Simulated Mutagenesis of the Hypervariable Loops of a Llama VHH Domain for the Recovery of Canonical Conformations

Camilo Velez-Vega,[†] Michael K. Fenwick,[‡] and Fernando A. Escobedo*

School of Chemical and Biomolecular Engineering, Department of Molecular Medicine, Cornell University, Ithaca, New York 14853

In this work, wildtype and mutated hypervariable regions of an anti-hCG llama VHH antibody were simulated via a molecular dynamics replica exchange method (REM). Seven mutants were simulated with the goal of identifying structural determinants that return the noncanonical H1 loop of the wildtype antibody to the type 1 canonical structure predicted by database methods formulated for conventional antibodies. Two cases with three point mutations yielded a stable type 1 H1 structure. In addition, other mutants with fewer mutations showed evidence of such conformations. Overall, the mutagenesis results suggest a marked influence of interloop interactions on the attainment of canonical conformations for this antibody. On the methodological front, a novel REM scheme was developed to quickly screen diverse mutants based on their relative propensities for attaining favorable structures. This multimutant REM (MMREM) was used to successfully identify mutations that stabilize a canonical H1 loop grafted on the llama antibody scaffold. The use of MMREM and REM for screening mutants and assessing structural stability may be useful in the rational design of antibody hypervariable loops.

Introduction

The chemical diversity of the antigen binding sites of antibodies plays a critical role in the immune system by facilitating the identification of countless foreign antigens. This diversity is due largely to the amino acid variability of the complementarity-determining regions (CDRs). Structurally, the binding sites of conventional antibodies are composed of three light chain and three heavy chain hypervariable regions (denoted L1, L2, L3, H1, H2, and H3, respectively), which represent six loops that pack together to form a versatile surface for molecular recognition. Accordingly, the problem of antigen binding site engineering is often simplified to the modeling and structure prediction of a set of six loops on a relatively conserved protein scaffold.

A variety of knowledge-based (KB),[1–19] ab initio,[20–33] and combined methods[34–36] have been used for predicting the structures of antibody hypervariable loops (for a comprehesive review of solutions for the more general problem of loop structure prediction for proteins see, e.g., Fiser et al.[37]). In particular, the use of KB methods has shown promising results[2,11] due to the limited number of conformations that have been observed for most antibody loop backbones. Al-Lazikani et al.[1] and Martin et al.[11] have categorized these loops into various canonical conformations based on available crystal structures. Shirai et al.[18] proposed a more rigorous classification for H3 loops. Despite the successful effort of grouping many of the experimentally resolved hypervariable structures, several loops do not fit within these classifications. Such is the case for some of the loops of the variable domains of camelid heavy chain antibodies (VHHs), which lack light chains.[5,38] These antibodies utilize a reduced biomolecular surface for antigen

binding that appears to have evolved increased hypervariable loop structural variability to compensate for the absence of the light chain.

Camelid VHHs have increasingly been used for engineered antigen binding given their various desired characteristics, which include small size, high expression level,[39] reversible folding after exposure to high temperatures,[40] and functionality as enzyme inhibitors.[41] Nevertheless, their general application still faces some limitations, such as changes in structure upon loop grafting or exposure to harsh conditions like increased temperature and low pH.[42] A computational method that can assist in the identification of key mutations that lead to affine structures with improved stability under different conditions is highly desirable.

A particular VHH amenable for systematic study of its key loop residues is the llama VHH raised against human chorionic gonadotropin (hCG). Its antigen-free structure has been resolved by Renisio et al.[43] via NMR spectroscopy (PDB code 1G9E and referred to as VHH-H14) and by Spinelli et al.[44] via X-ray crystallography (PDB code 1HCV). Whereas the crystal structure shows well-defined canonical class-2A H2 and noncanonical H1 conformations, the NMR structure ensemble of these loops is more indistinct and suggests significant flexibility in solution. Notably, KB methods provide either ambiguous or incorrect predictions for each of the three loops of the crystal structure, motivating the use of more sophisticated methods for structure prediction and computer analysis. For example, using the hybrid Monte Carlo replica exchange (HYMREX)[22] method, the crystal structure loop conformations were simulated and examples of loop flexibility were provided that are reflected in the dynamic equilibrium observed using NMR. Although prior studies of this system have suggested several residues to be key determinants of the loop conformations,[5,22,44] additional insights may be gained from site-directed mutagenesis.

In the present study, replica exchange molecular dynamics (REM) simulations of wildtype and mutant llama VHH hyper-

* To whom correspondence should be addressed. E-mail: fe13@cornell.edu.
† School of Chemical and Biomolecular Engineering.
‡ Department of Molecular Medicine.

variable loops were performed to quantify loop flexibility and identify residues that play a critical role in shaping and stabilizing these conformations. Much of the analysis is focused on the central H1 loop, which contains predominantly hydrophilic residues and adopts a backbone structure that is unique among the known H1 conformations in antibodies.[5] In contrast, the vast majority of H1 loops found in conventional antibodies adopt a well-defined conformation that is kinked in the center and presumably stabilized by key hydrophobic residues. One of our principal aims was to identify an effective strategy and a solvent force field for simulating the conformational equilibrium of such canonical and noncanonical antibody loops, and for identifying mutations that can account for their differing backbone structures. An All Pairs Exchange[45] adaptation of REM[46] was implemented in parallel using CHARMM[47] version 32 via the Multiscale Modeling Tools for Structural Biology (MMTSB) toolset.[48] Furthermore, a novel application of REM for rapid mutant screening was implemented in which various plausible mutations were evaluated in a single REM run.

This work is organized as follows: The crystal and NMR structures of the wildtype llama VHH hypervariable loops are reviewed first, followed by a brief analysis of the mutations selected and a description of the simulation details. The simulated conformational changes of the loops are then compared with those observed in experimental structures. Finally, the simulation results on the wildtype structure and the various mutants considered are presented and discussed in the context of stabilization of canonical structures. The paper ends with some concluding remarks about the structural determinants and the simulation method.

Llama VHH Loop Conformations

The available NMR[43] and crystal[44] structures of the anti-hCG llama VHH domain allow the identification of some key structural features of the hypervariable loops. For this study, H1, H2, and H3 comprise 1HCV residues 26−32 (GRTGSTY), 52,52a−56 (NWDSAR), and 95−102 (GEGGTWDS), respectively. Figure 1 shows the relative probabilities of occurrence for the rmsd values between the loops of the crystal structure and those of the twenty NMR structures reported (unless otherwise indicated, rmsd refers to backbone atom rmsd). For this calculation, each loop from the NMR structure ensemble was aligned to the corresponding loop of the crystal conformation. Noting that the crystal structure was obtained at lower temperatures with crystal packing whereas the NMR spectra were recorded at 300°K in solution, some discrepancy may be expected.

For the H1 loop, Figure 1A illustrates NMR configurations within a range of 1.4−2.2 Å from the crystal structure, the representative NMR structure deviating by 1.4 Å. In addition, Figure 1B shows the rmsd distribution for the NMR H1 loops relative to those of the crystal structure of a monoclonal antibody Fab fragment (PDB code 1DFB).[49] 1DFB adopts a type 1 H1 loop conformation and is therefore used as reference. Although the KB methods of Al-Lazikani et al.[1] and Martin et al.[11] predict a type 1 conformation for the H1 loop of the llama VHH, it is evident that no NMR conformation adopts a type 1 structure. This is also true for the crystal H1 loop, which has a rmsd value of 1.4 Å with respect to the corresponding 1DFB loop.

For the H2 loop, it can be seen in Figure 1A that the NMR structures are within a range of 0.8−1.8 Å from the crystal structure and the representative NMR structure differs by 1.6 Å. Comparisons are made with the H2 loop structures of two reference antibodies (Figure 1C), namely, 1DFB and the FV
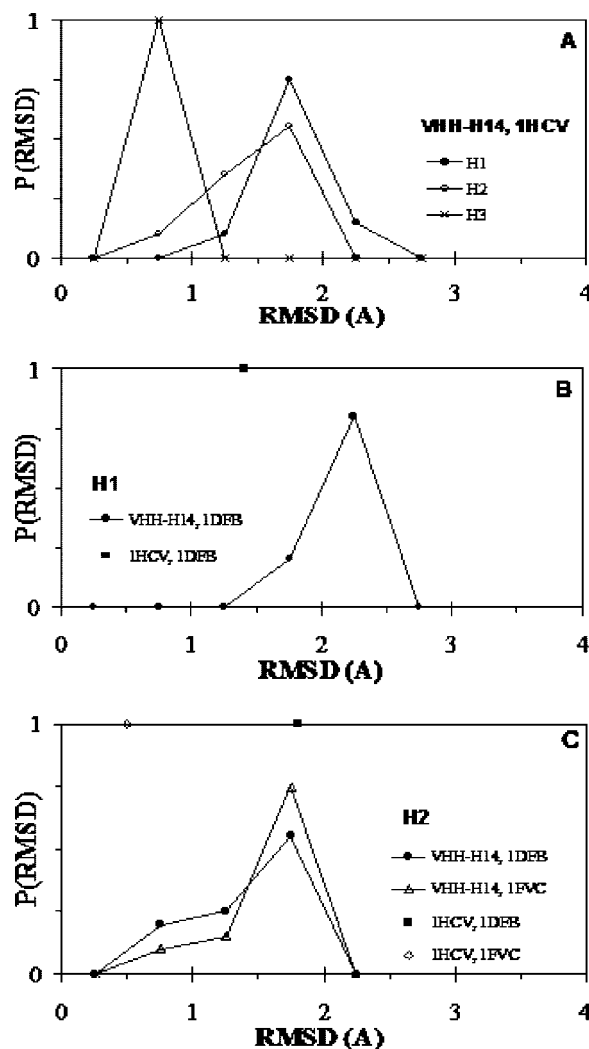

Figure 1. Probabilities of occurrence for the rmsd values of the loops of the crystal (1HCV) and twenty NMR (VHH-H14) structures relative to reference structures 1DFB and 1FVC. The plots show values for the VHH-H14 structures with respect to 1HCV for H1, H2, and H3 loops (A); the VHH-H14 and crystal structures with respect to 1DFB for the H1 loop (B); and the VHH-H14 and crystal structures with respect to 1DFB and 1FVC for the H2 loop (C). For every case, zero probability indicates that the loops did not sample conformations having such rmsd.

fragment of the humanized antibody 4D5[50] (PDB code 1FVC), which adopt a type 3 and a type 2A conformation, respectively. KB predictions for the llama VHH H2 loop indicate either a type 3[1] or a type 2A[11] conformation. Figure 1C shows that 45 and 25% of the conformations have rmsd values below 1.5 Å with respect to a type 3 and a type 2A structure, respectively. Conversely, the rmsd values of the 1HCV H2 loop with respect to the type 3 and type 2A reference structures are 1.8 and 0.5 Å, respectively. It can then be inferred that, even though H2 is markedly type 2A in the crystal structure, it may be flexible in solution and temporarily adopt a type 3 structure. However, it is noted that residues D53 and S54 were not assigned in the NMR spectra.

Finally, for the H3 loop, Figure 1A shows consistency between the crystal and NMR structures. This is indicative of an H3 conformation with reduced flexibility. All of the H3 structures adopt a kinked conformation in contrast to the extended structure predicted using the H3 KB rules.[22]

Hypervariable Loops of a Llama VHH Domain

J. Phys. Chem. B, Vol. 113, No. 6, 2009 **1787**

Selection of Candidate Mutants

Simulated site-directed mutagenesis experiments were performed in an attempt to return the llama VHH H1 loop structure to the type 1 conformation predicted by KB methods. Only sites within the H1 and H2 regions that are thought to directly shape the standard H1 conformation were mutated. As first noted by Spinelli et al.,[44] the 1HCV H1 amino acid sequence coincides with a type 1 loop in four of the seven residues. The three residues considered to be infrequent were R27, G29, and T31. T31 is observed in 10.6% (33/310) of the H1 loops of antibody structures analyzed, whereas R27 and G29 are found in less than 1% of the cases. Given the moderate likelihood of finding T31 in H1 type 1 structures, our study focused on positions 27 and 29 as key determinants of the structural features. Moreover, residue W52a in H2 was identified as being important for preventing the type 1 H1 structure formation, due to possible clashing with residues in this canonical form ("a" in W52a indicates that this residue is between landmark alignment positions N52 and D53). Residues R27, G29, and W52a were consequently selected as targets for mutation; residue N52 was also selected as a possible mutation site for reasons that are clarified below. Single point mutations at position 27 or 29 as well as double point mutations at these two locations were then assumed to be the ones likely to achieve the desired conformation when considering changes in the H1 loop alone. Three and four point mutations included changes in sites from the H2 loop.

The substitute residues for sites 27 and 29 were chosen on the basis of frequent occurrences (Y/F/G at position 27 and F/I/L at position 29) in canonical type 1 H1 loops.[11] Incidentally, mutations in these sites have been proposed as an alternative way for achieving an augmented affinity in VHHs.[51] For most of the simulations involving mutations at either or both of these positions, we selected F because (with human therapeutic applications in mind) it occurs at these positions in most of the approximately 51 human germline VH gene segments (26 and 33 of 51 for positions 27 and 29, respectively).[52] Furthermore, we observed that its side chain was easier to bury than that of Y at position 27 (perhaps due to its higher hydropathy index). As for mutations in H2, a W52aS replacement was simulated to examine the steric hindrance effect of this bulky hydrophobic residue on the attainability of a type 1 H1 conformation. The serine residue at this position was selected for its relatively small size and to avoid an evident bias toward an H2 type 3 (common residues D/P) or H2 type 2A (common residues P/T/A) conformation. Finally, a N52S mutation was simulated to evaluate the possible cooperative adoption of type 1 H1 and type 3 H2 loops, as occurs in antibody 1DFB, which also has a tryptophan at position 52a.

Simulation Method

A reduced model of the crystal structure (1HCV) previously defined[22] was used as the basis for all simulations. Briefly, the hypervariable regions and a subset of residues in the proximal framework regions of 1HCV were chosen for our simulations. To reduce computation time, distal framework residues that are unlikely to have a major influence on the structure of the hypervariable loops were excluded. Altogether, 57 of the 117 amino acids in 1HCV were included in our simulations. C_α, N, and C backbone atoms that do not belong to hypervariable loops were restrained with a harmonic constant K_{HARM} of 0.5 × (atomic mass), while backbone atoms from hypervariable regions and side chains from all residues included in the simulations were free to move. These restraints were imposed to simulate the limited motion displayed by backbone atoms of

TABLE 1: Mutations for the Different Cases Simulated Using Conventional REM

case	mutations
wildtype	
1-Fa	R27F
1-Fb	G29F
2-FF	R27F, G29F
2-FS	G29F, N52S
2-FL	R27F, G29L
3-FFSa	R27F, G29F, W52aS
3-FFSb	R27F, G29F, N52S

framework residues that are part of secondary (β-sheet) structures. Acetyl and N-methyl groups were used to cap the N- and C- terminal ends of each of the simulated fragments. The starting structure was slowly heated to 300 K via MD. Point mutations were introduced into the heated structure to realize the starting point for each of the mutant cases indicated in Table 5; this was followed by 1000 steps of steepest descent energy minimization and a short equilibration. The resulting conformation was the initial structure for each mutant. The velocity Verlet algorithm was chosen as the integrator with a time step of 2 fs. SHAKE[53] was used to constrain the lengths of bonds involving hydrogen atoms. Nonbonded interactions were calculated as described by Brooks et al.[47] with a cutoff for the nonbonded list generation of 20 Å, a cutoff for nonbonded interactions of 18 Å, and an onset of the switching function for nonbonded interactions of 16 Å. Other parameters used for REM are the default ones included in the MMTSB toolset.[48]

As mentioned in the introduction section, REM is a useful technique that has been applied to a wide variety of systems.[54] In general, M replicas of the system are simulated at M temperatures, and configurations are periodically exchanged in accordance with the Metropolis criterion. In this way, it is possible to sample large portions of phase space at high temperatures, producing structural changes than can favor visiting more constrained regions of phase space at low temperatures in an efficient manner. The configurations obtained at the lower temperatures of interest therefore reflect an enhanced sampling that would be very difficult to achieve using standard MD, even with a simulation time orders of magnitude longer. The All Pairs Exchange[45] variation enhances efficiency by redefining the generation probability in such a way that all possible replica pairs become candidates for exchange.

Various preliminary validation runs performed using CHARMM22-CMAP[55] with GBSW implicit solvent[56] and CHARMM19[57] with either GBMVA[58] or EEF[59] implicit solvent models are reviewed in the Supporting Information. It is noted that a solvent accessible surface area (SASA) calculation is included in the CHARMM GBMVA module. The results of these simulations favored the implementation of REM via the CHARMM19 force field with GBMVA. Interestingly, Olson et al.[60] have recently observed that CHARMM19 may in certain cases be more appropriate for loop modeling with implicit solvent than newer force fields such as CHARMM22. All of the REM simulations discussed in this work consisted of 12 replicas spanning a temperature range of 300−900 K. Swaps between temperatures were attempted every 500 MD steps, and configurations were stored with the same frequency. Individual temperatures were maintained using the Nose-Hoover thermostat. Two approaches were tested for obtaining REM temperatures that enhance equilibration of the entire system. In the first method, temperatures were chosen such that an acceptance ratio of approximately 30% was achieved for swap moves.[48,61,62] The second approach was that proposed by Trebst et al.,[63] which
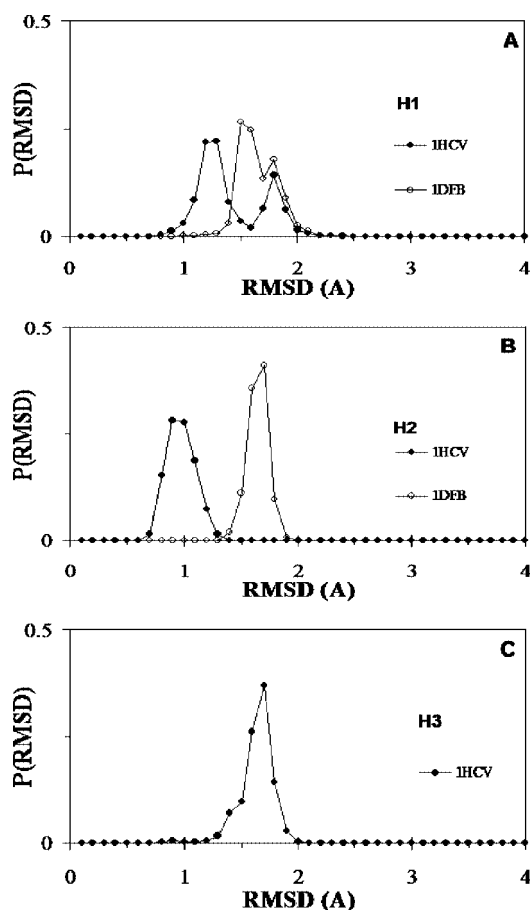
Figure 2. Probabilities of occurrence for the rmsd values of the simulated wildtype llama VHH H1 (A) and H2 (B), and H3 (C) loops at 300 K relative to reference structure 1DFB and 1HCV. The distributions account for the data evaluation period of a 10 ns run.

aims at the maximization of the number of round trips of replicas moving between the lowest and highest temperatures. Since no substantial improvement was observed when using the latter approach, the simulations reported below correspond to those obtained via the first technique.

Results and Discussion

Wildtype VHH Hypervariable Regions. The structural variability of the loops for the wildtype system was assessed via a 10 ns REM simulation. The energy minimized model of 1HCV was used as the initial conformation as well as the reference for the calculation of loop rmsd. Its rmsd values from the crystal H1, H2, and H3 loops are 0.91, 0.29, and 1.12 Å, respectively. The first 5 ns was considered to be an equilibration period; the results reported below correspond to the last 5 ns of the run, referred to as the data evaluation period.

Figure 2A shows the rmsd distribution for the simulated H1 loop at 300 K relative to 1DFB (H1 type 1) and 1HCV. Sampling of the crystal structure is evident with 35% of the configurations falling below 1.2 Å from the 1HCV H1 loop. Notably, the Y32 side chain of the low rmsd conformers lies flat under the loop, positioned analogously to that observed in the crystal structure. In contrast, the subensemble of structures corresponding to rmsd values of around 1.8 Å have their Y32 side chain exposed to the solvent, oriented toward the H3 loop. These higher rmsd values may further be caused by a kink observed in the backbone of residue 28, which is stabilized by an H-bond (absent in the crystal structure) between the side chains of T28 and Q3. Overall, the structures obtained lie within

TABLE 2: Percentage of Loop Conformations with rmsd Values Lower than 1.2 Å with Respect to Reference Structures for the Data Evaluation Period of Conventional REM Simulations

case	H1 (ref 1DFB)	H2 (ref 1FVC)
wildtype	0%	85%
1-Fa	10%	65%
1-Fb	11%	95%
2-FF	70%	96%
2-FS	6%	99%
2-FL	15%	89%
3-FFSa	99%	97%
3-FFSb	29%	86%

TABLE 3: Average of the rmsd Values with Respect to the Typical Structure for the Data Evaluation Period of Conventional REM Simulations

case	H1 (Å)	H2 (Å)	H3 (Å)	average for the three loops (Å)
wildtype	1.12	0.45	0.59	0.72
1-Fa	1.21	0.51	0.49	0.74
1-Fb	0.67	0.57	0.47	0.57
2-FF	0.51	0.58	0.49	0.53
2-FS	0.56	0.47	0.50	0.51
2-FL	1.05	0.49	0.53	0.69
3-FFSa	0.55	0.47	0.52	0.51
3-FFSb	1.29	0.58	0.73	0.87

a range of 0.5−2.5 Å from 1HCV, supporting the variability of this loop observed in the NMR study. None of the conformations achieve H1 rmsd values below 1.2 Å from a type 1 loop, supporting the noncanonical nature of this loop.

Figure 2B shows the rmsd probabilities for the H2 loop. The simulated conformations adopted by this loop are compared to the canonical type 3 (1DFB) and to the 1HCV loop (type 2A). None have rmsd values that fall below 1.2 Å from a type 3 loop, while 99% have rmsd values that are under this threshold when compared to the 1HCV loop. All of the H2 conformations display the N52-R56 H-bond observed in type 2A loops.

As evidenced from Figure 2C, the H3 loop structures obtained by simulation display moderate structural variation, with structures between 0.7−2 Å with respect to the 1HCV structure. These conformations resemble the kinked shaped NMR and crystal H3 loops and have two corresponding E96/T99 intraloop backbone-backbone H-bonds (not shown). However, an increased flexibility is observed for these conformers when compared to the NMR structures, perhaps due to model deficiencies in solvent mediated stabilizing interactions within the H3 loop.

Mutational Analysis. In an effort to identify key mutations of the llama VHH that lead to an H1 loop with a type 1 conformation, different mutants containing anywhere from 1 to 3 point mutations were simulated for 10 ns using conventional REM (see Table 1). The initial structures from which the mutants were simulated have rmsd values ranging from 0.93−1.05, 0.41−0.47, and 0.87−0.99 Å from the H1, H2, and H3 loops, respectively, of 1HCV. With respect to their corresponding canonicals, these initial structures have rmsd values of 1.63−1.78 Å from 1DFB for the H1 loop, and 0.39−0.49 Å from 1FVC for the H2 loop. Each mutant was simulated for 5 ns (equilibration period), and then assayed at 300 K during the following 5 ns (data evaluation period) to detect signs of a stable type 1 H1 structure. Two markers used for this purpose were (i) the percentage of configurations with H1 rmsd values lower
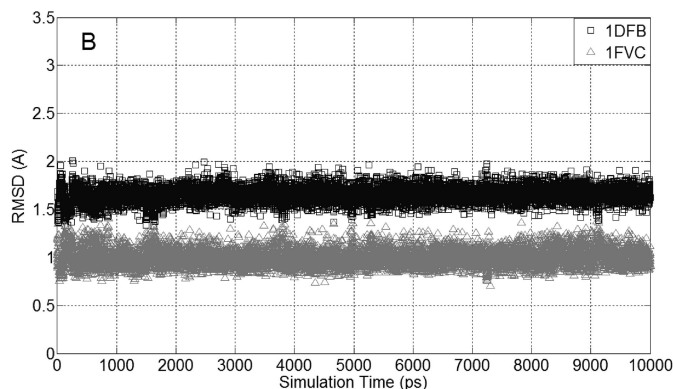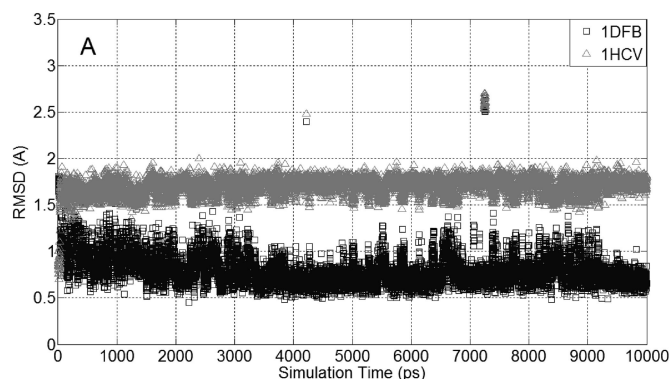
Hypervariable Loops of a Llama VHH Domain

J. Phys. Chem. B, Vol. 113, No. 6, 2009 **1789**

Figure 3. Rmsd values between the conformations of the simulated H1 loop and 1DFB, 1HCV (A) and the simulated H2 loop and 1DFB, 1FVC (B) of a 10 ns run for the 3-FFSa case.
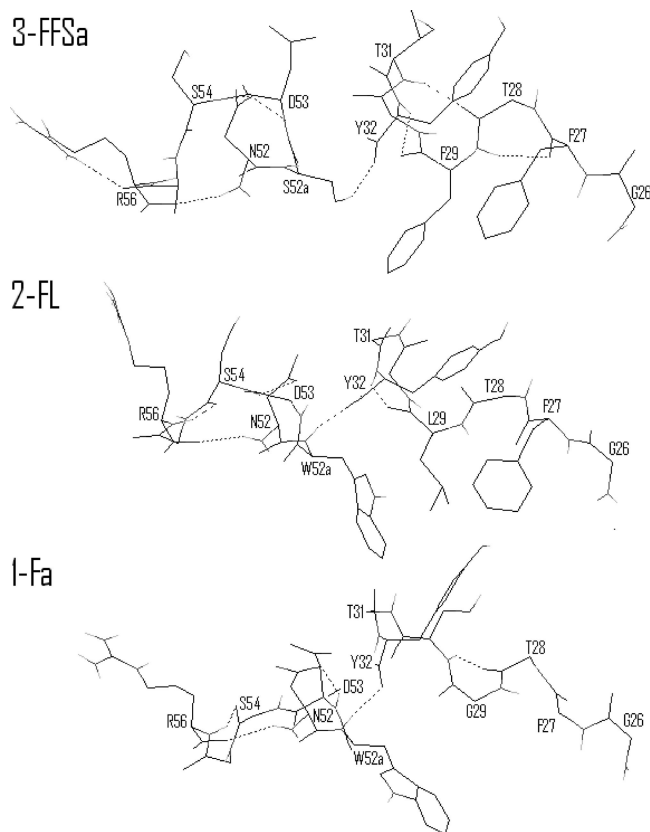

Figure 4. Simulated H1 (residues 26−32) and H2 (residues 52, 52a−56) loops from the low rmsd structures of mutants 3-FFSa, 2-FL, and 1-Fa, showing the H-bonding pattern for each case. Only those side chains that are required to display the H-bonds are illustrated.

Figure 5. Simulated loops of a representative 3-FFSa structure, as compared to the initial 3-FFSa conformation used for the simulation and the reference structures 1DFB (H1 loop), 1FVC (H2 loop), and 1HCV (H3 loop). H1 side chains 27, 29, and 32 are also shown.

than 1.2 Å with respect to 1DFB (Table 2), and (ii) the average rmsd of each of the three loops with respect to the "typical" simulated structure, which corresponds to the structure that is closest in rmsd to the mean rmsd of the simulated H1 conformations from the canonical 1DFB loop (Table 3). It was verified that each such typical structure was consistent with the structures from the most frequent ensemble observed for each mutant. It is noted that even though the experimentally observed canonical loops are expected to be highly stable, noncanonical stable structures may also be obtained through mutagenesis. Thus, the second marker is used here as an appropriate indicator of the structural stability of each mutant.

Tables 2 and 3 show that the 3-FFSa conformers adopt a stable H1 type 1 structure. Ninety-nine percent of the H1
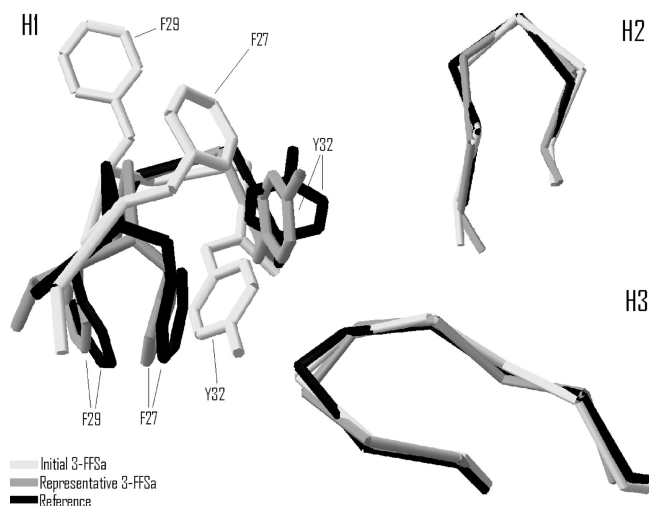
configurations have a rmsd value lower than 1.2 Å from 1DFB (Figure 3A), and a 0.55 Å average rmsd from the typical structure is obtained. Conversely, for the 2-FF case, despite having two key mutations at positions 27 and 29 thought to be sufficient to reshape the noncanonical H1 loop back to a canonical type 1 structure,[1,11] only 70% of the configurations are below the selected threshold. The results for these two cases highlight the importance of interloop interactions on the stabilization of the simulated H1 type 1 conformation (plots of the rmsd distributions at 300 K for 2-FF and 3-FFSa are given in Figure S1 of the Supporting Information). Likewise, the 2-FL and 1-Fa mutants achieve a modest number of conformations with low values of rmsd from 1DFB. Visual inspection of the low rmsd configurations for the 3-FFSa, 2-FL, and 1-Fa mutants confirms that type 1 H1 structures similar to the ones observed experimentally are obtained (Figure 4). For 3-FFSa, three H1 intraloop H-bonds are detected at the lowest temperature and coincide with those present in the H1 loop of the 1DFB crystal structure. However, this is in clear contrast with the H1 wildtype crystal structure, for which a single interloop H-bond is observed between G29 and Y32. The flexible nature of the wildtype conformer as compared to 3-FFSa is supported by the H-bonding pattern observed in the wildtype NMR and simulated structures. In the NMR structure ensemble, half of the structures have no H1 interloop H-bonds, whereas our simulations show no intraloop H-bonds for the majority of the conformers. For
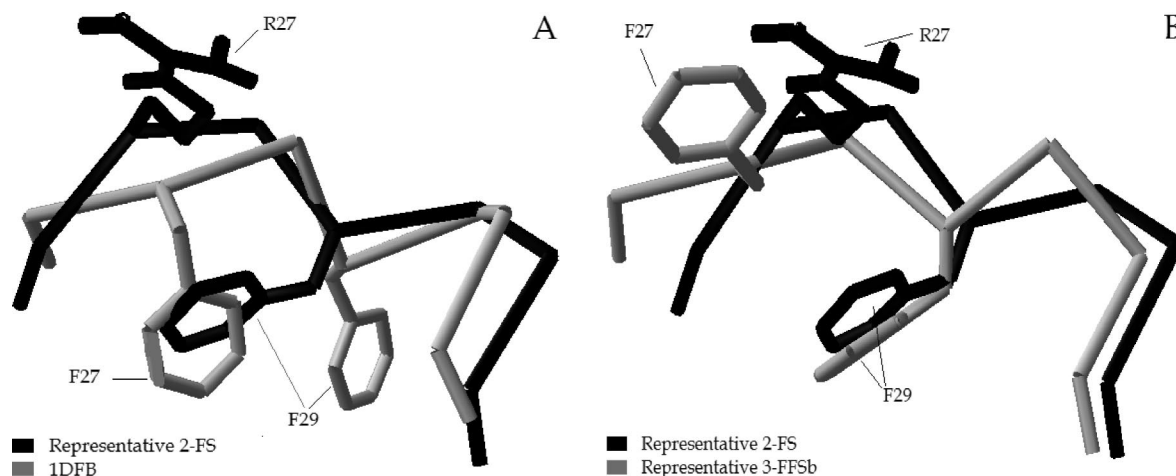
Figure 6. Representative structure of the simulated H1 loop of mutant 2-FS as compared to 1DFB H1 loop (A) and a low rmsd H1 loop of mutant 3-FFSb (B). Only relevant side chains are illustrated.
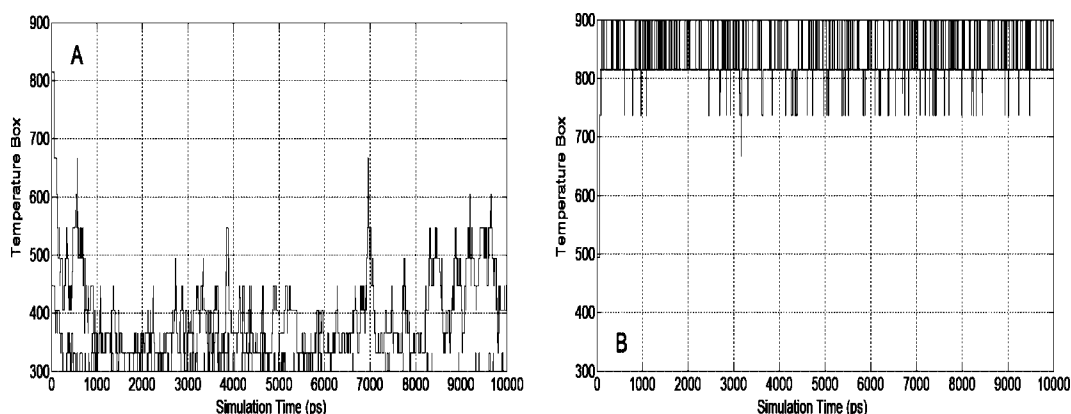

Figure 7. Walk over temperature space for the data evaluation period of a 6 mutant MMREM run. The two replicas for mutants 3-FFSa (A) and 3-FFSb (B) are displayed in each plot.

3-FFSa, there is also a well-conserved interloop H-bond between Y32 and the residue at position 52a, which is observed during the data evaluation period not only in the type 1 H1 structures of this mutant, but also in those attained by the other mutants that display this canonical type (see Figure 4). This specific interloop bond may be relevant for cooperative stabilization of H1 type 1 and H2 type 2A loops, given that it has been observed in some of the experimental structures that have these two canonical types (e.g., PDB structure 1TET).

Figure 5 shows a visual comparison for the three loops of a representative 3-FFSa stable conformer found by simulation, the initial 3-FFSa conformation of the simulation, and the corresponding reference structures. The H1 loop side chains of F27, F29, and Y32 have rearranged to positions that are almost coincident with those of the equivalent 1DFB residues. While the side chains of F27 and F29 have been buried in the interior of the loop, that of Y32 remains at the surface of the domain with its aromatic ring slightly shielded by loops H1 and H3 and its hydroxyl group pointing outward from the domain. No structures were found in which the Y32 side chain lies flat below the loop (as seen in the wildtype case), as the burial of side chains F27 and F29 precludes this from happening. This displacement of the Y32 side chain was also observed in all the low rmsd structures of other mutant cases, indicating that the burial of any hydrophobic side chain at position 27 or 29 likely has the same effect on Y32.

Altogether, these observations corroborate that a stable type 1 H1 structure for the 3-FFSa mutant was obtained with no noticeable perturbation of the H2 and H3 structures.

Considering the H1 loops of the remaining mutants, two particularly interesting cases are 1-Fb and 2-FS which, despite showing a reduced number of configurations with low H1 rmsd values from 1DFB, display a relatively stable behavior (see Table 3). A representative structure of the unconventional H1 loop displayed by these two mutants is shown in Figure 6A. The F29 side chain of the representative 2-FS structure is driven away from W52a (not shown) and toward the canonical location of F27. A similar position for the F29 side chain is observed in the relatively large number of low H1 rmsd conformers of 3-FFSb (29%), but in this case the F27 side chain is forced out toward the solvent (Figure 6B). Conformations with a nonburied F27 side chain are not likely to be energetically favorable since experimental structures show that this side chain is consistently buried in canonical H1 loops. The fact that 3-FFSb is found to be the most unstable mutant (see Table 3) supports the notion that its unfavorable F27 side chain positioning may lead to instability. Such mutants which cannot bury surface hydrophobic side chains may reduce VHH domain solubility (possibly promoting the aggregation of surface residues).

A plausible mechanism for the H1 noncanonical to canonical type 1 transition of the successful 3-FFSa was gathered from visual inspection of the trajectories of simulated conformations corresponding to 300 K (data not shown). Initially, two rapid
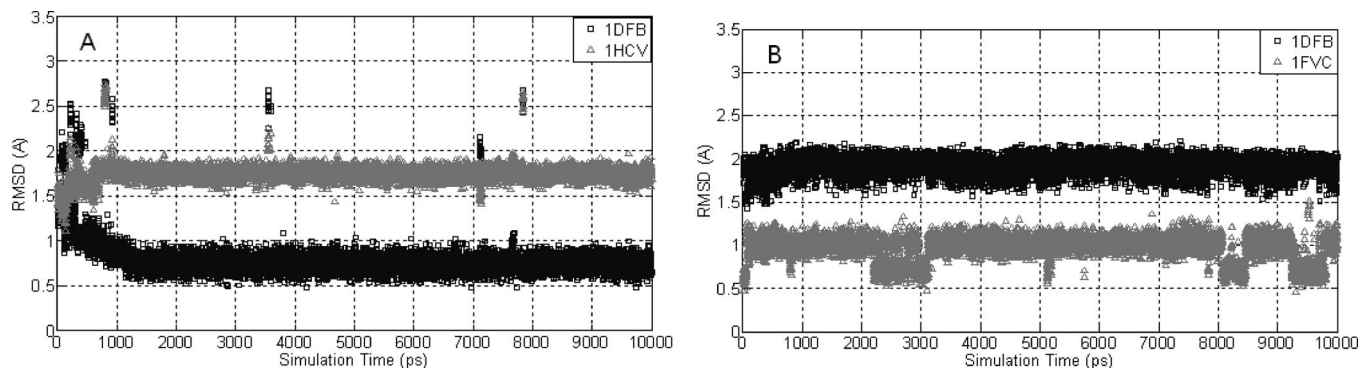
Hypervariable Loops of a Llama VHH Domain

J. Phys. Chem. B, Vol. 113, No. 6, 2009 **1791**

Figure 8. Rmsd values between the conformations of the simulated H1 loop and 1DFB, 1HCV (A) and the simulated H2 loop and 1DFB, 1FVC (B), of a 10 ns run for the 3-FFD case.

TABLE 4: Initial Temperature Distribution and Average between the Most Visited Temperatures for the Two Replicas of Each Mutant for the Data Evaluation Period of a 10 ns 6 Mutant MMREM Run

case	initial T for each replica (K)	average between most visited T (K)
1-Fa	300, 546	702
2-FF	366, 667	468
2-FS	447, 737	450
2-FL	332, 604	497
3-FFSa	405, 815	316
3-FFSb	494, 900	858

TABLE 5: Initial Temperature Distribution and Most Visited Temperature for Each Mutant for the Data Evaluation Period of a 10 ns 12-Mutant MMREM Run

case	initial T for each replica (K)	most visited T (K)
1-Fa	300	815
1-Fb	332	332
2-FF	366	604
2-FS	405	494
2-FL	447	546
2-FI	494	546
3-FFSa	546	366
3-FFSb	604	900
3-FFP	667	667
3-FFD	737	300
3-FLS	815	405
4-FFSS	900	737

and apparently cooperative changes take place: (i) the Y32 side chain is driven out of the interior of the H1 loop and is exposed to the solvent; (ii) the F29 side chain buries within the loop at a position analogous to that observed for the low rmsd 1-Fb, 2-FS, and 3-FFSb conformations (see 2-FS in Figure 6A). These rapid transitions are later followed by two slow events: (i) burial of the F27 side chain at its canonical position, displacing F29 from the location observed for 2-FS in Figure 6A to its canonical position (1DFB in Figure 6A); (ii) a shift of Y32 toward the H2 loop for subsequent formation of the stabilizing interloop H-bond with the residue at position 56.

The conformational behavior of H2 upon mutation is summarized in Tables 2 and 3. For the mutants studied, the type 2A conformation observed in the wildtype crystal structure is conserved. Upon inspection of the H2 loop of 3-FFSa (Figure 3B), it is apparent that this loop closely maintains its initial type 2A structure throughout the entire run. Figure 4 shows the characteristic H-bonding pattern of this mutant. The backbone–backbone H-bond between residues 52 and 56 is conserved in all H2 conformations examined. Table 3 shows average H2 rmsd values with respect to the typical structure ranging from 0.36−0.58 Å. A high degree of agreement between the simulated and reference structures of the H2 loop is also evident in Figure 5. It is worth noting that none of the cases studied show any evidence of type 3 structures.

Regarding the H3 loop conformations, the rmsd values from 1HCV for the mutants studied are similar to those observed for the wildtype and the structures conserve a kinked shape. In general, both markers suggest that the influence of the mutations on the H3 loop structure is not significant. With respect to the H3 of 3-FFSa, a structure that closely resembles that of the simulated wildtype H3 loop (Figure 2C) is observed (Figure 5). Accordingly, these mutant H3 loops preserve their kinked conformation and have the well-conserved pair of H-bonds between residues E96 and T99. However, it is noted that the

H3 loop of 3-FFSb was observed to be somewhat more variable than that of the other mutants (see Table 3) and the wildtype.

Overall, the simulated mutations appear to significantly affect H1 loop stability with average H1 rmsd values ranging from 0.51−1.29 Å from the typical structure. However, the overall stability of each mutant appears to be mainly, but not exclusively, due to the behavior of the H1 loop. Thus, the average variation of the three loops was used as an approximate measure of system stability. It can be seen in Table 3 that the highest variation is observed for 3-FFSb and 1-Fa, whereas 3-FFSa and 2-FS show the lowest deviation. The 2-FS H1 loop remains in a highly stable noncanonical conformation (see Figure 6A), making it attractive for further study.

Candidate Mutant Screening via Multiple Mutant REM. It is known that the large number of degrees of freedom in an antigen binding site hampers accurate prediction of the key interactions that lead to the stabilization of a given target structure. In the absence of detailed knowledge of a reliable stabilization mechanism, site-directed mutagenesis methods like the one implemented in this study can, along with guidance from antibody databases, greatly aid the identification of candidate mutations. Furthermore, tools that facilitate the effective selection of potentially useful mutants for subsequent studies can substantially reduce the number of mutants to be analyzed. Thus, a quick screening method is very desirable. In this context, an alternative application of REM that leads to a fast evaluation of the relative stability among various mutants is described below.

In REM, the recursive exchanges of different configurations (replicas) of a given mutant drive the lower energy replicas to
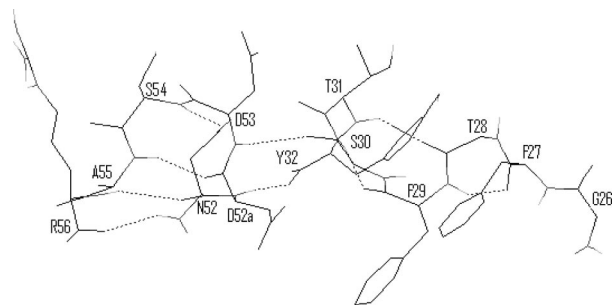
Figure 9. Simulated H1 (residues 26–32) and H2 (residues 52, 52a-56) loops of a representative 3-FFD structure, showing its H-bonding pattern.

the lower temperature boxes. An analogous behavior can be expected from a single REM run in which different mutants are placed at one or more temperature boxes. In this way, the overall system minimizes its free energy by preferentially placing the lower energy mutants in the lower temperature boxes, and the higher energy mutants in the higher temperature boxes. The replica swaps should still enable enhanced configurational sampling to allow those candidates whose mutations can potentially drive the loops back to energetically favorable conformations, to achieve such low-energy stable states faster and to successfully compete for the low temperature boxes. The net outcome from the application of this method, to be denoted as multiple mutant REM or MMREM, is a quick sorting of the relative stability of the multiple mutants simulated. For the systems studied in this work, MMREM was validated by comparing the mean temperatures of residence of competing replicas with results for structural stability determined from REM simulations. Two variations of this approach, a rigorous and a simplified version, which can be readily implemented in various scenarios, are discussed in detail in the Appendix.

The simplified version of MMREM was first applied to a 12-replica system with six of the mutants given in Table 1 with 1-Fb being excluded. Two replicas of each mutant were initially distributed throughout the 300–900 K range, as indicated in Table 4. Figure 7 illustrates the resulting walk over temperature space for both replicas of two representative mutants, 3-FFSa and 3-FFSb, during a 10 ns run. As observed, the relative energies of the mutants encourage replicas to compete for temperature boxes, with the more stable ones (i.e., the ones that can more readily sample low energies) tending to remain at the lower temperatures. This trend is captured by the average temperatures visited by replicas during the data evaluation period (see Table 4). The results are in good agreement with findings from the conventional REM simulations. For example, the relative average loop variability (Table 3) coincides with the positioning of the mutants according to their most frequent temperature of residence using MMREM. Furthermore, the relative location of the mutants whose low H1 rmsd conformations displayed a type 1 structure upon visual inspection, concurs with their frequency of adopting a type 1 H1 structure in REM (3-FFSa, 2-FF, 2-FL, and 1-Fa, in this order).

To further assess the MMREM approach, a system with 12 replicas each with a different mutant was simulated for a period of 10 ns. In addition to the cases used in the 6 mutant run described above, the mutants considered were 1-Fb, 2-FI (R27F,G29I), 3-FFD (R27F,G29F,W52aD), 3-FFP (R27F,G29F, W52aP), 3-FLS (R27F,G29L,W52aS), and 4-FFSS (R27F, G29F, W52aS, N52S). Table 5 shows the initial temperatures and most visited temperature for each mutant. As before, 3-FFSb

and 1-Fa locate themselves at the highest temperatures, now joined by 4-FFSS. The last mutant is unable to achieve low energy states, despite having the three stabilizing mutations R27F, G29F, and W52aS. On the other hand, the most stable cases found via conventional REM, namely 3-FFSa and 2-FS, have the third and fourth lowest temperatures of residence, respectively, preceded by mutants 3-FFD and 1-Fb. The latter case showed high stability during the REM runs, while the former is detailed below. Significantly, the two groups of mutants that are believed to have a similar behavior given the nature and position of their mutations, 3-FFSa/3-FLS and 2-FF/2-FI/2-FL, locate themselves at contiguous boxes.

A novel finding from the 12-MMREM simulation is that the replicas from the newly introduced 3-FFD mutant are found to maintain the lowest energies. Thus, in an attempt to validate the MMREM result, a 10 ns conventional REM run was conducted for this mutant. Figure 8 illustrates the rmsd values for the H1 and H2 loops with respect to their corresponding reference structures. The 3-FFD mutant H1 loop (Figure 8A) is seen to reach a stable H1 type 1 structure even faster than the successful 3-FFSa case (see Figure 3A), with 99% of the configurations having a rmsd value below 1.2 Å with respect to 1DFB and an average rmsd of the three loops from the typical structure of 0.49 Å (see Table 3 for comparison) for the data evaluation period. Figure 9 shows a visual inspection of the H1 and H2 loops from a representative 3-FFD structure with low rmsd values, highlighting their corresponding interactions. The structure, side chain location, and H-bonding pattern of loops H1 and H2 closely resemble those observed for low energy 3-FFSa structures (see Figure 4). In addition, 98% of the H2 loop structures have rmsd values below 1.2 Å with respect to 1FVC (Figure 8B), conformations that are stabilized by three intraloop backbone-backbone H-bonds. As in the 3-FFSa case, the 3-FFD H3 structures resemble those observed for the wildtype H3 loop. The increased stability of the 3-FFD mutant may be due to a slight structural change in the orientation of D53 encouraged by the extension of the D52a side chain toward the solvent, promoting a strong interaction between D53 and S30 that keeps the H1 loop conformations in place.

It is pointed out that the MMREM results reported here should be interpreted with care given that (i) the simulation periods were relatively short, (ii) only the simplified version of the method was implemented, and (iii) the simulation parameters were not optimized. For example, there is likely an optimal ratio of replicas to mutants that leads to faster equilibration (too many mutants may hamper a broad enough exploration of the temperature space for a given mutant to achieve ergodic sampling). The simplified MMREM is also likely to be most effective when mutants differ only by a few point mutations and when the system is sufficiently constrained so that a unique structure can be associated with the most energetically favorable conformation achievable. In this way, all mutants could be seen as competing for the same structure and where the "winner" is the one that best stabilizes that structure at lower temperatures. The method could be further refined to make it more specific in targeting a desired structure, for example, by introducing in addition to temperature, another "tempering" REM parameter that can more directly capture deviations from the target structure (like the rmsd values of the positions or angles of key residue atoms).

Final Remarks

The set of VHH antibody systems simulated in the present study constitute a challenging application for available simula-

Hypervariable Loops of a Llama VHH Domain

J. Phys. Chem. B, Vol. 113, No. 6, 2009 **1793**

tion protocols and force fields. The three loops of the wildtype and mutant systems display considerable variation in hydrophilicity and structural flexibility. Accordingly, we found that adequate study of their conformational equilibria demands the use of advanced methods and a careful examination of force fields. For the applications presented, we observed that REM in combination with the CHARMM19-GBMVA force field is capable of generating realistic structural ensembles from 10 ns simulations.

One of our primary goals was to use this combination of methods and force field to identify specific residues that reshape the H1 loop of the llama VHH domain 1HCV into a stable type 1 canonical conformation. Furthermore, by studying the structural changes of the three hypervariable loops of the wildtype and mutant structures, insights were obtained into the possible reasons underlying the conformational diversity of each particular system. A novel formulation of REM was also developed for screening multiple mutants based on their relative proclivity to adopt energetically favorable structures with associated reduced configurational dispersion. In general, the study of loop dynamics by simulation significantly aids in their structural typification and complements the information obtained from NMR and crystal analyses. In addition, the systematic approach implemented here may be useful in antibody engineering applications that require grafting loops having specific functionalities onto the framework regions of VHH domains.

The type 1 canonical structure of H1 is vastly conserved in human and mouse antibodies, implying that its backbone structural invariance has functional significance. In this work, it was observed that a type 1 H1 structure has increased stability over other H1 loop conformations, consistent with its high occurrence in databases of antibody crystal structures. Two cases with a highly stable type-1 H1 loop structure were obtained from the mutant simulations, each having three point mutations, namely 3-FFSa and 3-FFD. These results confirm the marked influence of highly hydrophobic residues (e.g., phenylalanine) at positions 27 and 29, whose stable side chain positioning drives the H1 backbone to a type 1 structure with a conserved H-bonding pattern. This appears to occur, however, only if clashing of residue W52a with residue 29 is avoided (in these two cases, the smaller residues S and D prevent clashing with F29). Mutants with one (1-Fa) and two (2-FF and 2-FL) mutations were also able to achieve the target H1 conformation, but with a moderate to low occurrence; such cases may be useful for further analysis in applications in which an increased number of mutations in a therapeutic antibody is detrimental to its immunocompatibility. It is worth noting that the achievement of an H1 type 1 conformation was observed to be dependent upon mutations in both the H1 and H2 loops, which indicates that interloop interactions may be relevant for the prediction of canonical conformations, at least for particular cases such as the ones studied in this work.

Overall, the average system stabilities measured in REM simulations are consistent with the average temperatures sampled by replicas in MMREM simulations. Furthermore, cases with successful mutations (i.e., that achieve high conversion to a type 1 H1 structure) have increased loop stability. Conversely, some of the tested mutants that show decreased stability display an H1 loop variability comparable to that of the wildtype fragment. Interestingly, two highly stable cases with a coincident noncanonical H1 loop were found; unique structures such as these may contribute to the design of novel binding domains.

In addition to the large binding repertoire observed for VHHs in camelids due to sequence variability, each loop possesses internal motion that promotes the sampling of distinct backbone conformations in some antibodies. In particular, our simulations of the llama anti-hCG VHH fragment suggest that its H1 loop has considerable backbone flexibility, in agreement with results of other studies.[22,43,44] Such flexibility presumably has functional significance, perhaps promoting induced-fit binding to hCG that results in a moderate affinity of $K_d = 300$ nM.[44]

In general, a lock and key recognition mechanism will be promoted by stable loops, leading to improved affinity for some antigens.[64] Given that the MMREM method is oriented toward evaluating the relative stabilities of different mutants, it may play an important role within a rational antibody design protocol aimed at engineering loops that bind antigens based on such a mechanism. Through the identification of stable hypervariable loops, the method may also be useful in therapeutic applications[65,66] for reducing antibody immunogenicity of otherwise variable noncanonical loops that may be able to bind an increased number of targets. In this way, experimental procedures such as VHH loop grafting of highly stable structures could be more effectively realized.

Ongoing efforts are focused on increasing the efficiency of REM and MMREM simulations toward explicit-solvent mutagenesis analyses of complex structures to more accurately describe water mediated interactions (known to be of crucial importance in some systems).

Appendix

Consider a certain mutant for which conventional MD REM[46] is used. Such system is studied within a generalized ensemble formed by M replicas of the same mutant, each simulated at a different temperature T. Since the replicas are noninteracting, the weight factor W_{REM} for state X is given by the product of the Boltzmann factors for each replica.

$$W_{REM}(X) = \exp\left\{ -\sum_{i=1}^{M} \beta_{m(i)} H(q^{[i]}, p^{[i]}) \right\} \quad (1)$$

In eq 1, H is the Hamiltonian, q and p are, respectively, the set of coordinates and momenta for the atoms in replica i, and β_m is the inverse temperature of replica i, which has a one-to-one correspondence with temperature m. Periodic exchanges between replicas i and j at T_m and T_n, respectively, that is

$$X = (..., x_m^{[i]}, ..., x_n^{[j]}, ...) \rightarrow X' = (..., x_m^{[j]'}, ..., x_n^{[i]'}, ...) \quad (2)$$

are performed to facilitate convergence toward an equilibrium distribution. For this purpose, the detailed balance condition is imposed on the transition probability $w(X \rightarrow X')$,

$$W_{REM}(X)w(X \rightarrow X') = W_{REM}(X')w(X' \rightarrow X) \quad (3)$$

The latter condition is satisfied by the Metropolis acceptance criterion

$$P_{acc}(X(..., x_m^{[i]}, ..., x_n^{[j]}, ...) \rightarrow$$

$$X'(..., x_m^{[j]'}, ..., x_n^{[i]'}, ...)) = \begin{cases} 1 & \text{for } \Delta \leq 0 \\ \exp(-\Delta) & \text{for } \Delta > 0 \end{cases} \quad (4)$$

$$\text{where } \Delta = (\beta_m - \beta_n) \times (U_j - U_i) \quad (5)$$

In eq 5, U_i and U_j are the potential energies of configurations i and j, with corresponding inverse temperatures β_m and β_n. The kinetic energy terms have been eliminated through velocity rescaling.

Consider now an exchange event between temperatures m and n of respective mutants A and B, using a criterion analogous to eq 4. For this multiple mutant replica exchange method (MMREM), detailed balance condition on the transition probability leads to

$$\frac{w(X \rightarrow X')}{w(X' \rightarrow X)} = \frac{w_A(q_A|\beta_n)w_B(q_B|\beta_m)}{w_A(q_A|\beta_m)w_B(q_B|\beta_n)} \quad (6)$$

where the probability weight for any given mutant is given by

$$w(q|\beta) = \exp(-\beta U(q))/Q(\beta, H) \quad (7)$$

In eq 7 Q, H, and U are the partition function, Hamiltonian, and energy of the system, respectively. Substituting the appropriate expressions of eq 7 into eq 6 and rearranging we have

$$\frac{w(X \rightarrow X')}{w(X' \rightarrow X)} = \frac{\exp[-(\beta_m - \beta_n)(U_B(q_B) - U_A(q_A))]}{\dfrac{Q(\beta_n, H_A)}{Q(\beta_m, H_A)} \dfrac{Q(\beta_m, H_B)}{Q(\beta_n, H_B)}} \quad (8)$$

Now, given that mutant A(B) (read A or B) is simulated in the canonical ensemble, its partition functions at T_m and T_n are related to the Helmholtz free energy **A** by the expressions:

$$\begin{aligned} \beta_m A_m^{A(B)} &= -\ln Q(\beta_m, H_{A(B)}) \\ \beta_n A_n^{A(B)} &= -\ln Q(\beta_n, H_{A(B)}) \end{aligned} \quad (9)$$

We can define a change in free energy $\Delta F^{A(B)}$ associated with a virtual temperature swap for mutant A(B) by subtracting the second equality of eq 9 from the first one

$$\Delta F^{A(B)} = F_n^{A(B)} - F_m^{A(B)} = \beta_n A_n^{A(B)} - \beta_m A_m^{A(B)} =$$

$$- \ln \frac{Q(\beta_n, H_{A(B)})}{Q(\beta_m, H_{A(B)})} \quad (10)$$

Equation 10 can then be introduced into eq 8 for both mutants A and B, and the result simplified to

$$\frac{w(X \rightarrow X')}{w(X' \rightarrow X)} = \exp[-(\beta_m - \beta_n)(U_B(q_B) - U_A(q_A)) +$$

$$(\Delta F^A - \Delta F^B)] \quad (11)$$

Equation 11 may be satisfied by the Metropolis acceptance criterion (eq 4), but with Δ now defined by

$$\Delta_{MMREM} = (\beta_m - \beta_n)(U_B(q_B) - U_A(q_A)) - (\Delta F^A - \Delta F^B) \quad (12)$$

If mutants A and B were identical, then $\Delta F^A - \Delta F^B = 0$ and Δ_{MMREM} would reduce to the Δ of the conventional REM (see eq 5). The excess free energy change (ΔF) associated with the temperature change for a given mutant can be evaluated by using the acceptance ratio method originally proposed by Bennett;[67] for example, applying the "unoptimized" version of the method to our system leads to

$$\Delta F^{A(B)} = -\ln\left[\frac{\langle P_{acc}^{A(B)}(\beta_m \rightarrow \beta_n)\rangle_{\beta_m}}{\langle P_{acc}^{A(B)}(\beta_n \rightarrow \beta_m)\rangle_{\beta_n}}\right] \quad (13)$$

In eq 13, the $\langle\rangle$ brackets denote ensemble averages taken at the subscripted β, and the acceptance probability (for the virtual temperature changes) can be found, for example, by application of Barker's rule[68] to obtain $P_{acc}^{A(B)} = e^{-(\beta_n - \beta_m)U_{A(B)}}/(1 + e^{-(\beta_n - \beta_m)U_{A(B)}})$. These $P_{acc}^{A(B)}$ values can be readily obtained (at no cost) from the data of the exchange attempts in the MMREM run. The $\Delta F^{A(B)}$ values would be initialized to zero at the beginning of the MMREM run, calculated on the fly, but updated only after a minimum of statistics have been accumulated. Note that, besides requiring no computational overhead, evaluation of $\Delta F^{A(B)}$ can be of interest independent of MMREM, for example, for thermodynamic analysis.

Rigorous execution of MMREM requires then the evaluation of the mutants' free energy changes (eq 13) for subsequent calculation of Δ_{MMREM}. This procedure, however, may be inconvenient to implement when using some of the available computational packages. An alternative, albeit approximate approach for analyses such as the one implemented in this study, exploits the fact that point mutations introduce relatively small perturbations to the basal energy of the system. If mutants A and B differ only by a small number of mutations, it is plausible to assume that $\Delta F^A - \Delta F^B$ in eq 12 will be small relative to the first term. Under these conditions, we can approximate $\Delta_{MMREM} \approx \Delta$ (with Δ as defined in eq 5); this variant was referred to as the "simplified" version in the main text. A quasi-rigorous version of the approach would entail separating the contributions to $U_{A(B)}$ into "common" interactions (among shared residues) and "mutant-specific" interactions and keeping constant the "temperature" of the latter; in such a case, ΔF^A would also approach ΔF^B as they would entail temperature changes of the common interactions only.

Ongoing research aims at evaluating the applicability of both MMREM schemes (rigorous and simplified versions) in varying scenarios.

Supporting Information Available: This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Al-Lazikani, B.; Lesk, A. M.; Chothia, C. *J. Mol. Biol.* **1997**, *273*, 927–948.

Hypervariable Loops of a Llama VHH Domain

J. Phys. Chem. B, Vol. 113, No. 6, 2009 **1795**

(2) Chothia, C.; Lesk, A. M. *J. Mol. Biol.* **1987**, *196*, 901–917.

(3) Chothia, C.; Lesk, A. M.; Gherardi, E.; Tomlinson, I. M.; Walter, G.; Marks, J. D.; Llewelyn, M. B.; Winter, G. *J. Mol. Biol.* **1992**, *227*, 799–817.

(4) Chothia, C.; Lesk, A. M.; Tramontano, A.; Levitt, M.; Smithgill, S. J.; Air, G.; Sheriff, S.; Padlan, E. A.; Davies, D.; Tulip, W. R.; Colman, P. M.; Spinelli, S.; Alzari, P. M.; Poljak, R. J. *Nature* **1989**, *342*, 877–883.

(5) Decanniere, K.; Muyldermans, S.; Wyns, L. *J. Mol. Biol.* **2000**, *300*, 83–91.

(6) Delapaz, P.; Sutton, B. J.; Darsley, M. J.; Rees, A. R. *EMBO J.* **1986**, *5*, 415–425.

(7) Ewert, S.; Honegger, A.; Pluckthun, A. *Methods* **2004**, *34*, 184–199.

(8) Johnson, G.; Wu, T. T. *Int. Immunol.* **1998**, *10*, 1801–1805.

(9) Mandal, C.; Kingery, B. D.; Anchin, J. M.; Subramaniam, S.; Linthicum, D. S. *Nat. Biotechnol.* **1996**, *14*, 323–328.

(10) Martin, A. C. R.; Cheetham, J. C.; Rees, A. R. *Methods Enzymol.* **1991**, *203*, 121–153.

(11) Martin, A. C. R.; Thornton, J. M. *J. Mol. Biol.* **1996**, *263*, 800–815.

(12) Monsellier, E.; Bedouelle, H. *J. Mol. Biol.* **2006**, *362*, 580–593.

(13) Morea, V.; Lesk, A. M.; Tramontano, A. *Methods* **2000**, *20*, 267–279.

(14) Morea, V.; Tramontano, A.; Rustici, M.; Chothia, C.; Lesk, A. M. *J. Mol. Biol.* **1998**, *275*, 269–294.

(15) Ramsland, P. A.; Guddat, L. W.; Edmundson, A. B.; Raison, R. L. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 453–461.

(16) Reczko, M.; Martin, A. C. R.; Bohr, H.; Suhai, S. *Protein Eng.* **1995**, *8*, 389–395.

(17) Schlessinger, A.; Ofran, Y.; Yachdav, G.; Rost, B. *Nucleic Acids Res.* **2006**, *34*, D777–D780.

(18) Shirai, H.; Kidera, A.; Nakamura, N. *FEBS Lett.* **1999**, *455*, 188–197.

(19) Smithgill, S. J.; Mainhart, C.; Lavoie, T. B.; Feldmann, R. J.; Drohan, W.; Brooks, B. R. *J. Mol. Biol.* **1987**, *194*, 713–724.

(20) Bruccoleri, R. E.; Haber, E.; Novotny, J. *Nature* **1988**, *335*, 564–568.

(21) Felts, A. K.; Gallicchio, E.; Chekmarev, D.; Paris, K. A.; Friesner, R. A.; Levy, R. M. *J. Chem. Theory Comput.* **2008**, *4*, 855–868.

(22) Fenwick, M. K.; Escobedo, F. A. *Biopolymers* **2003**, *68*, 160–177.

(23) Fine, R. M.; Wang, H.; Shenkin, D. L.; Yarmush, P. S.; Levinthal, C. *Proteins* **1986**, *1*, 342–362.

(24) Higo, J. I.; Collura, V.; Garnier, J. *Biopolymers* **1992**, *32*, 33–43.

(25) Hornak, V.; Simmerling, C. *Proteins Struct. Funct. Genet.* **2003**, *51*, 577–590.

(26) Kim, S. T.; Shirai, H.; Nakajima, N.; Higo, J.; Nakamura, H. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 683–696.

(27) Krol, M. *J. Comput. Chem.* **2003**, *24*, 531–546.

(28) Pellequer, J. L.; Chen, S. W. W. *Biophys. J.* **1997**, *73*, 2359–2375.

(29) Shenkin, P. S.; Yarmush, D. L.; Fine, R. M.; Wang, H. J.; Levinthal, C. *Biopolymers* **1987**, *26*, 2053–2085.

(30) Sinha, N.; Smith-Gill, S. J. *Cell Biochem. Biophys.* **2005**, *43*, 253–273.

(31) Tenette, C.; Ducancel, F.; Smith, J. C. *Proteins: Struct., Funct., Genet.* **1996**, *26*, 9–31.

(32) Voordijk, S.; Hansson, T.; Hilvert, D.; van Gunsteren, W. F. *J. Mol. Biol.* **2000**, *300*, 963–973.

(33) Zheng, Q.; Rosenfeld, R.; Delisi, C.; Kyle, D. J. *Protein Sci.* **1994**, *3*, 493–506.

(34) Martin, A. C. R.; Cheetham, J. C.; Rees, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9268–9272.

(35) Mas, M. T.; Smith, K. C.; Yarmush, D. L.; Aisaka, K.; Fine, R. M. *Proteins: Struct., Funct., Genet.* **1992**, *14*, 483–498.

(36) Webster, D. M.; Roberts, S.; Cheetham, J. C.; Griest, R.; Rees, A. R. *Int. J. Cancer* **1988**, *1*, 3–16.

(37) Fiser, A.; Do, R. K. G.; Sali, A. *Protein Sci.* **2000**, *9*, 1753–1773.

(38) Hamers-Casterman, C.; Atarhouch, T.; Muyldermans, S.; Robinson, G.; Hamers, C.; Songa, E. B.; Bendahman, N.; Hamers, R. *Nature* **1993**, *363*, 446–448.

(39) van der Linden, R. H. J.; Frenken, L. G. J.; de Geus, B.; Harmsen, M. M.; Ruuls, R. C.; Stok, W.; de Ron, L.; Wilson, S.; Davis, P.; Verrips, C. T. *Biochim. Biophys. Acta* **1999**, *1431*, 37–46.

(40) Perez, J. M. J.; Renisio, J. G.; Prompers, J. J.; van Platerink, C. J.; Cambillau, C.; Darbon, H.; Frenken, L. G. J. *Biochemistry (Moscow)* **2001**, *40*, 74–83.

(41) Lauwereys, M.; Ghahroudi, M. A.; Desmyter, A.; Kinne, J.; Holzer, W.; De Genst, E.; Wyns, L.; Muyldermans, S. *EMBO J.* **1998**, *17*, 3512–3520.

(42) Saerens, D.; Pellis, M.; Loris, R.; Pardon, E.; Dumoulin, M.; Matagne, A.; Wyns, L.; Muyldermans, S.; Conrath, K. *J. Mol. Biol.* **2005**, *352*, 597–607.

(43) Renisio, J. G.; Perez, J.; Czisch, M.; Guenneugues, M.; Bornet, O.; Frenken, L.; Cambillau, C.; Darbon, H. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 546–555.

(44) Spinelli, S.; Frenken, L.; Bourgeois, D.; deRon, L.; Bos, W.; Verrips, T.; Anguille, C.; Cambillau, C.; Tegoni, M. *Nat. Struct. Biol.* **1996**, *3*, 752–757.

(45) Brenner, P.; Sweet, C. R.; VonHandorf, D.; Izaguirre, J. A. *J. Chem. Phys.* **2007**, *126*, 074103.

(46) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(47) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(48) Feig, M.; Karanicolas, J.; Brooks, C. L., III. *MMTSB Tool Set, MMTSB NIH Research Resource*; The Scripps Research Institute: La Jolla, CA, 2001.

(49) He, X. M.; Ruker, F.; Casale, E.; Carter, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7154–7158.

(50) Eigenbrot, C.; Randal, M.; Presta, L.; Carter, P.; Kossiakoff, A. A. *J. Mol. Biol.* **1993**, *229*, 969–995.

(51) Nguyen, V. K.; Hamers, R.; Wyns, L.; Muyldermans, S. *EMBO J.* **2000**, *19*, 921–930.

(52) Tomlinson, I. M.; Walter, G.; Marks, J. D.; Llewelyn, M. B.; Winter, G. *J. Mol. Biol.* **1992**, *227*, 776–798.

(53) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(54) Earl, D. J.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910–3916.

(55) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(56) Im, W. P.; Lee, M. S.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1691–1702.

(57) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902–1921.

(58) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *J. Chem. Phys.* **2002**, *116*, 10606–10614.

(59) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133–152.

(60) Olson, M. A.; Feig, M.; Brooks, C. L., III *J. Comput. Chem.* **2008**, *29*, 820–831.

(61) Kone, A.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 206101.

(62) Yan, Q. L.; de Pablo, J. J. *J. Chem. Phys.* **1999**, *111*, 9509–9516.

(63) Trebst, S.; Troyer, M.; Hansmann, U. H. E. *J. Chem. Phys.* **2006**, *124*, 174903.

(64) Wedemayer, G. J.; Patten, P. A.; Wang, L. H.; Schultz, P. G.; Stevens, R. C. *Science* **1997**, *276*, 1665–1669.

(65) Helms, L. R.; Wetzel, R. *Protein Sci.* **1995**, *4*, 2073–2081.

(66) Lonberg, N. *Nat. Biotechnol.* **2005**, *23*, 1117–1125.

(67) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.

(68) Barker, A. A. *Aust. J. Phys.* **1965**, *18*, 119.