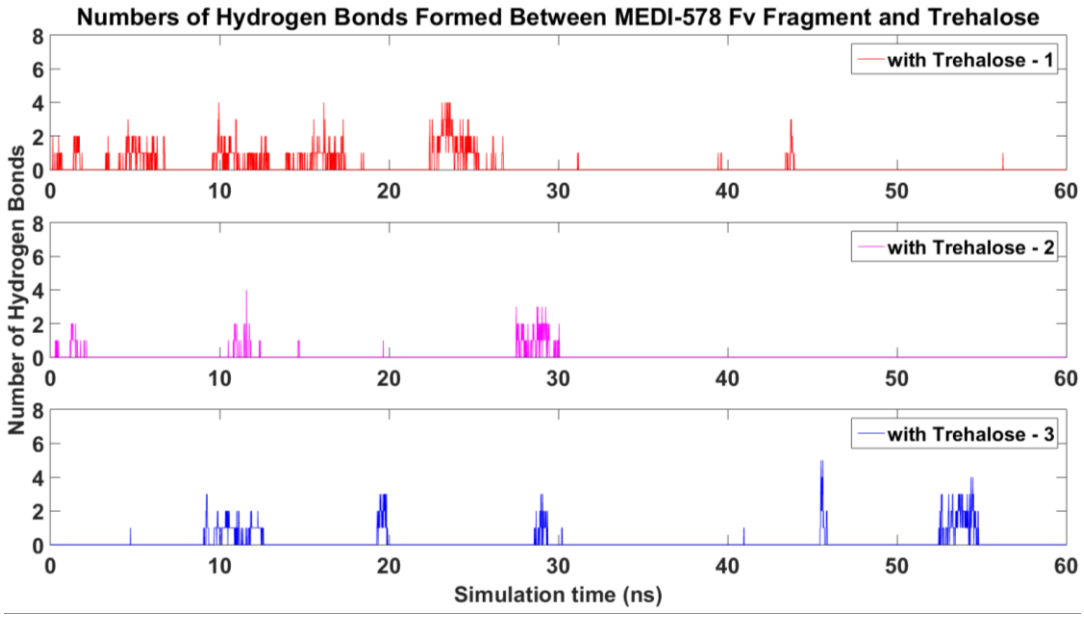
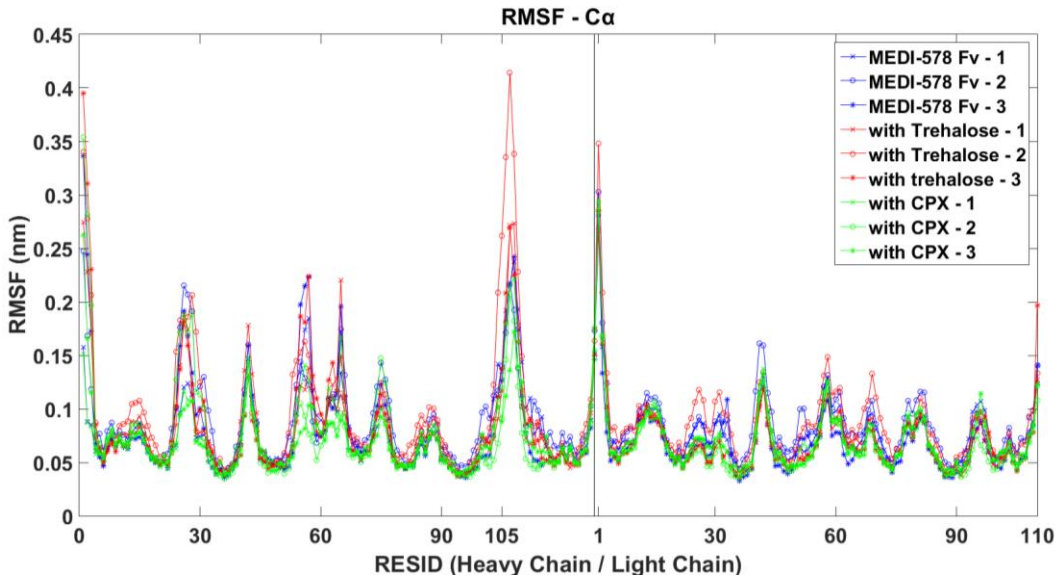


Figure 14: Hydrogen bond formation in MEDI-578 Fv fragment complexes as a function of time. a) Compound X, b) α,α-trehalose. Compound X forms hydrogen bonds with the MEDI-578 Fv fragment constantly whereas only intermittent hydrogen bonds were recorded. This suggests Compound X and MEDI-578 Fv fragment form stable complexes between but not trehalose.

a)



b)

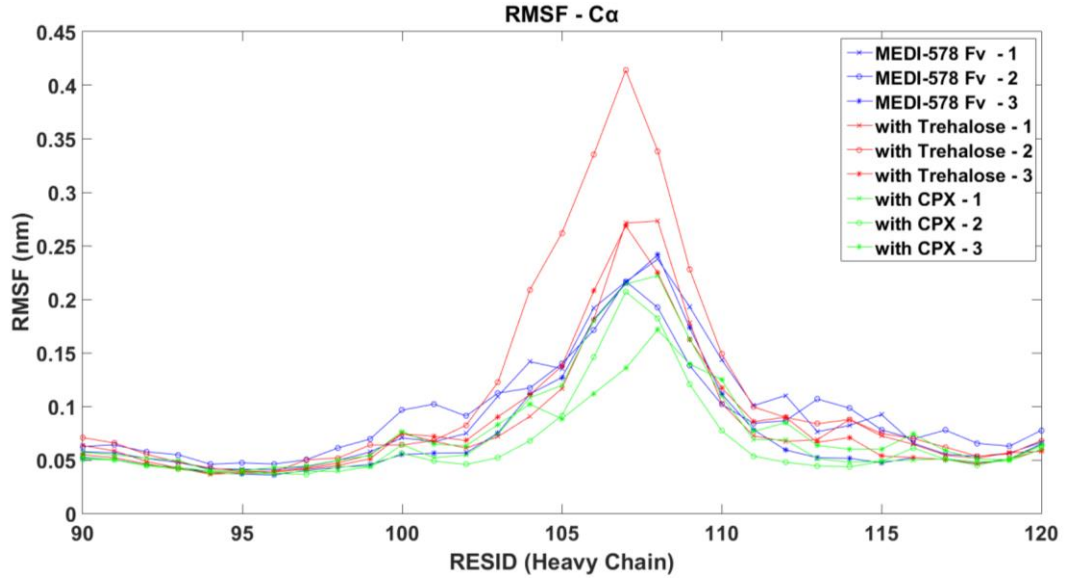


Figure 15: RMSF of Cα atoms of MEDI-578 Fv fragment complexes as a function of time. a) Overall RMSF, b) RMSF of the CDR3 on heavy chain. Fluctuations of the CDR3 loop increases with trehalose whereas fluctuations decreases with Compound X docked to the MEDI-578 Fv fragment.

2.4. Conclusions

This study provided insights about the interaction details of aggregation-prone region of an antibody and excipient binding using a set of compounds selected from ZINC database with both molecular docking and molecular dynamics simulations. There are three discrete findings observed in this study. First, *in-silico* data suggests Fv fragment- Fv fragment interactions are steered by hydrophobic interaction through the CDR. Second, commonly used disaccharides, such as α,α-trehalose, have weak binding affinity towards hydrophobic areas on Fv fragment of an antibody. Third, the CDR loops are very flexible and a compound selected from the ZINC database through virtual screening achieved a higher binding affinity towards the hydrophobic CDR loop and it provided the required features to stabilise the CDR loop. Further experimental evidence is required to confirm the hypothesis laid at the beginning of this chapter that potent small molecule which binds to surface accessible hydrophobic part of the mAb and stabilised the region can prevent antibody-antibody interactions and delay aggregation process *in-vitro*. Nevertheless, this study provided valuable *in-silico* contributions to the investigation of to design new excipients for therapeutic antibodies.

3. Experimental Verification of Compound X

3.1. Introduction

As already mentioned, protein aggregation is one of the most challenging problems for developing stable formulations of antibodies. The self-association of mAbs is directly related to aggregation which will affect the quality of final product in terms of its efficacy and safety [85-87]. In a liquid mAb formulation, the native mAb structure can be affected by a number of physical and chemical stresses such as heat, mAb concentration, ion concentrations and buffer systems which can affect the formation of irreversible aggregates. High concentration mAb formulations are becoming increasingly important in the pharmaceutical industry as they are often required for SC and IM administration to allow self-administration by patients. For high concentration mAb formulations, protein-protein interactions can adversely affect solubility, viscosity and aggregation [53]. In industrial practice, excipients such as disaccharides are added to formulation to reduce aggregate formation [54].

There are a range of analytical techniques that are used to monitor for signs of aggregation in mAb formulations, such as Dynamic Light Scattering (DLS), Size-exclusion Chromatography (SEC), Analytical Ultracentrifugation, Asymmetric field flow fractionation, Circular Dichroism (CD), Turbidimetry and Differential Scanning Fluorimetry [162]. Size exclusion chromatography (SEC) is the most commonly used analytical technique for quantifying and sizing soluble aggregates between 1 nm to 50 nm. SEC is a well-established technique for separations of mAb monomers and aggregates by size. It is worth noting that the British Pharmacopoeia recommends SEC for detection of dimers and higher order aggregates for some protein biopharmaceuticals [163]. Separation in SEC depends on pore size of the column and the retention time

depends on the efficiency of individual molecule penetrating through the pores of the stationary phase. Small molecules can penetrate into the pores so their retention time in the column is longer whereas larger molecules are excluded from pores and flow directly. Dynamic light scattering (DLS) is another commonly used technique to estimate the size and their hydrodynamic radius. Viscosity of a therapeutic antibody formulation is governed by protein-protein interactions. Connoly and co-workers have shown that viscosity at high concentrations can potentially be predicted using the interaction parameter k_D measured by DLS with concentrations between 1 and 10 mg/mL [164].

The previous chapter described the *in-silico* results of identification of the most aggregation prone region in MEDI-578 Fv fragment and Compound X (CPX) was identified as a potential excipient that can stabilise the dynamic fluctuation of such regions. Research presented in this chapter builds upon the conclusions of the previous chapter, and evaluates these *in-silico* data with SEC and DLS to study the antibody-antibody interactions and more importantly the long-term stability of antibody formulations of MEDI-578 in the presence of Compound X compared with α,α-trehalose and the original formulation as control.

3.2. Materials and Methods

3.2.1. Sample Preparation

MEDI-578 was obtained from MedImmune (Cambridge, United Kingdom), D-(+)-trehalose dehydrate, phosphate buffered saline and glacial acetic acid were purchased from Sigma-Aldrich (Poole, United Kingdom), Compound X was purchased from Life Chemicals Europe GmbH (Munich, Germany), sodium acetate trihydrate and sodium chloride were purchased from Fisher Chemicals (Loughborough, United Kingdom).

MEDI-578 was supplied in 50 mM sodium acetate, 100 mM sodium chloride at pH 5.5. Absolute concentration was determined by measuring the absorbance at 280 nm in triplicate using NanoDrop (ThermoFisher Scientific, Loughborough, United Kingdom) calculated with an extinction coefficient of 1.68 mL mg^{-1} cm^{-1} prior to use. Formulations of MEDI-578 with α,α-trehalose and with Compound X were prepared by dissolving the corresponding excipient in freshly prepared buffer and added to the MEDI-578 stock formulation to achieve an Antibody:Excipient molar ratio of 1:5. Subsequently, all formulations were prepared with pH adjustment followed by filtering through 0.22 μm sterile filters.

3.2.2. Dynamic Light Scattering

The diffusion interaction parameter, k_D for an antibody, was determined by dynamic light scattering (DLS) with dilute formulations of MEDI-578 ranging from 1 – 10 mg/mL and calculated according to the following equation:

$$D = D\_0\ (1 + k_D c) \qquad (3.1)$$

where D is the diffusion coefficient at a given protein concentration c and D_0 indicates the diffusion coefficient when c is at 0.

For each measurement, the diffusion coefficient, D, was determined by DLS and plotted against the concentration of antibody. The slope and the Y-intercept are determined by a linear regression model using the D values at various concentrations. The value of k_D was then calculated by dividing $k_D D_0$ (the slope) by D_0 (the Y-intercept).

The DLS measurements were made on a DynaPro PlateReader II system (Wyatt Technology, Santa Barbara, California, United States) with an 826.1 nm laser using a 96-well non-binding flat-bottom plate (Greiner Bio-One, Kremsmünster, Austria). Formulations were diluted with filtered fresh buffer for DLS measurements to antibody concentrations of 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 mg/mL. Samples of all 3 MEDI-578 formulations in each concentration was added to quadruplicate wells and each well contained 100 µL of the sample. Twenty acquisitions were collected for each well over 5 s. Only particles with hydrodynamic radius between 2 nm and 30 nm were considered for analysis to filter out signals from salts and dust.

3.2.3. Size-exclusion Chromatography

Samples of all 3 MEDI-578 formulations for SEC measurements were prepared by first diluting to achieve antibody concentrations of 1 mg/mL and were stored at 50°C for 28 days in glass vails. Measurements were taken immediately after sample preparation, and then after 7 and 28 days of storage. SEC was performed on an Agilent ZORBAX Bio Series GF-250 4 µm column (Agilent Technologies, Santa Clara, California, United States) on the Agilent Technology 1200 series HPLC system (Agilent Technologies, Santa Clara, California, United States) at a constant flow rate of 0.5 mL/min with a sample loading of 20 µL and UV detection at 280 nm. The mobile phase consisted of phosphate buffered saline (pH 7.4).

3.3. Results and Discussions

3.3.1. Determination of k_D by Dynamic Light Scattering

DLS measurements of MEDI-578 in 3 different formulations at 10 different concentrations ranging from 1 to 10 mg/mL was performed. The diffusion coefficients for each formulation were plotted against antibody concentration and $k_D D_0$ and D_0 were

determined with regression analysis, and the resulting k_D was calculated by dividing $k_D D_0$ with D_0. In general, smaller antibody monomers diffuse faster than larger particles such as antibody aggregates and result in a lower diffusion coefficient. Hence, a negative k_D indicates net attractive interactions between antibody molecules whereas a positive k_D indicates net repulsive interactions between antibody molecules.

The concentration-dependent diffusion coefficient of all MEDI-578 formulations are shown in Figure 16. The addition of NaCl salts to an antibody formulation in acetate buffer reduces the electrostatic interactions between antibody molecules. The effect of ionic strength at high concentration (> 100 mM) on k_D values were much smaller [165]. In the present case, the addition of NaCl to MEDI-578 can prevent changes in colloidal stability due to attractive or repulsive electrostatic interactions.

k_D values for MEDI-578 were negative in all formulations, suggesting a net attractive force exists between MEDI-578 molecules across this concentration range. The k_D for original MEDI-578 formulation in 50 mM sodium acetate, 100 mM sodium chloride at pH 5.5 was determined to be -9.03 mL/g. The addition of trehalose at a MEDI-578:trehalose at molar ratio of 1:5 resulted in a k_D increase to -6.33 mL/g which suggests that extended intermolecular attractions are reduced compared with the original MEDI-578 formulation. As the protein-protein interactions reduce, the viscosity of the solution reduces and likely indicate a reduced aggregation rate. The presence of Compound X at MEDI-578:CPX molar ratio of 1:5 also resulted in an increase of the k_D value. However the determined value for this formulation was -6.29 mL/g which is very similar to the determined k_D value for the formulation with trehalose. Therefore it is reasonable to suggest that the addition of Compound X or trehalose can reduce the attractive

interactions between MEDI-578 molecules at a low antibody:excipient ratio and leaves sufficient room for other excipient should the formulation require (i.e. surfactants).

The current DLS study revealed the strength of MEDI-578-MEDI-578 interactions in the presence of Compound X and trehalose. The *in-silico* data from the previous chapter suggested that Compound X stabilised the CDR3 on the heavy chain. However, this study does not provide information regarding how Compound X or trehalose affects the conformational stability of MEDI-578 compared to buffer-only conditions. Moreover, trehalose is commonly added to formulations at a much higher molar ratio to prevent aggregates formation [54] so a higher concentration may be required to achieve optimum k_D values.

	$k_D D_0$	D_0 (cm^2/s)	k_D (mL/g)
MEDI-578	-4.05E-06	4.48E-07	-9.03
MEDI-578 with Compound X	-2.82E-06	4.48E-07	-6.29
MEDI-578 with Trehalose	-2.85E-06	4.50E-07	-6.33

Table 6: Resolved diffusion interaction parameters and diffusion coefficients determined by DLS. The measured k_D values are negative suggesting net attractive interactions between antibody molecules. Addition of either Compound X or trehalose cause the k_D values become less negative, suggesting reduction of self-interactions of MEDI-578 in prescience of these molecules.

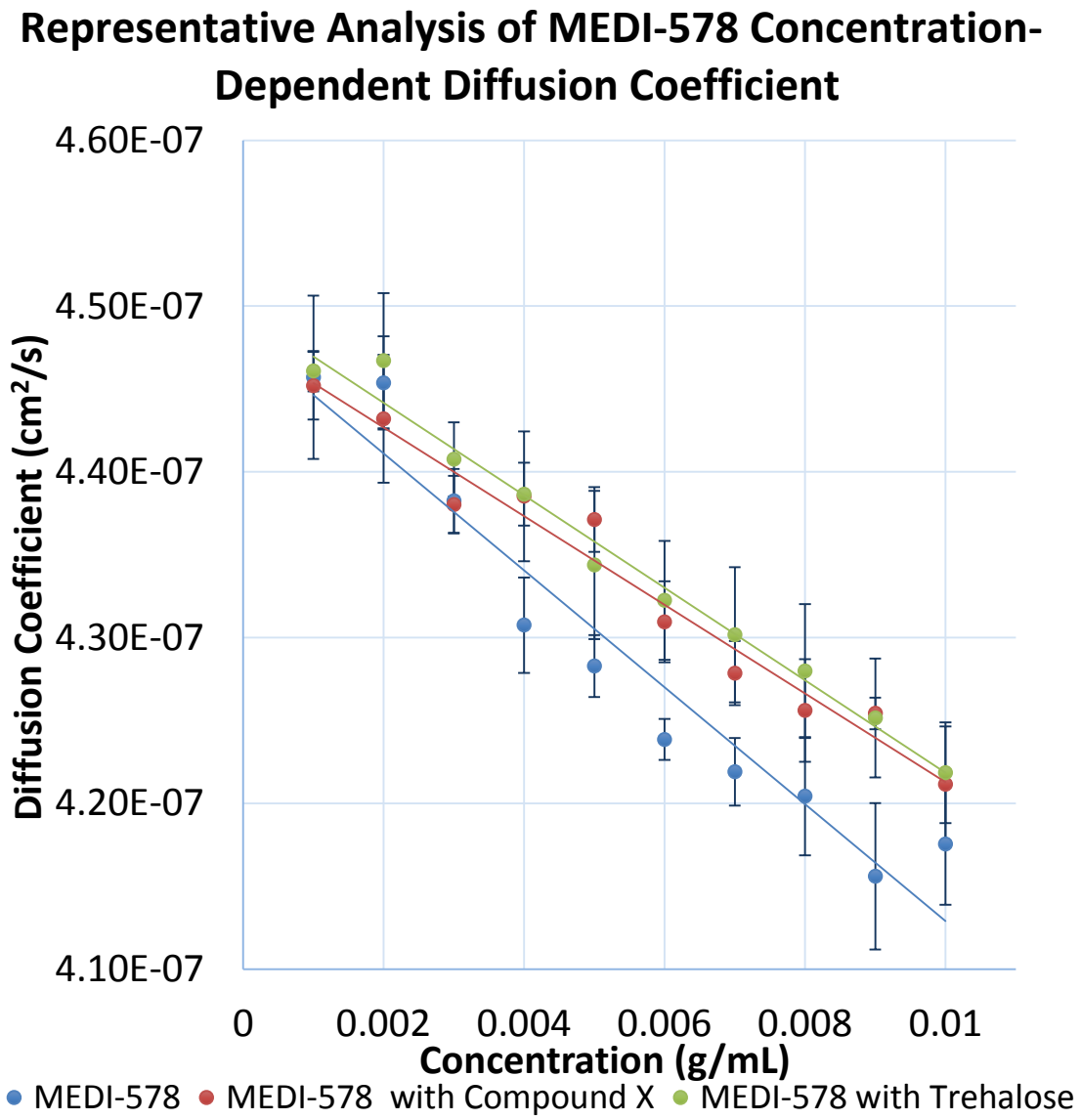


Figure 16: Change in diffusion coefficient versus concentration of MEDI-578 for a) native buffer conditions, b) with Compound X at 1:5 ratio, c) with Trehalose at 1:5 ratio. The diffusion coefficients decreases as the concentration increases in all three formulations of MEDI-578 suggesting attractive interactions between MEDI-578 molecules in all three formulations.

3.3.2. Aggregation Profile by Size-exclusion Chromatography

All chromatograms showed MEDI-578 eluted after 4.8 minutes. Compound X also showed absorbance at 280 nm and was eluted after 7.6 minutes while trehalose was not detected at this wavelength. All three formulations showed that there were no differences in the relative content of monomer and higher molecular weight aggregates. Incubation of the three MEDI-578 formulations at 50 °C for 7 days and 28 days also resulted in no difference in higher molecular weight aggregates. The experiments described here were

designed to assess the stability of MEDI-578 during storage at high temperature. MEDI-578 in in 50 mM sodium acetate, 100 mM sodium chloride at pH 5.5 appeared to be very stable at 50 °C and the addition of Compound X or trehalose do not appear to destabilise MEDI-578.

In therapeutic mAb formulations, storage at 50 °C does not appear to cause significant aggregation. Hernández-Jiménez and co-workers have stressed five marketed therapeutic mAbs products, Avastin, Erbitux, Remicade, MabThera and Herceptin, at 50 °C for 24 hours. Out of these products, Erbitux displayed significant aggregation after storage while the amounts of aggregates in other products were much lower [166]. The degradation kinetics may be dependent on the amino acid sequence, the three-dimensional structure or the formulation compositions. This study used a 1 mg/mL for conducting all these studies under thermal stress which is a lot less concentrated than most therapeutic mAbs products so the aggregation is less prominent.

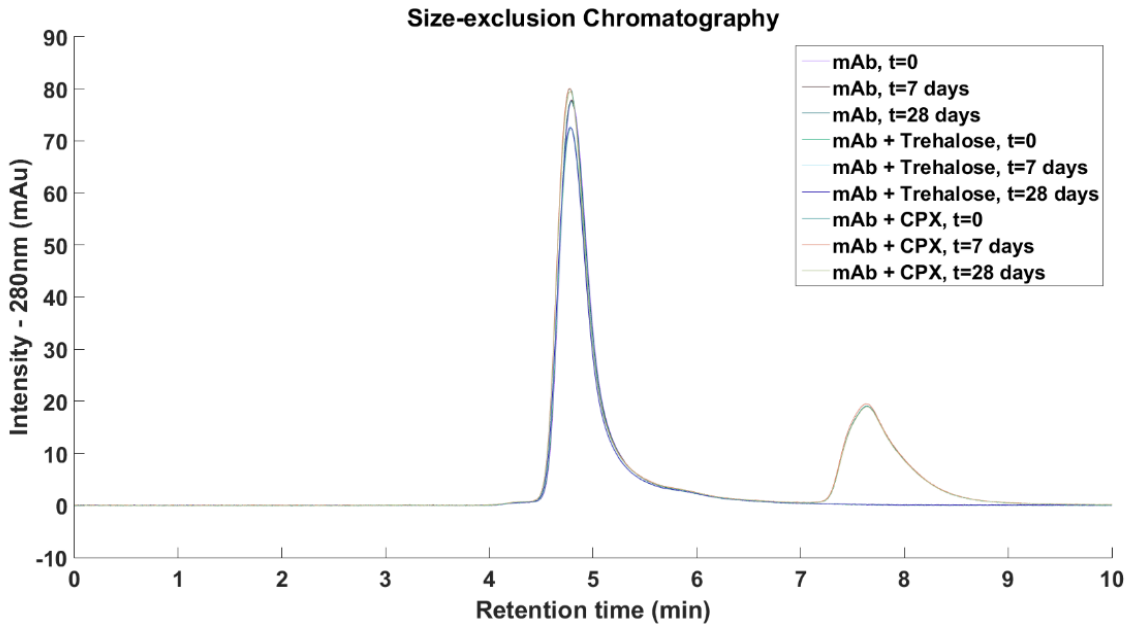


Figure 17: Representative size-exclusion chromatograms of different MEDI-578 formulations at different time points. MEDI-578 eluted after 4.8 minutes and Compound X eluted after 7.6 minutes. Only a single antibody peak was seen in all tested samples with no higher molecular weight aggregates were detected, suggesting MEDI-578 was stable at elevated temperature.

3.4. Conclusions

In summary, Compound X was tested in formulations of MEDI-578 and its effect as an anti-aggregation excipient compared with trehalose and MEDI-578 in 50 mM sodium acetate, 100 mM sodium chloride at pH 5.5 as a control. Unfortunately, the experimental data that emerge from the application of Compound X is confusing. A reduction of MEDI-578-MED-I578 attractive interactions have been seen in the presence of both Compound X and trehalose using DLS. The extent of the effect of Compound X, a potent small molecule that binds to surface accessible hydrophobic part of the mAb, do not exhibit significantly difference to traditional non-specific excipient trehalose. The native formulation of MEDI-578 at 1 mg/mL has proven to be a very stable and no higher molecular weight aggregates were found after 28 days storage at 50 °C. A more concentrated solution may be need to observe the formations of higher molecular weight aggregates. Therefore, the hypothesis that low molecular weight small molecule that binds to surface accessible hydrophobic part of the mAb can delay aggregation process cannot be verified experimentally. Since both Compound X and trehalose do not appear to destabilise MEDI-578 under thermal stress over the same period, it is reasonable to suggest that by disrupting antibody-antibody interactions by bind to the solvent accessible hydrophobic regions of an antibody can act as good as general excipient like trehalose.

4. Coarse-grained MD Simulations to Determine Antibody-antibody Interactions

4.1. Introduction

In Chapter 2, ZDOCK was used to study the formation of MEDI-578 Fv fragment dimer complex and the results suggested the CDR3 on the heavy chain is most likely accounted for most Fv fragment-Fv fragment interactions. Along with the SAP calculations, the CDR3 on the heavy chain was selected for virtual screening and Compound X was proven to reduce the fluctuation of this region in a short MD simulations while trehalose showed low affinity towards this area of the antibody and was unable to reduce the fluctuation. However experimentally, Compound X did not show a superior effect in terms of disrupting antibody-antibody interactions over trehalose. It is possible that there are multiple binding sites for disaccharides on the three-dimensional structure of an antibody responsible for antibody-antibody interactions [54]. Some of these binding sites may play a more important role in the formation of mAb-mAb complex than the most hydrophobic region on the structure. In addition, non-specific interactions as well as conformational changes of the mAb msy also contribute to mAb-mAb interactions in a formulation. Therefore a better understanding on how excipients may affect mAb-mAb interactions at the molecular level will allow a different and more effective approach to excipient design.

Proteins are dynamic in nature and they have a massive and complex energy landscape. ZDOCK is only a FFT-based rigid-body docking tool where the dynamic nature of the mAb was not considered. Docking methods typically rely on scoring function to identify the energetically most favourable conformation based on experimental or bioinformatics data. For this reason, docking methods may not be readily applicable to heterogeneous environments like a mAb formulation which contains a number of excipients.

MD simulations is an alternative computational approach to study interactions of molecules. However, it is extremely computationally expensive to compute the protein-protein interaction of two Fv fragments molecules with all atom force fields in an explicit solvation box with unbiased initial configurations. In Chapter 2, short 60 ns MD simulations were used to study the fluctuation of the CDR-3 on the heavy chain with the tested excipient docked to the region. If Compound X was not docked to the area at the beginning of the simulation, it would require a much longer simulation time to reach the site as components of the simulations can diffuse freely and there may be other local regions may exhibit affinity towards Compound X. A suitable simulation time-scale should be used for the evolution of protein complexes to cover as much of the energy landscape as possible. All atom explicit solvent models for MD simulations of such length are not possible with current computational power. An alternative MD simulation method was needed that could be used to give a suitably productive description of dimerization.

In this chapter, the formation of Fv fragment-Fv fragment complex is examined further using coarse-grained MD simulations in the presence of Compound X and trehalose to determine if these low molecular weight molecules can recognise the solvent accessible hydrophobic area of the mAb and prevent antibody-antibody interactions. It is also hoped that it can be shown that Compound X can reduce the local fluctuations of solvent accessible hydrophobic area to prevent formation of antibody dimers. DAFT was used to generate replicates. It is hoped these types of computational studies can be used to gain a more in-depth understanding of the molecular events, the mechanisms of Fv fragment-Fv fragment complex formations can be explained.

4.1.1. MARTINI Force Field

One way to decrease the degree of freedom in a MD system is through coarse-graining (CG) where a simplified representation is used to describe the system. These simplified representations reduce the computational demands significantly, allowing for simulating large and complex systems using a longer time-scale. The MARTINI force field is a popular CG force field for biological macromolecules and was first described by Marrink and co-workers in 2007 [167]. Initially the MARTINI force field was developed for lipids and it is now extended to a large range of biomolecular systems such as proteins, DNA, RNA, carbohydrates and polymers [168].

The philosophy of the MARTINI model is based on a four-to-one mapping, where four heavy atoms and their associated hydrogen atoms are represented by a single predefined bead. There are four main types of beads- polar (P), non-polar (N), apolar (C) and charged (Q). The beads are further subdivided into subtypes based on the hydrogen-bonding (d:donor, a:acceptor, da:both, 0:none) or the degree of polarity (1 to 5 in ascending order of polarity), yielding a total of 18 bead types (Table 7) [168]. Each bead type represents certain chemical functional groups of the CG force field and were validated against thermodynamic data from oil/water partitioning coefficients [167]. Following the same philosophy, water is parameterised as four water molecules mapped to a single CG water bead and ion beads represents the ion molecule along with the hydration shell. For ring-like fragments in amino acids, such as imidazole, benzene, phenol, indole and pyrrolidine, these are mapped with higher resolution up to two-to-one mapping with S bead [169-170].

The size of the beads and the strength of the interaction are determined by a Lennard-Jones potential and the R_{min} for normal beads and S beads are set at 0.47 nm and 0.43 nm

respectively. The ε for the S beads are scaled to 75% of the value of the normal beads to account for their smaller bead size and interaction forces. The large reduction of the degrees of freedom and simulations of proteins with MARTINI force field generally are stable at 20 fs, offering the possibility to study large, complex and slow protein systems.

MARTINI Bead Types			
Polar	**Non-polar**	**Apolar**	**Charged**
P1	Nda	C1	Qda
P2	Nd	C2	Qd
P3	Na	C3	Qa
P4	N0	C4	Q0
P5		C5	

Table 7: All 18 bead types in MARTINI model. There are four main beads types, polar (P), non-polar (N), apolar (C) and charged (Q). The beads are further subdivided into subtypes based on the hydrogen-bonding (d:donor, a:acceptor, da:both, 0:none) or the degree of polarity (1 to 5 in ascending order of polarity).

The topology for MARTINI model of the target molecule is generated through mapping the atomistic chemical structure to the corresponding CG bead types. Parameters for the MARTINI model can be derived from the underlying atomistic geometry or based on atomistic simulations data. When atomistic simulations data is used, the centre of masses of the corresponding atom groups are identified and converted into a mapped CG simulation. The distribution functions for each bonded terms are calculated and the CG parameters are optimised through comparing and systematically modification of individual terms until satisfactory agreement of the distribution functions is achieved.

4.1.2. Elastic Network

MARTINI force field trades atomic-resolution detail for computational cost reduction. The challenge is to reproduce accurately the structural and the dynamical properties of

the system studied. As previously mentioned, the philosophy of MARTINI model is based on mapping of a few heavy atoms and associated hydrogen atoms into a predefined bead. Hydrogen atoms accounts for roughly half of the atoms in a molecule of mAb and they are crucial for hydrogen bonds formations and hence defines the three-dimensional structure and the dynamics of the protein. ElNeDyn is an elastic network model of MARTINI protein force field that maintains the structural scaffold of the protein [171]. Within the ElNeDyn model, the overall structure of the protein is described as a network of point masses connected by springs coupled with harmonic oscillators with the cut-off distance between the point masses and force constant of the harmonic oscillators are defined. It is aimed to maintain the overall structure of the protein and to reproduce dynamical properties such as residue fluctuations as observed in atomistic simulations. The reported structure of MEDI-578 only contains the Fv fragment of the antibody [152]. ElNeDyn is particularly an advantageous approach for simulating the Fv fragment of MEDI-578 to prevent unfolding due to the lack of constant regions of the MEDI-578 Fv fragment model.

4.1.3. DAFT

Single protein molecule molecular dynamics simulations are widely used and are described in most published work. However such studies are rarely conducted with multiple protein molecules. For studying protein complexes and their formation without any physical structural information from X-ray crystallography or NMR, two protein molecules can be docked together before running a long atomistic MD simulation [172]. This will of course generate specific data on the protein complexes but that will only represent one of the possible orientations that two protein molecules can interact. In reality, protein molecules move freely in solution and interact with adjacent molecules in all possible orientations reversibly before forming a stable complex. It is important in

understanding the last stages in the formation of protein complexes, especially how the protein surface affects the recognition of the binding site.

There is a growing interest to place two protein molecules at a defined distance in a MD simulation allowing enough room for random translational and rotational positioning of one molecule in relation to other molecule to study dimer-complex formation [173-175]. These studies have found that using a single long simulation to study complex formation is unfeasible as once complex is formed, the formed complexes are stable for the remainder of the simulations. The disassociation rate may be too low to allow unbinding and binding events to occur hence sampling for the energy landscape is deemed as insufficient. Therefore it is more beneficial to perform a large number of MD simulations with random initial orientations at non-interacting distances to allow equal *a priori* probabilities of encounter so that the probabilities of complex formation at different orientations can be better assessed. Wassenaar and co-workers have developed the Docking Assay For Transmembrane components (DAFT) to aid setting up and performing multiple simulations to identify protein-protein interactions and binding orientations [176]. DAFT first converts the all-atom model into to MARTINI representation with martinize [170]. Then the molecular systems are built by adding protein molecules with random relative orientations. Subsequently, the molecular systems are solvated and ionised with insane [177]. The prepared systems are compatible with martinate [178], which performs energy minimisation, position restrained NVT equilibration, position restrained NPT equilibration and production run.

4.2. Materials and Methods

4.2.1. Parameterisation of Compound X

Atomistic topology of Compound X is obtained with Automated Topology Builder 2.2 (ATB) [179]. Reference atomistic simulation was performed using the GROMOS 54A7 force field [143]. The MD simulation system was prepared and ran using GROMACS 2016.3 [140]. A molecule of Compound X was surrounded with 1683 molecules of SPC water in a box of dimensions 3.75 × 3.75 × 3.75 nm^3. The system was first minimised and equilibrated for 1 ns and subsequently the production MD simulation was ran at 2 fs/step for a total of 100 ns with a snapshot taken every 1000 steps. Periodic boundary conditions were used along with constant temperature of 300 K and pressure at 1 bar using V-rescale thermostat and Parrinello-Rahman barostat respectively. Electrostatic and van der Waals interactions were calculated using a Verlet cutoff-scheme with Potential-shift-verlet as modifier with a cut-off of 1.4 nm.

The MARTINI two/three-to-one mapping approach was adopted for the rings in Compound X and four-to-one mapping was used for the remainder of the molecule (Figure 18). Bonded interactions were derived by mapping the all-atom simulation to CG resolution, the bond length, angle and dihedral angle between virtual particles over the entire simulation were measure by constructing histograms of respective bond or angle. In most cases, a gaussian distribution with a single peak was obtained. However, in some cases, the distribution may contain more than a single peak, so the most prominent peak was used.

Initially, the bond equilibrium length and angles were set to the centres of the histograms. A CG simulation was prepared with these parameters, where a molecule of Compound X is surrounded with 421 molecules of standard MARTINI water in a box of dimensions

3.70 × 3.70 × 3.70 nm³. The test systems were first minimised and equilibrated for 1 ns and subsequently the production MD simulation was ran at 20 fs/step for a total of 100 ns with all other conditions are exactly the same as all-atom simulation described above. The force constant and centre of the harmonic bond and angle potentials were adjusted iteratively until satisfactory agreement of the distribution functions with all-atom simulation was achieved.

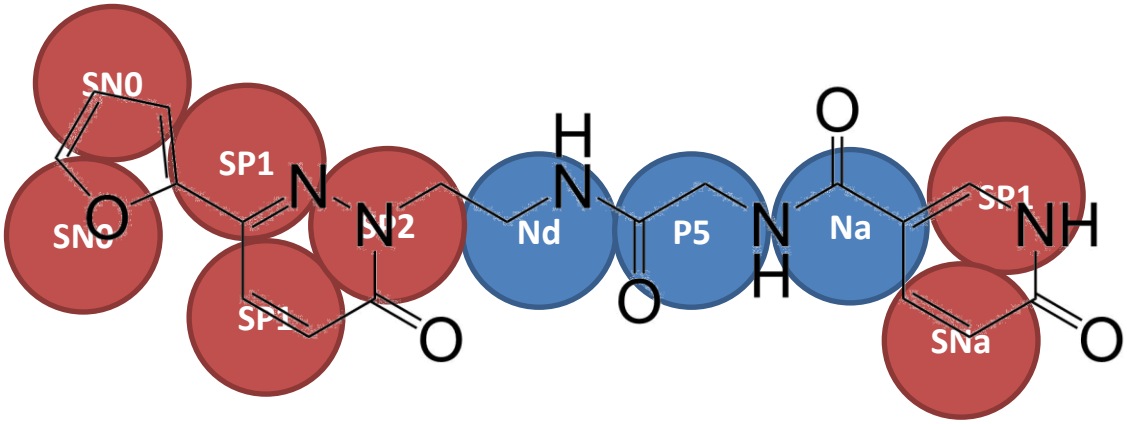


Figure 18: Coarse-grained mapping and bead assignments for Compound X. Ordinary beads were coloured in blue and S beads were coloured in red.

4.2.2. Coarse-grained Simulations with DAFT

The dimerisation simulations of MEDI-578 Fv fragments were prepared using the DAFT protocol [176] using MARTINI protein force field version 2.2 [169-170] with ElNeDyn [171] applied on the MEDI-578 Fv fragments using a cut-off distance of 0.9 nm and a force constant of 500 kJ mol^{-1} nm^{-2}. Three sets of simulations were prepared, two MEDI-578 Fv fragments in 100 mM NaCl, the same system with ten molecules of trehalose and the same system with ten molecules of Compound X through a bash script (Supplementary Information 8.4.). In each simulation, two MEDI-578 Fv fragments were placed at the same initial distance of 3.0 nm and the two MEDI-578 Fv fragments were randomly rotated, yielding 1024 different relative starting orientations for individual

simulations. Systems were subsequently solvated with standard MARTINI water and ionised using insane. The final systems contained two MEDI-578 Fv fragments, roughly 9150–9300 standard MARTINI water beads, 68 NA+ ion beads and 66 CL- ion beads [177] in a box of dimensions approximately 13.00 × 13.00 × 6.70 nm^3. For systems containing trehalose, ten molecules of MARTINI trehalose molecules were added to individual system by replacing standard MARTINI water beads with GROMACS 2016.3 through a bash script (Supplementary Information 8.6.) [140]. Topology and force field parameters for trehalose were taken directly from the MARTINI force field parameters for carbohydrate [180]. For systems containing Compound X, the topology and force field parameters were obtained as described earlier and ten molecules of Compound X were added in the same way as trehalose.

The prepared systems were minimised with martinate [178] using GROMACS 2016.3 [140] for 500 ps at 1 fs/step before undergoing a 100 ps NPT equilibration MD simulation at 20 fs/step. The final production runs were performed with a time step of 20 fs for 512 ns with a snapshot taken every 25000 steps. Periodic boundary conditions were used, a constant temperature of 300 K was maintained and pressure was controlled at 1 bar using V-rescale thermostat and Berendsen barostat respectively. Electrostatic and van der Waals interactions were calculated using a Verlet cutoff-scheme with a Potential-shift-verlet as modifier with a cut-off of 1.8 nm.

4.3. Results and Discussions

4.3.1. Parameterisation of Compound X

The coarse-graining of Compound X was developed using the bottom up approach, where the atomistic simulation data were mapped into virtual beads and the CG parameters were derived and refined through systematically modifying the individual terms until a high

degree of agreement to the atomistic simulation data using GROMOS 54A7 force field parameters generated from ATB 2.2 is achieved. The selection of MARTINI beads were chosen based on the principles of the original MARTINI publication [167]. Since there were no experimental values for the water/octanol partitioning coefficient of Compound X available, it is currently no possible to further validate Compound X with partition free energy calculations. The focus here is to reproduce the bonded parameters, bond length, angles and dihedral terms. The final parameters of Compound X in MARTINI force field file is listed in Supplementary Information 8.1. .

Figure 19 shows the overall bond length distributions for all CG bonds within Compound X in both models. In general, the section for the terms and force constants are representative to the behaviour of the atomistic simulation with GROMOS 54A7 force field parameters generated from ATB 2.2. The major exceptions were the bonds within pyridazine ring where the actual bond lengths within the beads were around 0.05 nm shorter than the actual length despite all the changes in the values for the bond length constraints.

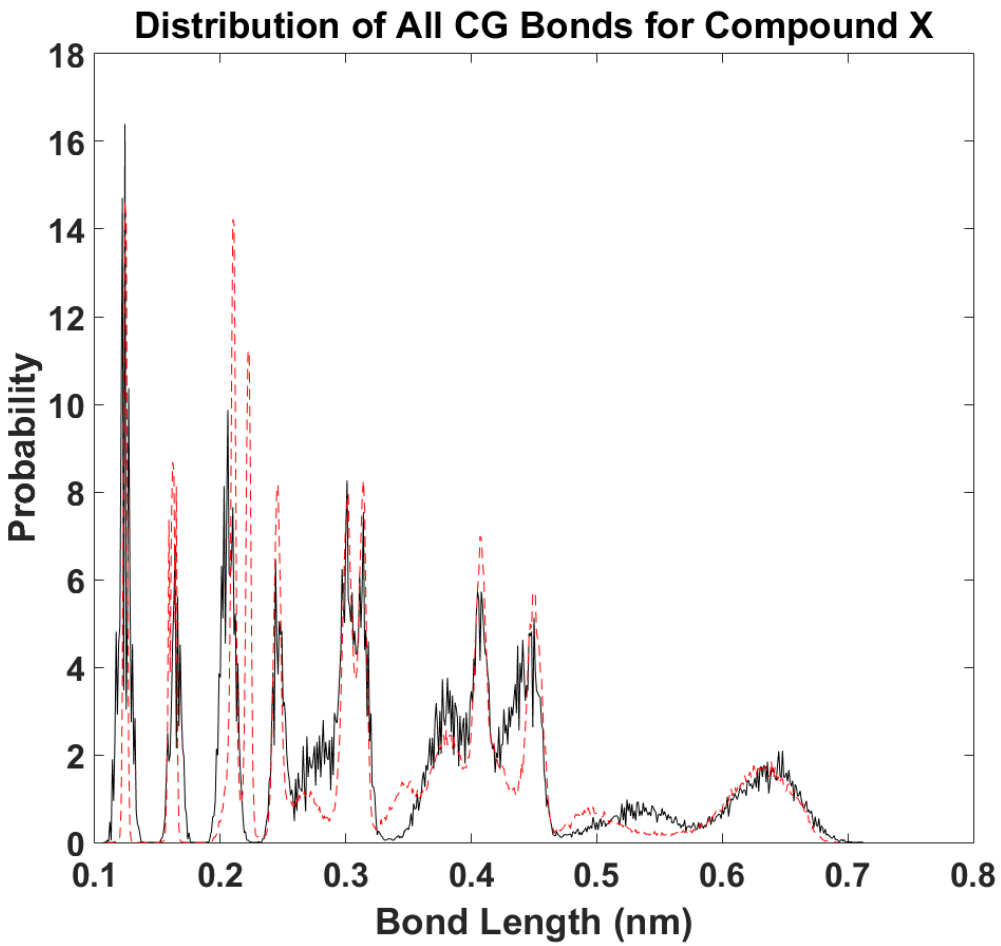


Figure 19: Distribution of bond lengths for all of the CG bonds within Compound X. The mapped CG bond lengths from all-atom simulations are coloured in red dotted lines and the black solid lines represent the CG bond lengths in the Compound X martini model.

All bond angles within Compound X were also analysis as part of the force field development process. For pyridazine and pyridine rings, they were represented by three S beads connected to each other so the bond angles within these rings were already defined by the bond lengths. Furan contains only five heavy atoms and it was represented by two S bead in our model so there was no angle parameter associated with it. The rest of the bond angles in the Compound X MARTINI model were in good agreement with the atomistic data using GROMOS 54A7 force field parameters generated from ATB 2.2. The peak and trough at 150^{o} - 160^{0} was caused by the connection between the ordinary bead outside of the pyridazine ring and the S beads in pyridazine ring. Further increase

of these force constants lead to destabilisation of the simulation. Despite both angle terms being defined, the model was unable to match the atomistic simulation data.

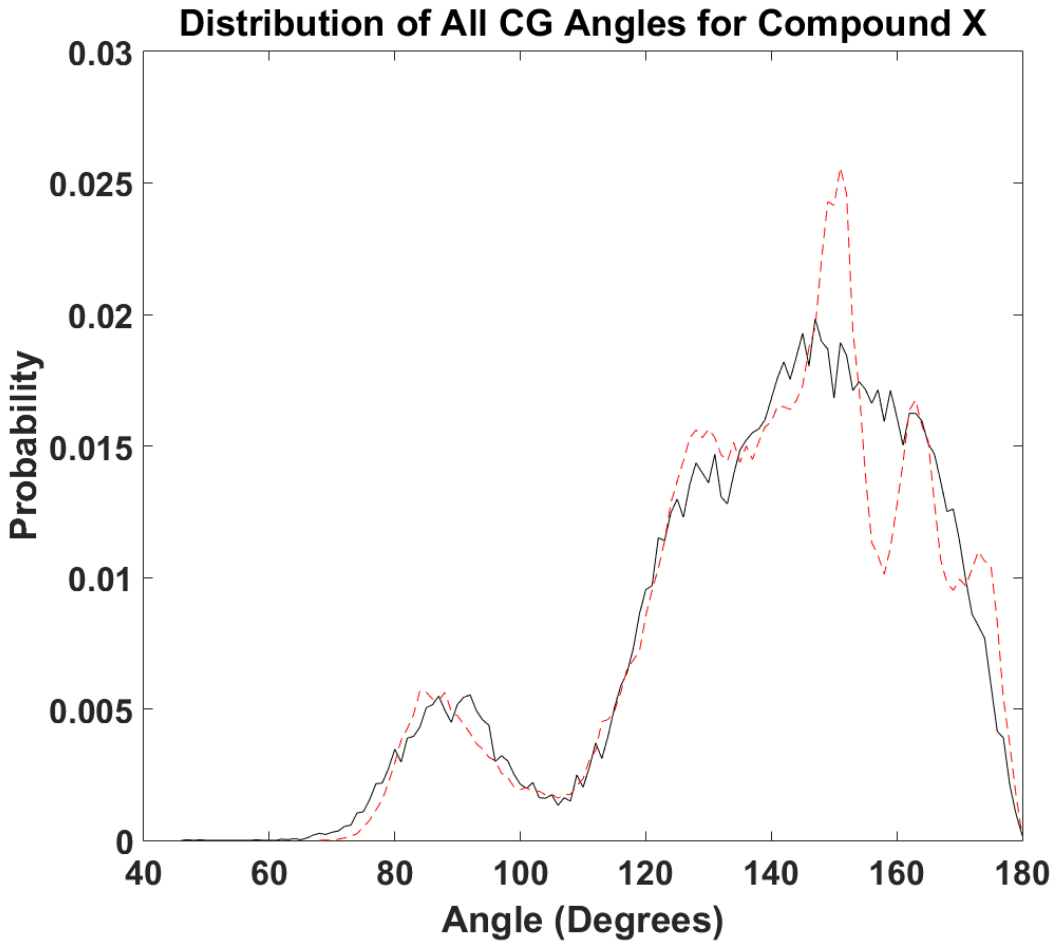


Figure 20: Distribution of angles for all of the CG angles within Compound X. The mapped CG angles from all-atom simulations are coloured in red dotted lines and the black solid lines represent the CG angles in the Compound X martini model.

Dihedrals are rarely included in standard MARTINI force fields because there are less rotatable groups due to coarse-graining. In MARTINI protein force field for proteins, the dihedral terms were only included in histidine, phenylalanine, tyrosine and tryptophan to prevent backflipping of the aromatic ring within the side chain [169]. When comparing the atomistic simulation with simulation data generated by the Compound X MARTINI model for every possible dihedral combinations, there were no significant differences observed given that there were no dihedral terms defined. The overall distribution for all

possible CG dihedral angles in the Compound X were already in reasonable agreement with atomistic simulation data without additional dihedral term definitions (Figure 21).

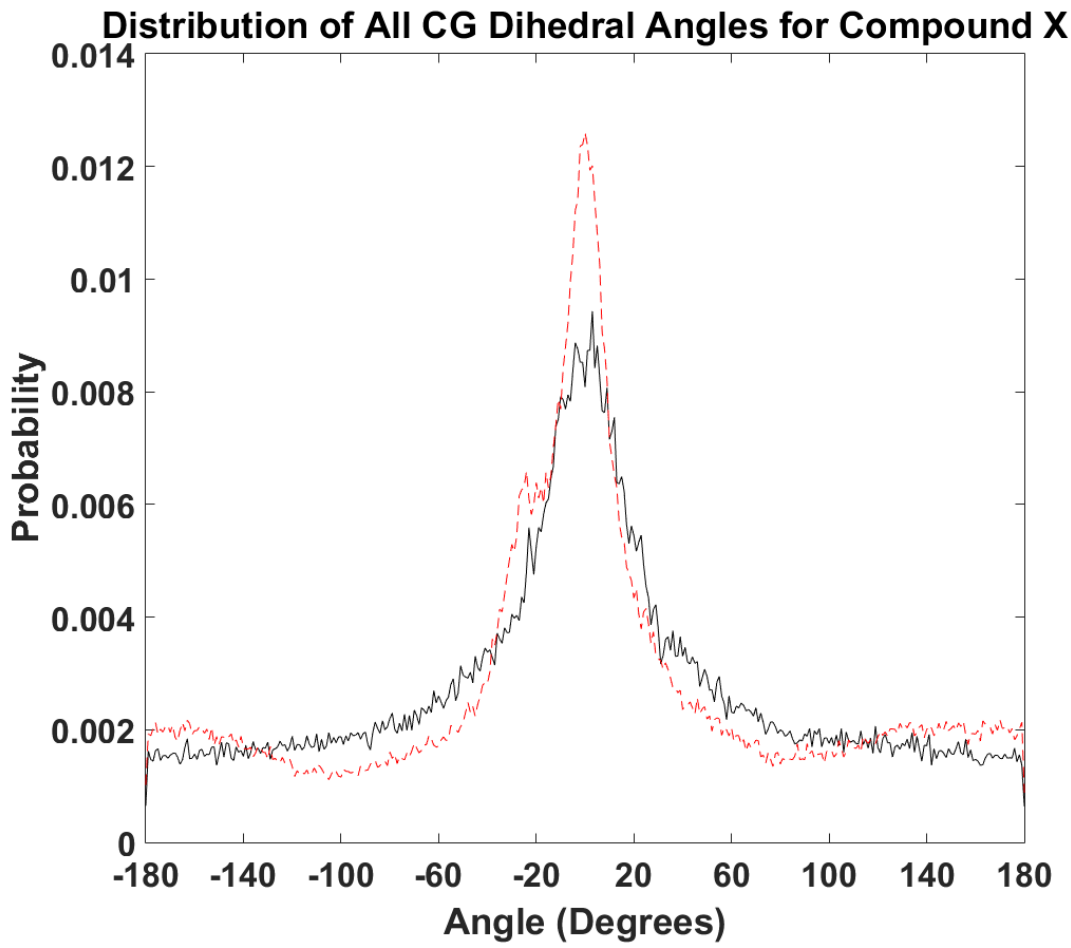


Figure 21: Distribution of angles for all possible CG dihedral angles within Compound X. The mapped CG dihedral angles from all-atom simulations are coloured in red dotted lines and the black solid lines represent the CG dihedral angles in the Compound X martini model.

4.3.2. Dimers Formation

4.3.2.1. Interaction Energies

To study how Compound X and trehalose affects the MEDI-578 Fv fragment dimer complex formation in a MARTINI MD simulation, DAFT was used to generate unbiased starting orientations. The energy profiles were predicted by calculating the sum of non-

bonded interaction energies, van der Waals and electrostatics, and they were plotted vs time (Figures 22, 23, 24). At the beginning of the simulations, the two MEDI-578 Fv fragments were separated at a defined distance of 3 nm and there were no interaction at this distance. During the course of the simulation, the two MEDI-578 Fv fragments diffuse and interact with other solutes in the simulations. When two MEDI-578 Fv fragments interact, the distribution of interaction energies shifts towards negative as binding is thermodynamically favourable. Dimerisation of MEDI-578 Fv fragments have been observed in most simulations with around 5% of all simulations not showing any interactions between the two molecules of MEDI-578 Fv fragment. The mean interaction energy was calculated with nonlinear least-squares fitting of the data. The distribution of energy values were represented with a spectral colour scheme in the form of vigintile (5%) points. The minimum and maximum values were coloured in pink, followed by a rainbow towards the central vigintile coloured in red with the mean value of all energies is shown as solid black line and calculated plateau values of the means were determined and shown in thick black dashed lines.

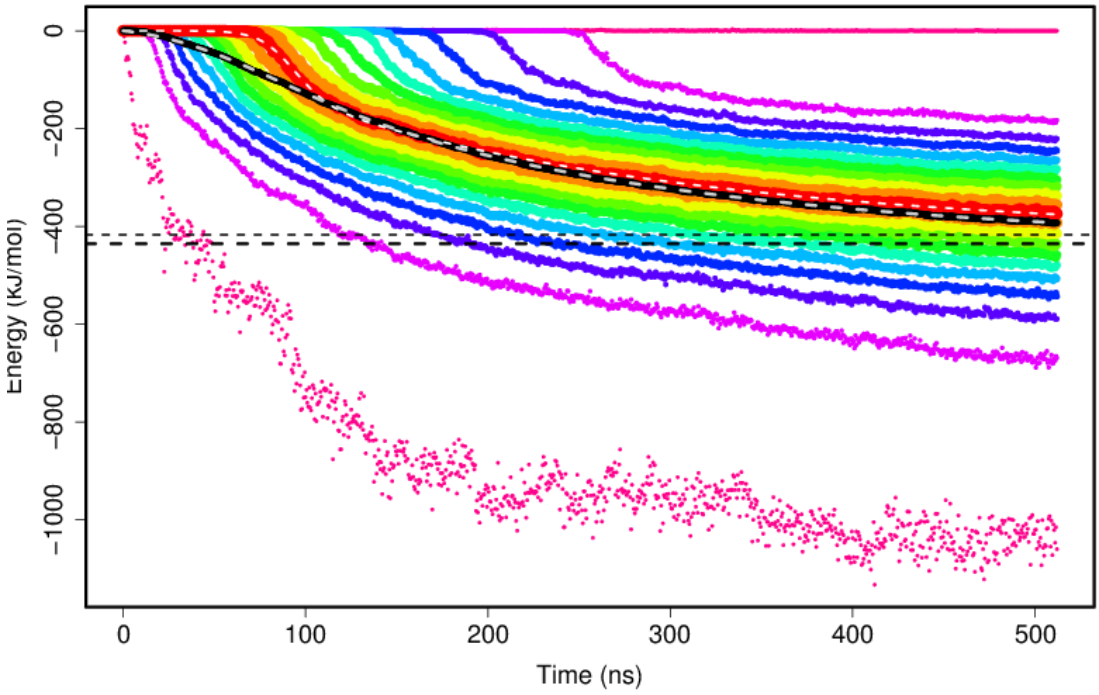


Figure 22: Interaction energy between MEDI-578 Fv fragment dimers in 100 mM NaCl over time presented in vigintile (5%) points. The central vigintile (median) was coloured in red. The solid black line indicates the mean interaction energies of all simulations and calculated plateau values of the means were determined and shown in thick black dashed lines. The calculated plateau value of the means for MEDI-578 Fv fragment dimerisation was determined to be -417.2 kJ/mol.

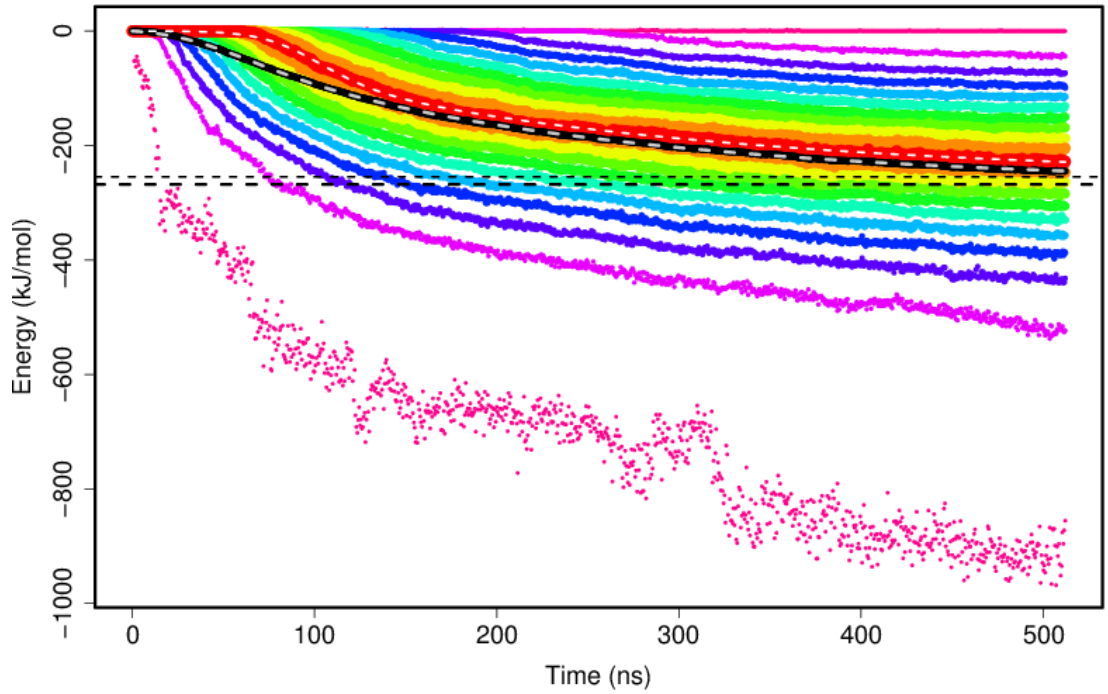


Figure 23: Interaction energy between MEDI-578 Fv fragment dimers in 100 mM NaCl and 10 molecules of Compound X over time. The calculated plateau value of the means for MEDI-578 Fv fragment dimerisation in presence of Compound X was determined to be -253.1kJ/mol.

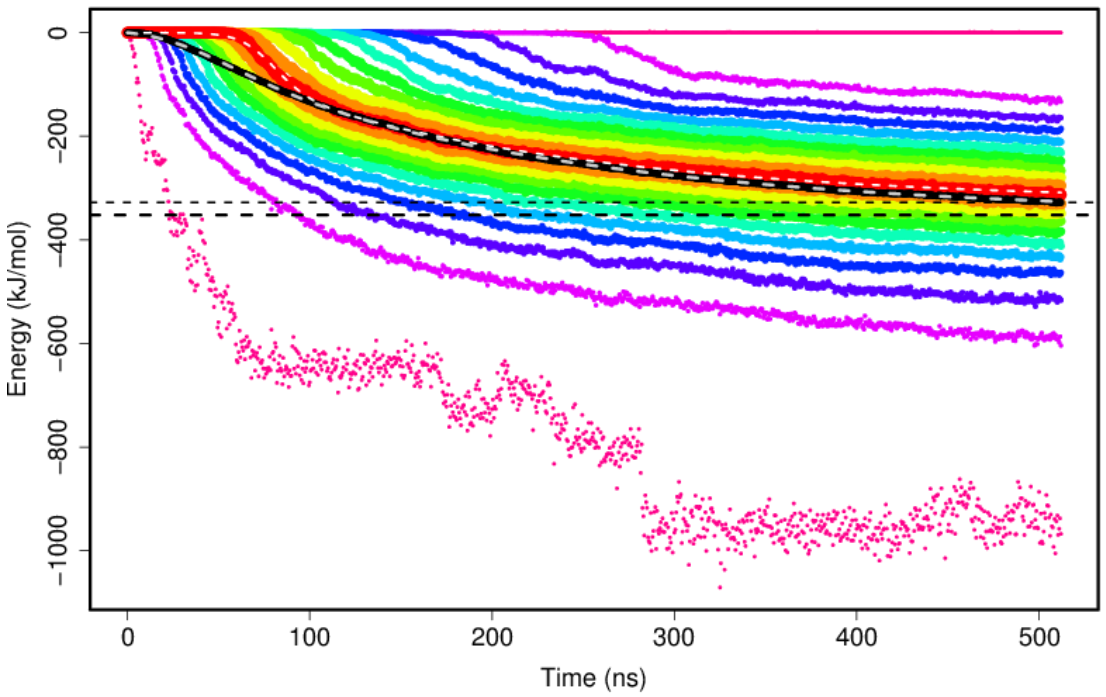


Figure 24: Interaction energy between MEDI-578 Fv fragment dimers in 100 mM NaCl and 10 molecules of Trehalose over time. The calculated plateau value of the means for MEDI-578 Fv fragment dimerisation in presence of trehalose was determined to be -319.8 kJ/mol.

The calculated plateau values of each set of simulations provided a simple overview of how Compound X and trehalose interferes with the protein-protein interaction energies (Table 8). The dimerisation of two MEDI-578 Fv fragments resulted in a wide distribution of interaction energies. The addition of either Compound X or trehalose molecules resulted in calculated mean interaction energies reductions. The presence of Compound X led to less negative mean interaction energies than did the presence of trehalose. This also suggests the intermolecular attractions between Fv dimers were reduced in the presences of either compared than for the two MEDI-578 Fv fragments alone.

When Compound X or trehalose binds to an Fv fragment at the interface of the dimer, these molecules act as physical barriers and reduce the interaction energies between Fv fragments. In this study, Compound X outperformed trehalose in terms of interaction energies. This may be due to the relative larger size of Compound X as Compound X and

trehalose were represented with a 10- and 6-bead model respectively so the size of the physical barriers are larger.

Excipient compound included in MD simulation	Calculated mean interaction energies (kJ/mol)
-	-417.2
10 Compound X molecules	-253.1
10 Trehalose molecules	-319.8

Table 8: The calculated plateau value from the mean interaction energy distributions in different systems. Addition of trehalose cause increase in the mean interaction energies and a higher mean interaction energies achieved in systems with Compound X, suggesting self-interactions of MEDI-578 Fv fragments are energetically less favourable in systems with trehalose or Compound X.

4.3.2.2. Interaction Patterns

The distance between any pair of beads from the respective of MEDI-578 Fv fragments was calculated for all frames for all simulations within a simulation set and location of excipients were tracked. Results were combined for identification of the most common interaction pairs. In contrast to the original *in-silico* studies, the self-association events in these MD simulations through the loop near residue VAL10 on the light chain occurred much more frequently than LEU105 on the heavy chain (Figure 26). Other interaction hot-spots were also identified, including VAL11, PHE55 and THR76 on the heavy chain. From a short 60 ns all-atom simulation (Figure 8), these regions had a negative SAP scores and were not considered as aggregation-prone region. ZDOCK also shown a much weaker signal from these regions (Figure 10). Top 10% of simulations with most negative interaction energies from the final snapshot have been evaluated for their confirmations (Figure 27). A strong signal has been seen from THR76 and to lesser extend from LEU105 on the heavy chain, suggesting Fv fragment-Fv fragment interactions through these regions are most energetic favourable. It is important to understand the locations of the interaction hot-spots within these MEDI-578 Fv dimer complexes as interruptions to

self-associations via these interaction sites will prevent stable dimer complexes formations. It is also worth to note that the formation of antibody dimers does not only depend on hydrophobicity and shape complementarity, surface charges of the antibody molecule are equally important.

Compound X was discovered through virtual screening using the CDR3 loop on the heavy chain as the target. Compound X was found to bind to the CDR3 on the heavy chain more frequently than trehalose, especially at ARG101 (Figure 25). Figure 28 shows a reduction of interaction events that were observed in the presence of Compound X through LEU105 on the heavy chain, the most hydrophobic residue from SAP calculations. These suggest Compound X can reach the intended area of interaction as defined in the virtual screening study, and that Compound X can acts as a protein-protein interaction breaker to prevent self-association. The results from the simulations also suggested that self-associations through heavy chain LEU105 were not the most favourable, the addition of Compound X shifted interactions towards VAL11 on the heavy chain and VAL10 on the light chain.

In the earlier all-atom simulations, trehalose was not able to bind to the region close to the CDR-3 on heavy chain. In this study, there were no reduction in self-association events through heavy chain LEU105 in presence of trehalose (Figure 29) compared to the systems with just two Fv fragments (Figure 26). The number of interaction events through VAL11 on the heavy chain and VAL10 on the light chain increases. There is a smaller increase in interaction events through PHE55 on the heavy chain. Both Compound X and trehalose do not appear to have any affinity towards these regions so they cannot prevent interactions between MEDI-578 Fv fragments at these sites. An increase of interaction between trehalose and THR76 on the heavy chain (Figure 25) as well as a decrease in

dimer complex formation through self-interaction at the same residue was observed in simulations with trehalose compared to simulation with Compound X (Figure 29).

This study also showed that the dynamic nature of mAb, effects of the charges and realistic formulation environments have to be taken account when studying mAb-mAb interactions and the exposure of the largest solvent-accessible hydrophobic patch is not the sole consideration for mAb-mAb interactions. In addition, a short 60 ns simulation with only one copy of the Fv fragment was insufficient to characterise the complex nature of mAb-mAb interactions.

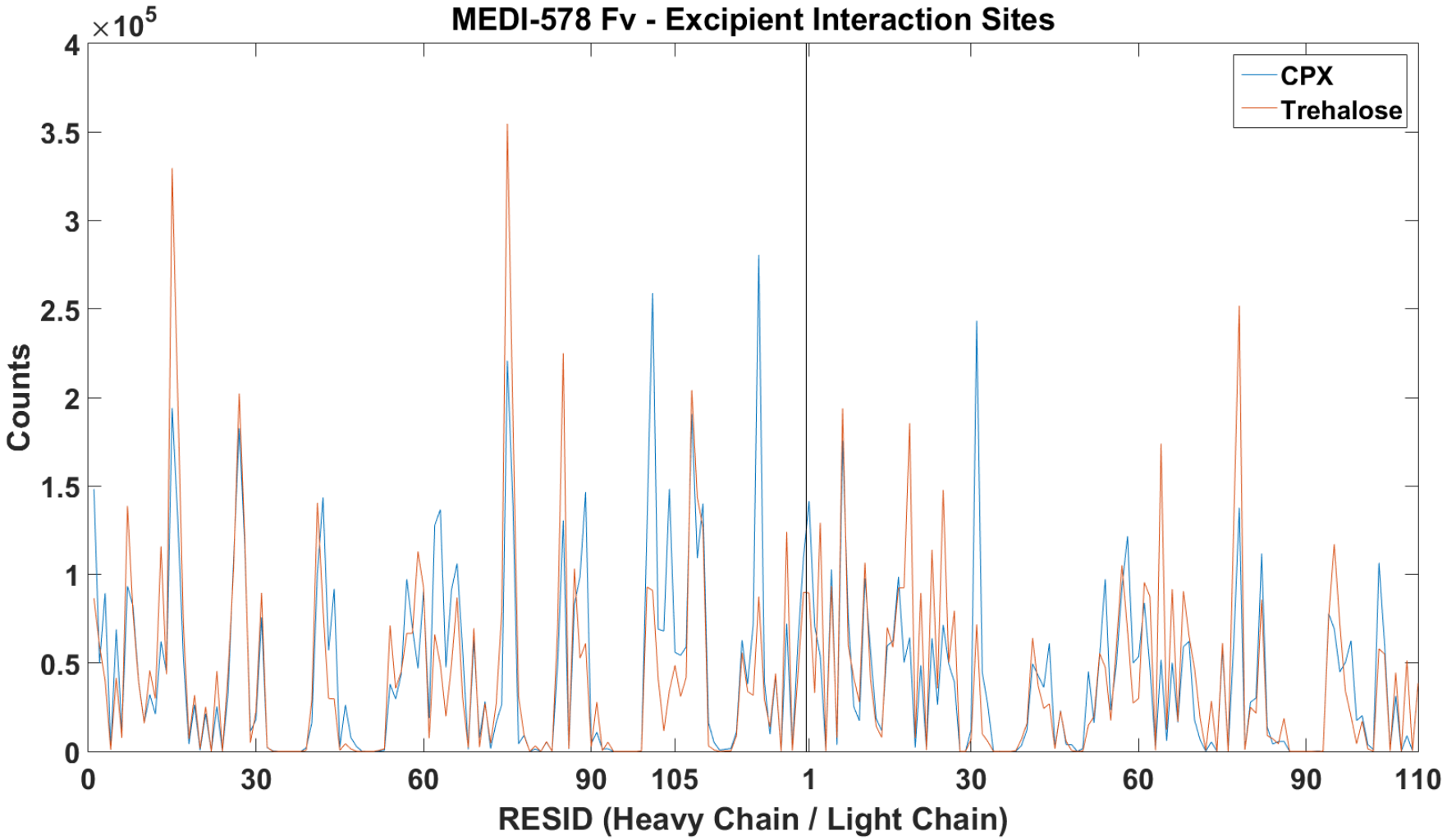


Figure 25: The locations of the excipients were tracked in each sets of 1024 MARTINI CG simulations with two MEDI-578 Fv fragments in 100 mM NaCl and 10 molecules of either Compound X (blue) or trehalose (orange). Compound X was found to bind to the CDR3 more frequently than trehalose.

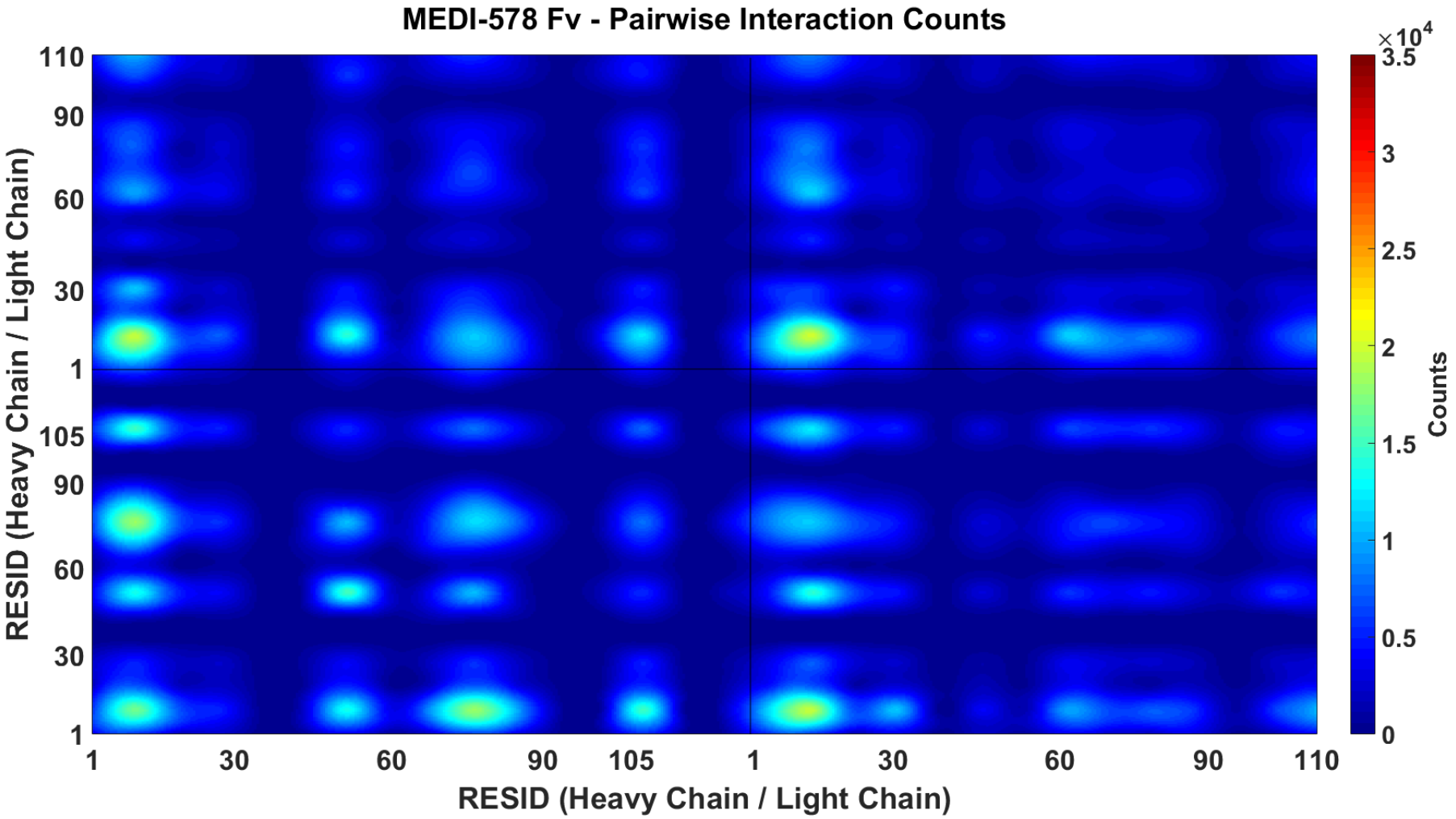


Figure 26: The most common binding pairs in a set of 1024 MARTINI CG simulations with two MEDI-578 Fv fragments in 100 mM NaCl. The occurrence of interaction pairs were coloured according to the colour bar on the right. Main self-association events occur in the CDR1 near residue VAL10 on the light chain. Apart from LEU105, alternative interaction hot-spots were also identified, including VAL11, PHE55 and THR76 on the heavy chain.

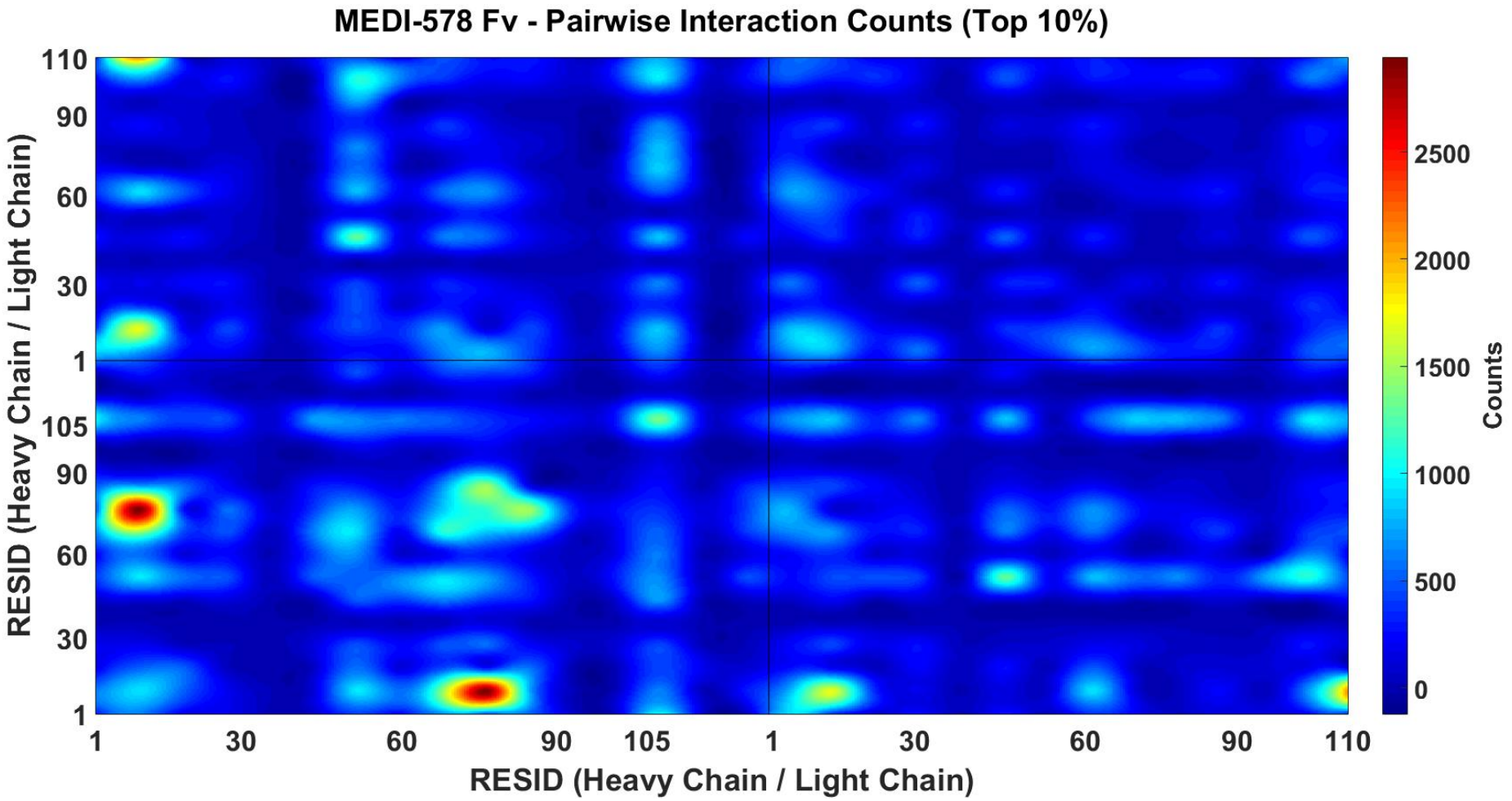


Figure 26: The most common binding pairs in top 10% of simulations with most negative interaction energies from a set of 1024 MARTINI CG simulations with two MEDI-578 Fv fragments in 100 mM NaCl. The area coloured in red denotes the binding pairs with most occurrence. The most energetically favourable MEDI-578 Fv fragment dimer complexes mostly interacts through THR76 and to lesser extend from LEU105 on the heavy chain.

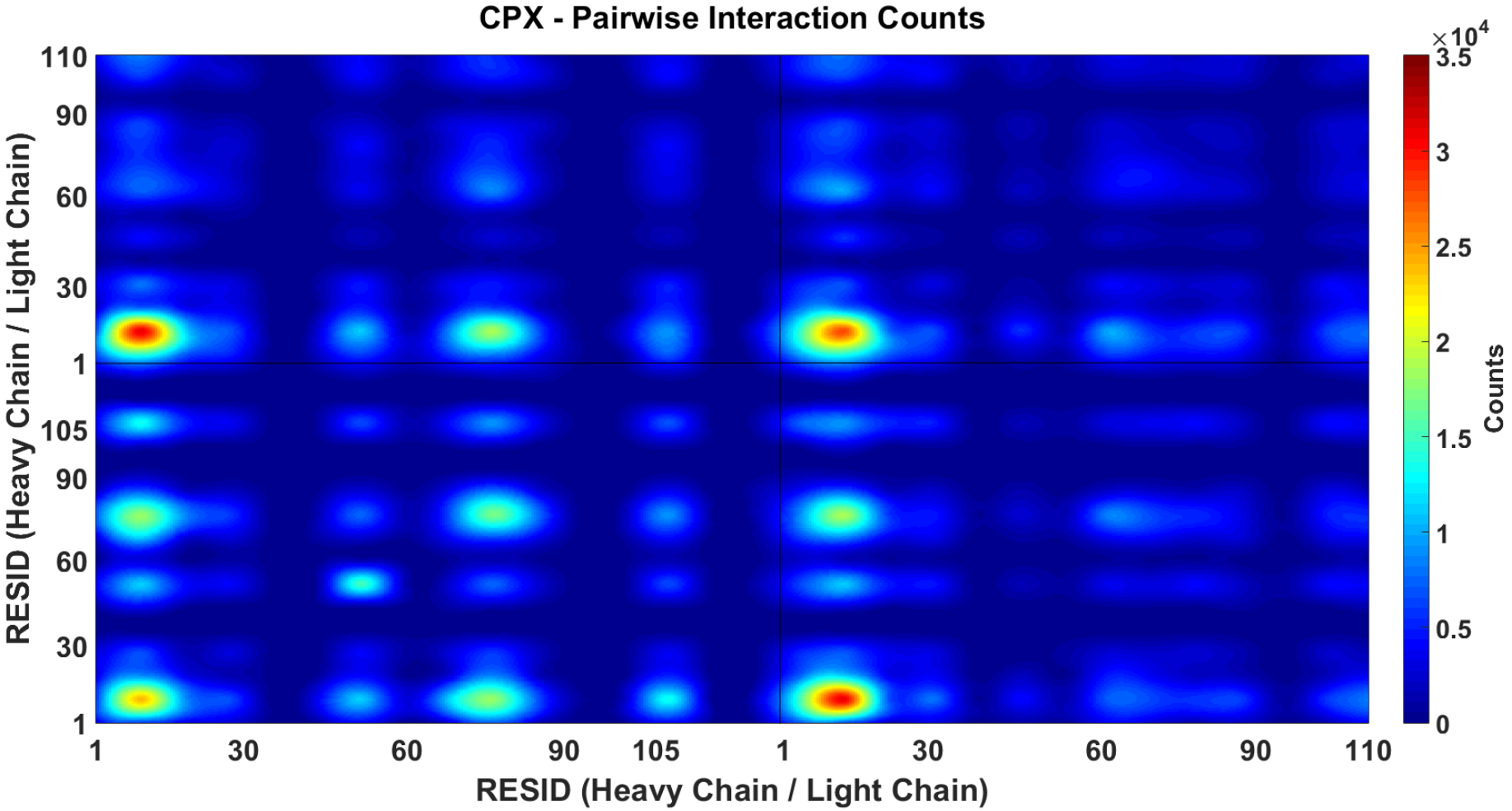


Figure 28: The most common binding pairs in a set of 1024 MARTINI CG simulations with two MEDI-578 Fv fragments in 100 mM NaCl and 10 molecules of Compound X. The most energetically favourable MEDI-578 Fv fragment dimer complexes mostly interacts through VAL11 on the heavy chain and VAL10 on the light chain while interactions through the most solvent-accessible hydrophobic patch (LEU105 on heavy chain) slightly reduces.

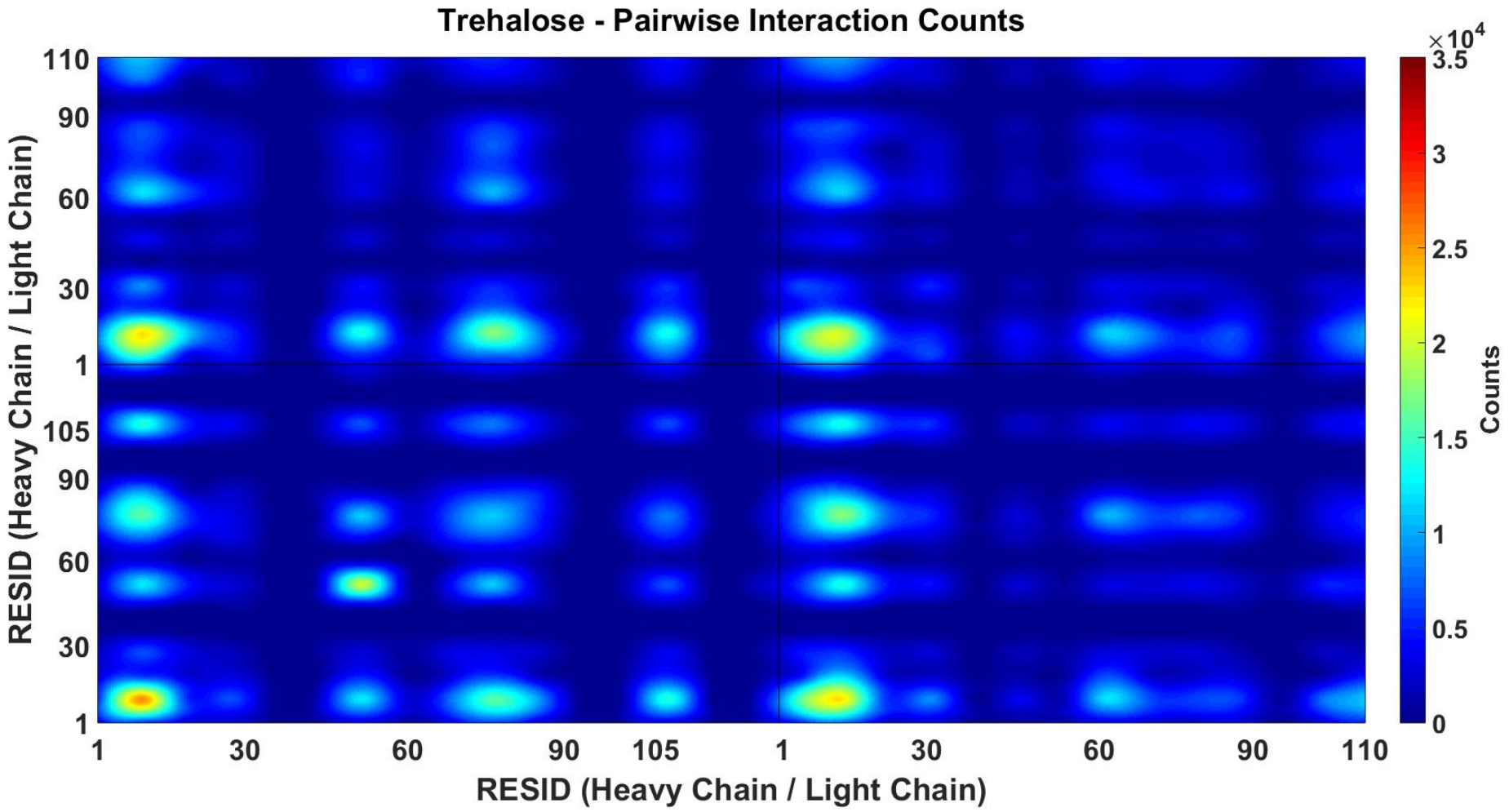


Figure 29: The most common binding pairs in a set of 1024 MARTINI CG simulations with two MEDI-578 Fv fragments in 100 mM NaCl and 10 molecules of trehalose. The area coloured in red denotes the binding pairs with most occurrence. The occurrences of self-interactions through LEU105 on heavy chain are similar to systems with only two MEDI-578 Fv fragments.

4.4. Conclusions

The computational work described in this chapter describes an alternative method to evaluate mAb-mAb interactions compared to Chapter 2. The hypothesis in Chapter 2 described that a low molecular weight excipient that binds to solvent accessible hydrophobic regions of an antibody to disrupt antibody-antibody interactions has been confirmed in this chapter. However, the experimental data did not verify that this low molecular weight excipient can delay aggregation process as shown in Chapter 3 compare with standard disaccharide excipient trehalose.

Electrostatics have equal importance as hydrophobicity and shape complementarity in the formation of antibody dimers. For simulating formulations of mAbs, a nanosecond scale short simulation may not be long enough to offer sufficient sampling of the system as the overall energy landscape is large and complex. Taking into account for the dynamic properties of the mAb, effects of solvation, charges and other excipients, the exposure of an aggregation prone region may change over time.

Ideally, an excipient should be able to reduce the interaction energies between mAb molecules and prevent the most favourable interactions. Coarse-graining using the MARTINI force field with the aid of DAFT to generate replicates with different orientations provided an effective protocol for simulating mAb-mAb interactions in presence of other excipients. As the computational power improves along with standardisation of the workflow, the use of the MARTINI force field provides an opportunity to predict interaction patterns with various excipients.

5. Free energy calculations for antibody self-associations simulated with MARTINI force field

5.1. Introduction

The effect of intermolecular antibody-antibody and antibody-solvent interactions undeniably play important roles in dimer formation. The non-covalent association of two MEDI-578 Fv fragments was evaluated as described in the previous chapter. The formation of antibody dimers depends on many factors including shape complementarity, residue hydrophobicity and surface charges of the antibody molecule.

Free energy calculations are often used to improve the predictions of binding energies from a MD simulation. In an aqueous environment, dimerisation of mAb molecules might exist in an equilibrium between the monomers and dimers which can be represented by the following equation.

$$A + A \rightleftharpoons AA \tag{5.1}$$

where A is monomer of an antibody and AA refers to their dimer complex.

The binding constant is defined as

$$K = \frac{[AA]}{[A] + [A]} \tag{5.2}$$

where [A] is the respective molar concentration of antibody monomer and [AA] refers to the molar concentration of their dimer complex.

The binding constant *K* is related to the changes in Gibbs free energy for the system which is defined as

$$\Delta G = \Delta H - T\Delta S \tag{5.3}$$

where ΔG is the Gibbs free energy change, ΔH is enthalpy change, T is absolute temperature and ΔS is entropy change.

The change in Gibbs free energy (ΔG) for dimer complex formation is related to the binding constant by the following expression.

$$\Delta G = -RTlnK \tag{5.4}$$

where ΔG is the free energy change, R is the ideal gas constant and K is the binding constant.

Formation of an antibody dimer complex is therefore depends on the changes of both enthalpy and entropy components of the system. The method used in Chapter 4 was based on generating a large number of simulation replicates with different starting configurations so each simulation would have an equal *a priori* probability for encounter. The energy values reported were the mean energy estimated with a nonlinear least squares function using the interaction energies between two molecules using a non-polarised MARTINI water model.

The enthalpy component of binding thermodynamics has been well addressed in the previous chapter by calculating the non-bonded interactions with the sums of Lennard-Jones and Coulomb potentials within the MARTINI CG model to study the formation of MEDI-578 Fv fragment dimers. Results from multiple replicates were pooled together and further evaluated with a model function to predict the mean interaction energy between the dimers. However, water is known to possess an extended hydrogen bond network in solution and undoubtedly plays important roles in binding energy predictions [181].

In the non-polarised MARTINI water model, four water molecules are represented by a single uncharged P4 bead [167]. The large hydrogen bond network of water is simplified

and may affect the accuracy of determining the enthalpy component as well as predicting the entropy gained from desolvation during antibody dimerisation. The method described in the previous chapter also neglects the entropic terms for the effects of solvation free energy [182] and the apolar energy contribution by SASA as describe in MM-PBSA [183], which has also been used to describe protein-protein interactions. This chapter aims to calculate these entropic contributions and evaluate their relative importance by comparing them to the sum of the Lennard–Jones and Coulomb potentials with a view to validate or improve the accuracy for quantifying the interaction between MEDI-578 Fv fragments in the MD simulation.

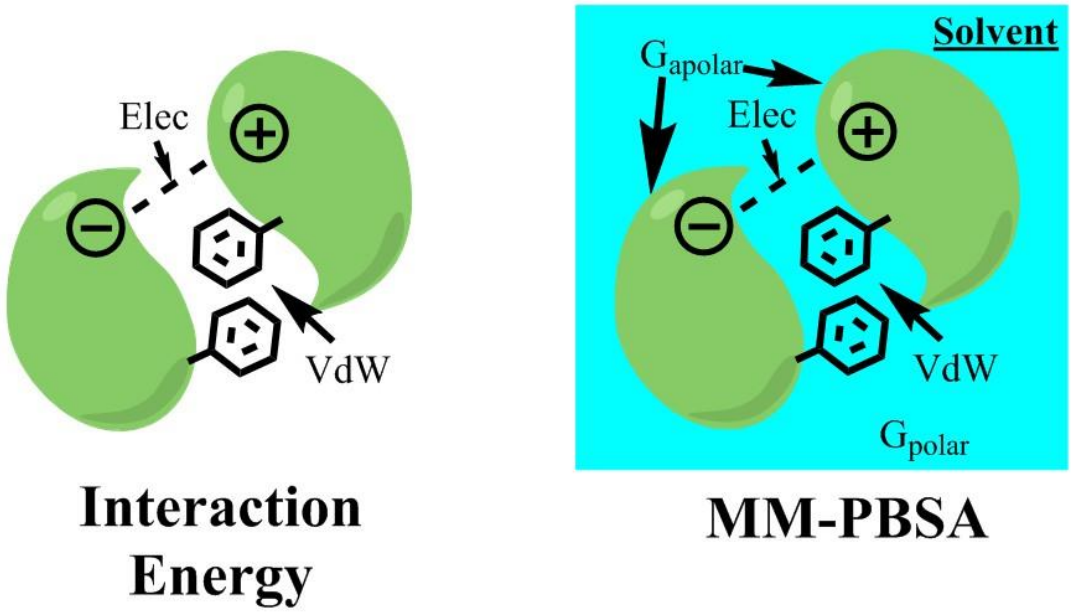


Figure 30: Graphical comparison of energy terms considered in interaction energy and MM-PBSA method.

5.1.1. Solvation Free Energy Differences

Antibody dimers have different solubility to monomers and the solvation free energy changes may relates to the stability of dimer complexes. The changes in the solubility of MEDI-578 Fv fragments in dimerisation along a MD trajectory can be predicted by solvation free energy simulation. Bennett acceptance ratio (BAR) is an efficient and

robust method to evaluate free energy based on energy difference between two ensemble states and has also been shown to be unbiased statistically as compared to exponential averaging and thermodynamic integration [184]. To calculate the free energy, a series of intermediate states (λ) are used to determine the interaction energies between these intermediate states, provided these intermediate states are close enough. In each λ point, the Lennard-Jones and Coulomb potentials between solutes and solvents are modulated accordingly. The overall solvation free energy are calculated from these intermediate states.

5.1.2. MM-PBSA

To increase the precision and accuracy of binding energy predictions, MM-GBSA and MM-PBSA methodologies are commonly employed to explore detailed energy composition of the binding-free energy of docked protein-ligand complexes [183]. In MM-GBSA and MM-PBSA calculations the free energy of protein-protein interactions are estimated by taking into account solvation energy contributions and molecular mechanics (MM) energy terms (bonded and non-bonded). The contribution of polar solvation energy is calculated based on an implicit solvent model (Generalized Born or Poisson–Boltzmann) and the apolar solvation energy is calculated from the solvent accessible surface area (SA) [183].

To the best of our knowledge, there has been no direct application of MM-GBSA and MM-PBSA calculations to the MARTINI CG model which may due to the following two reasons.

1. GB or PB implicit solvent models do not exist for the MARTINI force field. In dry MARTINI, the implicit solvent model for MARTINI and dry MARTINI have

drastically different parameters anddifferent non-bonded interaction levels are typically used. Differences in partitioning free energies, bead–bead potentials of mean force and radial distribution functions were observed in dry MARTINI compared to the original model [185].

2. The experimental value for water surface tension is 72 mN m^{-1} at 293 K [167]. However, the MARTINI CG model gives surface tensions of 45 and 30 mN m^{-1} depending on the size of the system [167]. This makes assigning accurate fitting parameters for apolar energy calculations based on the SASA difficult, especially since MARTINI beads are bigger than ordinary atoms.

5.2. Materials and Methods

5.2.1. Solvation Free Energy Calculations

The solvation free energy of the proteins were calculated on a selected MARTINI MD trajectory using the gmx bar script included with the GROMACS package [140]. In the implementation of the BAR method [184], the trajectory was sampled for multiple coupling states with a parameter λ to adjust the strength of the Van der Waals interactions between the proteins and the solvent. For our calculations, 10 points of λ were chosen from 0 to 1 with an increment of 0.1.

5.2.2. MM-PBSA with Backmap

Evaluating the binding energy using MM-GBSA or MM-PBSA in individual simulations is irrelevant. The two protein molecules were separated at the beginning of the simulation and the time taken for individual interaction events to occur differs. Therefore the binding energy of the Fv dimer complexes may be inaccurate due to the absence of interactions between Fv molecules at the beginning of individual simulation.

To take advantage of MM-PBSA for further evaluation of Fv dimer complexes formed, the final snapshot of a MARTINI CG simulation can be backmapped to an all atom representation [186-187] first and then used to calculate the binding energy using the g_mmpbsa tool [185].

5.3. Results and Discussions

5.3.1. Solvation Free Energy Differences

Significant computational resources were required for solvation free energy calculation. The changes in solvation free energy of MEDI-578 Fv fragments were calculated only for one snapshot per 8 ns on the original MARTINI trajectory of a selected replicate. In each replicate, the first interaction event was observed at 43 ns after the simulation started and became stable from 112 ns to 222 ns where further decrease of interaction energy was observed and became stable for the remaining of the simulation. Each snapshot was evaluated with Bennett Acceptance Ratio (BAR) calculations with 11 λ points (in steps of 0.1) [182]. Each λ point were simulated for 20000 steps and the overall free energy difference for each simulation snapshot from the output data were calculated with GROMACS BAR tool. The number of core-hours used to perform solvation free energy calculations for this simplified trajectory significantly exceeded the original core-hours used for genertating the MARTINI MD trajectory.

The calculated solvation free energies fluctuate between 8070 to 8210 kJ/mol (Figure 31). There were no observable differences in the solvation free energy before and after the interaction events occurred. It is reasonable to conclude that were no significant changes in the solvation free energy in the formation of weak intermittent protein complexes.

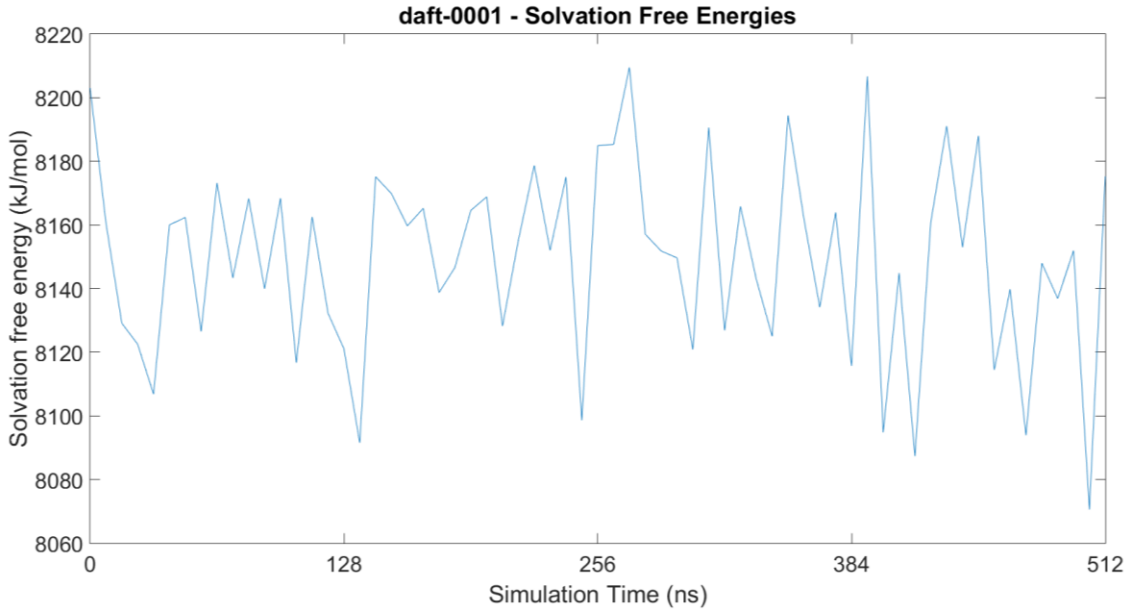


Figure 31: Solvation free energy changes of the two molecules of MEDI-578 Fv fragments in 100 mM NaCl over time of daft-0001. There were no observable difference in the solvation free energy before and after the interaction events occurred.

5.3.2. Backmap

From the MM-PBSA calculations of the backmapped final snapshots, the binding energy appeared to be energetically unfavourable with positive calculated binding energies (Figure 32). It appears there were no correlations between the binding energies from MM-PBSA calculations and the interaction energies as the calculated R^2 value was 0.0043 based on a linear regression model (Figure 32). A single replicate of our coarse-grained trajectory requires roughly 16 hours using 1 CPU core on HPC Grace. With the same hardware, this backmapping approach required approximately 10 minutes per frame using the standard protocol and additional 2 minutes for MM-PBSA calculations. This approach is not feasible as a total of 205 core-hours would be required to analyse a single trajectory. Each set of simulations described in Chapter 4 contains 1024 replicates that would require a total of 629760 core-hours to complete such an analysis. The computational resources required to conduct such calculations exceed the original resources used to generate the

trajectory. The solvation free energies calculated do not reflect the interaction events occurred in the coarse-grained trajectory so this method was not pursued further.

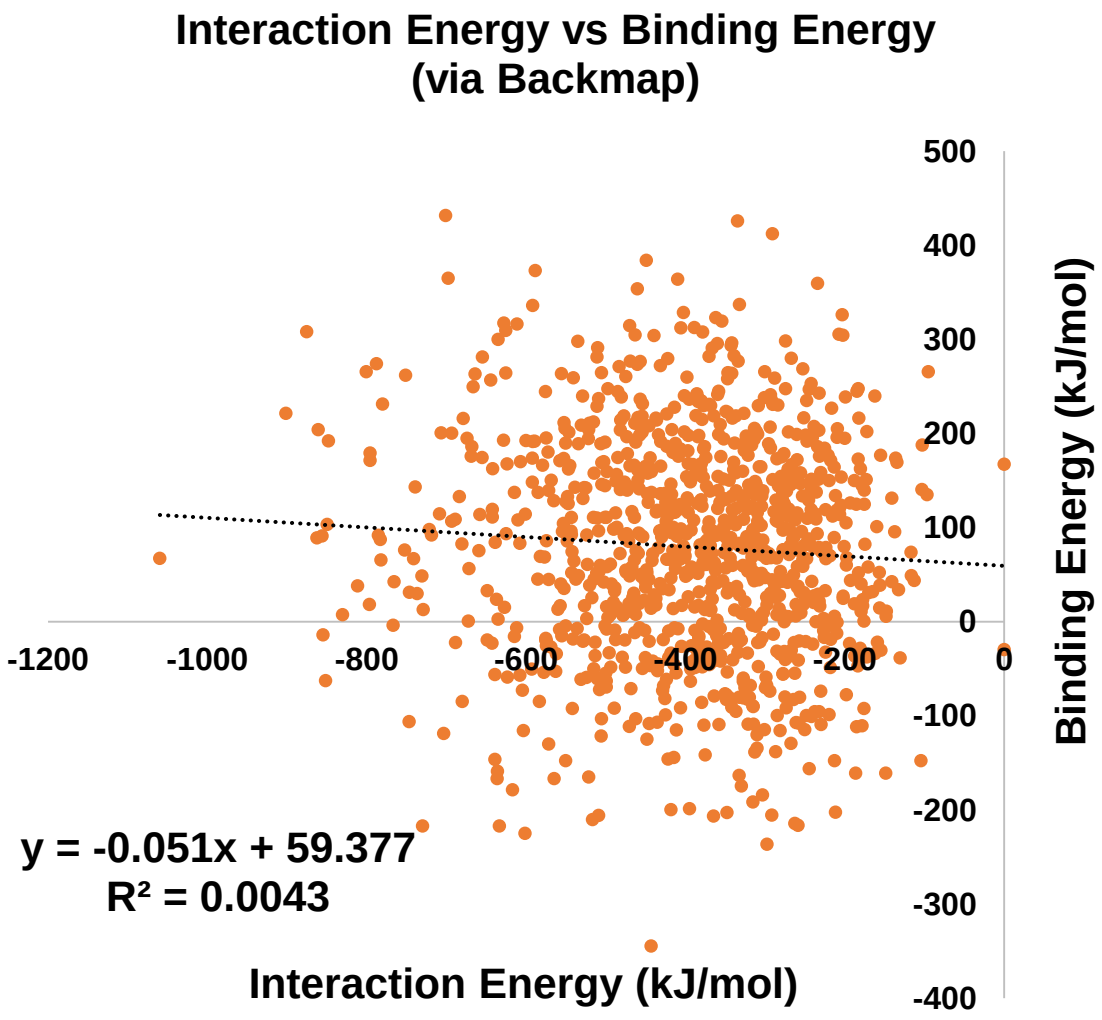


Figure 32: Relationship between the binding energies evaluated using backmap and MM-PBSA and the interaction energies of the end configurations in MARTINI CG simulations containing 2 MEDI-578 Fv fragments in 100mM NaCl. There were no correlations between the binding energies from MM-PBSA calculations and the interaction energies.

Snapshots were taken from a further 10 ns MD simulation were conducted with a backmapped protein complex (Figure 33). The energy terms, polar solvation, apolar solvation and molecular mechanics energy terms fluctuate throughout the simulation. The

polar solvation energy term was strongly positive and the apolar solvation energy term was slightly negative. The molecular mechanics energy terms were strongly negative, yielding slightly positive binding energies overall. The changes in the binding energies mirror the changes in the polar solvation energies as the energy changes in both apolar solvation and molecular mechanics energy terms were relatively minimal.

The same input files were used to generate a further 100 ns MD simulation trajectory to examine if additional simulation time is required for these energy terms to stabilise (Figure 34). The fluctuations in the polar solvation energy remained high and do not seem to stabilise during the 100 ns MD simulation (Figure 34). The binding energy was estimated to be 18 (±106) kJ/mol. The high error indicates simulating beyond 100 ns do not improve overall accuracy to estimate the binding energy using MM/PBSA.

The sum of Lennard-Jones and Coulomb potentials of the last frame of the MARTINI simulation was -1060 kJ/mol. However, during the 100 ns MD simulation of the dimer complex backmapped from MARTINI representations, the MM term fluctuates between -100 and -550 kJ/mol (Figure 34) which is significantly less negative than the calculated value in the MARTINI model. This suggests there are significant structural differences in the structure of the Fv dimer complex between CG and the all-atom representations. This is further confirmed by analysing the RMSD in the 100 ns MD simulation (Figure 35). The configuration of the dimer complex was shown gradually moved away from the original configuration formed at the end of the 512 ns MARTINI simulation.

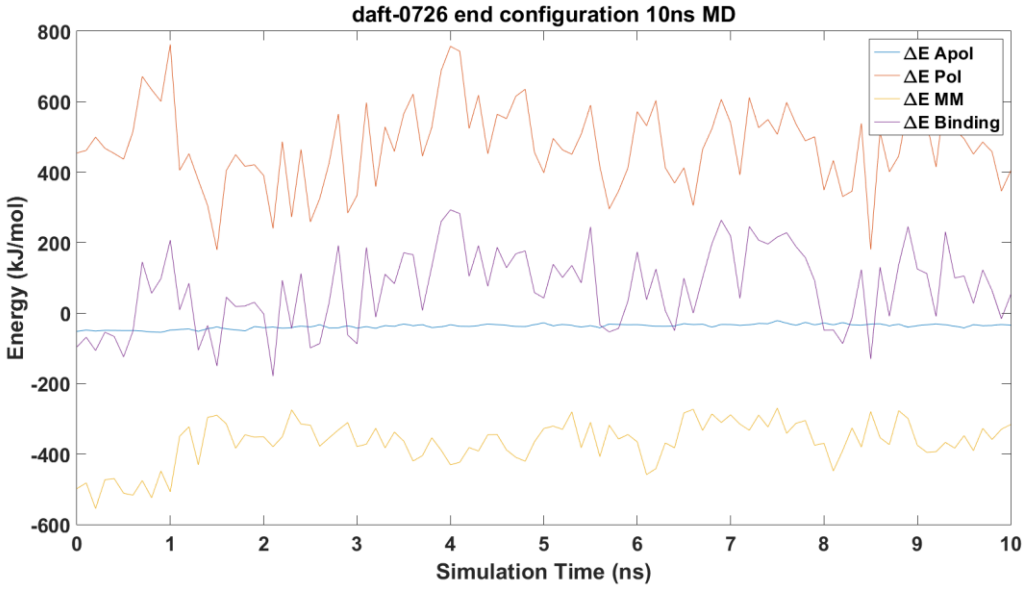


Figure 33: Changes in polar solvation, apolar solvation, molecular mechanics and binding energies over a short 10 ns all-atom MD simulation using MM-PBSA method from end configuration of daft-0726.

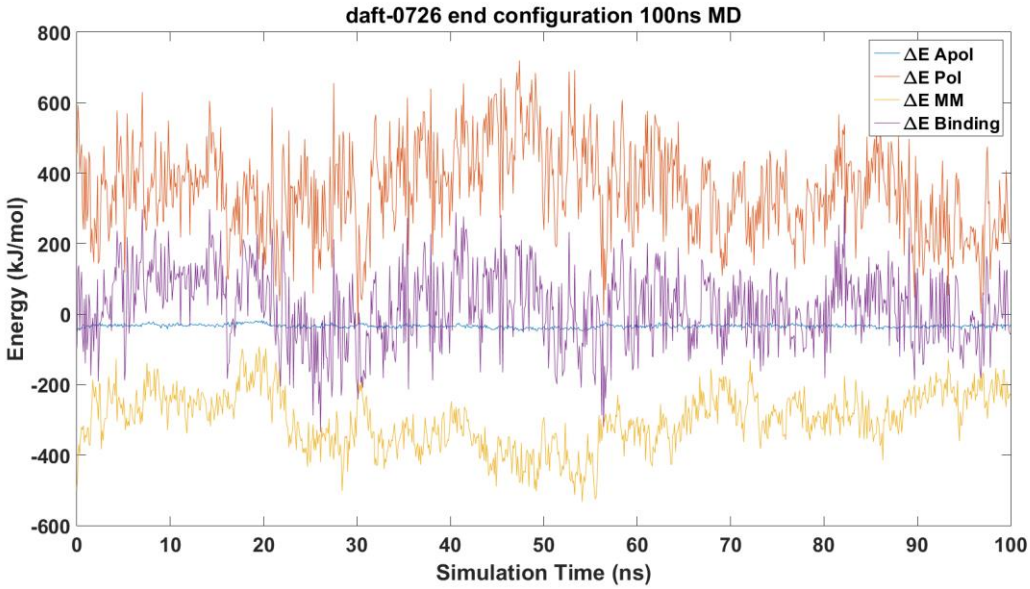


Figure 34: Changes in polar solvation, apolar solvation, molecular mechanics and binding energies over a 100 ns all-atom MD simulation using MM-PBSA method from end configuration of daft-0726.

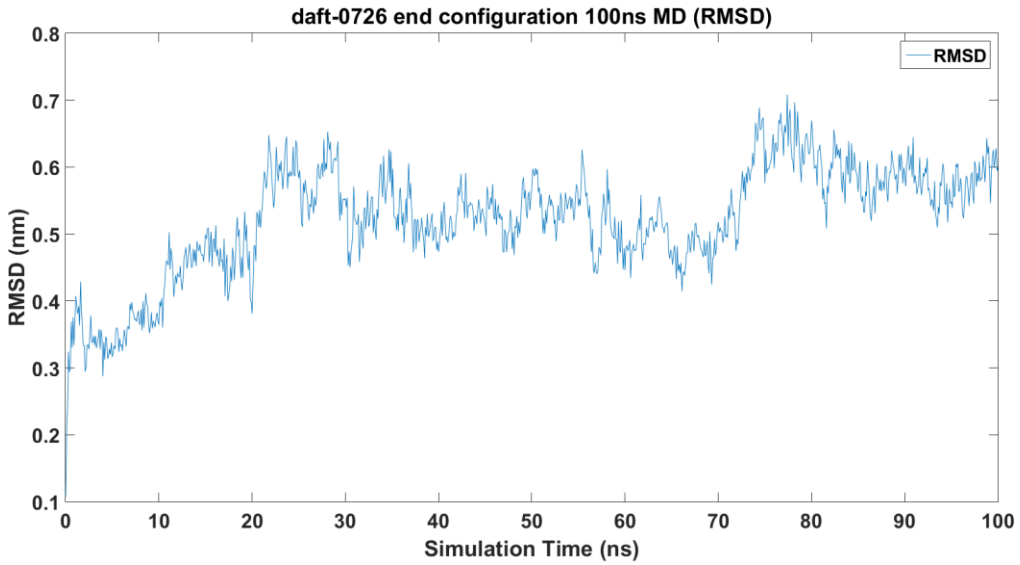
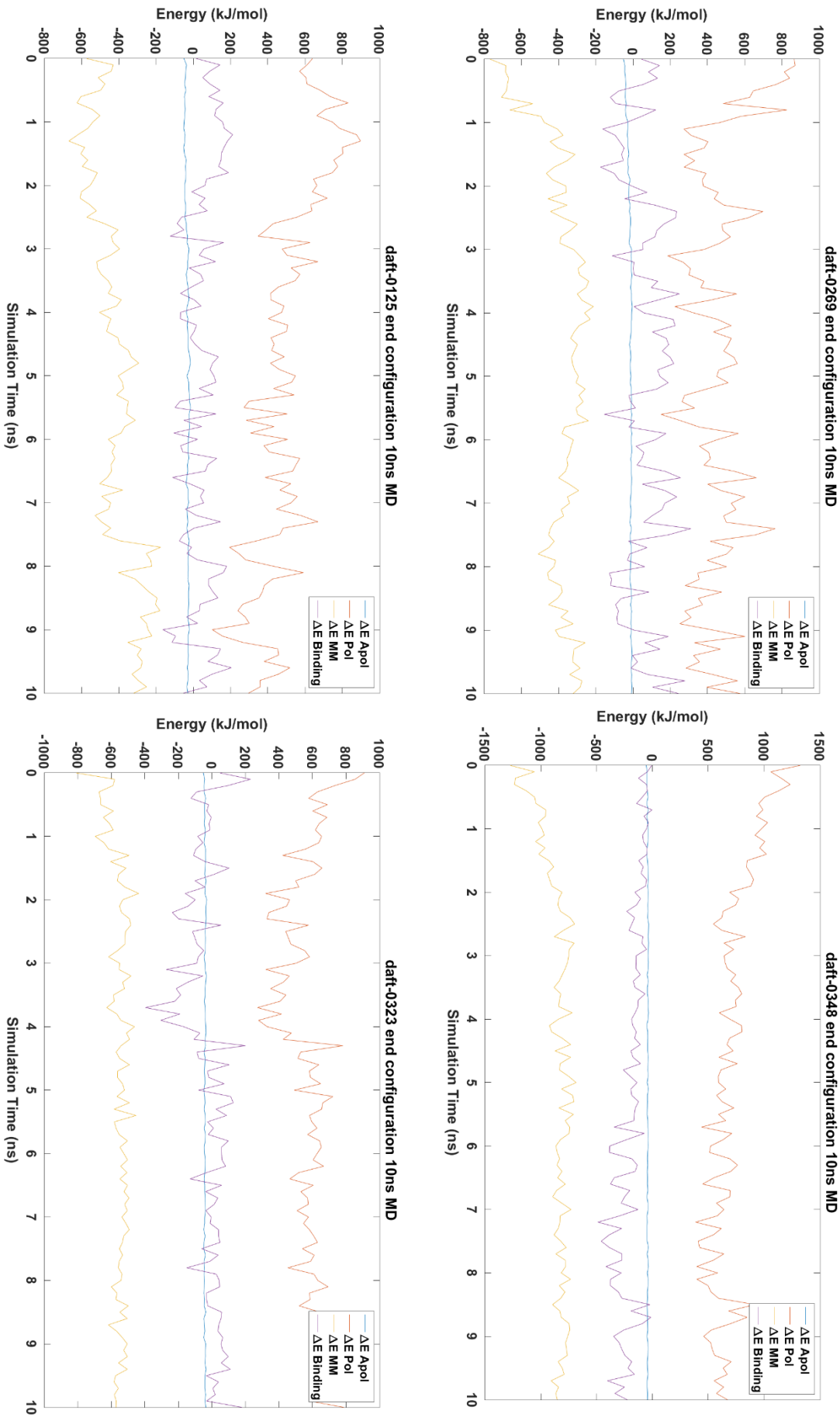
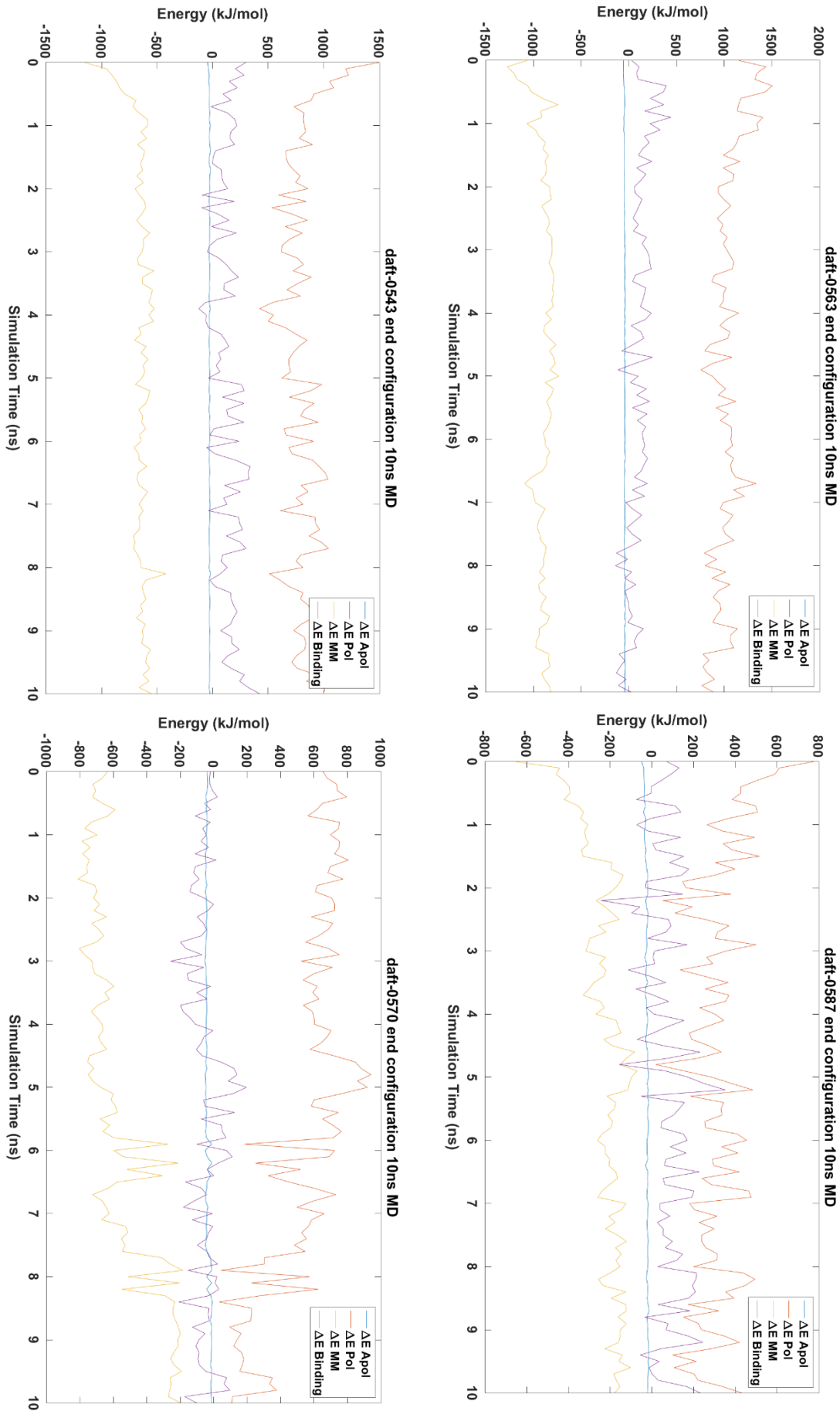
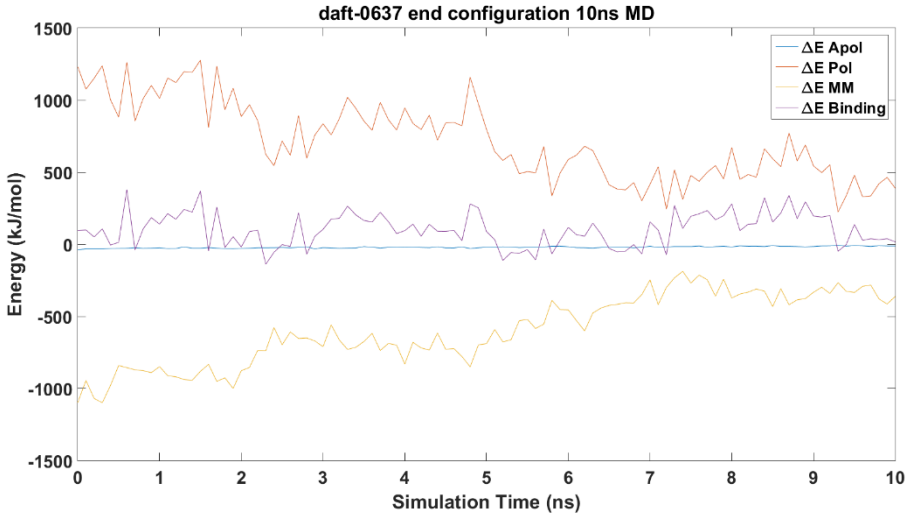


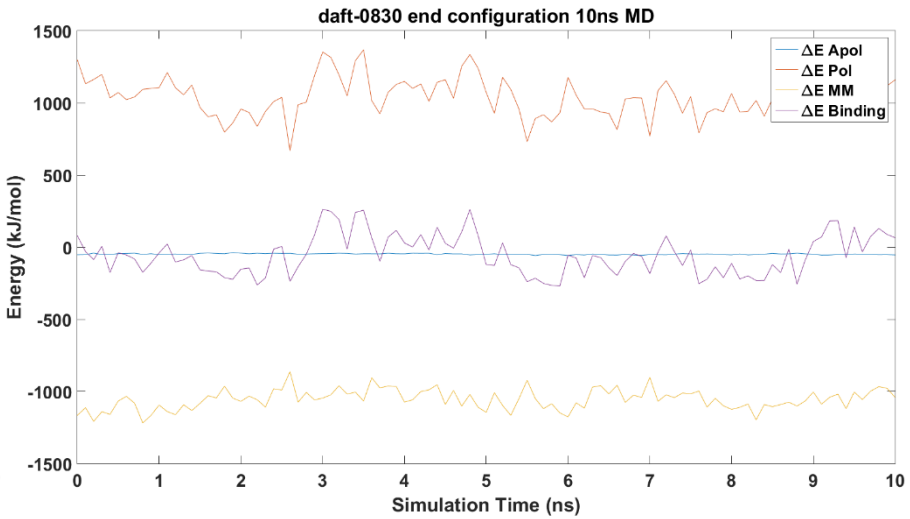
Figure 35: RMSD of Cα atoms of MEDI-578 Fv fragment dimer complexes over time in an all-atom 100 ns MD simulation from end configuration of daft-0726.

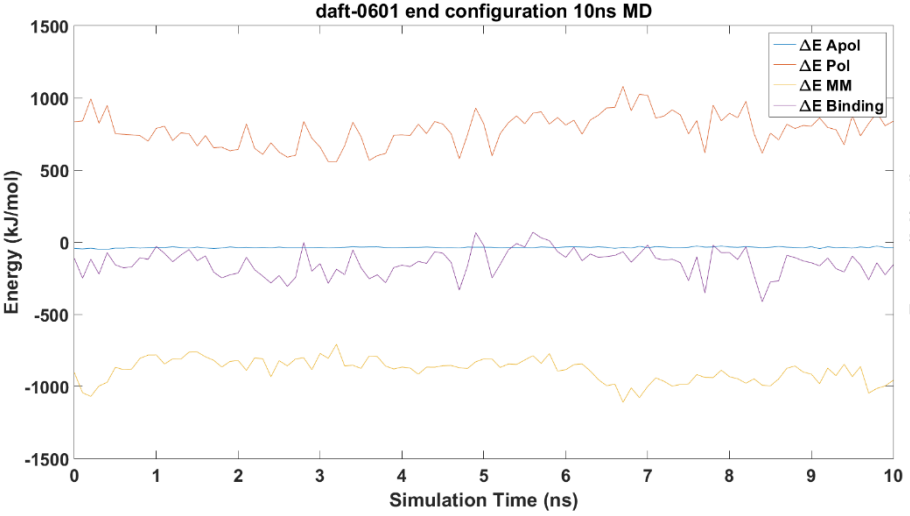
An additional 15 cases with the most negative sums of Lennard-Jones and Coulomb potentials in the final snapshot from MARTINI simulation were evaluated by conducting independent 10 ns all-atom MD simulations that were backmapped from the corresponding end configurations of the MARTINI simulations. The four energy terms, polar solvation, apolar solvation and molecular mechanics energy terms in daft-0323, daft-0348, daft-0543, daft-0563, daft-0601, daft-0830 and daft-0871 are relatively stable throughout the simulation while high fluctuations were seen in all other cases (Figure 36). Out of the 16 cases analysed, none of the MM terms in the all-atom model can reproduce the sum of Lennard-Jones and Coulomb in the MARTINI model (Table 9). Only 2 cases (daft-0348 and daft-0601) out of 16 are thermodynamically stable with binding energies constantly below 0 kJ/mol. There is no correlation between the sums of Lennard-Jones and Coulomb potentials in the MARTINI model and the binding energy evaluated with MM-PBSA using the backmapped structure.

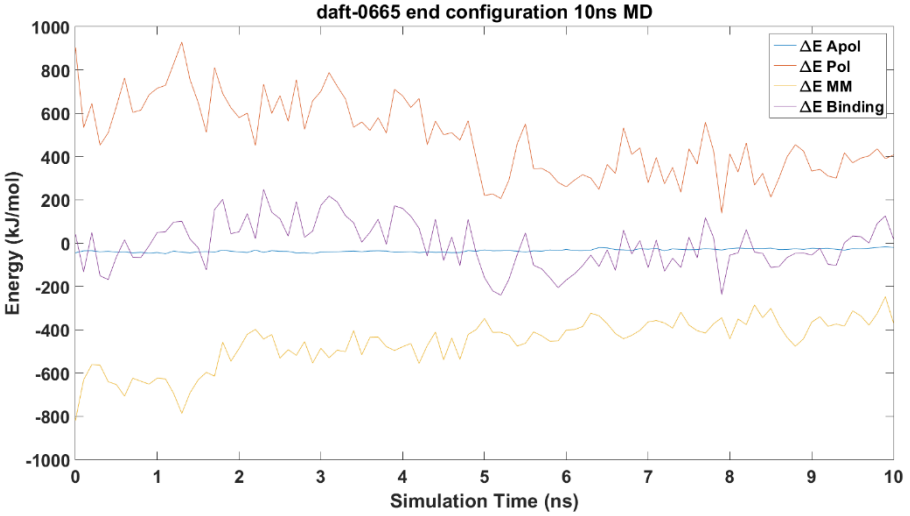

daft-0269 end configuration 10ns MD
Energy (kJ/mol)
Simulation Time (ns)
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding
daft-0348 end configuration 10ns MD
Energy (kJ/mol)
Simulation Time (ns)
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding
daft-0125 end configuration 10ns MD
Energy (kJ/mol)
Simulation Time (ns)
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding
daft-0323 end configuration 10ns MD
Energy (kJ/mol)
Simulation Time (ns)
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding

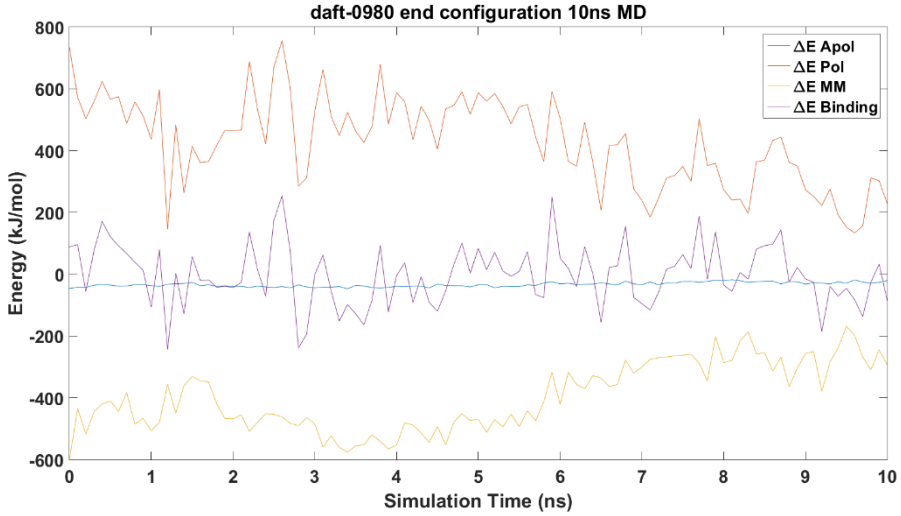

daft-0563 end configuration 10ns MD
Energy (kJ/mol)
Simulation Time (ns)
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding
daft-0543 end configuration 10ns MD
Energy (kJ/mol)
Simulation Time (ns)
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding
daft-0587 end configuration 10ns MD
Energy (kJ/mol)
Simulation Time (ns)
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding
daft-0570 end configuration 10ns MD
Energy (kJ/mol)
Simulation Time (ns)
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding

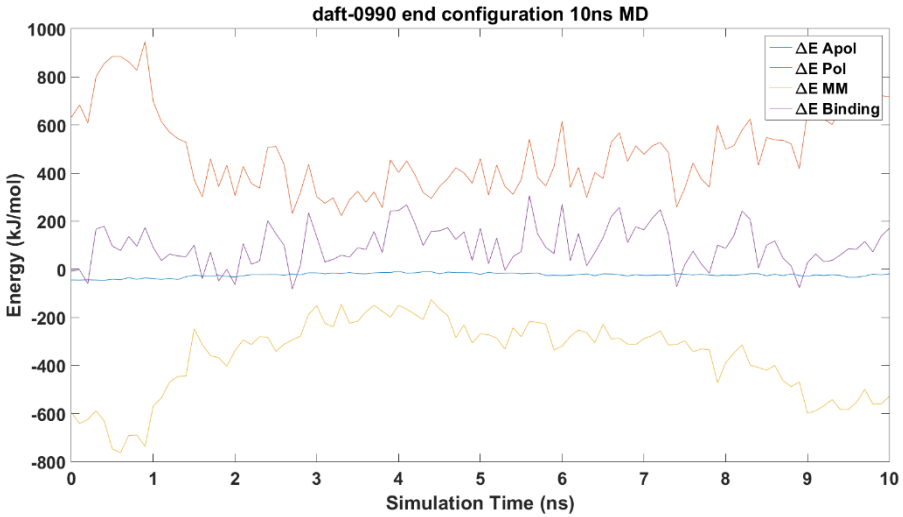

daft-0637 end configuration 10ns MD
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding
Energy (kJ/mol)
Simulation Time (ns)


daft-0830 end configuration 10ns MD
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding
Energy (kJ/mol)
Simulation Time (ns)


daft-0601 end configuration 10ns MD
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding
Energy (kJ/mol)
Simulation Time (ns)


daft-0665 end configuration 10ns MD
ΔE Apol
ΔE Pol
ΔE MM
ΔE Binding
Energy (kJ/mol)
Simulation Time (ns)





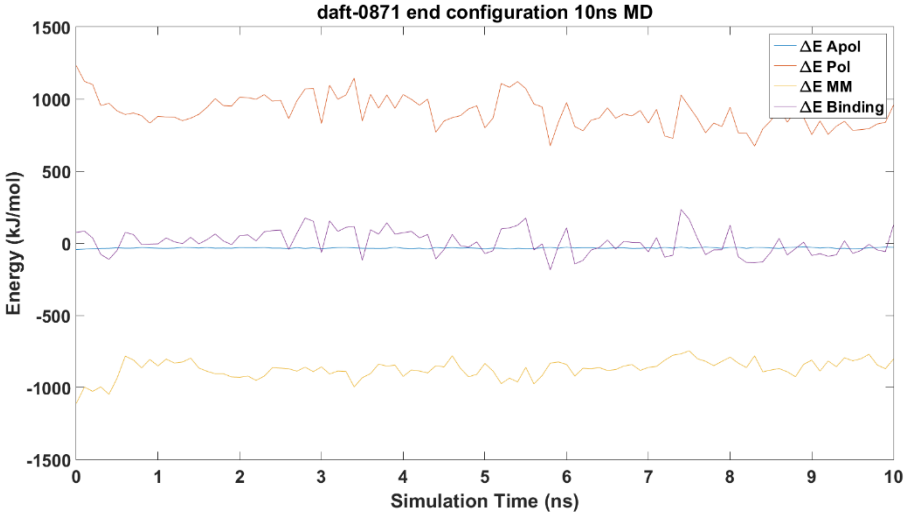


Figure 36: Changes in polar solvation, apolar solvation, molecular mechanics and binding energies over a short 10 ns all-atom MD simulation using MM-PBSA method from end configurations of selected simulations.

Replicate	Lennard-Jones + Coulomb (kJ/mol)	Binding Energy (kJ/mol)
daft-0125	-1060	46 +/- 86
daft-0269	-901	57 +/- 114
daft-0323	-875	-26 +/- 106
daft-0348	-863	-187 +/- 111
daft-0543	-861	128 +/- 114
daft-0563	-856	100 +/- 113
daft-0570	-855	-49 +/- 84
daft-0587	-852	71 +/- 98
daft-0601	-850	-143 +/- 90
daft-0637	-848	103 +/- 115
daft-0665	-831	-6 +/- 109
daft-0726	-811	70 +/- 111
daft-0830	-801	-55 +/- 136
daft-0871	-797	8 +/- 84
daft-0980	-796	-3 +/- 96
daft-0990	-796	96 +/- 84

Table 9. Summary of the sums of Lennard-Jones + Coulomb potentials and the binding energy evaluations from MM-PBSA method in the 16 cases studied.

5.4. Conclusions

The coarse-grained MARTINI model offers a large reduction in the degrees of freedom of the simulated system, allowing significant longer simulations and more replicates to be simulated using state-of-the-art computational resources. It remains possible in principle to calculate the initial model in an all-atom fashion to further examine the free energy changes in the formation of Fv dimer complexes but the computational resources required in practice are astronomical.

Solvation free energy calculations are computationally costly due to the series of intermediate points are required to calculate the overall changes in the free energies. The computational resources required to conduct such calculations exceed the original resources used to generate the trajectory. The solvation free energies calculated do not seem able to reflect the interaction events occurred in the coarse-grained trajectory.

To the best of our knowledge, there is currently no direct application of MM-GBSA or MM-PBSA for the analysis of MARTINI simulations in the literature due to lack of fundamental biophysical data underpinning these free-energy calculation methods. In the current work, the final snapshots of MARTINI CG simulation were converted to all-atom representations following a short MD simulation before conducting MM-PBSA calculations. The interaction energy values calculated from the MARTINI model did not correlate to the binding energy calculated with MM-PBSA. A further 10 ns of all-atom MD simulations were conducted on selected examples. Results indicated there were large fluctuations in the polar solvation energy term which remained highly positive and dominating the overall binding energies. Extending the simulation to 100 ns did not seem to stabilise the polar solvation energy term. Instead, the complex continues to deviate from the original structure as confirmed by analysing the RMSD.

This work was unable to find any differences in solvation free energies before and after the formation of the Fv dimer complex. In addition, no correlations between the calculated interaction energy values from MARTINI model and the binding energy calculated with MM-PBSA were found. It is important for computational modelling to provide practical, efficient and accurate methodology to model the dynamics of proteins. Evaluating transient protein dimer complexes with Lennard-Jones and Coulomb potentials remains a viable indicator for comparing protein-protein interactions between dimers complexes formed as this method uses minimal computational resources with no further post-simulation processing being required.

6. Coarse-Grained MD Simulations for Excipient Identification

6.1. Introduction

Amino acids are used in protein formulations and their effects have been well studied. One common example of amino acid excipient in protein formulations is arginine which is known for preventing aggregation, increasing solubility, reducing surface tension and reducing viscosity [105-106]. Beneficial effects of amino acid excipients are often seen in high concentration (i.e. 100 mM or above) so their application in therapeutic liquid protein formulations could be limited by the osmolality issues [188]. Hence, there is growing interest in protein formulation to discover new excipients with improved properties. Dipeptides are of particularly interest due to their low toxicities. Among all dipeptides, arginine-containing dipeptides are among the most studied in protein formulation [189-190]. Peptides are attractive molecules as potential excipients because there is a wide diversity of structures that remained to be explored which can potentially offer improved selectivity towards aggregation-prone regions of pharmacologically active proteins.

Virtual Screening using docking is an effective solution to reduce time and cost of screening a large library of small molecules for lead identification. However there are several limitations for its application to protein formulation. Many docking tools are highly dependent on the training set and often are only applicable to specific cases for drug discovery. Solvation and charges are important for determining protein-protein interactions in a formulation but water and ions are not included in these tools due to additional complexity and computational cost. Accuracy is also highly dependent on the quality of the inputted protein structures where predicted protein structures from homology modelling may lead to false positive results [191]. Coarse-grained MD

simulations can potentially overcome these challenges and predict the time-dependent behaviour of a system as shown in Chapter 4.

Research presented in this chapter builds upon the protocol based on MD simulation using the MARTINI CG force field as described in Chapter 4. MEDI-578 was again used as a model molecule to conduct virtual screening of excipients based on hydrophilic dipeptides, including ALA, ARG, GLN, GLY, HIS, LYS, PRO and SER. Only hydrophilic dipeptides were selected due to the high propensity for aggregation for hydrophobic peptides during synthesis, purification and storage. The *in-silico* results of the dipeptides were evaluated with DLS to study the antibody-antibody interactions in the presence of selected dipeptides to identify if a correlation exists between *in-silico* data and experimental results. The experience gained are particularly useful to guide the design of formulation of MEDI1912 through homology modelling of MEDI-578 Fv fragment to understand mAb-mAb interactions and to propose potential dipeptides to overcome formulation challenges as described in the earlier formulation work with this molecule [152]. This chapter also aim to establish whether it is advantageous to include the whole Fab fragment to improve the quality of the model and more importantly to identify whether there are any antibody-antibody interaction hot-spots at the constant domains of the Fab fragment.

6.1.1. MEDI1912

MEDI1912 was generated by *in-vitro* affinity maturation of MEDI-578. Whilst the binding affinity and specifically to NGF was increased, formulating MEDI1912 was challenging due to self-association, poor solubility, high viscosity and aggregation propensities [152]. SAP calculations identified three exposed hydrophobic residues, TRP30, PHE31 and LEU56, which were reverted back to the corresponding MEDI-578

amino acid residues SER30, THR31 and THR56 respectively to create a variant named as MEDI1912_STT. MEDI1912_STT displayed improved serum half-life, reduced mAb-mAb interactions, reduced aggregation and reduced viscosity [152].

6.1.2. Motavizumab

Motavizumab is a humanised IgG1 mAb that targets respiratory syncytial virus (RSV) fusion glycoprotein to treat serious lower respiratory tract disease caused by RSV in children [192]. Motavizumab has an approximately 76-fold greater binding avidity for RSV fusion glycoprotein than palivizumab, and motavizumab also shown a 19-fold improvement in neutralization of RSV *in-vitro*.

Despite the increase in binding affinity for RSV fusion glycoprotein, motavizumab did not show significant formulation challenges as in the development of MEDI1912. The marketed product of palivizumab (SYNAGIS) was formulated as 100 mg palivizumab in 1 mL 25 mM histidine buffer with 1.6 mM glycine at pH 6.0. Formulation of motavizumab is similar to palivizumab. The formulation of motavizumab used in the phase 2 clinical trial was formulated as 100 mg motavizumab in 1 mL 25 mM histidine buffer at pH 6.0 [193]. The X-Ray crystallography structure of the motavizumab Fab fragment had been resolved at a high resolution of 1.9 Å (PDB ID: 3QWO) making it is an ideal molecule to study alongside MED-I578 and MEDI1912 [192].

6.2. Materials and Methods

6.2.1. Antibody Models

6.2.1.1. MEDI-578 Fv Fragment

X-ray crystallographic structure of MEDI-578 Fv fragments-β-NGF complex at resolution of 3.4 Å (PDB ID: 5JZ7) was taken from protein data bank and the structure of MEDI-578 Fv fragment was used directly in this study [152].

6.2.1.2. Homology Model of MEDI1912 Fv Fragment

The three-dimensional structure of MEDI1912 has not been resolved. Therefore the three dimensional model of MEDI1912 was developed based on the MEDI-578 Fv fragment (PDB ID: 5JZ7) using Modeller 9.16 [194].

6.2.1.3. Motavizumab Fab

The three-dimensional structure of motavizumab epitope-scaffold bound to motavizumab Fab was obtained from protein data bank (PDB ID: 3QWO) and the structure of motavizumab Fab was used directly in this study [192].

6.2.2. Virtual Screening Simulations Setup

The dimerisation simulations were prepared using the DAFT protocol [176] using MARTINI protein force field version 2.2 [169-170] with ElNeDyn [171] applied on the Fv fragments using a cut-off distance of 0.9 nm and a force constant of 500 kJ mol^{-1} nm^{-2}. In each simulation, two antibody fragments were placed at the same initial distance of 3.0 nm and the two antibody fragments were randomly rotated, yielding different relative starting orientations for individual simulations. Systems were subsequently solvated with standard MARTINI water.

The all-atom three-dimensional structures of the dipeptides were generated through a VMD tcl script [195] and subsequently processed with martinize to generate the CG structures and topologies with a bash script (Supplementary Information 8.9. – 8.11.) [170]. Ten molecules of MARTINI dipeptide molecules were added to individual system by replacing standard MARTINI water beads and neutralised with GROMACS 2016.3 [140]. For MEDI-578 and MEDI1912, two Fv fragments were added into the simulation box and ionised to equivalent of 100 mM NaCl. In case of motavizumab, protonated histidine molecules were added to the simulation box to achieve an equivalent of 25 mM and subsequently neutralised. The entire process was automated by a bash script (Supplementary Information 8.12., 8.15., 8.18.)

All systems were minimised with martinate [178] with GROMACS 2016.3 [140] for 500 ps at 1 fs/step before undergoing a 100 ps NPT equilibration MD simulation at 20 fs/step. The final production runs were performed with a time step of 20 fs with a snapshot taken every 25000 steps. Periodic boundary conditions were used, a constant temperature of 300 K was maintained and pressure was controlled at 1 bar using V-rescale thermostat and Berendsen barostat respectively. Electrostatic and van der Waals interactions were calculated using a Verlet cutoff-scheme with Potential-shift-verlet as modifier with a cut-off of 1.8 nm.

6.2.3. Sample Preparation

MEDI-578 was supplied in 50 mM sodium acetate, 100 mM sodium chloride at pH 5.5 and Motavizumab was supplied in 25 mM histidine at pH 6.0 from MedImmune (Cambridge, United Kingdom). Glacial acetic acid, ALA-PRO, GLY-GLN, GLY-GLY, HIS-SER and histidine hydrochloride were purchased from Sigma-Aldrich (Poole, United Kingdom), ARG-ARG, HIS-HIS and HIS-LYS were purchased from Bacham

(Bubendorf, Switzerland), sodium acetate trihydrate and sodium chloride were purchased from Fisher Chemicals (Loughborough, United Kingdom).

The absolute concentrations of antibodies were determined by measuring the absorbance at 280 nm in triplicate using NanoDrop (ThermoFisher Scientific, Loughborough, United Kingdom) calculated with their respective extinction coefficient (MEDI-578: 1.68 mL mg^{-1} cm^{-1}, Motavizumab: 1.47 mL mg^{-1} cm^{-1}) prior to use. Formulations with dipeptides were prepared by dissolving the dipeptide in freshly prepared buffer and added to the stock formulation to achieve an antibody:dipeptide molar ratio of 1:5. Subsequently, all formulations were prepared with pH adjustment followed by filtering through 0.22 µm sterile filters.

6.2.4. Dynamic Light Scattering

The DLS measurements were made on a DynaPro PlateReader II system (Wyatt Technology, Santa Barbara, California, United States) with an 826.1 nm laser using a Corning 3540 384-well non-binding clear flat-bottom polystyrene plate (Corning, New York, United States). Formulations were diluted with filtered fresh buffer for DLS measurements to antibody concentrations of 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 mg/mL. Samples of all tested formulations in each concentration was added to sextuplicate wells and each well contained 30 µL of the sample. Twenty acquisitions were collected for each well over 5 s. Only particles with hydrodynamic radius between 2 nm and 30 nm were considered for analysis.

The diffusion interaction parameter, k_D for antibody was calculated according to the following equation by plotting the diffusion coefficient against the concentration of

antibody. The value of k_D was then calculated by dividing $k_D D_0$ (the slope) by D_0 (the Y-intercept) as described in Chapter 3.

$$D = D_0(1 + k_D c) \quad (3.1)$$

where D is the diffusion coefficient at a given protein concentration c and D_0 indicates the diffusion coefficient when c is at 0.

6.3. Results and Discussions

6.3.1. Screening Method Development

The amount of time and computational resources required to screen 64 dipeptides is enormous even with the use of coarse-graining. Therefore it was necessary to achieve a fine balance to obtain simulation accuracy with the available computational resources. Systems containing 2 MEDI-578 Fv fragments contain roughly 10000 beads that are comprised of 2 MEDI-578 Fv fragments, roughly 9250 water, 68 NA+ ion and 66 CL- beads.

The aim of this study was to establish what can be potentially achieved with limited computational resources using a smaller library of peptides derived from ALA, GLY and HIS, to generate a library of a total 9 dipeptides. Three sets of simulations were conducted, 1) 128 simulations of 1024 ns, 2) 256 simulations of 512 ns and 3) 512 simulations of 256 ns. The standard error of the mean (SEM) was high for 512 simulations of 256 ns compare to other sets of simulations (Table 10) suggesting 256 ns was too short to determine the formation of dimers in presence of dipeptides.

The plateau value calculated from the mean interaction energy fitted with the model function were similar in 128 simulations of 1024 ns and 256 simulations of 512 ns and the overall pattern calculated from these simulations were similar (Figure 37). Although the SEMs were higher in 256 simulations of 512 ns, the higher number of simulations would able to improve the resolutions of the MEDI-578 Fv fragment dimers about the distribution of interacting pairs and to dissect the molecular mechanisms of the dipeptides. For these reasons, 256 simulations of 512 ns were used to screen the dipeptides.

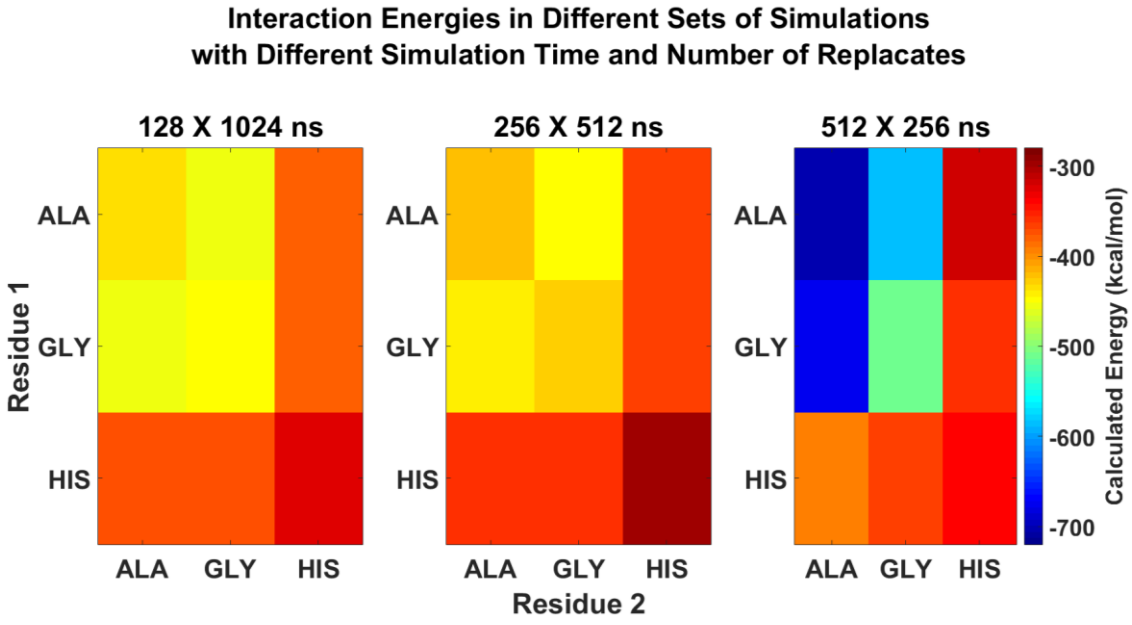


Figure 37: Visual representations of the calculated interaction energy between MEDI-578 Fv fragment dimers in systems with 10 molecules of dipeptides simulated for a) 128 simulations of 1024 ns, b) 256 simulations of 512 ns, c) 512 simulations of 256 ns. The overall pattern of interaction energies calculated from a) and b) were similar.

	128 × 1024 ns		256 × 512 ns		512 × 256 ns	
Dipeptide	Energy (kJ/mol)	SEM	Energy (kJ/mol)	SEM	Energy (kJ/mol)	SEM
AA	-437.5	0.52	-419.2	0.73	-704.3	13.52
AG	-453.0	0.27	-443.4	0.48	-675.1	10.49
AH	-371.1	0.35	-361.7	0.41	-390.9	4.11
GA	-454.6	0.24	-451.8	1.09	-586.9	6.44
GG	-446.5	0.34	-425.9	0.59	-508.7	3.97
GH	-374.5	0.19	-359.8	0.63	-369.0	1.45
HA	-376.8	0.22	-363.8	0.44	-321.0	3.22
HG	-382.2	0.39	-368.5	0.33	-360.1	2.73
HH	-322.9	0.28	-297.3	0.33	-337.2	1.97

Table 10: The plateau value calculated from mean interaction energy fitted with the model function and the standard error of the mean (SEM) in systems with 10 molecules of dipeptides simulated for a) 128 simulations of 1024 ns, b) 256 simulations of 512 ns, c) 512 simulations of 256 ns. The SEM is lower with longer simulations with less simulation setups.

6.3.2. MEDI-578 Dipeptide Library Screening

As demonstrated in the method development session, addition of different dipeptides to simulations with 2 MEDI-578 caused changes the interaction energies with different effects. In this screening study, eight hydrophilic amino acid residues were considered, ALA, ARG, GLN, GLY, HIS, LYS, PRO and SER. The pKa of histidine residues is approximately 6 and the pH in many formulations range from 5 to 8. At a pH close to 6, the amounts of ionised imidazole groups from histidine residues are derived from the Henderson–Hasselbalch equation and so histidine can either be protonated or neutral so both cases have to be considered. Therefore both protonated and neutral histidine residues in histidine-containing dipeptides have been have been studied, generating a total of 81 set of simulations. Protonated and non-protonated histidine residues were abbreviated as HSP and HIS respectively.

From the studies described in Chapter 4, the calculated mean interaction energies in MEDI-578 Fv fragments dimer complexes were calculated using the same method with

1024 simulations was -417.2 kJ/mol. Seventeen out of eighty-one sets of dipeptides displayed negative interaction energies, suggesting the addition of these dipeptides energetically favour tighter Fv fragment-Fv fragment interactions. None of these seventeen dipeptides contained the neutral histidine, HIS, or charged residues such as ARG, HSP and LYS (Table 11). These charged amino acids are commonly used in the final formulation of therapeutic antibodies (Table 1). HSP-HSP achieved the most positive calculated mean interaction energies among all these dipeptides. Arginine, histidine and lysine are known to reduce solution viscosity and reduce aggregates in protein formulations [93, 104-106, 196]. Among histidine-containing dipeptides, dipeptides with charged HSP residues usually result in a more positive calculated mean interaction energies compare to the neutral HIS residues.

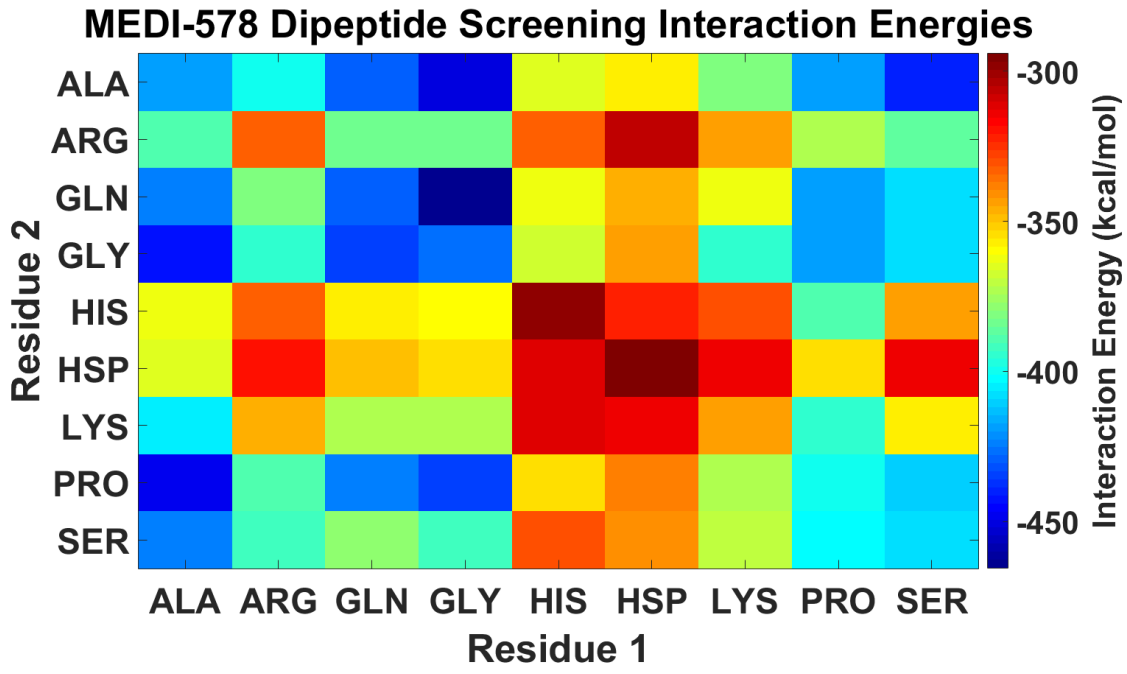


Figure 38: Visual representations of the plateau values calculated from mean interaction energies fitted with the model function in systems with 10 molecules of dipeptides simulated for 256 simulations of 512 ns with 2 MEDI-578 Fv fragments. Brown colour indicates the simulated dipeptide resulted in less energetically favourable dimer complexes and other colours refer to dimer complexes formed are more energetically favourable according to scale shown on the left.

Interaction Energies (kJ/mol)										
Residue 2	ALA	-419.2	-400.8	-429.5	-451.8	-363.8	-357.5	-380.6	-418.2	-441.0
	ARG	-390.0	-331.9	-383.4	-383.1	-332.4	-305.5	-342.8	-372.0	-386.8
	GLN	-423.4	-382.2	-428.2	-465.4	-361.9	-347.3	-362.9	-419.2	-408.8
	GLY	-443.4	-395.3	-435.5	-425.9	-368.5	-344.4	-393.8	-418.2	-407.9
	HIS	-361.7	-332.7	-356.2	-359.8	-297.3	-323.0	-330.5	-389.1	-343.1
	HSP	-364.9	-320.2	-350.1	-355.7	-312.4	-294.2	-313.7	-355.4	-313.3
	LYS	-406.5	-346.3	-373.4	-373.6	-311.9	-314.6	-342.8	-393.4	-358.4
	PRO	-447.6	-390.4	-424.8	-435.8	-353.2	-337.6	-373.1	-400.4	-410.0
	SER	-423.7	-391.3	-377.2	-391.2	-331.0	-340.9	-370.6	-403.2	-407.8
		ALA	ARG	GLN	GLY	HIS	HSP	LYS	PRO	SER
	Residue 1									

Table 11: The plateau value calculated from mean interaction energy fitted with the model function in systems with 10 molecules of dipeptides simulated for 256 simulations of 512 ns with 2 MEDI-578 Fv fragments. Among all tested dipeptides, addition of HSP-HSP have resulted in dimer complexes with less negative mean interaction energies. To lesser extent, similar effects were seen in dipeptides containing the neutral histidine, HIS, or charged residues such as ARG, HSP and LYS.

Wang and co-workers have studied the effect of seven amino acids, ALA, GLY, SER, PRO, ARG, HIS and LYS, on the viscosity in formulations of two antibodies [197]. In that study, the concentration of the first antibody is 200 mg/mL and the concentration of the second antibody is 165 mg/mL. Both antibodies were formulated in 20 mM histidine buffer at pH 6.0. Viscosities of the first antibody formulations with ALA, GLY, SER and PRO were within 10% relative viscosity of the original antibody formulation at 200 mg/mL in 20 mM histidine buffer at pH 6.0. For the second antibody these amino acids only shown a reduction in of relative viscosity by around 40% whereas in both antibody formulations, charged amino residues ARG, HIS and LYS reduce the relative viscosity by approximately 30 % and 70% respectively [197]. Antibody-excipient interactions can occur through charge-charge and hydrophobic interactions. The effects of individual excipient depends on the mechanisms of antibody-antibody interactions.

6.3.2.1. Interaction Patterns

Energetically favourable dimers are liable to be very stable and preventing formations of these dimers can reduce the extent of antibody-antibody interactions. In the previous study described in Chapter 4, the top 10% with most negative calculated mean interaction energies of MEDI-578 Fv fragment-Fv fragment complexes mostly interact through THR76 on the heavy chain. In this screening study, both protonated and non-protonated dihistidines have achieved the most positive calculated mean interaction energies among all screened dipeptides. In both cases, the signals from THR76 on the heavy chain are largely reduced in the presence of dihistidine compared with just two Fv fragments alone (Figure 39-42). For dipeptides those have achieved a more negative calculated mean interaction energies such as GLY-GLN (Figure 43), a more pronounced signal from THR76 on the heavy chain was observed.

In the case of HIS-HIS, MEDI-578 Fv fragment-Fv fragment interactions were observed through the region around VAL11 on the heavy chain whereas ALA13 on the light chain is more common with HSP-HSP (Figure 39-40). Synergistic effects were observed in HIS-HSP and HSP-HIS with interactions events through VAL11 on the heavy chain and ALA13 on the light chain being almost equal (Figure 41-42). The difference may be due to the positively charged HSP-HSP which is more energetically favourable to bind to negatively charged GLU10 on the heavy chain. Similarly, the region from the THR76 on the heavy chain may be additionally protected by masking the negative charge from GLU74 on the heavy chain with positively charged HSP-HSP. Similarly, a greater signal reduction was seen in protonated HSP-SER in the region from the THR76 on the heavy chain compared to neutral HIS-SER (Figure 44-45). This suggests using dihistidines in formulations with pH close to 6 may be beneficial as the four species HIS-HIS, HIS-HSP, HSP-HIS and HSP-HSP have different propensities towards both negatively charged and neutral residues on the antibody.

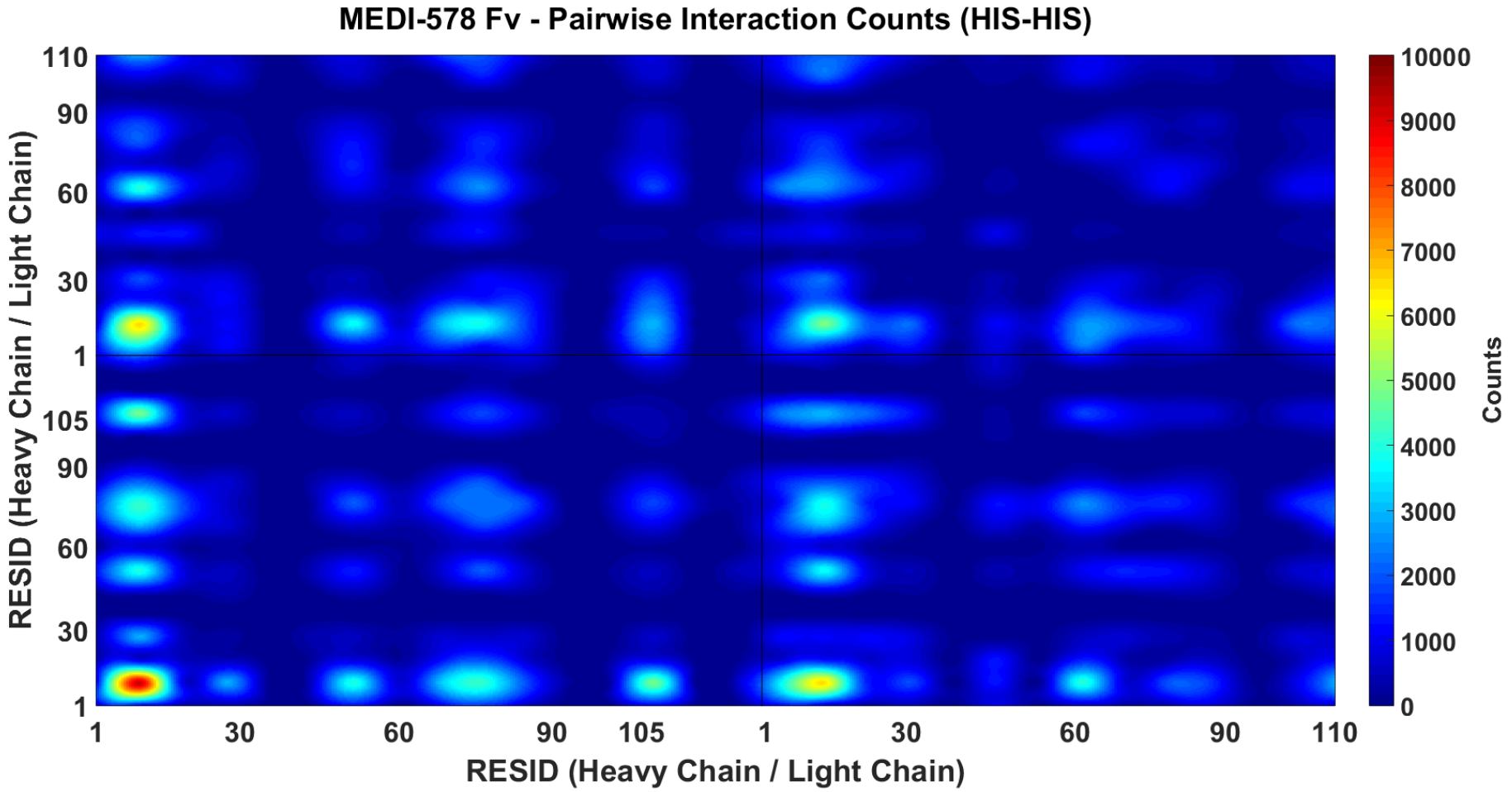


Figure 39: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI-578 Fv fragments and 10 molecules of HIS-HIS in 100 mM NaCl. Interaction events through the THR76 region on the heavy chain were reduced compared with the system without dipeptides. The VAL11 region on the heavy chain was the most common interaction hot-spot.

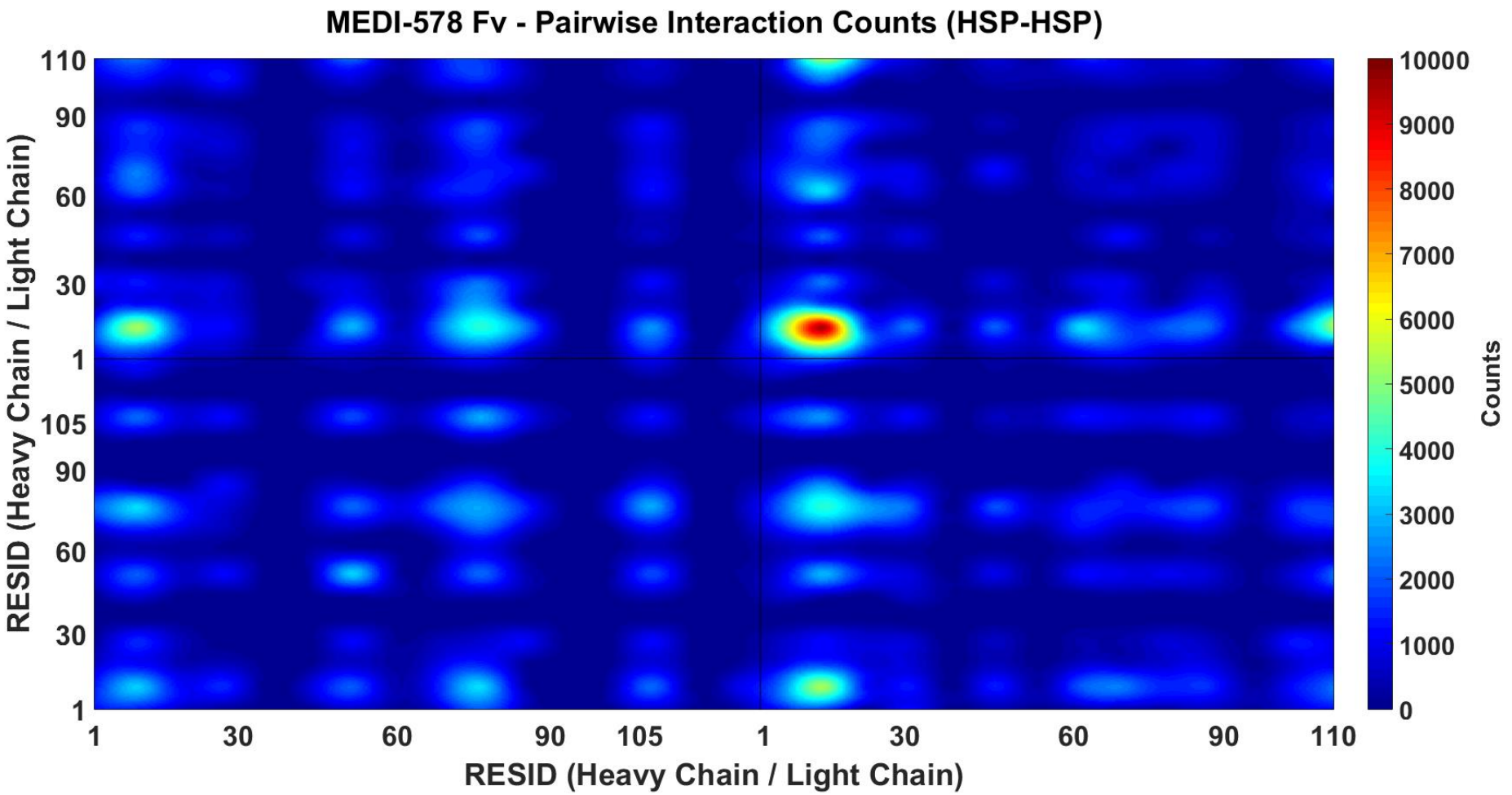


Figure 40: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI-578 Fv fragments and 10 molecules of HSP-HSP in 100 mM NaCl. Interaction events through the THR76 region on the heavy chain were reduced compared with the system without dipeptides. The ALA13 region on the light chain was the most common interaction hot-spot.

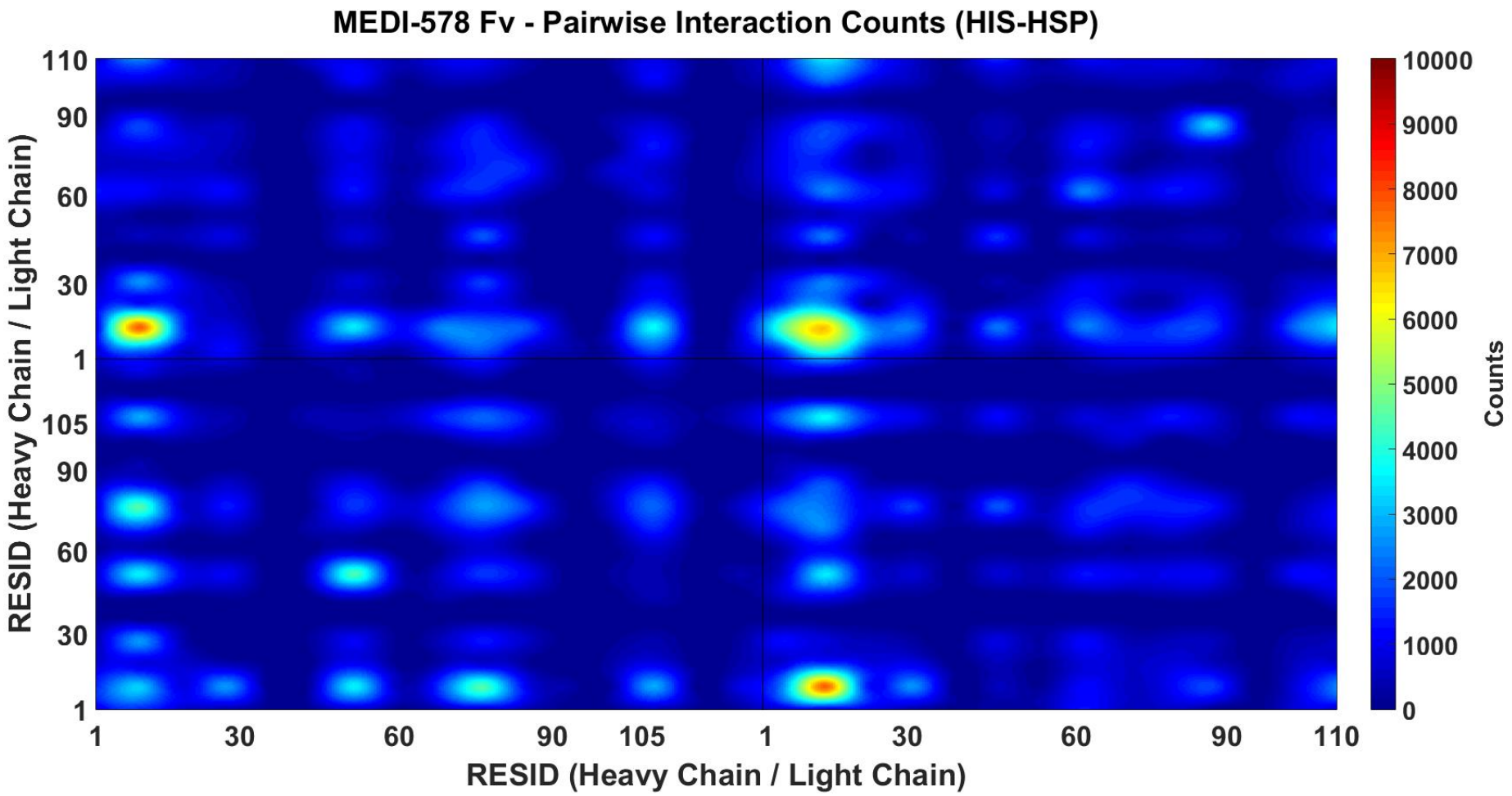


Figure 41: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI-578 Fv fragments and 10 molecules of HIS-HSP in 100 mM NaCl. Interaction events through the THR76 region on the heavy chain were reduced compared with the system without dipeptides. The VAL11 region on the heavy chain and ALA13 region on the light chain were the most common interaction hot-spots.

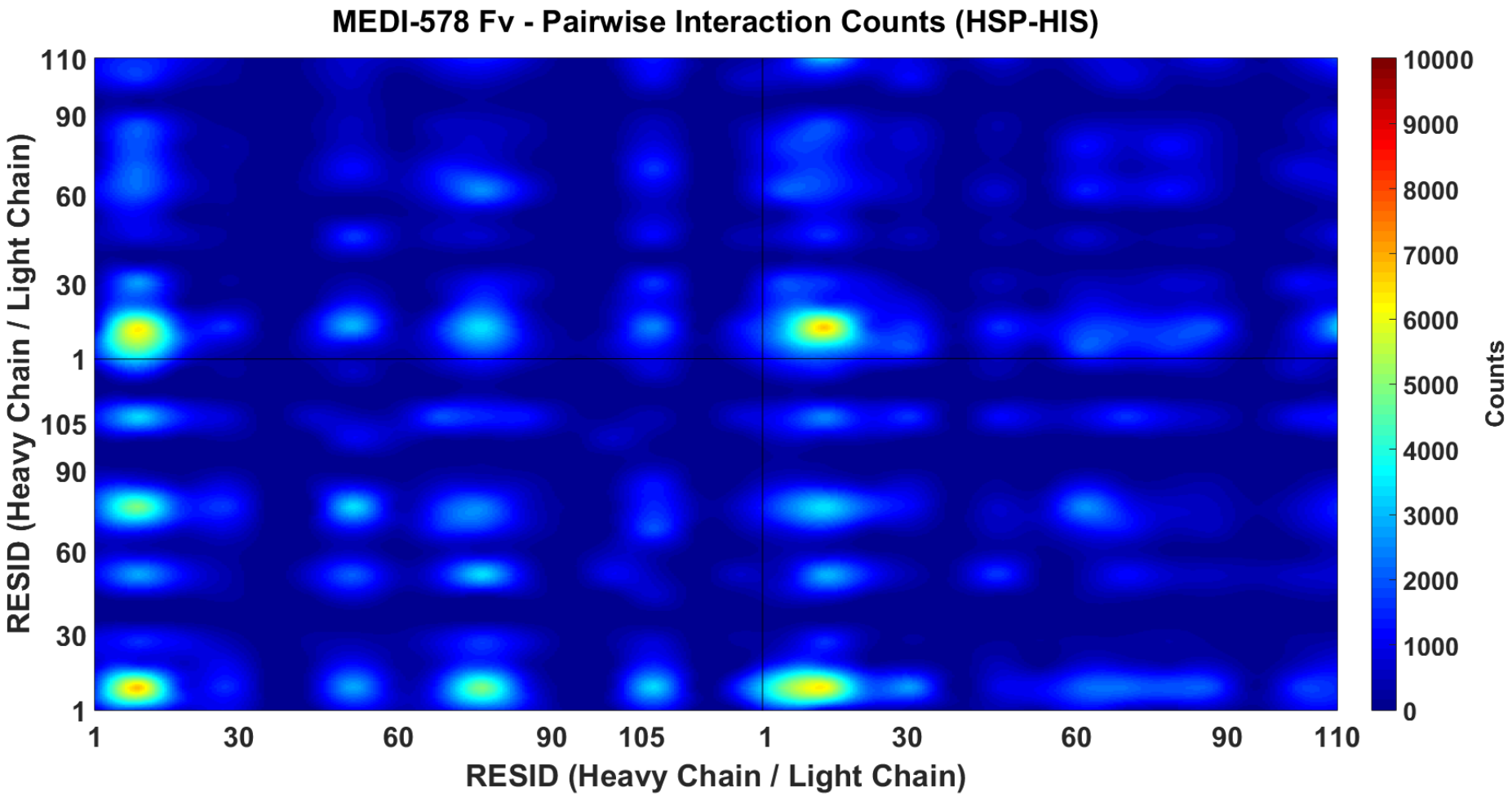


Figure 42: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI-578 Fv fragments and 10 molecules of HSP-HIS in 100 mM NaCl. Interaction events through the THR76 region on the heavy chain were reduced compared with the system without dipeptides. The VAL11 region on the heavy chain and ALA13 region on the light chain were the most common interaction hot-spots similar to HIS-HSP.

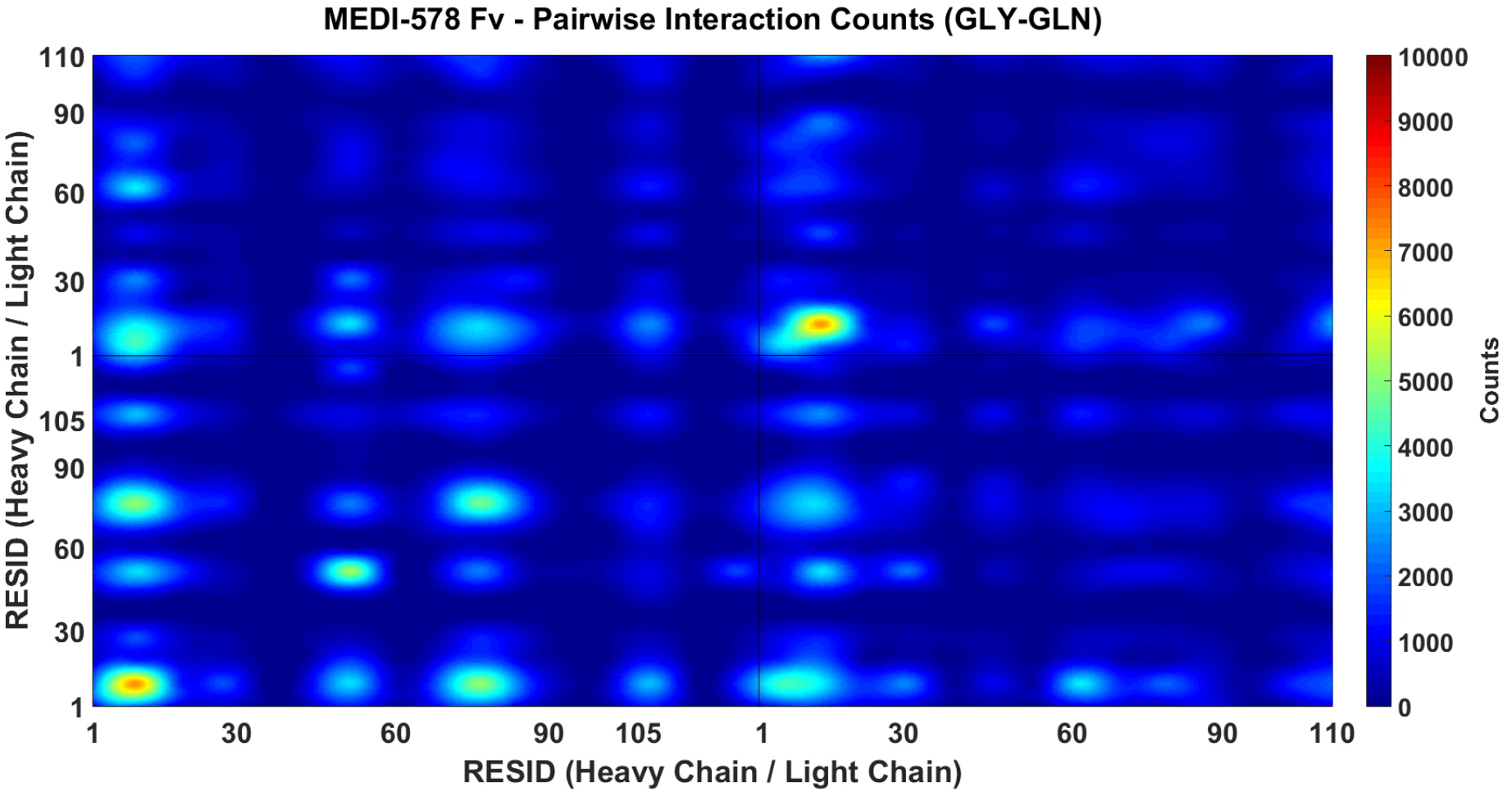


Figure 43: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI-578 Fv fragments and 10 molecules of GLY-GLN in 100 mM NaCl. Interaction events through the THR76 region on the heavy chain were increased compared with the system without dipeptides. Interaction events through the THR76 region on the heavy chain were increased compared with the system without dipeptides. In an earlier study, the THR76 region on the heavy chain was identified as the region that forms most energetically favourable MEDI-578 Fv dimer complexes.

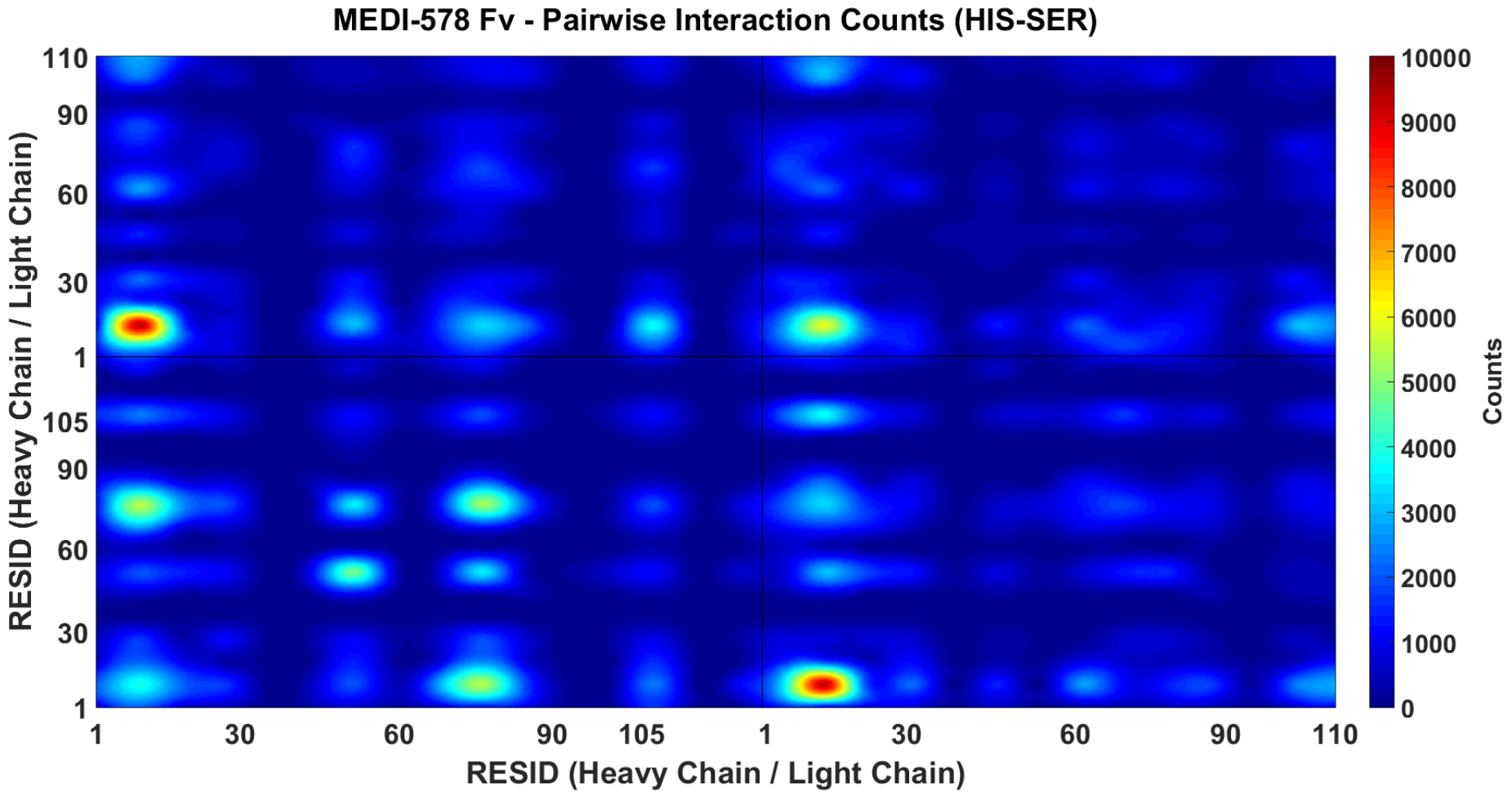


Figure 44: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI-578 Fv fragments and 10 molecules of HIS-SER in 100 mM NaCl. Interaction events through THR76 on the heavy chain happened more frequent than with HSP-SER.

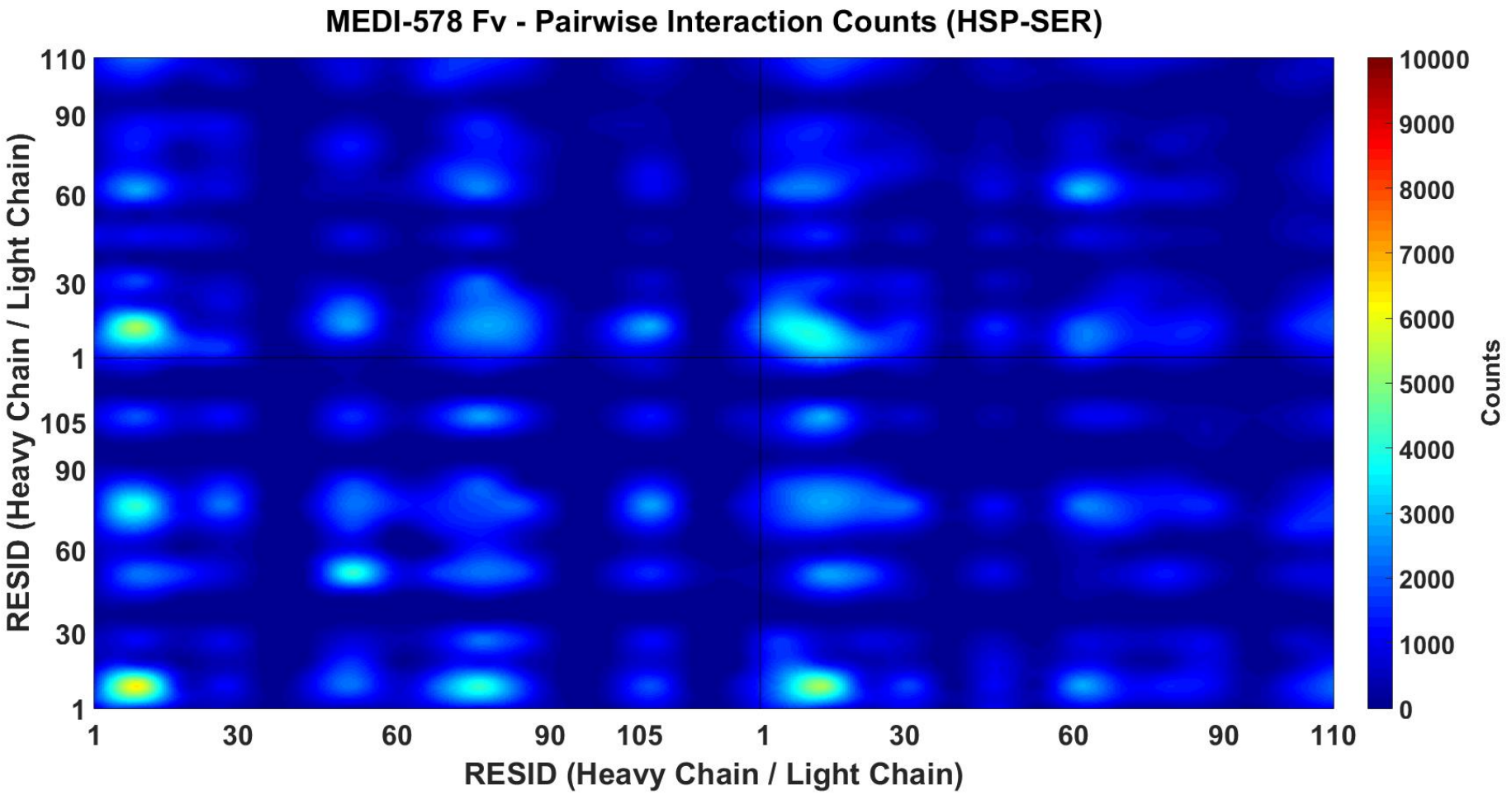


Figure 45: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI-578 Fv fragments and 10 molecules of HSP-SER in 100 mM NaCl. Interaction events through THR76 on the heavy chain happened less frequent than with HIS-SER.

6.3.2.2. Diffusion Interaction Parameter

Antibody-antibody interactions were measured with DLS and k_D values were calculated in each of the MEDI-578 formulations with dipeptides ALA-PRO, GLY-GLY, GLY-GLN, HIS-HIS, HIS-LYS, HIS-SER and ARG-ARG at 1:5 molar ratio. All samples with dipeptides displayed a negative k_D which suggests net attractive interactions between antibody molecules (Table 12). The k_D value for MEDI-578 in just 50 mM sodium acetate, 100 mM sodium chloride at pH 5.5 was determined to be -9.03 mL/g in Chapter 3. The addition of charged dipeptides, HIS-HIS, HIS-LYS, HIS-SER and ARG-ARG increased the k_D values suggesting the addition of these dipeptides reduces the net attractive interactions between MEDI-578 molecules .

A positive correlation was observed between the calculated mean interaction energies and the diffusion interaction parameters of different the MEDI-578 formulations with dipeptides (Figure 47) suggesting this model offers a good predictive power of antibody-antibody interactions through computing the interaction energies of the Fv fragment dimers formations. At pH 5.5, the majority of histidine residues were protonated and the regression model based on the calculated mean interaction energies of systems with protonated histidine-linked dipeptides achieved a predicted R^2 value of 0.9484 which is slightly higher than 0.9384 in model based on the values from neutral histidine-linked dipeptides (Figure 47), indicating that the MEDI-578-MEDI-578 interactions measured with DLS increase linearly with more negative calculated plateau values of MEDI-578 Fv fragment-MEDI-578 Fv fragment interactions in this coarse-grained model.

Dipeptide	Calculated mean interaction energies (kJ/mol)	Measured k_D (mL/g)
ALA-PRO	-447.6	-12.03
GLY-GLY	-425.9	-10.41
GLY-GLN	-465.4	-12.33
HIS-HIS (HSP-HSP)	-297.3 (-294.2)	-6.32
HIS-LYS (HSP-LYS)	-311.9 (-314.6)	-6.35
HIS-SER (HSP-SER)	-331.0 (-340.9)	-8.61
ARG-ARG	-331.9	-8.46

Table 12: The calculated plateau value from the mean interaction energy distributions in systems with different dipeptides and the measured k_D values in different formulations of MEDI-578 containing these dipeptides.

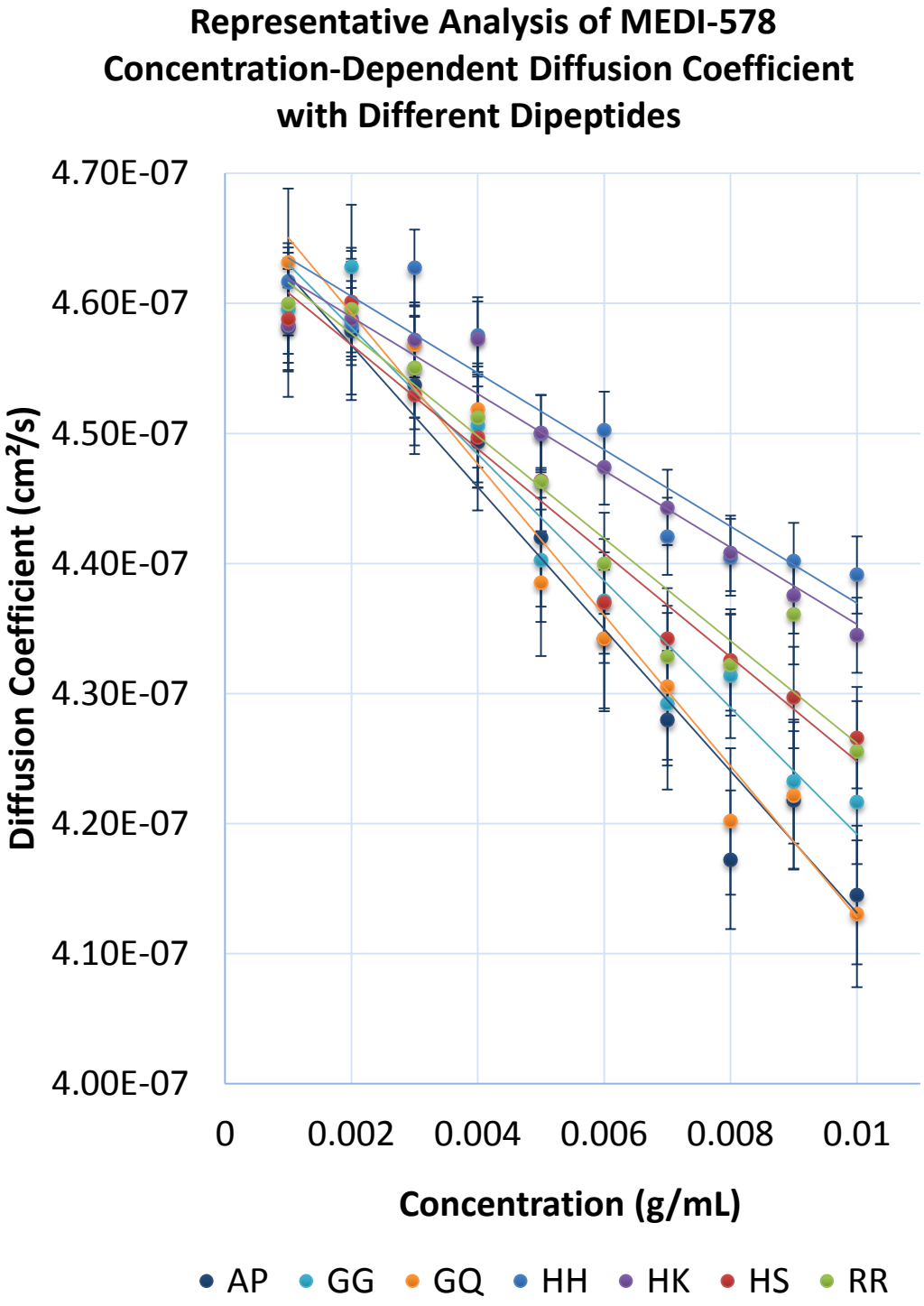


Figure 46: Change in diffusion coefficient versus concentration of MEDI-578 in with the following peptides at 1:5 molar ratio a) ALA-PRO, b) GLY-GLY, c) GLY-GLN, d) HIS-HIS, e) HIS-LYS, f) HIS-SER and g) ARG-ARG. The diffusion coefficient decreases as the concentration increase in all formulations of MEDI-578 with dipeptides suggesting attractive interactions between MEDI-578 molecules in all formulations.

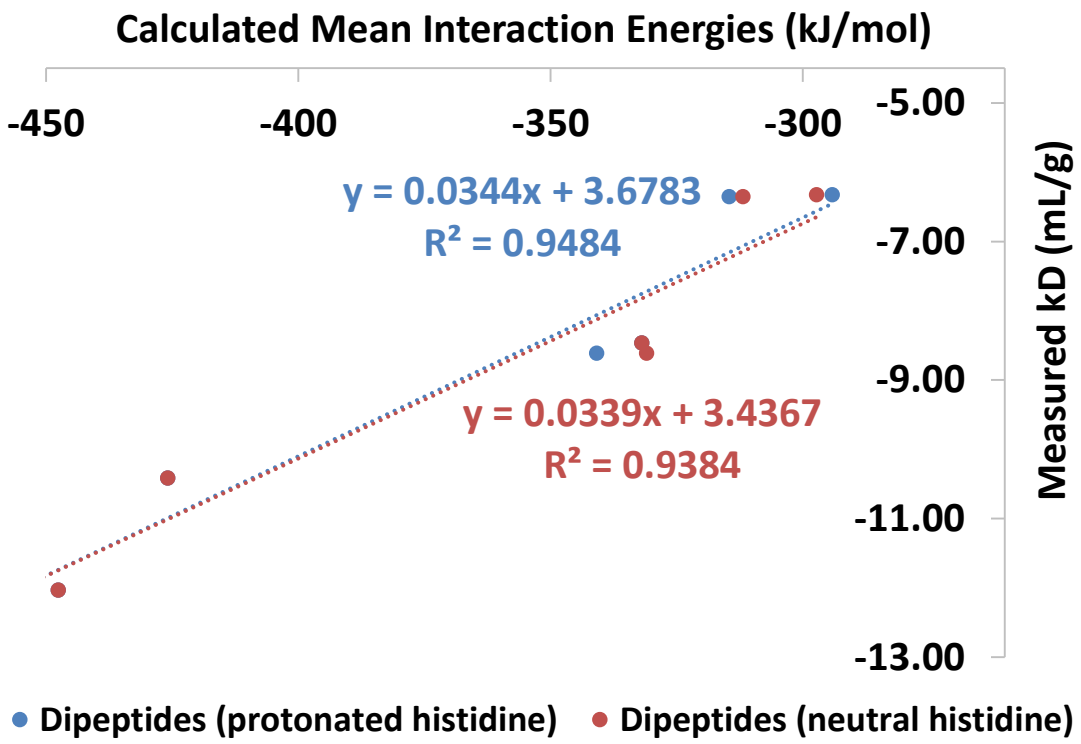


Figure 47: Correlations between the calculated mean interaction energies from CG simulations and measured k_D with different MEDI-578 formulations. A positive correlation was observed between the calculated mean interaction energies and the diffusion interaction parameters of different MEDI-578 formulations with dipeptides.

6.3.3. MEDI1912 Dipeptide Library Screening

6.3.3.1. MEDI1912 Dimerisation

A total of 256 independent simulations containing two Fv fragments of MEDI1912 were conducted for 512 ns based on a homology model built from the structure of MEDI-578. Similar to the simulations of MEDI-578 Fv fragments, the dimerisation of MEDI1912 Fv fragments were observed in most simulations. Approximately 5% of all simulations did not dimerise during the simulation period (Figure 48). The calculated plateau value from the mean interaction energy distributions for MEDI1912 Fv fragments dimerisation was -423.1 kJ/mol which was more negative than those calculated from MEDI-578 Fv fragments dimerisation (-417.2 kJ/mol). This suggests MEDI1912 Fv fragments dimerisation was more favourable energetically and possibly explain the increased aggregation propensity of MEDI1912.

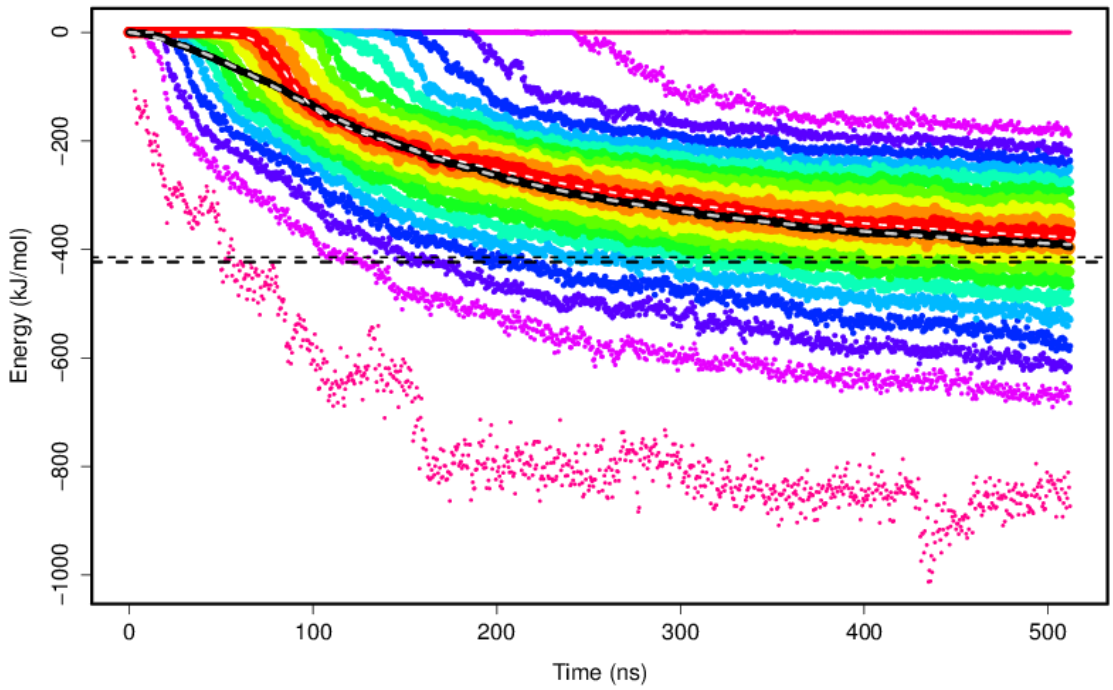


Figure 48: Interaction energy between MEDI1912 Fv fragment dimers in 100 mM NaCl over time. The calculated plateau value of the means for MEDI1912 Fv fragment dimerisation was determined to be -423.1 kJ/mol.

Interactions through aggregation hot-spots identified in an earlier study, TRP30, PHE31 and LEU56 [152], were observed and the frequency of interactions through these aggregation hot-spots were greater than those simulations from MEDI-578 Fv fragments dimerisation. When analysing the overall interaction patterns from MEDI1912 Fv fragments simulations, the interactions through the regions close to VAL11 from the heavy chain were the strongest. Other interaction sites include regions near to TRP30, PHE31, LEU56, ILE59, VAL78, LEU105 on the heavy chain and ALA13, ASN31, ARG62 and LEU110 on the light chain (Figure 49). Stronger signals were detected in regions around VAL11, TRP30, PHE31, LEU56, VAL78 and LEU105 on the heavy chain and ASN31 on the light chains in the top 10% simulations in terms of most negative interaction energies (Figure 50). These interacting regions were consistent with the identified interaction regions within the molecule of MEDI1912 using hydrogen/deuterium exchange - mass spectrometry (HDX-MS) in the earlier aggregation hot-spots identified study by Dobson and co-workers [152].

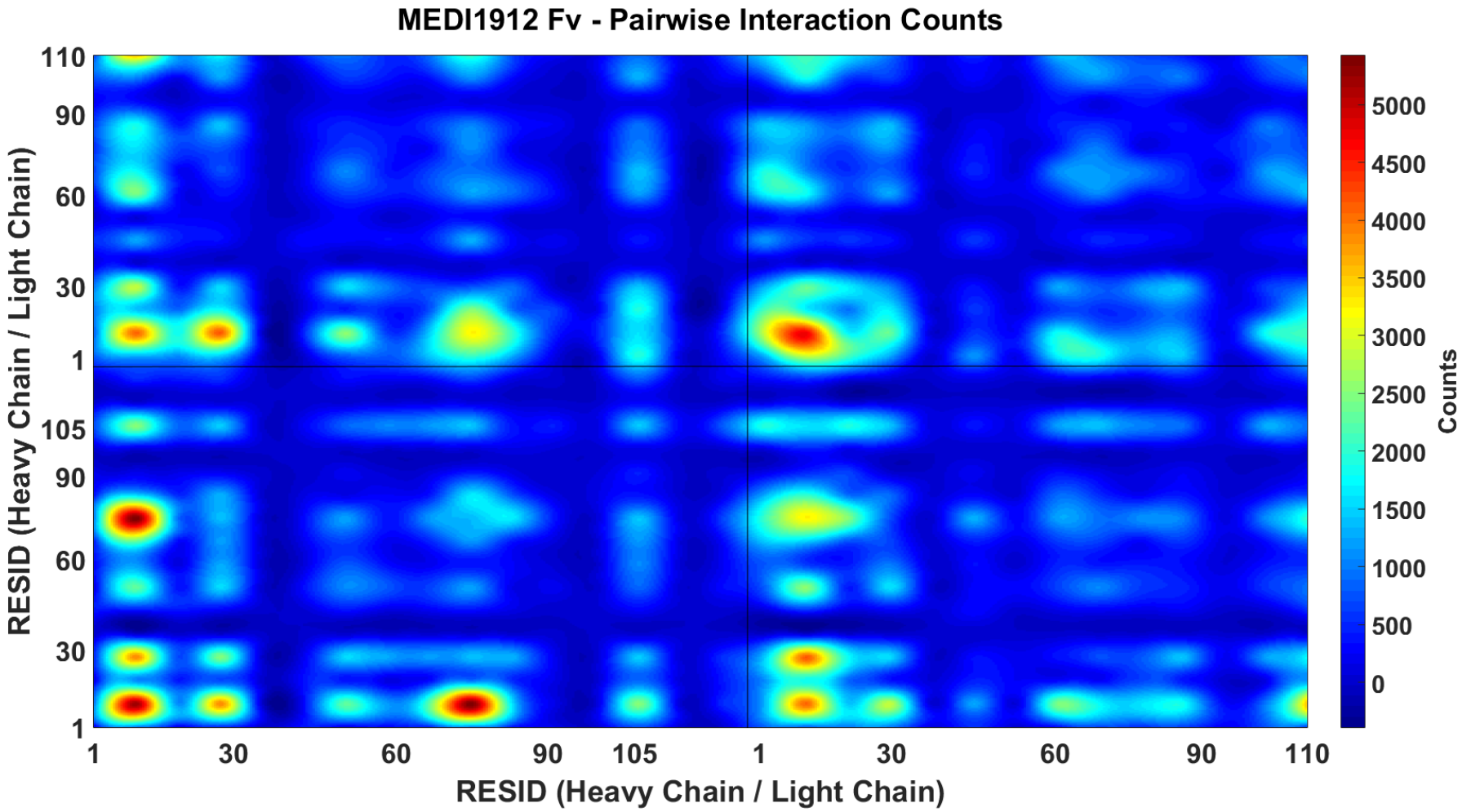


Figure 49: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI1912 Fv fragments from a homology model in 100 mM NaCl. A number of interaction hot-spots were identified, including TRP30, PHE31, LEU56, ILE59, VAL78, LEU105 on heavy chain and ALA13, ASN31, ARG62 and LEU110 on the light chain. TRP30, PHE31 and LEU56 were mutated in an earlier study by Dobson and co-workers to increase solubility [152].

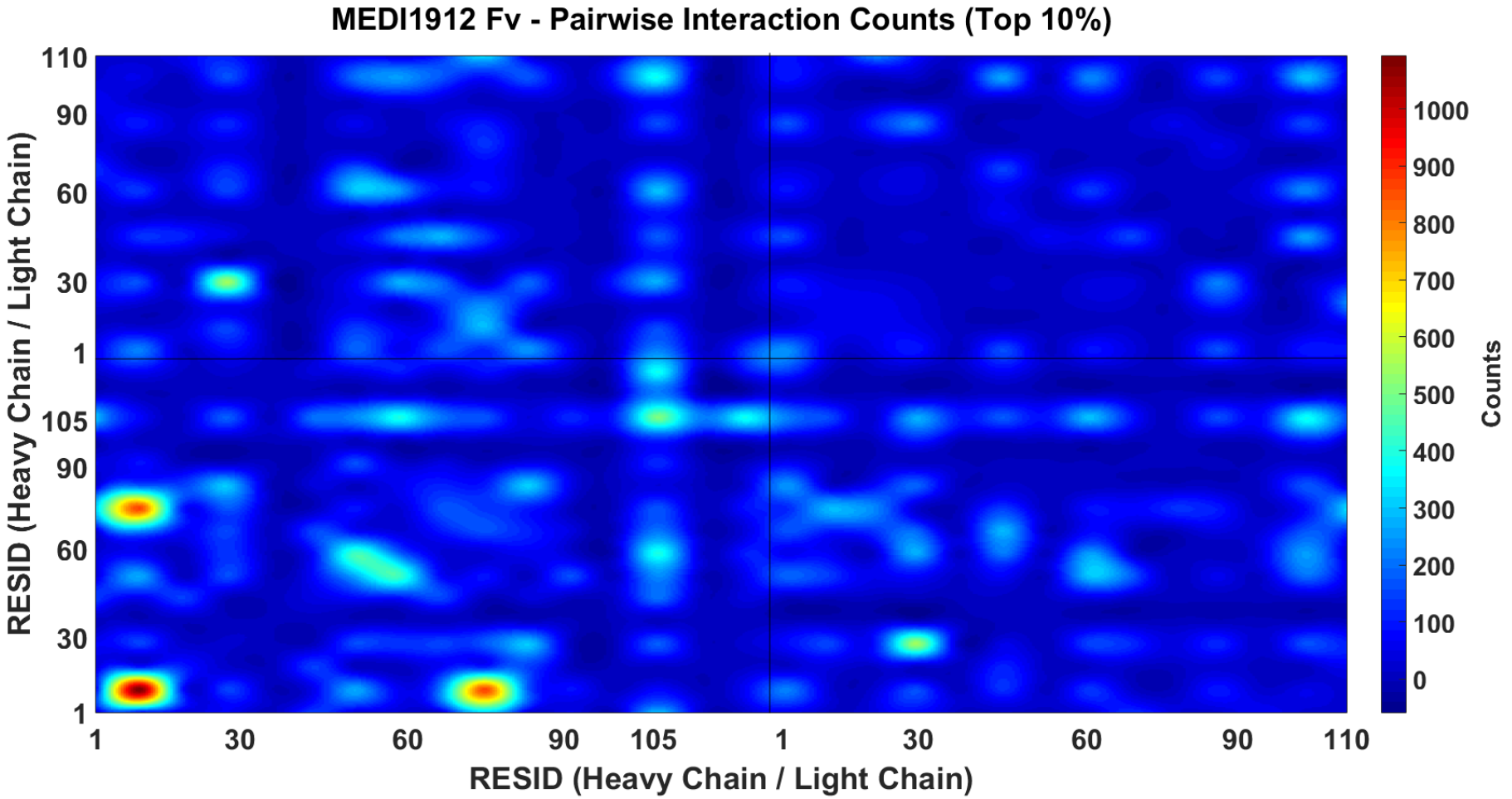


Figure 50: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI1912 Fv fragments from a homology model in 100 mM NaCl. Stronger signals have been seen in regions VAL11, TRP30, PHE31, LEU56, VAL78 and LEU105 on heavy chain and ASN31 on the light chains in the top 10% simulations in terms of most negative interaction energies.

6.3.3.2. MEDI1912 Dipeptide Screening Results

In the studies with MEDI-578, more than half of the dipeptides screened with this method have resulted in a less negative interaction energies suggesting dipeptides carry the potentials to reduce antibody-antibody interactions and hence reducing aggregation in MEDI1912. The same eight hydrophilic amino acid residues were then considered, ALA, ARG, GLN, GLY, HIS (HSP), LYS, PRO and SER to generate a total of 81 set of simulations. The overall mean interaction energies pattern among these systems are similar to the screening with MEDI-578 (Figure 51). Sixty-one out of eighty-one sets of dipeptides displayed more positive calculated mean interaction energies. Among these sixty-one dipeptides, charged residues ARG, HSP and LYS-linked dipeptides always resulted a less negative calculated mean interaction energies suggesting addition of these dipeptides can reduce tight MEDI1912 Fv dimer formations.

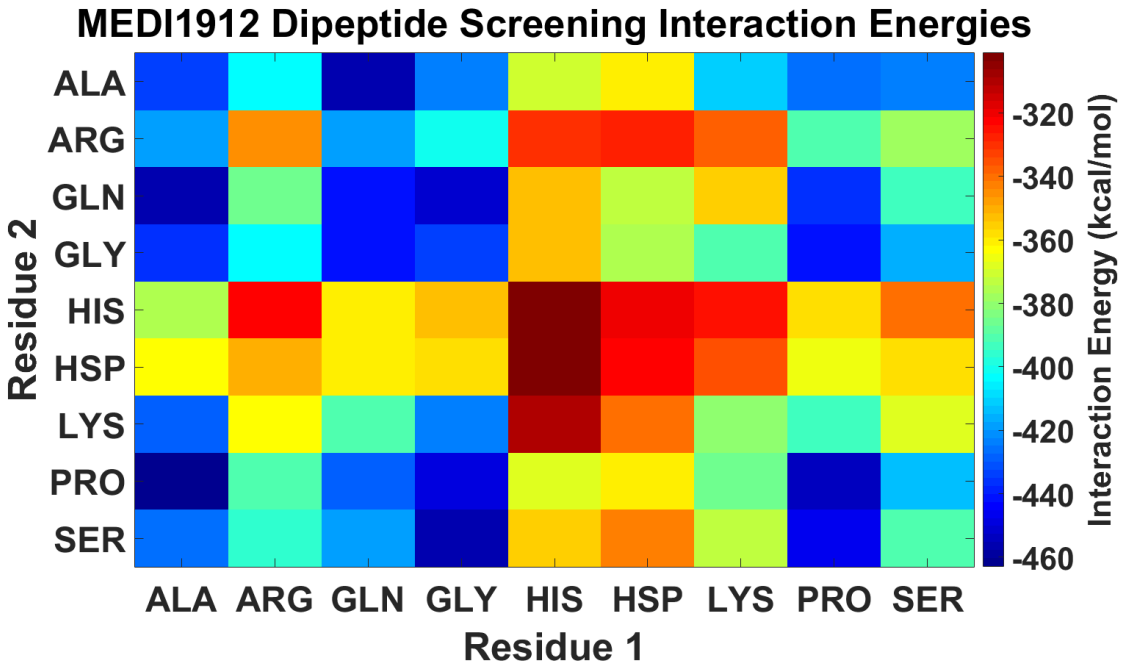


Figure 51: Visual representations of the plateau values calculated from mean interaction energy fitted with the model function in systems with 10 molecules of dipeptides simulated for 256 simulations of 512 ns with 2 MEDI1912 Fv fragments from a homology model. The overall pattern of plateau values calculated from mean interaction energies were similar to the screening with MEDI-578 Fv fragment.

Interaction Energies (kJ/mol)										
Residue 2	ALA	-432.9	-403.0	-455.6	-423.5	-369.9	-360.6	-410.0	-425.4	-422.5
	ARG	-419.3	-345.3	-419.7	-402.1	-329.5	-326.6	-338.8	-391.4	-379.1
	GLN	-456.3	-386.8	-442.2	-451.8	-352.1	-371.9	-354.2	-436.6	-392.3
	GLY	-437.4	-404.2	-441.0	-433.2	-353.6	-374.9	-391.2	-441.9	-416.5
	HIS	-375.9	-321.8	-361.4	-353.5	-301.1	-318.9	-325.6	-358.3	-339.9
	HSP	-362.8	-351.6	-359.9	-357.9	-302.1	-321.9	-335.6	-364.8	-357.4
	LYS	-428.6	-362.8	-391.5	-422.7	-310.3	-340.0	-381.2	-393.2	-367.6
	PRO	-462.8	-389.7	-427.9	-448.3	-368.7	-361.0	-385.7	-453.6	-412.6
	SER	-427.0	-396.6	-417.8	-455.9	-355.5	-342.2	-372.2	-445.7	-391.1
		ALA	ARG	GLN	GLY	HIS	HSP	LYS	PRO	SER
	Residue 1									

Table 13: The plateau value calculated from mean interaction energy fitted with the model function in systems with 10 molecules of dipeptides simulated for 256 simulations of 512 ns with 2 MEDI1912 Fv fragments. Similar to the results from the screening with MEDI-578 Fv fragments, addition of dipeptides containing the neutral histidine, HIS, or charged residues such as ARG, HSP and LYS resulted in dimer complexes with less negative mean interaction energies with HIS–HIS achieved the least less negative mean interaction energies among all screening dipeptides.

Systems with 10 molecules of PRO-ALA showed significant interaction events through the regions VAL11, TRP30, LEU56 and VAL78 on the heavy chain and ALA13 on the light chain (Figure 52). These regions on the heavy chain were found to be potentially responsible for dimer formation in MEDI1912 by Dobson and co-workers [152]. TRP30, PHE31 and LEU56 on heavy chain were mutated to MEDI1912_STT to improve solubility, reduce viscosity and reduce aggregation. The addition of PRO-ALA will not prevent MEDI1912 self-interactions in these two regions and PRO-ALA is unlikely to solve the formulation challenges of MEDI1912.

HIS-HIS showed a reduction of interactions through the VAL11 and VAL78 regions on the heavy chain (Figure 53). Charged dihistidine (HSP-HSP) have displayed a greater

prevention of MEDI1912 self-interaction through VAL11 on the heavy chain, possibly through binding to the adjacent negatively charged GLU (Figure 54). Systems with HSP-HSP shifted the interactions towards VAL78 and LEU105 regions on the heavy chain. Neutral dihistidine (HIS-HIS) showed a significantly weaker signal on the VAL78 and LEU105 regions on the heavy chain and interactions of MEDI1912 Fv fragment favour VAL11 on the heavy chain. In both cases, the signal from TRP30, PHE31 and LEU56 on the heavy chain are much weaker compared to PRO-ALA. Shifting from more to less energetically favourable interaction hot-spots indicates HIS-HIS prevents formation of energetically favourable MEDI1912 Fv fragment dimers. Hence, suggesting HIS-HIS may offer the opportunity to tackle the formulation challenges in MEDI1912 although further studies required to work out the ideal antibody:excipient ratios and pH for ratios for the ionisation state for imidazole groups within HIS-HIS.

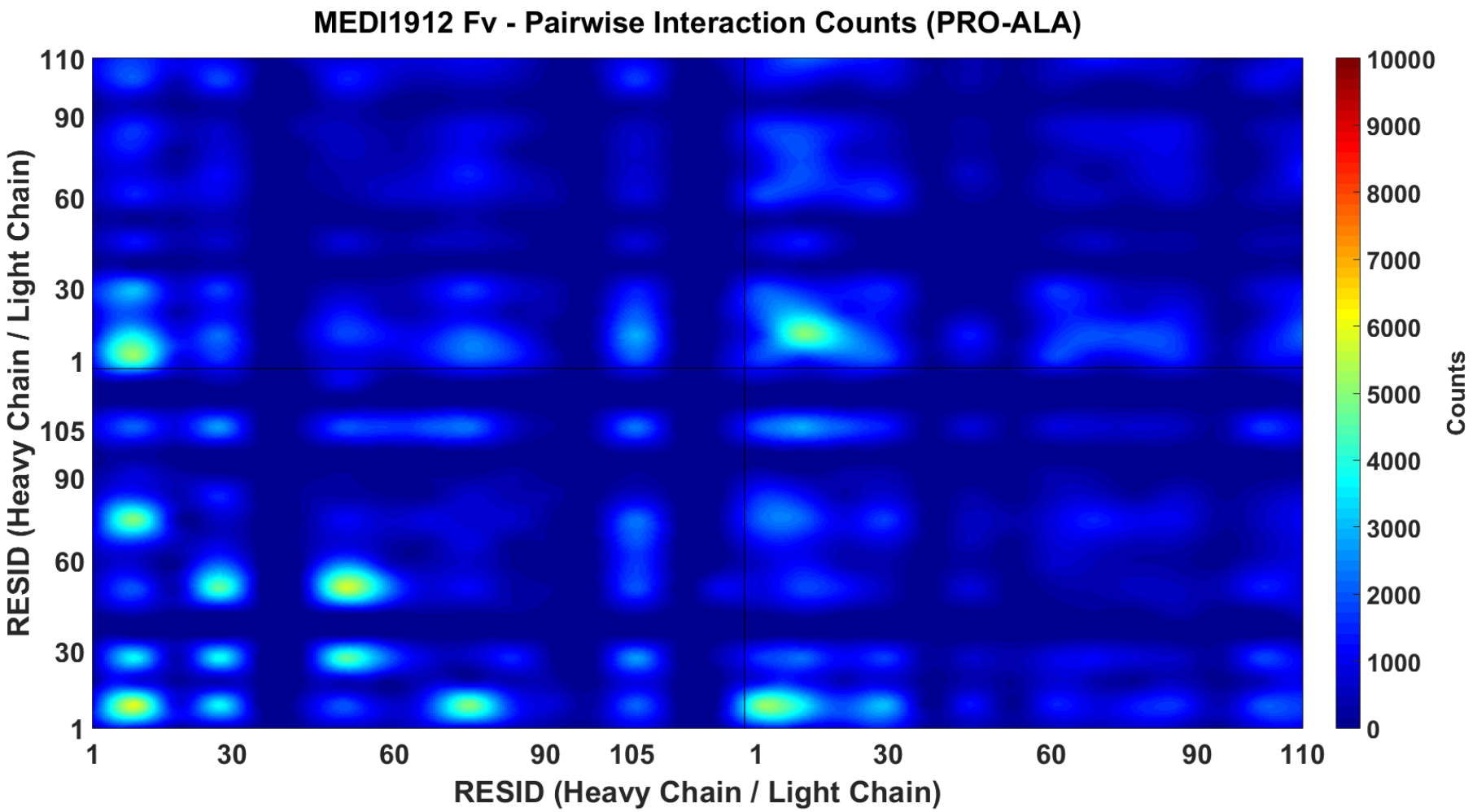


Figure 52: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI1912 Fv fragments from a homology model in 100 mM NaCl with 10 molecules of PRO-ALA. Significant interaction events were observed through the regions TRP30, PHE31 and LEU56 which were mutated to increase solubility in MEDI1912_STT.

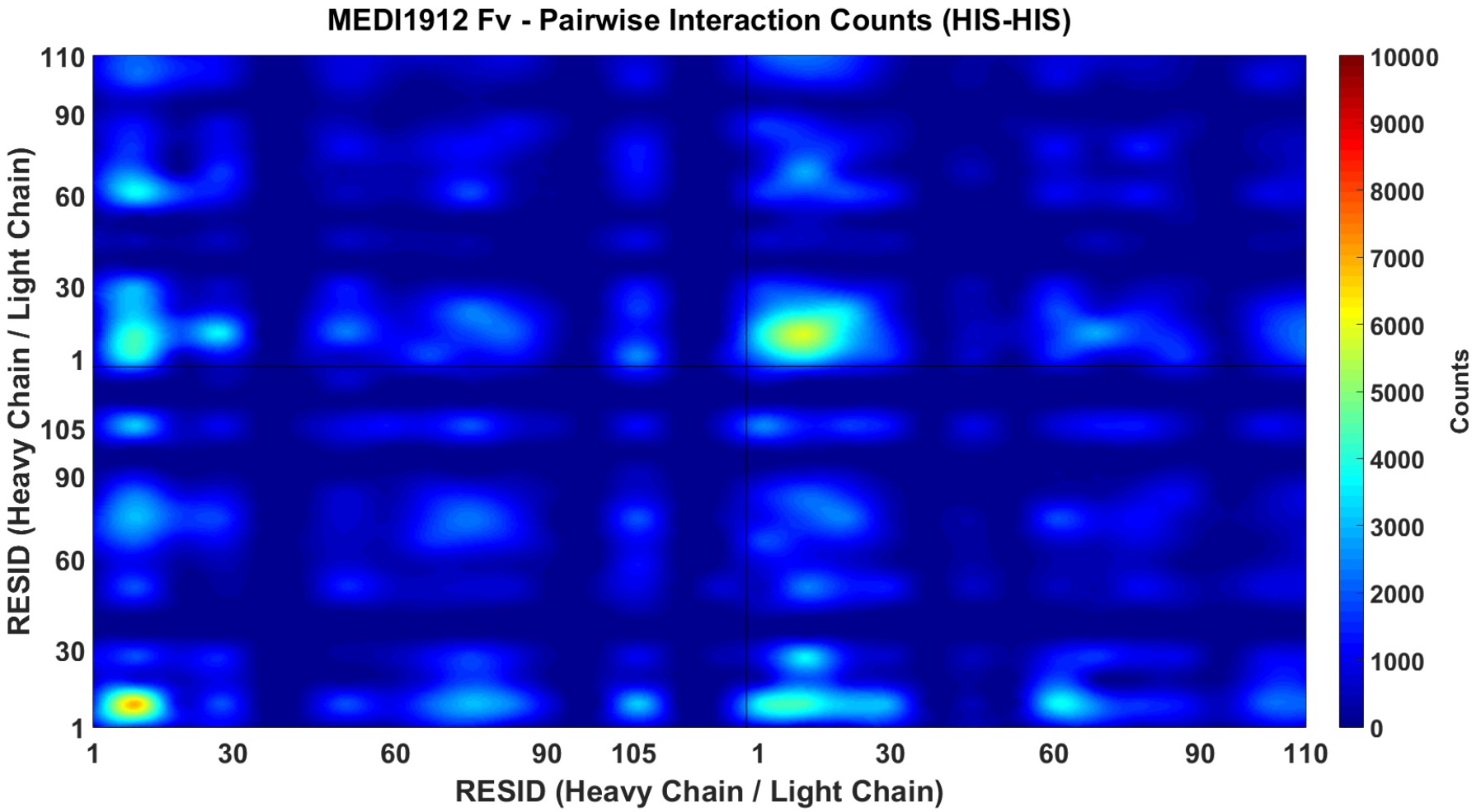


Figure 53: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI1912 Fv fragments from a homology model in 100 mM NaCl with 10 molecules of HIS-HIS. The occurrence of interaction events through the VAL78 and LEU105 regions on the heavy chain were reduced compared with PRO-ALA and shifted towards VAL11 on the heavy chain.

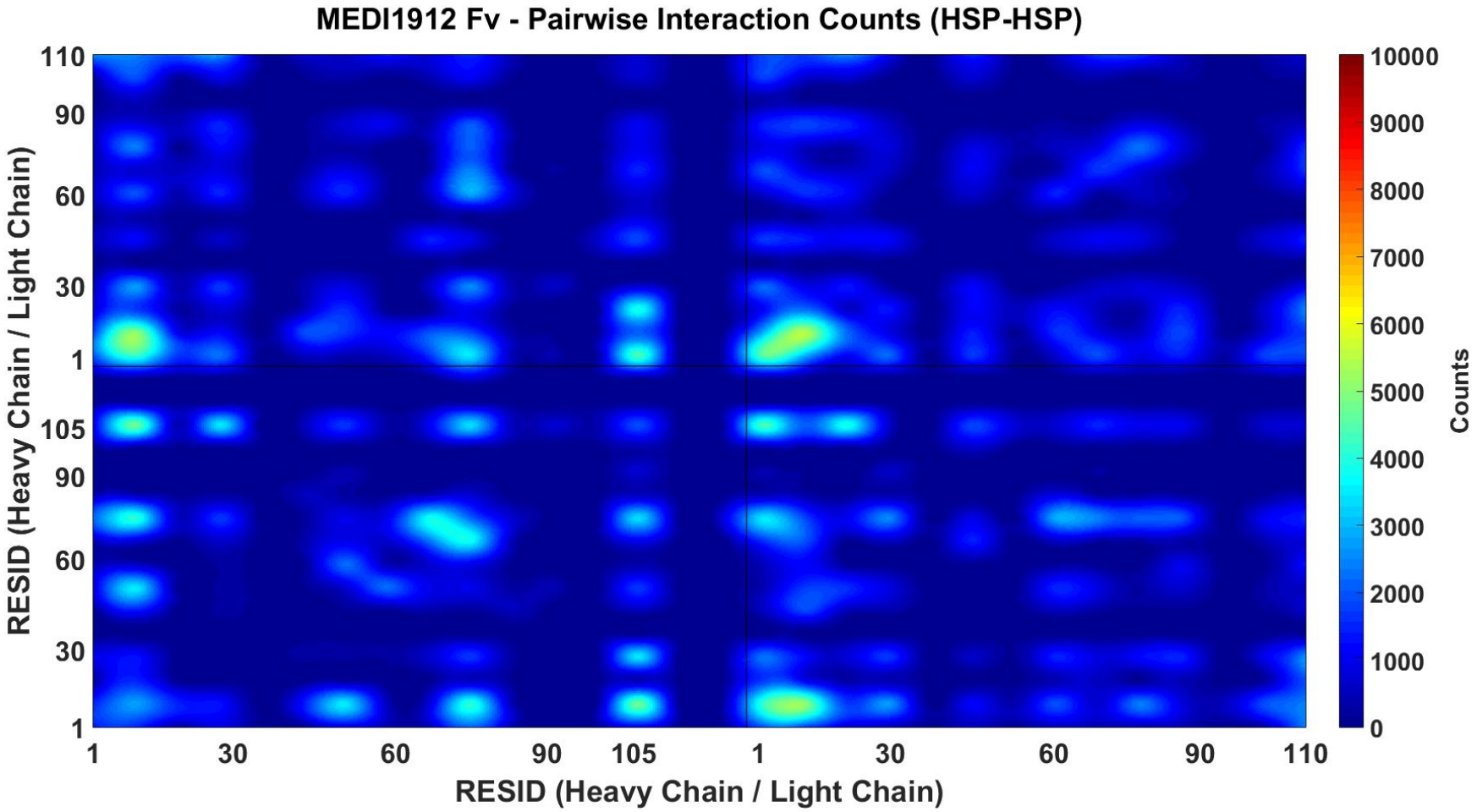


Figure 54: The most common binding pairs in a set of 256 MARTINI CG simulations with two MEDI1912 Fv fragments from a homology model in 100 mM NaCl with 10 molecules of HSP-HSP. The occurrence of interaction events through the VAL11 on the heavy chain were lower compared to HIS-HIS, possibly through binding to the adjacent negatively charged GLU.

6.3.4. Motavizumab Dipeptide Library Screening

6.3.4.1. Motavizumab Dimerisation

To identify the self-interaction patterns of motavizumab Fab, a set of simulations with just 2 motavizumab Fab molecules were prepared. Each simulation contained 2 motavizumab Fab molecules, 12 CL- ion molecules and roughly 18000 water beads, approximately in total of 20000 MARTINI beads which is roughly two times bigger than systems with 2 Fv fragments. The calculated plateau value from the mean interaction energy distributions for motavizumab dimerisation was -268.9 kJ/mol.

The larger systems required doubled computational time on the same computer system to simulate 512 ns of a system containing 2 motavizumab Fab molecules compared to a system containing 2 MEDI-578 Fv fragments. Only approximately 30% of simulations formed motavizumab Fab dimer complexes. During 512 ns of simulation time, the mean interaction energies did not level off towards the end of simulation and mean interaction energies also did not overlap with the central vigintile, suggesting the convergence did not complete (Figure 55).

Although convergence did not fully complete within 512 ns, the motavizumab Fab dimer complexes formed in the top vigintile (top 5%). Most negative interaction energies were still important as these dimer complexes may provide valuable insights on the locations of interaction hot-spots so these can be identified to give stable dimer complexes (Figure 56). These interaction hot-spots were located at the regions of the CDR2 loop (i.e. ASP54), the loop at the side of the Fab connecting the β-sheets that forms the CDR2 and CDR3 (i.e. TRP84) and the last β-sheet on the CH1 domain (i.e. LYS206) on the heavy chain. The only interaction hot-spot seen on the light chain is located at the last loop connecting two β-sheets (i.e. LYS190) on the CH1 domain on the light chain.

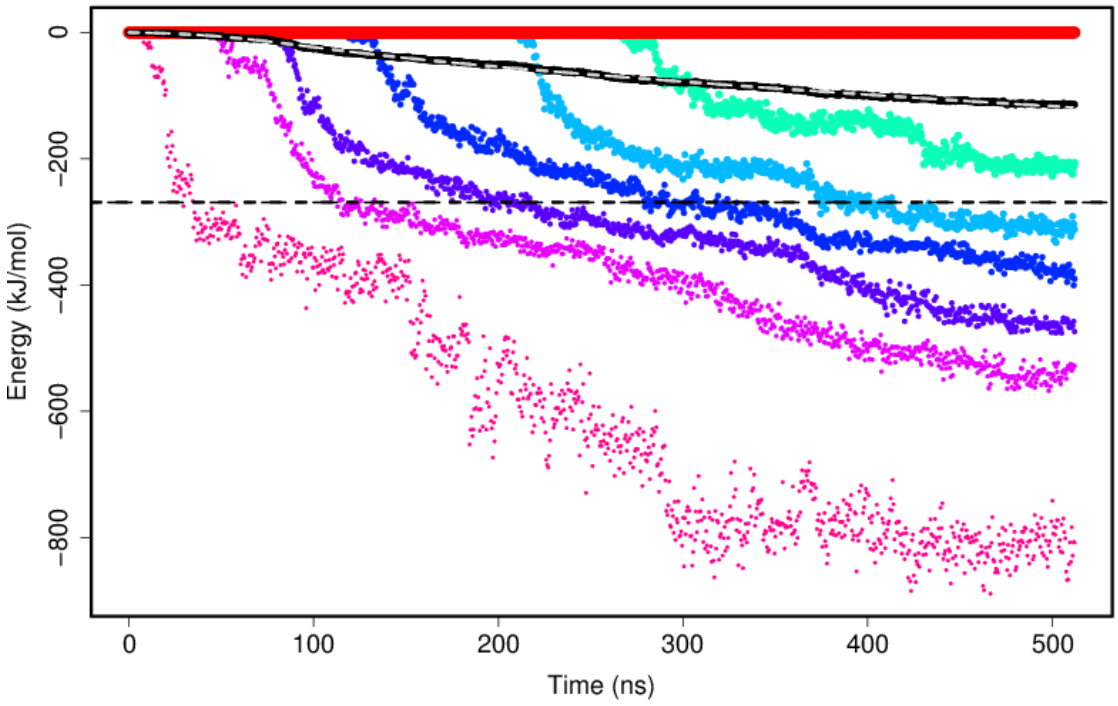


Figure 55: The energy distributions between the motavizumab Fab dimers were plotted vs time. The mean value of all energies is indicated as black line. Only around 30% of 256 simulations formed motavizumab Fab dimer complexes. During 512 ns of simulation time, the mean interaction energies did not level off towards the end of simulation and mean interaction energies also did not overlap with the central vigintile (red colour), suggesting the convergence did not complete.

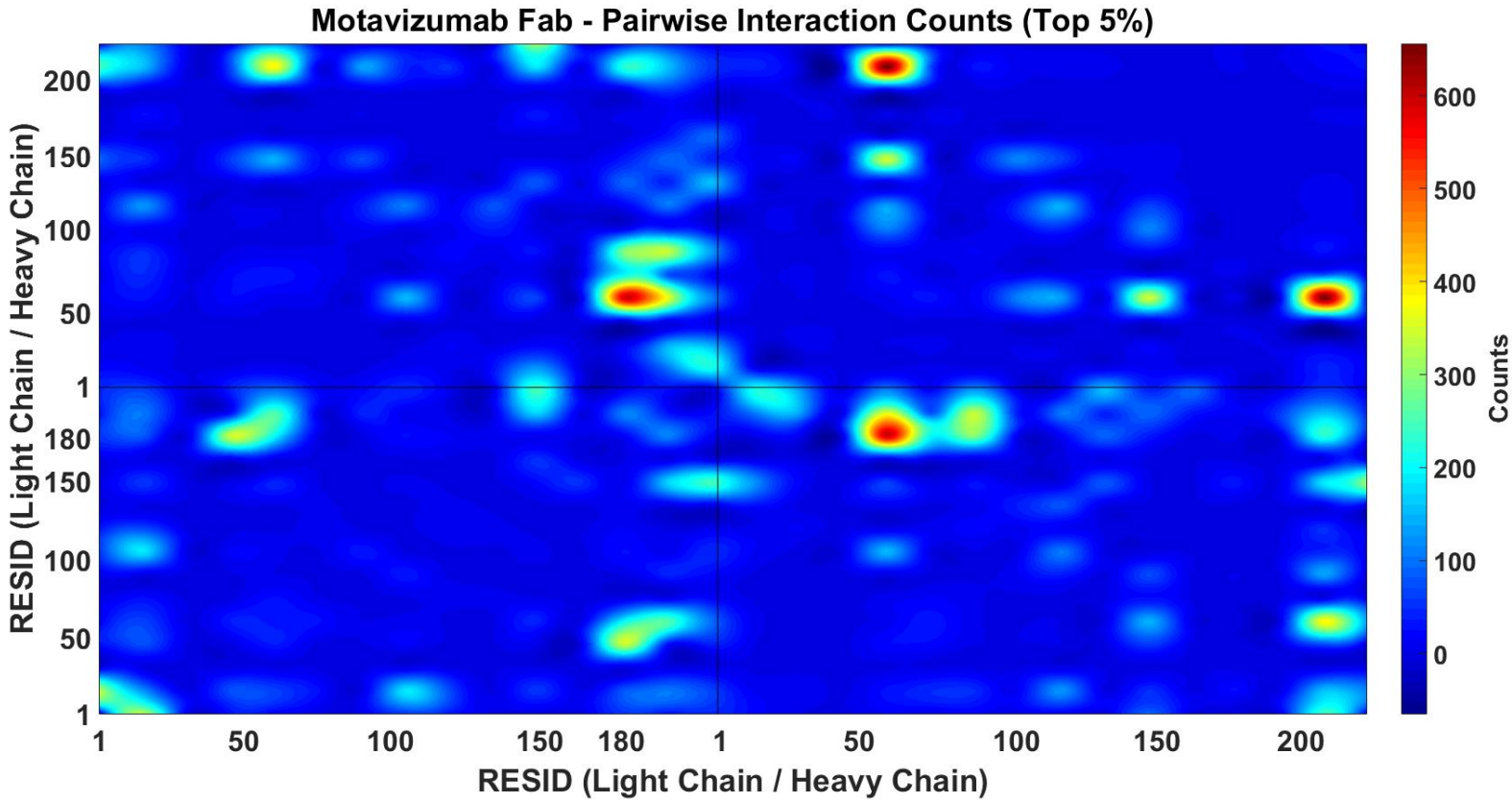


Figure 56: The most common binding pairs in a set of 256 MARTINI CG simulations with two Motavizumab Fab fragments. The area coloured in red denotes the binding pairs with most occurrence. These two main interaction hot-spots were located at the regions close to ASP54 on the heavy chain and LYS206 on the light chain.

6.3.4.2. Motavizumab Dipeptide Screening Results

Motavizumab is formulated in 25 mM Histidine at pH 6.0. To account for the effect of the salts, protonated histidine residues (HSP) has been were added to the simulation box. Similar to the motavizumab Fab dimerisation simulations, only around 30% of all simulations formed dimer complexes during each 512 ns simulation. Despite incomplete convergence within this simulation time, differences in interaction energies between Fab fragments can still be observed. Charged amino acid residues ASP, HIS or LYS-linked dipeptides can lead to achieve a less negative mean interaction energy with HIS-HSP achieved the least negative mean interaction energies among all screened dipeptides (Table 14).

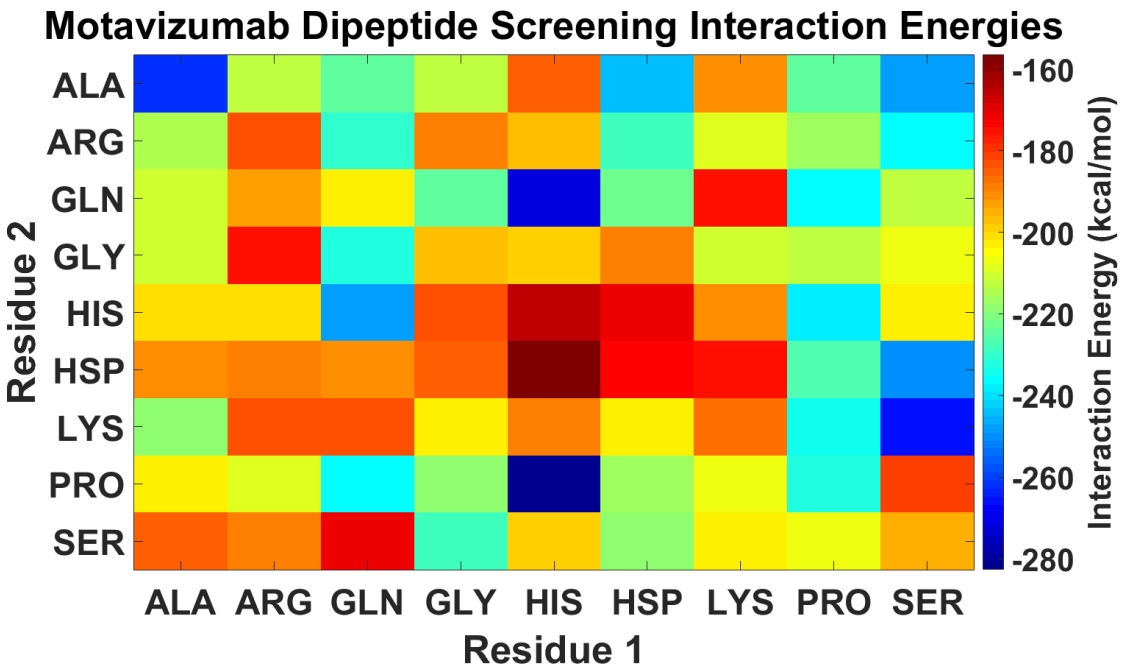


Figure 57: Visual representations of the plateau value calculated from mean interaction energy fitted with the model function in systems with 10 molecules of dipeptides simulated for 256 simulations of 512 ns with 2 Motavizumab Fab fragments. Due to insufficient convergence and high standard errors of the means associated with 512 ns of simulation time, there was a lack of consistent trend as seen in screening with MEDI-578 and MEDI1912.

Interaction Energies (kJ/mol)										
Residue 2	ALA	-262.8	-213.5	-224.8	-212.1	-184.9	-245.0	-191.4	-224.0	-247.4
	ARG	-215.1	-182.6	-231.3	-188.1	-197.0	-227.6	-208.2	-216.8	-236.9
	GLN	-209.8	-193.5	-202.1	-223.8	-272.3	-222.0	-175.8	-236.5	-212.6
	GLY	-210.1	-174.8	-231.9	-196.5	-199.2	-189.1	-211.4	-213.4	-207.7
	HIS	-201.6	-199.9	-248.3	-183.2	-165.8	-171.1	-190.7	-238.2	-202.6
	HSP	-191.9	-189.7	-190.9	-185.2	-156.6	-174.2	-174.9	-226.3	-251.0
	LYS	-219.3	-183.8	-183.5	-202.2	-189.4	-201.9	-187.7	-233.6	-265.3
	PRO	-202.4	-209.2	-236.1	-218.1	-282.5	-216.4	-207.0	-232.4	-180.7
	SER	-185.3	-190.0	-170.7	-228.8	-198.2	-219.2	-203.5	-207.4	-194.3
		ALA	ARG	GLN	GLY	HIS	HSP	LYS	PRO	SER
	Residue 1									

Table 14: The plateau value calculated from mean interaction energy fitted with the model function in systems with 10 molecules of dipeptides simulated for 256 simulations of 512 ns with 2 Motavizumab Fab molecules. HIS-HSP achieved the least less negative mean interaction energies among all screening dipeptides.

6.3.4.3 Diffusion Interaction Parameter

Self-interactions of motavizumab were measured with DLS and k_D values were calculated in each of the motavizumab formulations with dipeptides ALA-PRO, GLY-GLY, GLY-GLN, HIS-HIS, HIS-LYS, HIS-SER and ARG-ARG at 1:5 molar ratio. The formulation of motavizumab in 25 mM histidine buffer at pH 6.0 has a positive k_D of 29.65 mL/g. The positive k_D suggests net repulsive interactions between motavizumab molecules in the formulation.

Even with the net repulsive nature of between motavizumab molecules, the addition of dihistidine resulted in a largest increase of the k_D values among the six dipeptides tested *in-vitro* suggesting the addition of dihistidine can increase the net repulsive interactions of motavizumab. At pH 6, half of the imidazole groups are protonated and the other half

are neutral so it is technically possible to use dihistidine as a buffer that prevents antibody self-interactions.

Similar to the studies with MEDI-578, a positive correlation is also seen between the calculated mean interaction energies and the measured k_D. The regression model based on the calculated mean interaction energies of systems with protonated and neutral histidine-linked dipeptides achieved predicted R^2 values of 0.9311 and 0.8202 respectively (Figure 59). These predicted R^2 values are lower than those reported in the studies with MEDI-578, possibly due to the large errors associated with the large SEMs due to incomplete convergences. Nevertheless, this study reconfirms this approach can offer reasonable predictive power of predicting excipient's effects on antibody-antibody interactions.

Dipeptide	**Calculated mean interaction energies (kJ/mol)**	**Measured k_D (mL/g)**
ALA-PRO	-202.4	28.86
GLY-GLY	-196.5	30.49
GLY-GLN	-223.8	23.61
HIS-HIS (HSP-HSP)	-165.8 (-174.2)	34.18
HIS-LYS (HSP-LYS)	-189.4 (-201.9)	29.80
HIS-SER (HSP-SER)	-196.5 (-219.2)	28.29
ARG-ARG	-182.6	30.77

Table 15: The calculated plateau value from the mean interaction energy distributions in systems with different dipeptides and the measured k_D values in different formulations of Motavizumab containing these dipeptides.

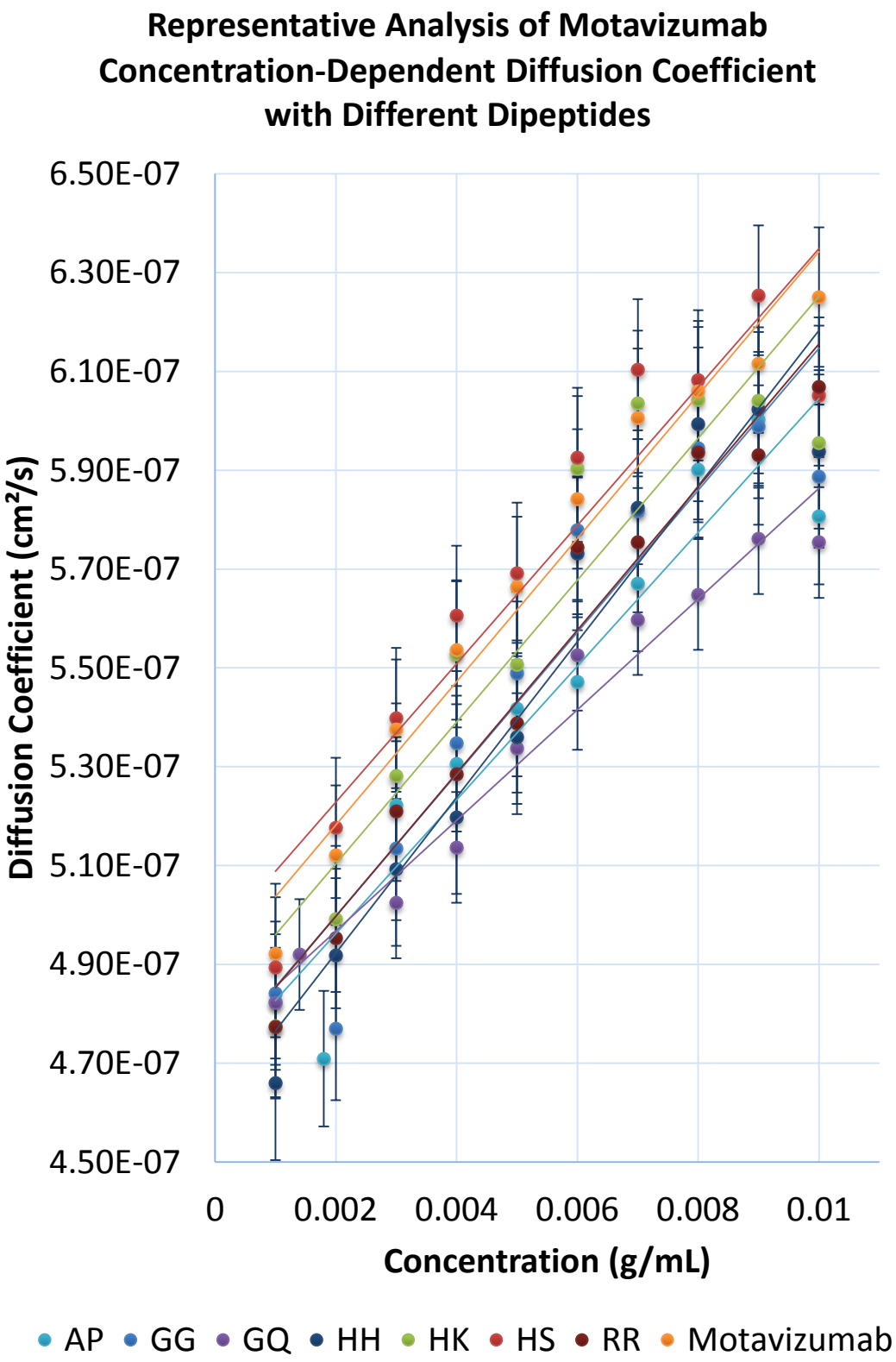


Figure 58: Change in diffusion coefficient versus concentration of Motavizumab in 25mM Histidine buffer at pH 6 and the following dipeptides were added at 1:5 molar ratio a) ALA-PRO, b) GLY-GLY, c) GLY-GLN, d) HIS-HIS, e) HIS-LYS, f) HIS-SER and g) ARG-ARG. The diffusion coefficients increase as the concentration increase in all formulations of Motavizumab suggesting repulsive interactions between Motavizumab molecules in all formulations.

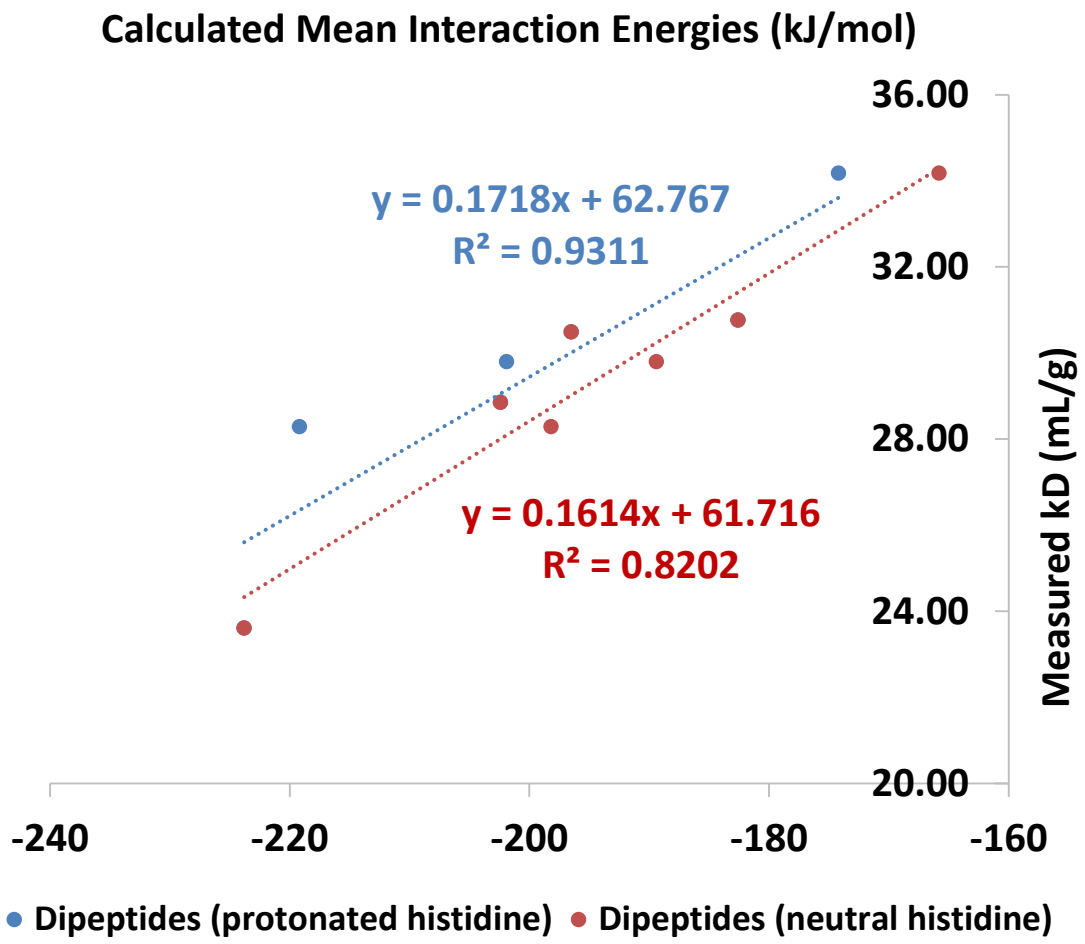


Figure 59: Correlations between the calculated mean interaction energies from CG simulations and measured k_D with different Motavizumab formulations. A positive correlation was observed between the calculated mean interaction energies and the diffusion interaction parameters of different Motavizumab formulations. The regression model based on the calculated mean interaction energies of systems based on protonated histidine-linked dipeptides achieved an adjusted R^2 value of 0.8202 compared to 0.9311 in model based on the values from neutral histidine-linked dipeptides.

6.4. Conclusions

In summary, MARTINI CG simulations demonstrated in this study were able to reasonably predict the effect of selected dipeptide excipients on self-interactions of the two antibody used in this study. This study focussed on how selected dipeptides affect the self-interactions of antibody molecules. The synergistic effects of the neutral and protonated imidazole groups in dihistidine can prevent antibody self-interactions at different parts of the antibody within the Fv fragment. Further work would require to work out the optimum ratios of different dihistidine species through pH adjustments. This method can be also be extended to investigate mixed excipient systems. This approach also offers great potential of reducing the time and resource on formulation development for challenging antibodies through homology modelling which materials are often limited in early stages of formulation development.

The variability in sequence is limited to the Fv fragment while the CH1 and VH1 domains are constant. Significant computational resources are required to conduct a screening study using the structure of a Fab. Results presented in this chapter suggested screening study using the structure of Fv fragment may be sufficient to conduct a virtual screening for potential excipients and achieved satisfactory agreement with experimental data.

Bereau and Kremer have developed an automated algorithm for parametrisation of small organic compounds in MARTINI force field [198] which opens up the potential to explore the potentials of innovative excipients by applying this approach to study their effects on protein-protein interactions and to narrow down the number of potential excipients for *in-vitro* testing.

7. Final Discussion and Future Perspectives

Liquid antibody formulations are usually more favourable than freeze-dried formulations as there is no requirement for reconstitution but antibodies are susceptible to aggregation, which is one of the most common physical degradation pathways that occurs during processing and storage. Antibody aggregation involving the irreversible self-association of antibody molecules depends on their three-dimensional structure, sequence and solvent-accessible hydrophobic and charged residues. Aggregation is also influenced by a number of environmental factors such as concentration of antibody, viscosity and formulation constitutions. An important class of remedies is to identify suitable excipients that can improve solubility, reduce viscosity and delay aggregation.

A combination of *in-silico* technique have been used to characterise antibody-antibody interactions of antibody. Hydrophobic CDRs are possible aggregation-prone regions for antibody. Mutation of the CDRs will affect binding of antibody to its target so stability can potentially be more readily improved via better design of the antibody formulation.

The first study described in Chapter 2 identified aggregation-prone regions on MEDI-578 Fv fragment through SAP calculations to identify solvent-accessible hydrophobic patches and by protein-protein docking to identify likely dimer conformations. The variability in sequence is limited to the Fv fragment of an antibody so the distributions of aggregation-prone regions on the Fv fragment would be different for different mAbs. Combining the results from these approaches allowed us to identify the aggregation-prone regions with greater confidence to generate a plausible design of a specific excipient. *In-silico* screening of a diverse library of compounds using public databases by applying various filters was used to obtain excipient-like compounds for the Fv fragment for consideration.

DLS and SEC studies indicated the best docked compound (CPX) to the most hydrophobic region of MEDI-578 did not destabilise MEDI578. It was believed that CPX would be a good candidate excipient to but it failed experimentally to show significant difference to trehalose, a traditional excipient.

Proteins are dynamic in nature so the binding of two antibody molecules results in complex intramolecular and intermolecular interactions which can lead to changes in conformation and exposure of aggregation-prone regions. In this context, all-atom MD simulations can simulate antibody-antibody interactions explicitly by including the water and ion molecules. All-atom MD simulations can probe these protein-protein interactions with atomistic details, revealing non-covalent interactions between different molecules and predict energy profiles which can be used to compare antibodies in different environments. However, the computational cost is a major burden for all-atom simulations. Coarse-grained MD simulations can largely reduce the computational cost but a single long simulation will generally only be able to form stable dimer complexes which does not dissociate due to low dissociation rates. To overcome this challenge, it has been shown the energy landscape of dimers formations can be simulated with multiple simulations with unbiased starting orientations for shorter timescale using of DAFT [176] and MARTINI protein force field [169].

Further examination of CPX with coarse-grained MD simulations revealed that CPX has the propensity to prevent antibody-antibody interactions at the most hydrophobic site as identified with SAP. As there are multiple aggregation-prone regions on an antibody, coarse-grained MD simulations also identified other sites with greater interaction energies at different locations to the most solvent-accessible hydrophobic patch that was initially used to identify CPX. Antibody-antibody interactions are not limited to the most solvent-

accessible hydrophobic patch and therefore it is necessary to take into account of the dynamic nature of mAb, effects of the charges and realistic formulation environments for the energy landscape of antibody dimer complexes.

Chapter 5 was unable to find any differences in solvation free energies before and after the formation of dimer complex. In addition, no correlations between the calculated interaction energy values from the MARTINI model and the binding energy calculated with MM-PBSA were found. It is important for computational modelling to be useful in a practical way in experimental research to provide a fast, efficient and accurate way to model the dynamics of proteins. Evaluating transient mAb dimer complexes with Lennard-Jones and Coulomb potentials remain as viable indicators for comparing mAb-mAb interactions between dimer complexes formed as this method uses minimal computational resources with no further post-simulation processing being required.

Coarse-grained MD simulations provided better coverage of the energy landscape for antibody-antibody interactions. The potential to virtually screen for novel excipients inhibit or interfere with antibody-antibody interactions that could result in antibody aggregation is possible using MD simulations. The DLS results described suggested there was satisfactory agreement with the *in-silico* data. Dihistidine was found to be the best dipeptide among the dipeptide library screening. This dipeptide has different propensities towards different interaction sites on MEDI-578 structure. Given the similarity in the structure of antibodies based molecules that were studied, formulations consisting of a full-length mAb can potentially be designed using moderate computational resource (e.g Fv fragment). The study described in Chapter 6 has already demonstrated that the energy landscape for Fabs are large and complex where the free-energy is harder to be determined by the same MD approach. The experience gained in this study will be of

assistance in discovering innovative excipients other than short peptides as an automated algorithm for parametrisation of small organic compounds in MARTINI force field is available [198], allowing exploration of compounds without MARTINI parameters.

The approach taken in this thesis is to combine and improve various computational and experimental methods in order to develop a rational method of developing excipients to inhibit specific degradation pathways for complex biological medicines. The results developed here represent a promising approach for designing tailored excipients to interact preferentially with precisely those local regions on the macromolecule that are involved in aggregation. This is a computationally demanding process, but the coarse-grained methods used here along with suitable experimental testing of the results provides a method for designing simple tailored excipients, such as dipeptides. Such simple excipients can be made to the very high purity levels needed for drug product additives in a comparatively inexpensive process. As the complexity of such tailored impurities increases, the costs involved in manufacturing them to 99.9% purity will increase dramatically. It is therefore important that a simple dipeptide was able to inhibit aggregation for an industrially relevant antibody. Such approaches are likely to increase in importance for the next generation of biological medicines, which are likely to include more complex, multi-domain macromolecules and complexes of such macromolecules. Currently popular products of increased complexity include ADCs, bispecific antibodies and peptibodies.

8. References

[1] Mullard A. 2017 FDA drug approvals. Nat Rev Drug Discov. 2018; 17:81-85.

[2] Grillo-López AJ. The first antibody therapy for cancer: A personal experience. Expert Rev Anticancer Ther. 2013; 13:399-406.

[3] Behring E, Kitasato S. Ueber das zustandekommen der diphtherie-immunität und der tetanus-immunität bei thieren dtsch. Med Wschr. 1890; 16:1113–1114.

[4] Tiselius A, Kabat EA. An electrophoretic study of immune sera and purified antibody preparations. J Exp Med. 1939; 69:119-131.

[5] Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci USA. 1973; 70:3240-3244.

[6] Kohler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. Nature. 1975; 256:495–497.

[7] Smith SL. Ten years of orthoclone OKT3 (muromonab-CD3): A review. J Transpl Coord. 1996; 6:109–119.

[8] Clark M. Antibody humanization: a case of the 'Emperor's new clothes'? Immunol. Today. 2000; 21:397-402.

[9] Rodgersa KR, Chou RC. Therapeutic monoclonal antibodies and derivatives: Historical perspectives and future directions. Biotech Adv. 2016; 34:1149-1158.

[10] Narciso JE, Uy ID, Cabang AB, Chavez JF, Pablo JL, Padilla-Concepcion GP, Padlan EA. Analysis of the antibody structure based on high-resolution crystallographic studies. N Biotechnol. 2011; 28:435-447.

[11] Hay SR. Immunological mechanisms of allergic eye disease. Clin Eye & Vison Care. 1999; 11: 21-24.

[12] Wang W, Singh S, Zeng DL, King K, Nema S. Antibody structure, instability, and formulation. J Pharm Sci. 2007; 96:1-26.

[13] Keizer RJ, Huitema AD, Schellens JH, Beijnen JH. Clinical pharmacokinetics of therapeutic monoclonal antibodies. Clin Pharmacokinet. 2010; 49:493-507.

[14] Leibiger H, Wüstner D, Stigler RD, Marx U. Variable domain-linked oligosaccharides of a human monoclonal IgG: Structure and influence on antigen binding. Biochem J. 1999; 338:529-538.

[15] Shinkawa T, Nakamura K, Yamane N, Shoji-Hosaka E, Kanda Y, Sakurada M, Uchida K, Anazawa H, Satoh M, Yamasaki M, Hanai N, Shitara K. The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. J Biol Chem. 2003; 278:3466-2473.

[16] Jung ST, Kang TH, Kelton W, Georgiou G. Bypassing glycosylation: engineering aglycosylated full-length IgG antibodies for human therapy. Curr Opin Biotechnol. 2011; 22:858-867.

[17] Kayser V, Chennamsetty N, Voynov V, Forrer K, Helk B, Trout BL. Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies. Biotechnol J. 201; 6:38-44.

[18] Li CH, Narhi LO, Wen J, Dimitrova M, Wen ZQ, Li J, Pollastrini J, Nguyen X, Tsuruda T, Jiang Y. The effect of pH, temperature and salt on the stability of E. coli and CHO derived IgG1 Fc. Biochemistry. 2012; 51:10056-10065.

[19] Nimmerjahn F, Gordan S, Lux A. FcγR dependent mechanisms of cytotoxic, agonistic, and neutralizing antibody activities. Trends Immunol. 2015; 36:325-336.

[20] Razinkov VI, Treuheit MJ, Becker GW. Accelerated formulation development of monoclonal antibodies (mAbs) and mAb-based modalities. J Biomol Screen. 2015; 20:468-483.

[21] Mordenti J, Cuthbertson RA, Ferrara N, Thomsen K, Berleau L, Licko V, Allen PC, Valverde CR, Meng YG, Fei DT, Fourre KM, Ryan AM. Comparisons of the intraocular

tissue distribution, pharmacokinetics, and safety of 125I-labeled full-length and fab antibodies in rhesus monkeys following intravitreal administration. Toxicol Pathol. 1999; 27:536-544.

[22] Coleman L, Mahler SM. Purification of Fab fragments from a monoclonal antibody papain digest by Gradiflow electrophoresis. Protein Expr Purif. 2003; 32:246-251.

[23] Ward ES, Ober RJ. Chapter 4 Multitasking by Exploitation of Intracellular Transport Functions: The Many Faces of FcRn. Adv Immunol. 2009; 103:77-115.

[24] Nisonoff A, Rivers MM. Recombination of a mixture of univalent antibody fragments of different specificity. Arch Biochem Biophys. 1961; 93:460-462.

[25] Spiess C, Zhai Q, Carter PJ. Alternative molecular formats and therapeutic applications for bispecific antibodies. Mol Immunol. 2015; 67:95-106.

[26] Jäger M, Schoberth A, Ruf P, Hess J, Lindhofer H. The trifunctional antibody ertumaxomab destroys tumor cells that express low levels of human epidermal growth factor receptor 2. Cancer Res. 2009; 69:4270-4276.

[27] Hurvitz SA, Dirix L, Kocsis J, Bianchi GV, Lu J, Vinholes J, Guardino E, Song C, Tong B, Ng V, Chu YW, Perez EA. Phase II randomized study of trastuzumab emtansine versus trastuzumab plus docetaxel in patients with human epidermal growth factor receptor 2–positive metastatic breast cancer. J Clin Oncol. 2013; 31:1157-1163.

[28] Shpilberg O, Jackisch C. Subcutaneous administration of rituximab (MabThera) and trastuzumab (Herceptin) using hyaluronidase. Br J Cancer. 2013; 109:1556-1561.

[29] Jackisch C, Müller V, Dall P, Neumeister R, Park-Simon TW, Ruf-Dördelmann A, Seiler S, Tesch H, Ataseven B. Subcutaneous Trastuzumab for HER2-positive breast cancer – evidence and practical experience in 7 german centers. Geburtshilfe Frauenheilkd. 2015; 75:566-573.

[30] Richter WF, Bhansali SG, Morris ME. Mechanistic determinants of biotherapeutics absorption following sc administration. AAPS J. 2012; 14:559-570.

[31] Supersaxo A, Hein WR, Steffen H. Effect of molecular weight on the lymphatic absorption of water-soluble compounds following subcutaneous administration. Pharm Res. 1990; 7:167-169.

[32] Harrington L. Administer single-site 30-ml intramuscular injection? Medsurg Nurs. 2005; 14:379-382.

[33] Muzammil S, Mabus JR, Cooper PR, Brezski RJ, Bement CB, Perkinson R, Huebert ND, Thompson S, Levine D, Kliwinski C, Bradley D, Hornby PJ. FcRn binding is not sufficient for achieving systemic therapeutic levels of immunoglobulin G after oral delivery of enteric-coated capsules in cynomolgus macaques. Pharmacol Res Perspect. 2016; 4:e00218.

[34] des Rieux A, Fievez V, Garinot M, Schneider YJ, Préat V. Nanoparticles as potential oral delivery systems of proteins and vaccines: a mechanistic approach. J Control Release. 2006; 116:1-27.

[35] Hussack G, Hirama T, Ding W, Mackenzie R, Tanha J. Engineered single domain antibodies with high protease resistance and thermal stability. PLoS One. 2011; 6:e28218.

[36] Shimizu M, Miwa Y, Hashimoto K, Goto A. Encapsulation of chicken egg yolk immunoglobulin G (IgY) by liposomes. Biosci Biotechnol Biochem. 1993; 57:1445-1449.

[37] Bitonti AJ, Dumont JA, Low SC, Peters RT, Kropp KE, Palombella VJ, Stattel JM, Lu Y, Tan CA, Song JJ, Garcia AM, Simister NE, Spiekermann GM, Lencer WI, Blumberg RS. Pulmonary delivery of an erythropoietin Fc fusion protein in non-human primates through an immunoglobulin transport pathway. Proc Natl Acad Sci USA. 2004; 101:9763-9768.

[38] Agu RU, Ugwoke MI, Armand M, Kinget R, Verbeke N. The lung as a route for systemic delivery of therapeutic proteins and peptides. Respir Res. 2001; 2:198-209.

[39] Goldberg T, Wong E. Afrezza (Insulin Human) inhalational powder. A new inhaled insulin for the management of type-1 or type-2 diabetes mellitus. PT. 2015; 40:735–741.

[40] Spadiut O, Capone S, Krainer F, Glieder A, Herwig C. Microbials for the production of monoclonal antibodies and antibody fragments. Trends Biotechnol. 2014; 32:54-60.

[41] Moussavou G, Ko K, Lee JH, Choo YK. Production of monoclonal antibodies in plants for cancer immunotherapy. Biomed Res Int. 2015; 2015:306164.

[42] Andersen DC, Reilly DE. Production technologies for monoclonal antibodies and their fragments. Curr Opin Biotechnol. 2004; 15:456-462.

[43] Li F, Vijayasankaran N, Shen AY, Kiss R, Amanullah A. Cell culture processes for monoclonal antibody production. MAbs. 2010; 2:466-479.

[44] Kim NS, Byun TH, Lee GM. Key determinants in the occurrence of clonal variation in Humanised antibody expression of CHO cells during dihydrofolate reductase mediated gene amplification. Biotechnol Prog. 2001; 17:69-75.

[45] Yang Z, Xiong H. Culture conditions and types of growth media for mammalian cells. Biomedical Tissue Culture. 2012; 3-18.

[46] Franek F, Eckschlager T, Katinger H. Enhancement of monoclonal antibody production by lysine-containing peptides. Biotechnol Prog. 2003; 19:169-174.

[47] Aldington S, Bonnerjea J. Scale-up of monoclonal antibody purification processes and it is one of the regulatory requirements. J Chromatogr. B. 2007; 848:64-78.

[48] Liu HF, Ma J, Winter C, Bayer R. Recovery and purification process development for monoclonal antibody production. MAbs. 2010; 2:480-499.

[49] Duhamel RC, Schur PH, Brendel K, Meezan E. pH gradient elution of human IgG1, IgG2 and IgG4 from protein A-sepharose. J Immunol Methods. 1979; 31:211-217.

[50] Huse K, Böhme H, Scholz GH. Purification of antibodies by affinity chromatography. J Biochem Biophys Methods. 2002; 51:217-231.

[51] Sommerfelda S, Strubeb J. Challenges in biotechnology production—generic processes and process optimization for monoclonal antibodies. Chem Eng Process. 2005; 44:1123-1137.

[52] Lugtenburg P, Avivi I, Berenschot H, Ilhan O, Marolleau JP, Nagler A, Rueda A, Tani M, Turgut M, Osborne S, Smith R, Pfreundschuh M. Efficacy and safety of subcutaneous and intravenous rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone in first-line diffuse large B-cell lymphoma: the randomized MabEase study. Haematologica. 2017; 102:1913-1922.

[53] Shire SJ, Shahrokh Z, Liu J. Challenges in the development of high protein concentration formulations. J Pharm Sci. 2004; 93:1390-1402.

[54] Cleland JL, Lam X, Kendrick B, Yang J, Yang TH, Overcashier D, Brooks D, Hsu C, Carpenter JF. A specific molar ratio of stabilizer to protein is required for storage stability of a lyophilized monoclonal antibody. J Pharm Sci. 2001; 90:310-321.

[55] Bedu-Addo FK. Understanding lyophilization formulation development. Pharmaceu Tech. 2004; 10-18.

[56] Kasper JC, Friess W. The freezing step in lyophilization: Physico-chemical fundamentals, freezing methods and consequences on process performance and quality attributes of biopharmaceuticals. Eur J Pharm Biopharm. 2011; 78:248-263.

[57] Bhatnagar BS, Bogner RH, Pikal MJ. Protein stability during freezing: Separation of stresses and mechanisms of protein stabilization. Pharm Dev Technol. 2007; 12:505-523.

[58] Gharsallaoui A, Roudaut G, Chambin O, Voilley A, Saurel R. Applications of spray-drying in microencapsulation of food ingredients: An overview. Food Res Int. 2007; 40:1107–1121.

[59] Vehring R. Pharmaceutical particle engineering via spray drying. Pharm Res. 2008; 25:999–1022.

[60] Maltesen MJ, van de Weert M. Drying methods for protein pharmaceuticals. Drug Discov Today Technol. 2008; 5:e81-8.

[61] Kroon DJ, Baldwin-Ferro A, Lalan P. Identification of sites of degradation in a therapeutic monoclonal antibody by peptide mapping. Pharm Res. 1992; 9:1386-93.

[62] Depreter F, Pilcer G, Amighi K. Inhaled proteins: Challenges and perspectives. Int J Pharm. 2013; 447:251-280.

[63] Manning MC, Chou DK, Murphy BM, Payne RW, Katayama DS. Stability of protein pharmaceuticals: An update. Pharm Res. 2010; 27:544-575.

[64] Wakankar AA, Borchardt RT. Formulation considerations for proteins susceptible to asparagine deamidation and aspartate isomerization. J Pharm Sci. 2006; 95:2321–2336.

[65] Robinson NE. Protein deamidation. Proc Natl Acad Sci USA. 2002; 99:5283–5288.

[66] Harris RJ, Kabakoff B, Macchi FD, Shen FJ, Kwong M, Andya JD, Shire SJ, Bjork N, Totpal K, Chen AB. Identification of multiple sources of charge heterogeneity in a recombinant antibody. J Chromatogr B Biomed Sci Appl. 2001; 752:233-245.

[67] Terashima I, Koga A, Nagai H. Identification of deamidation and isomerization sites on pharmaceutical recombinant antibody using $H_2{}^{18}O$. Anal Biochem. 2007; 368:49-60.

[68] Wakankar AA, Borchardt RT, Eigenbrot C, Shia S, Wang YJ, Shire SJ, Liu JL. Aspartate isomerization in the complementarity-determining regions of two closely related monoclonal antibodies. Biochemistry. 2007; 46:1534-1544.

[69] Oliyai C, Borchardt RT. Chemical pathways of peptide degradation. IV. Pathways, kinetics, and mechanism of degradation of an aspartyl residue in a model hexapeptide. Pharm Res. 1993; 10:95-102.

[70] Cacia J, Keck R, Presta LG, Frenz J. Isomerization of an aspartic acid residue in the complementarity-determining regions of a recombinant antibody to human IgE: identification and effect on binding affinity. Biochemistry. 1996; 35:1897-1903.

[71] Folzer E, Diepold K, Bomans K, Finkler C, Schmidt R, Bulau P, Huwyler J, Mahler HC, Koulov AV. Selective oxidation of methionine and tryptophan residues in a therapeutic IgG1 molecule. J Pharm Sci. 2015; 104:2824-2831.

[72] Kim YH, Berry AH, Spencer DS, Stites WE. Comparing the effect on protein stability of methionine oxidation versus mutagenesis: Steps toward engineering oxidative resistance in proteins. Protein Eng. 2001; 14:343-347.

[73] Wang W, Vlasak J, Li Y, Pristatsky P, Fang Y, Pittman T, Roman J, Wang Y, Prueksaritanont T, Ionescu R. Impact of methionine oxidation on the binding of human IgG1 to FcRn and Fcγ receptors. Mol Immunol. 2011; 48:860-866.

[74] Wu Y, Levons J, Narang AS, Raghavan K, Rao VM. Reactive impurities in excipients: profiling, identification and mitigation of drug–excipient incompatibility. AAPS PharmSciTech. 2011; 12:1248-1263.

[75] Gao X, Ji JA, Veeravalli K, Wang YJ, Zhang T, Mcgreevy W, Zheng K, Kelley RF, Laird MW, Liu J, Cromwell M. Effect of individual Fc methionine oxidation on FcRn binding: Met252 oxidation impairs FcRn binding more profoundly than Met428 oxidation. J Pharm Sci. 2015; 104:368-377.

[76] Darby N, Creighton TE. Disulfide bonds in protein folding and stability. Methods Mol Biol. 1995; 40:219-252.

[77] Andya JD, Maa YF, Costantino HR, Nguyen PA, Dasovich N, Sweeney TD, Hsu CC, Shire SJ. The effect of formulation excipients on protein stability and aerosol performance of spray-dried powders of a recombinant Humanised anti-IgE monoclonal antibody. Pharm Res. 1999; 16:350-358.

[78] Wu H, Kroe-Barrett R, Singh S, Robinson AS, Roberts CJ. Competing aggregation pathways for monoclonal antibodies. FEBS Lett. 2014; 588:936-941.

[79] Doye JPK. The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. J Chem Phys. 1999; 110:6896-6906.

[80] Menzen T, Friess W. High-throughput melting-temperature analysis of a monoclonal antibody by differential scanning fluorimetry in the presence of surfactants. J Pharm Sci. 2013; 102:415-428.

[81] He F, Hogan S, Latypov RF, Narhi LO, Razinkov VI. High throughput thermostability screening of monoclonal antibody formulations. J Pharm Sci. 2010; 99:1707-1720.

[82] Pereira VL, Fernandes JO, Cunha SC. Mycotoxins in cereals and related foodstuffs: A review on occurrence and recent methods of analysis. Trends Food Sci Tech. 2014; 36: 96–136.

[83] Perevozchikova T, Nanda H, Nesta DP, Roberts CJ. Protein adsorption, desorption, and aggregation mediated by solid-liquid interfaces. J Pharm Sci. 2015; 104:1946-1959.

[84] Vlasak J, Ionescu R. Fragmentation of monoclonal antibodies. MAbs. 2011; 3:253–263.

[85] Lowe D, Dudgeon K, Rouet R, Schofield P, Jermutus L, Christ D. Aggregation, stability, and formulation of human antibody therapeutics. Adv Protein Chem Struct Biol. 2011; 84:41-61.

[86] Shire SJ. Formulation and manufacturability of biologics. Curr Opin Biotechnol. 2009; 20:708-714.

[87] Hermeling S, Crommelin DJ, Schellekens H, Jiskoot W. Structure-immunogenicity relationships of therapeutic proteins. Pharm Res. 2004; 21:897-903.

[88] Roberts CJ. Therapeutic protein aggregation: mechanisms, design, and control. Trends Biotechnol. 2014; 32:372-380.

[89] Roberts CJ. Protein aggregation and its impact on product quality. Curr Opin Biotechnol 2014; 32:372-380.

[90] Cromwell ME, Hilario E, Jacobson F. Protein aggregation and bioprocessing. AAPS J. 2006; 15:572-579.

[91] Sharma VK, Chih HW, Mrsny RJ, Daugherty, A.L. The formulation and delivery of monoclonal antibodies. Therapeutic monoclonal antibodies: from bench to clinic. 2009; 671-705.

[92] Salinas BA, Sathish HA, Shah AU, Carpenter JF, Randolph TW. Buffer-dependent fragmentation of a Humanised full-length monoclonal antibody. J Pharm Sci. 2010; 99:2962-2274.

[93] Zheng JY, Janis LJ. Influence of pH, buffer species, and storage temperature on physicochemical stability of a Humanised monoclonal antibody LA298. Int J Pharm. 2006; 308:46-51.

[94] Lam XM, Oeswein JQ, Ongpipattanakul B, Shahrokh Z, Wang SX, Weissburg RP, Wong RL. Antibody formulation. Patent US 6991790. 1997.

[95] Zhang B, Yang Y, Yuk I, Pai R, McKay P, Eigenbrot C, Dennis M, Katta V, Francissen KC. Unveiling a glycation hot spot in a recombinant Humanised monoclonal antibody. Anal Chem. 2008; 80:2379-2390.

[96] Chang LL, Pikal MJ. Mechanisms of protein stabilization in the solid state. J Pharm Sci. 2009; 98:2886-2908.

[97] Fischer S, Hoernschemeyer J, Mahler HC. Glycation during storage and administration of monoclonal antibody formulations. Eur J Pharm Biopharm. 2008; 70:42-50.

[98] Elbein AD, Pan YT, Pastuszak I, Carroll D. New insights on trehalose: a multifunctional molecule. Glycobiology. 2003; 13:17-27.

[99] O'Brien J. Stability of trehalose, sucrose and glucose to nonenzymatic browning in model systems. J Food Sci. 1996; 61:679-682.

[100] Meyer JD, Nayar R, Manning MC. Impact of bulking agents on the stability of a lyophilized monoclonal antibody. Eur J Pharm Sci. 2009; 38:29-38.

[101] Mahler HC, Müller R, Friess W, Delille A, Matheus S. Induction and analysis of aggregates in a liquid IgG1-antibody formulation. Eur J Pharm Biopharm. 2005; 59:407-417.

[102] Thirumangalathu R, Krishnan S, Ricci MS, Brems DN, Randolph TW, Carpenter JF. Silicone oil and agitation-induced aggregation of a monoclonal antibody in aqueous solution. J Pharm Sci. 2009; 98:3167-3181.

[103] Chou DK, Krishnamurthy R, Randolph TW, Carpenter JF, Manning MC. Effects of Tween 20 and Tween 80 on the stability of albutropin during agitation. J Pharm Sci. 2005; 94:1368-1381.

[104] Chen B, Bautista R, Yu K, Zapata GA, Mulkerrin MG, Chamow SM. Influence of histidine on the stability and physical properties of a fully human antibody in aqueous and solid forms. Pharm Res. 2003; 20:1952-1960.

[105] Arakawa T, Ejima D, Tsumoto K, Obeyama N, Tanaka Y, Kita Y, Timasheff SN. Suppression of protein interactions by arginine: A proposed mechanism of the arginine effects. Biophys Chem. 2007; 127:1-8.

[106] Liu J, Nguyen MD, Andya JD, Shire SJ. Reversible self-association increases the viscosity of a concentrated monoclonal antibody in aqueous solution. J Pharm Sci. 2005; 94:1928-1940.

[107] Arakawa T, Philo JS, Tsumoto K, Yumioka R, Ejima D. Elution of antibodies from a Protein-A column by aqueous arginine solutions. Protein Expr Purif. 2004; 36:244-248.

[108] Tian F, Middaugh CR, Offerdahl T, Munson E, Sane S, Rytting JH. Spectroscopic evaluation of the stabilization of Humanised monoclonal antibodies in amino acid formulations. Int J Pharm. 2007; 335:20-31.

[109] Flanagan RJ, Jones AL. Fab antibody fragments: some applications in clinical toxicology. Drug Saf. 2004; 27:1115-1133.

[110] Abuchowski A, van Es T, Palczuk NC, Davis FF. Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. J Biol Chem. 1977; 252:3578-3581.

[111] Abuchowski A, McCoy JR, Palczuk NC, van Es T, Davis FF. Effect of covalent attachment of polyethylene-glycol on immunogenicity and circulating life of bovine liver catalase. J Biol Chem. 1977; 252:3582-3586.

[112] Chapman AP. PEGylated antibodies and antibody fragments for improved therapy: a review. Adv Drug Deliv Rev. 2002; 54:531-545.

[113] Weir AN, Nesbitt A, Chapman AP, Popplewell AG, Antoniw P, Lawson AD. Formatting antibody fragments to mediate specific therapeutic functions. Biochem Soc Trans. 2002; 30:512-516.

[114] Mordenti J, Thomsen K, Licko V, Berleau L, Kahn JW, Cuthbertson RA, Duenas ET, Ryan AM, Schofield C, Berger TW, Meng YG, Cleland J. Intraocular pharmacokinetics and safety of a Humanised monoclonal antibody in rabbits after intravitreal administration of a solution or a PLGA microsphere formulation. Toxicol Sci. 1999; 52:101-106.

[115] Wang J, Chua KM, Wang CH. Stabilization and encapsulation of human immunoglobulin G into biodegradable microspheres. J Colloid Interface Sci. 2004; 271:92-101.

[116] Schweizer D, Serno T, Goepferich A. Controlled release of therapeutic antibody formats. Eur J Pharm Biopharm. 2014; 88:291-309.

[117] Breen ED, Curley JG, Overcashier DE, Hsu CC, Shire SJ. Effect of moisture on the stability of a lyophilized Humanised monoclonal antibody formulation. Pharm Res. 2001; 18:1345-1353.

[118] Duddu SP, Dal Monte PR. Effect of glass transition temperature on the stability of lyophilized formulations containing a chimeric therapeutic monoclonal antibody. Pharm Res. 1997; 14:591-595.

[119] Moore GE. Cramming more components onto integrated circuits. P IEEE. 1998; 86:82-85.

[120] Veurink M, Westermaier Y, Gurny R, Scapozza L. Breaking the aggregation of the monoclonal antibody bevacizumab (Avastin®) by dexamethasone phosphate: insights from molecular modelling and asymmetrical flow field-flow fractionation. Pharm Res. 2013; 30:1176-1187.

[121] Barata TS, Zhang C, Dalby PA, Brocchini S, Zloh M. Identification of Protein-Excipient Interaction Hotspots Using Computational Approaches. Int J Mol Sci. 2016; 17:853.

[122] Chennamsetty N, Voynov V, Kayser V, Helk B, Trout BL. Design of therapeutic proteins with enhanced stability. Proc Natl Acad Sci USA. 2009; 106:11937-1142.

[123] Buck PM, Kumar S, Singh SK. Insights into the potential aggregation liabilities of the b12 Fab fragment via elevated temperature molecular dynamics. Protein Eng Des Sel. 2013; 26:195-205.

[124] Chaudhri A, Zarraga IE, Yadav S, Patapoff TW, Shire SJ, Voth GA. The role of amino acid sequence in the self-association of therapeutic monoclonal antibodies: insights from coarse-grained modeling. J Phys Chem B. 2013; 117:1269-1279.

[125] Buck PM, Chaudhri A, Kumar S, Singh SK. Highly viscous antibody solutions are a consequence of network formation caused by domain-domain electrostatic complementarities: insights from coarse-grained simulations. Mol Pharm. 2015; 12:127-139.

[126] Conchillo-Solé O, De Groot NS, Avilés FX, Vendrell J, Daura X, Ventura S. AGGRESCAN: A server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics. 2007; 8:65.

[127] Gasior P, Kotulska M. FISH Amyloid - a new method for finding amyloidogenic segments in proteins based on site specific co-occurrence of aminoacids. BMC Bioinformatics. 2014; 15:54.

[128] Trovato A, Seno F, Tosatto SC. The PASTA server for protein aggregation prediction. Protein Eng Des Sel. 2007; 20:521-523.

[129] Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol. 2004; 22:1302-1306.

[130] Tartaglia GG, Vendruscolo M. The Zyggregator method for predicting protein aggregation propensities. Chem Soc Rev. 2008; 37:1395-1401.

[131] Wang X, Das TK, Singh SK, Kumar S. Potential aggregation prone regions in biotherapeutics. MAbs. 2009; 1:254-267.

[132] Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, Güntert P. Optimal isotope labelling for NMR protein structure determinations. Nature. 2006; 440:52-57.

[133] Mintseris J, Pierce B, Wiehe K, Anderson R, Chen R, Weng Z. Integrating statistical pair potentials into protein complex prediction. Proteins. 2007; 69:511-520.

[134] Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit. 1996; 9:1-5.

[135] Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. J Comput Chem. 2005; 26:1668-1688.

[136] Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem. 2009; 30:1545-1614.

[137] Todorov IT, Smith W, Kostya Trachenko, Doveb MT. DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism. J Mater Chem. 2006; 16:1911-1918.

[138] Plimpton S. Fast parallel algorithms for short-range molecular dynamics. J Comput Phys. 1995; 117:1-19.

[139] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem. 2005; 26:1781-1802.

[140] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005; 26:1701-1718.

[141] Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc. 1995; 117:5179-5197.

[142] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998; 102:3586-3616.

[143] Scott WRP, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krüger P, Van Gunsteren WF. The GROMOS biomolecular simulation program package. J Phys Chem A. 1999; 103:3596-3607.

[144] Ryckaert J P, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comp. Phys. 1977; 23:327–341.

[145] Andersen HC. Rattle: a 'velocity' version of the shake calculations algorithm for molecular dynamics. J. Comput. Phys. 1983; 52:24–34.

[146] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. J. Comput. Chem. 1997; 18:463–1472.

[147] T. Darden, D. York and L. Pedersen. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. J Chem Phys. 1993; 98:10089-10092.

[148] R. Hockney and J. Eastwood. Computer Simulation Using Particles. IOP, Bristol, 1988.

[149] Meredith SC. Protein denaturation and aggregation: Cellular responses to denatured and aggregated proteins. Ann N Y Acad Sci. 2005; 1066:181-221.

[150] Fernández A. What factor drives the fibrillogenic association of beta-sheets? FEBS Lett. 2005; 579:6635-6640.

[151] Vos MJ, Zijlstra MP, Carra S, Sibon OC, Kampinga HH. Small heat shock proteins, protein degradation and protein aggregation diseases. Autophagy. 2011; 7:101-103.

[152] Dobson CL, Devine PW, Phillips JJ, Higazi DR, Lloyd C, Popovic B, Arnold J, Buchanan A, Lewis A, Goodman J, van der Walle CF, Thornton P, Vinall L, Lowne D, Aagaard A, Olsson LL, Ridderstad Wollberg A, Welsh F, Karamanos TK, Pashley CL, Iadanza MG, Ranson NA, Ashcroft AE, Kippen AD, Vaughan TJ, Radford SE, Lowe DC. Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo. Sci Rep. 2016; 6:38644.

[153] Pierce BG, Hourai Y, Weng Z. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. PLoS One. 2011; 6:e24657.

[154] Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; 31:455-461.

[155] Sterling T, Irwin JJ. ZINC 15--ligand discovery for everyone. J Chem Inf Model. 2015; 55:2324-2337.

[156] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem. 2009; 16: 2785-2791.

[157] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25:1605-1612.

[158] Raman EP, Guvench O, MacKerell AD Jr. CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. J Phys Chem B. 2010; 114:12981-12994.

[159] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force field. J Comput Chem. 2010; 31:671-690.

[160] Tetko IV, Tanchuk VY. Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. J Chem Inf Comput Sci. 2002; 42:1136-1145.

[161] Andya JD, Hsu CC, Shire SJ. Mechanisms of aggregate formation and carbohydrate excipient stabilization of lyophilized Humanised monoclonal antibody formulations. AAPS PharmSci. 2003; 5:E10.

[162] Hamrang Z, Rattray NJW, Pluen A. Proteins behaving badly: emerging technologies in profiling biopharmaceutical aggregation. Trends Biotechnol. 2013; 31:448-458.

[163] British Pharmacopoeia 2018. London: Stationery Office. 2017.

[164] Connolly BD, Petry C, Yadav S, Demeule B, Ciaccio N, Moore JM, Shire SJ, Gokarn YR. Weak interactions govern the viscosity of concentrated antibody solutions: high-throughput analysis using the diffusion interaction parameter. Biophys J. 2012; 103:69-78.

[165] Yadav S, Shire SJ, Kalonia DS. Specific ion and buffer effects on protein–protein interactions of a monoclonal antibody, viscosity behavior of high-concentration monoclonal antibody solutions: correlation with interaction parameter and electroviscous effects. J Pharm Sci. 2012; 101:998-1011.

[166] Hernández-Jiménez J, Martínez-Ortega A, Salmerón-García A, Cabeza J, Prados JC, Ortíz R, Navas N. Study of aggregation in therapeutic monoclonal antibodies subjected to stress and long-term stability tests by analyzing size exclusion liquid chromatographic profiles. Int J Biol Macromol. 2018; 118:511-524.

[167] Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B. 2007; 111:7812-7824.

[168] Marrink SJ, Tieleman DP. Perspective on the Martini model. Chem Soc Rev. 2013; 42:6801-6822.

[169] Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse-grained force field: Extension to proteins. J Chem Theory Comput. 2008; 4:819–834.

[170] de Jong DH, Singh G, Bennett WF, Arnarez C, Wassenaar TA, Schäfer LV, Periole X, Tieleman DP, Marrink SJ. Improved parameters for the MARTINI coarse-grained protein force field. J Chem Theory Comput. 2013; 9:687-697.

[171] Periole X, Cavalli M, Marrink SJ, Ceruso MA. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics and intermolecular recognition. J Chem Theory Comput. 2009; 5:2531-2543.

[172] Chowdhury N, Bagchi A. Identification of ligand binding activity and DNA recognition by RhlR protein from opportunistic pathogen Pseudomonas aeruginosa-a molecular dynamic simulation approach. J Mol Recognit. 2018; e2738.

[173] Periole X, Knepp AM, Sakmar TP, Marrink SJ, Huber T. Structural determinants of the supramolecular organization of G protein-coupled receptors in bilayers. J Am Chem Soc. 2012; 134:10959-10965.

[174] Sengupta D, Marrink SJ. Lipid-mediated interactions tune the association of glycophorin A helix and its disruptive mutants in membranes. Phys Chem Chem Phys. 2010; 12:12987-12996.

[175] Psachoulia E, Marshall DP, Sansom MS. Molecular dynamics simulations of the dimerization of transmembrane α-helices. Acc Chem Res. 2010; 43:388-396.

[176] Wassenaar TA, Pluhackova K, Moussatova A, Sengupta D, Marrink SJ, Tieleman DP, Böckmann RA. High-throughput simulations of dimer and trimer assembly of membrane proteins. The DAFT approach. J Chem Theory Comput. 2015; 11:2278-2291.

[177] Wassenaar TA, Ingólfsson HI, Böckmann RA, Tieleman DP, Marrink SJ. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. J Chem Theory Comput. 2015; 11:2144-2155.

[178] Wassenaar TA, Ingólfsson HI, Priess M, Marrink SJ, Schäfer LV. Mixing MARTINI: electrostatic coupling in hybrid atomistic-coarse-grained biomolecular simulations. J Phys Chem B. 2013; 117:3516-3530.

[179] Koziara KB, Stroet M, Malde AK, Mark AE. Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. J Comput Aided Mol Des. 2014; 28:221-233.

[180] López CA, Rzepiela AJ, de Vries AH, Dijkhuizen L, Hünenberger PH, Marrink SJ. MARTINI coarse-grained force field: extension to carbohydrates. J Chem Theory Comput. 2009; 5:3195-3210.

[181] Zhu YL, Beroza P, Artis DR. Including explicit water molecules as part of the protein structure in MM/PBSA calculations. J Chem Inf Model. 2014; 54:462-469.

[182] König G, Bruckner S, Boresch S. Absolute hydration free energies of blocked amino acids: implications for protein solvation and stability. Biophys J. 2013; 104:453-462.

[183] Homeyer N, Gohlke H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. Mol Inform. 2012; 31:114-122.

[184] Shirts MR, Pande VS. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. J Chem Phys. 2005; 122:144107-144114.

[185] Kumari R, Kumar R; Open Source Drug Discovery Consortium, Lynn A. g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model. 2014; 54: 1951-1962.

[186] Arnarez C, Uusitalo JJ, Masman MF, Ingólfsson HI, de Jong DH, Melo MN, Periole X, de Vries AH, Marrink SJ. Dry Martini, a coarse-grained force field for lipid membrane simulations with implicit solvent. J Chem Theory Comput. 2015; 11:260-275.

[187] Pluhackova, K., Wassenaar, T. A., and Böckmann, R. A. Molecular Dynamics Simulations of Membrane Proteins. Methods Mol Bio. 2013; 1033:85-101.

[188] Arakawa T, Ejima D, Tsumoto K, Obeyama N, Tanaka Y, Kita Y, Timasheff SN. Suppression of protein interactions by arginine: a proposed mechanism of the arginine effects. Biophys Chem. 2007; 127:1-8.

[189] Nuhu MM, Curtis R. Arginine dipeptides affect insulin aggregation in a pH- and ionic strength-dependent manner. Biotechnol J. 2015; 10:404-16.

[190] Reform Biologics LLC. Viscosity-reducing excipient compounds for protein formulations. US9605051B2. 2015.

[191] Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. Curr Top Med Chem. 2014; 14:1923-1938.

[192] McLellan JS, Correia BE, Chen M, Yang Y, Graham BS, Schief WR, Kwong PD. Design and characterization of epitope-scaffold immunogens that present the motavizumab epitope from respiratory syncytial virus. J Mol Biol. 2011; 409:853-866.

[193] Fernández P, Trenholme A, Abarca K, Griffin MP, Hultquist M, Harris B, Losonsky GA. A phase 2, randomized, double-blind safety and pharmacokinetic assessment of respiratory syncytial virus (RSV) prophylaxis with motavizumab and palivizumab administered in the same season. BMC Pediatr. 2010; 10:38.

[194] Webb B, Sali A. Comparative protein structure modeling using MODELLER. Curr Protoc Bioinformatics. 2014; 47:5.6.1-32.

[195] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996; 14:33-38.

[196] Inoue N, Takai E, Arakawa T, Shiraki K. Arginine and lysine reduce the high viscosity of serum albumin solutions for pharmaceutical injection. J Biosci Bioeng. 2014; 117:539-543.

[197] Wang S, Zhang N, Hu T, Dai W, Feng X, Zhang X, Qian F. Viscosity-lowering effect of amino acids and salts on highly concentrated solutions of two IgG1 monoclonal antibodies. Mol Pharm. 2015; 12:4478-4487.

[198] Bereau T, Kremer K. Automated parametrization of the coarse-grained MARTINI force field for small organic molecules. J Chem Theory Comput. 2015; 11:2783-2791.

9. Supplementary Information

9.1. MARTINI Force Field Parameters for Compound X

```
[moleculetype]
; molname       nrexcl
  CPX           2

[atoms]
; id    type    resnr   residue  atom    cgnr    charge
    1     P5      1     CPX       P01     1         0
    2     Na      1     CPX       N01     2         0
    3     SP1     1     CPX       S01     3         0
    4     SNa     1     CPX       S02     4         0
    5     Nd      1     CPX       N02     5         0
    6     SP2     1     CPX       S03     6         0
    7     SP1     1     CPX       S04     7         0
    8     SP1     1     CPX       S05     8         0
    9     SN0     1     CPX       S06     9         0
    10    SN0     1     CPX       S07     10        0

[bonds]
; i j     funct     length     force.c.
  1 2       1         0.382     15000
  1 5       1         0.304     15000
  5 6       1         0.280     15000

[constraints]
;  i    j     funct    length
   2    3      1       0.301
   3    4      1       0.125
   2    4      1       0.314
   6    7      1       0.122
   7    8      1       0.210
   6    8      1       0.204
   9   10      1       0.165
   8    9      1       0.245
```

[angles]

; i j k funct angle force.c.

3 2 1 2 180 125.0

4 2 1 2 120 150.0

2 1 5 2 95 125.0

1 5 6 2 150 60.0

5 6 8 2 130 100.0

5 6 7 2 160 125.0

6 8 9 2 150 250.0

7 8 9 2 180 250.0

8 9 10 2 180 1000.0

[dihedrals]

; i j k l funct angle force.c.

9.2. Coarse-grained Coordinates for Compound X

CPX

10

1CPX P01 1 1.832 2.156 2.128 0.0957 0.0654 0.0116

1CPX N01 2 1.894 1.792 2.217 0.1529 -0.0293 0.0861

1CPX S01 3 1.832 1.511 2.305 0.0174 -0.0154 0.0361

1CPX S02 4 1.947 1.514 2.352 -0.0017 -0.0596 0.0862

1CPX N02 5 2.079 2.128 1.967 0.5704 -0.1574 0.0712

1CPX S03 6 2.022 2.031 1.772 0.0939 -0.0045 0.0786

1CPX S04 7 2.058 1.974 1.671 0.1023 -0.1258 0.1482

1CPX S05 8 1.871 2.065 1.639 0.0347 -0.2877 0.0674

1CPX S06 9 1.695 2.088 1.460 -0.3013 -0.1319 -0.2231

1CPX S07 10 1.630 2.198 1.358 0.1870 0.1349 -0.2783

3.70927 3.70927 3.70927

9.3. Coarse-grained Coordinates for Trehalose

Trehalose

6

1TREH B1 1 4.181 4.489 4.918 -0.0727 -0.1038 0.0979

1TREH B2 2 4.227 4.319 5.053 0.1427 0.0380 0.2070

1TREH B3 3 4.146 4.108 5.150 -0.2139 0.1489 0.1633

```
  1TREH    B4    4   4.605   4.065   4.819 -0.0842 -0.0204 -0.1003
  1TREH    B5    5   4.774   4.214   4.769 -0.1414  0.1776  0.2694
  1TREH    B6    6   4.472   3.900   4.879  0.0771 -0.1224 -0.0177
 9.25395   9.25395   9.25395
```

9.4. Bash Script to Set Up Simulations with Compound X

```
#!/bin/bash
./daft.sh -aa MEDI578=5jz7-CD.pdb -c 2 -3d -gmxrc GMXRC -dssp dssp-2.0.4-linux-amd64 -pymol pymol -n 1024 --insane{-salt=0.1} --martinize{-ff=elnedyn22} --martinate{-T=300} -d 3 -time 512 -ff elnedyn22 -sol W
cd MEDI578-MEDI578
mv daft cpx
cd cpx
for d in */ ; do
    echo "$d"
    cd "$d"
    cp ../../../martini_v2.2.itp .
    mv martini_v2.2.itp martini.itp
    cp ../../../martini_v2.0_ions.itp .
    cp ../../../martini_v2.2_aminoacids.itp .
    cp ../../../CPX.itp .
    cp ../../../daft_sample.top tmp
    cp ../../../daft_sample.ndx .
    nanum=$(grep -c "NA+   " daft.gro)
    clnum=$(grep -c "CL-   " daft.gro)
    gmx insert-molecules -f daft.gro -ci ../../../CPX.gro -o daft_CPX.gro -n daft.ndx -replace Solvent -nmol 10 -radius 0.26
    while true; do
        martininanum1=$(grep -c "NA+   " daft_CPX.gro)
        martininanum2=$(grep -c "CL-   " daft_CPX.gro)
        if (("$martininanum1" == "$nanum")); then
            if (("$martininanum2" == "$clnum")); then
                break
            else
                rm daft_CPX.gro
                gmx insert-molecules -f daft.gro -ci ../../../CPX.gro -o daft_CPX.gro -n daft.ndx -replace Solvent -nmol 10 -radius 0.26
                martininanum1=$(grep -c "NA+   " daft_CPX.gro)
```

```
            martininanum2=$(grep -c "CL-   " daft_CPX.gro)
            fi
        else
            rm daft_CPX.gro
            gmx insert-molecules -f daft.gro -ci ../../../CPX.gro -o daft_CPX.gro -n daft.ndx -
replace Solvent -nmol 10 -radius 0.26
            martininanum1=$(grep -c "NA+   " daft_CPX.gro)
            martininanum2=$(grep -c "CL-   " daft_CPX.gro)
            fi
    done
    echo -e "keep 0\nq\n" | gmx make_ndx -f daft_CPX.gro -o daft_CPX_system.ndx
    echo -e "keep 11\nq\n" | gmx make_ndx -f daft_CPX.gro -o daft_CPX_solvent.ndx
    sed "s/non-Protein/Solvent/" daft_CPX_solvent.ndx > daft_CPX_sol.ndx
    cat daft_sample.ndx daft_CPX_sol.ndx daft_CPX_system.ndx > all.all
    rm *.ndx
    mv all.all daft.ndx
    martiniwaternum=$(grep -c "  W" daft_CPX.gro)
    sed -i "\$aW  $martiniwaternum" tmp
    sed -i "\$aNA+      $martininanum1" tmp
    sed -i "\$aCL-       $martininanum2" tmp
    sed -i "\$aCPX      10" tmp
    rm daft.top
    mv tmp daft.top
    rm daft.gro
    mv daft_CPX.gro daft.gro
        source martinate.sh
    cd ..
done
```

9.5. daft_sample.top for Simulations with Compound X

```
#include "martini.itp"
#include "martini_v2.2_aminoacids.itp"
#include "martini_v2.0_ions.itp"
#include " CPX.itp"
#include "MEDI578_C.itp"
#include "MEDI578_D.itp"

[ system ]
```

DAFT setup for MEDI578-MEDI578

[molecules]
; name number
MEDI578_C 1
MEDI578_D 1;
; name number
MEDI578_C 1
MEDI578_D 1;

9.6. Bash Script to Set Up Simulations with Trehalose

```
#!/bin/bash
./daft.sh -aa MEDI578=5jz7-CD.pdb -c 2 -3d -gmxrc GMXRC -dssp dssp-2.0.4-linux-amd64 -pymol pymol -n 1024 --insane{-salt=0.1} --martinize{-ff=elnedyn22} --martinate{-T=300} -d 3 -time 512 -ff elnedyn22 -sol W
cd MEDI578-MEDI578
mv daft trehalose
cd trehalose

for d in */ ; do
   echo "$d"
   cd "$d"
   cp ../../../martini_v2.2.itp .
   mv martini_v2.2.itp martini.itp
   cp ../../../martini_v2.0_ions.itp .
   cp ../../../martini_v2.2_aminoacids.itp .
   cp ../../../martini_v2.0_sugars.itp .
   cp ../../../daft_sample.top tmp
   cp ../../../daft_sample.ndx .
   nanum=$(grep -c "NA+   " daft.gro)
   clnum=$(grep -c "CL-   " daft.gro)
   gmx insert-molecules -f daft.gro -ci ../../../TRE.gro -o daft_TRE.gro -n daft.ndx -replace Solvent -nmol 10 -radius 0.26
   while true; do
      martininanum1=$(grep -c "NA+   " daft_TRE.gro)
      martininanum2=$(grep -c "CL-   " daft_TRE.gro)
      if (("$martininanum1" == "$nanum")); then
         if (("$martininanum2" == "$clnum")); then
```

```
            break
        else
            rm daft_TRE.gro
            gmx insert-molecules -f daft.gro -ci ../../../TRE.gro -o daft_TRE.gro -n daft.ndx
-replace Solvent -nmol 10 -radius 0.26
            martininanum1=$(grep -c "NA+   " daft_TRE.gro)
            martininanum2=$(grep -c "CL-   " daft_TRE.gro)
            fi
      else
        rm daft_TRE.gro
        gmx insert-molecules -f daft.gro -ci ../../../TRE.gro -o daft_TRE.gro -n daft.ndx -
replace Solvent -nmol 10 -radius 0.26
        martininanum1=$(grep -c "NA+   " daft_TRE.gro)
        martininanum2=$(grep -c "CL-   " daft_TRE.gro)
        fi
  done
  echo -e "keep 0\nq\n" | gmx make_ndx -f daft_TRE.gro -o daft_TRE_system.ndx
  echo -e "keep 11\nq\n" | gmx make_ndx -f daft_TRE.gro -o daft_TRE_solvent.ndx
  sed "s/non-Protein/Solvent/" daft_TRE_solvent.ndx > daft_TRE_sol.ndx
  cat daft_sample.ndx daft_TRE_sol.ndx daft_TRE_system.ndx > all.all
  rm *.ndx
  mv all.all daft.ndx
  martiniwaternum=$(grep -c "  W" daft_TRE.gro)
  sed -i "\$aW  $martiniwaternum" tmp
  sed -i "\$aNA+      $martininanum1" tmp
  sed -i "\$aCL-       $martininanum2" tmp
  sed -i "\$aTREH      10" tmp
  rm daft.top
  mv tmp daft.top
  rm daft.gro
  mv daft_TRE.gro daft.gro
        source martinate.sh
  cd ..
done
```

9.7. daft_sample.top for Simulations with Trehalose

```
#include "martini.itp"
#include "martini_v2.2_aminoacids.itp"
```

#include "martini_v2.0_ions.itp"
#include "martini_v2.0_sugars.itp"
#include "MEDI578_C.itp"
#include "MEDI578_D.itp"

[system]
DAFT setup for MEDI578-MEDI578

[molecules]
; name number
MEDI578_C 1
MEDI578_D 1;
; name number
MEDI578_C 1
MEDI578_D 1;

9.8. daft_sample.ndx Simulations with Compound X and Trehalose

[A_MEDI578]
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88
89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111
112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131
132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151
152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171
172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191
192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211
212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231
232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251
252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271
272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291
292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311
312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331
332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351
352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371
372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391
392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411
412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431

432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451
452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471
472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491
492 493 494 495 496 497 498 499 500 501 502

[B_MEDI578]

503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522
523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542
543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562
563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582
583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602
603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622
623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642
643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662
663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682
683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702
703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722
723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742
743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762
763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782
783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802
803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822
823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842
843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862
863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882
883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902
903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922
923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942
943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962
963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982
983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001
1002 1003 1004

[Solute]

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88
89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111
112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131
132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151
152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171

172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191
192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211
212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231
232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251
252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271
272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291
292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311
312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331
332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351
352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371
372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391
392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411
412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431
432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451
452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471
472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491
492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511
512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531
532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551
552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571
572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591
592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611
612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631
632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651
652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671
672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691
692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711
712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731
732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751
752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771
772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791
792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811
812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831
832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851
852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871
872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891
892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911
912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931
932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951

952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971
972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991
992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004
[Membrane]

9.9. Bash Script to Generate Dipeptides

```
#!/bin/bash
# non-protonated histidine residues
declare -a aminoacid1=("ALA" "GLY" "HIS" "LYS" "PRO" "GLN" "ARG" "SER")
declare -a aminoacid2=("ALA" "GLY" "HIS" "LYS" "PRO" "GLN" "ARG" "SER")
for i in "${aminoacid1[@]}"
do
  for j in "${aminoacid2[@]}"
  do
      pep=$i"-"$j
      echo "$pep"
      sed -e "s/XXX/"$i"/g" ref/XY.pdb > tmp1
      sed -e "s/YYY/"$j"/g" tmp1  > $pep"-aa.pdb"
      sed -e "s/XXX/"$i"/g" ref/create_dipeptides.tcl > tmp2
      sed -e "s/YYY/"$j"/g" tmp2 > tmp3
      vmd -dispdev text -e tmp3
      sed -i "s/HSP/HIS/g" $pep"-aa.pdb"
      rm tmp1 tmp2 tmp3
      python ref/martinize.py -f $pep"-aa.pdb" -name $pep -o $pep".top" -x $pep"-
CG.pdb" -ff martini22 -ss EEE -nt
    done
  done

# protonated histidine residues
declare -a aminoacid3=("HIS")
declare -a aminoacid4=("ALA" "GLY" "LYS" "PRO" "GLN" "ARG" "SER")
for i in "${aminoacid3[@]}"
do
  for j in "${aminoacid4[@]}"
  do
      pep1=$i"-"$j
      pep2=$j"-"$i
```

```
        echo -e "0\n" | python ref/martinize.py -f $pep1"-aa.pdb" -name "HSP-"$j -o "HSP-"$j".top" -x "HSP-"$j"-CG.pdb" -ff martini22 -ss EEE -nt -his
        echo -e "0\n" | python ref/martinize.py -f $pep2"-aa.pdb" -name $j"-HSP" -o $j"-HSP.top" -x $j"-HSP-CG.pdb" -ff martini22 -ss EEE -nt -his
      done
   done
echo -e "0\n0\n" | python ref/martinize.py -f HIS-HIS-aa.pdb -name HSP-HSP -o HSP-HSP.top -x HSP-HSP-CG.pdb -ff martini22 -ss EEE -nt -his
echo -e "0\n1\n" | python ref/martinize.py -f HIS-HIS-aa.pdb -name HSP-HIS -o HSP-HIS.top -x HSP-HIS-CG.pdb -ff martini22 -ss EEE -nt -his
echo -e "1\n0\n" | python ref/martinize.py -f HIS-HIS-aa.pdb -name HIS-HSP -o HIS-HSP.top -x HIS-HSP-CG.pdb -ff martini22 -ss EEE -nt –his
```

9.10. Coordinate File for the Dipeptide Backbone

```
ATOM      1  N   XXX    22      -1.195   0.201  -0.206  1.00  0.00           N
ATOM      2  CA  XXX    22       0.230   0.318  -0.502  1.00  0.00           C
ATOM      3  C   XXX    22       1.059  -0.390   0.542  1.00  0.00           C
ATOM      4  O   XXX    22       0.833  -0.270   1.749  1.00  0.00           O
ATOM      8  N   YYY    23       2.171  -1.248   0.107  1.00  0.00           N
ATOM      9  CA  YYY    23       2.773  -1.785   1.324  1.00  0.00           C
ATOM     10  C   YYY    23       4.279  -1.793   1.224  1.00  0.00           C
ATOM     11  O   YYY    23       4.866  -2.186   0.213  1.00  0.00           O
END
```

9.11. Tcl Scipt for Generating All-atom Coordinate file for Dipeptides

```
package require psfgen
resetpsf
topology ref/top_all36_prot.rtf
pdbalias residue HIS HSD
segment pep {pdb XXX-YYY-aa.pdb}
coordpdb XXX-YYY-aa.pdb pep
guesscoord
writepdb XXX-YYY-aa.pdb
exit
```

9.12. Bash Script to Set Up MEDI-578 Simulations with Dipeptides

```
#!/bin/bash
declare -a aminoacid1=("ALA" "GLY" "HIS" "HSP" "LYS" "PRO" "GLN" "ARG"
"SER")
declare -a aminoacid2=(""ALA" "GLY" "HIS" "HSP" "LYS" "PRO" "GLN" "ARG"
"SER")
npep=10
for i in "${aminoacid1[@]}"
do
  for j in "${aminoacid2[@]}"
  do
      pep=$i"-"$j
      ./daft.sh -aa MEDI578=5jz7-CD.pdb -c 2 -3d -gmxrc GMXRC -dssp ./dssp-2.0.4-
linux-amd64 -pymol pymol -n 256 --insane{-salt=0} --martinize{-ff=elnedyn22} --
martinate{-T=300} -d 3 -time 512 -ff elnedyn22 -sol W
      cd MEDI578-MEDI578
      mv daft $pep
      cd $pep
      for d in */ ; do
         echo "$d"
         cd "$d"
         cp ../../../martini_v2.2.itp .
         mv martini_v2.2.itp martini.itp
         cp ../../../martini_v2.0_ions.itp .
         cp ../../../martini_v2.2_aminoacids.itp .
         cp ../../../dipeptides/$pep".itp" .
         cp ../../../daft_sample.top tmp.top
         cp ../../../daft_sample.ndx .
         gmx insert-molecules -f daft.gro -ci ../../../dipeptides/$pep"-CG.pdb" -o
daft_insert.gro -n daft.ndx -replace Solvent -nmol $npep -radius 0.26
         sed "s/NA+    NA+/W       W/" daft_insert.gro > daft_DPD.gro
         echo -e 'keep 0\nq\n' | gmx make_ndx -f daft_DPD.gro -o daft_DPD_system.ndx
         echo -e '!0&!1\nname 4 Solvent\nkeep 4\nq\n' | gmx make_ndx -f daft_DPD.gro
-n daft_sample.ndx -o daft_DPD_sol.ndx
         cat daft_sample.ndx daft_DPD_sol.ndx daft_DPD_system.ndx > all.all
         rm *.ndx
         mv all.all daft.ndx
         martiniwaternum=$(grep -c "  W" daft_DPD.gro)
```

```
        sed -i '6 i\#include "'$pep'.itp"' tmp.top
        sed -i "\$aW  $martiniwaternum" tmp.top
        sed -i "\$a$pep       $npep"  tmp.top
        rm daft_insert.gro daft.gro daft.top
        mv tmp.top daft.top
        gmx grompp -f ../../../dipeptides/ref/water_min.mdp -p daft.top -c daft_DPD.gro
-o ion -n daft.ndx
        echo -e "W\n" | gmx genion -s ion.tpr -p daft.top -nname CL- -pname NA+ -
conc 0.1 -neutral -o daft.gro
        rm daft_DPD.gro mdout.mdp ion.tpr
        rm *#
        source martinate-run.sh
        cd ..
        done
      cd ../..
    done
  done
```

9.13. daft_sample.top for MEDI-578 Simulations

```
#include "martini.itp"
#include "martini_v2.2_aminoacids.itp"
#include "martini_v2.0_ions.itp"
#include "MEDI578_C.itp"
#include "MEDI578_D.itp"


[ system ]
DAFT setup for MEDI578-MEDI578

[ molecules ]
; name       number
MEDI578_C    1
MEDI578_D    1;
; name       number
MEDI578_C    1
MEDI578_D    1;
```

9.14. daft_sample.ndx for MEDI-578 Simulations

[A_MEDI578]
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88
89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111
112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131
132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151
152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171
172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191
192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211
212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231
232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251
252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271
272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291
292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311
312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331
332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351
352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371
372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391
392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411
412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431
432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451
452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471
472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491
492 493 494 495 496 497 498 499 500 501 502
[B_MEDI578]
503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522
523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542
543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562
563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582
583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602
603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622
623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642
643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662
663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682
683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702
703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722

723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742
743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762
763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782
783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802
803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822
823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842
843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862
863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882
883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902
903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922
923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942
943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962
963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982
983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001
1002 1003 1004

[Solute]

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88
89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111
112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131
132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151
152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171
172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191
192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211
212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231
232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251
252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271
272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291
292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311
312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331
332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351
352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371
372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391
392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411
412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431
432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451
452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471
472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491

492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511
512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531
532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551
552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571
572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591
592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611
612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631
632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651
652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671
672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691
692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711
712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731
732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751
752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771
772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791
792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811
812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831
832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851
852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871
872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891
892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911
912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931
932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951
952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971
972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991
992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004
[Membrane]

9.15. Bash Script to Set Up MEDI1912 Simulations with Dipeptides

```
#!/bin/bash
declare -a aminoacid1=("ALA" "GLY" "HIS" "HSP" "LYS" "PRO" "GLN" "ARG"
"SER")
declare -a aminoacid2=(""ALA" "GLY" "HIS" "HSP" "LYS" "PRO" "GLN" "ARG"
"SER")
npep=10
for i in "${aminoacid1[@]}"
do
```

```
  for j in "${aminoacid2[@]}"
  do
      pep=$i"-"$j
      ./daft.sh -aa MEDI1912= MEDI1912.pdb -c 2 -3d -gmxrc GMXRC -dssp ./dssp-2.0.4-linux-amd64 -pymol pymol -n 256 --insane{-salt=0} --martinize{-ff=elnedyn22} --martinate{-T=300} -d 3 -time 512 -ff elnedyn22 -sol W
      cd MEDI1912-MEDI1912
      mv daft $pep
      cd $pep
      for d in */ ; do
         echo "$d"
         cd "$d"
         cp ../../../martini_v2.2.itp .
         mv martini_v2.2.itp martini.itp
         cp ../../../martini_v2.0_ions.itp .
         cp ../../../martini_v2.2_aminoacids.itp .
         cp ../../../dipeptides/$pep".itp" .
         cp ../../../daft_sample.top tmp.top
         cp ../../../daft_sample.ndx .
         gmx insert-molecules -f daft.gro -ci ../../../dipeptides/$pep"-CG.pdb" -o daft_insert.gro -n daft.ndx -replace Solvent -nmol $npep -radius 0.26
         sed "s/NA+    NA+/W        W/" daft_insert.gro > daft_DPD.gro
         echo -e 'keep 0\nq\n' | gmx make_ndx -f daft_DPD.gro -o daft_DPD_system.ndx
         echo -e '!0&!1\nname 4 Solvent\nkeep 4\nq\n' | gmx make_ndx -f daft_DPD.gro -n daft_sample.ndx -o daft_DPD_sol.ndx
         cat daft_sample.ndx daft_DPD_sol.ndx daft_DPD_system.ndx > all.all
         rm *.ndx
         mv all.all daft.ndx
         martiniwaternum=$(grep -c "  W" daft_DPD.gro)
         sed -i '6 i\#include "'$pep'.itp"' tmp.top
         sed -i "\$aW  $martiniwaternum" tmp.top
         sed -i "\$a$pep       $npep"  tmp.top
         rm daft_insert.gro daft.gro daft.top
         mv tmp.top daft.top
         gmx grompp -f ../../../dipeptides/ref/water_min.mdp -p daft.top -c daft_DPD.gro -o ion -n daft.ndx
         echo -e "W\n" | gmx genion -s ion.tpr -p daft.top -nname CL- -pname NA+ -conc 0.1 -neutral -o daft.gro
         rm daft_DPD.gro mdout.mdp ion.tpr
```

```
        rm *#
        source martinate-run.sh
        cd ..
        done
      cd ../..
    done
  done
```

9.16. daft_sample.top for MEDI1912 Simulations

```
#include "martini.itp"
#include "martini_v2.2_aminoacids.itp"
#include "martini_v2.0_ions.itp"
#include "MEDI1912_C.itp"
#include "MEDI1912_D.itp"



[ system ]
DAFT setup for MEDI1912-MEDI1912

[ molecules ]
; name        number
MEDI1912_C  1
MEDI1912_D          1;
; name        number
MEDI1912_C  1
MEDI1912_D          1;
```

9.17. daft_sample.ndx for MEDI1912 Simulations

```
[ A_MEDI1912 ]
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87
88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110
111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130
131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190
```

191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230
231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250
251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270
271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290
291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310
311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330
331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350
351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370
371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390
391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410
411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430
431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470
471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490
491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507

[B_MEDI1912]

508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527
528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547
548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567
568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587
588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607
608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627
628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647
648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667
668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687
688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707
708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727
728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747
748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767
768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787
788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807
808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827
828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847
848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867
868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887
888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907
908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927
928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947

948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967
968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987
988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005
1006 1007 1008 1009 1010 1011 1012 1013 1014
[Solute]
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87
88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110
111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130
131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170
171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190
191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230
231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250
251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270
271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290
291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310
311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330
331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350
351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370
371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390
391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410
411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430
431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470
471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490
491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510
511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530
531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550
551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570
571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590
591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610
611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630
631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650
651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670
671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690
691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710

711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730
731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750
751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770
771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790
791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810
811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830
831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850
851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870
871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890
891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910
911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930
931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950
951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970
971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990
991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007
1008 1009 1010 1011 1012 1013 1014
[Membrane]

9.18. Bash Script to Set Up Motavizumab Simulations with Dipeptides

```
#!/bin/bash
declare -a aminoacid1=("ALA" "GLY" "HIS" "HSP" "LYS" "PRO" "GLN" "ARG"
"SER")
declare -a aminoacid2=("ALA" "GLY" "HIS" "HSP" "LYS" "PRO" "GLN" "ARG"
"SER")
hisnum=27
npep=10
for i in "${aminoacid1[@]}"
do
  for j in "${aminoacid2[@]}"
  do
      pep=$i"-"$j
      ./daft.sh -aa MOTA=Fab.pdb -c 2 -3d -gmxrc GMXRC -dssp dssp-2.0.4-linux-
amd64 -pymol pymol -n 256 --insane{-salt=0} --martinize{-ff=elnedyn22} --
martinate{-T=300} -d 3 -time 512 -ff elnedyn22 -sol W
      cd MOTA-MOTA
      mv daft $pep
      cd $pep
```

```
for d in */ ; do
    echo "$d"
    cd "$d"
    cp ../../../martini_v2.2.itp .
    mv martini_v2.2.itp martini.itp
    cp ../../../martini_v2.0_ions.itp .
    cp ../../../martini_v2.2_aminoacids.itp .
    cp ../../../martini_v2.0_sugars.itp .
    cp ../../../HEX.itp .
    cp ../../../daft_sample.top tmp.top
    cp ../../../daft_sample.ndx .
    cp ../../../dipeptides/$pep".itp" .
    sed -e "s/CL-    CL-/W        W/g" daft.gro > daft1.gro
    sed -e "s/NA+    NA+/W        W/g" daft1.gro > daft2.gro
    rm daft.gro daft1.gro
    mv daft2.gro daft.gro
    gmx insert-molecules -f daft.gro -ci ../../../dipeptides/$pep"-CG.pdb" -o
daft_insert.gro -n daft.ndx -replace Solvent -nmol $npep -radius 0.26
    echo -e "keep 13\nq\n" | gmx make_ndx -f daft_insert.gro -o daft_solv.ndx
    gmx insert-molecules -f daft_insert.gro -ci ../../../HEX.gro -o daft_HIS.gro -n
daft_solv.ndx -replace W -nmol $hisnum -radius 0.26 -try 10000
    echo -e "keep 0\nq\n" | gmx make_ndx -f daft_HIS.gro -o daft_HIS_system.ndx
    echo -e '!2\nname 6 Solvent\nkeep 6\nq\n' | gmx make_ndx -f daft_HIS.gro -o
daft_HIS_solvent.ndx -n daft.ndx
    cat daft_sample.ndx daft_HIS_solvent.ndx daft_HIS_system.ndx > all.all
    rm *.ndx
    mv all.all daft.ndx
    martiniwaternum=$(grep -c "  W" daft_HIS.gro)
    sed -i '7 i\#include "'$pep'.itp"' tmp.top
    sed -i "\$aW  $martiniwaternum" tmp.top
    sed -i "\$a$pep       $npep"  tmp.top
   sed -i "\$aHEX       $hisnum" tmp.top
    rm daft_insert.gro daft.gro daft.top
    mv tmp.top daft.top
    gmx grompp -f ../../../dipeptides/ref/water_min.mdp -p daft.top -c daft_HIS.gro -
o ion -n daft.ndx
    echo -e "W\n" | gmx genion -s ion.tpr -p daft.top -nname CL- -pname NA+ -
neutral -o daft.gro
    rm daft_HIS.gro mdout.mdp ion.tpr
```

```
        rm *#
        cd ..
        done
      cd ../..
    done
  done
```

9.19. daft_sample.top for Motavizumab Simulations

```
#include "martini.itp"
#include "martini_v2.2_aminoacids.itp"
#include "martini_v2.0_sugars.itp"
#include "martini_v2.0_ions.itp"
#include "MOTA_H.itp"
#include "MOTA_L.itp"
#include "HEX.itp"

[ system ]
DAFT setup for MOTA-MOTA

[ molecules ]
; name       number
MOTA_L       1
MOTA_H       1;
; name       number
MOTA_L       1
MOTA_H       1;
```

9.20. daft_sample.ndx for Motavizumab Simulations

```
[ A_MOTA ]
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88
89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111
112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131
132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151
152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171
172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191
```

192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211
212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231
232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251
252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271
272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291
292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311
312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331
332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351
352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371
372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391
392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411
412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431
432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451
452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471
472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491
492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511
512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531
532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551
552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571
572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591
592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611
612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631
632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651
652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671
672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691
692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711
712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731
732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751
752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771
772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791
792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811
812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831
832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851
852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871
872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891
892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911
912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931
932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950
[B_MOTA]

951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970
971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990
991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008
1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024
1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056
1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072
1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088
1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104
1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120
1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136
1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152
1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168
1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184
1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216
1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232
1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248
1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264
1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296
1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312
1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328
1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344
1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360
1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376
1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392
1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408
1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424
1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440
1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456
1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472
1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488
1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504
1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520
1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536
1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552
1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568
1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584

1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600
1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616
1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632
1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648
1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664
1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680
1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696
1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712
1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728
1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744
1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760
1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776
1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792
1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808
1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824
1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840
1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856
1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872
1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888
1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900

[Solute]

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88
89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111
112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131
132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151
152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171
172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191
192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211
212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231
232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251
252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271
272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291
292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311
312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331
332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351
352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371
372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391

392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411
412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431
432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451
452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471
472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491
492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511
512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531
532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551
552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571
572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591
592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611
612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631
632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651
652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671
672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691
692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711
712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731
732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751
752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771
772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791
792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811
812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831
832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851
852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871
872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891
892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911
912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931
932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951
952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971
972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991
992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008
1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024
1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056
1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072
1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088
1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104
1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120
1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136

1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152
1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168
1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184
1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216
1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232
1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248
1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264
1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296
1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312
1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328
1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344
1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360
1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376
1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392
1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408
1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424
1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440
1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456
1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472
1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488
1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504
1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520
1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536
1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552
1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568
1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584
1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600
1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616
1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632
1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648
1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664
1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680
1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696
1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712
1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728
1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744
1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760

1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776
1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792
1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808
1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824
1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840
1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856
1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872
1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888
1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900
[Membrane]