

TOWARDS THE RATIONAL DESIGN OF A TARGET-SPECIFIC ANTIBODY

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Sai Pooja Mahajan

February 2016

© 2016 Sai Pooja Mahajan

TOWARDS THE RATIONAL DESIGN OF A TARGET-SPECIFIC ANTIBODY

Sai Pooja Mahajan, Ph. D

Cornell University 2016

Antibodies and antibody-fragments have emerged as promising tools for many therapeutic and biotechnological applications. Antibody fragments (e.g., scFvs, Fabs, VHHs) derive functionality via their variable domains, which bind to a target (antigen) of interest. Antibody fragments obtained from conventional antibodies (i.e., human or mice IgGs) comprise two chains: variable heavy and variable light. Nanobodies (hereafter VHHs) are unique antibodies found in camelids. VHHs are the smallest naturally occurring binding domains and derive functionality via a single variable domain on a heavy chain. Only 3 hypervariable loops (H1, H2, H3) form the antigen-binding surface as opposed to 6 loops in conventional antibody fragments (3 from heavy and 3 from light chain). Due to their small size and surprising ability to bind a wide range of antigens with high specificity and affinity, VHHs are excellent candidates for antibody engineering. Despite their recent discovery, many engineered VHHs have already entered into clinical trials for treatment of a range of human diseases.

It is our aim to rationally engineer VHHs with specificity for a target antigen by tailoring the hypervariable loops. As a first step toward such a goal, the design of loops with a desired conformation was considered. As proof-of-concept and to build our understanding of the binding loops of VHH antibodies, the study focused on the H1 loop of the anti-human Chorionic Gonadotropin (hCG) llama VHH that exhibits a noncanonical conformation. This loop was redesigned to “tilt” the stability of the loop structure from a noncanonical conformation to a (humanized) type 1 canonical conformation by studying the effect of selected mutations to the amino acid sequence of the H1, H2, and proximal residues.

To test and extend our understanding of antigen-binding by VHHs, a dual modeling-experimental approach was pursued for designing a VHH specific to Alpha-Synuclein (AS). AS is the main pathological marker and perhaps the causative agent of Parkinson's disease. Starting from an immunized Camelid library against the Non-amyloid component (NAC) region of A53T mutant of AS (A53T), a bacteria-based selection technique was used to obtain a NAC-specific VHH, followed by computational modeling of the VHH and the VHH-NAC binding. The use of FLITRAP (an E. Coli based high-throughput screening technique) allows us to select for a soluble and intracellularly stable VHH (intrabody). Furthermore, using computational modeling the following tasks were completed: 1) Propose possible conformations of the VHH binding region, 2) Postulate viable modes of VHH binding to the NAC region, 3) Propose mutations that would enhance binding and ultimately, 4) Validate the proposed predictions through experiments. Counterintuitively, it was found that while mutations targeting the central hydrophobic NAC region only led to weak binding affinities, mutations, at the periphery of the binding site, that target the charged flanking hydrophilic residues of NAC are key to substantially increase binding affinity.

The main goal of this research was hence to demonstrate the possibility of developing a model of binding *in-silico* starting from the amino acid sequence of the Antibody and the antigen and using it to predict affinity-enhancing mutations. This work differs from many other structure-based design studies in that the crystal structure of neither the Antibody, the antigen, or the complex is known; it hence tackles a much more challenging (but common) situation. This work also differs from high-throughput screening techniques based on multiple rounds of screening to obtain a high-affinity binder. Our dual experimental-modeling approach can be considered as an important step towards developing rational design strategies based on *ab-initio* modeling and bottom-up design approaches, which would ultimately enable us to gain a deeper understanding of protein surfaces and interactions.

BIOGRAPHICAL SKETCH

Sai Pooja Mahajan was born and brought up in Chandigarh, India. She completed her Undergraduate and Masters' studies in Chemical Engineering from the Indian Institute of Technology Bombay (IITB), Mumbai, India in August 2008. She worked with Swasth India Pvt. Services from February 2009 to July 2009 as a project manager incharge of setting up a pilot microhealth insurance scheme in 10 villages in Latur, Maharashtra. In August 2009, she joined Cornell University to pursue a PhD in Chemical and Biomolecular Engineering. Over the last 6 years she has worked on multiple problems important in understanding the nature of protein surfaces and their ability to bind other proteins with remarkable affinity and specificity. She has also worked on problems related to phase-transitions in colloids and the problem of vapor nucleation in metastable liquids.

I dedicate my thesis to my Parents and my dear sister for their unwavering support and unconditional love; to my mentors for their confidence in me despite numerous setbacks and failures and to my dear friends whose kind words and love kept me going through the many cold winters and sunless, chilly springs of Ithaca.

ACKNOWLEDGMENTS

I would like to acknowledge my mentors Dr. Fernando Escobedo, Dr. Matthew Delisa and Dr. Roger Loring for their support and guidance. Dr. Escobedo is one of the best teachers I have had the opportunity of learning from. His class, *Advanced Thermodynamics*, sparked a sense of wonder and excitement to delve deeper into the many mysteries of thermodynamics. I discovered the joy of deriving *fancy Fs* and *Gs* and learning about the ever-so-mysterious driving force of entropy and the fun of playing with biased die. Professor Delisa's vision is truly inspiring. He amazes me with his incredible knowledge of research literature and ability to sift the grain from the chaff, pinpointing the most important research questions, ever-expanding the impact of his research by exploring new frontiers.

Over the last 6 years, I have been fortunate to meet and interact with some of the most wonderful and inspiring people at Cornell. They have elevated my graduate school experience by their unique personalities and their passion and smarts. Poornima has been a good friend and a constant source of support. Her presence can brighten the gloomiest of days and the darkest corners of Olin Hall. Vikram's dedication to his work is awe-inspiring. Beth will always be a role-model for her superhuman ability to live her life to the fullest – a scientist, a provider, a mother, a wife, an insane ultra-marathon runner and a wonderful friend and source of inspiration. I couldn't have survived a day in the lab without the guidance and support of DD, Jason, Alyse, Dr. Lee, MP, Tommy, May, Aravind, Pao and Xiaolu. The grueling and many-a-times frustrating experience of working in the lab was brightened with laughter and camaraderie by their presence. Tommy's childlike optimism and enthusiasm, MP's dedication and perseverance, May's focus and positive attitude, Xiaolu's superwoman-ness and almost-always true refrain "it-works" and Aravind's displays of

loquaciousness and repurposing poetry to explain mundane events in the lab will never be forgotten!

I would also like to acknowledge Sushmit, Bryan and Jonathan for being wonderful colleagues and invaluable sounding boards. Priyamvada and Taomo have been mentors, confidants and friends. Their inspiring presence over the last couple of years and advice on matters of research and life has been very valuable.

I would like to acknowledge computational grant from Xsede - TG-MCB130059 and funding from NSF grant CBET-0933092.

TABLE OF CONTENTS

1	Introduction	1
2	Conformational equilibria of an anti-hcg llama nanobody and its mutants	6
2.1	Introduction	6
2.2	Methods	9
	Model for vHH	9
	Mutagenesis Strategy	10
	Thermodynamics and Free Energy Simulations: Metadynamics	12
	Collective Variables	13
	Details of the BE Simulations	16
	Calculation of percentage of conformations belonging to different clusters	17
	Kinetics: Kinetic Monte Carlo	17
	Validation of Force field	20
2.3	Results and Discussion	21
	Overview	21
	Wildtype (Wt)	23
	2.3.1.1 Mutant Fa	25
	2.3.1.2 Mutant FF	26
	Mutants FFSa, FFD	27
	Mutants FFPA and FFTL	30
	Pseudo-Kinetics Results	35
	Role of key-residues and proximal residues	39
2.4	Conclusions	41
3	Design of an anti-alpha synuclein VHH: From selection to model development	45
3.1	Introduction	45
3.2	Methods Experiments	46
	Alpaca Immunization	46
	Bacterial Strains, Plasmids and Growth Conditions	47
	Cloning and Plasmid construction	47
	Selective Growth Assays	48
	Subcellular fractionation and Western Blot analysis	48
	Enzyme-linked immunosorbent assay (ELISA)	49
3.3	Methods Simulations	50
	Model for vHH	50
	Model for AS-NAC region	51
3.4	Results and Discussion	59
	Selection of a vHH Antibody against A53T NAC peptide	59
	Developing an atomistic model for vhh 4C and its complex with NAC peptide	61
4	Improving affinity of binding using rational design	64
4.1	Introduction	64
4.2	Methods Simulations	64
	Umbrella Sampling to Obtain Potential of Mean Force	64

PBEQ	66
Mutagenesis Strategy	75
4.3 Methods Experiments	77
Expression, Purification and Size Exclusion Chromatography (SEC)	77
Binding rates and dissociation constant determination using SPR	77
4.4 Results	78
Utilizing hydrophilic residues in the NAC peptide to enhance energy of binding:	
Targeting electrostatic interactions at the binding interface	78
Predicting affinity-enhancing mutations by improving electrostatic	
complementarity of the binding interface for Model III	83
Testing computational predictions in experiments: Long-range electrostatic	
interactions in vhh-NAC binding are important.....	84
Predicting affinity-enhancing mutations by improving electrostatic	
complementarity of the binding interface for Model V	84
5 Conclusions and future work.....	87
5.1 Summary and Conclusions	87
5.2 Future Work	91
Appendix A: Sequence of anti-AS VHHs.....	92
References.....	93

LIST OF FIGURES

<p>Figure 2.1 Plausible steps in the transition from the F (floppy) -> I (intermediate) -> C (canonical type-1); Sidechains F27, F29 and Y32 (from left to right) are shown to illustrate this transition in two arbitrarily delineated steps; F → I describes step 1, and I → C describes step 2</p>	14
<p>Figure 2.2 2-D Free Eenergy profiles for CAB-CA05 VHH. Black circles mark basins corresponding to the type-1 canonical conformation</p>	21
<p>Figure 2.3 Summary of FE study of a) Wt, b) Fa and c) FF. Left: 2-dimensional FE profiles in the 2 most important CVs for the H1 loop of the Wt and mutants; flags mark the center of the stable conformations. Center: Iso-surfaces of minimum FE clusters in 3 CVs; color of each isosurface corresponds to a particular set of conformations. The colors of flags (in the 2-dimensional profiles), isosurfaces (in the 3-dimensional profiles) and the conformations agree. Canonical basin/conformation, if observed, is shown in red.....</p>	23
<p>Figure 2.4 Summary of FE study of d) FFSa, e) FFD and f) FFPA. Left: 2-dimensional FE profiles in the 2 most important CVs for the H1 loop of the Wt and mutants; Flags mark the center of the stable conformations. Center: Isosurfaces of minimum FE clusters in 3 CVs; color of each isosurface corresponds to a particular set of conformations. The colors of flags (in the 2-dimensional profiles), isosurfaces (in the 3-dimensional profiles) and the conformations agree. See axes and structure guide in Figure 2 for more details. The canonical conformation, if observed, is shown in red. Isosurface and conformation shown in FFPA mutant are the closest to those of the type-1 canonical structure.....</p>	27
<p>Figure 2.5 2-D FE profiles for Wt and mutants in CVs d_1, ψ_{S30}, ψ_{T31}.....</p>	32
<p>Figure 2.6 2-D FE profiles for mutants FFSa, FFD, FFPA and FFTL in CVs d_1, ψ_{S30}, d_3 and d_4.....</p>	33

Figure 2.7 H1 type-1 canonical conformations obtained from BE Metadynamics for a) FF b) FFSa c) FFD d) FFPA mutants. The H1 loop is colored black, and residues marked in green (carbon), blue (nitrogen) and red (oxygen) correspond to important proximal residues within loops H2 and H4. The text in red highlights mutation sites.....	34
Figure 2.8 RMSD from an H1 type-1 canonical conformation of a reference structure (PDB code 1DFB), for the type-1 canonical basins of FF, FFSa, FFD, FFPA and FFTL.....	34
Figure 2.9 Relevant transition pathways proposed for the F ->C transition of the FFSa mutant. Also reported, is the percentage of pathways visiting basins I4/I1/I2 before reaching basin C. Color scheme same as the one used in Figure 4d.....	39
Figure 2.10 Evidence of conformational isomers of the type-1 canonical conformation in experimentally characterized antibody SPE7 juxtaposed with I4 conformation of the FFSa mutant. <i>Left</i> : Conformational isomer (I, from PDB structure 1OCW) for antibody SPE7. <i>Right</i> : Conformational isomer (I4) for mutant FFSa	41
Figure 3.1 Workflow for generation and rational design of anti-AS vHH	46
Figure 3.2 Sequence of AS and peptides used for experiments and modeling	52
Figure 3.3 Selection of candidate vHH, expression and binding A) Selective plating of serially diluted cells expressing different antigen-antibody pairs on Carb as indicated. Each spot corresponds to 5uL of serialy diluted cells. B) Western blot analysis of periplasmic and cytoplasmic fractions generated from WT cells carrying pSALect and expressing ssTorA-vHH-FLAG. Detection of vHH constructs was with anti-FLAG antibody, and detection of GroEL using anti-GroEL antibody and served as a fractionation marker. C. ELISA binding assay shows vHH binds A53T.	53
Figure 3.4 Model for vHH 4C. A. Homology Model for vHH 4C based on PDB: 3P0G. B. Enhanced conformational sampling in REM simulations C. Top two conformations obtained from REM simulations (N and ST).	54

Figure 3.5 Hydrophobic residues clustered at the interface of H1 and H3 loops for vHH 4C N and ST models. The Solvent Accessible Surface (probe radius 1.4 Å) and the Cartoon model are shown. Only the hydrophobic residues are shown as Licorice with element C in green, H in white, N in blue and O in red.	55
Figure 3.6 Probability of occurrence of different complexes in the lowest temperature replica in REMD Simulations for the vHH4C exhibiting conformation N and Alpha-Helical Ag (N-AH).	57
Figure 3.7 Probability of occurrence of different complexes in the lowest temperature replica in REMD Simulations for the vHH4C exhibiting conformation N and Alpha-Helical Ag (N-AH) for “newsite”.	57
Figure 3.8 Probability of occurrence of different complexes in the lowest temperature replica in REMD Simulations for the vHH4C exhibiting conformation ST and Alpha-Helical Ag (ST-AH).....	58
Figure 3.9 Probability of occurrence of different complexes in the lowest temperature replica in REMD Simulations for the vHH4C exhibiting conformation ST and Random Coil Ag (ST-RC).	58
Figure 3.10 Probability of occurrence of different complexes in the lowest temperature replica in REMD Simulations for the vHH4C exhibiting conformation N and Random Coil Ag (N-RC).	59
Figure 3.11 Top 4 complexes obtained from Simulations for the vHH-NAC complex. Complexes I and II were obtained for a 19-residue peptide (AAs 61-78) and Complexes III and IV were obtained for a 13-residue peptide (AAs 66-78). The vHH is shown in metallic Tan and the NAC peptide is shown in Cyan.	62
Figure 3.12 Energy of interaction between vHH-Ag for the 4 models and the reference Ab-Ag case.	63
Figure 4.1 Schematic for Umbrella Sampling (US) scheme to obtain Potential of Mean Force (PMF)	66

Figure 4.2 US Results for Model III for the 13-residue (hydrophobic only) and 24-residue peptides for the Ag.....	78
Figure 4.3 Electrostatic Potential of Binding Surfaces for Model III A) Complex of vHH 4C with Ag. The Complex and the Ag are shown. In the complex, the vHH is shown with a solid surface and the Ag is shown with a wire-mesh. In the Ag, the surface is shown as a solid surface. Blue indicates an electrostatic potential of 1 kT/e and red indicates an electrostatic potential of -1 kT/e. B) The electrostatic FE of binding calculated using PBEQ for vHH 4C and mutants AT and E. C) Electrostatic Potential of vHH 4C and mutants AT and E.....	80
Figure 4.4 Putative models to explain SPR results for K75AM77T and K75E mutants	80
Figure 4.5 Electrostatic Potential of Binding Surfaces for Model V A) Complex of vHH 4C with Ag. The Complex and the Ag are shown. In the complex, the vHH is shown with a solid surface and the Ag is shown with a wire-mesh. In the Ag, the surface is shown as a solid surface. Blue indicates an electrostatic potential of 1 kT/e and red indicates an electrostatic potential of -1 kT/e. B) The electrostatic FE of binding calculated using PBEQ for vHH 4C and mutants AT, E and N76D. C) Electrostatic Potential of vHH 4C and mutants AT, E and N76D.....	81
Figure 4.6 The electrostatic FE of binding calculated using PBEQ for vHH 4C and mutants AT, E and N76D for Model III, Model V and Model II.....	81
Figure 4.7 US Results for Model III (Top) and Model V (Bottom)	82
Figure 4.8 Coarse-Grained Extrapolation (Right) of Model V (Left). In the complex, the vHH is shown with a solid surface and the Ag is shown with a wire-mesh. The N and C terminal residues of A53T or AS not included in the computational model are shown as Black Lines to mark their location.	83

LIST OF TABLES

Table 2.1 Mutations for some of the cases previously simulated in implicit solvent ¹ and two new 4-point mutants	8
Table 2.2 3-D basins for FF mutant	36
Table 2.3 3-D basins for FFSa mutant	37
Table 2.4 3-D basins for FFD mutant	37
Table 2.5 MFPTs for transitions to and from the H1 type-1 canonical basin for the FF, FFSa and FFD mutants. Transition time reported in MC steps.	38
Table 4.1 SPR results for vHH 4C and mutants	82

1 INTRODUCTION

Antibodies are gamma globulin proteins found in the blood and other bodily fluids and are used by the immune system to identify and neutralize foreign objects or antigens. Antibodies and their recombinant fragments – scFv, Fab, and engineered variants - diabodies, triabodies, minibodies and others – have emerged as powerful tools in biosensing, bioimaging and targeting specific biomolecules in treatment of cancer², infectious and inflammatory diseases³.

Conventional antibodies typically comprise two large heavy chains (VH) and two small light chains (VL). Each of these chains contains an antigen-binding domain with high sequence variability, which consists of three heavy (H1, H2, H3) and three light (L1, L2, L3) hypervariable loops. In camelids (e.g., llamas, camels), a subset of their naturally occurring antibodies is devoid of the light chain. The variable heavy domains of these antibodies are called VHHs or nanobodies due to their small size. The nanobodies' antigen binding regions are thus formed solely by the 3 hypervariable loops H1, H2 and H3, hence eliminating the need for pairing of the heavy and light domains as in the case of conventional antibodies. We note that the Kabat scheme⁴ has been used for numbering the positions of the amino-acids throughout this document.

Nanobodies can be efficiently expressed as soluble non-aggregating recombinant proteins in bacteria⁵, and other microorganisms⁶. Nanobodies can be regarded as evolution-engineered single-domain binding agents with several appealing properties such as high solubility, facile genetic manipulation, recognition of hidden-sites, clefts and mimicry of other biomolecules, all of which are either difficult to achieve in isolated VH domains or even with scFvs and Fab, or haven't been observed in conventional antibodies. Moreover, the 80% sequence similarity

between the human VH domain and the VHH makes the latter an excellent candidate for application to human therapeutics. However, a more thorough analysis of the properties of the antigen-binding region of the nanobody is required to understand: i) the significance of its structural deviations from the conventional VH domain, and ii) how a single domain achieves immunogenicity comparable to that observed in conventional antibodies⁷. Understanding the latter question(s) will be a step towards engineering soluble single-domain binding reagents with high specificity and affinity. Extensive studies^{8,9,10,11} on the human and murine VH domain suggest that the repertoire of conformations of the hypervariable H1 and H2 loops of the VH domain is highly conserved and is limited to a few conformations. These studies encompassed both VH germline and expressed genes after somatic mutations¹². Furthermore, Barre *et. al.*¹³ showed that cartilaginous fish, which is the most distantly related species to human that still has an immune system, also exhibits similar canonical structures and key residues for the H1 and H2 loops of the VH domain. This implies that canonical hypervariable loop structures arose very early in the stages of evolution of the immune system and have since been highly conserved. Three key questions then follow: Would VHH domains with a non-congruent set of H1 and H2 conformational repertoire from the VH domain be suitable candidates for human therapeutics? Is it possible to “humanize” the H1 and H2 loops of the VHH domain? How will such a process affect specificity and affinity of these domains? In Chapter 2, we specifically address the question of “humanizing” the H1 loop of a VHH with a non-standard H1 loop conformation.

Antibodies are considered a paradigm for high-affinity, target-specific binding. Hence, the next design question was whether it is possible to use computational modeling to develop a model for binding between an Antibody and an Antigen and use it to enhance binding. In

Chapters 3 and 4, we address the problem of rational design of a target-specific antibody in the vHH format.

Protein-protein interactions are at the core of almost all cellular processes including signal transduction, cell division, DNA replication, biosynthesis, and degradation. Hence, the ability to modulate protein-protein interactions is of great interest to both fundamental science and applied research such as drug design. However, our understanding of these interactions is limited. Bottom-up, rational design of target-specific protein surfaces will be a stringent test of our understanding of the biophysical phenomena that govern these interactions. While many studies have demonstrated the ability to use structure-based information to rationally improve protein-protein interactions, the design of target-specific binders is largely an unsolved problem. Previously, multiple groups have demonstrated the use of rational design to enhance protein-protein interactions with variable success by 1) Improving electrostatic complementarity¹⁴⁻¹⁶ and/or 2) Optimizing packing/contacts between the surfaces^{17,18} Recently, Tinberg *et. al.*¹⁹ have demonstrated the ability to design high affinity binders against a steroid molecule using a pre-organized framework. In this approach, a scaffold protein is used as the framework in which target-specific interactions are introduced in the binding pocket. While their approach is quite successful in targeting small molecule like steroids, which bind to convex, pocket-like surfaces, it may not be generally applicable to target diverse protein surfaces. The problem of designing target-specific binders becomes especially challenging in the absence of high-resolution 3D structures. This is especially true in the early stages of evolution of target-specific binders against weakly immunogenic targets wherein many hits with low affinities are obtained from standard selection/screening techniques. At this stage, it may not be feasible to obtain structures of these

binders in complex with the target. The ability to develop in-silico models of binding, to evolve these binders towards higher affinities, could be an invaluable tool at this stage.

Alpha-Synuclein (AS) is a pathological marker of Parkinson's disease (PD) and related disorders²⁰. It is natively disordered, with many conformational forms upon interacting. Misfolding and subsequent aggregation of AS and other proteins into soluble and insoluble oligomers and high molecular weight fibrils has been implicated in PD and other neurodegenerative diseases. Prevention of aggregation and misfolding of such aggregation-prone species has been suggested as a viable therapeutic strategy for reducing the pathogenesis in various neurodegenerative diseases. The central hydrophobic domain (AAs 61-95) is termed as the nonamyloid component (NAC)²¹. The NAC domain is aggregation-prone and can form beta-sheets by self-association²². A transgenic *Drosophila* model expressing wild type AS without the NAC domain was shown to significantly reduce toxicity, making it an attractive target for treatment such as recombinant antibody technology^{21,23,24}. The NAC region has been previously targeted in other studies for inhibiting the aggregation and/or reducing toxicity of alpha-synuclein. Various methods such as yeast surface display (scFv NAC32 and VH14)²³, phage display^{25,26} and, more recently, rational design approaches^{27,28} have been used for generating antibody fragments for this purpose. In particular, the rational design approaches (as opposed to high-throughput screening) have focused on designing NAC-specific peptides and grafting them into the CDR regions of a variable heavy domain (vHH) framework.

In Chapter 3, we discuss the selection of a candidate vHH from an Immunized library using a *E. Coli* based selection technique, followed by experimental characterization and in-silico model for the candidate vHH and Antigen. In Chapter 4, we discuss the development of a model for the complex followed the prediction of affinity enhancing mutations using various strategies.

Experimental binding assays are used to test our predictions and improve the model. A combination of advanced Molecular Dynamics (MD) techniques, numerical techniques and experimental techniques were employed to build a coherent picture of the vhh-NAC binding.

The main goal of this work is to demonstrate the plausibility of developing a model of binding in-silico starting from the amino acid sequence of the Antibody and the antigen and using it to predict affinity-enhancing mutations. Moreover, we want to demonstrate the use of detailed advanced molecular techniques (as opposed to knowledge based approaches or techniques employing minimum energy models) with atomistic-detail to make experimentally relevant predictions. This work differs from many other structure-based design studies where the crystal structure of the Antibody and/or antigen and/or complex is known *a priori*. This work also differs from high-throughput screening techniques based on multiple rounds of screening to obtain a high-affinity binder. Our dual approach can be considered as a small step towards developing rational design strategies based on *ab-initio* modeling and bottom-up design approaches which would ultimately enable us to gain a deeper understanding of protein surfaces and interactions.

In Chapter 5, we present a summary of the work and present ideas to guide future work in the field of rational design.

2 CONFORMATIONAL EQUILIBRIA OF AN ANTI-HCG LLAMA NANOBODY AND ITS MUTANTS

2.1 Introduction

In this study, we focus on the H1 loop of the anti-hCG llama nanobody that exhibits a noncanonical conformation. We aim to “tilt” the stability of the H1 loop structure from a noncanonical conformation to a (humanized) type 1 canonical conformation by studying the effect of selected mutations to the amino acid sequence of the H1, H2, and proximal residues.

The llama VHH raised against the alpha-subunit of the human chorionic gonadotropin (hCG) has been characterized by Renisio *et. al.*²⁹ via NMR spectroscopy (PDB code 1G9E and referred to as VHH-H14) and by Spinelli *et al.*³⁰ via X-ray crystallography. A hybrid Monte Carlo Replica Exchange (HYMREX)³¹ was used to simulate the loop conformations of this system providing further evidence for conformational flexibility of the loops observed in the NMR study. Several studies have also been conducted to determine key-residues determining loop conformations for this system³⁰⁻³³. A novel *in-silico* screening/mutagenesis analysis was performed using implicit solvent¹, in which mutants with 1, 2 and 3 mutations in the H1 and H2 loops of the llama VHH domain were suggested as good candidates to recover the type-I canonical conformation observed in conventional VH domains. That study further showed that two mutants with 3 mutations formed a stable type-1 canonical H1 loop, and that alternate non-canonical H1 conformations may also be stabilized. In the present work, we study the behavior of the wildtype and mutant loops in explicit solvent, to not only validate the implicit solvent results

but also provide evidence of the presence of “intrinsic disorder”³⁴ in the conformational landscape of the H1 loop.

We take advantage of the combination of a small system size and the power of a biasing molecular dynamics technique – metadynamics³⁵⁻³⁷ – to study in detail the conformational isomerism of the hypervariable loops of the VHH system using all-atom, explicit solvent simulations. We find that the simulated free-energy profiles of the H1 loop reveal intrinsic conformational disorder. We place such finding in the context of many experimental studies which have demonstrated the conformational flexibility and isomerism of the hypervariable loops of the antibody variable domains³⁸⁻⁴³.

The rest of the chapter is organized as follows. First the mutants selected for study are reviewed and the simulation details described. The results are then presented for the VHH loops in the folded framework, followed by a discussion of the free-energy landscapes of various mutants and implications of the conformational variability of the H1 loop on binding. The paper ends with remarks on the significance of these results for antibody engineering.

VHH llama loop conformations have previously been reviewed in detail^{33,44}. Briefly, VHH hypervariable loops exhibit a wider diversity in length and conformations than the conventional VH loops found in humans and mice. It has been hypothesized that this diversity may perhaps compensate for the lack of the light variable domain. It has been observed that affinity improvement of a VHH domain involves mutations in residues in all three hypervariable regions⁴⁵. The H1 and H2 conformations exhibited by VHHs that cannot be categorized as per the canonical types^{10,12,13} are referred to as non-canonical conformations. The llama VHH has a type-2A/3 canonical conformation for the H2 loop and a non-canonical conformation for the H1 loop,

in the crystal and NMR structures. The length of the llama VHH H1 loop corresponds to that of the type-1 canonical conformation in conventional antibodies; however, this loop lacks “typical” residues in key-positions (27, 29) found in a type-1 canonical H1 loop conformation. As described in the next section, mutations at key positions in the H1 and other loops were introduced in a previous molecular simulations study¹, in an attempt to “recover” the type-1 canonical conformation of the H1 loop.

Table 2.1 Mutations for some of the cases previously simulated in implicit solvent¹ and two new 4-point mutants

Case	Mutations
Wildtype (wt)	---
1-Fa	R27F
2-FF	R27F, G29F
3-FFSa	R27F, G29F, W52aS
3-FFD	R27F, G29F, W52aD
4-FFPA	R27F, G29F, W52aP, R71A
4-FFTL	R27F, G29F, W52aT, R71L

2.2 Methods

Model for vHH

A reduced model of the crystal (1HCV), which has been defined previously¹, was used for computational economy. Briefly, 57 out of 117 amino acids of the VHH were included in our simulations. These residues comprise the hypervariable loops (H1, H2, and H3) along with proximal residues which may be important in determining and stabilizing the loop conformations; *i.e.*, distal framework residues which are unlikely to have influence on the structure of the hypervariable loops were excluded. The backbone atoms of the residues in H4 (flanking positions 71-77) are not restrained in the present study (in contrast to previous conformational and mutagenesis studies from our group^{31 1}). The C_α, N and C backbone atoms that do not belong to the hypervariable loops were restrained (except H4) with a harmonic force constant of 2000 kJ/nm, while backbone atoms in H1, H2, H3 and H4 along with side-chains from all residues were free to move. These restraints were imposed to simulate the limited motion of these residues in the folded domain. While such an approach may not provide the most accurate description of the folding dynamics of the H1 loop during folding of the VHH domain *in vivo*, it captures 1) the stability of the type-1 canonical conformation with respect to other possible H1 loop conformations, 2) the relative ease with which such a conformation is subject to rearrangement in the VHH domain for a mutant and, 3) evidence of the potential role of the H1 loop conformation and proximal residues in antigen recognition and binding. In all simulations, the N and C terminals of all simulated fragments were capped by acetyl and N-methyl groups respectively.

As in the previous study¹, this reduced model of 1HCV and the force field adopted (detailed below) were validated by conducting molecular dynamics and Bias Exchange

metadynamics simulations^{36,3746-4849,50} for the VHH Cab-CA05 domain (PDB code 1F2X)^{44,51} in explicit solvent. From these validation runs we confirmed that, in agreement with experiments, its H1 loop exhibits the type-1 canonical conformation. Additional details of this analysis are given in the SI section.

All simulations employed the GROMACS^{52,53} software using the CHARMM27^{54,55} force field and the TIP3P model for water molecules. We run metadynamics simulations using the plugin PLUMED⁵⁶. The starting structures for all mutants were generated by introducing the corresponding mutations in the reduced 1HCV crystal, followed by energy minimization, solvation and equilibration steps. All simulations were conducted in the NVT ensemble at T=300 K. The leap-frog algorithm was used as the integrator with a time-step of 2 fs. The LINCS and SETTLE algorithms in GROMACS were used to constrain the length of all non-water⁵⁷ and water bonds, respectively. Electrostatics were treated with particle-mesh Ewald (PME)⁵⁸, using a short-range cutoff of 1.2 nm, and the van der Waals interactions were switched off between 1.0 and 1.2 nm.

Mutagenesis Strategy

In this work we revisit some of the mutants previously studied in implicit solvent, and suggest two new mutants to further illustrate the role of key proximal residues to the H1 loop.

In the implicit solvent study, the residues at positions 27, 29 and 52a were chosen as key mutation sites for several reasons. Briefly, an analysis of H1 loop sequences of conventional antibodies initially indicated that the residues at positions 27 (R) and 29 (G) in the wildtype llama VHH were rarely observed in H1 loops of VHs. Incidentally, mutations in these two sites have

been proposed as a way to achieve augmented affinity in VHHs⁵⁹. Similarly, residue 52a in H2 was reported¹ as being potentially important for preventing the formation of a type-1 H1 conformation, due to possible clashing with the side-chain of a buried residue at position 29. The substitute residues for sites 27 and 29 were chosen therein¹ on the basis of frequent occurrences (Y/F/T at position 27 and F/I/L at position 29) in the canonical type-1 H1 loops^{8,1012}. Residue W52a was mutated to S/D in an attempt to decrease the steric hindrance caused by the bulky W side-chain at this position, by replacing this amino acid with ones having a smaller side-chain.

The implicit solvent analysis showed that mutants with F substitutions at positions 27 and 29 exhibited stability and higher propensity for the type-1 canonical conformation, and that the W52aS and W52aD mutations further increased the rigidity of this H1 canonical type. Consequently, these successful mutations for residues 27, 29 and 52a were also studied in this work and are summarized in Table 1. We further study two new mutants with 4 mutations – 2 mutations in H1 at positions 27 (R27F) and 29 (G29F), 1 mutation in H2 at position 52a and 1 mutation at position 71 (see Table 1). We chose position 71 as mutation site mainly because in preliminary simulations with mutant FFD we found that the R71 side-chain formed transient hydrogen bonds with the D52a and K76, and we wanted to test whether a charged-to-neutral substitution at position 71 would lead to a faster non-canonical to canonical transition by eliminating kinetic traps due to these hydrogen bonds. Moreover, it has been previously suggested that the size of the residue at position 71 is a major determinant of the spatial location and structure of the H2 loop of conventional VHs⁶⁰. Based on these two precedents, we selected residues A and L, which have small hydrophobic side-chains, as substituents of R71 for the two 4-mutants in Table 1. In general, residues 71-77 are believed to combine with H1, H2 and H3 to form a larger paratope^{61,62}. This is further supported by affinity maturation studies which show

mutations in this region^{63,64}. In a recent study³³ which analyzed 27 VHH domains, it was found that these residues have a significantly higher structural divergence compared to VH loops. We will refer to these residues as the “H4” loop. We note that while the role of the H4 loop in antigen-binding may be less significant than that of loops H1, H2 and H3, previous studies and our preliminary simulations suggested that residue 71 on this loop can influence the structure of the H1 and H2 loops. Our choice of substituents for site 52a for the two 4-mutants in Table 1 stemmed from work by Chothia *et. al.*¹⁰, who studied the protein sequences of 83 VH open reading frames (ORFs) and found that the most frequent combinations of residues in VHs with H1-type-1 and H2-type-2A loops are P/T/A and A/L/T, at positions 52a and 71, respectively.. Hence, the combinations P-A (see for instance antibody with PDB code 1MFA) and T-L (see for instance nanobody Cab-CA05, PDB code 1F2X) were chosen for the 52a-71 pair, to determine if a generally conserved combination would lead to an increased stability of the final canonical conformation (if exhibited) of the H1 loop.

Thermodynamics and Free Energy Simulations: Metadynamics

In BE metadynamics, multiple molecular dynamics (MD) simulations (replicas) are run simultaneously, each biased with the metadynamics potential in one or two collective variables. Every few MD steps, a Monte-Carlo (MC) exchange-move is attempted based on:

$$\min\left\{1, \exp\left[\frac{1}{T}\left(V^a_G(x^a, t) + V^b_G(x^b, t) - V^a_G(x^b, t) - V^b_G(x^a, t)\right)\right]\right\} \quad 2.1$$

where x^a and x^b are the coordinates of replica a and b and $V^r_G(x,t)$ is the metadynamics potential acting on replica $r(a/b)$. This treatment allows for an efficient conformational search in a high-dimensional collective variable space without the requirement of biasing each simulation in more than two collective variables. The MC exchange step enhances conformational search in a fashion similar to that in REM. From each simulation, one obtains low-dimensional projections of the free-energy which can then be used to obtain an estimate of the higher-dimension free-energy in CVs of interest. The low-dimensional free-energy projections after a sufficiently long time have been shown to fluctuate around an average which estimates the FE of the system in the chosen CV space. However, this is only true when the choice of CVs ensures that all “slow” degrees of freedom that may be important to the conformational transitions in the system of interest have been sampled.

Collective Variables

The set of collective variables (CVs) used in our study is very different from the global collective variables that have been used in previous bias-exchange metadynamics studies for protein folding. We found that a combination of global and local parameters is required to provide a complete description of the loop mobility and conformational landscape. The transition, starting from the non-canonical conformation, for mutants with mutations at position 27 and position 29, displays a particularly rugged conformational landscape and the collective variables defined below are able to describe kinetic traps related to the conformation of the H1 loop. It is instructive to visualize the overall transition from the non-canonical to canonical conformation to understand the collective variables used. The transition can be broadly described to consist of two essential structural changes, in no particular order (see Figure 1):

1) Floppy \rightarrow Intermediate (F \rightarrow I): The burial of the hydrophobic residue at position 29. This involves the formation of a 3-residue 3-10 helix-like structure in which the ϕ dihedrals become progressively negative and the ψ dihedrals become progressively positive at positions 29, 30 and 31. The non-canonical (ϕ , ψ) combinations for residues 29, 30 and 31, in degrees, are (-66,-23), (-88,-1), (-101,148) and those for the canonical type-1 conformation are (-54,-40), (-71,-9), (-98,-6), e.g., the canonical dihedrals corresponding to a 3-residue 3-10 helix.

2) Intermediate \rightarrow Type-1 canonical (I \rightarrow C): The collapse of the phenylalanine side-chain at position 27, to interact with the aromatic ring of Y at position 32 forming a compact type-1 canonical conformation of the H1 loop.

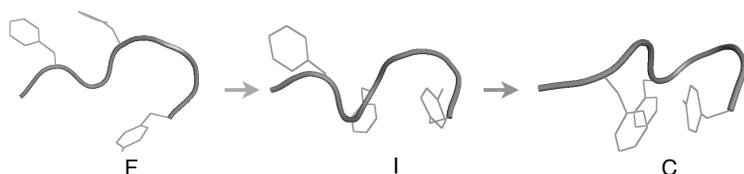


Figure 2.1 Plausible steps in the transition from the F (floppy) \rightarrow I (intermediate) \rightarrow C (canonical type-1); Sidechains F27, F29 and Y32 (from left to right) are shown to illustrate this transition in two arbitrarily delineated steps; F \rightarrow I describes step 1, and I \rightarrow C describes step 2

Keeping these two transitions in mind, we used the following set of collective variables for our simulations:

- 1) d_1 – The distance of the residue at position 29 from the hydrophobic core, measured as the distance between the center of mass (COM) of the heavy atoms in residue 29, and the COM of the backbone atoms in residues I69 and T77.
- 2) ψ_{S30} – The C_α -C backbone dihedral of residue at position 30 (S).

3) ψ_{T31} – The C_{α} -C backbone dihedral of residue at position 31 (T).

CV d_1 captures the $F \rightarrow I$ transition. The type-1 canonical conformation with a kink at position 29 and a buried hydrophobic side-chain is characterized by a small value of d_1 . The CVs ψ_{S30} and ψ_{T31} are part of the 3-residue 3-10 helix (at positions 29, 30, 31) in a type-1 canonical conformation and have characteristic negative values in this state. CVs 1, 2 and 3 can completely define the burial of a hydrophobic phenylalanine residue at position 29.

4) d_2 – The distance between the side-chain at position 32 (Y) and the side-chain at position 27 (R/F). This captures changes in step $I \rightarrow C$ (see Figure 1).

5) d_3 – The distance between the COM of the side-chains of the residue at position 52a on the H2 loop and that of residue at position 76 (K) on the H4 loop. This describes a local feature that prevents the hydrophobic side-chain of residue 29 from packing into the hydrophobic core, and captures the potential influence of loops H2 and H4 on the burial of the side-chain at position 29.

6) d_4 – The distance between the COM of the side-chain at position 71 (R/A/L) and the COM of the backbone atoms of residue 34. It was observed in preliminary simulations (run without this CV) that the side-chain at position 71 frequently inhabits the place that the side-chain of residue 29 occupies in the type-1 canonical conformation. This is the second CV that captures the dependence of the burial of the F29 side-chain on local features of the protein – in this case – the residues in the H4 loop.

Details of the BE Simulations

For bias-exchange (BE) metadynamics, we used 6 replicas for mutants FFSa, FFD and FFTL, 1 unbiased and the other 5 biased over different CV pairs: d_1 and ψ_{S30} , d_2 and ψ_{S30} , d_1 and d_3 , d_1 and d_4 , and ψ_{S30} and ψ_{T31} , respectively. For mutant FFPA, we used 5 replicas, 1 unbiased and the other 4 biased in d_1 and ψ_{S30} , d_2 and ψ_{S30} , d_1 and d_3 , and ψ_{S30} and ψ_{T31} , respectively. The d_1 and d_4 biased replica was not used for this mutant because the A71 side-chain (unlike R and L) is small and we do not expect its distance from the H1 loop (d_4) to be a crucial factor in determining the F29 side-chain positioning for the type-1 canonical conformation. For the wildtype and the Fa mutant, we used 4 replicas, 1 unbiased and the other 3 biased in d_1 and ψ_{S30} , d_2 and ψ_{S30} , and ψ_{S30} and ψ_{T31} , respectively. Since Wt and Fa have G at position 29, the role of parameters d_3 and d_4 was not considered important in hindering the position of the side-chain of the residue at position 29.

One-dimensional FE profiles were also hard to interpret as multiple conformations could correspond to the same CV; e.g., multiple conformations may have a d_1 value that corresponds to the type-1 canonical conformation, but only a specific value of the ψ_{S30} dihedral for small d_1 values distinguishes it from the rest. Hence, a replica in which d_1 and ψ_{S30} dihedral were biased simultaneously resulted in FE profiles that were easy to interpret. Moreover, the CVs describing local features were combined with a global CV, resulting in FE profiles which were informative as to the correspondence of particular local features with global features (e.g., both d_2 and d_4 distances only become important for small d_1 values). Since we are trying to encapsulate the high dimensional space of atomic coordinates that define loop conformation basins into the low-dimensional space of a few CVs, the minimal set of CVs needed to sample the type-1 canonical

conformation, starting from a floppy/non-canonical conformation, was found from preliminary two-dimensional metadynamics simulations of mutants FFSa and FFD in d_1 and ψ_{S30} . In particular, we examined 22 CVs, including the dihedrals of all backbone residues in the H1 loop, the RMSD of the H1 loop from a type-1 canonical conformation, the number of native contacts of all c-alpha and c-gamma atoms, d_1 - d_4 , the number of protein-water contacts, and the dihedral correlation for S30 and T31 residues.

Gaussian hills of height 0.1 kJ/mol were added every ps. The BE simulations converged in ~50-100 ns for the different mutants and were run for additional 10 ns at a reduced hill deposition rate of 0.01 kJ/mol per ps to obtain statistics for the final free-energy “FE” landscapes.

Calculation of percentage of conformations belonging to different clusters

The percentage of conformations belonging to a particular cluster is found by calculating area under the FE curve in a three-dimensional CV space formed by CVs d_1 , ψ_{S30} and ψ_{T31} . The limits for the integration were determined by the largest cube encompassing all conformations with an FE under $3k_B T$ from the FE minima for that cluster.

Kinetics: Kinetic Monte Carlo

From the low-dimensional FE obtained from each replica, we constructed a multi-dimensional FE to account for all the important CVs. We then followed an approach similar to that described previously⁵⁰ to construct a Markovian kinetic model based on binning the FE data in the CV

space. This model assumes that the transitions between bins can be considered as diffusion in the CV space biased by the FE. The rate constants in such a case can be expressed as^{65,66:}

$$k_{\alpha\beta} = k_{\alpha\beta}^0 \exp\left(-\frac{1}{2} \frac{(F_\beta - F_\alpha)}{T}\right) \quad 2.2$$

where $k_{\alpha\beta}^0 = k_{\beta\alpha}^0$, are the rates associated with a flat FE surface. With the purpose of obtaining approximate relative rates of transition between important macrostates of the system, we set all $k_{\alpha\beta}^0 = O(1)$. While this is a simplifying assumption, we find it sufficient for the purpose of estimating relative rates (assuming that the exponential factor is dominant), and describing the important transitional steps in going from the floppy to the compact type-1 canonical conformation. This approach also allows us to qualitatively compare the mutants' propensities to sample different conformations, which may be a measure of their inherent flexibility. Once the data is gridded and rate constants are determined for transitions between nearest neighbor bins based on equation (1), we use both an analytical and a numerical (Kinetic Monte Carlo or KMC) approach for obtaining mean first passage times between important macrostates or basins.

The analytical solution is based on the equations derived for the milestoning^{67,68} approach. The cited method can be approximately mapped into ours by identifying their milestones with our "bins" in CV space. The authors derive these equations for the case when the jumps between the bins can be described as a Markov jump process, such that the probability of a system to jump from state i at time t , to state j at time Δt , is independent of all states before i . The rate of such a transition can be described in terms of rates q_{ij} for each pair of bins. Then the master equation of the Markov jump process can be written as,

$$\frac{d\rho(t)}{dt} = Q\rho(t) \tag{2.3}$$

where $\rho(t)$ is a row vector in which row i is the probability of finding the system in state i at time t and Q is the rate matrix such that its off-diagonal entries are given by q_{ij} and diagonal entries are given by $-\sum_{j \neq i} q_{ij}$.

To obtain an estimate of the mean first passage times, the authors derive an analytical expression which only depends on Q :

$$\hat{Q}T^N = -\mathbf{1} \tag{2.4}$$

where \hat{Q} is the $(N-1) \times (N-1)$ matrix obtained by deleting the row and column corresponding to the N^{th} milestone (bin) rendering it a cemetery state, T^N is a column vector with $(N-1)$ entries, i^{th} entry corresponding to the mean first passage time (MFPT) of the i^{th} milestone to the N^{th} milestone. It is also shown that if one uses optimal milestones (i.e. milestones in which successive transitions are statistically independent), equation S3 gives the exact expression for the mean first passage time to milestone T^N .

We use this approach to obtain MFPTs for transitions among different basins in the CV space, assuming that for a certain optimal grid spacing the transitions become Markovian. Accordingly, we found the maximum grid spacing for which the ratio of transition rates between different basins in a high-dimension CV space remains constant. For each grid-spacing, the FE for each microstate was evaluated as the average over all conformations that lie within the 3D cube formed by the 3 CVs. As the grid-spacing increases, we average over larger regions in the

3D CV space, obtaining increasingly higher coarse-grained estimates of the FE. When the grid spacing is coarse-grained further (potentially erasing important local features of the landscape), we see a reversal of trend in the rates of transition between different basins and the analytical, and KMC results start diverging. We therefore aim to choose a grid spacing which is sufficiently small to accurately describe the changes in the FE in the chosen CV space and sufficiently large such that “fast” motions within each bin can be ignored. We tested this approach with a model potential (Muller potential), which has also been used in a previous work⁶⁸. We were able to obtain identical results from KMC and equation S3 using rate constants between microstates (or bins) of the form 2 (in the main text). By recording the sequence in which the events occur in the KMC simulation, we generate an ensemble of pathways for the transition.

Validation of Force field

To validate the use of the CHARMM27-CMAP force field in explicit solvent for our system, we conducted molecular dynamics and BE metadynamics simulations of the VHH Cab-CA05 (PDB code 1F2X). This VHH exhibits the type-1 canonical conformation of the H1 loop. We used the pre-folded framework approach as described in the main article. The FE profiles obtained from BE metadynamics are shown in Figure 2. These simulations were started from the type-1 canonical conformation. The FE profiles confirm the presence of a deep basin corresponding to the stable canonical conformation.

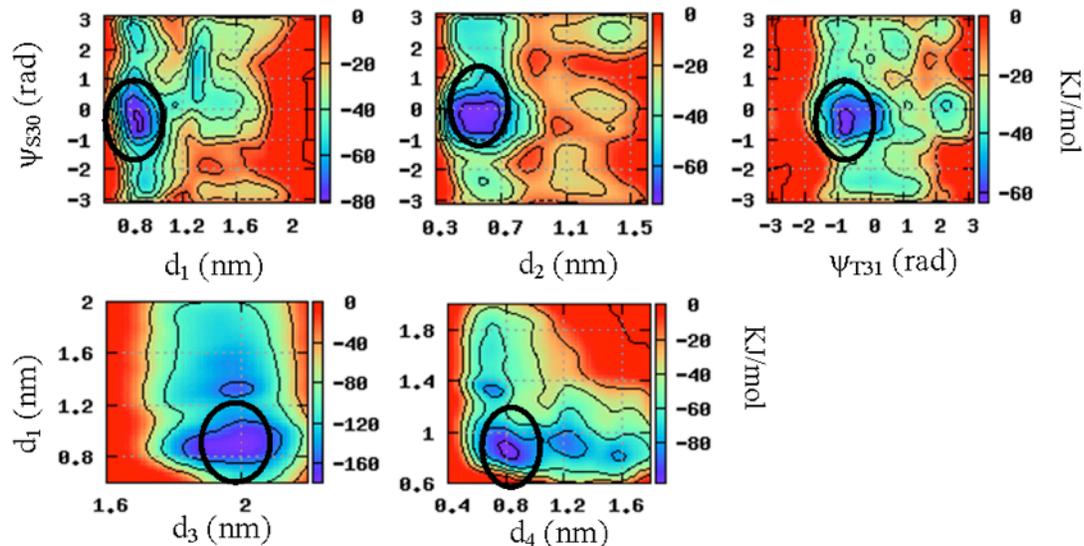


Figure 2.2 2-D Free Energy profiles for CAB-CA05 VHH. Black circles mark basins corresponding to the type-1 canonical conformation

2.3 Results and Discussion

Overview

The results of biased-exchange metadynamics for Wt and the mutants are summarized in Figures 3, 4, 5 and 6. In Figures 2a-2c (Wt, Fa, FF) and 3a-3c (FFSa, FFD, FFPA), we show the FE profiles obtained for replicas biased in the CV spaces of ψ_{S30} vs. d_1 (2-dimensional) and the isosurfaces obtained for FE values $\sim 3-5kT$ from the lowest FE in the 3-dimensional CV space d_1 , ψ_{S30} and ψ_{T31} . As mentioned earlier, these are the three most important CVs and are sufficient to compare the conformational equilibria of the mutants. The figure also shows the conformations corresponding to the important basins observed in the FE profiles (the conformations are color coded to represent specific clusters of stable conformations). These basins are obtained by combining FE results for CVs d_1 , ψ_{S30} and ψ_{T31} , producing FE basins in a three-dimensional CV space. For the Wt H1 loops, floppy conformations are mainly observed, characterized by large

values of d_1 and broad basins. For the mutants, new basins are created for smaller d_1 values. Moreover, as the number of mutations increases from one to three (Fa to FFSa), the canonical basin (for smaller d_1 values) becomes deeper, i.e., more stable. The Fa mutant with R27F mutation has a new FE basin (with respect to the Wt FE basins) corresponding to smaller d_1 values, though it does not correspond to the type-1 canonical conformation. Interestingly, for the mutants with two or more favorable mutations we observed a FE basin corresponding to the H1 type-1 canonical conformation; in particular, mutants FF, FFSa, FFD and FFPA resided in such a basin with a probability of 10%, 22%, 5%, and 8%, respectively. In effect, one can generate H1 loops with varying propensity towards a type-1 structure by mutating residues at these key positions. We explain this in more detail in the following sections, by discussing important features obtained from our analysis of the FE profiles for Wt and each mutant.

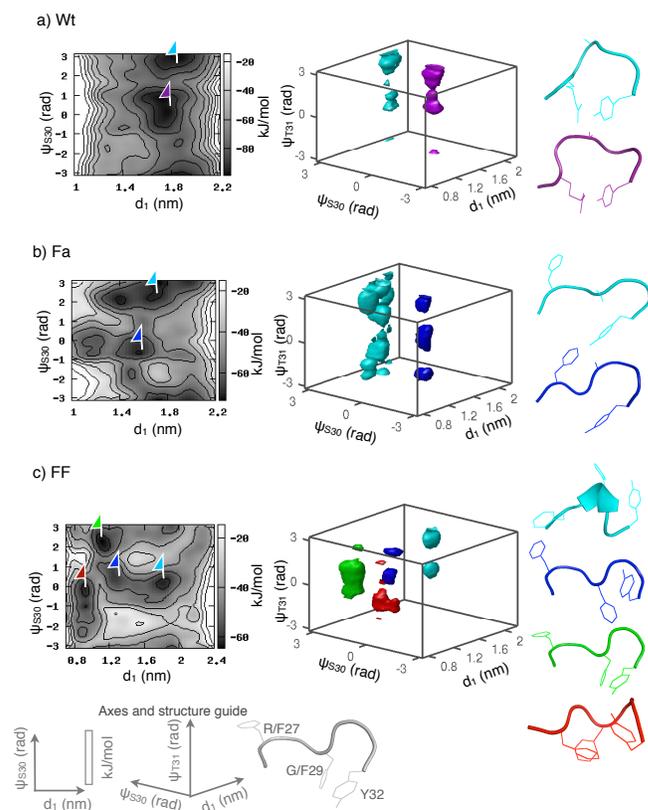


Figure 2.3 Summary of FE study of a) Wt, b) Fa and c) FF. Left: 2-dimensional FE profiles in the 2 most important CVs for the H1 loop of the Wt and mutants; flags mark the center of the stable conformations. Center: Iso-surfaces of minimum FE clusters in 3 CVs; color of each isosurface corresponds to a particular set of conformations. The colors of flags (in the 2-dimensional profiles), isosurfaces (in the 3-dimensional profiles) and the conformations agree. Canonical basin/conformation, if observed, is shown in red.

Wildtype (Wt)

The crystal and NMR structures for the wildtype llama VHH have been solved. The backbones of the two structures superimpose well, with an RMSD value of 1.68 Å.

In a previous implicit solvent study, it was found from the last 5ns of a 10 ns REM simulation that 35% of the H1 configurations sampled by the replica at 300 K had an RMSD $< 1.2 \text{ \AA}$ from the 1HCV crystal structure and no conformations were found with an RMSD within 1.2 \AA from the type-1 canonical structure.

A threshold value of 1.5 \AA is used for reporting the RMSD from reference structures in this study. Like in the NMR study, we see that the H1 loop in the wildtype VHH appears to have many possible conformations characterized by multiple broad FE wells (see ψ_{S30} vs. d_1 FE plot) accommodating many H1 loop conformations. The structures from the cluster analysis of the broad free-energy minima are shown in Figure 2a. 57% of the conformations from the basin at $d_1 \sim 1.7 \text{ nm}$ and $\psi_{S30} \sim 0.0$ radians have an RMSD $> 2.0 \text{ \AA}$ and only 19% have an RMSD $< 1.5 \text{ \AA}$ with respect to 1HCV. A narrow but shallow free-energy minimum at $d_1 \sim 1.4 \text{ nm}$ is representative of the 1HCV H1 loop with an RMSD $> 2 \text{ \AA}$ for 10% and $< 1.5 \text{ \AA}$ for 62% of the conformations.

The ψ_{S30} and ψ_{T31} dihedrals in the 20 best structures calculated from the NMR study lie in the range ~ -1.3 to 0.87 and -0.52 to 0.73 radians (calculated from PDB structures, 1G9E), respectively. The ψ_{S30} and ψ_{T31} FE plot (Figure 3a) shows that this region corresponds to FE minima, also coinciding with the one for the ψ_{S30} vs. d_1 plot.

The ψ_{S30} vs. d_2 FE plot shows that there is a global FE minimum in this CV space for small values of d_2 . While d_2 is the distance between the R27 and Y32 side-chains, in the conformations representing these small d_2 values it is the interaction between the side-chains of R27 and Y32, E96 and D101 that seems to cause a small d_2 value, exhibiting the association between the H1 and H3 loops through side-chain interactions. These findings cannot be verified since the R27

side-chain is absent in 1HCV, due to the low density of this X-ray crystal structure. The NMR conformations, on the other hand, show a similar association in only 1 out of the 20 best structures. Our observations also corroborate the implicit-solvent simulation results of Fenwick and Escobedo³¹ who found that the R27 residue forms a salt bridge with D101 residue that mimics the effect of an arginine at position 94 (common residue at position 94 in H3 loops of conventional antibodies, absent here). We also observe, in the representative conformations from the main basins in the Wt landscape, R27, D101 and E96 interact closely (H-bonds between N and O residues on the side-chains, lengths $< 4 \text{ \AA}$). One can further hypothesize that R27 in the H1 loop interacts with D101 and E96 in the H3 loop, and Y32 in the H1 loop interacts with W52a in the H2 loop, thereby “coordinating” the cooperative motion of the three loops.

2.3.1.1 Mutant Fa

In the implicit solvent study¹, the Fa mutant was found to exhibit both type-1 canonical (10% of equilibrium conformations) and wildtype non-canonical H1 conformations. The results of the biased metadynamics simulations are summarized in Figure 3b (and SI Figures 5, 6). Upon comparison of the ψ_{S30} vs. d_1 FE plots for Wt and Fa, it is clear that a new FE basin is generated for $d_1 \sim 1.2 \text{ nm}$. For this basin, 15% of the conformations have an RMSD $< 1.5 \text{ \AA}$ from the type-1 canonical conformation. However, the most populated cluster for this basin has a representative conformation with an alpha-helix formed by residues F27, T28, G29, and S30. In the same plot, the FE basin at $d_1 \sim 1.4 \text{ nm}$ and $\psi_{S30} \sim -0.8$ radians (Figure 3b, blue isosurface and conformation), has an RMSD $\sim 3.8 \text{ \AA}$ from the 1HCV non-canonical conformation. The mutation R27F also disrupts the electrostatic interactions between R27, D101 and E96 residues which were observed in the Wt conformations; on the other hand, one sees now more variability in the H2 loop conformation which correlates with the variability in the H1 loop conformation. In short, these

results suggest that one mutation at position 27 in the H1 loop increases the propensity for conformations with lower d_1 values, albeit not sufficient to form a stable type-1 canonical basin.

2.3.1.2 Mutant FF

This mutant has two mutations at the key positions 27 and 29 in the H1 loop. In the implicit solvent study¹, the FF mutant was found to exhibit both type-1 canonical (70% of equilibrium conformations) and wildtype non-canonical conformations). In the BE metadynamics simulations, a new basin is seen at $d_1 \sim 0.9$ nm and $\psi_{S30} \sim -1.0$ radians in the ψ_{S30} vs. d_1 FE plot (Figure 3c). The conformation corresponding to this basin is shown in Figure 3c (red) and 5a, where we observe the burial of F at position 29. While the side-chains of F27, F29 and Y32 may have some deviation from a type-1 canonical conformation, 99% of the conformations corresponding to this basin have an RMSD < 1.5 Å from the type-1 canonical conformation (Figure 8). Moreover, while the canonical basin in the d_1 vs. ψ_{S30} FE profile is narrow and well-defined, it is not so in the ψ_{S30} and ψ_{T31} profile as it merges with another basin, exhibiting a tendency towards a conformation with dihedrals that deviate from those for the canonical basin (Figure 5c).

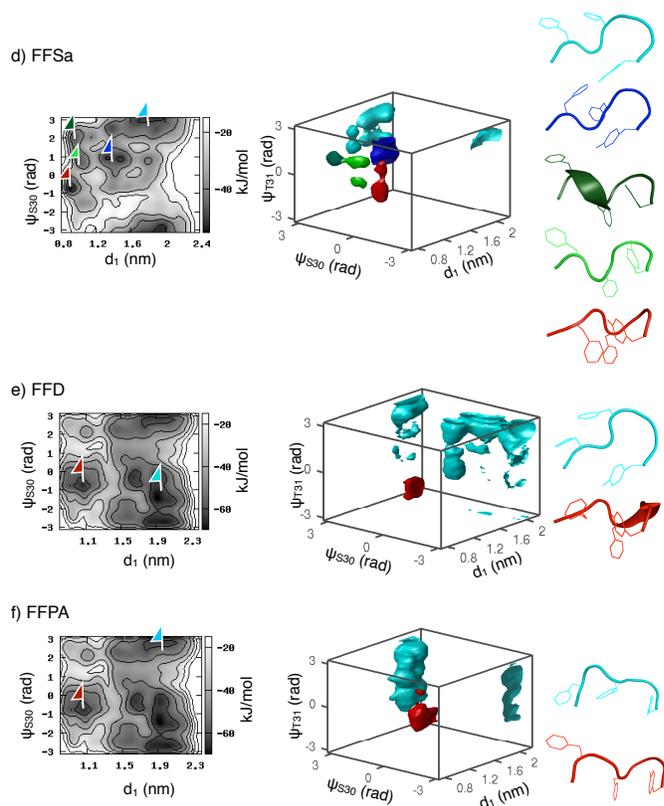


Figure 2.4 Summary of FE study of d) FFSa, e) FFD and f) FFPA. Left: 2-dimensional FE profiles in the 2 most important CVs for the H1 loop of the Wt and mutants; Flags mark the center of the stable conformations. Center: Isosurfaces of minimum FE clusters in 3 CVs; color of each isosurface corresponds to a particular set of conformations. The colors of flags (in the 2-dimensional profiles), isosurfaces (in the 3-dimensional profiles) and the conformations agree. See axes and structure guide in Figure 2 for more details. The canonical conformation, if observed, is shown in red. Isosurface and conformation shown in FFPA mutant are the closest to those of the type-1 canonical structure.

Mutants FFSa, FFD

In the implicit solvent simulations, these mutants were found to adopt a stable type-1 H1 canonical conformation. The FE profiles obtained from the two BE metadynamics simulations are shown in Figures 4d and 4e. The type-1 canonical conformation is represented by a narrow, stable basin in these plots, partially validating the findings of the implicit solvent study.

For the FFSa mutant, the FE landscape appears extremely rugged with more than one stable conformation accessible to the H1 loop. Let us first consider the FE basin for large values of d_1 . In the ψ_{S30} vs. d_1 (Figure 4d), d_3 vs. d_1 and d_4 vs. d_1 plots (see Figure 6 for details on the FE profiles for CVs d_3 and d_4), we observe a FE basin corresponding to $d_1 \sim 1.8$ nm and $\psi_{S30} \sim 3.14$ radians, $d_3 \sim 1.0$ nm and $d_4 \sim 0.8$ nm. The representative conformation is shown in Figure 4d (cyan); this is an extended structure and the H1 loop side-chains are fully exposed to the solvent, as in the non-canonical conformation. Another FE basin exists at $d_1 \sim 1.2$ nm which corresponds to an intermediate basin discussed in more detail in the section on pseudo-kinetics. The FE basins at $d_1 \sim 1$ nm are the most interesting; the most stable one at $\psi_{S30} \sim -0.8$ radians represents a type-1 canonical structure (red, others green). The conformations around this basin in the d_1 vs. ψ_{S30} free-energy profile are structurally close to the type-1 canonical conformation and, as we show later, may exist in equilibrium with this canonical type (and play a role in antigen-binding). These FE profiles (Figure 4d) illustrate that the flexibility of the loops makes it difficult to transition from the initial floppy conformation (large d_1 , root mean square fluctuation “RMSF” $\sim 0.2 \pm 0.02$ nm) to the final type-1 canonical conformation (small d_1 , RMSF $\sim 0.13 \pm 0.01$ nm) due to the multiple kinetic traps separating the two states. The overall transition can be characterized as a disorder-to-order transition, especially when one considers the importance of packing the F29, S52a, R71 and K76 side-chains in the cavity formed by the H1, H2 and H4 loops, together with the proper positioning of the F27 and Y32 side-chains, shielding the hydrophobic core on all sides. This packed cavity lowers the FE of the type-1 basin relative to those of the surrounding basins, in which the cavity is partially exposed to water molecules: the aromatic-aromatic⁶⁹ and aromatic-amino⁷⁰ interactions among the aromatic rings (27F, 29F, 32Y) and the amino groups in arginine and lysine (R71, K76) and hydrophobic interactions. This canonical basin also gains

additional entropy by shielding the F29 side-chain from water, increasing the entropy of water molecules that would otherwise be ordered around these side-chains to minimize unfavorable interactions with aromatic rings. However, this effect may not be sufficient to stabilize the loop in this conformation entirely, thereby leading to many conformational isomers. At the end of this section, we discuss this in the context of the hydrogen bonding patterns observed for the H1 type-1 canonical conformation obtained for mutants FF, FFSa, FFD and FFPA (Figure 5).

The FE profiles of the FFD mutant are summarized in Figure 4e. Similar to the FFSa case, FFD has multiple FE basins; however, the type-1 canonical basin observed at $d_1 \sim 1.0$ nm is not surrounded by other basins. Moreover, in the ψ_{S30} vs. d_1 FE plot (Figure 4e), this basin is not characterized by a narrow FE well. The d_3 vs. d_1 and d_4 vs. d_1 FE plots offer additional details. The distributions in RMSD values from a type-1 H1 structure (PDB code 1DFB), for the basins representing this H1 loop type are shown in Figure 8, for FF, FFSa and FFD mutants.

Mutants FFPA and FFTL

In the type-2A canonical conformation for the H2 loop observed in conventional antibodies, residue combinations P-A and T-L are commonly observed at positions 52a and 71. As explained earlier, the purpose of making similar mutations to the H2 loop is to determine if a better conserved combination provided higher stability to the final canonical conformation of the H1 loop. The FE profile for the FFPA mutant is shown in Figure 4f. For the specific region characterizing the canonical basin ($d_1 \sim 1.0$ nm, $d_2 \sim 0.5$ nm, $\psi_{S30} \sim -1.0$, $\psi_{T31} \sim -0.3$), we did not find a stable FE well for the FFPA or FFTL mutants.

A closer look at the stable basins closest to the type-1 canonical conformation in our simulations suggests that the K76 side-chain interacts with the H1 backbone (Figure 7d). This prevents the F27 side-chain from collapsing to form the compact canonical conformation. To understand the role of the residue at position 76, we investigated which is the most frequently occurring amino-acid at this position in conventional VH ORFs. We found that while the two combinations (P-A and T-L) are observed in gene families with both H1 type-1 and H2 type-2A canonical loops, *a serine rather than a lysine is conserved at position 76 for all ORFs corresponding to these gene families*. In fact, *S is observed almost exclusively at position 76 when P is at position 52a and A is at position 71*. Two representative cases are antibodies 1MFA and 1TET (PDB codes), which have the combinations P-A and T-L, respectively, and exhibit H1 type-1 and H2 type-2A canonical conformations, Antibody 1MFA has a T at position 76, whereas 1TET has an S at that site. Interestingly, both PDB structures reveal hydrogen-bonding patterns similar to those observed for FFPA (Figure 7 d) except that the H-bond between the side-chain of residue 76 and the H1 backbone is missing. This might be due to the smaller size of the T and S

side-chains, which prevents them from being within H-bonding distance of the H1 backbone, as observed for the longer K76 side-chain of the llama VHH. While we haven't conducted any simulations to test the effects of the K76S mutation, our results for FFPA and FFTL show that the H1 type-1 canonical conformation has not become more stable (relative to the FF mutant) by adding mutations that disregard the packing of side-chains in the cavity formed by the H2 and H4 loops.

It is also informative to compare the hydrogen bonding patterns of the canonical loops obtained for mutants FF, FFSa, FFD and the canonical-like conformation for mutant FFPA shown in Figure 7. It appears that the number of hydrogen bonds is maximized for the FFD mutant (17), followed by FF and FFSa (11 and 10 respectively) and then FFPA (8). Both inter-loop and intra-loop bonds are maximized for the canonical conformation of FFD. Besides H-bonding, non-electrostatic interactions such as hydrophobic, amino-aromatic and aromatic-aromatic contacts also play an important role in the stabilization of these compounds. Also, despite that one of our markers of stabilization towards an H1 type-1 loop is the burial of hydrophobic side-chains F27 and F29, there exists an interplay between the decrease in protein degrees of freedom upon burial of these side-chains, and the increase in entropy upon releasing (to bulk) the waters that are ordered around these hydrophobic side-chains while they are exposed to solvent. This interplay can lead to different outcomes; for instance, we observe that for mutants FF, FFSa and FFD, floppy as well as compact-canonical conformations are stabilized to varying extents.

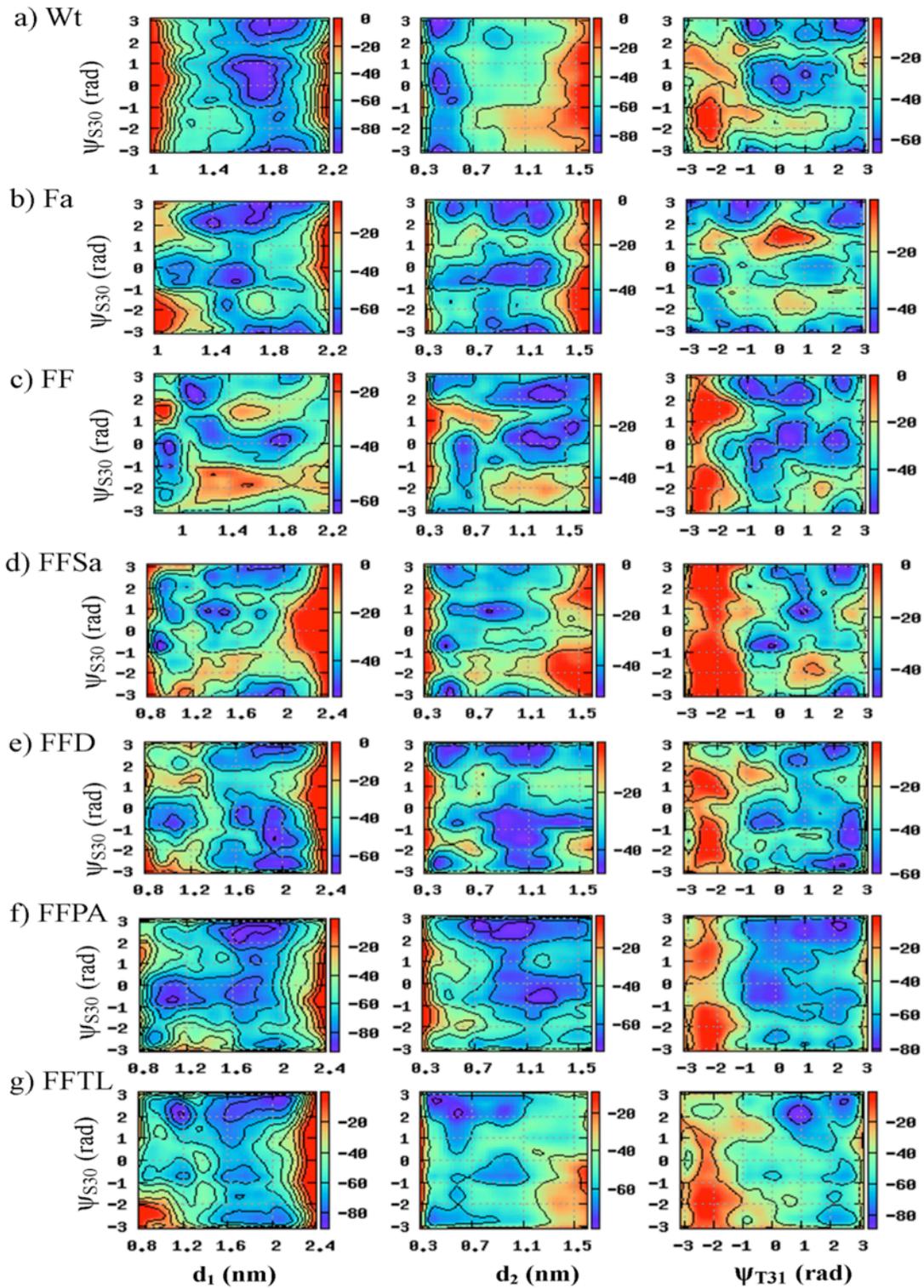
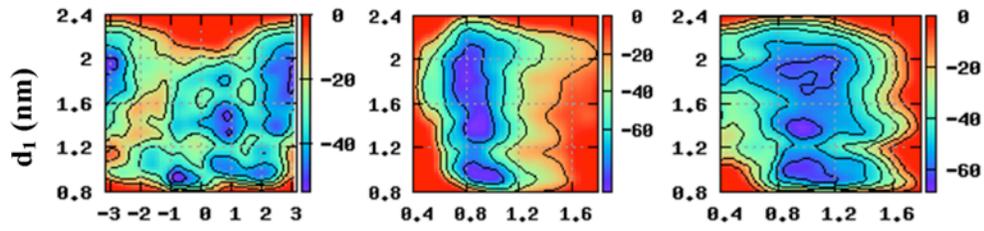
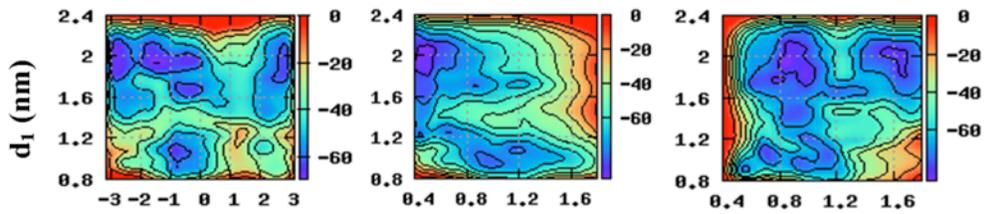


Figure 2.5 2-D FE profiles for Wt and mutants in CVs d_1 , ψ_{S30} , ψ_{T31}

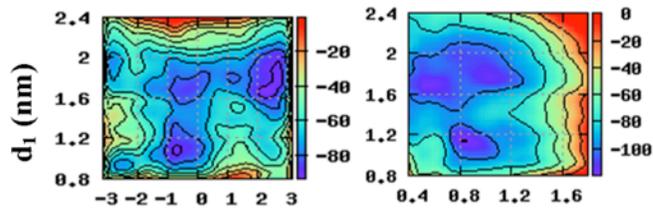
a) FFSa



b) FFD



c) FFPA



d) FFTL

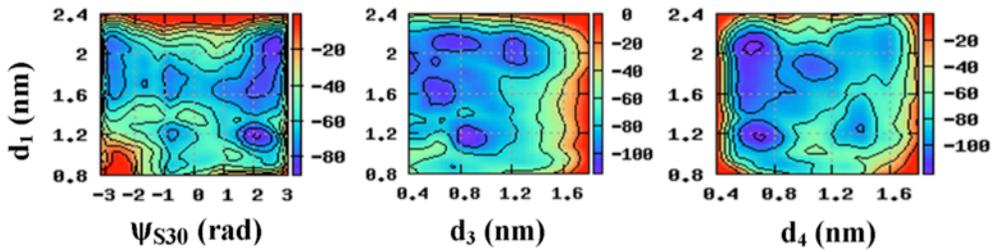


Figure 2.6 2-D FE profiles for mutants FFSa, FFD, FFPA and FFTL in CVs d_1 , ψ_{S30} , d_3 and d_4

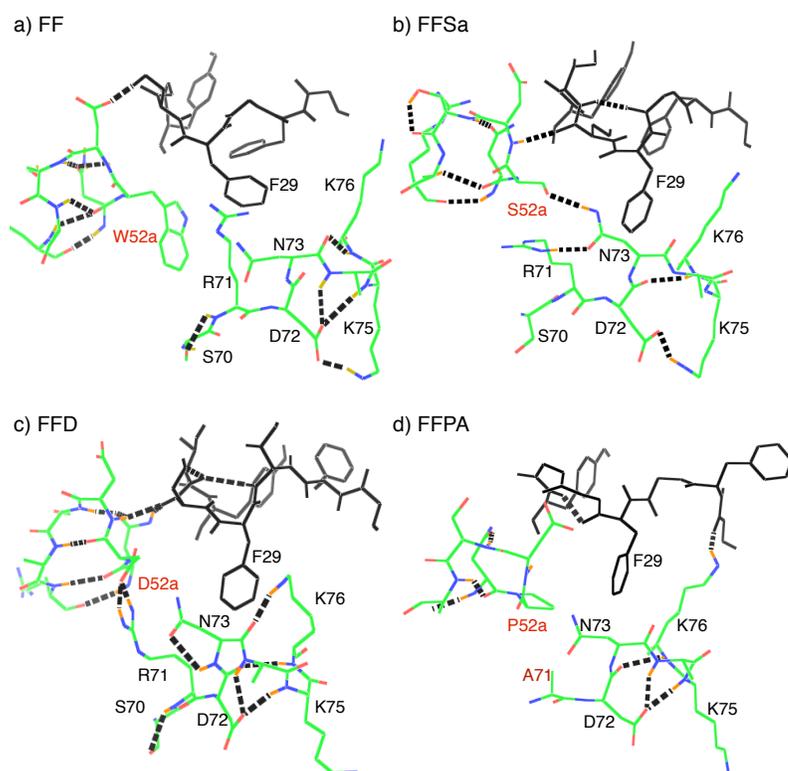


Figure 2.7 H1 type-1 canonical conformations obtained from BE Metadynamics for a) FF b) FFSa c) FFD d) FFPA mutants. The H1 loop is colored black, and residues marked in green (carbon), blue (nitrogen) and red (oxygen) correspond to important proximal residues within loops H2 and H4. The text in red highlights mutation sites.

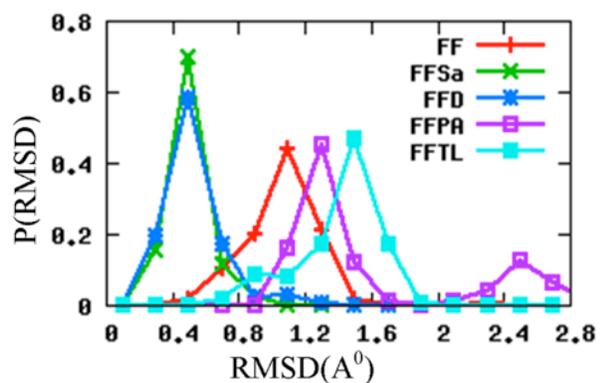


Figure 2.8 RMSD from an H1 type-1 canonical conformation of a reference structure (PDB code 1DFB), for the type-1 canonical basins of FF, FFSa, FFD, FFPA and FFTL.

Pseudo-Kinetics Results

For the purpose of constructing a Markov model, we use the 3-D CV space in d_1 , ψ_{S30} , and ψ_{T31} , and compare the kinetics for FF, FFSa and FFD mutants which we take to be a representative set (e.g., embodying different FE profiles and propensities toward the H1 type-1 canonical structure). The principal basins in the 3-D CV space are defined in Tables 2, 3 and 4, and the corresponding conformations in the 3-D CV space are shown in Figures 3c, 4d, and 4e. While FF and FFSa seem to have multiple intermediate (I) basins between the floppy (F) basin (characterized by a large value of d_1) and the type-1 canonical basin (C), the FFD mutant seems to exhibit no intermediate basins. This is further reflected in the mean first passage times (MFPTs) summarized in Table 5. The MFPTs obtained from both KMC and via an analytical solution are in good agreement. Let us first compare the $F \leftrightarrow C$ transition for FF, FFSa and FFD. The MFPT for the $F \rightarrow C$ transition in FF and FFSa is comparable to (though slightly higher than) the reverse transition, whereas for FFD it is an order of magnitude higher. The $C \rightarrow F$ (and to other I states) transition, on the other hand, provides information on the accessibility and ease of transition of the H1 loop from the type-1 canonical state to nearby partially “folded” non-canonical conformations that may play a role in cooperative changes in the H2 and H3 loops (relevant to binding). In this context, FFD may have the most “stable” type-1 canonical conformation for the H1 loop if it were accessed before the F basin during an actual folding event as its $C \rightarrow F$ transition has a much higher MFPT than FF and FFSa, and has the least flexible H1 loop. In other words, in FFD the canonical basin provides the strongest kinetic trapping. Moreover, as noted in the previous section, FFD exhibits the most number of hydrogen- bonds in the cavity formed by H1, H2 and H4 loops (Figure 7c). This may be a contributing factor in the increased stability of

its canonical conformation and the lack conformational isomers as in the case of FFSa. This complements the idea that the canonical to non-canonical transition of the H1 loop is an order-to-disorder transition in which involves the breakage of both intra- and inter-loop hydrogen-bonds.

We also generated an ensemble of paths for these transitions, by recording the sequence of events in the KMC simulation. We did this for the F→C transition of the FFSa mutant. The most probable pathways thus obtained are shown in Figure 9. Out of 1682 pathways, 1418 (84%) visit basin I4, 190 (11%) visit basin I2, 74 visit basin I1 and only 1 visits basin F just before reaching basin C. These pseudo-pathways suggest that the preferred route for this transition is through the intermediate basin I4.

Table 2.2 3-D basins for FF mutant

Basin	Boundaries $d_1, \psi_{S30}, \psi_{T31}$	% of all conformations	Color code in Figure 2c
F	(1.7,1.9), (-1.0,0.0), (-1.0, 0.0)	25	Cyan
I1	(1.25,1.4), (0.0,1.0), (-1.0,3.14)	4	Blue
I2	(1.05,1.15),(2.0,3.14),(-1.0,3.14)	20	Green
C	(0.85,1.0),(-1.0,0.0),(-1.0,0.0)	10	Red

Table 2.3 3-D basins for FFSa mutant

Basin	Boundaries $d_1, \Psi_{S30}, \Psi_{T31}$	% of all conformations	Color code in Figure 3a
F	(1.3,2.0), (2.4,3.14), (0.5,2.8)	37	Cyan
I1	(1.25,1.6), (0.0,1.2), (0.5,1.8)	18	Blue
I2	(1.0,1.2), (-2.5,-1.0), (- 1.5,1.5)	0.1	Not shown
I3	(0.95,1.15), (1.5,3.0), (0.5,2.0)	2	Dark Green
I4	(0.95,1.15), (0.2,1.5), (- 2.0,1.5)	3	Green
C	(0.85,1.0), (-1.2,-0.2), (-1.0,1.8)	22	Red

Table 2.4 3-D basins for FFD mutant

Basin	Boundaries $d_1, \Psi_{S30}, \Psi_{T31}$	% of all conformations	Color code in Figure 3b
-------	---	------------------------	-------------------------

F	(1.7,2.1), (-3.14,0.0), (1.0,3.14)	84	Cyan
C	(0.85,1.1),(-1.0,0.0),(- 1.0,0.5)	5	Red

Table 2.5 MFPTs for transitions to and from the H1 type-1 canonical basin for the FF, FFSa and FFD mutants. Transition time reported in MC steps.

Transition	FF	FFSa	FFD
C→I1 (C ← I1)	239 (495)	418 (1172)	---
C→I2 (C ← I2)	225 (533)	1540 (19)	---
C→I3 (C ← I3)	---	514 (1084)	---
C→I4 (C ← I4)	---	409 (1059)	---
C →F (C ← F)	309 (533)	557 (1282)	2423 (67251)

The conformations sampled in our simulations correspond to an “initial condition” where the framework region is prefolded, and need not match the transient configurations occurring when the VHH domain is folding from some initial state (e.g., a random coil). Our prefolded-framework simulations do reveal the different conformational basins that can occur at equilibrium, but cannot determine which basin may be encountered first (given some initial conditions).

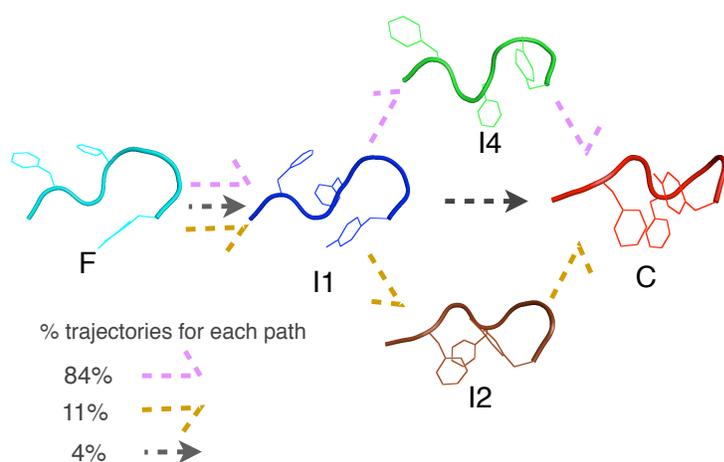


Figure 2.9 Relevant transition pathways proposed for the F ->C transition of the FFSa mutant. Also reported, is the percentage of pathways visiting basins I4/I1/I2 before reaching basin C. Color scheme same as the one used in Figure 4d.

Role of key-residues and proximal residues

For the transition studied, we found that packing of the hydrophobic residue at position 29 in the H1 loop depends on residues in loops H2 and H4. Depending on the size and nature (hydrophobic, hydrophilic or aromatic) of the residues comprising these loops and how they collectively pack in the cavity formed by the H1, H2 and H4 loops, the H1 and H2 loops may have “rigid” or floppy

conformations. For hypervariable regions whose H1 and H2 loops do not rearrange during antigen binding, one would expect to have less floppy loops – characterized by strong interactions between the side-chains of residues in this cavity. The interactions between side-chains in this cavity may also give rise to cooperative rearrangement of the three hypervariable loops. In the wildtype llama, the H1 loop has a G at position 29, in contrast to the phenylalanine primarily observed in type-1 canonical loops. The NMR rate constants suggest that the H1 loop is the floppiest, and that the H1 and H2 loops can be associated with slow conformational changes that may take place during binding. This indirectly suggests a specific role for the residue at position 29. By replacing the F (I/L) present in conventional H1 loops with a G, the llama antibody may achieve a more flexible H1 loop allowing it to compensate for a rather short H3 loop. Moreover, the H2 loop has a W at position 52a in the wildtype and its side-chain does not directly interact with G at position 29, given that this residue lacks a side-chain. The overall motion of the H1 loop is more closely coupled with that of the residues at positions 31 and 32, which have been conjectured to participate in binding. Incidentally, residues 31 and 32 and residues from the H2 loop have high R2 (Hz) and Rex (Hz) values which correspond to slow conformational changes. Thus, H1 may play the role of orchestrating conformational changes in the H2 and H3 loops through residues R27 and Y32.

Overall, our results support the idea that the “folding” of the H1 loop from a floppy or random coil conformation to a type-1 canonical conformation, can be broadly described by two transitions: (i) the burial of a hydrophobic side-chain at position 29, and (ii) the collapse of the side-chain at position 27 (usually F/Y). Upon investigating the role of residue 71 through mutants FFSa, FFD, FFPA and FFTL, we found that this residue interacts with others in the H1 and H2 loops to alter the stability of the H1 loop conformation. It was initially expected that

mutants FFPA and FFTL would further stabilize the canonical conformation; however, this was not the case. This behavior is likely due to unfavorable interactions between the K76 side-chain and the H1 loop backbone. It is hypothesized, based on a subsequent sequence analysis of conventional antibodies with combinations P-A and T-L at positions 52a-71, that a smaller hydrophilic residue like serine or threonine may remedy this.

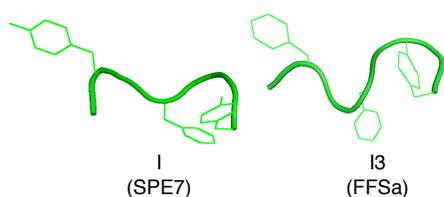


Figure 2.10 Evidence of conformational isomers of the type-1 canonical conformation in experimentally characterized antibody SPE7 juxtaposed with I4 conformation of the FFSa mutant. *Left:* Conformational isomer (I, from PDB structure 1OCW) for antibody SPE7. *Right:* Conformational isomer (I4) for mutant FFSa

2.4 Conclusions

In this work we have shown that the FE landscape of the llama VHH H1 loop tilts from a stable non-canonical conformation for the Wt, to a stable type-1 canonical conformation for the 3-point mutant FFSa. We have also studied other mutants with different intermediate degrees of propensity towards a type-1 H1 canonical structure. The FE profiles provide evidence of the conformational isomerism of the H1 loop. Proximal residues at positions 52a, 71 and 76 have been identified in our simulations as playing an important role in determining the conformational states of the H1 loop. The results obtained via pseudo-kinetic simulations of a Markovian model suggest that while FFSa samples the H1 type-1 canonical basin with the highest frequency among all the mutants studied, it may be susceptible to transition between conformational isomers; in contrast, FFD may have a more rigid and kinetically stable canonical conformation. This

highlights the importance of a single mutation at position 52a in determining the conformational flexibility of the H1 loop.

Our results serve to validate some of the most important results of the earlier VHH mutagenesis study conducted in implicit solvent. Indeed, we also found that mutants FF, FFSa and FFD exhibit a FE basin for the H1 type-1 canonical conformation, and that residues at positions 27, 29 and 52a play a key role in determining the stability of this canonical type. However, our results for mutant Fa differ from those reported in the implicit solvent study, in that we did not observe a well-populated H1 type-1 canonical basin. We believe that this discrepancy is only partially due to differences in the solvation and/or energy models, given that the results for mutants FF, FFSa and FFD are qualitatively consistent with those of the implicit solvent study. More importantly, we think that the CV d_1 is not well suited to describe the F \rightarrow C transition for Fa mutant because, unlike the other mutants which have a hydrophobic sidechain at position 29, Fa does not; in fact, its residue G29 lacks a side-chain which hence renders d_1 a poor indicator for the quality of hydrophobic core formation. Overall, for the 1Fa mutant, a combination of backbone dihedrals for residue 29 (instead of d_1) along with the other CVs may have been a more suitable choice to sample the conformational space pertinent to the H1 type-1 canonical basin. As additional findings, our explicit-solvent simulation results unveiled the diverse range of conformational isomerism that the H1 loop can exhibit and demonstrated the key role played by residues proximal to the H1 loop in determining the most accessible conformational isomers. This could be an important step in understanding cross-reactivity of antibodies and improve their rational design.

Our work also leads us to conclude that replica exchange simulations in implicit solvent may be a fast and reliable way to predict stable conformations of VHH loops, since it agrees

fairly well with our explicit solvent simulation results, at least in terms of predicting the relative propensities towards a type-1 canonical conformation. However, explicit solvent simulations reveal a more detailed and realistic picture of the conformational landscape of the wildtype VHH and its mutants, at the expense of requiring more extensive sampling to negotiate the increased ruggedness of the energy landscape. .

Many studies^{34,38,39,41-4371-75} have found that antibodies may exist in multiple configurations or isomers even before exposure to antigens. Our results also highlight such antibody isomerism pre-binding. Specifically, a conformational landscape for the H1 loop (and/or other loops) that is characterized by multiple stable/metastable states separated by small energy barriers, may have evolved to undergo conformational rearrangement during binding (e.g., FFSa and Wt), or may pre-exist in equilibrium with its isomers even before the antigen comes into the picture. On the other hand, an antibody with a stable structure separated from other metastable states by a large energy barrier may be less likely to exhibit conformational isomerism before the antigen approaches it (e.g., FFD). Moreover, one may conjecture that such an antibody would be less likely to undergo conformational rearrangement, and thus more prone to bind via a lock-key type mechanism. To illustrate the potential role that the isomerism we observe could have in binding, consider the H1 loop conformation corresponding to two conformational isomers obtained by James et al.^{39,41} (Figure 12). The SPE7 antibody studied in this reference binds to different antigens through different pre-existing conformations. As shown in Figure 12, for SPE7, a closely related conformational isomer of the type-1 conformation of the H1 loop exists as in the case of FFSa or FF. The H1 loop from the PDB structure 1OCW, which is the conformational isomer of the H1 type-1 canonical conformation is analogous to the I4 and I3 isomers seen in our simulations of the FFSa mutant. Hence, our study also illustrates similar conformational

isomerism in molecular detail, and shows that the sequence of the loops and their proximal residues modulates the intrinsic disorder of the VHH hypervariable regions. This observation aligns with the idea that intrinsic disorder plays an important role in the function of proteins – especially for molecular recognition. The presence of multiple conformational basins separated by moderate free-energy barriers, as observed in our simulations, imparts higher loop flexibility and may be a feature that is more prevalent in single-domain antibodies than in conventional two-domain antibodies. Moreover, it is expected that the interactions with a second domain (as in conventional VHs) may impose additional geometric or enthalpic constraints that favor H1 loop rigidity.

3 DESIGN OF AN ANTI-ALPHA SYNUCLEIN VHH: FROM SELECTION TO MODEL DEVELOPMENT

3.1 Introduction

In this Chapter, we describe our attempt to develop a model for an anti-Alpha-Synuclein vHH. Alpha-Synuclein (AS) is a pathological marker of Parkinson's disease (PD) and related disorders²⁰. It is natively disordered, with many conformational forms upon interacting. Misfolding and subsequent aggregation of AS and other proteins into soluble and insoluble oligomers and high molecular weight fibrils has been implicated in PD and other neurodegenerative diseases. Prevention of aggregation and misfolding of such aggregation-prone species has been suggested as a viable therapeutic strategy for reducing the pathogenesis in various neurodegenerative diseases. The central hydrophobic domain (AAs 61-95) is termed as the nonamyloid component (NAC)²¹. The NAC domain is aggregation-prone and can form beta-sheets by self-association²². Recently, rational design approaches^{27,28} have been used for generating antibody fragments for this purpose. In particular, the rational design approaches (as opposed to high-throughput screening) have focused on designing NAC-specific peptides and grafting them into the CDR regions of a variable heavy domain (vHH) framework.

We have used a dual approach. Figure 3.1 illustrates our approach in the form of a flowchart. Starting from an immunized Camelid library against the NAC region of A53T mutant of AS (henceforth referred to as A53T), we used a bacteria-based selection technique to obtain a NAC-specific vHH, followed by computational modeling of the vHH and the vHH-NAC binding.

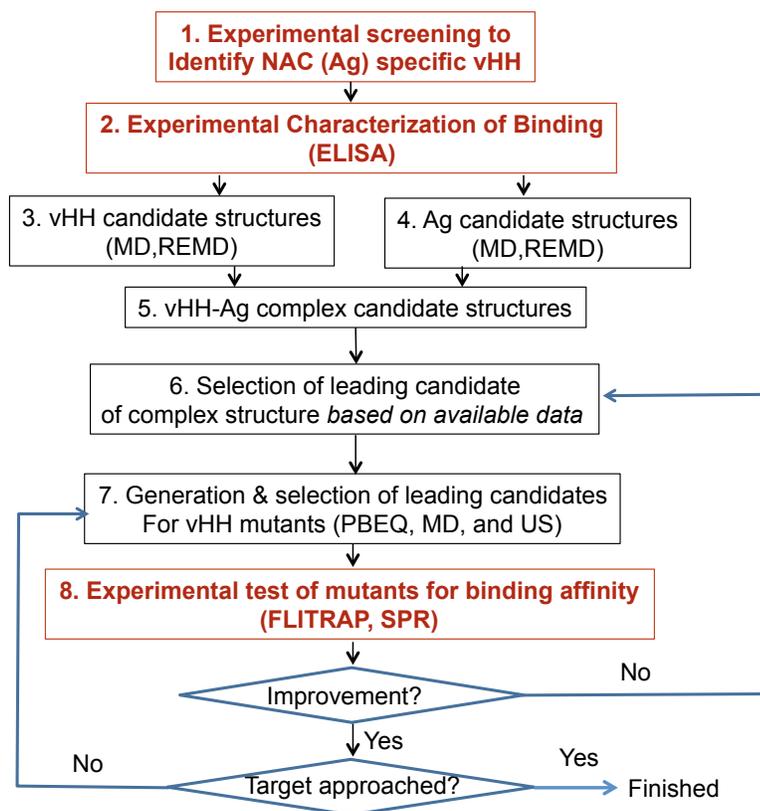


Figure 3.1 Workflow for generation and rational design of anti-AS vHH

3.2 Methods Experiments

Alpaca Immunization

The immunized library from an Alpaca was generated by the Messer group (Albany). Briefly, the peptide AAs 53-87 was synthesized as a biotin fusion and used for immunizing Alpaca. Methods for Alpaca immunization are described in detail elsewhere⁷⁶.

Bacterial Strains, Plasmids and Growth Conditions

Wild-type *E. Coli* strain MC4100 was used for FLITRAP⁷⁷ experiments and subcellular fractionation. BL21 (DE3) was used for cytoplasmic expression of proteins. Cultures were grown in LB medium (except for purification) supplemented with the appropriate antibiotic, and protein expression was induced with IPTG (0.1 mM) or arabinose (0.2%-1% w/v) depending on the plasmid used. Antibiotics were supplemented at the following concentrations: ampicillin (100 $\mu\text{g}/\text{mL}$), chloramphenicol (20 $\mu\text{g}/\text{mL}$), and kanamycin (50 $\mu\text{g}/\text{mL}$).

Cloning and Plasmid construction

The immunized vHH library was PCR amplified from the JSC phagemid and cloned into the pDD18-ssTorA-X-FLAGtag (described elsewhere⁷⁷). The library was cloned in place of “X”. The ssTorA is the signal sequence for targeting proteins to the Twin-Arginine-Translocation (Tat) pathway located in the inner membrane of *E. Coli*. The pDD322-TatABC-Y-Bla cassette construct elsewhere was used to clone the antigen (A53T) in place of “Y”. The pDD18 and pDD322 constructs were used for FLITRAP screening⁷⁷.

To demonstrate periplasmic localization of vHH carrying the Tat signal sequence, a pSALECT construct⁷⁸ was used for its relatively low copy number (in comparison to pDD18). The hits from library selection were subcloned into pSALECT plasmid with a C-terminal FLAG tag using XbaI and NotI sites flanking the target gene.

To express proteins of interest in the cytoplasm, we subcloned hits from FLITRAP selection into a pET21a(+) plasmid with a C-terminal 6 \times -His tag using NdeI and NotI sites flanking the target gene.

Selective Growth Assays

For FLITRAP experiments, cells carrying FLITRAP plasmids (pDD18 and pDD322 constructs) were grown overnight in LB containing 25 $\mu\text{g}/\text{mL}$ Cm and 10 $\mu\text{g}/\text{mL}$ Tet. Screening of cells was performed by spreading an equivalent number of serially diluted overnight cells directly onto LB agar plates supplemented with 0.2% arabinose and 0-100 $\mu\text{g}/\text{mL}$ Carb and incubating at 37 °C for 36-48 h. Library selections were performed by electroporating the plasmid DNA libraries into E. Coli cells followed by electroporation with pDD322 –TatABC-A53T-Bla. Cells were incubated at 37 °C for 1h without antibiotics; then subcultured overnight into fresh LB containing 25 $\mu\text{g}/\text{mL}$ Cm and 10 $\mu\text{g}/\text{mL}$ Tet to ensure that cells contained both plasmids. After \approx 16 h, cells were spun down and normalized in fresh LB to $\text{OD}_{600} = 2.5$ followed by direct plating 100 μL of diluted cells onto LB agar supplemented with 1% arabinose and 50-100 $\mu\text{g}/\text{mL}$ Carb. Hits were randomly picked after incubation at 30 °C for 40-48 h. Randomly chosen positive clones were screened by spot plating to confirm Carb resistance and then sequenced.

Subcellular fractionation and Western Blot analysis

To prepare subcellular fractions for Western Blot analysis, we pelleted and washed \sim 20ml (vHH samples normalized by $\text{OD}_{600} \sim 75$) of induced culture with subcellular FB (30mM Tris-HCl, 1mM EDTA, and 0.6 M sucrose). Cells were resuspended in 1 mL subcellular FB and then incubated for 20 min at room temperature. After addition of 400 μL of 5 mM MgSO_4 , cells were incubated for 10 min on ice. Cells are spun down, and the supernatant was taken as the periplasmic soluble fraction. The pellet was sonicated thrice for 20 seconds. Following centrifugation at 14000 rcf at 4°C, the second supernatant was taken as the cytoplasmic soluble

fraction, and the pellet was the insoluble fraction. Proteins were separated by 12% SDS-polyacrylamide gels (Rio-Rad), and Western blotting was performed according to standard protocols. Briefly, proteins were transferred onto polyvinylidene fluoride membranes, and membranes were probed with either anti-FLAG antibodies conjugated to horseradish peroxidase (HRP) or anti-6x-His antibodies conjugated to HRP (Abcam).

Enzyme-linked immunosorbent assay (ELISA)

ELISA was used to evaluate the binding of lysates and purified vHHs to A53T. ELISA plates were coated overnight at 4°C with 1ug/ul A53T (Genway) in a bicarbonate buffer. Plates were then blocked at room temperature for 2 h with 2% nonfat milk in TBS. After plates were washed using TBS supplemented with 0.05% Tween 20 (TBST), lysate/purified protein samples serially diluted in TBS with 1% BSA (TBS-BSA) were added to the plates (40 μ L /well). Plates were incubated for 1 h at room temperature and then washed with TBST. HRP-conjugated anti-6x-His antibody (Abcam) in TBS-BSA was added to the plates (50 μ L /well). After 1 h of incubation at room temperature, plates were washed and then incubated with TMB HRP substrate (Thermo Fisher Scientific) for 10-20 mins. The reaction was quenched with 2M H₂SO₄, and the absorbance of the wells was measured at 450nm.

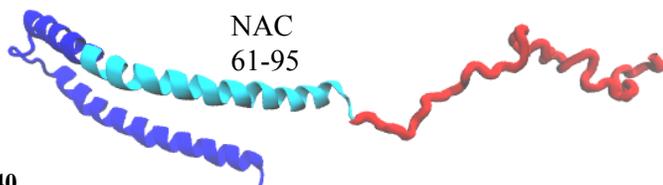
3.3 Methods Simulations

Model for vHH

The starting structures for vHH 4C were generated using Homology model engine of Swiss Model⁷⁹⁻⁸¹. The homology model was generated for 121 out of 122 amino acid residues of the vhh (Q at N-terminal was not present in the template structure). All simulations employed GROMACS using the CHARMM27 force field with CMAP with the OBC implicit solvent model. The SD integrator of the GROMACS package was used with a timestep of 2fs. The LINCS algorithm in GROMACS was used to constrain the length of all bonds. The inverse friction constant of the SD integration was set to 91 ps^{-1} ⁸². All interactions were run with infinite cutoff. The homology model was subjected to a short steepest descent energy minimization followed by a short equilibration at a constant temperature of 300K. For the REMD⁸³ simulations, all residues except those in the hypervariable loops H1, H2, H3 and H4 were backbone-restrained^{1,84}. The C_{α} , N and C backbone atoms that do not belong to the hypervariable loops were restrained with the harmonic force constant of 2000 kJ/nm, while backbone atoms in H1-H4 along with side chains from all the residues were free to move. These restraints were imposed to simulate the limited motion of these residues in the folded domain. 24 replicas at temperatures ranging from 300K to 900K were simulated for 20ns (per replica) for the REMD simulations with an exchange attempt every 1ps (500 steps). The `g_cluster` program in GROMACS with the linkage algorithm was used to cluster structures at the lowest temperature replica (300K) to obtain the top 8 conformations. These conformations were used to start a new REMD, and simulated for another 8ns to obtain the top 2 replicas.

Model for AS-NAC region

To develop a model for binding, we need a model for the NAC peptide. While it is widely accepted that native AS occurs as a disordered monomer⁸⁵, the possibility that transient local structural motifs/features^{86,87} are present has not been ruled out. In fact, multiple studies have reported presence of transient alpha-helical character in the NAC region. Consequently, two models were considered for the NAC region – 1) alpha-helical peptide (residues) obtained from the PDB: 1XQ8⁸⁸, the membrane associated X-ray structure of AS and 2) random coil peptide (generated by piecing together the AAs one-by-one). The aim of considering specific secondary conformations of the NAC peptide is to be able to generate a wide spectrum of possibilities for the final structure of the complex. Moreover, if the AS molecule is in a state of dynamic equilibrium between multiple transient conformations⁸⁷, binding might be specific to only one particular conformation. Alternatively, binding might result in arresting NAC in a particular conformation, as the circular dichroism spectra for AS is consistent with transient formation of α -helices in the first 100 N-terminal residues⁸⁷. Peptides of different lengths were used for modeling at different stages. For example, initially a 19-residues peptide (AAs 61-78) was considered for docking calculations. However, after equilibration and REMD steps, none of the random-coil containing complexes were found in the top 8 complexes obtained from REMD (described in more detail in the next section). This could primarily be a result of the unwieldy nature of the random-coil model for NAC peptide compared to the alpha-helical model. After the top complexes were obtained for the 19-residues peptides, the peptide was extended to 24 residues (AAs 57-80) or to 27 residues (AAs 57-83) (Figure 3.2).



AS 1-140
 MDVFMKGLSK AKEGVVAAA E KTKQGVAAA GKTKEGVLYV
 GSKTKEGVVH GVATVAEKT **EQVTNVGGAV VTGVTAVAQK**
TVEGAGSIAA ATGFVKKDQL GKNEEGAPQE GILEDMPVDP
 DNEAYEMPSE EGYQDYEPEA

NAC peptide 61-95
EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFV
Immunization peptide 53-87 (AS)
ATVAEKTKEQVTNVGGAV VTGVTAVAQK TVEGAGS
13 residue peptide 66-78 (A53T)
VGGAVVTGVTAVA
19 residue peptide 61-79 (A53T)
EQVTNVGGAV VTGVTAVAQ
24 residue peptide 57-83 (A53T)
EKTKEQVTNVGGAV VTGVTAVAQK
27 residue peptide 57-83 (A53T)
EKTKEQVTNVGGAV VTGVTAVAQKTVE

Figure 3.2 Sequence of AS and peptides used for experiments and modeling

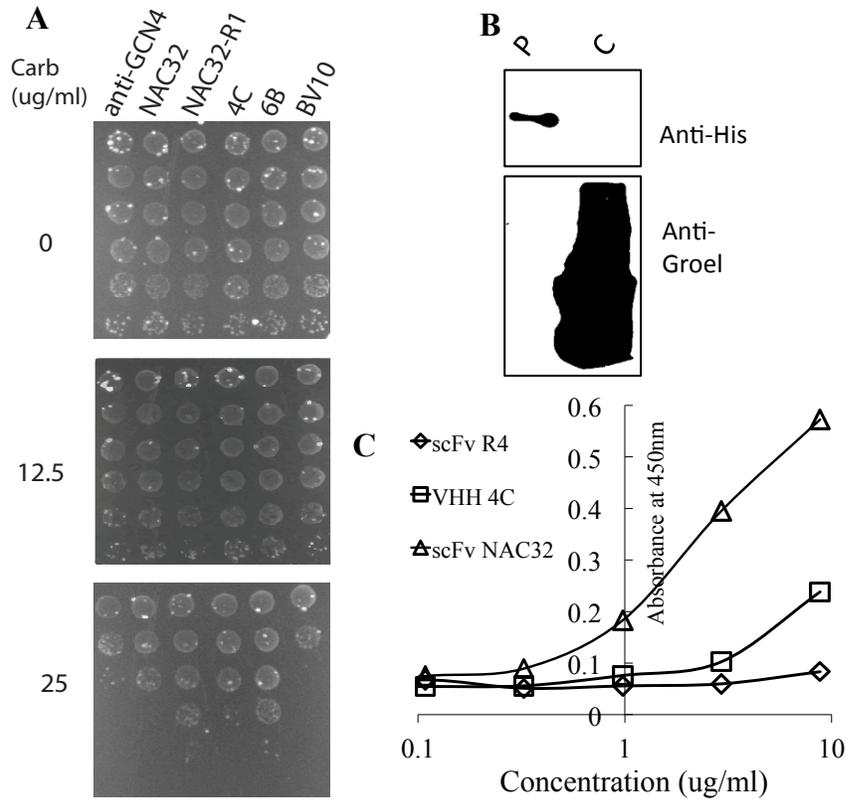


Figure 3.3 Selection of candidate vHH, expression and binding A) Selective plating of serially diluted cells expressing different antigen-antibody pairs on Carb as indicated. Each spot corresponds to 5uL of serially diluted cells. B) Western blot analysis of periplasmic and cytoplasmic fractions generated from WT cells carrying pSAlect and expressing ssTorA-vHH-FLAG. Detection of vHH constructs was with anti-FLAG antibody, and detection of GroEL using anti-GroEL antibody and served as a fractionation marker. C. ELISA binding assay shows vHH binds A53T.

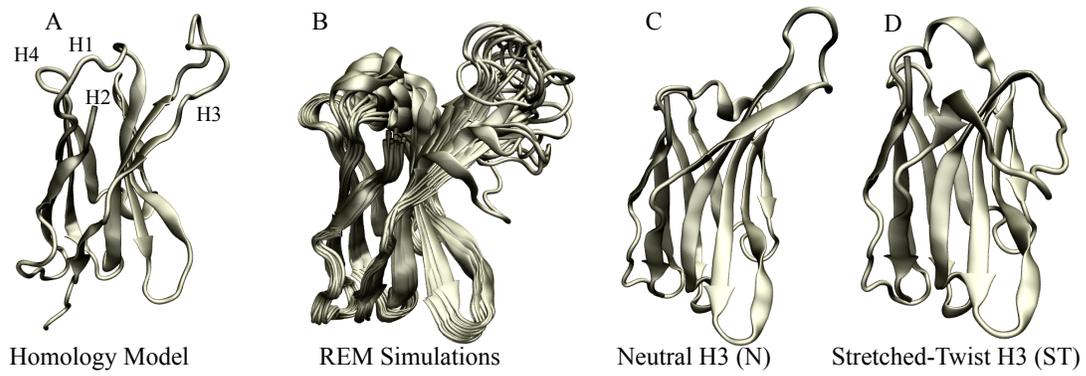


Figure 3.4 Model for vHH 4C. A. Homology Model for vHH 4C based on PDB: 3P0G. B. Enhanced conformational sampling in REM simulations C. Top two conformations obtained from REM simulations (N and ST).

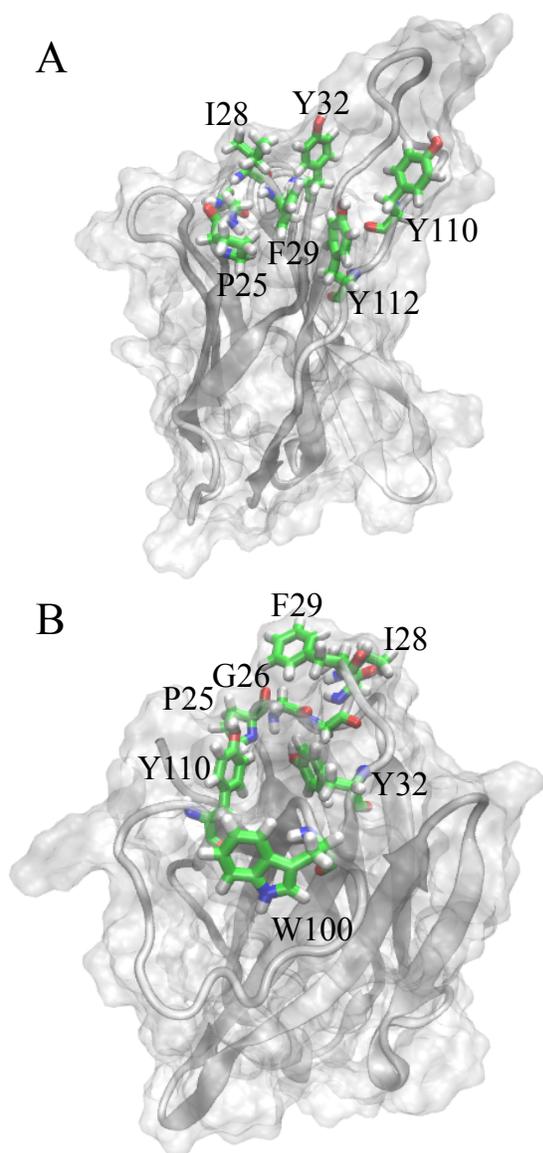


Figure 3.5 Hydrophobic residues clustered at the interface of H1 and H3 loops for vHH 4C N and ST models. The Solvent Accessible Surface (probe radius 1.4 Å) and the Cartoon model are shown. Only the hydrophobic residues are shown as Licorice with element C in green, H in white, N in blue and O in red.

Model for vhh-NAC complex

The two conformations of vHHs (N, ST) and 2 models of NAC (alpha-helical (AH) and random-coil (RC), 19-residues) were used as starting structures to generate docked structures

using the ClusPro docking software⁸⁹. 38 structures were generated (4 possible permutations and ~10 structures per vhh-NAC pair). In order to rank these structures from the most stable to the least stable, all structures were energy minimized and equilibrated for 5ns in the NVT ensemble at 300K. Only the complexes that were stable in the equilibration runs were retained for further simulations. For each vhh-NAC pair, a separate REMD was conducted. REMD was used as a technique to rank the complexes from the most stable (lowest temperature box 300K; also temperature of interest) to the least stable (highest temperature 450K). Three stable structures were obtained for N-AH and two for ST-AH. None of the complexes generated for the RC model from ClusPro were stable in the equilibration simulations. This could be a result of the unwieldy nature of the RC peptide generated by piecing together residues. To generate candidate structures for the N-RC and ST-RC, a 13-residues peptide (66-78) was used to generate candidate structures from ClusPro followed by equilibration and REMD as described above. After top structures for the complexes were obtained from REMD simulations, the 13-residue peptides were extended to longer peptides as desired.

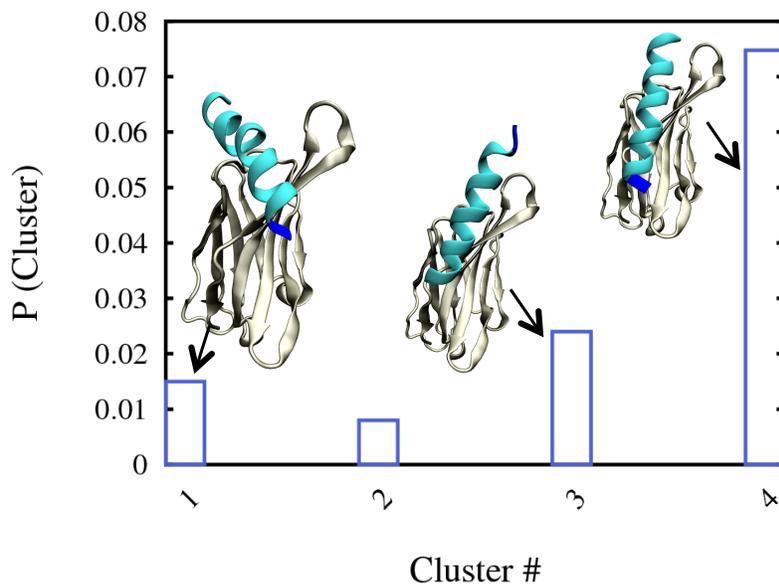


Figure 3.6 Probability of occurrence of different complexes in the lowest temperature replica in REMD Simulations for the vHH4C exhibiting conformation N and Alpha-Helical Ag (N-AH).

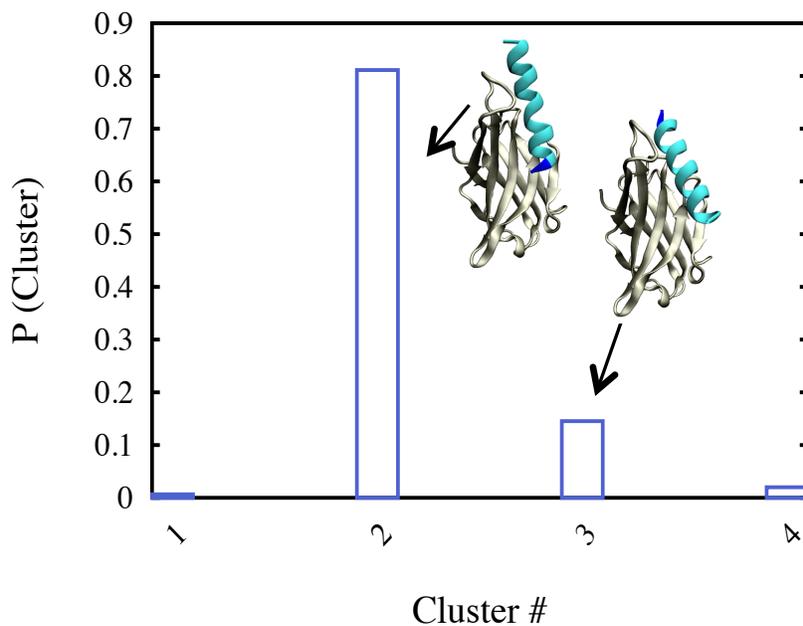


Figure 3.7 Probability of occurrence of different complexes in the lowest temperature replica in REMD Simulations for the vHH4C exhibiting conformation N and Alpha-Helical Ag (N-AH) for “newsite”.

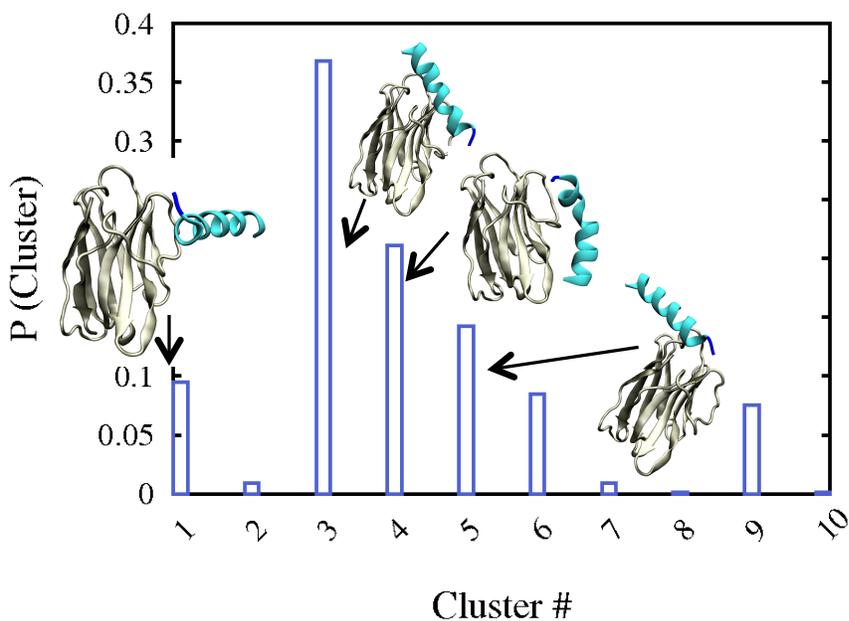


Figure 3.8 Probability of occurrence of different complexes in the lowest temperature replica in REMD Simulations for the vHH4C exhibiting conformation ST and Alpha-Helical Ag (ST-AH).

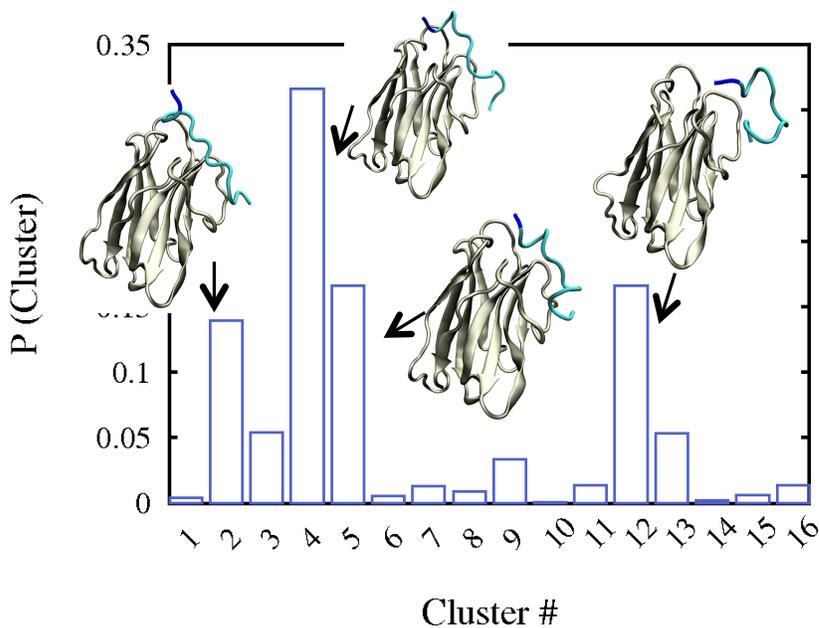


Figure 3.9 Probability of occurrence of different complexes in the lowest temperature replica in REMD Simulations for the vHH4C exhibiting conformation ST and Random Coil Ag (ST-RC).

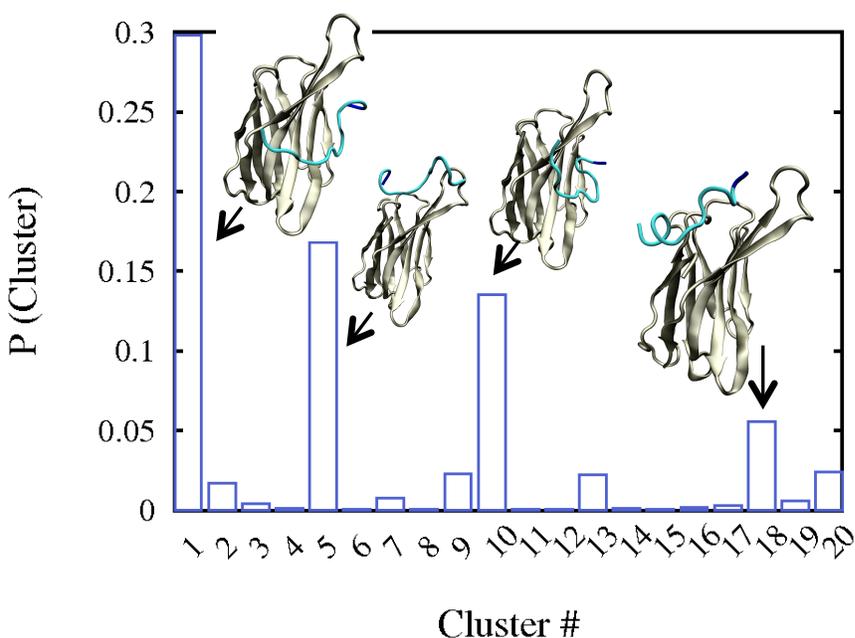


Figure 3.10 Probability of occurrence of different complexes in the lowest temperature replica in REMD Simulations for the vHH4C exhibiting conformation N and Random Coil Ag (N-RC).

3.4 Results and Discussion

Selection of a vHH Antibody against A53T NAC peptide

To obtain a candidate vHH, an immunized library from an Alpaca was subjected to the Twin Arginine translocation (TAT)-based selection technique FLI-TRAP⁷⁷ (Figure 3.3). The immunized library was generated against the NAC region of A53T by the Messer Group in Albany. This immunized library was then subjected to a single round of FLI-TRAP. The FLITRAP technique is a TAT-based selection technique which has been previously used to screen for folded, soluble and antigen-specific scFvs. Briefly, FLI-TRAP entails one plasmid, encoding an ssTorA-X-FLAGtag, and a second plasmid, encoding the Y-βlactamase fusion. The vhh library was cloned at the X position and A53T was cloned into the Y position. The ssTorA

signal sequence directs the vHH to the TAT pathway. The β lactamase (Bla) fused to A53T acts as a reporter protein conferring carbenicillin resistance when localized in the periplasm. If the vHH binds to A53T in the cytoplasm, the whole complex is translocated across the Tat pathway to the periplasm conferring carbenicillin resistance to the cells.

6 positive clones were thus isolated and sequenced. Sequencing revealed 2 unique full-length sequences – vHH 6B (6B) and vHH 4C (4C). Growth phenotype was further confirmed by retransforming the isolated clones and spot-plating. scFvs previously found to bind A53T – scFv-NAC32 and scFv-NAC32-R1⁹⁰ – were used as positive controls, a randomly chosen vHH from the immunized library – vHH-BV10 and a non-specific scFv, anti-GCN4⁹¹ were used as negative controls. The FLI-TRAP hits vHH 4C and vHH 6B confer higher carbenicillin resistance compared to the negative controls and comparable to the positive controls. The results are summarized in 3.3A.

The soluble expression of vHhs 4C and 6B was tested by expression of the ssTorA-vhh-FLAGtag construct in DH5alpha cells. Only a soluble and folded vHH will be translocated to the periplasm by the TAT pathway. In Figure 3.4B, we show the anti-FLAG blots of the subcellular fractions. vHH 4C is found in the periplasm and the fractionation is verified by anti-GroEL blot in which the GroEL (subcellular marker) is localized in the cytoplasm as expected. 6B did not express well and was not seen on the Western Blot. vHH 4C was further found to bind A53T in an ELISA assay with an affinity lower than scFv NAC32 but higher than a non-specific scFv (Figure 3.4C). Hence, vhh 4C was chosen as a suitable candidate for affinity improvement using computational methods.

Developing an atomistic model for vhh 4C and its complex with NAC peptide

By using REMD simulations for enhanced conformational sampling of the loop regions, we obtained the two main structures for the binding loops. Notably, the H3 loop exhibits Neutral (N) (42% of all conformations sampled) and a stretched-twist (ST) (22% of all conformations sampled) conformations in the two main structures (Figure 3.5). Both structures were retained for further simulations and analyses.

The N and ST conformations lead to two very distinct binding surfaces as shown in Figure 3.5. In both conformations, we see a clustering of hydrophobic residues at the interface of H1 and H3 loops (I28, F29, Y32, Y110, Y112; see Figure 3.6). This is a preliminary indication of the possible location of the paratope on the vHH. Both conformations exhibit groove-like cavities, as one would expect for proteins that bind to peptides.

Top models for the complex from each REMD simulation are summarized in Figures 3.7, 3.8, 3.9, 3.10 and 3.11. The top 5 models (top 1-2 models selected for each case) for the complex are shown in Figure 3.12. To characterize the models, we obtained the average energy of interaction between the vHH and the Antigen peptide (Ag) for each model. The energy of interaction can be calculated separately for the VanderWaals (VdW)-type interactions and Coulombic interactions. The results are shown in Figure 3.13. For the sake of comparison, we also calculated the energy of interaction for a reference case. We used the anti-CemX antibody⁹² as the reference case. In this case, an antibody binds a disordered peptide with a nanomolar K_D . For the reference case, the total energy of interaction is significantly higher than any of our models (Figure 3.13). Moreover, the coulombic interactions contribute significantly larger energy to binding than the VdW interactions. This is probably due to the large number of hydrophilic

and charged residues on the target peptide CemX unlike the hydrophobic NAC peptide. This observation and other factors described in the next Chapter led us to the idea to optimize coulombic interactions at the vhh-NAC interface to enhance affinity.

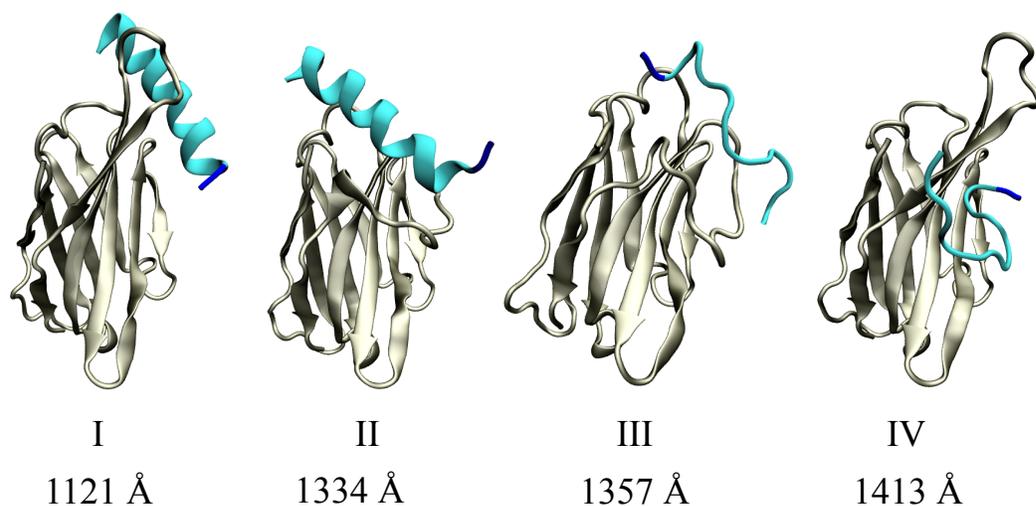


Figure 3.11 Top 4 complexes obtained from Simulations for the vHH-NAC complex. Complexes I and II were obtained for a 19-residue peptide (AAs 61-78) and Complexes III and IV were obtained for a 13-residue peptide (AAs 66-78). The vHH is shown in metallic Tan and the NAC peptide is shown in Cyan.

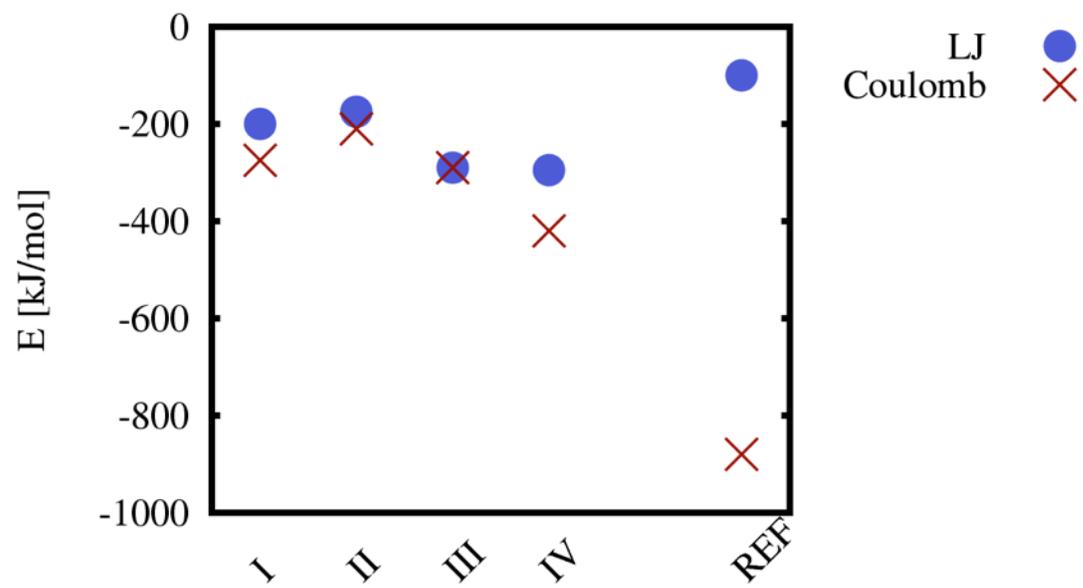


Figure 3.12 Energy of interaction between vHH-Ag for the 4 models and the reference Ab-Ag case.

4 IMPROVING AFFINITY OF BINDING USING RATIONAL DESIGN

4.1 Introduction

A combination of advanced Molecular Dynamics (MD) techniques (Replica Exchange Molecular Dynamics (REMD), Umbrella Sampling (US), Weighted Histogram Method (WHAM)), numerical techniques (Finite Difference form of Poisson Boltzmann Equation (PBEQ)) and experimental techniques (high throughput screening, binding assays, Surface Plasmon Resonance (SPR)) were employed to build a coherent picture of the vhh-NAC binding.

4.2 Methods Simulations

Umbrella Sampling to Obtain Potential of Mean Force

US simulations were used to obtain PMF of binding as a function of the radial distance between the vHH and the Ag (Figure 4.1). The initial configurations for the US windows were obtained by using the Pulling code in GROMACS. ~80-96 windows were placed at a distance of 0.03 nm for windows at small separations (<2.5nm) to 0.10 nm for windows at large separations (>4nm) between the vhh and the Ag. Each window was run for 5-15ns depending on the time it took for the FE profiles to converge within each window. Overall, the simulations were run for a cumulative time of 500ns to 1.5 μ s. The g_wham⁹³ program was used to obtain the PMF as a function of distance using WHAM⁹⁴. The idea of WHAM is to estimate the statistical uncertainty of the unbiased probability distribution $P(\xi)$ along the reaction coordinate or order parameter ξ given the umbrella histograms, and subsequently to compute the PMF that corresponds to the smallest uncertainty. The WHAM equations read,

$$P(\xi) = \frac{\sum_{i=1}^{Nw} g_i^{-1} h_i(\xi)}{\sum_{j=1}^{Nw} n_j g_j^{-1} \exp[-\beta(w_j(\xi) - f_j)]} \quad 4.1$$

and

$$\exp(-\beta f_j) = \int d\xi \exp[-\beta w_j(\xi)] P(\xi) \quad 4.2$$

Here, β denotes the inverse temperature $1/k_B T$, with the Boltzmann constant k_B and the temperature T , and n_j is the total number of data points in histogram h_j . The statistical inefficiency g_i is given by $g_i = 1 + 2\tau_i$, with the integrated autocorrelation time τ_i of umbrella window i (in units of the simulation frame time step). If all windows have the same autocorrelation time then the g_i term cancels out. Otherwise the g_i term assigns lower weights to windows with longer correlation times. The umbrella sampling scheme shown in Figure 4.1 implemented to obtain the PMF of binding along the separation (r) between the vHH and the antigen (Ag). The PMF can be considered as a one dimensional projection of the three-dimensional free-energy of binding.

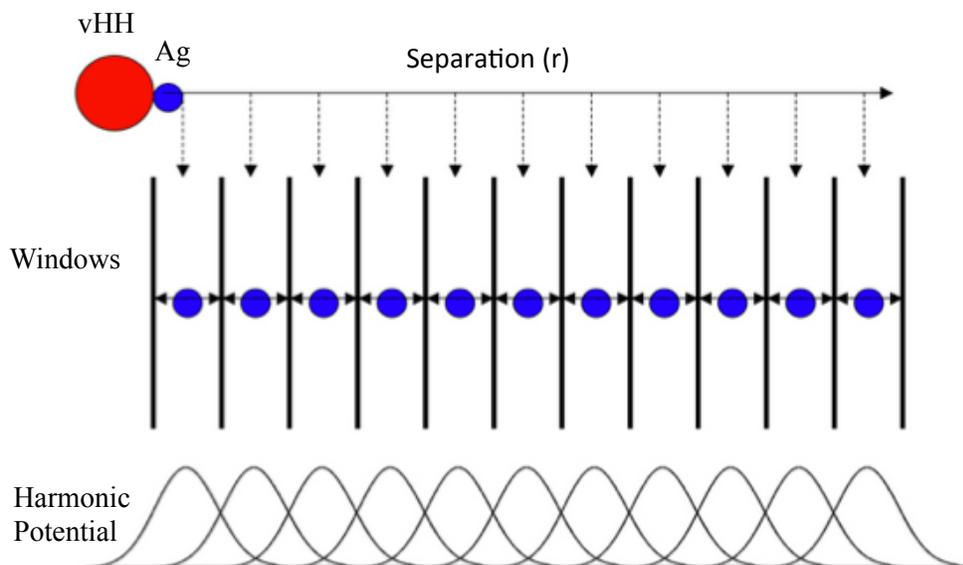


Figure 4.1 Schematic for Umbrella Sampling (US) scheme to obtain Potential of Mean Force (PMF)

PBEQ

The Delphi Software⁹⁵⁻¹⁰⁰ was used to calculate the electrostatic free energy of binding. It can solve the Finite Difference form of the Poisson Boltzmann distribution.

The theory below was compiled from a number of references listed here¹⁰¹⁻¹⁰³.

The Poisson-Boltzmann Equation is based on the fundamental equation of electrostatics, the Poisson equation:

$$\nabla^2 \varphi(\vec{r}) = -4\pi\rho(\vec{r})/\epsilon \quad 4.3$$

where φ is the potential, ρ is the charge density, ϵ is the dielectric constant, and \vec{r} is the position vector and the potential is given by,

$$\varphi(\vec{r}) = \sum_i q_i / |\vec{r} - \vec{r}_i| \quad 4.4$$

If a material responds in an average way to electric field it has a uniform susceptibility, where \vec{P} is the polarizability and χ is the susceptibility and \vec{E} is the electric field.

$$\vec{P} = \chi \vec{E} \quad 4.5$$

$$\epsilon = 4\pi\chi + 1 \quad 4.6$$

If ϵ varies through space, then the Poisson equation is modified as,

$$\nabla \cdot \epsilon(\vec{r}) \nabla \varphi(\vec{r}) = -4\pi \rho(\vec{r}) \quad 4.7$$

The above equation is also valid when the potential and charge density are mean values given by some statistical mechanical calculation

$$\nabla \cdot \epsilon(\vec{r}) \cdot \langle \nabla \varphi(\vec{r}) \rangle + 4\pi \langle \rho(\vec{r}) \rangle = 0 \quad 4.8$$

Where the brackets $\langle \rangle$ denote mean value. The mean values represent averaging over all possible conformations in solution. When mobile ions are present, the charge density can be written as $\rho = \rho^m + \rho^f$, where the superscripts m and f refer to the mobile and fixed charge components. In this context, what is meant by mobile is that the distribution of these charges is affected the thermal motions i.e. their mean position is one of the variables to be determined, along with the potential. In contrast, the term fixed indicates that the charges are not subject to this averaging;

i.e., their positions are assumed to be known exactly, or at least form one of the inputs to the calculation.

In biophysical applications the source of the fixed charges (ρ^f) are usually the macromolecules. In addition, it is assumed that the amount of fixed charge is constant; i.e. we consider the case of constant surface charge, as opposed to constant surface potential.

To determine the distribution of mobile ions we require at equilibrium that the chemical potential of each ion species μ_i is uniform throughout the solution

$$\mu_i(\vec{r}) = \mu_i^0 + k_B T \ln a_i(\vec{r}) + z_i e \varphi(\vec{r}) = \mu_i^0 + k_B T \ln a_i^b \quad 4.9$$

where μ_i^0 is the standard-state chemical potential, e is the proton charge, T is the absolute temperature, z_i is the valence of the ion, a_i is the activity of the ion in the bulk solution, i.e., far from the colloid where the potential is zero. It is usually assumed that the activity coefficient is unity and independent of potential and concentration; thus $a_i = c_i$, the concentration. This assumption is equivalent to the neglect of the mutual Debye-Huckel interaction between the ionic atmosphere ions. This approximation is later referred to as the Boltzmann approximation. The above equation can now be rewritten in the familiar-form of the Boltzmann expression (wherein the probability density of ions – here concentration- depends on the exponential of the electrostatic potential) as

$$\langle c_i(\vec{r}) / c_i^b \rangle = \exp(-z_i e \langle \varphi(\vec{r}) \rangle / kT) \quad 4.10$$

Where the exponent is an estimate of the potential of mean force between the ion and the colloid. Assuming only a 1-1 salt, for which the PB equation is most applicable. For this case this leads to the NLPB equation (in terms of the dimensionless potential $\phi = e\varphi/kT$)

$$\nabla \cdot (\vec{r}). \nabla \phi(\vec{r}) + f(\phi) = 0 \quad 4.11$$

where

$$f(\phi) = -\epsilon \kappa^2 \sinh(\phi(\vec{r})) + 4\pi e \rho^f(\vec{r})/kT \quad 4.12$$

and $\kappa^2 = 1/\lambda^2 = 8\pi e^2 I / \epsilon kT$, where λ is the Debye length, and I is the ionic strength of the bulk solution. Analytical or numerical solutions of this equation yield the potential distribution throughout space. Sharp and Honig¹⁰² have used the calculus of variations to derive the function F such that it satisfies the following equations:

$$f(\phi) = \partial F / \partial \phi \quad 4.16$$

Rewriting the Poisson-Boltzmann equation in terms of F,

$$\begin{aligned} -\nabla \cdot (\vec{r}) \nabla \phi(\vec{r}) \\ = \partial(\partial F / \partial \phi_x) / \partial x + \partial(\partial F / \partial \phi_y) / \partial y + \partial(\partial F / \partial \phi_z) / \partial z \end{aligned} \quad 4.13$$

Where ϕ_x, ϕ_y, ϕ_z are the partial differentials of potential with respect to the three spatial variables, x, y, z . A theorem from the calculus of variations states that, for any differential equation that can be put in this form, F is the unique function whose integral over the independent variables x, y and z is the minimized solution to the differential equation. Since the condition for equilibrium is that the free energy of the system be at a minimum, the electrostatic free energy can be identified as,

$$\Delta G_{el} = \int F(x, y, z, \phi, \phi_x, \phi_y, \phi_z) dv \quad 4.14$$

Where dv is the volume element and the integral is over all space. For NLPB, it is straightforward to show that¹⁰²,

$$F = 4\pi\rho^f\phi - \epsilon\kappa^2 \cosh(\phi) - \epsilon(\nabla\phi)^2/2 \quad 4.15$$

Substituting back into the expression for ΔG_{el} , and adjusting units and setting the reference state of zero energy at $\rho^f = 0$, we get,

$$\Delta G_{el} = \int (\rho^f\phi - kTc^b\{2\cosh(\phi) - 2\} - \vec{E} \cdot \vec{D}/2)dv \quad 4.16$$

where the integral is over all space, ΔG is in units of kT , $\vec{D} = \epsilon\vec{E}/4\pi$ is the electric displacement, and c^b is the bulk salt concentration. The term $c^b\{2\cosh(\phi) - 2\}$ is the excess concentration of ions at any point in solution compared to bulk. Rewriting this term as $\sum_i \Delta\pi_i$,

$$\Delta G_{el} = \int (\rho^f - \sum_i \Delta \pi_i - \vec{E} \cdot \vec{D}/2) dv \quad 4.17$$

And for Linear Poisson-Boltzmann (LPB) assuming that $\phi \ll 1$, when the exponential term may be expanded and truncated at the linear term giving the LPB equation,

$$\Delta G_{el} = \int (\rho^f \phi + \rho^m \phi/2 - \vec{E} \cdot \vec{D}/2) dv \quad 4.18$$

Now consider the NLPB. The expression can be further simplified by using the well-known expression¹⁰¹,

$$\int \rho^{free} \phi/2 dv = \int \vec{E} \vec{D}/2 dv \quad 4.19$$

In the above expression, the right hand side is the work required to “assemble a system” starting with an unpolarized dielectric in place, moving the free charges and allowing the dielectric to respond as it sees fit. The left hand side is the more explicit expression for the work required to assemble a charge density ρ^{free} (note that the superscript *f* denotes fixed and the superscript *free* denotes free) in a volume dv with a potential ϕ . Also, this expression is strictly true for linear dielectrics.

Also, the free (non polarization) charge is the sum of the fixed charge on the molecule and the ionic charge in the solvent, $\rho^{free} = \rho^{fixed} + \rho^{solv}$

Plugging this back into the NLPB equation, we get

$$\Delta G_{el} = \int \left(\frac{\rho^f \phi}{2} - \sum_i \Delta \pi_i - \rho^{solv} \phi / 2 \right) dv \quad 4.20$$

In the above expression, the second term is an osmotic pressure (due to excess ions) and the third is an electrostatic stress. It has been shown that in the case of linearized PB equation, the last two terms cancel out so that the free energy is given by the first term.

For the solution of the finite difference form of the LPB, the first term can be rewritten and expanded in the form,

$$\frac{\rho^f \phi}{2} = \frac{1}{2} \sum_j q_j \phi(\vec{r}_j) \quad 4.21$$

Where the potential is the one generated by all the charges, of any kind, except for the one located at \vec{r}_j . The software Delphi uses the NLPB equation (Finite-difference solution). The software solves the LPB first and uses the solution as a starting point for the NLPB. Within Delphi, there are two methods to calculate ΔG_{el} :

1. Grid-based: The charge at a grid point is multiplied by the potential at that point. This however introduced precision errors as the charge distribution at a grid point depends on the grid spacing used. The results thus obtained are found to be rather sensitive to grid spacing.
2. Partitioning-Based: The potential at a given point is instead partitioned into the potential due to 1) the direct effect of real charges, 2) from the surface polarization charges

(reaction field term) and from mobile ions in solution. Thus, the potential at the position of charge can be written as

$$\varphi(\vec{r}_j) = \varphi_{coul}(\vec{r}_j) + \varphi_{react}(\vec{r}_j) + \varphi_{ion}(\vec{r}_j) \quad 4.22$$

The Coulombic potential, generated by other fixed charges at the j -th charge position, is calculated in a hypothetical homogeneous media with the dielectric constant ϵ_i of the region where charge is located

$$\varphi_{coul}(\vec{r}_j) = \sum_{i \neq j} \frac{q_i}{4\pi\epsilon_0\epsilon_i|\vec{r}_j - \vec{r}_i|} \quad 4.23$$

The “corrected reaction field” term, arising from the polarization of the boundary between different media is given by

$$\varphi_{react}(\vec{r}_j) = \sum_p \frac{\delta_p}{4\pi\epsilon_0|\vec{r}_j - \vec{r}_p|} \quad 4.24$$

Where p goes over the locations of polarization surfaces and δ_p is the polarization charge at position r_p on the surface. These polarization charges are calculated using Gauss’s law and the potential map obtained from the finite difference algorithm.

The last term accounts for the potential which is generated by the mobile ions in solution. Under the Boltzmann approximation (Equation 4.9), these charges are a function of the potential,

obtained by solving the PB equation, and are screened by the polarizable solvent, whose dielectric constant is $\epsilon = \epsilon_{solv}$

$$\varphi_{solv}(\vec{r}_j) = \sum_k \frac{h^3 d\rho_k^{ion}}{4\pi\epsilon_0\epsilon_{solv}|\vec{r}_j - \vec{r}_k|} \quad 4.25$$

where h is the grid spacing, $d\rho_k^{ion}$ is the net (excess) ion charge density at the k -th grid point in solution, and k goes over the grid points in solution.

The procedure outlined above makes it possible to use real charges rather than their grid representation in calculating the free energy of the system. Once the surface and excess ion charges are calculated, then the potential at the position of all fixed charges can be obtained from the above three terms. This leads to major improvements in precision because the calculation of free energies involves only using Coulomb's law for real, polarization and mobile charges. This also eliminates the so-called grid energy, which is an artificial energy term that comes from partitioning the point charges at neighboring lattice points.

In Delphi, the LPB is first solved and then non-linearity is gradually added. This is accomplished by adding the non-linear term multiplied by 0.05 to the initially linear equation. Thus, after 20 blocks of 10 iterations each, the algorithm begins to solve the full nonlinear equation.

For all our calculations using Delphi, we used the following inputs:

- 1) PBEQ was solved for the complex (and not the encounter complex). This method has been previously implemented with success¹⁰⁴.
- 2) PARSE parameters were used for partial atomic charges and radii.
- 3) A dielectric constant of 4 was used for protein and explicit water, and 80 for implicit solvent regions. Ionic strength was set at 0M (no salt) and 0.150M (with salt), modeled

with a 2.0-Å Stern Layer and a molecular surface generated with a 1.4-Å probe sphere. The number of grid points per Angstrom (gpa) was set to 1.5.

For all calculations, electrostatic free energy was calculated for 50-100 structures and the estimated value was obtained by averaging over them. The electrostatic free energy of binding is calculated as:

$$\Delta\Delta G_{el} = \Delta G_{el}^{Complex} - \Delta G_{el}^{vHH} - \Delta G_{el}^{Ag} \quad 4.26$$

And the electrostatic free energy for the complex, vHH and Ag are calculated using the partition-based method described above.

Mutagenesis Strategy

Initial US free energy profiles suggested the importance of long-range electrostatic interactions between the vHH and the Ag. Furthermore, from SPR, we found that vHH 4C has an unusually small rate of association. The association rate improves on reducing the salt in the buffer. This indicates the importance of long-range electrostatic interactions. From the potential energy surfaces of the vHH and Ag obtained from PBEQ calculations, we were able to identify 2 residues – Lysine (K) at position 75 and K at position 99 which could be responsible for long-range repulsive interactions. We thus suggested the mutants K75A and K75E and K99Q (Q for Glutamine). Also mutations were introduced at position 77, which commonly accompany substitutions at position 75 in vHHs. For example, while K is a highly conserved residue at position 75, GeneBank:gbAGH30283, PDB:3POG and PDB:4LDE all occur with either an Glutamate (E) or an Alanine (A) at this position, all of which are accompanied by a Threonine

(T) at position 77. The additional mutation at position 77 might alleviate any steric hindrance and mimic the naturally occurring residues in vHHs. Two mutants, AT and ATQ, were found to be significantly better than vHH 4C from PBEQ free energy analysis. The PMFs for vHH 4C and both these mutants were calculated using US to quantify the change in binding FE upon mutations. The improvement was found to be significant and both mutants were prepared and their dissociation constants and rates of association and dissociation quantified using SPR. Contrary to expectation, the K_D worsened and so did the rate of association. Hence, by introducing these mutations, we had made the long-range interactions more repulsive.

This led to the conclusion that the model under consideration (model III; Chapter 3 PREF), may not be the right one. We scanned the all the REMD results (not just the top model from each REMD) and considered two models (II, V-new) that may be able to explain the worsening of K_D upon making the above mutations.

We ran 2-5ns MD simulations for vHH 4C and mutants K75A+M77T and K75E. From these simulations, we obtained contribution of the new mutant residue at the vHH-Ag interface to the energy of binding. Mutants were compared to vHH 4C and mutations that led to a more favorable energy contribution were selected. We found N76D (Asparagine to Aspartate) to have favorable contributions. Further mutations were made to the N76D residue. Mutations were made in the vicinity of the D76 residue in order to improve the affinity of binding. Two mutations were found to be promising – Y72T (Y for Tyrosine) and Q81R (R for Arginine).

4.3 Methods Experiments

Expression, Purification and Size Exclusion Chromatography (SEC)

After expressing the vHHs in the cytoplasm and obtaining soluble fractions as described above, we purified 6x-His-tagged vHHs using protocols provided with the Ni-NTA protein purification resin (Qiagen) using a gravity column. Soluble lysates were incubated with the resin for 1 h at 4°C on a rotating platform to improve mixing. The lysates with the resin were poured into columns, until all the lysate and resin flowed into the column. The flow-through was poured again over the resin and allowed it to flow-through again. The columns were washed four times with a buffer containing 20mM NaH₂PO₄, 300mM NaCl, and 90mM imidazole (pH 7.5). vHHs were eluted using a buffer containing 20mM NaH₂PO₄, 300mM NaCl, and 250mM imidazole (pH 7.5). After affinity purification, the elutes were concentrated followed by SEC. The 2.5ml of concentrated protein (1ml at a time) was injected into a Superdex 75 column (GE Healthcare) attached to a FPLC system (GE Healthcare) and samples were run at a flow-rate of 0.5 mL/min at 4 °C in buffer containing 15mM Tris-HCl, 500mM NaCl. The high salt concentration was used to prevent aggregation in the concentrated vHHs. The migration of vHHs was monitored by absorbance at 280nm, and 250 µL fractions were collected.

Binding rates and dissociation constant determination using SPR

SPR experiments were performed using a Biacore 3000 biosensor (GE Healthcare). Purified vHHs were immobilized on CM5 sensor chips at 10 µg/ml using an amine coupling kit at pH 4.0 until reaching 1200 RU, according to the manufacturer's instructions. The binding analyses were carried out in HBS-EP Buffer (GE Healthcare Life Sciences) at a flow rate of 30 µl/min. Each

vHH protein was associated with ligands for 3 min, and then dissociation phases were observed. During SPR experiments, flow cells were regenerated by injection of 10 mM glycine (pH 4.0) followed by thorough washing with the running buffer. Additionally, a blank cell on the same sensor chip was used as a reference to correct for non-specific binding. Kinetic parameters were calculated using BIAevaluation software 3.2. All sensorgrams were fitted to a 1:1 Langmuir binding model using a simultaneous non-linear program.

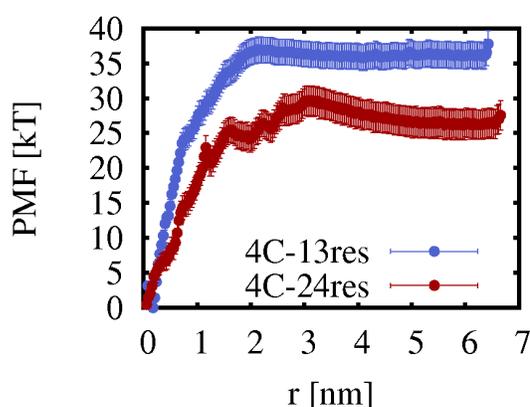


Figure 4.2 US Results for Model III for the 13-residue (hydrophobic only) and 24-residue peptides for the Ag.

4.4 Results

Utilizing hydrophilic residues in the NAC peptide to enhance energy of binding:

Targeting electrostatic interactions at the binding interface

There are multiple factors to consider in when aiming to improve binding to a disordered protein (assuming it is not accompanied by a folding event) – 1) there is a considerable loss of conformational entropy of the Ag peptide upon binding 2) some of this decrease in conformational entropy will be compensated for by the gain in entropy upon the desolvation of

the hydrophobic residues in the Ag peptide 3) ultimately, the gain in enthalpy due to new contacts formed at the interface should compensate for the net loss in entropy to lead to a favorable free-energy of binding.

The NAC region is primarily hydrophobic. However, the hydrophobic peptide (AAs 66-78 VGGAVVTGVTAVA) is flanked on the N-terminus by peptide P1 (AAs: 57-65 EKTKEQVTN) and on the C-terminus by peptide P2 (AAs: 79-83 QKTVE). One strategy to improve binding would be to maximize the potential for electrostatic contacts between the vHH and the Ag. Compared to hydrophobic interactions, energetic contributions from electrostatic interactions are much more significant. This strategy might work particularly well for targeting the NAC region, which is primarily composed of hydrophobic residues and thereby leads to low-energy binding surfaces. To test this hypothesis from a purely theoretical perspective, we ran umbrella sampling (US) simulations to obtain an estimate for the Potential of Mean Force (PMF) as a function of separation between the vhh and the Ag for two versions of the Ag – 1) only the 13-residue hydrophobic peptide (66-78), 2) 24-residue peptide (57-80). Sample simulations were run for model III. For this model, US simulations revealed a less favorable complex with the 24-residues peptide (Figure 4.2). This implied that the electrostatic interactions might be suboptimal and thereby leading to a less favorable PMF for complex formation for this complex. As a first step towards improving affinity, we aimed to minimize this apparent electrostatic repulsion between the vHH and Ag. To analyze the electrostatics interactions, we used the PBEQ implemented in the Delphi software.

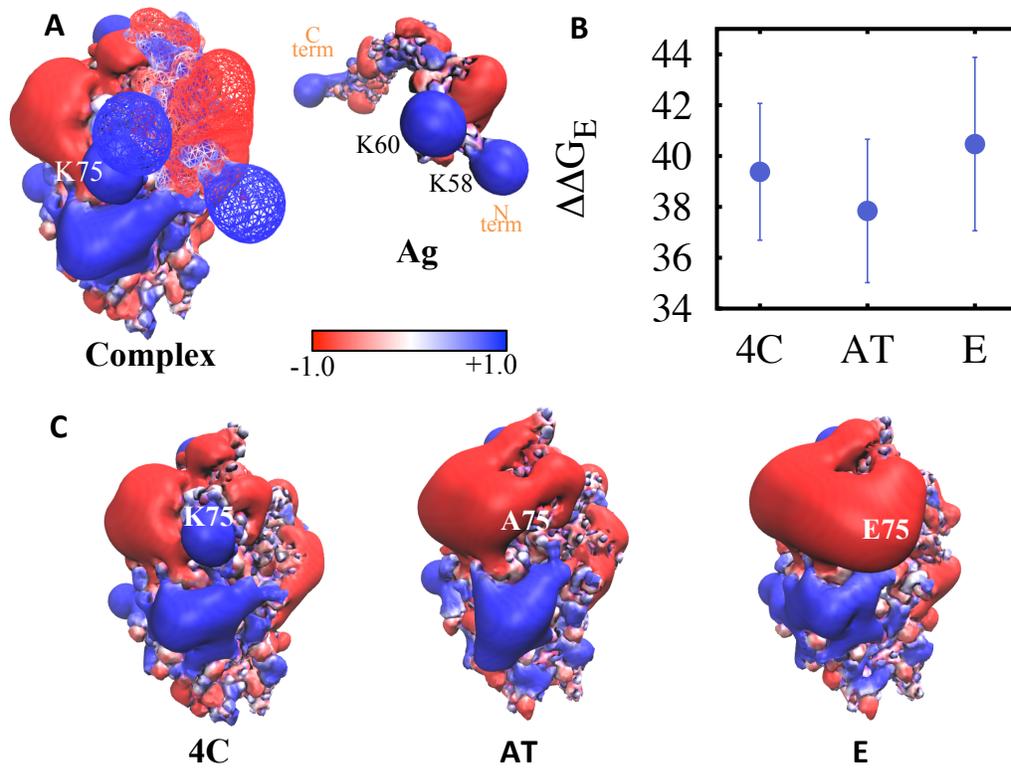


Figure 4.3 Electrostatic Potential of Binding Surfaces for Model III A) Complex of vHH 4C with Ag. The Complex and the Ag are shown. In the complex, the vHH is shown with a solid surface and the Ag is shown with a wire-mesh. In the Ag, the surface is shown as a solid surface. Blue indicates an electrostatic potential of 1 kT/e and red indicates an electrostatic potential of -1 kT/e. B) The electrostatic FE of binding calculated using PBEQ for vHH 4C and mutants AT and E. C) Electrostatic Potential of vHH 4C and mutants AT and E.

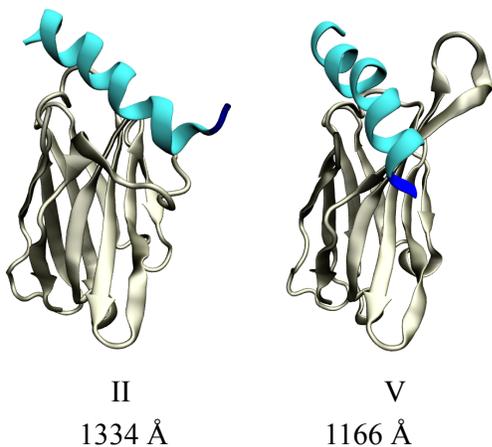


Figure 4.4 Putative models to explain SPR results for K75AM77T and K75E mutants

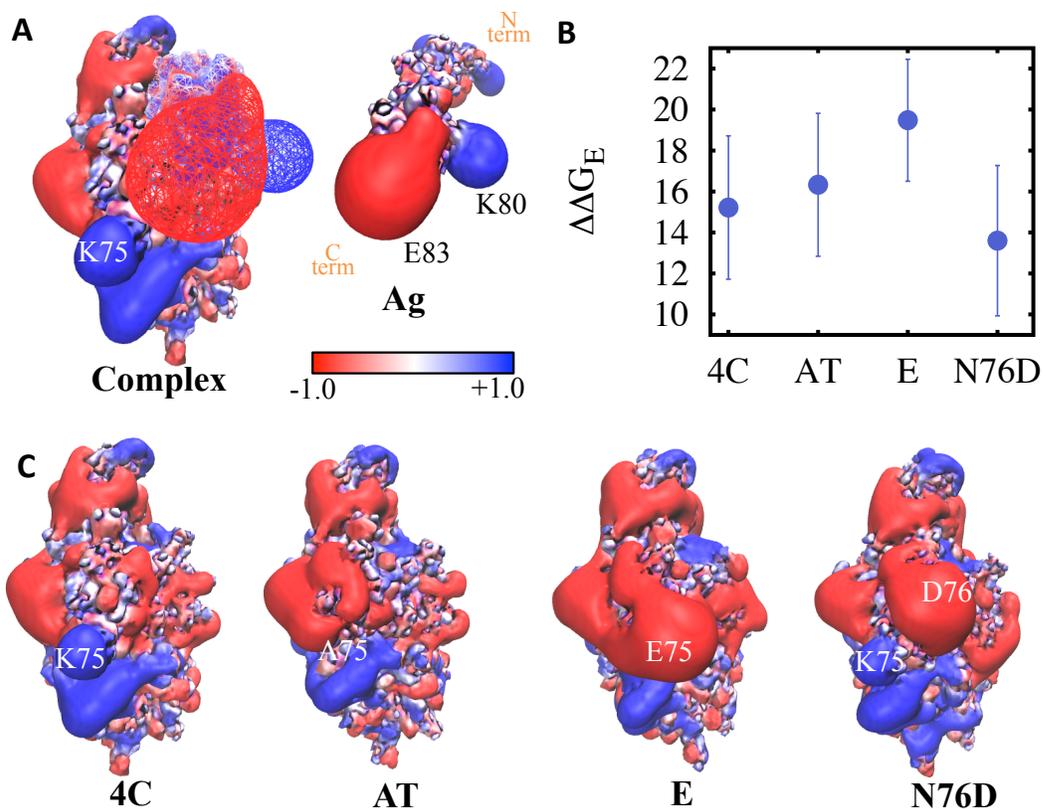


Figure 4.5 Electrostatic Potential of Binding Surfaces for Model V A) Complex of vHH 4C with Ag. The Complex and the Ag are shown. In the complex, the vHH is shown with a solid surface and the Ag is shown with a wire-mesh. In the Ag, the surface is shown as a solid surface. Blue indicates an electrostatic potential of 1 kT/e and red indicates an electrostatic potential of -1 kT/e. B) The electrostatic FE of binding calculated using PBEQ for vHH 4C and mutants AT, E and N76D. C) Electrostatic Potential of vHH 4C and mutants AT, E and N76D.

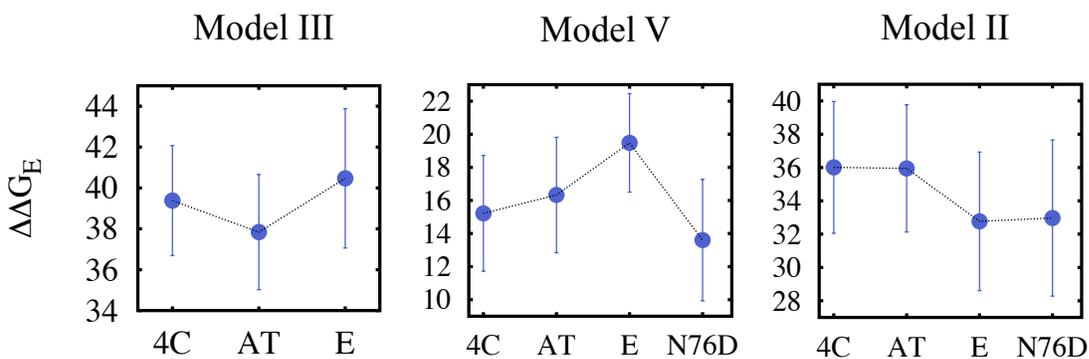


Figure 4.6 The electrostatic FE of binding calculated using PBEQ for vHH 4C and mutants AT, E and N76D for Model III, Model V and Model II.

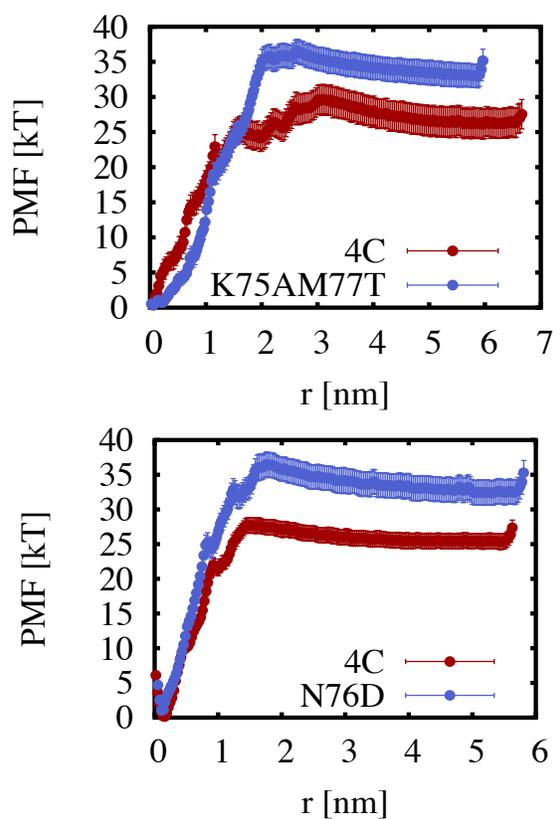


Figure 4.7 US Results for Model III (Top) and Model V (Bottom)

Table 4.1 SPR results for vHH 4C and mutants

vHH	Salt Concentration (mM)	k_A (1/Ms)	k_D (1/s)	K_D (M)	χ_2
4C	150mM	1.31E+03	4.22E-03	3.22E-06	1.18
AT	150mM	11.9	0.011	9.23E-04	0.424
E	150mM	3.92	5.36E-03	1.37E-03	1.29
4C	75mM	1.36E+04	3.3E-03	2.43E-07	2.11
AT	75mM	3.43E+03	0.0101	2.96E-06	2.9
N76D	150mM	4.05E+04	2.69E-03	6.64E-08	2.35
N76DY72T	150mM	2.59E+04	2.04E-03	7.87E-08	4.04
N76DQ81R	150mM	1.70E+04	1.90E-03	1.12E-07	4.72

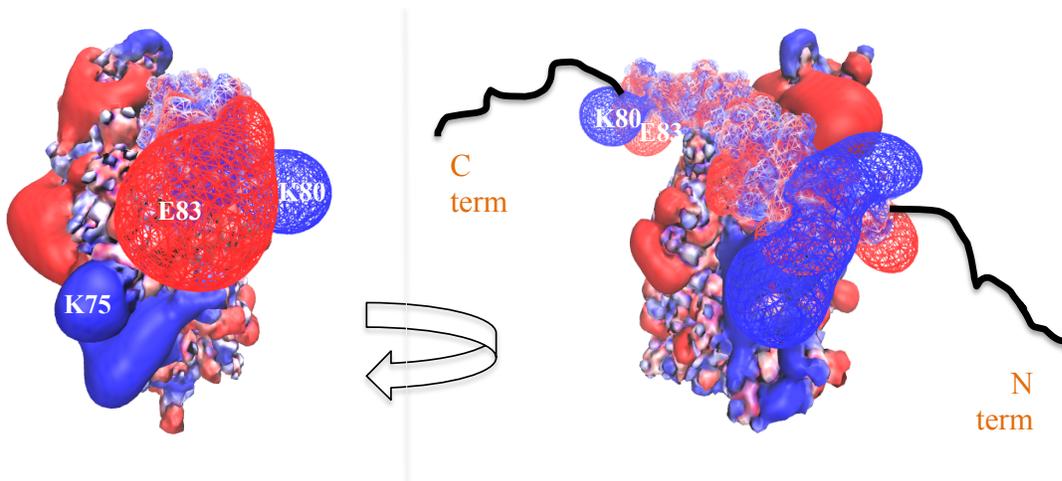


Figure 4.8 Coarse-Grained Extrapolation (Right) of Model V (Left). In the complex, the vHH is shown with a solid surface and the Ag is shown with a wire-mesh. The N and C terminal residues of A53T or AS not included in the computational model are shown as Black Lines to mark their location.

Predicting affinity-enhancing mutations by improving electrostatic complementarity of the binding interface for Model III

We obtained the PBEQ surface for the vHH and Ag corresponding to Model III. It revealed 2 main residues on the vHH that were interacting unfavorably with the Ag. These residues were identified as Lysine at position 75 (K75) and Lysine at position 99 (K99). This is due to the proximity of the K75 residue on the vHH to residues K58 and K60 on the Ag (Figure 4.4A, 4.4B). Similarly, K99 residue on vHH is proximal to the K80 residue on Ag. Since PBEQ is a numerical calculation for a structure (averaged over multiple equilibrium structures), it is much faster than a US simulation and was used as a first step for scanning multiple mutants for improved affinity. We thus proposed 5 mutants – K75A (A), K75E (E), K75A+M77T (AT), K75E+M77T (ET), K75A+M77T+K99Q (ATQ) aimed at alleviating electrostatic repulsion at the vHH-Ag interface, followed by equilibration simulations and PBEQ analysis of the equilibrated structures. Mutants A, AT and ATQ were found to have an improved electrostatic FE of binding over vHH 4C while

mutants E and ET showed little or no change (Figure 4.3). Subsequent US simulations for these mutants confirmed AT and ATQ as suitable candidates for enhanced affinity (Figure 4.6).

Testing computational predictions in experiments: Long-range electrostatic interactions in vhh-NAC binding are important

To test these findings, we expressed and purified vhh4C, AT and E. Surface Plasmon Resonance was used to determine the rates of association, dissociation and the dissociation constant of the purified proteins. These results are reported in Table 1. Change in salt concentration of the buffer led to a significant change in rates of association estimated by SPR indicating the importance of long-range electrostatic interactions supporting our hypothesis. However, the trend observed for affinity was opposite from that predicted in simulations. The association rates were found to be in the following order: 4C>AT>E. Hence, instead of observing improved binding for mutant AT and little or no-change for mutant E, we observed a considerable worsening of binding for mutants AT and E with respect to vHH 4C.

Predicting affinity-enhancing mutations by improving electrostatic complementarity of the binding interface for Model V

The above results seemed to suggest that the positively charged lysine at position 75 was important for binding and the rate of association became worse on changing the residue from positive to neutral to negative. Two possibilities were considered: 1) K75 could be important for long-range interaction with the negatively charged C-terminus of A53T and/or 2) K75 could be interacting with certain negatively charged residues in the NAC region. We scanned the remaining models and found Model V to be a reasonable candidate (and Model II). In this model

(and Model II), peptide P2 rather than P1 is in proximity to residue K75. Compared to P1, P2 is less positively charged and residue E83 (on P2) interacts residue K75 in this model. Moreover, in this model (and Model II), the negatively charged C-terminus of A53T might be closer to K75 on the vHH and may also influence long-range interactions. Hence, PBEQ calculations were performed for vHH 4C and mutants AT and E for the Model V (and Model II). The results are summarized in Figure 4.4. The trend for rate of association is $4C > AT > E$ as observed in experiments (Figure 4.4B). However, the difference between AT and 4C is rather small. One reason for this slight discrepancy could be simulation of a partial model of A53T. There may be some effect of the highly negatively charged C-terminus of A53T in long-range electrostatic interactions with the vhh. This would explain why mutant E is considerably worse than mutant AT. The electrostatic surface generated from PBEQ analysis for vHH 4C was also used to propose the mutation N76D, to introduce a new interaction between residue at position 76 on the vHH and residue K80 on P2. It is apparent from the PBEQ surface in Figure 4.4C that residues K75 and D76 (as in N76D mutant) on the vHH will form a surface complementary to residues E83 and K80 on the Ag respectively. Subsequently, US simulations were performed to obtain PMF for vHH 4C and mutants AT and N76D for Model V. (Running simulations for AT results for N76D shown). The PMF of binding for N76D is about 10kT larger than that for 4C indicating stronger binding (Figure 4.6).

Two other mutants N76DD72T and N76DQ81R were also tested. The rationale behind the additional D72T and Q81R is as follows. The residue D72 on vHH interacts with K75 on the vHH. By replacing the former with a T, we were hoping to “free” the K75 residue on the vHH to interact exclusively with the E83 residue on the Ag. The mutation Q81R was introduced to change the overall charge on the vHH and also to compensate for the N76D mutation that makes

the vHH more negatively charged wrt 4C. This was done to make the vHH more positively charged in order to improve long-range electrostatic interactions with a negatively charged A53T.

To test Model V, we synthesized, expressed and purified **mutant N76D** followed by SPR analysis. The results are reported in Table I. **Moreover, by making a single mutation, we have been able to improve the affinity of vhh4C by over an order of magnitude (50 fold).** Figure 4.5 compares all three models (II, III and V) discussed above) and only model V agrees with the experimental observations. The mutations N76DD72T and N76DQ81, however, led to slight worsening in affinity as seen from SPR results. While the D72T mutation was introduced to increase interaction of K75 on vHH with the E83 residue on the Ag, it also increases the entropy of the K75 residue in the uncomplexed vHH, which is lost upon binding. The Q81R mutation was based on a rather coarse-grained description of the system: negatively charged vHH interacting with positively charged vHH (Figure 4.7).

5 CONCLUSIONS AND FUTURE WORK

In this work, we have used computational modeling to understand, assess and redesign the binding surface of a vHH domain to achieve two different design goals. Most protein-engineering and drug-discovery work aimed towards therapeutic applications focuses on generating faster-associating or lower-affinity or aggregation-resistant or more-stable proteins; there seems to be a lag in the effort to understand the fundamental nature of protein surfaces and how they achieve these goals. Evidently, it is a hard problem for many reasons, not least of which is the highly functionalized nature of building units of proteins, compared to other macromolecules like DNA and RNA (ignoring modifications like phosphorylation, glycosylation etc.).

Even incremental gain toward a more fundamental description of these systems can lead to faster growth of many exciting applications like biomimetics, artificial surfaces capable of complex motions at the nano- and micro-scale, and perhaps in the distant future, surfaces that can respond to minute changes in temperature, salt-concentration, pH, mechanical stress, osmotic pressure and various other stimuli. Essentially, here lies the key that opens the door to the possibility of generating an endless variety of biologically inspired nanotransducers and biomaterials.

5.1 Summary and Conclusions

In our attempt to gain a deeper insight into the binding behavior of vHHs, we first redesigned a hypervariable loop in the binding region of a vHH to obtain mutations

exhibiting varying propensities for the targeted conformation and more interestingly, with different thermodynamic landscape and dynamics.

Our work on the hypervariable loop illustrates conformational isomerism in molecular detail, and shows that the sequence of the loops and their proximal residues modulates the intrinsic disorder of the vHH hypervariable regions. Since the motions of the loops, especially H1 and H3, are highly correlated, modulation in one loop leads to an overall change in the binding surface formed by the hypervariable loops. This observation aligns with the idea that intrinsic disorder plays an important role in the function of proteins – especially for molecular recognition. The presence of multiple conformational basins separated by moderate free-energy barriers, as observed in our simulations, imparts higher loop flexibility and may be a feature that is more prevalent in single-domain antibodies than in conventional two-domain antibodies. Moreover, it is expected that interactions with a second domain (as in conventional VHs) may impose additional geometric or enthalpic constraints that favor H1 loop rigidity.

Our results also highlight antibody isomerism pre-binding. Specifically, a conformational landscape for the H1 loop (and/or other loops) that is characterized by multiple stable/metastable states separated by small energy barriers, may have evolved to undergo conformational rearrangement during binding (e.g., FFSa and Wt), or may pre-exist in equilibrium with its isomers even before the antigen comes into the picture. On the other hand, an antibody with a stable structure separated from other metastable states by a large energy barrier may be less likely to exhibit conformational isomerism before the antigen approaches it (e.g., FFD).

We also demonstrated the role played by residues in a 4th hypervariable loop (H4). Slight variations in these residues can modify the shape and properties of the binding surface. This was further demonstrated when a single mutation in the anti-AS vHH led to an order of magnitude improvement in affinity.

As far as methods are concerned, from our conformational equilibria work using explicit solvent simulations, we found that the results agreed well with implicit solvent REMD simulations. This allowed us to use implicit-solvent REMD, computationally inexpensive compared to explicit solvent BE metadynamics, for vHH 4C. Thus, our work on redesigning the conformation of the H1 loop for anti-hCG vHH guided us in our work in developing a model for the anti-AS vHH and improving its affinity.

Equipped with the right set of methods and a better understanding of the binding surface of vHHs, we attempted to develop a three-dimensional, all-atomistic *in-silico* model of the vHH 4C and its complex with the NAC region of AS with the ultimate goal of using this model to improve its affinity for NAC.

We have previously described the main challenges associated with developing such a model. However, the problem is somewhat simplified by targeting a linear epitope on AS. This simplified design goal allows us to target an otherwise challenging target; a disordered protein which exhibits partial secondary structure, the propensity for which changes with environmental conditions and binding partners. To ensure that none of the plausible options were left out, we considered multiple conformations for the vHH and the antigen. However, we only considered peptides from the NAC region (and flanking residues), the implicit assumption here being that the immunized library should

be primarily specific to the NAC region. This assumption may be too simplistic, however, and future studies should explore it further.

The main achievement of the model was to illuminate the importance of electrostatic contributions, especially for targeting a primarily hydrophobic fragment which can contribute a maximum FE of desolvation directly proportional to its surface area. Given that the hydrophobic residues in the NAC region are small (short side-chains, low surface area), this contribution, even when maximized, may not be sufficient to overcome the loss in conformational entropy upon binding. This becomes especially significant when binding involves a vHH, which has a considerably smaller surface area compared to a conventional antibody. By targeting the flanking hydrophilic residues, we were able to overcome this “hydrophobic binding maxima” by introducing electrostatic complementarity at the binding interface, and effectively increasing the size of the binding surfaces.

The correctness of our model is demonstrated by the qualitative agreement between simulations, numerical PBEQ estimates and experimental results. However, within the framework of the present model, there is room for further investigation. For example, keeping the same model and relative orientation of the vHH and the Ag, one can explore a partially helical peptide for the Ag (instead of a fully helical peptide). For example, circular dichroism data on the NAC region can be used to generate a peptide with a probabilistically correct percentage of alpha-helical character. This may improve the predictions of the model and lead to other affinity enhancing mutations. Furthermore, we may also consider a hybrid atomistic and coarse-grained description for the NAC region and the rest of AS respectively.

5.2 Future Work

There are multiple challenging and intriguing problem statements arising from our work and similar work done over the last couple of years. They are listed below:

- 1) **Structural Motifs:** Finding common structural motifs for various NAC-specific antibodies (vHH 4C, NAC32, NAC32-R1, NAC32-R2, VH14, D5, D10 etc.).
- 2) **Theoretical Upper Limit:** Prediction of a theoretical upper limit of free-energy of binding for a fixed Antibody (or vHH) for a given linear target, taking into account the desolvation, configurational entropy and enthalpy. This can be useful in guiding experimental evolution as well.
- 3) **Experimental Evolution of vHH 4C:** A parallel evolution of vHH 4C, using experimental techniques, with full characterization of change in binding rates and dissociation constants at each stage. Which mutations show up and how many rounds of evolution are needed? Can these mutations be explained by the computational model?
- 4) **Targeting toxic AS species:** How can we modify a vHH to target various forms of AS – monomer, oligomeric-soluble, oligomeric-insoluble, fibrillar?
- 5) **Highly evolved antibodies:** Broadly-Neutralizing Antibodies (bNAbs) are neutralizing antibodies which are effective against multiple HIV viral strains as opposed to non-bNAbs which are specific for individual viral strains. Studying the conformational equilibria of these bNAbs and understanding how these antibodies have evolved (some over years) to become effective against various hidden, highly conserved epitopes on the HIV.

APPENDIX A: SEQUENCE OF ANTI-AS VHHS

Sequence of anti-AS vHHS discussed in Chapters 3 and 4:

```

4C  QVQLVESGGGGLVQPGGSLRLSCVTPGGIFSDYAMGWYRQAPGKQRELVADITVSD-TTR 59
6A  QLQLVESGGGGLVQPGGSLRLSRAASGSIFSDYAMGWYRQAPGKQRELVADITVSD-TTR 59
2A  QVSSWSRAGG-LVQPGGSLRLSCAASGSIFSDYAMGWYRQAPGKQRESVADITVSD-TTR 58
6B  QVQLVESGGG-LVQPGESLRLTCTVSGVTSDDYFVCGWFRQAPRAEREGIASISIEHGDIY 59
    *:. . .** ***** *****: ..* _** :*:***** :** :* .*:::

4C  YSDPVKGRFTISRDNKQNMVYLQMNSLKPEDAAIYTCNAQK--WEDRYGS-SERYDYWGQ 116
6A  YSDPVKGRFTISRDDAKNMVYLQMNSLKPEDTAIYTCNAQK--WEDRYGS-SERYGYWGQ 116
2A  YSDPVKGRFTISRDNKQNMVYLQMNSLKPEDTAIYTCNAQK--WEDRYGS-SERYDYWGQ 115
6B  YSHDVRGRFTISRGSAKNTVDLQMNMLKPADTGVYICAAVQHLYSDGFELQPNEWNYWGR 119
    **_ *::*****_*** * *****_* * :::* * * : . * : .:::****:

4C  GTQVTVSSAHHSEDP- 131
6A  GTQVTVSSAHHSEDP- 131
2A  GTQVTVSSAHHSEDP- 130
6B  GTQVTVSSEPKTPKPQ 135
    ***** : : .*
  
```

Figure 1 Full hits sequences obtained from 1st round of FLITRAP

REFERENCES

1. Velez-Vega, C.; Fenwick, M. K.; Escobedo, F. A. Simulated mutagenesis of the hypervariable loops of a llama VHH domain for the recovery of canonical conformations. *J Phys Chem B* **2009**, *113*, 1785-1795.
2. Revets, H.; De Baetselier, P.; Muyldermans, S. Nanobodies as novel agents for cancer therapy. *Expert Opin. Biol. Ther.* **2005**, *5*, 111-124.
3. Holliger, P.; Hudson, P. J. Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol.* **2005**, *23*, 1126-1136.
4. Kabat, E. A.; Wu, T. T.; Perry, H. M.; Gottesman, K. S.; Foeller, C. *Sequences of Proteins of Immunological Interest*; U. S. Department of Health and Human Services: Bethesda, MD, 1991; Vol. 5.
5. Arbabi Ghahroudi, M.; Desmyter, A.; Wyns, L.; Hamers, R.; Muyldermans, S. Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Lett.* **1997**, *414*, 521-526.
6. Joosten, V.; Lokman, C.; Van Den Hondel, C. A.; Punt, P. J. The production of antibody fragments and antibody fusion proteins by yeasts and filamentous fungi. *Microb. Cell. Fact.* **2003**, *2*, 1.
7. De Genst, E.; Saerens, D.; Muyldermans, S.; Conrath, K. Antibody repertoire development in camelids. *Dev. Comp. Immunol.* **2006**, *30*, 187-198.
8. Chothia, C.; Lesk, A. M. Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **1987**, *196*, 901-917.
9. Tramontano, A.; Chothia, C.; Lesk, A. M. Structural determinants of the conformations of medium-sized loops in proteins. *Proteins* **1989**, *6*, 382-394.
10. Chothia, C.; Lesk, A. M.; Gherardi, E.; Tomlinson, I. M.; Walter, G.; Marks, J. D.; Llewelyn, M. B.; Winter, G. Structural repertoire of the human VH segments. *J. Mol. Biol.* **1992**, *227*, 799-817.
11. Wu, S.; Cygler, M. Conformation of complementarity determining region L1 loop in murine IgG lambda light chain extends the repertoire of canonical forms. *J. Mol. Biol.* **1993**, *229*, 597-601.
12. Al-Lazikani, B.; Lesk, A. M.; Chothia, C. Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* **1997**, *273*, 927-948.
13. Barre, S.; Greenberg, A. S.; Flajnik, M. F.; Chothia, C. Structural conservation of hypervariable regions in immunoglobulins evolution. *Nat. Struct. Biol.* **1994**, *1*, 915-920.
14. Kangas, E.; Tidor, B. Charge optimization leads to favorable electrostatic binding free energy. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **1999**, *59*, 5958-5961.
15. Lee, L. P.; Tidor, B. Barstar is electrostatically optimized for tight binding to barnase. *Nat. Struct. Biol.* **2001**, *8*, 73-76.
16. Lee, L. P.; Tidor, B. Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. *Protein Sci.* **2001**, *10*, 362-377.

17. Clark, L. A.; Boriack-Sjodin, P. A.; Eldredge, J.; Fitch, C.; Friedman, B.; Hanf, K. J.; Jarpe, M.; Liparoto, S. F.; Li, Y.; Lugovskoy, A.; Miller, S.; Rushe, M.; Sherman, W.; Simon, K.; Van Vlijmen, H. Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. *Protein Sci.* **2006**, *15*, 949-960.
18. Clark, L. A.; Boriack-Sjodin, P. A.; Day, E.; Eldredge, J.; Fitch, C.; Jarpe, M.; Miller, S.; Li, Y.; Simon, K.; van Vlijmen, H. W. An antibody loop replacement design feasibility study and a loop-swapped dimer structure. *Protein Eng. Des. Sel.* **2009**, *22*, 93-101.
19. Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Nelson, J. W.; Schena, A.; Jankowski, W.; Kalodimos, C. G.; Johnsson, K.; Stoddard, B. L.; Baker, D. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **2013**, *501*, 212-216.
20. Cookson, M. R.; van der Brug, M. Cell systems and the toxic mechanism(s) of alpha-synuclein. *Exp. Neurol.* **2008**, *209*, 5-11.
21. Giasson, B. I.; Murray, I. V.; Trojanowski, J. Q.; Lee, V. M. A hydrophobic stretch of 12 amino acid residues in the middle of alpha-synuclein is essential for filament assembly. *J. Biol. Chem.* **2001**, *276*, 2380-2386.
22. Cookson, M. R. alpha-Synuclein and neuronal cell death. *Mol. Neurodegener* **2009**, *4*, 9-1326-4-9.
23. Lynch, S. M.; Zhou, C.; Messer, A. An scFv intrabody against the nonamyloid component of alpha-synuclein reduces intracellular aggregation and toxicity. *J. Mol. Biol.* **2008**, *377*, 136-147.
24. Periquet, M.; Fulga, T.; Myllykangas, L.; Schlossmacher, M. G.; Feany, M. B. Aggregated alpha-synuclein mediates dopaminergic neurotoxicity in vivo. *J. Neurosci.* **2007**, *27*, 3338-3346.
25. Zhou, C.; Emadi, S.; Sierks, M. R.; Messer, A. A human single-chain Fv intrabody blocks aberrant cellular effects of overexpressed alpha-synuclein. *Mol. Ther.* **2004**, *10*, 1023-1031.
26. Emadi, S.; Barkhordarian, H.; Wang, M. S.; Schulz, P.; Sierks, M. R. Isolation of a human single chain antibody fragment against oligomeric alpha-synuclein that inhibits aggregation and prevents alpha-synuclein-induced toxicity. *J. Mol. Biol.* **2007**, *368*, 1132-1144.
27. Perchiacca, J. M.; Ladiwala, A. R.; Bhattacharya, M.; Tessier, P. M. Structure-based design of conformation- and sequence-specific antibodies against amyloid beta. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 84-89.
28. Sormanni, P.; Aprile, F. A.; Vendruscolo, M. Rational design of antibodies targeting specific epitopes within intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 9902-9907.
29. Renisio, J. G.; Perez, J.; Czisch, M.; Guennegues, M.; Bornet, O.; Frenken, L.; Cambillau, C.; Darbon, H. Solution structure and backbone dynamics of an antigen-free heavy chain variable domain (VHH) from Llama. *Proteins* **2002**, *47*, 546-555.
30. Spinelli, S.; Frenken, L.; Bourgeois, D.; de Ron, L.; Bos, W.; Verrips, T.; Anguille, C.; Cambillau, C.; Tegoni, M. The crystal structure of a llama heavy chain variable domain. *Nat. Struct. Biol.* **1996**, *3*, 752-757.
31. Fenwick, M. K.; Escobedo, F. A. Hybrid Monte Carlo with multidimensional replica exchanges: conformational equilibria of the hypervariable regions of a llama VHH antibody domain. *Biopolymers* **2003**, *68*, 160-177.

32. Decanniere, K.; Desmyter, A.; Lauwereys, M.; Ghahroudi, M. A.; Muyldermans, S.; Wyns, L. A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. *Structure* **1999**, *7*, 361-370.
33. Sircar, A.; Sanni, K. A.; Shi, J.; Gray, J. J. Analysis and modeling of the variable region of camelid single-domain antibodies. *J. Immunol.* **2011**, *186*, 6357-6367.
34. Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* **2005**, *18*, 343-384.
35. Biarnes, X.; Bongarzone, S.; Vargiu, A. V.; Carloni, P.; Ruggerone, P. Molecular motions in drug design: the coming age of the metadynamics method. *J. Comput. Aided Mol. Des.* **2011**, *25*, 395-402.
36. Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, 054103.
37. Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435-13441.
38. Foote, J. Immunology. Isomeric antibodies. *Science* **2003**, *299*, 1327-1328.
39. James, L. C.; Roversi, P.; Tawfik, D. S. Antibody multispecificity mediated by conformational diversity. *Science* **2003**, *299*, 1362-1367.
40. James, L. C.; Tawfik, D. S. The specificity of cross-reactivity: promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. *Protein Sci.* **2003**, *12*, 2183-2193.
41. James, L. C.; Tawfik, D. S. Structure and kinetics of a transient antibody binding intermediate reveal a kinetic discrimination mechanism in antigen recognition. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12730-12735.
42. Stanfield, R. L.; Takimoto-Kamimura, M.; Rini, J. M.; Profy, A. T.; Wilson, I. A. Major antigen-induced domain rearrangements in an antibody. *Structure* **1993**, *1*, 83-93.
43. Stanfield, R. L.; Wilson, I. A. Antigen-induced conformational changes in antibodies: a problem for structural prediction and design. *Trends Biotechnol.* **1994**, *12*, 275-279.
44. Decanniere, K.; Muyldermans, S.; Wyns, L. Canonical antigen-binding loop structures in immunoglobulins: more structures, more canonical classes? *J. Mol. Biol.* **2000**, *300*, 83-91.
45. Davies, J.; Riechmann, L. Affinity improvement of single antibody VH domains: residues in all three hypervariable regions affect antigen binding. *Immunotechnology* **1996**, *2*, 169-179.
46. Micheletti, C.; Laio, A.; Parrinello, M. Reconstructing the density of states by history-dependent metadynamics. *Phys. Rev. Lett.* **2004**, *92*, 170601.
47. Piana, S.; Laio, A. A bias-exchange approach to protein folding. *J Phys Chem B* **2007**, *111*, 4553-4559.
48. Todorova, N.; Marinelli, F.; Piana, S.; Yarovsky, I. Exploring the folding free energy landscape of insulin using bias exchange metadynamics. *J Phys Chem B* **2009**, *113*, 3556-3564.
49. Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148-154.

50. Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Comput. Biol.* **2009**, *5*, e1000452.
51. Lauwereys, M.; Arbabi Ghahroudi, M.; Desmyter, A.; Kinne, J.; Holzer, W.; De Genst, E.; Wyns, L.; Muyldermans, S. Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. *EMBO J.* **1998**, *17*, 3512-3520.
52. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26*, 1701-1718.
53. Hess, B.; Kutzner, C.; van, d. S.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447.
54. MacKerell, A. D.; Bashford, D.; Bellott, D.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; WiÅ³rkiewicz-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.
55. Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6*, 459-466.
56. Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961-1972.
57. Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116-122.
58. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
59. Nguyen, V. K.; Hamers, R.; Wyns, L.; Muyldermans, S. Camel heavy-chain antibodies: diverse germline V(H)H and specific mechanisms enlarge the antigen-binding repertoire. *EMBO J.* **2000**, *19*, 921-930.
60. Tramontano, A.; Chothia, C.; Lesk, A. M. Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J. Mol. Biol.* **1990**, *215*, 175-182.
61. Kehoe, J. M.; Capra, J. D. Phylogenetic aspects of immunoglobulin variable region diversity. *Contemp. Top. Mol. Immunol.* **1974**, *3*, 143-159.
62. Bond, C. J.; Wiesmann, C.; Marsters, J. C., Jr; Sidhu, S. S. A structure-based database of antibody variable domain diversity. *J. Mol. Biol.* **2005**, *348*, 699-709.
63. Baca, M.; Presta, L. G.; O'Connor, S. J.; Wells, J. A. Antibody humanization using monovalent phage display. *J. Biol. Chem.* **1997**, *272*, 10678-10684.
64. Carter, P.; Presta, L.; Gorman, C. M.; Ridgway, J. B.; Henner, D.; Wong, W. L.; Rowland, A. M.; Kotts, C.; Carver, M. E.; Shepard, H. M. Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 4285-4289.

65. Hummer, G. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J. Phys.* **2005**, *7*, 34.
66. Bicout, D. J.; Szabo, A. Electron transfer reaction dynamics in non-Debye solvents. *J. Chem. Phys.* **1998**, *109*, 2325.
67. Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. On the assumptions underlying milestoning. *J. Chem. Phys.* **2008**, *129*, 174102.
68. Vanden-Eijnden, E.; Venturoli, M. Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.* **2009**, *130*, 194101-194101-13.
69. Burley, S. K.; Petsko, G. A. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **1985**, *229*, 23-28.
70. Burley, S. K.; Petsko, G. A. Amino-aromatic interactions in proteins. *FEBS Lett.* **1986**, *203*, 139-143.
71. Wilson, I. A.; Stanfield, R. L.; Rini, J. M.; Arevalo, J. H.; Schulze-Gahmen, U.; Fremont, D. H.; Stura, E. A. Structural aspects of antibodies and antibody-antigen complexes. *Ciba Found. Symp.* **1991**, *159*, 13-28; discussion 28-39.
72. Wilson, I. A.; Ghiara, J. B.; Stanfield, R. L. Structure of anti-peptide antibody complexes. *Res. Immunol.* **1994**, *145*, 73-78.
73. Wilson, I. A.; Stanfield, R. L. Antibody-antigen interactions: new structures and new conformational changes. *Curr. Opin. Struct. Biol.* **1994**, *4*, 857-867.
74. Wilson, I. A.; Stanfield, R. L. A Trojan horse with a sweet tooth. *Nat. Struct. Biol.* **1995**, *2*, 433-436.
75. Stanfield, R.; Cabezas, E.; Satterthwait, A.; Stura, E.; Profy, A.; Wilson, I. Dual conformations for the HIV-1 gp120 V3 loop in complexes with different neutralizing fabs. *Structure* **1999**, *7*, 131-142.
76. Maass, D. R.; Sepulveda, J.; Pernthaner, A.; Shoemaker, C. B. Alpaca (Lama pacos) as a convenient source of recombinant camelid heavy chain antibodies (VHHs). *J. Immunol. Methods* **2007**, *324*, 13-25.
77. Waraho, D.; DeLisa, M. P. Versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3692-3697.
78. Karlsson, A. J.; Lim, H. K.; Xu, H.; Rocco, M. A.; Bratkowski, M. A.; Ke, A.; DeLisa, M. P. Engineering antibody fitness and function using membrane-anchored display of correctly folded proteins. *J. Mol. Biol.* **2012**, *416*, 94-107.
79. Benkert, P.; Biasini, M.; Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **2011**, *27*, 343-350.
80. Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **2003**, *31*, 3381-3385.
81. Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, *22*, 195-201.

82. Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6*, 459-466.
83. Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999**, *314*, 141-151.
84. Mahajan, S. P.; Velez-Vega, C.; Escobedo, F. A. Tilting the balance between canonical and noncanonical conformations for the H1 hypervariable loop of a llama VHH through point mutations. *J Phys Chem B* **2013**, *117*, 13-24.
85. Fauvet, B.; Mbefo, M. K.; Fares, M. B.; Desobry, C.; Michael, S.; Ardah, M. T.; Tsika, E.; Coune, P.; Prudent, M.; Lion, N.; Eliezer, D.; Moore, D. J.; Schneider, B.; Aebischer, P.; El-Agnaf, O. M.; Masliah, E.; Lashuel, H. A. alpha-Synuclein in central nervous system and from erythrocytes, mammalian cells, and Escherichia coli exists predominantly as disordered monomer. *J. Biol. Chem.* **2012**, *287*, 15345-15364.
86. Bartels, T.; Choi, J. G.; Selkoe, D. J. alpha-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature* **2011**, *477*, 107-110.
87. Wang, W.; Perovic, I.; Chittuluru, J.; Kaganovich, A.; Nguyen, L. T.; Liao, J.; Auclair, J. R.; Johnson, D.; Landeru, A.; Simorellis, A. K.; Ju, S.; Cookson, M. R.; Asturias, F. J.; Agar, J. N.; Webb, B. N.; Kang, C.; Ringe, D.; Petsko, G. A.; Pochapsky, T. C.; Hoang, Q. Q. A soluble alpha-synuclein construct forms a dynamic tetramer. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17797-17802.
88. Ulmer, T. S.; Bax, A.; Cole, N. B.; Nussbaum, R. L. Structure and dynamics of micelle-bound human alpha-synuclein. *J. Biol. Chem.* **2005**, *280*, 9595-9603.
89. Brenke, R.; Hall, D. R.; Chuang, G. Y.; Comeau, S. R.; Bohnuud, T.; Beglov, D.; Schueler-Furman, O.; Vajda, S.; Kozakov, D. Application of asymmetric statistical potentials to antibody-protein docking. *Bioinformatics* **2012**, *28*, 2608-2614.
90. Waraho-Zhmayev, D.; Gkogka, L.; Yu, T. Y.; DeLisa, M. P. A microbial sensor for discovering structural probes of protein misfolding and aggregation. *Prion* **2013**, *7*, 151-156.
91. der Maur, A. A.; Zahnd, C.; Fischer, F.; Spinelli, S.; Honegger, A.; Cambillau, C.; Escher, D.; Pluckthun, A.; Barberis, A. Direct in vivo screening of intrabody libraries constructed on a highly stable single-chain framework. *J. Biol. Chem.* **2002**, *277*, 45075-45085.
92. Chu, H. M.; Wright, J.; Chan, Y. H.; Lin, C. J.; Chang, T. W.; Lim, C. Two potential therapeutic antibodies bind to a peptide segment of membrane-bound IgE in different conformations. *Nat. Commun.* **2014**, *5*, 3139.
93. Hub, J. S.; de Groot, B. L.; van, d. S. g_wham: A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713-3720.
94. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992**, *13*, 1011-1021.
95. Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, *23*, 128-137.

96. Li, C.; Li, L.; Zhang, J.; Alexov, E. Highly efficient and exact method for parallelization of grid-based algorithms and its implementation in DelPhi. *J. Comput. Chem.* **2012**, *33*, 1960-1966.
97. Smith, N.; Witham, S.; Sarkar, S.; Zhang, J.; Li, L.; Li, C.; Alexov, E. DelPhi web server v2: incorporating atomic-style geometrical figures into the computational protocol. *Bioinformatics* **2012**, *28*, 1655-1657.
98. Li, L.; Li, C.; Sarkar, S.; Zhang, J.; Witham, S.; Zhang, Z.; Wang, L.; Smith, N.; Petukh, M.; Alexov, E. DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys.* **2012**, *5*, 9-1682-5-9.
99. Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126-2136.
100. Li, C.; Petukh, M.; Li, L.; Alexov, E. Continuous development of schemes for parallel computing of the electrostatics in biological systems: implementation in DelPhi. *J. Comput. Chem.* **2013**, *34*, 1949-1960.
101. Griffiths, D. *Introduction to Electrodynamics*; Prentice Hall Inc.: New Jersey, 1999; Vol. 3, pp 191.
102. Sharp, K. A.; Honig, B. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem.* **1990**, *94*, 7684-7692.
103. Rocchia, W.; Alexov, E.; Honig, B. Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J Phys Chem B* **2001**, *105*, 6507-6514.
104. Lippow, S. M.; Wittrup, K. D.; Tidor, B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat. Biotechnol.* **2007**, *25*, 1171-1176.